Creating a dashboard-based visualization tool to enhance data-based consultancy within the food safety sector

Jesse van Gameren

First Supervisor: Daniel Braun
Second Supervisor: Ieke Schrader
Company supervisor: Max van Leeuwen

University of Twente
Department of IEBIS
BSc Industrial Engineering and Management
20th of January 2025

Management Summary

Introduction

Eurofins FSS is a food testing and consultancy company that gives advice based on product testing to clients. The process of testing has three major steps: creating the testing plan, testing products based on the testing plan and consulting the client om basis of test results. When creating a testing plan, it is hard to quantify which time of food safety tests should and should not be included, as not much data is public about food testing data, and expertise within the sector cannot expect everything.

To solve this problem, Eurofins FSS wants to analyze and visualize their historic data, to get more knowledge on the food safety of different products and tests. Given that the test database of Eurofins is very big, the data can be of big help when it comes to analyzing patterns and changes within food safety.

To do this, an easy to use and visualization system needs to be built to following certain company criteria, resulting with the following question needing to be answered: *"How can Eurofins Food Safety Solutions implement a data visualizing system to give the best insight in their historic data to analyze current risk?"*

To collect all the pieces of data needed, first the important data needed to be identified. To select the important data that is implemented within the visualization tool. Managerial meetings were held as well as interviews to ask future application users and current employees about the data they are missing within the company and what data would be the most important to visualize. Secondly, through the process of managerial meetings, software was chosen in the form of Power Bi to make the visualization tool in, as this program is compatible with Eurofins FSS.

Data Preparation

After knowing what data is important to the company, the next step into creating and implementing a fitting data visualization system was analyzing the data that is present in the database. The data in Eurofins FSS' database was cleaned by deleting all errors out of their database and clearing all data that is not usable within this research. This results in Eurofins FSS having a clean research dataset for future purposes. For this research, a dataset was created from all existing data with proper the right formatting, and deleted errors causing variables that were still present within the dataset. Lastly, the data validity of the dataset of Eurofins was checked, resulting in a trustworthy and valid database of Eurofins FSS.

Modeling

To create the visualization tool, the first step was designing the tool. An analytical dashboard was made, due to the high usability of dashboards as Eurofins FSS is already familiar with the use of dashboards, together with the ease of use for new individuals when learning how to use a dashboard. The pillars of importance when looking at the design of the dashboard are both usability and clarity.

To ingrain the principle of clarity and usability within the dashboard, the dashboard is designed around familiarity, trying to use the same layout and options everywhere in the application. Using standardized values and colors results in a very easy to see and use application.

The back end of the application is designed to result in the least error within the application, together with ordering the data as clear as possible to make it easy to add new functionalities to the application. The semantic model in Power BI follows the rules of the third normal form in Normalization, with exceptions from the import entities, as they import all data into the application and cannot be normalized without losing data.

In total the application is fully designed for the results to be connected to the workflow of Eurofins FSS. Results found within the application are in the same format as the rapport of Eurofins FSS, to ensure easy usage of the application in the current workflow.

The created dashboard is gives consultants within Eurofins a new quantitative source of data to use to make decisions to create testing plans. The application gives the user information about the risk levels different products have, how product errors change over time and more crucial information that can be used to monitor the food safety within the Netherlands. The application results in Eurofins FSS being able to use internal data as a source of data for research and the creation of plans, a big step compared to the previously needed use of outside data sources. The application is a big step in data self-sufficiency of Eurofins FSS.

Implementation

The implementation of the application is the last step of successfully creating a new visualization system. For implementation, a few steps have been taken. First of all, the application has been fully transferred to the IT team of Eurofins FSS, together with a step plan and biweekly meetings to integrate the applications within Eurofins FSS' workflow.

Next to this new project members have been added to the project to follow up on the development of the application and testing the application in actual client related usage cases. Both these implementation steps are to be combined to create a good way to implement the application company wide.

To help with the implementation of the application, a user manual has been made to help new users that do not know certain Power Bi functionalities, as well as to help with certain functions that are forgotten over time.

Lastly, there are a few improvements documented to increase the value of the dashboard. The following improvements are the main focus to maximize value: Adding the ability to see what results are coming from what specific Eurofins research laboratories, the ability to compare Eurofins FSS data with European data in one visualization compared to multiple screens and the ability to make the dashboard anonymous to be able to show the data directly to the customer. Another point of improvement is making a standardized filter system within the database of Eurofins FSS, to have standardized usable data across Eurofins FSS.

Table of Contents

Chapter 1: Introduction

1.1 - Background

Eurofins is a testing and consultancy company that consults clients within almost every sector looking at products used in production or human consumption. Eurofins is an international company, active in sixty-two countries with over 62,000 members of staff, helping to perform 450 million test each year. (*About Eurofins Scientific*, n.d.) Within the region Benelux Eurofins has twelve business units, ranging from chemical testing to consumer food testing. The department of Food Testing has a sister company called Eurofins Food & Safety Solutions (Eurofins FSS), that test different sorts of food products meant for human consumption based on legal standards and benchmarks, sold within retail from retailers and wholesalers.

Currently, the process of consultancy and testing goes as follows:
First, the client company and Eurofins come in contact with each other, resulting in a deal for either short term or long-term testing. This results in the start of the consultancy and testing. Eurofins makes a testing plan for the company based partially on intuition, and partially on governmental testing data of error within a certain sector. This plan will be discussed together by both parties. After both parties agree, the testing will begin. The results of the testing will then be put in a rapport, which is sent to the client, following up with a consultation about the found results and how to move further based on the tests done. *A BPMN 2.0 model of the consultancy process can be found in section 1.4 and appendix F.*

With the Eurofins group being a fast-growing testing and consultancy company globally, being world leading in multiple sectors (*About Eurofins Scientific*, n.d.), Eurofins FSS uses inhouse consultant expertise based on experience as their main source of information for decision making. This experience-based information source for making decisions is hard to quantify, resulting Eurofins FSS sometimes getting questioned about their way of determining what tests are useful to do within a product group. Given that their way of choosing what test needs to be done is not transparent and the test amounts and sorts influence the costs of the client, some clients want more transparency about the way the tests get selected.

Eurofins FSS wants to better their market position in the upcoming two years by getting bigger and more clients, but also by increasing the quality of their consultancy process. The way Eurofins FSS want to improve their consultancy process is by changing from a non-transparent and more intuition-based consultancy plan model to a data-driven approach, having showable historical data that gives insight into the sector and its risks. This development in transparency and quality will be led by the development of a risk assessment tool, to determine the relevance of certain tests done on food, so that the testing plan shown to customers can be backed by data, together with being transparent for the client.
As of right now, Eurofins FSS is standing at the beginning of becoming a data-driven company, yet to start becoming more data based, the historical data collected over the last few years needs to be analyzed and standardized. The following step being the creation and implementation of the different data driven tools Eurofins FSS wants to use to increase their process of work.

1.2 - Problem Identification

1.2.1 - Action problem

The main focus of Eurofins FSS when it comes to improvement in the coming years is increasing the quality of the consultancy they give, by making the consultancy more data focused, resulting in an increase in reliability within their judgement of tests to be done on clients, and making the process of constructing the consulted testing plan more transparent. This all is done to try and solve the following action problem:

Eurofins Food Safety Solutions' testing and consultancy process is not transparent enough to facilitate the wanted market share.

1.2.2 - Problem cluster

After uncovering the action problem Eurofins FSS is trying to improve, internal research into what the root of the problem is began. When talking with both the operational manager and the director of Eurofins FSS, a collection of problems was uncovered within a few meetings. This process of uncovering problems within data usage was done in within a week of research, as the management within Eurofins FSS had already thought about what points of interest are related to the action problem, and what smaller problems contribute to the bigger problem. After, individual research was done to check all parts of the problem, to see if all important problems were mentioned, and nothing was left out.

In total, the following problem cluster could be organized (Figure 1.1). The action problem of not enough trust from clients and transparency for clients can be divided into two main sources, a low transparency of the creation of the testing plan given to the client, and the little amount of data the initial consultancy and the testing plan is based on. The consultancy process is originally based on little data, as the testing data points are not integrated in the analysis system of Eurofins FSS, leaving little data to show via the program used. Next to this, the data gathered cannot easily be used for consultancy as it is not ready to be shown to the client. A side factor in this was that the company had little time to work out the historical data, due to difficult deadlines that were set with the setup for the department. The short comings that result in having no data ready for use in consultancy, is that the quality of the database and its management is low, due to little updates being done and a low form of normalization and standardization resulting in hard to manage data. Lastly, data is not visualized for both internal and external use, resulting in missing clear interpretable data for consultants to use in consultancy tasks.

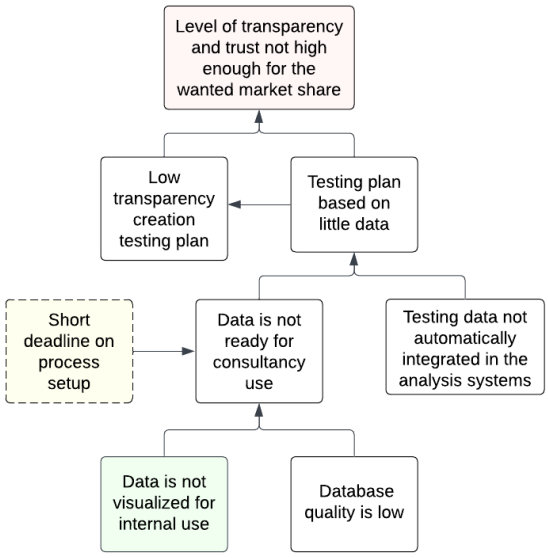


Figure 1.1: *problem cluster analysis trustworthiness and transparency of consultation process Eurofins FSS.*

1.2.3 - Core problem

The problem cluster stated before shows that there are multiple source problems that lead to the action problem that results in stagnation within the company branch growing. When looking at the origin of the action problem, the core exists mostly out of problems that are present due to neglect of the existing data organization, as well as not having put a lot of effort into digitalizing the consultancy process.

Looking at the visualization of data for internal use, the value of direct improvements comes in both efficiency and quality of work for Eurofins FSS employees. With improved data visualization in the form of ease to use visuals, table and documents, Eurofins FSS employees can more easily understand data that is present within the company without needing a fast background in data research, as well as be able to use the data visualized to set up and prepare the consultancy better, being able to transform and use the data within consultancy.
Eurofins FSS problem owners want to tackle this problem, due to the impact improvements can have, as well as the ability to work further on this type of research. This is because even at a barebones state, the visualization of data can help drastically with showing customers at a later stage the need and value of the product Eurofins FSS can provide. The value that solving this core problem creates is higher than solving another core problem from the company, as well as this core problem resulting in a directly usable product compared to the increase in level of database quality resulting in more direct value from a one time project.

Eurofins FSS is as of right now looking at increasing the transparency and quality of their product via visualization enough to reach their set benchmarks. This results with the following core problem to solve to get the best for the company:
The data present within Eurofins FSS' database is not visualized to a level of satisfaction for use within Eurofins Food Safety Solution's consultancy process resulting in a insufficient level of reliability and transparency within consulting.

1.3 - Problem-solving Approach

The following problem-solving approach will be split into two parts, research design and research limitations. Research design will be about the way the total research is designed looking at a work and research flow perspective, it will mention what the main goals are of the research, spread out in a main research question, and multiple sub questions, that together should result in a complete answer to the problem at hand. The research design will go into how the knowledge found via the sub questions asked will be used to create a deliverable that is usable for Eurofins FSS. Research limitations will go into the limitations of availability, validity and reliability of data that is used or gathered during the research.

1.3.1 - Research design

Research design is based around the main problem that needs to be addressed, via use of research questions and sub questions. The research question that is connected to the problem statement mentioned before is: *How can Eurofins Food Safety Solutions implement a data visualizing system to give the best insight in their historic data for consultancy purposes?*

To answer the question properly, a problem-solving method should be applied. With the problem being based around the premise of data, CRISP-DM is the problem-solving methodology that will be used to structure the sub questions present and answer to the research question. (Schröer et al., 2021) CRISP-DM (**CR**oss **I**ndustry **S**tandard **P**rocess of **D**ata **M**ining) is as the name suggests a methodology usable in almost every sector to do data-based research to get to a particular goal that a party wants to try and reach. This is done via six steps: Business understanding, Data understanding, Data preparation, Modeling, Evaluation and Deployment. Sub questions will be divided underneath the different steps of CRISP-DM to result in well-structured research that results in a solution for the presented problem. In figure 1.2 a diagram of CRISP-DM can be found.

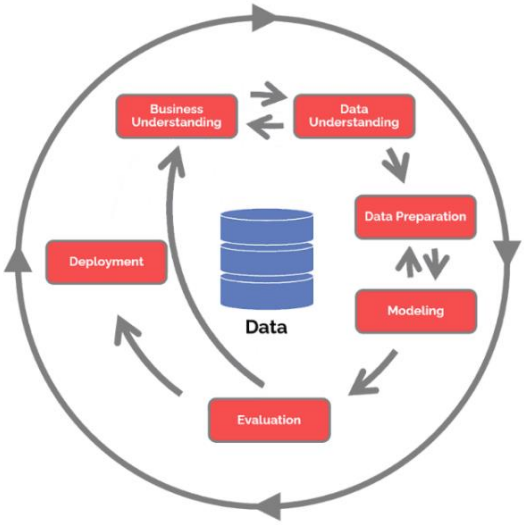


Figure 1.2: CRISP-DM methodology model.

Business understanding

Step one of CRISP-DM is business understanding. This step is the preparation step where all general data is gathered about the current situation. (Schröer et al., 2021) In the case of this thesis, chapter 1 is the biggest part of the business understanding as the problem identification together with the plan of approach is the biggest part of the contents of business understanding.

The first step in creating a visualization tool is to gain insight into both the wanted product idea, as well as the opinion of the employees within the business to gain the best business understanding.

The opinion of the people that will be using the intended system have not been thoroughly questioned about what they want in the analytical system, only the higher ups that will use and supervise the system have been questioned. To get a better view of what is useful for the company, interviews will be done with employees in different roles of the company to gain a total overview of all roles in the company and their

needs. This business understanding can then be related to theoretical research that has been researched to get insight into what visualization framework works the best within the company.

1. *What type of data analysis dashboards and designs are there to visualize data that fit Eurofins FSS' needs?*

Data understanding

After gaining insight into the Eurofins FSS itself and its goal of the application, it is time to look at the next step of understanding which is the data. Looking at data understanding, this is the step that connects the current situation to the existing data and reviews the quality, quantity and the possibilities lying within the data present. (Schröer et al., 2021) The data that will be used for my thesis will be a mix of historical data collected over the years by Eurofins FSS, data gathered and created by myself through interviewing for a better business understanding and calculating new data to the already existing database. The addition of this data is something that needs to be done very carefully, as the additional data should not influence the truth that is present through the already existing data. Lastly, identifying the state of the data that will be used in terms of how clean it is also important. This will be the groundwork that needs to be done before preparing the data for modeling. The sub questions linked to this part of CRISP-DM is:

2. *What parts of Eurofins' historical data are most useful for Eurofins FSS' employees during the consultancy process?*
3. *How can dashboarding and proper design best be applied to create the best data insight for Eurofins FSS?*

The questions above both need to be stemmed out of a mix of understanding the data present, understanding the company, together with researching literature to come to the best conclusion on how to visualize the date present within Eurofins FSS. An important factor in this step is the possible limitations that are related to data usage within the company.

Data preparation

The biggest amount of time of the thesis will be spend on data preparation, as this step in CRISP-DM is very important for the quality of outcome of the project. This step organizes the process of converting data to be able to use it for modeling in a few steps. These steps are divided as follows; Selecting relevant data, Cleaning the usable data, constructing new data if needed, integrating new data if new data enters the picture and formatting the data as needed. (Schröer et al., 2021) To guide the process of cleaning the data, the following sub questions should be answered before, and during the cleaning of the data.

4. *How should Eurofins FSS clean and manage their historical data to be able to dashboard their most important data?*

This research question will be answered by going over a plan on how Eurofins FSS should structure their database as improvement together with creating a step-by-step plan as deliverable on how to clean Eurofins FSS' datasets for visualization in a universal way for all Eurofins data processing. This will be done by the steps mentioned and documenting them in an easy-to-use way.

Modeling

The fourth and most impactful step is Modeling. In this step, the main point is the creation of a tool to give insight into the data present, together with all formulas that come with it. (Schröer et al., 2021) The tool that will be created will be a dashboard, as this is an easy-to-use way to gain insight into important factors, together with having the benefit of being relatively easy to implement in existing workflows. The idea is that the dashboard automatically updates depending on the database, together with certain customizable aspects for different occasions. To be able to create such a dashboard, the following question needs to be answered.

5. *How can the design requirements be visualized best to create a functioning dashboard that supplements the data research of the consultancy process of Eurofins FSS?*

Finding the correct design requirements for creating a quality dashboard that is in line with the view of Eurofins FSS will be done by collecting information from employees during the previously stated sub questions, together with applying the theory collected within sub question 2. Applying theories collected results in a concept of a dashboard that should both be functional as well as implementable into the workflow of Eurofins FSS.

Evaluation

After creating the envisioned dashboard, evaluation starts. With the created model, the idea is that a lot becomes possible for Eurofins FSS, yet the question always remains if this is the case. This is why evaluation is important.
In the evaluation section of the CRISP-DM methodology, possibilities within the new system get tested together with the new results found due to the use of the new system. (Schröer et al., 2021) Something that should also not be forgotten is usability, as the maker of an application is more proficient than the average user of the system created. This all together results in the following question needing to be answered.

6. *To what extent does the created dashboard fulfill the needs within data analysis of Eurofins FSS' employees?*

The evaluation will be based on a mix of analyses. Firstly, looking at how trustworthy the created system is, with an eye on the level of accuracy looking at risk assessment. Secondly, looking at the usability aspect of the dashboard. This can be tested by interviewing employees, then reviewing the statements made about the dashboard.

Deployment

Deployment is a very important part of CRISP-DM, that people sometimes forget to work out properly. For a model/application to work correctly, the system needs to be integrated, either within the software, or within the workflow of the employee. (Schröer et al., 2021) This integration is most of the time harder than the average person expects. Depending on the costs, time, training needed, and overall change in workflow, it is hard to manage the whole implementation without any problems occurring. Due to the limitations of this thesis regarding the scope of the project, together with the size of Eurofins, deployment falls out of the scope of the project. A usable version of a tool to visualize data will be made, with the purpose to be used, yet companywide implementation does not fall within the scope of the thesis.

To give a total overview of all sub questions, there specifications and the process of the research being conducted, the following table has been created:

Sub question	CRISP-DM Stage	Type of Research	Population	Data gathering method	Qualitative or Quantitative
1. *What type of data analysis dashboards and designs are there to visualize data that fit Eurofins FSS' needs?*	Business understanding	Descriptive Research	Eurofins FSS employees: consultants, managers and analysts, (Literature)	Interviews and Systematic Literature Review	Qualitative
2. *What parts of Eurofins' historical data are most useful for Eurofins FSS' employees during the consultancy process?*	Data understanding	Exploratory Research	Eurofins FSS employees: consultants, managers and analysts, (Eurofins FSS' historical data)	Interviews	Qualitative
3. *How can dashboarding and proper design best be applied to create the best data insight for Eurofins FSS?*	Data understanding	Exploratory Research	Eurofins FSS employees: consultants, managers and analysts, (Literature and Eurofins FSS' historical data)	Previously collected interview data and literature research.	Qualitative
4. *How should Eurofins FSS clean and manage their historical data to be able to dashboard their most important data?*	Data preparation	Mix Exploratory and Explanatory Research	(Literature and Eurofins FSS' historical data)	Literature research and researching Eurofins FSS data via applying said research	Qualitative and Quantitative

5. *How can the design requirements be visualized best to create a functioning dashboard that supplements the needs of Eurofins FSS?*	Modelling	Mix Exploratory and Explanatory Research	Eurofins FSS employees: consultants, managers and analysts, (Literature)	Previously collected interview data and literature research	Qualitative
6. *To what extent does the created dashboard fulfill the needs within data analysis of Eurofins FSS' employees?*	Evaluation	Evaluative	Eurofins FSS employees: consultants, managers and analysts	Interview and research into compatibility with Eurofins FSS' existing systems	Qualitative

Table 1.3: Table with complete research structure of all sub questions.

1.3.2 - Deliverables

The goal of this research is to be able to deliver the following information to Eurofins FSS to give order within their data systems, together with an analysis tool that can help with the risk assessment of their consultancy testing from a data perspective. This goal is built up out of the following deliverables:

1. A data cleaning plan worked out step-by-step for Eurofins FSS' use case.
2. A dashboard that gives insight into the historical data of Eurofins FSS by giving risk assessment insights, implemented in the current workflow at Eurofins FSS.
3. A user guide to guide users using the data visualization system in the usage of the created interface.

1.4 Current Situation

This section discusses the current situation in Eurofins FSS, looking at data analysis that is currently being performed to aid the consultancy process to retailers by Eurofins. In this chapter, an overview of the consultancy process is given, with indication of what points of consultancy can be enhanced using data analysis. Followed by a small summary of parts of the organization that can be helped by using data analysis. In the end interviews are established to find the most important data to visualize for Eurofins FSS.

1.4.1 - Data usage within Eurofins FSS

Eurofins FSS uses data gathered in a multitude of ways, from their own research data, to using open data of the food testing sector found by other institutions.

To overview the quality of Eurofins FSS data usage, DAMA's (Data Management body of knowledge, New Jersey) assessment of data management maturity of six levels is applied to Eurofins FSS. **(**Caballero et al., 2023) The DAMA assessment of data management maturity is split up into the following levels.

- **Level 0: No capabilities.** There is no organized data management system or workflow throughout the organization to aid in data management.
- **Level 1: Initial.** Using a limited set of tools, general purpose data is managed on only the required level. Data is not used outside of experts that are required to manage the data.
- **Level 2: Repeatable.** Data management tools are getting implemented in the company and role specific data support arises for certain functions. Data management tools are working towards centralization.
- **Level 3: Defined.** The introduction of data management systems and processes that are scalable and can enable the organization in different ways. Data management has policies, formal process definitions, Data quality increases due to more control of data.
- **Level 4: Managed.** Data management results in increased knowledge within the company. Data can be used to help new project results of new projects. Data-related risks need to be managed. KPI's become an important part of data management.
- **Level 5: Optimized.** Data management is optimized. automation and technology change management are managed closely. All processes are screened and improved using data. Forecasting can be applied to important processes.

When looking at Eurofins FSS, the level of data management that corresponds to the company is around level 2 and 3 depending on the process at hand. Underneath a list of important departments of Eurofins FSS is named where data management is a keystone of the department and can be improved. All departments are reviewed on their level of data management.

Reporting

First of all, due to Eurofins FSS only being only one department out of twelve in Eurofins Netherlands, and Eurofins Netherlands only being one country out of the total Eurofins Group reaching sixty-two countries in the world, Eurofins FSS has very big corporate structure surrounding the company. This structure makes sure that all business units align with the view of the Eurofins Group, to keep the same goal in mind or the total company.

With the big corporate structure of Eurofins Netherlands, reporting in between business units needs to be done for higher-ups to be able to have a view of everything that is going on in all business units. To be able to do this properly, data needs to be collected and kept safe to guarantee trustworthiness in the data reported to the higher-ups. This database is a massive dataset that includes every test handled by Eurofins FSS, combined with other important and usable testing data such as the company from which the test is, the date the test was taken, what parameters have been tested of the product, and more. As of right now this data is only used for administrative purposes, yet it has potential to be used more generally in other processes as well.

In total the process would be around level three looking at the DAMA assessment of data management maturity. The process overall is scalable and has clear guidelines and processes lined out yet is still in a stage of low automation. (Caballero et al., 2023)

Consultancy

Consultancy is the other part of Eurofins FSS that uses data to get to the conclusion they need. Consultancy works closest together with the client companies of Eurofins FSS as they are a point of communication for a client company, as well as the designers of the testing plans that are given to client companies.

To summarize the consultancy process, the consultant in contact with a company visits for a number of days to the client, and does research on the products. With this research together with public data on food testing and the expertise of knowing certain products specifics, a testing plan is set up. This testing plan is a list of different tests that should be done according to Eurofins FSS expertise to safely produce certain products and meet the regulatory requirements of testing the products sold. This plan is discussed with the client and changed to the client's wishes. After the testing plan is set in stone, the testing process starts. The testing process is a series of planned dates and tests whereas as a result, all test data is reported in a general report ranging over all the documented tests. This rapport is discussed with the client, giving consultation on what the problems are and how to deal with the problems presented. A BPMN 2.0 of the process can be found underneath.

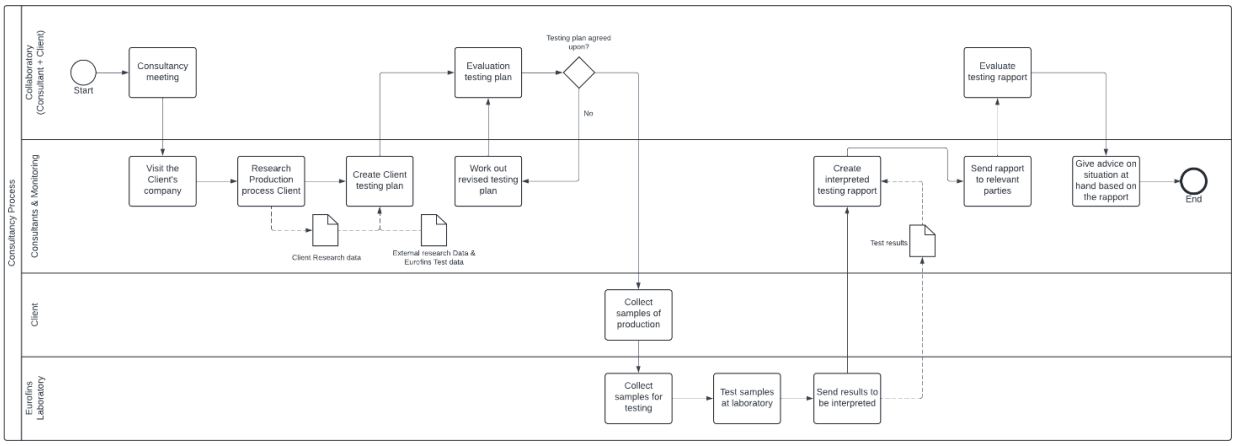


Figure 1.4: BPMN 2.0 model of the consultancy process of Eurofins FSS. (A bigger version can be found in Appendix F)

Within the total consultation process a lot of data is used, yet all data is gathered by outside sources, together with this data research being done repeatedly by hand. This process can be optimized by having different data sources available in a direct manner, instead of needing to search for data time after time again. The list of all data sources used in consultation are as follows:

-RASFF
-RiskPlaza
-EFSA
-NVWA
-RIVM
-Dutch Branch organizations (A list off all relevant branch organization can be found in Appendix C)

In total the Consultancy department uses a lot of data that is not automated or included in a data management system. Overall, I would say that the DAMA data management maturity level of consultancy between level one and level two lies. The automation and use of Eurofins FFS' own data within this data would result in a big improvement of quality of consultancies as well as a more efficient workflow for the consultants themselves. **(**Caballero et al., 2023)

Quality checks

Lastly, when discussing the aspects in which data currently is used within Eurofins FSS, total management checks are the last point on the list. In total, management within Eurofins FSS is interested to know what is going on within the company, employee-wise, but also business-wise. Management at Eurofins FSS wants to know what is measured over the years, together with learning how data can be used to help processes within the company itself. The data that is present as of right now is mainly used to report to higher ups, yet management also uses the data to keep track of the most important points of data, to keep an overview of the company, as well as looking at opportunities to use the data that is saved to possibly evolve the company to be more data and statistic driven. One of those opportunities is the research that I am performing.

Overall the process of companywide quality checks has an expected data management maturity level of two on the scale of DAMA, given that the data is present and there are systems to help the user with creating the documentation needed for the quality checks, yet all systems are not automated and done mostly via manual work using the systems. **(**Caballero et al., 2023)

1.4.2 Improvement possibilities

As stated before, Eurofins FSS has multiple process parts that can be streamlined better using data that is present in the database of Eurofins FSS. The main focus of Eurofins FSS is consultancy and testing. This is also the place where the most positive changes can be made to the total process by implementing data-analytics. Looking at the DAMA data management maturity levels, Eurofins FSS has improvement possibilities in multiple sectors when looking at data management, with consultancy being the lowest performer. To find all the opportunities for improvement within Eurofins FSS, managerial meetings were held, to find the weaker points of the company per department. This information was relayed to a data analysis project at the beginning of the research track of this thesis. This gave insight into deciding what the best improvement possibilities were within the company looking at data visualization.

First of all, during the consultation process, a testing plan needs to be made by the consultant with regard to information about the different food sectors that need to be tested, together with extra information and requests that are given by the client.
As said before, all the information about the food sectors and risk factors is gathered externally, resulting in a lot of work for consultants finding and grouping all information together to get to an end resulting to base their testing plan on. This is very inefficient and can be optimized better.
First of all, grouping all data together should be able to be done automatically if all consultants rely on data that still needs to be found during the creation of testing plans. Depending on the source of the data, the data can either be qualitative or quantitative, resulting in it being harder to make a general overview of all data automatically. As of right now, all data, both quantitative and qualitative, needs to be reviewed fully by hand to help get to a conclusion. A realistic step in optimizing the consultancy process is automatically analyzing the quantitative data that can be found within the data sources present in food testing. This step should increase the quality of the process of finding data for consultancy, as an automatic script is faster and more reliable when it comes to finding quantitative data in a database compared to humans.

Secondly a change could be made to the data that is used to create the testing plan. As mentioned before, a lot of data on the previously done testing is saved in the database of Eurofins FSS as a safety measure to report and ensure good data. Due to the number of tests that are performed by Eurofins FSS, the historical data can be used to do analysis into what type of products or tests are more important to test looking at the relevant data that was found. With the implementation of own data in the analysis done for consultancy, the relevance of other data sources goes down, resulting in more independence as a company, together with a faster consultancy system as the system can be made to the liking of the consultants themselves.
With all the prior mentioned optimizations, also the quality guarantee increases, as the quality of the data is monitored by Eurofins themselves, whereas right now most data is taken from external parties, resulting in no connected evaluating factor that evaluates the quality of the data that is presented.
An extra benefactor that is present when having an own data system is that it becomes possible to show the personalized data to clients to show the significance of the research done into the results found in. To optimize the process of giving consultancies, the data that is used to make the testing plans should be able to convince the client of the relevancy of the data when shown correctly.

In total, there is one part of the consultation process that can be improved to increase the efficiency and quality of the consultancy of Eurofins FSS, and that is the retrieval of the information and the creation of the testing plan. This part of the consultation process can be seen in figure 1.5. This part of the consultancy process can be optimized using visualization and automated applications to increase the speed, quality and independence of Eurofins FSS' consultants. As stated above, improvement possibilities such as implementing a system where Eurofins FSS' own historic data is used to aid the data analysis for consultants to create a testing plan can increase the quality and effectiveness of the total consultancy process drastically. Compared to other options, using collected testing data to help with the data analysis per group of variables is the most useful, as it solves the most problems that are connected to the consultancy process.

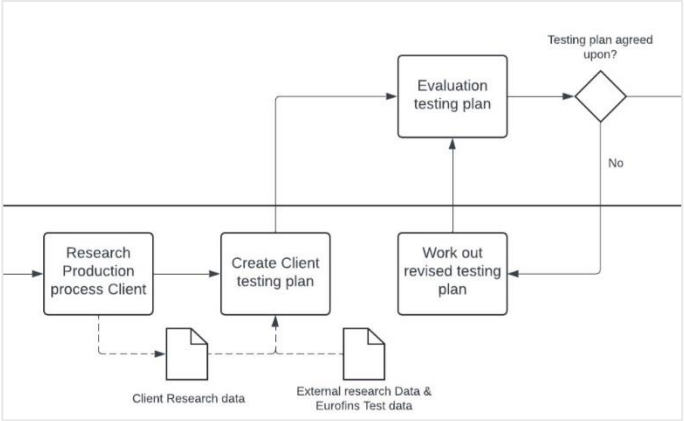


Figure 1.5: Enlarged view of the consultancy process of Eurofins FSS part that is the main focus of improvement in efficiency and quality.

1.4.3 Required data for useful visualization

With the data that is present reviewed, it is now clear what resources are available to create an information dashboard that supplies Eurofins FSS' consultancy with data-driven analysis to help make a testing plan. The next step of the process is to look at what data is useful to analyze to create a good working data tool. This data will be found by interviewing the specialists in the fields of work that are of importance, leaving management and consultancy.

The reason for choosing meetings and interviewing to find the right solution is because the specialists within the company know best the needs of the clients and the company itself when it comes to getting useful data. Together with the previous point, an interview gives the possibility for the interviewee to give extra comments on the subject at hand, together to ask further questions if a certain subject comes up that is important.

The interview is based around ten questions together with a consent form to give consent for the use and the storage of the audio taping that will be done. The interview has two parts, first going into the consultancy process as it is running as of right now. Secondly, going into the possibilities within consultancy at Eurofins FSS with extra data-driven systems available. The total interview plan and consent form can be found in appendix A & B.

Chapter 2: Theoretical framework

This chapter is the theoretical framework used to write the thesis. This framework goes over both technical jargons used within data analysis, as well as explaining company related concepts that are used often within Eurofins FSS. This chapter is divided into two parts, general theory and visualization techniques.

2.1 – General Theory

Dashboard

The definition of a dashboard as Velcu et al. (2012) states is a "visual and interactive performance management tool that displays on a single screen the most important information to achieve one or several individual and/or organizational objectives, allowing the user to identify, explore, and communicate problem areas that need corrective action".

With this definition set, different sorts of dashboards can be categorized in groups. Damyanov et al. (2019) researched the different primary sorts of dashboards that exist and gave three main categories of dashboards that exist within companies: analytical, operational and strategic. Analytical dashboards are used for the statistical analysis of data within a company, operational dashboards are used to oversee a certain (part of an) operation looking at different points of interest depending on the operation at hand. Lastly, strategic dashboards are meant to give a strategic overview of the long- and short-term goals of the company. These sorts of dashboards can also be used together to give the exact wanted overview. More on this topic can be found in section 4.1, visualization techniques.

Design

Design within data visualization applications is an important field that cannot go unnoticed. For people that need to use the designed application it is very important how the application is structured and looks, as the effects of design are significant. Nieto et al. (2020) says that both the clarity of the data shown, together with the concentration of the person viewing the data are significantly connected to how visually pleasing the designed dashboard is for that person.

Within design also layout and interaction come up as subjects. Both factors are important to the quality of use of the user. Dashboards should give a feeling of self-service, giving the user all the information that they need, without any need for a deeper understanding of the mechanics behind the dashboard. (Damyanov, 2019) More on this topic can be found in section 2.2, visualization techniques.

Historical data

In this thesis, collected historical data will be used. In the context of this thesis, historical data is data collected prior to the making of the dashboard during a three-year period ranging from 2021 to 2024. With historical data, data quality and relevancy need to be discussed. (Zhou et al., 2021) When working with data collected before planned research, it is important to keep track of quality of data as the quality of data can differ per timeframe. Methods of testing might have changed, or ways of noting data in detail have changed, possibly reducing the quality of data. This needs to be considered when processing the data within the present research.

Data Cleaning

Data Cleaning is important to manage properly when doing analysis on historical data. With older data, formatting and notation can change over time resulting in data that does not match together to do analyses properly. To solve this problem, data cleaning needs to be performed. This is done via systematically going through the data and following steps to make all data usable in one standardized way. (Oladipupo et al., 2023; Zhou et al., 2021) The step-by-step plan for cleaning Eurofins FSS' data will be created in chapter 3 of this thesis based on literature and data found within the corresponding chapter.

Risk Assessment Tool

The goal of this research is to give insight into the historical data of Eurofins FSS to use the data present with making a data-driven decision on the risks within food testing and the frequency of testing that is needed. Therefore, the risk assessment of Eurofins FSS needs to be compatible with the dashboard and analysis done.

For risk assessment, Eurofins FSS uses a risk assessment tool called the Risk Assessment Matrix. This matrix is often used for risk assessment in many different sectors of work. This matrix is built on two axes, the first axis displaying the likelihood the risk shows up, and the second axis displaying the severity of the risk. Both axes get a score ranging from 1 to 5. And the matrix is based on the multiplication of both axis scores, resulting in a place in the matrix with a certain score. In general, different groups are made based on the final scores present. (Benešová et al., 2018) An example of a risk assessment matrix with related scoring can be seen in table 2.1 and below:

Risk Assessment Matrix						
		Likelihood				
		1. rare	2. unlikely	3. possible	4. probable	5. highly probable
Severity	1. very low	1	2	3	4	5
	2. low	2	4	6	8	10
	3. medium	3	6	9	12	15
	4. high	4	8	12	16	20
	5. very high	5	10	15	20	25

Table 2.1: Example of a risk assessment matrix.

- Scores ranging from 1 to 3 are low risks.
- Scores ranging from 4 to 9 are medium risks.
- Scores ranging from 10 to 20 are high risks.
- Scores ranging from 21 to 25 are critical risks.

The idea of risk groups is to have an overview of what risks need to be dealt with first, and what problems can be discussed at later notice. (Benešová et al., 2018) Companies can have protocols connected to the different groups of risk assessment scores to answer different problems on a needed level.

Scoring likelihood and severity in the right category is very important to have a well-functioning risk assessment matrix, therefore most companies use a precise calculation to categorize said risks properly before using the matrix to create an overview of risks. Within Eurofins FSS the risk assessment matrix is structured differently, as food testing has different parameters, and severe cases are very important, the matrix scores per square are fully customized to fit the scope of food testing and the safety of the general population.

Power BI

To create dashboards and manage data well, Power BI will be used. Power BI is the Business Intelligence application from Microsoft, which is meant to be used to manage and use collected data to visualize goals or aspects selected by the user. (Metre et al., 2023) Power BI is used quite regularly to generate business insights via data driven dashboards, due to the high number of ways to visualize data. Together with this Power BI also has integrated systems to convert or transform different sorts of data into wanted calculations.

Power BI is user-friendly and very informative for a broad range of people, as the visualization that Power BI creates, makes all data very clear to all users. Therefore, Power BI is a favorite among companies to use for data analysis. (Metre et al., 2023) Within the company, Power BI is used to manage the current data collected, therefore the use of Power BI is expected from the company.

Power BI is a Business Intelligence application that is also data secure. Power BI falls underneath the umbrella of applications that are related to Office 365. Which all have a high standard of data security. Given that other Office 365 applications are also used within Eurofins FSS to structure the used data, the transfer of data from the source to the application is also very secure as the same format can be used, resulting in the possibility of a direct link between source and application.

Normalization

Normalization of a database is the process of changing the structure of a database to create relations between data to avoid problems with data organization. (Lien, 2009, pp. 51–126**)** To create a normalized database that is best ordered to reduce the number of errors found within keeping data, a theoretical step plan is used called the DBMS normalization.

The DBMS Normalization uses different levels of Normalizations to create an easy-to-follow step by step plan for improving database normalization. DBMS has seven levels of Normalization as shown underneath:

- First Normal Form (1NF)
- Second Normal Form (2NF)
- Third Normal Form (3NF)
- Boyce-Codd Normal Form (BCNF)
- Fourth Normal Form (4NF)
- Fifth Normal Form (51NF)
- Sixth Normal Form (6NF)

The DBMS normal form gets increasingly better for the data, the higher the form is, yet with this upside, the downside of work and effort comes along. The most used Normal form is the third normal form, as this is seen as the best consensus between time and effort compared to quality. Within this research, the third normal form is also used as a guideline for database structure.

If a Database is in the third normal form, the database has the following important features:

- The database should only hold unique records, no repeating groups.
- Removing redundant information by ordering different sorts of data in different tables, this requires the data to be linked using a key structure to connect data tables to each other.
- Make sure that all data I independent from one another.

2.2 - Visualization Techniques

To be able to create the best fitting visualization application, the first step is to collect all information surrounding the possibilities of creating a data visualization system.
Due to the application being designed for a company to use, usability is one of the main components within the design philosophy. With an eye on usability and the width of the scope of the company to get people to understand visualization software, dashboard-based systems will be the focus of the research. Dashboards are used a lot in the world due to the low entry bar of understanding for the user, resulting in an easy to pick up and use system for broader numbers to use.

Dashboard types

The term dashboard, as Damyanov et al. (2019) uses, means "a visual and interactive performance management tool that displays on a single screen the most important information to achieve one or several individual and/or organizational objectives, allowing the user to identify, explore, and communicate problem areas that need corrective action" Defined by Yigitbasioglu and Velcu in 2012. Based on the given meaning of the word dashboard, types of dashboards are formed. The first two update categories within dashboards are dynamic and static dashboards.
Dynamic dashboards change with time, and are therefore mostly designed online, as data can be updated at a moment's notice. Static dashboards are quite the opposite, being a rapport based on a certain time. This can be a dashboard online, yet also certain quality rapports can be put under this category.

Stehle et al. (2020) comments on the effects of automated updating dashboards compared to manual updating dashboards. Dynamic dashboards are always synchronized up with real time yet have the problem of change blindness occurring.

Change blindness is the increase of users not seeing changes occur within the data, due to the constant updating of the data. This problem occurring with gradually changing data can result in possible wrong interpretation of data, as small data changes can sometimes be impactful. The problem of change blindness can be counteracted by using detectable flags to show that a data point has changed, yet this is not perfect in all scenarios. The opposite applies to static dashboards, where change blindness is not present, yet not all data will always be real time accurate.

Damyanov et al. (2019) states that there are three main different types of dashboards: analytical, operational and strategic. Analytical dashboards have the main function to be able to process big data while extracting this data online. Resulting in being able to see quality analytical calculations in a clear way given from the extracted data. A lot of information is the main resource of an analytical dashboard, and can be updated either live, or every so often when necessary.
Strategic dashboards monitor KPI's of long term and short-term goals, to keep track how the improvement process within a company goes. These dashboards are more global, and less interactive.
An operational dashboard gives an overview of a certain operational system, showing KPI's of that certain operation. Damyanov et al. does not go in depth about operational dashboards.

Dashboard design

A dashboard has a goal of outputting visual results via calculation. Zhang et al. (2022) created the following figure to visualize the process of a data visualization system.

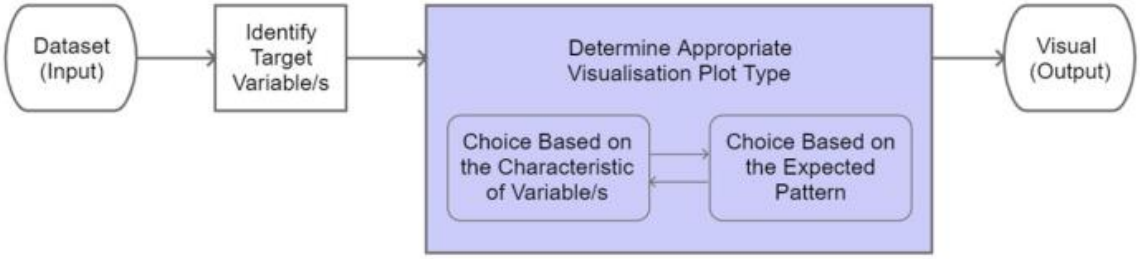


Figure 2.2: Analysis oriented process of data visualization. (Zhang et a.l, 2022)

Together with the figure shown above, Zhang et al. (2022) created the following overview of important factors to think about when designing a data visualization system. The given models show the process that needs to be designed, together with factors that need to be kept in mind when designing a dashboard.

Looking at the designing attributes, a dashboard can be designed and structured in a lot of different ways.

	Data Visualization System
Orientation	Analysis oriented
Audience	Professionals
Basic Units	Variables
Informative Feature	Simple message
Entertainment	Thin
Visual Space	2D static visual space
Visual Complexity	Straight forward

Table 2.3: List of important factors of a data visualization system.

given the number of options when designing a dashboard. There are a lot of small aspects that need to be held in mind when designing a dashboard. Damyanov et al. (2019) made a dashboard features diagram, to visualize all points of interest looking at dashboard design. This diagram can be found in figure 2.4

The biggest difficulty is getting all wanted information clear on a screen. Therefore, all information needs to be shown in an effective way, which changes depending on the company and clients that are worked with. (Damyanov et al., 2019) This works together with the interactivity of dashboards, depending on the client, interactivity of the dashboard can become a more or less important aspect of the dashboard, which is a big impact on the use of the provided dashboard. Gonçalves et al. (2023) mentions that there are two simple standardized points of quality creation when designing a dashboard. These are the following two points:

1. Visuals should create a clear overview that in function employees should understand, with text as extra description to increase clarity. This is the best way to divide text and visuals to keep clarity as high as possible.
2. The dashboard needs to be visually attractive, with proper color coding to increase the understanding of visuals.

Nieto at al. (2022) agrees with the points mentioned by Gonçalves, and elaborates that pre-attentive attribute, attributes you take in without putting conscious effort in it, such as the general visualization design, are important to gain extra concentration towards the created dashboards. Aspects that fall within this list of pre-attentive attributes are variations in color, shape, enclosure, intensity, hue or size.

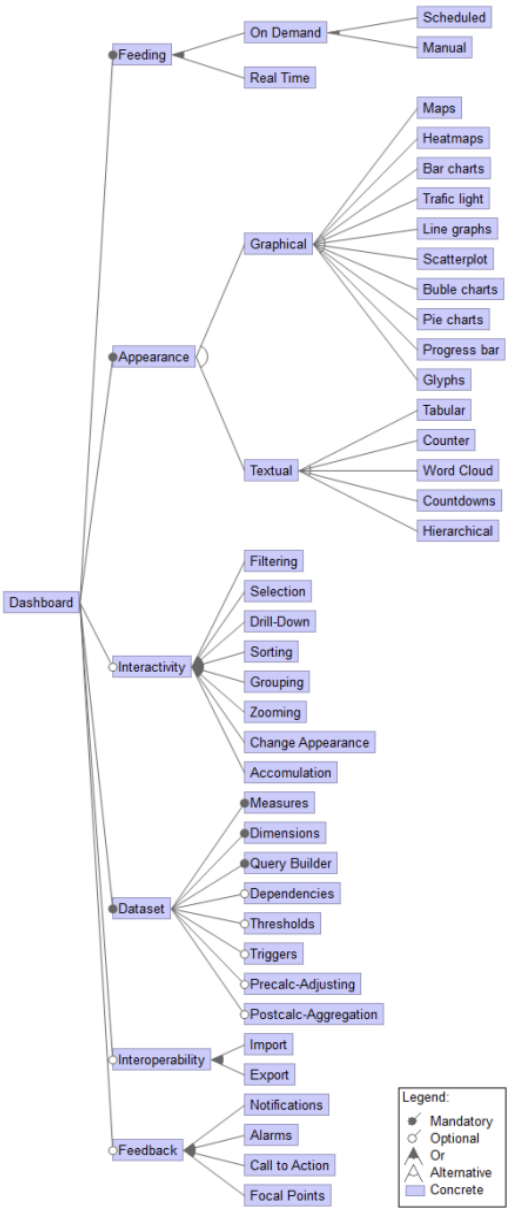


Figure 2.4: tree of all design aspects for designing a dashboard. (Damyanov et al., 2019)

Damyanov et al. (2019) provided a list of the most important principles of visualization within dashboard design. This list shows all the most important factors that need to be thought about and provided properly to ensure a quality dashboard. The list of principles can be found in figure 2.5.

Area	Principle
Functional	1) Defining the problem, 2) Providing structure, 3) Providing clarity, 4) Providing simplicity, 5) Providing emphasis, 6) Providing unity.
Administrative	1) Defining the problem, 2) Providing structure, 3) Providing clarity, 4) Providing simplicity, 5) Providing emphasis, 6) Providing unity.
Aesthetic	1) Harmony, 2) Aesthetic proportion.
Cognitive	1) Facilitating attention, 2) Facilitating perception, 3) Facilitating processing, 4) Facilitating memory.

Figure 2.5: A list of all Principles of dashboard design. (Damyanov et al., 2019)

Dashboard usability

Usability is an aspect that very important when designing an application, as the application needs to be usable in a proper way for the goal it was created for.
First of all, one of the most important usability points in this day and age is the possibility to access the dashboard at all times in any location. (Damyanov et al., 2019) Given that working from home also has gotten more popular, this aspect has become more important than ever. This also means that the created dashboard needs to work universally no matter what type of device or screen is used. Over time, companies and their data most of the time changes, therefore a dashboard needs to be able to fit different data sources as well.

Dashboards can both be interactive or not, which needs to be dependent on the function of the part of the dashboard. In total interactivity has not right or wrong. The only balance that needs to be held is the feeling that the user has self-service over the data, meaning that the data is there for the user, and that the user can see all data they are interested in. this dynamic way of working for the user increases the total usability of the application. (Damyanov et al., 2019)

Nieto et al. (2022) mentions a similar effect of usability for the user depending on the amount of visual clutter. If the dashboard is too full, and therefore it is hard to understand the dashboard. The usability and motivation of the user both go down, resulting in a less useful dashboard.

- As mentioned before, change blindness can be the effect of dynamic dashboarding, due to the constantly changing data. (Stehle et al., 2020) This can reduce the quality of output of the system, as users are less focused on the data changes, as it happens very frequently, resulting in a lower quality of outcome within the total process.
  Looking at the refresh process of a dashboard, a refresh can be disturbing for the user if the dashboard is not usable while the dashboard is updating its data sources. (Stehle et al., 2020) This is a usability hazard for static dashboards that update once every time period. This factor is important for systems that need to be running all the time, which is more common under operational dashboards.

Chapter 3: Data Preparation

After the goals of the business are properly structured and priorities are set, it is important to prepare data for analysis and visualization. Within Chapter 3 the following steps of data preparation will be followed. The first step of data analysis is analyzing the data that is present. After that is done, the next step is scaling the data to the project limits, so all data that falls outside the scope of the project is not taken into the process of data cleaning and data transforming. The next step is data cleaning, which will be done via creating a step-by-step plan, together with using this step-by-step plan to clean the data that is present in the project. Lastly, a summary of all possible validity issues is given.

3.1 - Present Data

Before actions can be taken, the data that is present needs to be evaluated. For the project at hand a dataset has been handed over which needs to be transformed and cleaned to conclude using data analysis. The data consists of a data table with forty-one columns regarding tests that are done in the calendar year 2023. In total, the dataset has 811.885 initial entries of data, which will be lowered to later in the process while cleaning the data in part 3.3 of this chapter.

The exact data that is saved falls underneath the trade secret of Eurofins FSS, therefore the exact data cannot be named within the report written. To give insight into the sort of data that is present in the dataset of Eurofins FSS, I will describe the information that is available.

The table underneath shows an overview of the data available in the dataset given together with the ability to use the data for research purposes. The expected relevance of the data has been determined out of meetings with managers and consultants expected to use the system to look at the needed variable within the dashboard, as well as systematic requirements for big data analysis.

	# of columns	Sort of data (Qualitative or Quantitative)	Expected relevance (Usable/ Not usable)
Where the test took place	2 out of 41 columns	Quantitative	Usable in the future, this information is usable if internal research is wanted, if that is not wanted, this data becomes useless for the determined end product.
What sample/test group does the test fall under	3 out of 41 columns	Quantitative	Usable, this data is very important to be able to distinguish all data that is being processed.
What test is performed	3 out of 41 columns	Quantitative	Usable, this is very important information, as this is the one of the main ways to see what aspects in testing are going right or wrong. This together with the results is the basis of the possible research.

What clients are involved with the test	7 out of 41 columns **(4 out of 7 columns useful)**	Quantitative	Usable to a degree, looking at what companies and suppliers make and what mistakes is very important, as well as to be able to compare different companies with one another. Yet more data then needed is in this group, 3 columns are double data, therefore not needed.
What type of product	4 out of 41 columns **(2 out of 4 columns useful)**	Quantitative and Qualitative	Usable, very useful information, yet some data is not standardized and qualitative, which cannot be used (2 columns)
Process related date administration	6 out of 41 columns **(1 out of 6 columns useful)**	Quantitative	Usable to a degree, as putting a time and date onto a test, to see what changes over time is a very important step to see a trend over time. But not all data is needed. Given that 1 time is needed to be linked to the tests to a time, all 5 other points are not useful, together with consistency issues in date-based data within the dataset.
Result administration	1 out of 41 columns	Quantitative	Not usable, this goes into what rapport the data is used for.
Results of the test	3 out of 41 columns	Quantitative and Qualitative	Usable in future research, the data present is the direct value collected out of the tests, together with the benchmark data, difference calculations can be done, yet this falls outside the scope of the process.
Testing benchmarks	2 out of 41 columns	Quantitative and Qualitative	Usable in future research, the data present is the benchmark or the test, together with the resulting data, difference calculations can be done, yet this falls outside the scope of the process.
Reporting result	2 out of 41 columns	Quantitative	Usable, the data sorts result of tests into categories of how severe problems are.
Comments on the test	1 out of 41 columns	Qualitative	Not usable, the data is fully qualitative.
Reference test to regulation that is in effect.	1 out of 41 columns	Quantitative	Usable in the future if research into regulations wants to be done statistically later.
Columns with no entries	6 out of 41 columns	-	Not usable, no data present.

Table 3.1: Overview of all data attributes present within Eurofins FSS' dataset given for research.

The table above results with a mix of data from which parts can and cannot be used to come to a result. in total there are thirteen different groups of data. Eighteen out of forty-one columns are to no use in total, eight out of forty-one data columns are useful in research that might be conducted in a later state of the process, Leaving 20 columns of data usable for the project.

3.2 - Data Cleaning step plan

Before data can be used within data analysis, the data needs to be processed to the proper formatting and level of cleaning that is compatible with the application used. Within data cleaning, different apply depending on the software used. To make data cleaning easier to follow, and to create a dataset that is more universally usable, certain standard data cleaning steps can be taken to decrease error percentage within the dataset. The following data cleaning step plan is used to reduce the number of errors within the dataset:

Removing Irrelevant Data
The first step for cleaning data is removing all data which is not useful to the dataset. Removing irrelevant data results in less data where errors can occur in, as well as limiting human mistakes made in data that is not relevant to the initial goal. The removal of irrelevant data also results in a faster analysis system, as less data needs to be read by the software. (Gao et al., 2020)

Removing Duplicates
As mentioned in the theoretic framework, deleting duplicates within a dataset is very important to ensure reliable analysis. (Gao et al., 2020) Removing the duplicates within a dataset results in a faster dataset as well as a higher reliability of results, as wrongful coded calculations cannot mistake data that might be noted duplicate.

Deleting Blank Values
Within datasets, sometimes blank values are passed through the system, either by software or by human made mistakes. (Gao et al., 2020)(Globa et al., 2022) Blank values can result in clutter within visualization systems, as well as problems when using count measures, as a blank values is still counted as an entry, skewing the results towards zero.
It can also be that certain parts of data entries are not filled in properly, resulting in certain critical information being blank. This data is not reliable and should not be taken into account when doing data analysis, therefore all data entries with blank values in important areas have been removed from the dataset.

Datatype Conversion
To ensure proper calculations resulting the dataset, it is important to have matching datatypes linked to all data values. (Gao et al., 2020) Wrong datatypes can result in wrong interpretation of data values from a software. For example, if a date is not noted as a date datatype, it will be seen as a regular number, which is not correct and can result in mistake when additional calculations come in frame.

Checking related Values
Within data entries, values can have different relationships with each other. This can range from either one-to-one, one-to-many or many-to-many relationships. (Gao et al., 2020) (Globa et al., 2022) With one-to-one relationships, data points from one entry need to adhere to the one-to-one relationship that is present between certain data points. For example, a company name needs to be linked one-to-one with their respective company code, as no other company can have the same code. If this is not the case, an error is present within the dataset. These errors need to be deleted, which is done by evaluating the data relationships present within the dataset, and deleting all anomalies present within the dataset.

Standardizing data

With big data, it is important that data is as standardized as possible. With every extra loose variable, getting reliable output that is usable within data analysis. (Globa et al., 2022) To be able to group variables together properly, all the same properties should be formulated in a standard manner. This is called standardization. For Eurofins FSS' dataset, this is mostly needed for certain term usages. In the system, different locations companies have different ways of noting their results of tests, this needs to be aligned properly to do data analysis reliably. The second part where standardization is something that needs to be done, is using different languages in the same system. Within the database of Eurofins FSS, some terms have been changed over time from Dutch to English, to make the system accessible for everyone, yet this was not done for the historic data. This results in the historic data still needing to be translated using standardized terms.

3.3 - Data cleaning steps taken

With the dataset present within Eurofins FSS the steps above were taken to clean the data to a level which is usable in the data analysis system that will be created. The dataset of Eurofins FSS is decently organized, with a lot of information stored that is not useful for most instances, yet still accessible for specific use. The Eurofins FSS dataset is in total has a few small consistent transformations which are needed before being able to be used within data analysis. The following steps have been taken to create a consistent dataset which always works without errors.

1. Deleting all unnecessary information around the dataset grid. Extra information was always present when downloading a dataset to keep track of the origin of the dataset.
2. Deleting all unnecessary Columns as stated above. This makes sure the created program does not have problems with certain new data types as well as reduces unnecessary load for the analytics application.
3. Standardizing phrasing of certain key variables, as data from beginning 2023 and the end of 2023 have different ways of communicating the same idea.
4. Changing empty entries into their related entry. Blank entries stand for certain values within the company which are used very often and therefore faster to leave empty.
5. Delete Values that have Blank in unexpected places, which can occur due to handling certain data manually.

3.4 - Data Validity

When checking datasets, data validity is a very important aspect. The dataset used within this research is a preexisting dataset of Eurofins FSS, encapsulating all tests Eurofins FSS has done over the last calendar year. This dataset is used by Eurofins FSS to be able to find certain results of tests back from certain companies or tests. To be able to do this in a proper manner, the data needs to be valid to a certain extent. With Eurofins being a major corporate company, it is assumed that the data is valid with little errors, as the reporting structure of the different company layers within the Eurofins Group have precise ruling about the need of reporting data summaries to higher company layers.
While checking on errors possibly present within the dataset of Eurofins, the following validity issues where found and removed:

- The dataset includes all tests performed under the name of Eurofins FSS from the last year. This also included the test values used to test new test codes and systems. These values were removed to increase the validity of the test.
- Within the dataset, the data was screened on bugs within the data. One of these bugs was that for one company no times were available of when the test was performed, therefore this inconsistent data was removed.
- Within the data of 2023 of Eurofins FSS, a three-month bug appeared in the saving of data, resulting in the data saving double, this problem has also been removed from the dataset.

With all the changes made to clean the data, the total validity of the data is high. All data values within the dataset are as far as known within both the IT and Management layer of Eurofins FSS correct values that represent tests performed during the last calendar year. This results in the data being seen as valid.

Chapter 4: Modeling

The next step in the process of developing a data visualization system that can aid consultants of Eurofins FSS to consult other companies is the modeling of the application itself. Modeling the application will be done in two steps. The steps are as followed:

- **Modeling the Application:** The modeling of the application will be divided in three parts that will be discussed one-by-one:
    - Semantic model
    - The application
    - Usability
- **Steps of integration:** The created application needs to be integrated into the workflow of Eurofins FSS before being able to be used properly.

Using these steps within a CRISP-DM framework will result in an application that can be used by Eurofins FSS to aid in their consultancy process. The modeling of the visualization system will be done on basis of requirements and insights gained in the previous steps of the CRISP-DM framework.

4.1 - Modeling the Application

The following part of the chapter is about the creation of the application. First it will go into detail on the data structure of the model, it will explain what type of visuals are visible and how these are created and explain the how the focus of the model, usability, is implemented throughout the application.

4.1.1 - Semantic model

Behind the application being created, the data needs to be managed properly. Within Power BI, this is done via the semantic model of the application. The semantic model of a Power BI application can be compared to an entity-relationship model, which is used to visualize the structure of a database.

The semantic model of the power BI application is split into sixteen entities, some of which are fully standing alone, whereas most are connected to one another using a primary key structure. In Appendix D, a picture of the semantic model can be found. The following sixteen entities are structured within the semantic model.

First of all, all connected Eurofins FSS data related entities:

- **Export**: this is the main export file which has been exported from the Eurofins Database and imported into the power BI application. This is the source of all main data, therefore most points within the semantic model are connected to this root. In this entity, certain new calculations were made to structure other entity keys well, as well as creating certain measures that are needed for visualization of the base data.
- **ParentCompanies & ClientCompanies**: Both the Parent company and Client company entity have all company related statistics per company, this is important for analysis looking at the statistics per company, to see how companies perform compared to the baseline. These entities are only linked to the Export entity.
- **Parameters**: This entity has all loose parameters that are tested and their statistics within it. This entity contains very detailed information that is not used for many cases and is not too big. It is connected to only the export entity.
- **Parameter Severity Level**: This entity goes over the numbers that are based on the result groups of the parameters that are performed. To goes over all statistics related to the pass and fail rates of parameters. This entity is only connected to the export entity as well.
- **Product Category Management**: This entity is an extra imported file that links product names to certain categories within food products. For example, Milk is categorized under Dairy. This category system is used in the front-end of the system, yet not saved in the back end of the system or database. This addition must be made manually once, yet after the list is added, the system the data is taken to the export data and added to the standard dataset entries as a value.
- **Product Category Statistics**: This entity holds the statistics related to the product categories, making it able to show how the product categories do against each other, and being able to see what type of product has flaws in what sector. This is also only linked to the export entity.
- **Sample** & **Sample Severity Level**: The Sample entity has the whole export dataset changed to only the total sample tests that are done, as a singular sample gets tested on multiple parameters.

The sample-based results are the main way for Eurofins to see what problems occur per product er category, therefore this transcript and calculated entity is very important.
It is created out of the export entity, taking the sample severity, and taking all the higher layered categories within the data that are given through via the export, yet not used within the export entity to calculate measures for total analysis. Sample Severity Level looks at the results that are passed through the Sample entity, it is connected to only the sample entity.

- **Tests & Results per Test**: The Tests entity is connected to the export entity, similar to the sample entity, the test entity holds all data per test that is performed. For example, the statistics are ordered per test, so it is possible to see the results of singular tests, such as salmonella, or lead for example. This entity is very important to keep track of what tests are done often, what tests are more problematic, and related to what products these test problems occur. Results per Test give an overview of the total scores of the tests done. This entity is only related to the test's entity.
- **List of all TestCodes**: List of All TestCodes links all test codes to test categories. Tests within Eurofins FSS have a separate code for all tests done within Eurofins. All tests are ordered in different test categories, ranging from microbiology, labeling and much more. These categories are not put in the database, therefore need to be added via this entity giving all tests their category.
- **Test Categories**: This entity connects all statistics with the test categories introduced via the List of all test codes. This entity holds all information around these categories.

Second of all, all the RASFF related data entities.

- **RASFF_window_results**: The RASFF window results entity is the entity that holds all imported data from the RASFF database. This entity is used for the additional data visualization of RASFF, which is used to compare to the Eurofins data. This data is separate to the Eurofins data and is only linked to the RASFF_MergeCatergoryHazard entity.
- **RASFF Dates**: RASFF Dates is an entity that changes all dates within RASFF to a date without different hours during the day, as this disrupts the visualization of Power BI, to keep the application as much as possible in Power BI, this change is made in one of the entities.
- **RASFF_MergeCategoryHazard**: The last entity is an additional entity which is needed to filter the Categories and Hazards from the power BI properly in one of the used visualizations. This additional entity is the dataset of the main table shown in the RASFF part of the application, where both categories and hazards based on time needed to be separated, which was done using this entity and calculations that are added to the semantic model.

In total, every different primary key within the data that has additional data connected to itself has its own entity, with a link to the origin entities. The way that the semantic model is structured, results in a Third Normal form semantic model. This good as stated in the theoretical context, given that this increases quality of the data management drastically as well as reduces the number of possible errors to a low number that can be insignificant to most uses of data. (Lien, 2009, pp. 51–126**)** The Third Normal form and the Boyce-Codd Normal form are used most in companies as the quality to time and effort ratio is the most optimal. In the semantic model the only exceptions of the Third Normal form are the origin export documents of the data, as these contain a singular sheet with all imported data. The semantic model is made this way to decrease errors, keep clarity within the model, as well as calculating as much as possible within the model, resulting in less work for Eurofins FSS before being able to use the application.

4.1.2 - The application

The application made has been developed following a process of full freedom on the design and way of functionality, with regular meetings with managers and consultants within Eurofins FSS to advice on sector specific matters, as well as give usability insight about the consultants that will become the users of the application. With this way of communication, all necessary functionalities were implemented into the application, while keeping an outside perspective of designing a visualization tool in an efficient manner. Underneath, the application will be explained following an order of different parts of the application. The first part of the application is the toolbar and user interface at the top of the application. This is the main navigation hub for the application. Second of all the visualization of the Eurofins FSS data visualizations will be explained. The visualization of Eurofins FSS' data is divided into two overviews, one based on a sample level, and one on an individual parameter level. Lastly, as third the RASFF open data visualization will be discussed.

The Toolbar

The Eurofins Data Analysis Tool works around the use of filters to use big batches of data to come to a specific conclusion within the data. To aid with different views and keeping filter functions logical and user friendly, every page of the analysis tool has the same toolbar on the top to create a familiar way of working in the system. The reason that this design is chosen is due to the theory behind familiarity and change blindness. (Stehle et al., 2020) Small changes per page can result in misinterpretation of data due to blindness of subtle changes. Next to this, using a toolbar is familiar to the users given the repeated exposure to this form of design when using websites such as web shops. Underneath the functions of the toolbar will be explained. Underneath a figure (figure 4.1) of the Data Analysis tool can be found with the toolbar highlighted.

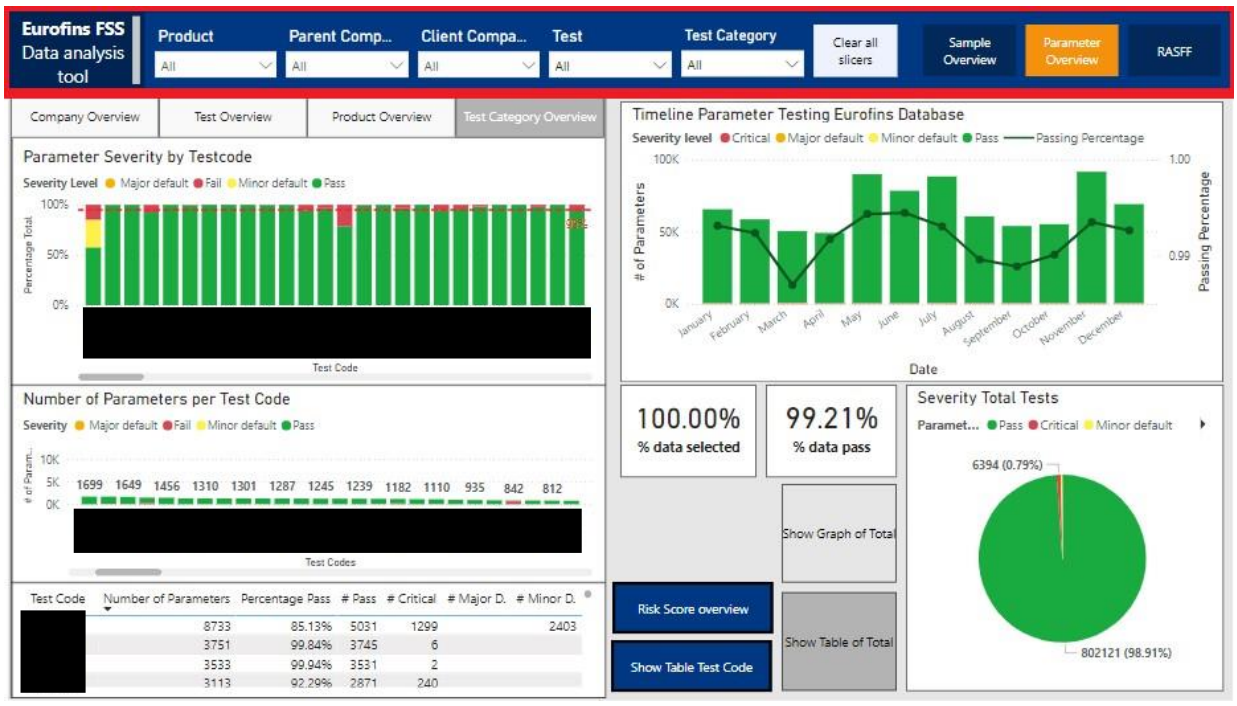


Figure 4.1: Overview of the dashboard with the toolbar marked in red.

In the toolbar, the most important thing found is the row of the five main filters which are used within the visualization tool to filter all the present data. The filters function at the top of the page is the main way of selecting what type of data the user wants to see, resulting in being the main tool used to look at specific important case-based data, such as results of certain products or companies present. These filters apply to all visualizations and are globally activated. To increase usability, the button layout has been color-coded, and functions such as ‘clear all’ have been added for an easy-to-use filter system. The page navigation is standard button layout with a vibrant color for an active page for ease of use of the average user. The order of the buttons is decided based on importance. When someone reads a view, they will take a natural flow of either left to right or top to bottom, given the reading order that most Europeans are used to. For the toolbar, the main function is the row of filters, therefore located on the left side of the toolbar. The page menu is less important, as this is something that will be used less in analysis situations, rather before starting to do analysis, and is laid out with vibrant colors to jump out from the toolbar. (Spotorno & Tatler, 2024)

Eurofins FSS Overview

The main body of the dashboard tool is the sample and parameter overview that encapsule all Eurofins FSS’ testing data, with the most useful part of the data analysis tool being the sample overview. The sample overview is the main overview, giving a concrete overview of all data on the scale of samples tested, which is the main scale used to do analysis within Eurofins FSS. The parameter is a more specific overview which can be used to get parameter specific data. A sample test can have ten to thousands of parameters to be tested, which results are saved to see what part of a test or sample actually the problem for a company or sector is. The parameter data is too specific to be consistently useful for big data, as the groups of data are too specific resulting in low numbers of results to be analyzed, showing results which are not statistically significant. The parameter overview and the sample overview are designed in the same way to keep familiarity, with only a few differences per page which are a necessity to capture all information on different levels, The design philosophy comes through the theory of placing the same functions on the same location to give the result of better muscle memory on where certain functions are located as well as limiting unexpected change to give less thinking steps to the user. (Grabinger et al., 2023) For this reason, both overviews will also be explained at the same time.

The Eurofins FSS data visualization screens are divided into a few parts to make a clear distinction between the different parts of the visualization software. Looking at the visual below, both the sample overview and parameter overview are based around the same principle. The screen division stems from the principle of reading direction and the clear grouping of functions to instill logic in the user. Using a clear distinction between left and right by grouping the graphs with borders, results in a clearer working environment for the user. (Grabinger et al., 2023) (Van Geert & Wagemans, 2023)

Underneath the division of the analysis screen can be found:
- The Right side is for standard statistics *(Red)*
- The left side is for looking through data via specific graphs based on X variable *(Orange)*
- The buttons at the bottom are extra functionalities *(Green)*

This will also be the division between paragraphs of text. The screen division can be seen below:

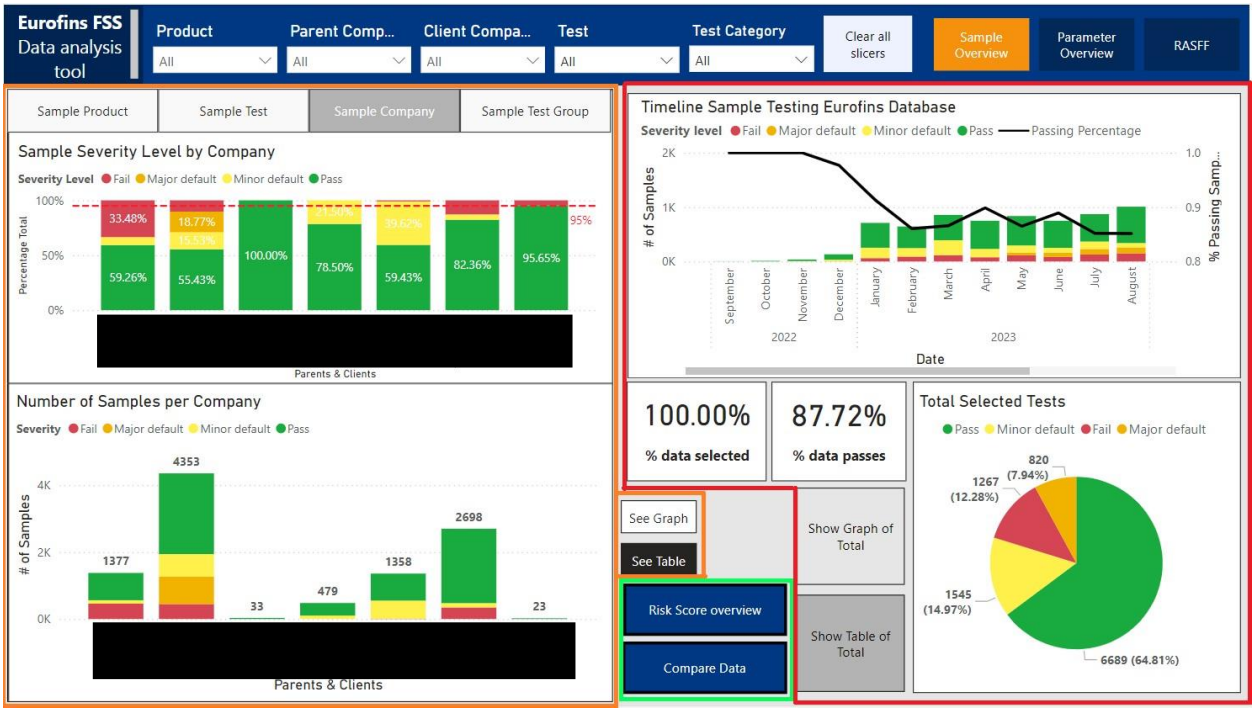


Figure 4.2: Overview of the total dashboard with colored zones showing the different parts of the visualization software. Orange to the left, Red to the right, Green in the bottom middle.

Right side overviews: Standard visuals

The right side of the screen is meant to show visuals that change together with filters, showing basic statistics which are easy to use to overview the total dataset or to export to show to clients for reporting purposes, as well as usage to make general reports about the different food sectors as briefing. Variables such as test pass rate, severity grade of tests and more are calculated to create this overview depending on the filters that are used. For both the Sample overview and parameter overview, the given information is completely identical.

Underneath a screenshot with numbering can be found with all de visuals numbered. Per number the visuals will be explained shortly.

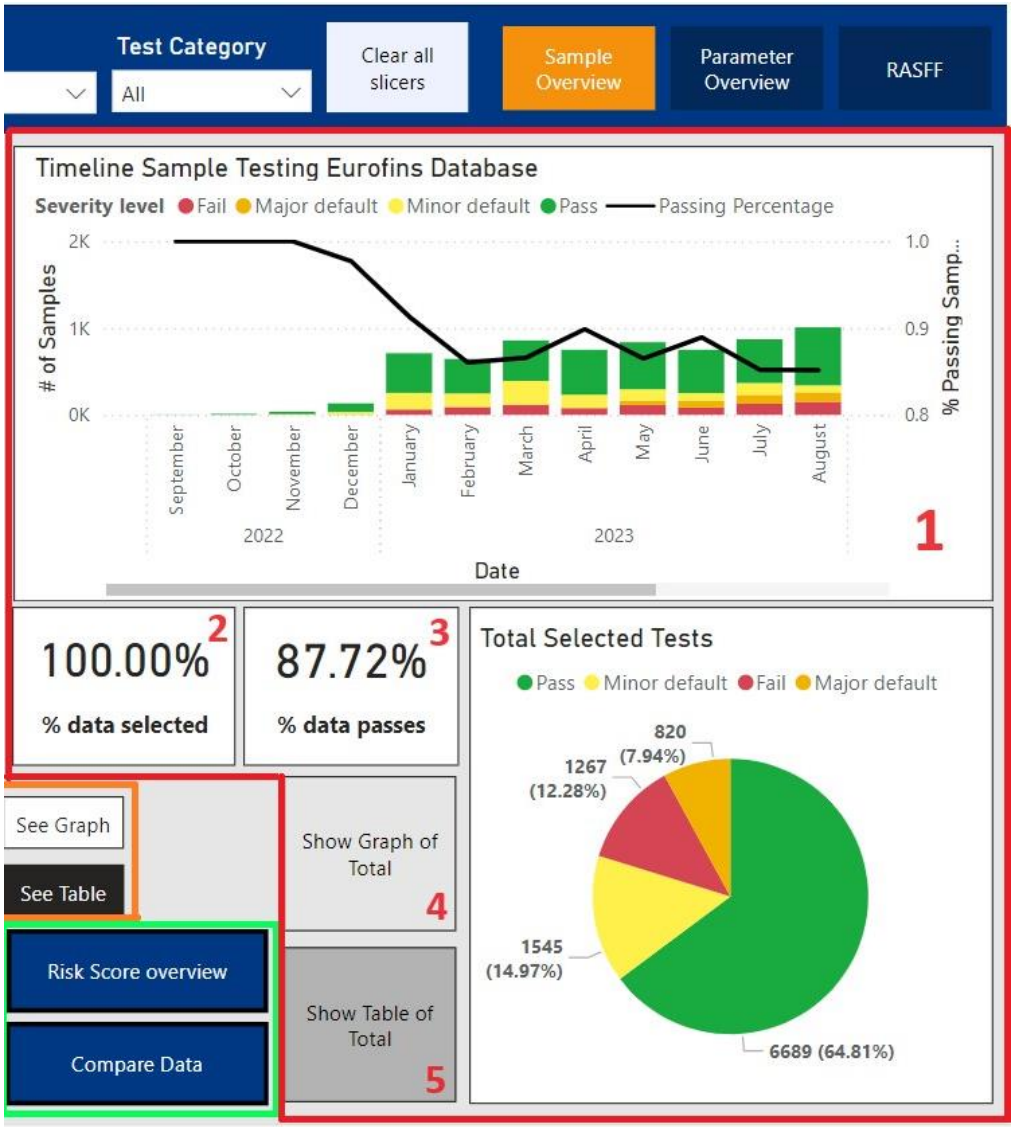


Figure 4.3: Overview of the right side of the dashboard with standard visuals for all data of the Sample based dashboard.

1. In the top right visual a timeline can be found that shows the selected group of results severity score in testing based on time. This visual can look at total years (drill up) or at monthly values (drill down). The worth of this visual lay within the possibility to see the number of tests done per moment of time, as well as the distribution of severity grade over time, filtered by filters selected. The results of this can be used to see the increasing and decreasing frequency of certain failing tests and parameters. This is very useful in a dynamic testing sector such as food safety, as the priority for research into food safety keeps changing with the changing severity of certain parameters over time. The graph design was taken as the graph both gives insight into both percentage of tests, which is very useful to see due to percentage being a universal way to look at data errors, as well as giving the amount of tests done in a timeframe, to check the amount of entries of the data to see if the test is performed a lot, and if the percentage of errors is based on a significant number of tests to give a clear conclusion based on the data.

2. The Percentage selected graph shows the percentage of data that all visuals still use after using filters. This visual helps with keeping track with the relevancy of the results found in the visual, as a conclusion found using 0.05% of all data is less trustworthy than one based on 50.0% of the data. This graph was created due to the feedback given by Eurofins FSS about tracking data relevancy in the application, as this was a problem that occurred in the testing of the program. This graph should lower the amount of change blindness that occurs when selecting different filters by making clear that the selection in the dataset has changed, resulting in a better realization that the dataset is changing. This visual is calculated by taking the selected datapoint and dividing though the total amount of data stored in the semantic model. (Grabinger et al., 2023) (Stehle et al., 2020)

3. This Visuals shows the passing rate the selected data. So, if the data is filtered on a company, it is the total passing rate of the company. This is a fast way to see or copy the percentage of passing tests, resulting in the user not having to calculate important data themselves. This visual is important as an easy to see end conclusion, as this percentage is the main performance indicator of food testing in Eurofins FSS. The visual has enlarged numbers together with visual two, to give a centered and jumping out design so that the results are clear for the user. (Grabinger et al., 2023)

4. Total Selected Tests (Graph) shows the division of severity grades of the selected data in a pie chart. This graph is easy to interpret, together with giving a clear overview of the percentages of samples with various levels of severity. This data is frequently used in communication with clients; therefore, this graph is created to aid in having visuals of the data present for presentation.

5.Total Selected Tests (Table) shows the same data as Total Selected Tests (Graph) but than in a table. It is an easy overview of the severity of tests scores, making it easy to see how a certain group of products or companies is performing on average. The table is open for copying data, resulting in a fast and high usability to copy and rapport the quality of certain companies and sectors to clients.

Overall, the information in both visual four and five is very simple yet useful. Due to the information being nice to have for showing to clients, showing the ratios of pass and fail test clearly, the table is more useful overall to copy all clear data. Out of meetings came that having a simple visual is nice to use towards clients, therefore the buttons to switch from graph to table were created. The visual overall is not a visual what needs to be analyzed to find patterns within data that give insight into the testing field, therefore the graph is less important and only used with pre-determined intention, this is why the visuals are place in the bottom right corner. (Spotorno & Tatler, 2024)

The total layout has been structured with all visuals separated from each other, as they are not all directly connected to one another. This design should keep this idea of all graphs working separately for the user. The borders in between the graphs also result in less visual clutter than connected graphs, which increases the clarity of the dashboard. (Nieto et al., 2022)

Left side overviews: specific data overview based on X variable.

Looking at the left side view of the analysis tool, both the sample overview and the parameter overview have the same graphs and visuals, but other slightly other Layouts.

The focus will be the set on the sample overview as this is the overview that yields the most results when it comes to the analysis.

Underneath an overview of the visuals/buttons is given per number to show the similarities between pages. After the visuals will be discussed per number.

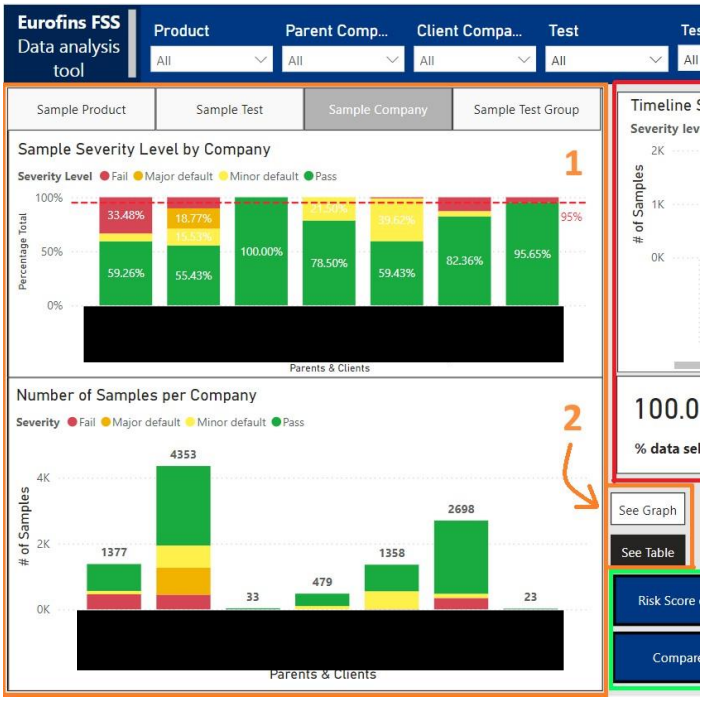


Figure 4.4: Visualization of the specific data overview of the Sample based data dashboard. No table is present but buttons are present to change view.

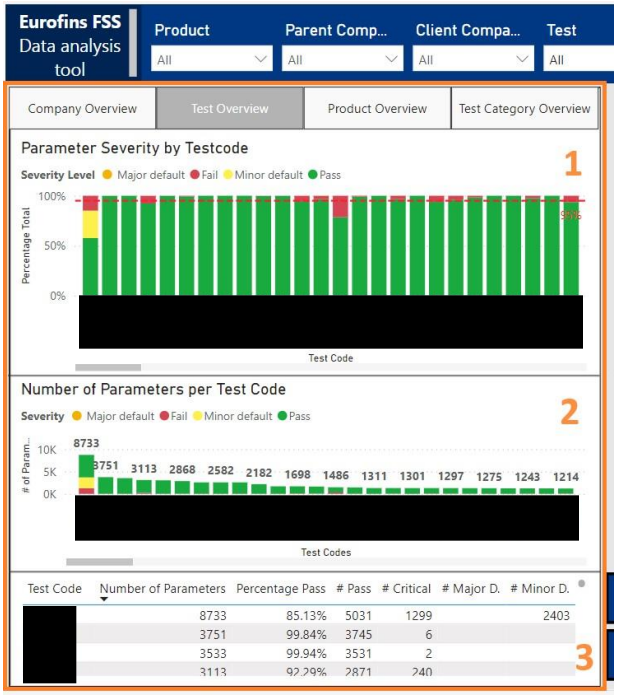


Figure 4.5: Visualization of the specific data overview of the Parameter based data dashboard. Both table and graph are visible together.

The difference present is that on the sample overview, the table and lower graph can be switched in view via a button, whereas this is not the case for the parameter overview, where both visuals are shown stacked on top of each other. The difference is present due to the different use cases. The sample overview can be used for general research into the dataset, or to create a testing plan for clients, whereas the parameter overviews use case is specific points within the data. Because of this, the graphs of the parameter overview can be made smaller, due to already having a directed search with less information in the graphs and tables. Given that the graph and table are on the same page in the parameter overview it is easier to pinpoint important information due to not having to switch in between views.

When looking at the left side view, it is important to know that this view is based on a tab system. The left side is divided into three visuals, with multiple different tabs that have different information that is used as the main value. This main value will be called value X in the explanation about the visuals. Every tab has the same graphs, based around different values.

The values that can be used to analyze the data by using the tab system are:

- **Company overview:** here both parent companies and client companies can be analyzed using drill-down functions.
- **Test overview:** here test codes can be analyzed.
- **Product overview:** here product groups and loose products can be analyzed using drill-down functions.
- **Test category overview:** here test categories/groups can be analyzed.

The categories chosen to be seen in this overview are a result of the feedback and interviews resulting from both Eurofins FSS management and consultants.

The reason the overview is structured this way is to find the balance between having important information clear on screen while not cluttering the screen too much. The clearest way with the highest usability is using a tab system such as the one used which is located on the same part of the page no matter what. Using a tab system is ingrained in most people after years of digitalization, making it easy to use. With this layout, all related graphs can be seen at the same time, which is needed for a clear view of the data. Secondly, the graphs needed to be big enough to capture all data clearly for analysis. (Grabinger et al., 2023)
Underneath all graphs are explained:

1. **Severity Level by Value X**

This visual is a percentage of total visual, showing the selected value X in columns of 100% per group. This visual shows the percentage division based on the selected tab, showing the actual rate of passing, without blinding the viewer with statistics that are skewed due to the number of tests differing.

2. Number of Samples per Value X

This Graph shows the amount and division of the selected values groups in severity levels. This graph is used to back up the previous graph by showing how many tests have been done in a certain sector.

3. Table Overview Value X

This Visual is a table showing data divided based on Value X. The value it shows are the passing %, The number of Passes, major defaults, minor defaults, fails and critical values, as well as the number of results per grouping of the selected tab. This table is an easy to read and copy overview of a group of tests, company, or sector, which is good for making overviews on paper about a certain sector.

The visualizations that have been chosen as this was the data that Eurofins FSS wants to look at in a more efficient way. The graphs are chosen to complement the sort of data it is. Using a percentage graph with an adjustable threshold line to look at anomalies and using a column chart to look at amounts in order of most used to least used, with the ability to filter down to individual results using the filters on the top. The order of the graphs and tables are determined based on importance from top to bottom in the reading direction. (Spotorno & Tatler, 2024)

Button Overview: Extra Functionalities

Both the Sample overview and the Parameter overview have two blue buttons each with crucial functionalities. They are located at the bottom of the screen. The buttons are big and blue to jump out of the dashboard, so they do not disappear in the background of the design. This is important due to the important functions that are behind the buttons. The functions behind are too big to fit in the main screen and result in visual clutter, therefore the buttons are used. In this paragraph these buttons will be discussed, first explaining the Sample Page Buttons uses and calculations, then the Parameter Page Buttons.

Sample Page: Risk Score Overview ***(See figure 4.6 on the next page)***

This button shows a matrix per category of the risk matrix value of that certain group. This value is calculated using the data of the last 365 days, taking the frequency of non-passing values as the measure to calculate the risk matrix value as explained in the theoretical context. This number is an estimation of the chance of a failing test being present in a certain group. This matrix can be used to argue about the level of risk behind a certain test or product group. The data used is only the last 365 days to keep relevancy of the data accurate. The higher the number the higher the chance of a violation as expected from recent historical data.

This function is important as the use of the risk matrix created by Eurofins FSS is main source of data when making decisions within Eurofins FSS when it comes to food risk management. During the process of developing the dashboard, the request came to implement an automated risk frequency matrix tool. With this addition later into the process and the importance behind the functionality, a space needed to be found to develop this feature. Due to the process this function is a button-based screen cover that uses numbers and a green to red color gradient to clearly indicate the risk frequency per product per group, using the psychology behind perceiving green as good, and red as bad. (Kawai et al., 2022)

Sample Page: Compare Data View ***(See figure 4.7 on the next page)***

Compare Data opens a window with a similar visual structure to the left side of the analysis tool. This window is a fully independent working window which view can be altered by using its own filters. This window gives the possibility to compare data between two different selections. This function can help with seeing differences between datasets, to conclude comparison against other data groups or the average. This function was thought of during testing an earlier version of the dashboard within the company, coming to the conclusion that being able to compare certain parts of the data would be ideal. As for the testing, this was possible when opening the application twice on multiple screens. To solve this problem an internal comparing function was created.

Parameter Page: Risk Score Overview ***(See figure 4.8 two pages down)***

This button opens a page which shows how parameters are performing in their test group, as well as what test group the parameters are in. This is a function useful if you want to dive deeper into the data after sample data is not being precise enough. This data view is calculated using the frequency of errors during the last 365 days.

The functionality of the parameter risk score overview was created as an addition on the sample risk overview. With the created risk matrix on sample level the interest in find out what parameter are behind sample errors was created. The use of this function is a bit niche and not for general analysis, only rather specified searches, yet the insight it gives is concrete about what parameters are the reason for failing safety test. This precision is the main value that it brings. Given the niche use cases for narrow searches, the function is well stored behind a button to reduce clutter unless it is used with the prior intention of looking at specific data.

Parameter Page: Show table Value X ***(See figure 4.9 two pages down)***

The last button on the Parameter page is a button that opens a window to give a better view of the table located at the bottom left. As the view is not as important to be seen in general the table is small, yet with the big amount of data that is present in the analysis tool, it is wanted to have an option to make the table bigger and get a clear overview of the table when searching for specifics.

This button is mostly available for two reasons. First of all, accessibility, to make the small table better readable in case that someone does not have a focused reason yet to search precisely in the data, and all data needs to be shown clearly. Secondly to keep the same balance as present in the Sample overview. Having one blue button less in the bottom throws off the balance of the dashboard due to the changing view for the user. This lowers the feeling of familiarity, as it is a small change that feels off. (Grabinger et al., 2023)

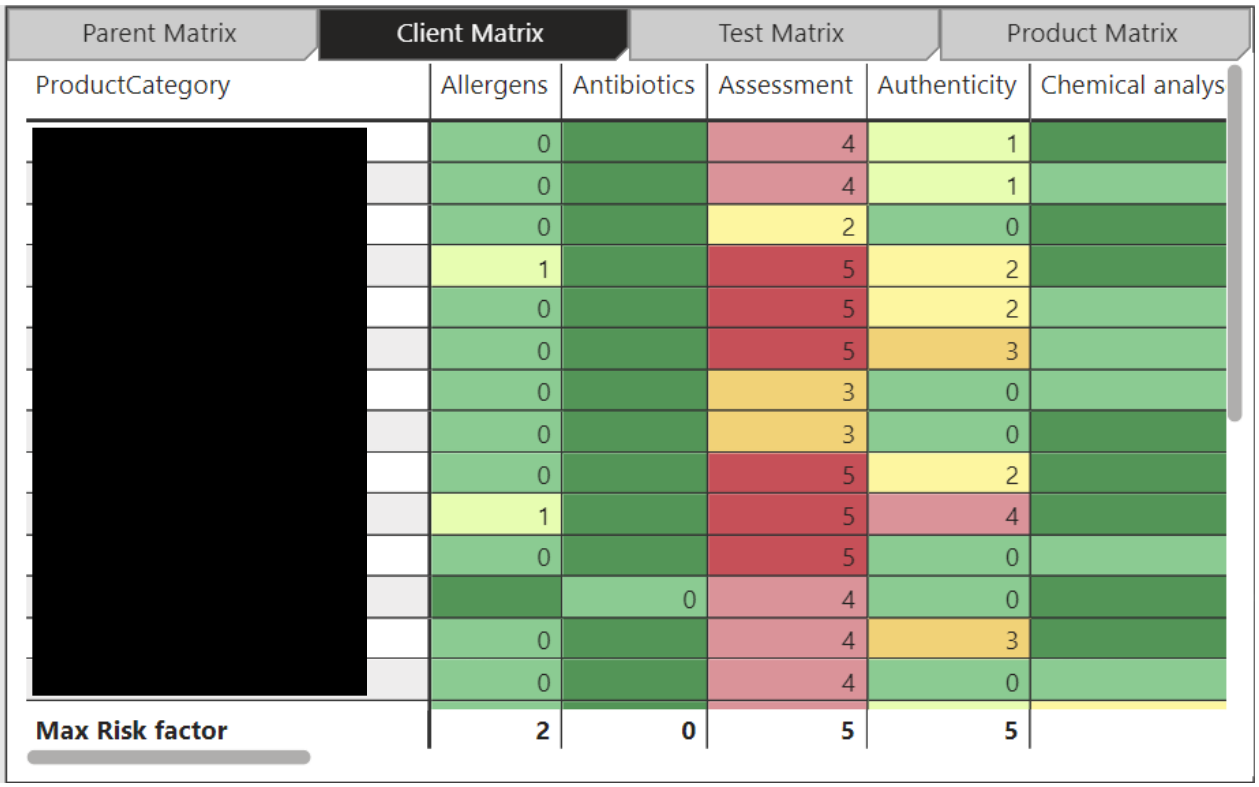
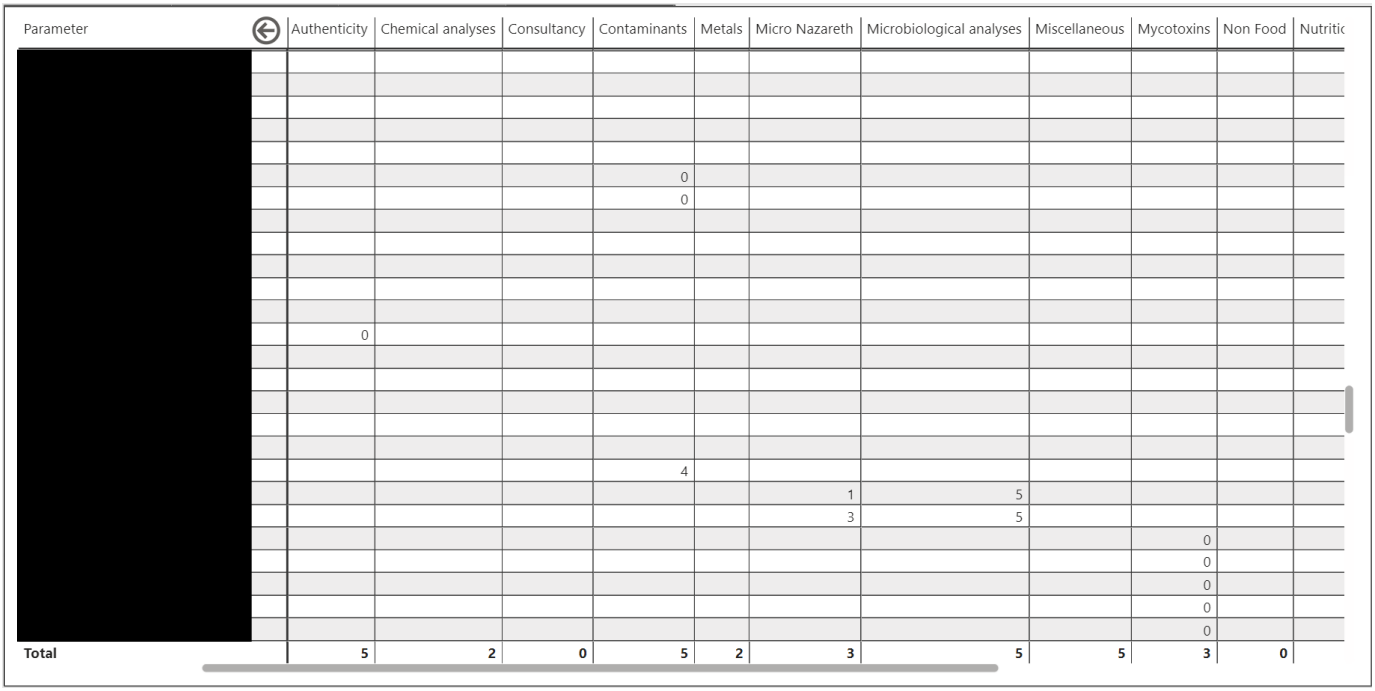


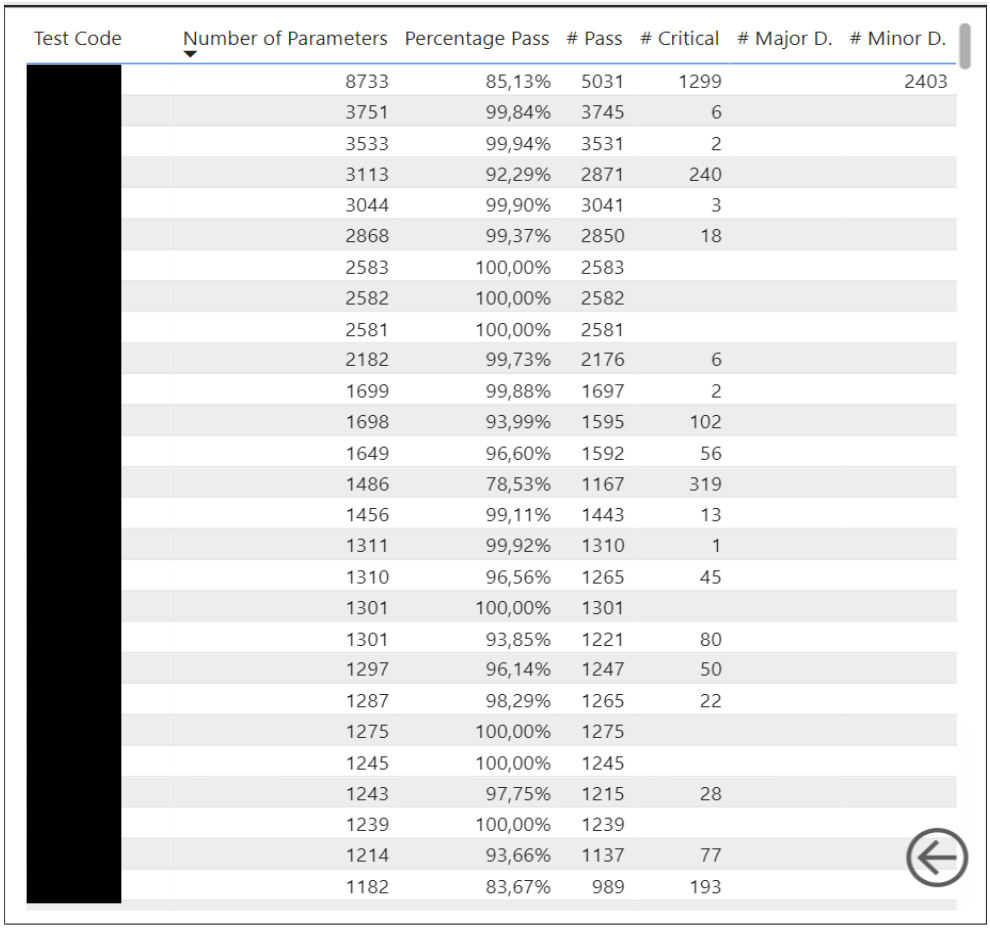
Figure 4.6: Sample Risk score overview visual. Scores represent the frequency of a violation happening per group based on the risk matrix concept.

Figure 4.7: Compare function Sample overview. The two visualizations can show different data resulting in the possibility to compare different parts of the dataset to come to conclusions easier.



Parameter	Authenticity	Chemical analyses	Consultancy	Contaminants	Metals	Micro Nazareth	Microbiological analyses	Miscellaneous	Mycotoxins	Non Food	Nutritic
[redacted]											
				0							
				0							
	0										
				4							
						1	5				
						3	5				
									0		
									0		
									0		
									0		
									0		
Total	5	2	0	5	2	3	5	5	3	0	

Figure 4.8: Risk score overview per parameter. Giving what group of tests said parameter fall under and its frequency level.



Test Code	Number of Parameters	Percentage Pass	# Pass	# Critical	# Major D.	# Minor D.
[redacted]	8733	85,13%	5031	1299		2403
	3751	99,84%	3745	6		
	3533	99,94%	3531	2		
	3113	92,29%	2871	240		
	3044	99,90%	3041	3		
	2868	99,37%	2850	18		
	2583	100,00%	2583			
	2582	100,00%	2582			
	2581	100,00%	2581			
	2182	99,73%	2176	6		
	1699	99,88%	1697	2		
	1698	93,99%	1595	102		
	1649	96,60%	1592	56		
	1486	78,53%	1167	319		
	1456	99,11%	1443	13		
	1311	99,92%	1310	1		
	1310	96,56%	1265	45		
	1301	100,00%	1301			
	1301	93,85%	1221	80		
	1297	96,14%	1247	50		
	1287	98,29%	1265	22		
	1275	100,00%	1275			
	1245	100,00%	1245			
	1243	97,75%	1215	28		
	1239	100,00%	1239			
	1214	93,66%	1137	77		
	1182	83,67%	989	193		

Figure 4.9: Show table Value X, the button that shows the table in the bottom left of the parameter overview in a bigger and clearer format.

RASFF Overview

RASFF is a European institution that deals with food testing errors. Within RASFF European regulation violations get documented, open for companies and individuals to use within research. Given that the data from RASFF is trustworthy, the RASFF data can add a new layer of vision to Eurofins FSS perspective of food testing within Europe. Therefore, RASFF data is implemented in the Analysis tool as an extra overview to add to the depth of different data sources for Eurofins FSS to research. The RASFF overview shows all RASFF data usable for data analysis in Eurofins FSS' sector.

Eurofins FSS has used RASFF data to help with their consultancies and data analysis over the last few years and have been looking at integrating the RASFF data directly into their analysis data. Using multiple different analysis screens and software is not efficient to either use or learn, therefore integrating both views into a singular application is the best course action. This extra implementation that Eurofins FSS has been hoping for has the major benefit for consultants, with data ready in the application instead of needing to download all the RASFF data separately constantly.

An important thing to note is that RASFF does use different notation for the violations they collect than Eurofins FSS. This results in slightly differently structured data, and the use of different filters to filter the data consequently. The filter function works the same as the Sample and Parameter overview yet is stated slightly different using different categories.

Another thing to keep in mind with RASFF data is that all RASFF data is taken from a violation database, resulting in almost all data points being violations (not passing instances). This means that percentage-based tests and visuals showing percentages do not have any merit and are for that reason not included. Frequency compared to other moments in time however are a good way to see changes over time, as the number of tests does not change drastically, yet only the violations get documented by RASFF.
RASFF data can be used to confirm possible conclusions made on other data, plus it can give insights on what sort of data to look out for in the Netherlands.

Underneath the layout of the RASFF Overview can be found:

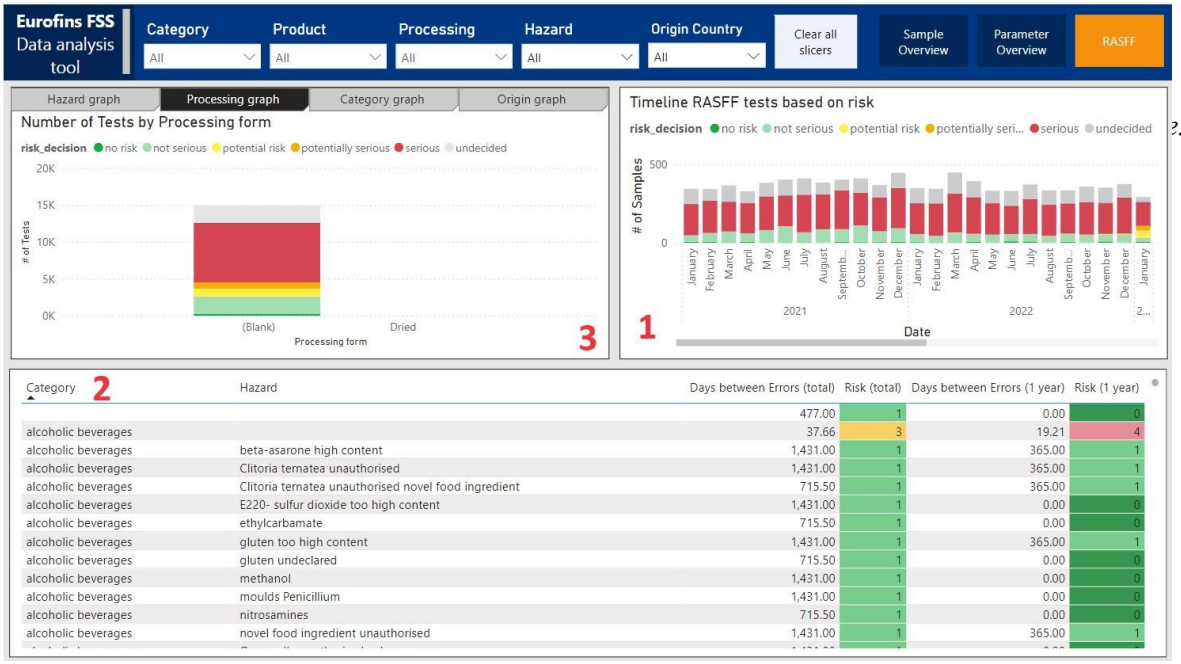


Figure 4.10: Overview of the RASFF dashboard with numbers per visual.

1. The timeline of RASFF data is the visual that shows the timeline of data from RASFF over the last 3 years. with filters this timeline can be used to collect important data about subjects such as if there are changes per season in certain groups of products, or if there are changes over time in general in certain aspects, and to see if the volume of tests has changed of RASFF.
This graph is in place to check if certain errors are showing up more or less over the years, as well if certain countries for example are doing better with the amount of exceedances of legal limits the last year compared to before. Based on this, clients can get insight into what business relationships can be more likely to give problems in food safety.

2. RASFF data table, this table investigates the mix of products and hazards that is mentioned within the RASFF database. This table takes the product group and lists the hazards it can have behind it. After that it lists the frequency of violations per hazard per product group, as well as the risk factor that is connected to this value. The graph shows both the risk value and frequency over the total database import (3 years) and the last year (365 days) to see if in short term changes are visible per product group or hazard. This can show important trends within the data based on time-related events. Having data from two sources feels more trustworthy for clients to understand the problem that exists within the food testing sector.
This graph is a mixture of a risk matrix as well as an information table about what hazards are present for what type of products, therefore the visual shows a lot of information. Given the large amount of information, and this visual being the main visual with all information put together in one table, most of the screen is used for this visual. This visual is intended to be used as a searchable table, which is searched based on a predetermined product or interesting information found in the other visuals, therefore this visual is placed at the bottom of the screen, as it does not need to be the first visual the user looks at. (Grabinger et al., 2023) (Van Geert & Wagemans, 2023) The reason all this data is taken together is because it creates a clear systematic table, which is concise and not scattered around the visual. Given that the data source is external, the understanding and clarity of the data source of the user is expected to be lower, therefore grouping all important data together is chosen as the best way to make the information the most usable.

3. Number of Tests by TAB, this visual is a tabbed visual looking into ordering the total data within groups to analyze a layer deeper. This visual has four tabs that can be used to show the data, these options are:

- Hazard
- Processing
- Product category
- Origin

Using filters and the mentioned tabbed column chart that shows the different groups of data, different combinations can be made to gain proper insight into relations such as the sorts of products together with origin country, or the effect of way of processing on hazards. As of right now this graph does not have the greatest impact on data analysis, yet this graph is the start of the following step within Eurofins, where certain newer added details of foods are introduced to the database, such as origin country and Processing form. After these new details are developed, this graph can look at the European data of a multitude of attributes. This graph is the first sign of the next big step that Eurofins FSS wants to take in data analysis; combining external datasets with their dataset, working them out to be compatible with one another, and creating a bigger and more wide spread dataset, strafing to create a data analysis system that looks into relationships in food safety through the whole of Europe.

4.1.3 - Usability

The application created has been designed to be used regularly during research for consulting clients from Eurofins FSS. Given that this process needs to be as efficient as possible to have a high value for the company, a low entry level for users is required to give the application use. During the creation process of the application, many usability factors have been discussed within the company to design an application that suits the users within Eurofins FSS the best. Underneath the measure taken to increase the usability of the software can be found.

Universal Design

Looking at the application, one of the first principles of design is making an application that needs to be used for a longer period of time similar in many regards, so that all parts of the application transfer knowledge on the usage of the application as a whole.

The first step taken to keep the design simple and universal is the toolbar at the top being the same for all different pages. The only exception is the filters on the RASFF visualization page, as the data is stored differently. Next to this, all the functionalities are the same for all pages, resulting in familiarity in the use of the application.

The concept of familiarity has been extended to the layout of the visualization pages as well. When looking at the Sample Overview and Parameter Overview, both pages are designed in similar fashion to create a familiar and natural user interface. The layout of all pages is structured under the same principle of having timeline-based graphs in the top right, having changeable graphs with more detailed views at the top left, and more detailed points of interest ranging over the whole screen to capture the view of the user. To emphasize the importance of the results which can be found in the visualization. The graphs all use the same green-to-red color scheme to make it clear how severe the exceedances are, which comes over natural due to the ingrained connections between colors and meaning. (Kawai et al., 2022)

User Manuals

With the Application, two different manuals were created to ensure that all information that might be needed is documented for the user to read. The two manuals that are created are about the usage of the visualization tool, and the steps that need to be taken to create the dataset that is used within the visualization tool. An altered User manual without company data can be found in appendix E. The step-plan for data preparation is not able to be found in the appendix due to the step-plan being too closely related to private data which is not publicly available from within Eurofins FSS.

The User manual of the visualization tool is a fourteen page document which goes over all of the different visualizations which as able to be seen within the visualization software, as well as the important Power BI function that are important to know to use the visualization tool to its full potential. The manual should work as a document which can be used when knowledge about the visualization tool is missing, and functionalities are unknown to the user. The manual is not directed at finding conclusions directly from data inserted, rather the user manual is aimed at explaining the user how to use the application to the visualization’s potentials, to be able to create conclusions themselves.

Dynamic Semantic Model

The created application has a dynamic semantic model, meaning that the visualizations and data links automatically update when new data is added to the semantic model. This means that the data structures do not have to be recoded every time a new dataset is uploaded to the system. Additional data that is added to the system will automatically be fully worked out and cleaned as the semantic model uses, resulting in a high usability when a company changes its naming structure, or adds new variables to their internal systems. Additionally, newly added data attributes can be imported into the application without resulting in errors. The data will not be used until the application is internally updated, yet while making changes to the way data is saved or used within Eurofins FSS, the data visualization tool will keep working with new data as intended.

Specific functions within Power BI

One of the downsides of the visualization tool created is the usage of specific functions within Power BI to create the optimal visualization application. In Power BI there are a decent number of integrated functions that are helpful when making a visualization application. These functions are standardized possibilities within Power BI, which results in a familiar experience for Power BI users, which is a good aspect of the usage of Power BI functionalities. The problem with Power Bi functionalities is that for a new user, all icons and functionalities can feel very unnatural. Power BI has a lot of tools which help data visualization, yet these tools need to be learned by heart, which does not help when introducing a new application to a bigger group of employees.
To limit the problems regarding the usability of Power BI functionalities, All useful functionalities are mentioned within the user manual given to the user, yet this does not solve the inherent problem of the functionalities feeling unnatural, and therefore harder to learn.

An upside that is present when using Power BI as analytical software is the constant uptime of the analytical software as well as the high accessibility. Power BI can be accessed on all devices and accounts that are allowed by security to access the data. This can be done from anywhere in the world at any time of day, which is a great upside when working with a lot of different clients in different locations.

4.2 - Steps of integration

With a finished application, the last step in data research is integration. The development of the data visualization tool has been separated from the standard workflow of both monitoring teams and consultants to ensure a steady workplace for the employees of the company. To integrate the created application, changes will have to be made in the workflow to ensure usage of the visualization tool. During the research period, steps have been taken to implement the created software to enhance the data collection part of the consultancy process.

IT Team Transfer

With the finished research, the information gathered, and systems set in place have gotten significant value for the company, resulting in the need of proper transfer to the company. A transfer meeting and track has been set up, to transfer all human knowledge and all applications to the IT team of Eurofins. This meeting resulted in most of the data being transferred to the IT team of Eurofins, together with keeping me as contact point for integration during the following week. All information including User manuals and cleaning scripts have been transferred during the transfer period, which lasted a week. Within this transfer period, conversations have been had about the improvement steps which are possible looking at the near future, which will be talked about more in detail in paragraph 5.2 ‘Improvements’.

New Project Members

Given that the application has a lot of use within the company, and the change of workflow within Eurofins FSS is guaranteed over time with changes in employees, a few employees that have worked with consultancy and IT have been assigned to be the first users of the application that has been developed to improve the created dashboard. These people have been fully briefed on the application, from how to use the application to how to edit the application. Some people have only been briefed on the application after creation, whereas others have been taken through the process of designing the application. In total this group of employees will be the group which will test the application, together with further development to enhance the tool created. As a starting point, bi-weekly meetings have been set up to discuss both the IT side of the application as well as the usability aspect.

Chapter 5: Evaluation

This chapter Evaluates over the course of the creation of the application together with the result that has been created. This is done via internal interview within Eurofins FSS, as well as by reflecting on the created dashboard.

5.1 - Managerial Evaluation

To evaluate the dashboard created as a result of the research done within the company, an interview was with the operational manager of Eurofins FSS was held. In the interview important topics such as the worth of the dashboard for the company and further plans for the dashboard have been discussed to review the value of the created dashboard. Underneath the important points from the interview have been summarized.

Progress in data analysis of Eurofins FSS
Looking at the progress of data analysis, Eurofins FSS was looking at diving further into using data actively within the company. One of the first steps to the increase of data usage at Eurofins FSS was the hire of university students that will do internships. The goal of Eurofins FSS was to make use of almost all historical data that Eurofins has, yet to what extent is still an iterative process, as there was little active use of historical data before. With the dashboard the first big step in data usage has been made within Eurofins FSS. The creation of the dashboard has shown the possibilities that are present with the use of data, for example the automation of the risk matrix that is used in the company. The application is still not automated, which is the end goal of Eurofins FSS, yet the application has been created resulting in the first multi-functional dashboard to get data insights. Eurofins FSS is overall very happy with the insights into their data which are possible using the dashboard.

Implementation and integration
The implementation and integration of the application within Eurofins has been discussed internally and has been planned over the course of a three month to half year time period, with a few to standardize the dashboard and make it applicable to all Eurofins FSS laboratories within the Netherlands. A reoccurring problem with internship based expansion within Eurofins FSS is that when an intern makes usable software or models, the knowledge about the use and further expansion on the topic gets lost after the intern leaves the company. In the case of the created dashboard, an internal project group is set up and an IT team is planned to be created to improve and standardize the dashboard for use on a bigger national scale.

Updates to the System
Employees from Eurofins have the overall conclusion that the dashboard created has all basis functionalities wanted to start using data analysis within the company. The focus points of updating the dashboard are the following: 1. Researching if other variables should be included in the dashboard. 2. Automating the data insertion within the system from the database. 3. Changing the layout to a more Eurofins known layout.
All the named aspects are seen as satisfactory for now, yet for longer time use updates to these aspects are seen as favorable by Eurofins FSS management. Given the way the dashboard was designed for automated data insertion and analysis, the dashboard has already been automated to refresh data, as well as the application already being changed to a more Eurofins FSS known layout.

Overall conclusion on the application

Overall, the dashboard is more than sufficient for what the management of Eurofins FSS expected. The dashboard has a lot of options and visually it works and looks good. Given that the project was made within a team of employees that do not have a focus in IT, the knowledge of how everything is computed within the system can sometimes be hard for new users. With all data now available there are some questions that arose looking at the design such as: What screens should be standard as display, and what screens should be accessible via buttons to increase the usability for all employees instead of the project group?

One change that has to be made is within the Eurofins FSS' categorization of data, as it is not fully compatible with data analysis, given that the data groups become too small. This was known following the research done yet still needs to be implemented.

Overall the Dashboard is a big step up from physical yearly rapports, and the first step within the company to create efficient data insights. Yet with the increase of available data, the increase in wanted functions also increase. Therefore, the project within Eurofins FSS keeps growing and growing.

5.2 – Future improvements

With the creation of the application, the discussion of when an application is done comes into the picture. With the creation of a visual data analysis application, the main focus of the conversation lies when there are enough ways to analyze the data. With the application created, the decision was made that the tool has been worked out as far as possible looking at time constraints and value-effort balances. Yet over a longer period of time, other additional implementations can be done to increase the value of the application made.

IT automation

First of all, the data used within the application is a dataset exported from a Power BI semantic model. Instead of importing the exported dataset into Power BI, the dataset can be linked directly to the semantic model of the application. This was not doable due to security, yet with IT overseeing the project, this implementation can be the first step to an automated analysis system.

Better Filter options

Secondly, the options of data filtering could be expanded upon. Looking at the filter options that are available now, most of the important options are there, yet certain datasets are hard to give the proper scope of data, therefore new labels can be implemented to segregate all data in a way which is most efficient. Next to this, being able to select a timeline of data that wants to be used within the analysis would be a nice addition. This option is not natural in Power BI and needs to be coded in using external code. This is doable on a bigger timescale, but creating this function without having any problems in data segregations is unrealistic on a small timescale.

Additional Analysis

Lastly, the increase in new ideas for additional analysis. While creating a data visualization tool, it is very common to get new ideas to pop up to implement while modeling the sketched design. This increasing number of ideas can hinder the process of creating the needed application, due to wanting to implement options which are not as useful, and therefore not worth the time to implement. In the end of the process of creating the application, a hand full of ideas of visualization was left behind which are more specific additions that can be useful to get to the total goal. This hand full of specific additions are the following analysis: An addition to see what results are coming from what specific Eurofins research laboratories, the ability to compare Eurofins FSS data with European data in one visualization compared to multiple screens and the ability to make the dashboard anonymous to be able to show the data directly to the customer.

Chapter 6: Conclusion and Limitations

This chapter will discuss both the total conclusion and the limitations of the performed research.

6.1 Conclusion

To conclude the research question of this research, a group of six sub-questions were formulated. For each sub question an answer has been found to find the conclusion of the research question. Underneath all sub questions are answered together with an overall conclusion of the main research question.

1. *What type of data analysis dashboards and designs are there to visualize data that fit Eurofins FSS' needs?*

When analyzing the theory of data visualization software, the first thing to research is what the requirements are for the company. In the case of Eurofins FSS the most important requirement are both usability, as a multitude of people needs to be able to use the application, as well as making being able to see an overview of all the different sorts of data in their respective functional way. Therefore, dashboards were a topic of interest by Eurofins FSS, as they are the most known and easy-to-use visualization applications used within the company to rapport data. Next to this it is also easily possible to show different sorts of data in one view or system. When looking at the different sorts of dashboards, it is generally discussed that there are three different forms of dashboards, with the analytical dashboard being the best suited data visualization tool for the case at hand. Looking at design, there are a lot of different ways to design an application. Design can be focused on multiple different main points, such as presentation, clear conclusions or having large amounts of data on one screen. Given that the dashboard needs to be used by a multitude of people without a high level of prior knowledge about data analysis, the main points of design are clarity and usability.

2. *What parts of Eurofins' historical data are most useful for Eurofins FSS' employees during the consultancy process?*

Within the database of Eurofins FSS, parts of the data are not complete due to the lack of standardization within the front- and back-end of the database. After analysis of what type of data was present, an overview of all the data was made, which was discussed in meetings. Using the knowledge found within the company the data structure of Eurofins FSS, it was concluded that the main points of interest within the data of Eurofins, the company from which the product is, the product specifics, the date when the analysis was, the tests which were performed on certain products, what parameters where tested and what categories tests are grouped in. This data is the base of data analysis within Eurofins FSS, due to it being the needed data to make risk assessments matrixes, as well as the data being needed to make client directed risk assessments based on company details.

3. *How can dashboarding and proper design best be applied to create the best data insight for Eurofins FSS?*

Using an analytical dashboard designed after clarity and usability certain design choices are crucial. The level of usability can be increased by making clear distinctions and functions within the application, as well as trying to keep familiarity as much as possible to increase usability over time. To increase clarity, it is important to format and title visualizations well, together with using the right graphics for the right

analysis to increase familiarity resulting in more clarity. Given that clients need to be able to understand all data as well when shared, clarity is a very important pillar which should be taken as the main design principle. Given that the amount of relevant data within Eurofins FSS is bigger than fits within one clear dashboard, a report design has been chosen, which encapsulates multiple different dashboards in the form of different pages based on different data levels, which are connected to create the best user experience.

4. *How should Eurofins FSS clean and manage their historical data to be able to dashboard their most important data?*

The dataset from Eurofins FSS has a great body of data yet has certain management issues that need to be addressed. All unused attributes within the data should be deleted from the database as this only lowers the data reliability and increases the possibility of errors. The second point is standardizing more attributes which are standardizable without system errors, resulting in less problems in grouping data within data research. More frequent error checkups should also be done. Research shows that there is a lack of knowledge of errors present within the database. Lastly to keep data refreshment within analysis applications as consistent as possible, both the user end of the database system, as well as the back-end of the database system should have one-on-one relationships between the data, so that no data needs to be inserted by hand due to certain terms being used in the user end of the company, yet not collected within the database.

5. *How can the design requirements be visualized best to create a functioning dashboard that supplements the data research of the consultancy process of Eurofins FSS?*

The visual and data requirements which are needed for creating a dashboard within Eurofins FSS range from security to usability points. First of all, security is very important. Eurofins FSS' data cannot be online on most web services, only excluding certain trusted sources. Secondly, given that the most important pillar of design is usability, the application should be clear to use for all users, as well as understandable for clients when shown. Looking at the specific visualizations within the analytical dashboard, the most important thing is compatibility with the existing way of working within Eurofins FSS. The data shown within the visualization application need to be transferrable to other systems within Eurofins FSS, such as the use of the risk matrix as end conclusion or using the severity scales created within Eurofins FSS. The application which has all wished for functionalities is Power BI. Power BI was previously used within Eurofins FSS, aiding in the usability of the application as well as certain people within Eurofins FSS already having used the application. Next, the application needs to be automatically maintainable. Given that time is scarce when it comes being able to maintain an application manually for a long time, a self-sufficient application is ideal to be able to use the output data, while not having a time drawback as an investment. This is ensured by meetings and a progress plan set up within Eurofins with the IT team to connect the main database to the application. With this the administrator rights have been transferred to the IT department of Eurofins to ensure proper management of the application. Lastly, every two weeks an evaluation meeting with both managers and consultants as well as the IT team is planned to see the state of the dashboard, as well as give feedback on the workflow. Due to these meetings, the design of the analytical dashboard could be fully customized to the best use case for Eurofins FSS consultants.

6. *To what extent does the created dashboard fulfill the needs within data analysis of Eurofins FSS' employees?*

A few steps were taken to evaluate the created dashboard. First the IT department takeover meeting resulted in the takeover of the application to Eurofins FSS data team to incorporate the last few changes to link the database of Eurofins FSS to the visualization application. In this meeting all specifics of the application were mentioned and the plan to implement the application were discussed. Within this meeting a bi-weekly meeting schedule was created between Eurofins FSS and the IT department to discuss the steps for implementation of the application into the workflow of Eurofins FSS. Next to this, a group of employees got tasked with the usage of the application to evaluate points that might still need to be redesigned or added by the IT team. From the last managerial meeting the conclusion can be drawn that the application has value for the company, yet the question of how to incorporate the new application into the workflow of the existing consultants is hard to answer. The value of the application is acknowledged as an implementation plan has been made with the IT team to increase the project size and implement the created systems into the workflow of Eurofins FSS.

With the answers stated above the main research question can be answered.

- *How can Eurofins Food Safety Solutions implement a data visualizing system to give the best insight in their historic data for consultancy purposes?*

To gain insight into Eurofins' historic data Eurofins FSS wanted to have a data visualization system to gain insight in their data. To gain this insight a dashboard application was created.

First, different sorts of historic data were chosen on base of company interviews and database analysis on the basis of what data is standardized and as enough use for the company to be implemented within the data analysis tool. Priorities of the application have been set by having regular meetings with the users of the application, as well as the managers of the department. Using a biweekly evaluation system, every two weeks new goals could be set for what can and should be added to the application, with new input coming from the side of the researcher, and feedback coming from the managerial side of Eurofins FSS. Using feedback loops and determined core principles determined by management and the consultancy department, a data visualization application was made that followed the key ideas of the company, with extra enhancements thought out by the research.

After creating the dashboard. Integration is the next step, resulting in a series of meetings and integration plans that are created together with IT departments, managers of Eurofins FSS and people working on the analysis project. These meetings result in an integrated, automatically updating visualization system that is connected to Eurofins FSS' database, ready to be implemented in the day-to-day consultancy workflow of Eurofins FSS.

The implemented dashboard gives the wanted insights into the historical data of Eurofins FSS by creating a full overview of all determined relevant data of Eurofins FSS. Using a multitude of different visualization methods to cover all possibilities of data insight. By creating a universal data analysis dashboard, all historical data can be visualized, making Eurofins FSS less reliant on other companies to do data visualization. The dashboard also helps Eurofins FSS by automatically making the wanted data graphs that are used in report making, and give an easy way to show data and data relevancy to clients when working on a testing plan.

6.2 Research limitations

During the performance of the research, limitations came into play that have impact on the quality of the research performed. Underneath all limitations are explained.

The biggest limitation to this research project is the time that is available to complete the research itself. With the small margin of time for the research being 10 weeks, finishing a big project with multiple aspects can be quite difficult. Due to this, the project needed to be scaled down correctly, resulting in less possibilities within the project. This limitation resulted in me needing to cut certain parts of the project out of the plan, such as looking in normalizing the database of Eurofins FSS, and looking into feedback about what clients' opinion was on the developed dashboard. Lastly, this also impacted on the hardware and software which was usable within the research, as getting new applications or hardware took crucial time (averaging 3.5 weeks to go from request to usable).

Data availability is the second point of limitation that was prevalent for the research being conducted. When looking at the data being collected and used in the research, it is important to note that almost all data that is visualized is data from within Eurofins FSS itself. This means that certain rules are bound to the use of the data. For example, a lot of data that is used within the research and creation of the application is anonymized within the research paper, as names of clients contractually cannot be named in either research or systems that are used or seen by other companies.
To add on to this, certain pieces of data could not fall outside of Eurofins FSS' databases due to company confidentiality, meaning that certain data cannot be used in for example this thesis precisely, while still being very important to the final product that is being created. This variable dependent classification policy resulted in a mix of hard to lead back rules and variables, which needed to be managed well.

When looking at the validity and reliability of the data that was used, a few important aspects of the data should be mentioned. The data used can be divided in two groups, data provided by the company and data gathered via interviews.
When looking at the data provided by the company, there are a few important points to talk about. First of all, the data given by the company is a database that is not normalized, resulting in a higher chance of errors within the database provided by the company.
Next to this the database given had a few problems that where known that needed to be resolved before being able to use the data reliable. A few non-confidential problems are problems such as empty attribute sets within the entity of the back-end, Blank entries that should have a standardized value, and a lot of non-standardized terminology, resulting in data groups under certain attributes that should be grouped together via a standardized group names where small mismatches occur in naming. This made the data provided prone to errors. The problems stated can be solved using steps of normalization and standardization myself, yet preexisting errors within the historical can possibly not be solved during the process of data cleaning, as not all database errors are known within the company. Expected is that little errors may still be present in the dataset, and given that the dataset is very big, the effect of singular errors is expected to be minimal.

Next to the given dataset, interviews and meetings have taken place to gather information on what type of data the user (analyst or client) would like to see back in the dashboard that was being created. The interviews were held in-person, with in-person consent within the introduction of the interview. When it comes to data gathered via interviews, it is important to know that this data is hard to quantify, yet pretty reliable. This was not a problem for the project, as the interviews were meant to gain an idea what types of data were the most important to be able to access easily within the dashboard, as well as what variables needed to be included in the dashboard, and which data points were less important for proper use. The interviews gathered opinions on the matter by different sorts of people with different roles, which may not have led to a more quantifiable outcome, but gave insight to the different sorts of people that will work with the system, which was be used together with theory to create a dashboard with a higher usability. It is important to note that the interviews done are confidential, as the information discussed within the interviews is about the business private business practices and data usage within Eurofins FSS.

References

Benešová, A., Hirman, M., Steiner, F., & Tupa, J. (2018). *Current trends in the management of diagnostic processes in testing laboratories*. 2018 International Conference on Diagnostics in Electrical Engineering (Diagnostika). https://doi.org/10.1109/diagnostika.2018.8526136

Caballero, I., Gualo, F., Rodríguez, M., Piattini, M. (2023). *A Maturity Model for Data Governance, Data Quality Management, and Data Management*. In: Naiouf, M., Rucci, E., Chichizola, F., De Giusti, L. (eds) Cloud Computing, Big Data & Emerging Topics. JCC-BD&ET 2023. Communications in Computer and Information Science, vol 1828. Springer, Cham. https://doi.org/10.1007/978-3-031-40942-4_12

Damyanov, I., & Tsankov, N. (2019). *On the Possibilities of Applying Dashboards in the Educational System*. DOAJ (DOAJ: Directory of Open Access Journals). https://doi.org/10.18421/tem82-15

About Eurofins scientific. (n.d.). Eurofins Scientific. https://www.eurofins.com/about-us/

Eurofins Worldwide Laboratory Testing services. (n.d.). Eurofins Scientific. https://www.eurofins.com/

Gao, P., Han, Z., & Wan, F. (2020). Big Data Processing and Application Research. *2020 2nd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM)*. https://doi.org/10.1109/aiam50918.2020.00031

Globa, L., Novogrudska, R., & Grebinichenko, M. (2022). The comprehencive approach to big data preprocessing. In *Lecture notes in networks and systems* (pp. 119–142). https://doi.org/10.1007/978-3-031-16368-5_6

Gonçalves, C. T., Gonçalves, M. J., & Campante, M. I. (2023). *Developing integrated performance dashboards visualizations using Power BI as a platform*. Information, 14(11), 614. https://doi.org/10.3390/info14110614

Grabinger, L., Hauser, F., & Mottok, J. (2023). *On the perception of graph layouts*. Journal of Software Evolution and Process, 36(5). https://doi.org/10.1002/smr.2599

Kawai, C., Zhang, Y., Lukács, G., Chu, W., Zheng, C., Gao, C., Gozli, D., Wang, Y., & Ansorge, U. (2022). The good, the bad, and the red: implicit color-valence associations across cultures. *Psychological Research*, *87*(3), 704–724. https://doi.org/10.1007/s00426-022-01697-5

Lien, Y. E. (2009). Relational database design. In *Elsevier eBooks*. https://doi.org/10.1016/b978-0-12-374730-3.x0001-0

Metre, V., Mathur, A., Prakash, R., Dahake, Bhapkar, Y., Ghadge, J., Jain, P., & Gore, S. (2023). *An Introduction to Power BI for Data Analysis*. International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING, 142–147. https://ijisae.org/index.php/IJISAE/article/view/3401/1988

Nieto, G. M. F., Kitto, K., Shum, S. B., & Martínez-Maldonado, R. (2022). *Beyond the Learning Analytics Dashboard: Alternative Ways to Communicate Student Data Insights Combining Visualization, Narrative and Storytelling*. LAK22. https://doi.org/10.1145/3506860.3506895

Oladipupo, M. A., Obuzor, P. C., Bamgbade, B. J., Adeniyi, A. E., Olagunju, K. M., & Ajagbe, S. A. (2023). *An Automated Python Script for Data Cleaning and Labeling using Machine Learning Technique*. Informatica, 47(6). https://doi.org/10.31449/inf.v47i6.4474

Schröer, C., Kruse, F., & Gómez, J. M. (2021). *A Systematic Literature Review on Applying CRISP-DM Process model*. Procedia Computer Science, 181, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Spotorno, S., & Tatler, B. W. (2024). What's left of the leftward bias in scene viewing? Lateral asymmetries in information processing during early search guidance. *Cognition*, *254*, 106009. https://doi.org/10.1016/j.cognition.2024.106009

Stehle, S., & Kitchin, R. (2020). *Real-time and archival data visualization techniques in city dashboards. International Journal of Geographical Information Science*, 34(2), 344–366. https://doi.org/10.1080/13658816.2019.1594823

Van Geert, E., & Wagemans, J. (2023). *Prägnanz in visual perception*. Psychonomic Bulletin & Review. https://doi.org/10.3758/s13423-023-02344-9

Yigitbasioglu, O. M., & Velcu, O. (2012). *A review of dashboards in performance management: Implications for design and research*. International Journal of Accounting Information Systems, 13(1), 41-59. doi: 10.1016/j.accinf.2011.08.002.

Zhang, Y., Reynolds, M. A., Lugmayr, A., Damjanov, K., & Hassan, G. M. (2022). *A visual data Storytelling framework. Informatics*, 9(4), 73. https://doi.org/10.3390/informatics9040073

Zhou, S., Zhang, R., Chen, D., & Zhu, X. (2021). *A novel framework for bringing smart big data to proactive decision making in healthcare*. Health Informatics Journal, 27(2), 146045822110246. https://doi.org/10.1177/14604582211024698

APPENDIX

Appendix A: Interview Setup

Introduction:

This interview is a conversation about the role that data could play within Eurofins FSS, With the main focus laying on consultancy. Within this interview I will ask some questions about both the role in the company you have, together with questions that are about the most important aspects of the company. Over time I will ask more about what you think about using data from the company to help the company move forward. The interview will either be conducted in English or Dutch depending on the wants of the interviewee.

First, I would like you to read and sign the following consent form. That describes what will be done with the data that will be collected via this interview. **Gives consent form**

Now that that is out of the way I will start with the interview.
I will record the interview as mentioned in the consent form to track and save the interview performed.

At any moment you can opt out of the interview, without any repercussions.

Now, let us begin:

First a small introduction, I am Jesse, and I will be asking your opinion about matters regarding the process of Consultancy as of right now, as well as asking about possible changes within the use of data to facilitate Eurofins FSS's consultancy. Do you understand?

Then let me start with the first question:

1. What is your function within Eurofins FSS?
2. What is your relationship with the process of consultancy of Eurofins FSS? (Do you do personal consultancy, or are you further away from the process)
3. Can you describe how the initial consultancy process goes in a broad way?
4. During consultancy, a testing plan has to be made, do you know how this is done?
5. Do you agree with the way a testing plan is made? Is it right/wrong? Do you think it needs improvement?
6. What type of improvements could the consultancy process regarding making a testing plan have?

Now I will tell you a bit about what I do within Eurofins at this moment. I work on creating a system to visualize historic data, to give a better insight into the total testing that Eurofins does, as well as helping with creating a system that can help decide what tests need to be held within the consultation that is held. The idea is that this system makes the reasoning behind what tests are done with clients more data-driven, to help consultants with their consultancy. I will now ask some questions about the subject at hand.

7. Do you think a data-driven system that gives data insights into certain sectors and clients can help Eurofins FSS' consultancy?
8. What sort of data do you want to see in an assisting tool for consultancy at Eurofins FSS.
9. How would you want to use a data-driven system in a consultancy situation, if you want to use them?

10. What features would you like to see in a data-driven analysis system for consulting for Eurofins FSS?

That is the end of the interview. If you have any questions or have anything to add feel free to do so, else I will stop the recording and the interview officially.

.

Appendix B: Consent Form Interview

Consent Form for an interview surrounding the topic of data-driven analysis tools and the information that should be available within such systems at Eurofins FSS

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated 19-05-2024, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	○	○
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	○	○
Use of the information in the study		
I understand that information I provide will be used for designing and creating a data-driven analytical system that will be presented to Eurofins Food Safety Solutions, to increase the use of data within the consultancy process of Eurofins Food Safety Solutions.	○	○
I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team.	○	○
Consent to be Audio/video Recorded		
I agree to be audio/video recorded.	○	○
Future use and reuse of the information by others		
I give permission for the anonymized answers in the interview about using data-driven analytics in Eurofins Food Safety Solutions that I provide to be archived in Jesse van Gameren' personal laptop storage so it can be used for future research and learning. The data will only be used for Jesse van Gameren' personal Thesis from 2024 connected to the University of Twente.	○	○

Signatures

_____	_____	_____
Name of participant	Signature	Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____	_____	_____
Researcher name	Signature	Date

Study contact details for further information

Jesse Yannik van Gameren, j.y.vangameren@student.utwente.nl

Contact Information for Questions about Your Rights as a Research Participant

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee/domain Humanities & Social Sciences of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-hss@utwente.nl

Appendix C: list of branch organization that can be used for data analysis

Relevante spelers in het domein levensmiddelen van plantaardige oorsprong

- CBL: Centraal Bureau voor Levensmiddelenhandel
https://www.cbl.nl/onderwerpen/voedselveiligheid/ (geen onderzoeksdata)
- FNLI: Federatie Nederlandse Levensmiddelenindustrie waaronder vele grote en kleine brancheorganisaties ressorteren
https://fnli.nl/voeding-gezondheid/voedselveiligheid (geen onderzoeksdata)
- Culinaria NL: Nederlandse Vereniging van Fabrikanten van Culinaire Producten (geen website)
- FWS: Nederlandse Vereniging Frisdranken, Water, Sappen
- KNVKT: Koninklijke Nederlandse Vereniging voor de Koffie en Thee
- NEA: Vereniging van de Geur- en Smaakstoffenfabrikanten
- NEBAFA: Vereniging van Nederlandse Fabrikanten van Bakkerijgrondstoffen
- Nederlandse Brouwers
- Nederlandse Vereniging van Sausfabrikanten
- NVB: Nederlandse Vereniging voor de Bakkerij
- VAVI: Vereniging voor de Aardappelverwerkende Industrie
- VBZ: Vereniging voor de Bakkerij en Zoetwarenindustrie
- VIGEF: Vereniging van de Nederlandse Groenten- en Fruit verwerkende Industrie
- Groenten en fruithuis
- NVS: Nederlands Vereniging Specerijenhandel
- NZV: Nederlandse Vereniging voor de Handel in Gedroogde Zuidvruchten, Specerijen en Aanverwante Artikelen
- AKSV: Algemene Kokswaren en Snackproducenten Vereniging
- MVO: Ketenorganisatie voor oliën en vetten
- NBOV: Nederlandse Brood- En Banketbakkers Ondernemers Vereniging
- NBC: Stichting Nederlands Bakkerij Centrum
- CTGB: College voor de Toelating Gewasbeschermingsmiddelen en Biociden

Relevante spelers in het domein levensmiddelen van dierlijke oorsprong, vlees.

- Federatie Nederlandse levensmiddelenindustrie waaronder vele grote en kleine Brancheorganisaties ressorteren (FNLI)
- Centraal Bureau voor levensmiddelenhandel (CBL)
- Centrale Organisatie voor de Vleessector (COV)
- Vereniging van Slachterijen en Vleesverwerkende bedrijven (VSV)
- Koninklijke Nederlandse Slagers (KNS)
- Vereniging voor de Nederlandse Vleeswarenindustrie (VNV)
- Vereniging van de Nederlandse Pluimvee verwerkende Industrie (Nepluvi)
- Nederlandse Bond van Poeliers en Wildhandelaren (NBPW)
- AKSV: Algemene Kokswaren en Snackproducenten Vereniging
- MVO: Ketenorganisatie voor oliën en vetten

Relevante spelers in het domein samengestelde levensmiddelen

- CBL: Centraal Bureau voor Levensmiddelhandel
- Federatie Nederlandse Levensmiddelenindustrie (FNLI) waaronder vele grote en kleine Brancheorganisaties ressorteren
- Culinaria NL: Nederlandse Vereniging van Fabrikanten van Culinaire Producten
- FWS: Nederlandse Vereniging Frisdranken, Water, Sappen
- NEA: Vereniging van de Geuren
- Smaakstoffenfabrikanten
- NEBAFA: Vereniging van Nederlandse Fabrikanten van Bakkerijgrondstoffen
- Nederlandse Brouwers
- Nederlandse Vereniging van Sausfabrikanten
- NVB: Nederlandse Vereniging voor de Bakkerij
- NZO: Nederlandse Zuivelorganisatie
- VAVI: Vereniging voor de Aardappelverwerkende Industrie

- VBZ: Vereniging voor de Bakkerij en Zoetwarenindustrie
- VIGEF: Vereniging van de Nederlandse Groenten en Fruit verwerkende Industrie
- Visfederatie: Federatie van organisaties op het gebied van visverwerking en visgroothandel
- Groenten en fruithuis
- NVS: Nederlands Vereniging Specerijenhandel
- NZV: Nederlandse Vereniging voor de Handel in Gedroogde Zuidvruchten, Specerijen en Aanverwante Artikelen
- AKSV: Algemene Kokswaren en Snackproducenten Vereniging
- MVO: Ketenorganisatie voor oliën en vetten
- NBOV: Nederlandse Brood- En Banketbakkers Ondernemers Vereniging
- NBC: Stichting Nederlands Bakkerij Centrum
- VBZ: Vereniging voor de Bakkerij- en Zoetwarenindustrie
- Spirits Nl

Relevante spelers in de domeinen voedingssupplementen en kruidenpreparaten en voeding voor specifieke groepen

- VNFKD: Vereniging van Nederlandse fabrikanten van kinder- en dieetvoedingsmiddelen
- CBL: Centraal Bureau Levensmiddelen
- FNLI: Federatie Nederlandse Levensmiddelen Industrie
- NPN: de brancheverenigingvoor producenten, grondstofleveranciers, groothandelaren, importeurs en distributeurs van voedingssupplementen, zoals vitamine- en kruidenpreparaten
- Neprofarm: de brancheorganisatie van fabrikanten en importeurs van zelfzorgproducten

Appendix D: Power BI Semantic Model

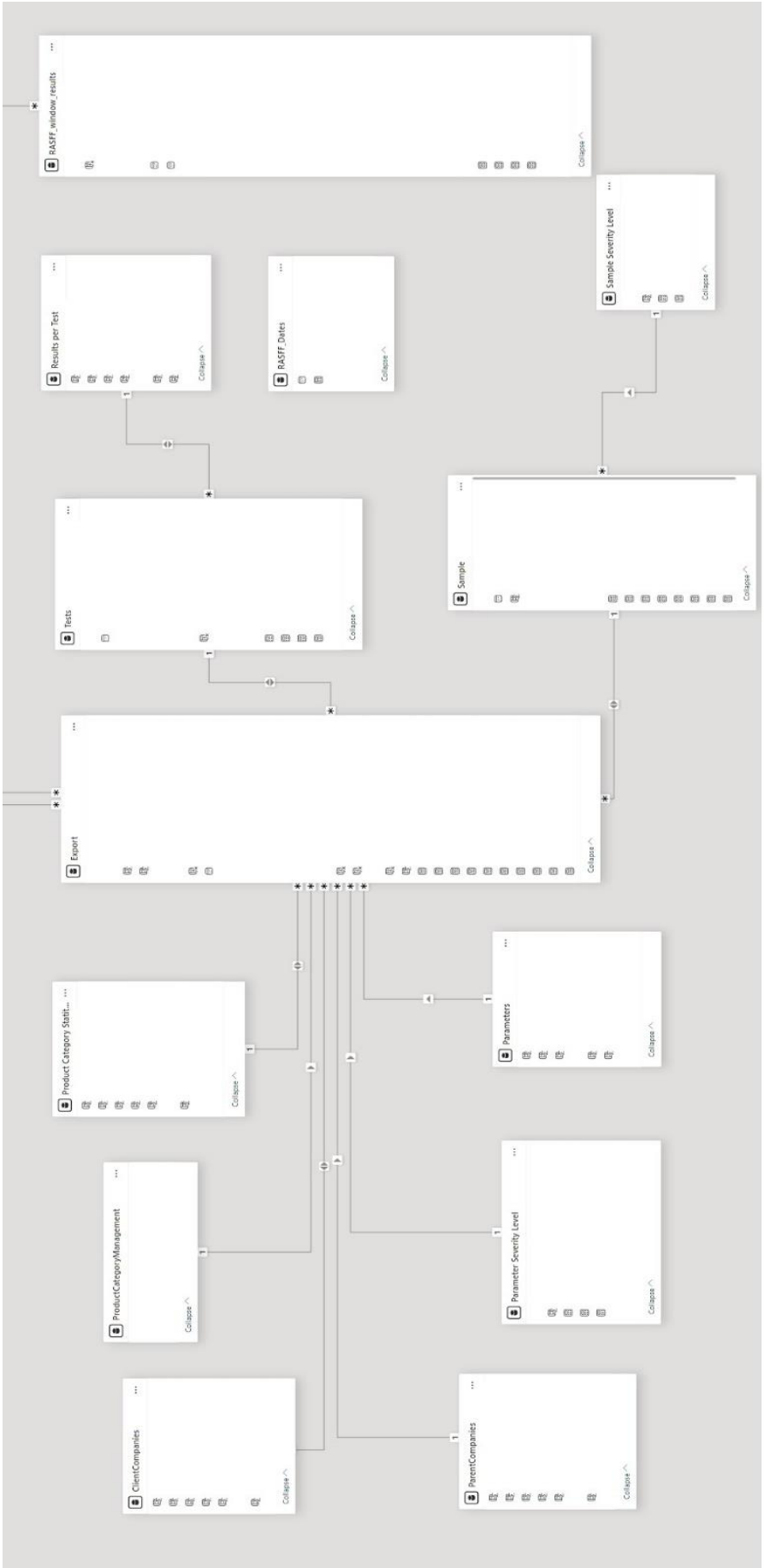


Figure D.1: Power BI semantic model that manages the back end of the dashboard system.

All names of variables are deleted due to the names being classified under company policy.

Appendix E: User Guide

See next page.

User Guide

Eurofins Data Analysis Tool

Updated Last:
05/07/2024 by Jesse van Gameren

Introduction

This document is a user guide to aid users while using the Eurofins Data Analysis Tool created in summer 2024 to help with getting insight into the analytical data that is present within the database of Eurofins. This Guide will show all functions of the analysis tool, together with telling how the present values, graphs and tables got calculated.

Application

The Eurofins Data Analysis Tool is a Business Intelligence tool created using software from Power BI to get insight into the food sector data that Eurofins has. This application was created as an internship project and picked up by the Eurofins BI team. The application uses automated calculations and filtering options to create an overview of important points that can be found within data to help with making statements about the state of food testing on a real time level.

Table of Contents

1. General Parts of the Analysis tool

1.1 Power BI Interaction Functionalities

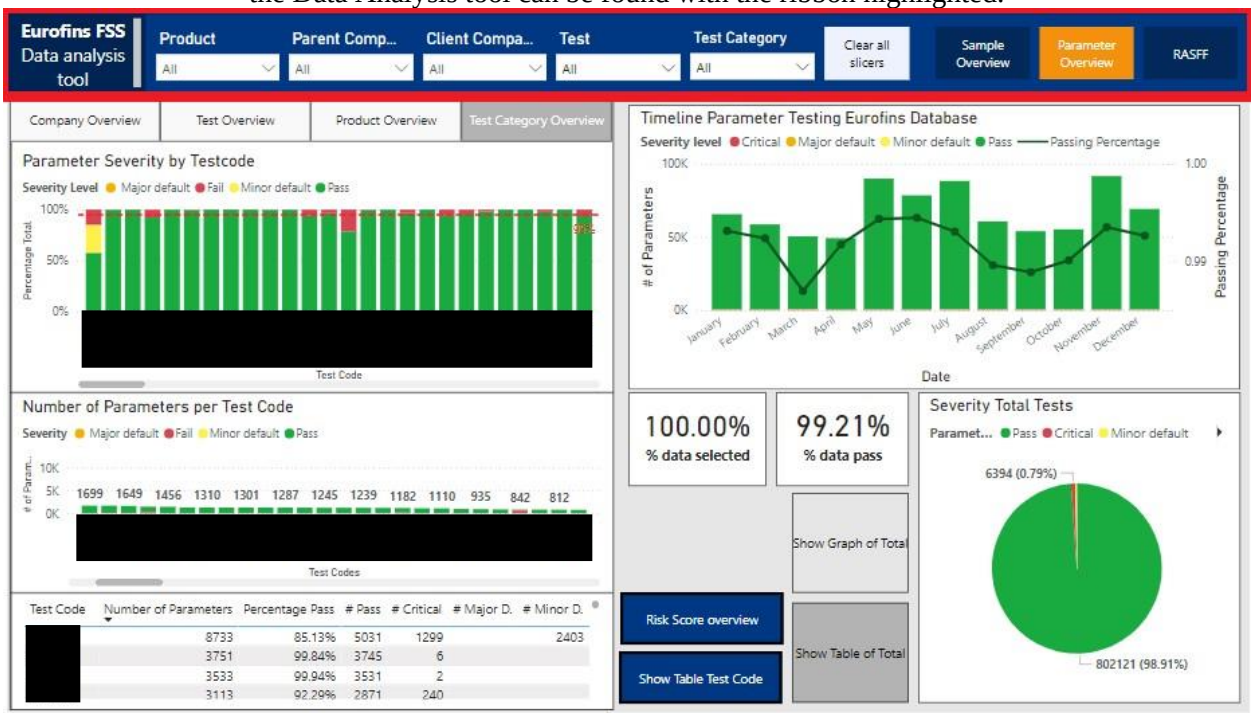
Power Bi has a lot of potential and has a lot of options to investigate data thoroughly with a lot of different functions. Some of these functions are very interesting, and others are not as useful for the application at hand. In This paragraph a few functions will be explained to look at certain graphs that have these options.

In the top right of a graph or table all options with the visual are found, these are important to be used properly when used, therefore This short list of explanations will be given about the options in power Bi that are useful. See the numbering below.

Function 1: Drill up is the function that makes the visual go up 1 level of depth. (Example: a visual looks at Client companies, clicking on the arrow will show the total Parent companies instead (a level of depth higher))
Function 2: Drill Down is the arrow down that if activated, if you click on a value within the graph (for example a test group), it will instead of highlighting the value drill down 1 level for that value. (Test codes from that test group)
Function 3: Go to next Hierarchy goes to down a level of hierarchy just like drill down, yet it takes all values together, not specific to 1 group value.
Function 4: Expand all down one level shows the next layer of hierarchy divided per value of the hierarchy above (So it will show all test codes, but ordered under the test groups they belong under) Overall this is a better version of function 3.
Function 5: Copy This copies the visual or table.
Function 6: Filter This shows al filters active on the visual.
Function 7: Expand This expands the visual to full screen.
Function 8: More under more a few self-explainable helpful functions lie, such as export to export the data, and show as table to transform a graph of data into a table. Other functions are overall not useful.

1.2 Top Ribbon Overview

The Eurofins Data Analysis Tool works around the use of filters to use big batches of data to come to a specific conclusion within the data. To aid with different views and keeping filter functions logical and user friendly, every page of the analysis tool has the same Ribbon on the top to create a familiar way of working in the system. Underneath the functions of the Ribbon will be explained. Underneath a picture of the Data Analysis tool can be found with the ribbon highlighted.

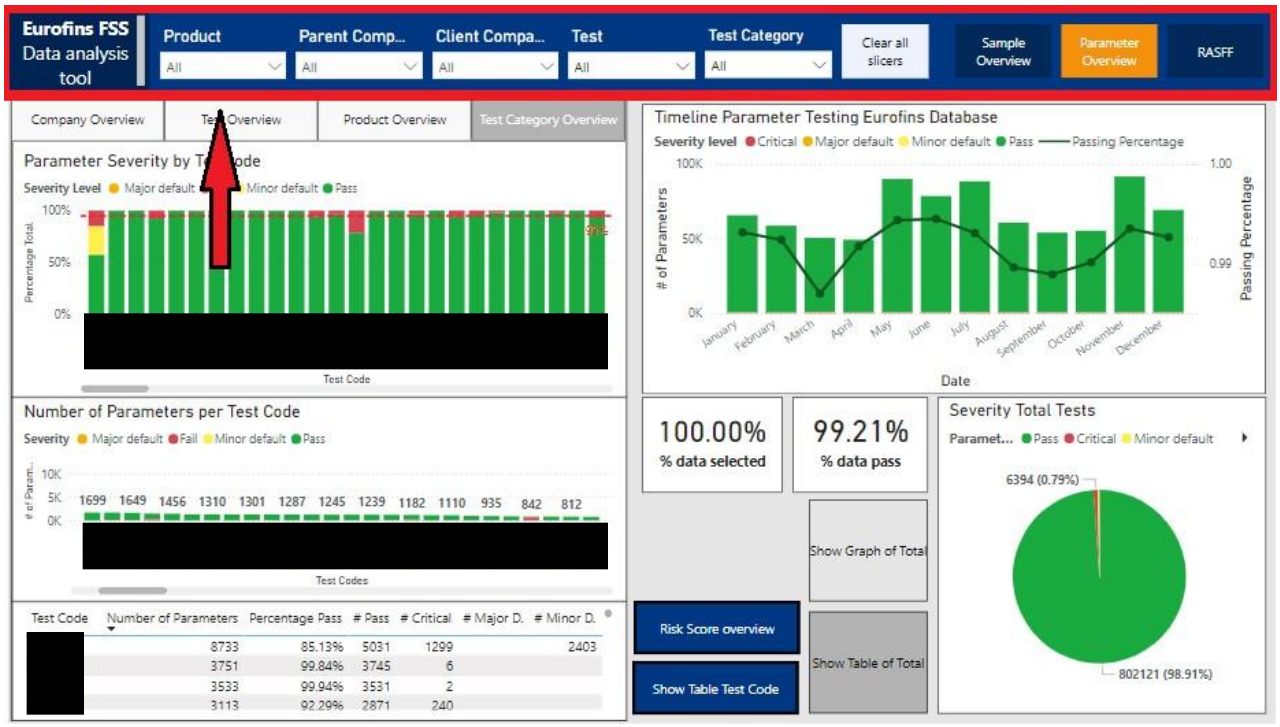


1.2.1 Filters

In the top ribbon, the first notable thing found is the row of 5 filters that can be found at the top of the page. (picture below with arrow pointing at the filters). The filters function at the top of the page is the main way of selecting what type of data the user wants to see. These filters filter all the data used to the sections selected within the filters, resulting in updated graphs based around the selected data.

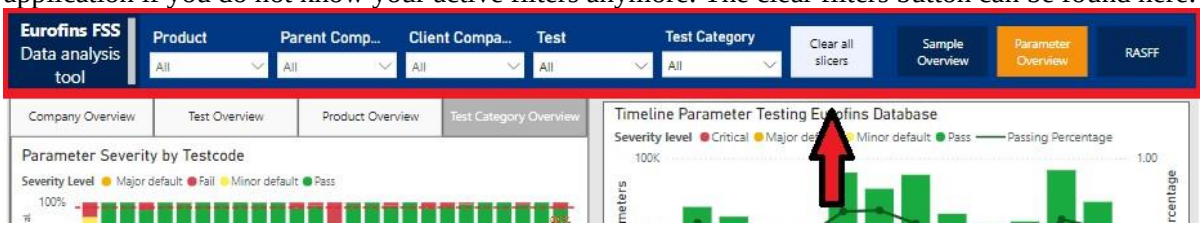
There are 5 filters present:
- **Product**: Here the product group or loose products can be selected to be analyzed.
- **Parent Company:** Here single or multiple Parent Companies can be selected.
- **Client Company:** Here single or multiple Client Companies can be selected.
- **Test:** here loose test codes or names can be selected for analysis.
- **Test Category:** here a group of tests within a single domain can be taken (contaminants or pesticides for example)

All filters have a search bar to find certain filters directly, together with having a select all button.
***Note Multi-Selection:** If you want to select multiple values in a singular filter, you have to select the values while holding **Left Ctrl**. (So **Left Ctrl + Left mouse button**)



1.2.2 Clear all Filters

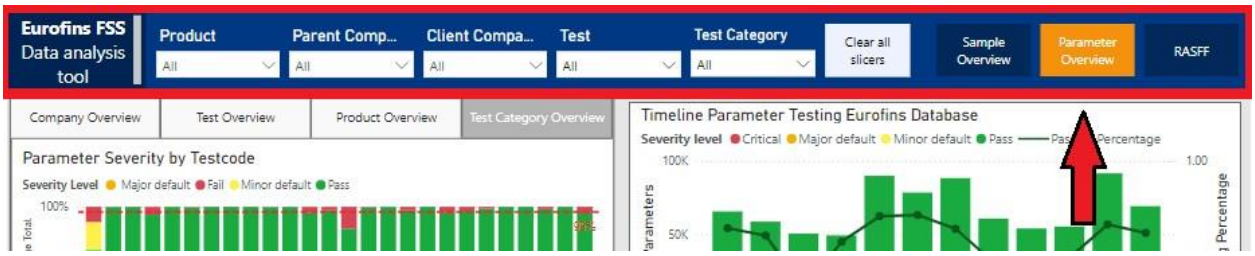
The clear all filters button does what the name suggests and clears all active filters on the top. This will reset the application to the beginning, which is useful to either start over with searching, or to reset the application if you do not know your active filters anymore. The clear filters button can be found here:



1.2.3 Navigation System

The Analysis Tool has 3 different pages which get used to show data. To switch between pages a tab system can be found on the top right of the screen (as shown below). Clicking on the tabs switches the page visible.

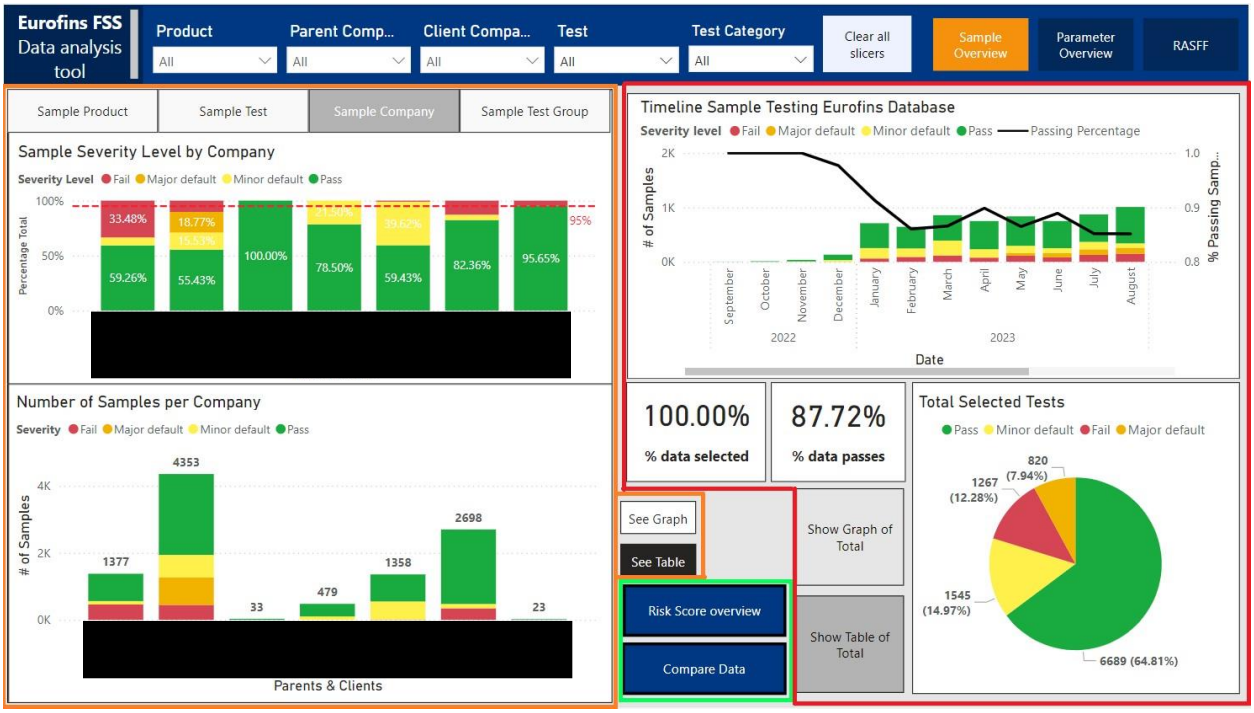
Keep in Mind! If you use filters on page A and do not clear them before switching to page B, they will stay active even on the other page.



2. Sample and Parameter Overview

The main body of the Analysis Tool is the Sample and Parameter Overview.
The most useful part of the Data analysis tool is the Sample Overview. This should be used 90% of the time while using the analysis tool created, with the parameter overviews as an extra layer of depth to look into specific parameters if extra conclusions are needed. Both the Parameter Overview and the Sample overview are quite similar, with only a few differences per page, therefore this part of the guide combines both views together.

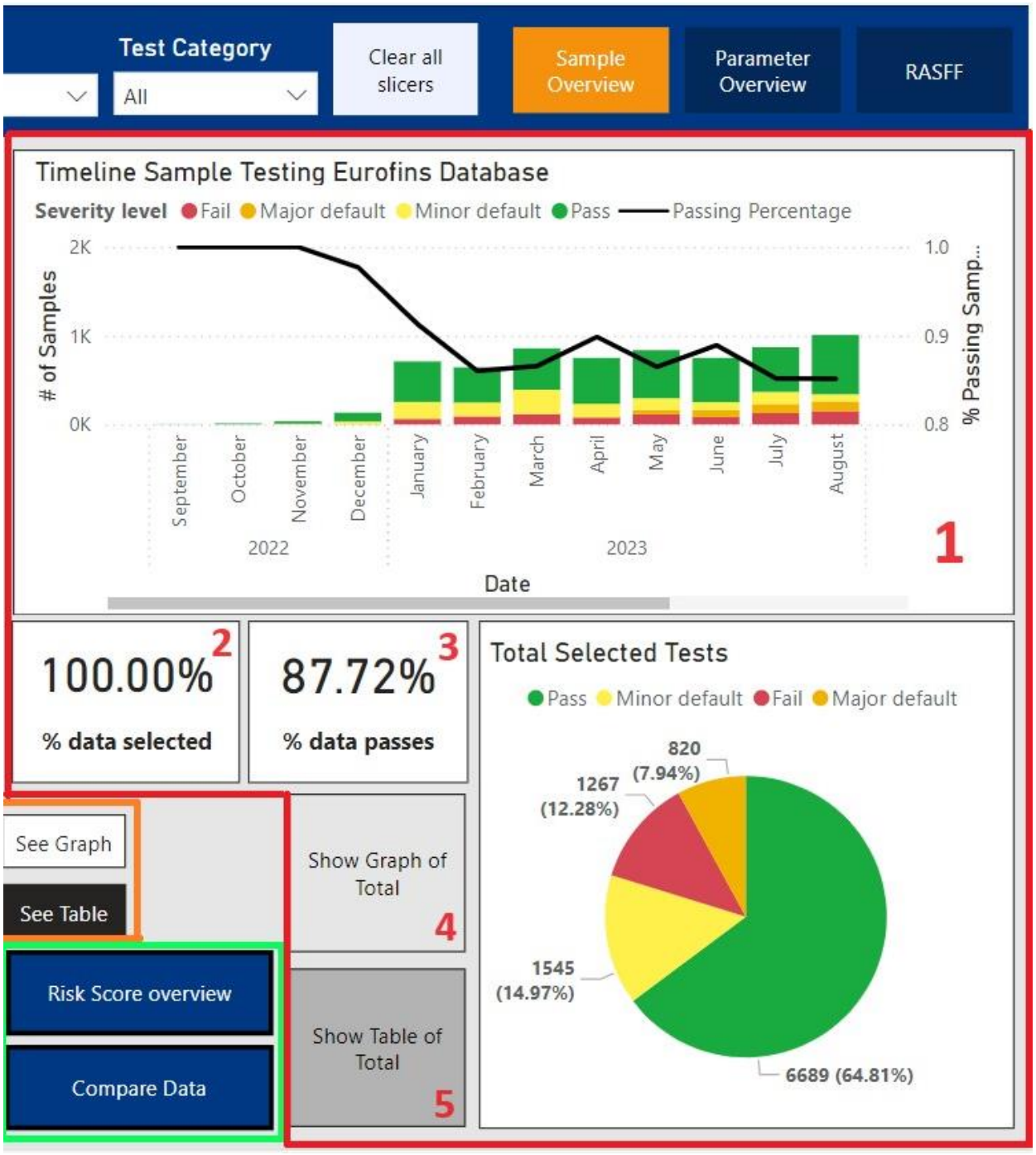
The Sample and Parameter overviews are both based around the same principle:
- The Right side is for standard statistics
- The left side is for looking through data
- The buttons at the bottom are extra functionalities
This will also be the division between paragraphs within the guide to keep clarity. The screen division can be seen below:



2.1 Right side Overviews: Standard visuals

The right side of the screen is meant to show visuals that change together with filters, showing basic statistics which are handy and easy to use to overview the situation or to export to show to clients. Statistics such as pass rate, severity division and more. **For both the Sample overview and parameter overview, the given information is completely identical.**

Underneath a screenshot with numbering can be found with all de visuals numbered. Per number the visuals will be explained.



2.1.1 Timeline

In the top right visual a timeline can be found that shows the selected group of results severity score in testing based on time. This visual can look at total years (Drill up) or at monthly values (drill down).

The columns are based on the left Y axis, showing the number of samples per unit of time. The line is the average passing rate which needs to be reviewed using the right Y axis.

**Note that if Drill up/down are not clear, see paragraph 1.1 of the User Guide*

2.1.2 Percentage selected.

The Percentage selected graph shows the percentage of data that all visuals still use after using filters. With this you can for example see what percentage of tests is done by a certain company or within a certain test group. This visual helps with keeping track with the relevancy of the results found in the visual, as a conclusion found using 0.05% of all data is less trustworthy than one based on 50.0% of the data.

2.1.3 Data Passed

This Visuals shows the passing rate the selected data. So, if the data is filtered on a company, it is the total passing rate of the company.

2.1.4 Total Selected Tests (Graph)

This shows the division of test conclusions of the selected data in a pie chart. This graph is easy to interpret, together with giving a clear overview of the percentages of samples with different levels of severity.

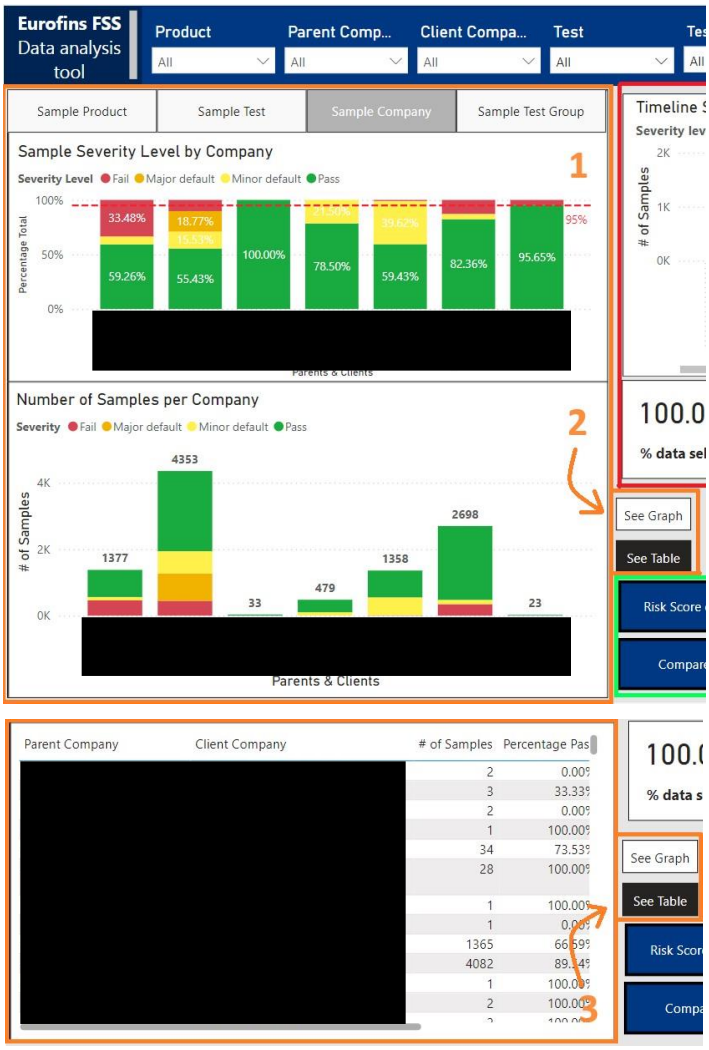
2.1.5 Total Selected Tests (Table)

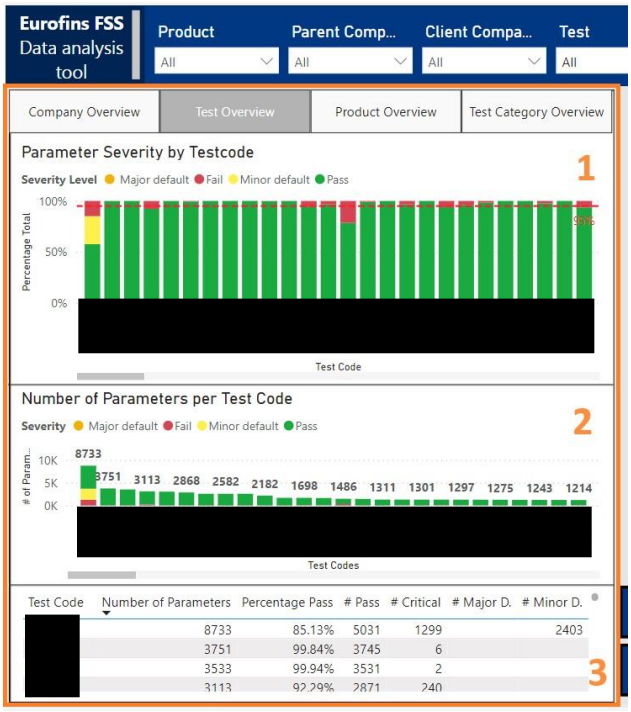
This Table shows the same data as Total Selected Tests (Graph) but than in a table. It is an easy overview of the severity of tests scores, making it easy to see how a certain group of products or companies is performing on average.

2.2 left side Overviews: data views

Looking at the left side view of the analysis tool, both pages have the same graphs and visuals, but other Layouts. Therefore, first both layouts will be laid next to each other to show the differences. The focus will be the set on the Sample overview as this is the overview that yields the most results when it comes to the analysis.

Underneath an overview of the visuals/buttons is given per number to show the similarities between pages. After the visuals will be discussed per number.

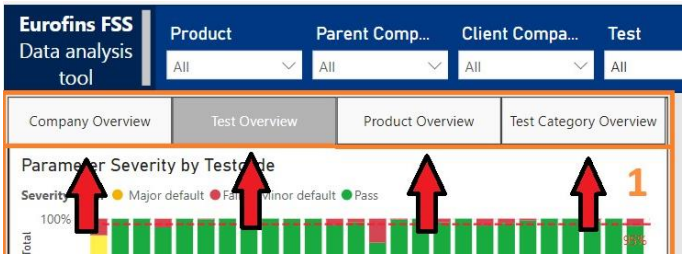




The difference present is that on the Sample overview, The table and Lower Graph can be switched in view via a button, whereas this is not the case for the Parameter Overview, where both visuals are shown stacked on top of each other.

When looking at the left side view, it is important to know that this view is based on a tab system. The left side is divided into 3 visuals, with multiple different tabs that have different information that is used as the

main value. This main value will be called **value X** in the explanation about the visuals. So, every tab has the same graphs, based around different values. For example, the percentages of severity scores per Value X (company, testcode, etc.). To the right the tabs can be seen.



The values that can be used to analyze the data by using the tab system are:

- **Company overview:** here both parent companies and client companies can be analyzed in different ways.
- **Test overview:** here test codes can be analyzed in different ways.
- **Product overview:** here product groups and loose products can be analyzed in different ways.
- **Test category overview:** here test categories/groups can be analyzed in different ways.

2.2.1 Severity Level by Value X

This visual is a percentage of total visual, showing the selected value X in columns of 100% per group. This visual shows the percentage division based on the selected tab, showing the actual rate of passing, without blinding the viewer with statistics that are skewed due to the number of tests differing. Given that the test is based on percentage, this visual accurately shows difference in passing, and can be used well to show how certain groups of values (companies or products for example) are doing compared to the rest of the values.

2.2.2 Number of Samples per Value X

This Graph shows the amount and division of the selected values groups in severity levels. This graph is used to back up the previous graph (2.2.1) by showing how many tests have been done in a certain sector. Given more data values increases reliability, this graph shows well if the numbers are representative, as well with showing if the spread in tests over a sector is balanced (not letting one company do all tests in the meat sector for example).

2.2.3 Table Overview Value X

This Visual is a table showing data divided based on Value X. The value it shows are the passing %, The number of Passes, major defaults, minor defaults, fails and critical values, as well as the number of results per grouping of the selected tab. This table is an easy to read and copy overview of a group of tests, company, or sector, which is good for making overviews on paper about a certain sector.

2.2 Button Overview: Extra Functionalities

Both the Sample Overview and the Parameter Overview have 2 Blue buttons each with useful extra functionalities. They are located at the bottom of the screen (in the green square in the introduction of Chapter 2., page 7). In this paragraph These Buttons will be discussed, first explaining the Sample Page Buttons uses and calculations, then the Parameter Page Buttons.

2.3.1 Sample Page: Risk Score Overview

This button shows a matrix per category of the risk matrix value of that certain group. This value is calculated using the data of the last 365 days, taking the frequency of non-passing values as the measure to calculate the risk matrix value. This number is an estimation of the chance of a failing test being present in a certain group. This matrix can be used to argue about the level of risk behind a certain test or product group. The data used is only the last 365 days to keep relevancy of the data accurate. The higher the number the higher the chance of a violation. If a square does not have a value, it has never occurred that such a mix of variables was tested. If the value is 0, this test mix has not been violated last year at all.

2.3.2 Sample Page: Compare Data View

Compare Data opens a window with a similar visual structure to the left side of the analysis tool. This window is a fully independent working window which view can be altered by using its own filters. This window gives the possibility for the user to make 2 different filter selections and compare the data between the left and right side of the screen. This function can help with seeing the differences between a certain company or the general dataset, or the difference between certain test groups to find conclusions within the data.

2.3.3 Parameter Page: Risk Score Overview

This button opens a page which shows how parameters are performing in their test group, as well as what test group the parameters are in. This is a function useful if you want to dive deeper into the data after sample data is not being precise enough. Given that a group of tests does bad, you can see what parameter the problem of the test group might be using this view. Like the Risk Score Overview of the Sample page, this risk score is calculated looking at the frequency of not passing test over the last 365 days.

2.3.4 Parameter Page: Show table Value X

The last button on the Parameter page is a button that opens a window to give a better view of the table located at the bottom left. As the view is not as important to be seen in general the table is small, yet with the big amount of data that is present in the analysis tool, it is wanted to have an option to make the table bigger and get a clear overview of the table when searching for specifics. Therefore, this button is in place. The button changes with the changes in the tab system on the left side to always display the proper data.

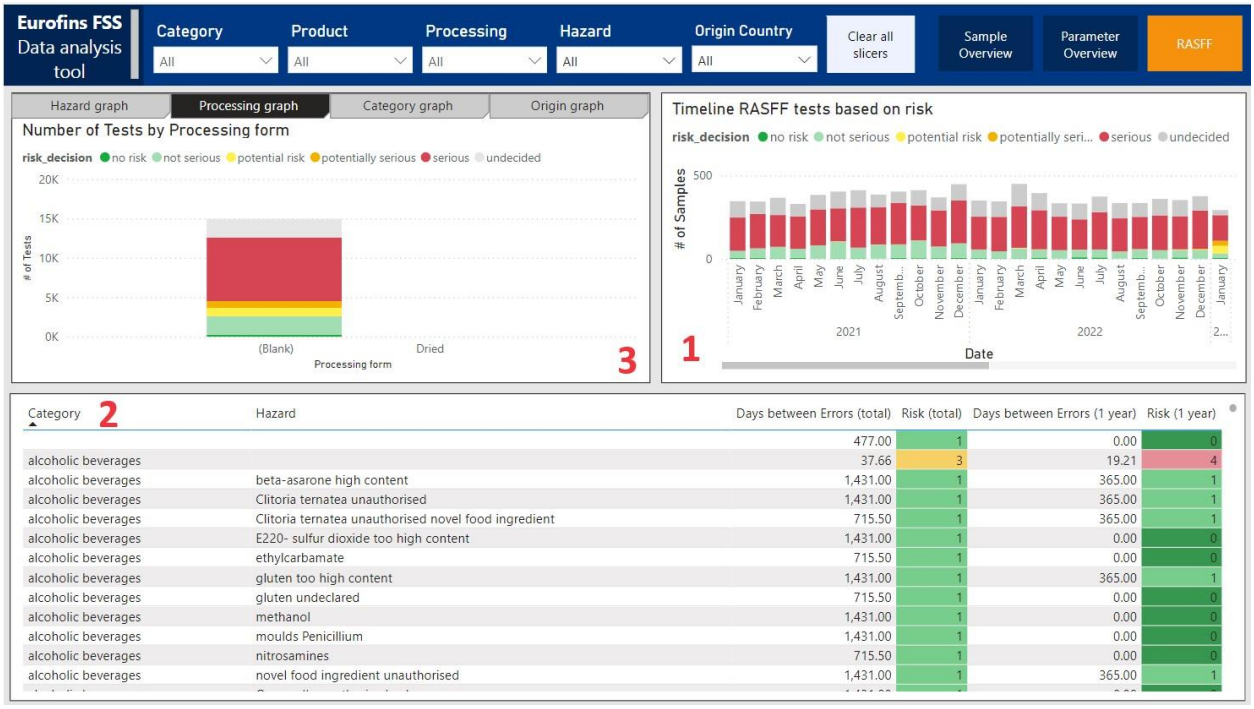
3. RASFF Overview

Lastly RASFF is a European institution that deals with food testing errors. Within RASFF European regulation violations get written down, which is open data to be used by everyone. Given that the data from RASFF is trustworthy and adds a new layer of vision to Eurofins perspective of food testing, this data is also partially implemented in the Analysis tool. The RASFF Overview shows the interesting parts of the data that RASFF collects.

An Important thing to note is that RASFF uses different names and notations for the violations they note. This results in slightly differently structured data, and the use of different Filters to filter the data consequently. The filter function works the same as the Sample and Parameter Overview yet is stated slightly different.

Another thing to keep in mind with RASFF data is that all RASFF data is taken from a violation database, resulting in almost all data points being violations (not passing instances). This means that percentage-based tests and visuals showing percentages do not have any merit and are for that reason not included. Frequency compared to other moments in time however are a good way to see changes over time, as the number of tests does not change drastically, yet only the violations get documented by RASFF. RASFF data can be used to confirm possible conclusions made on other data, plus it can give insights on what sort of data to look out for in the Netherlands.

Underneath the layout of the RASFF Overview can be found:



The Tables and Graphs will be discussed one by one following the number order shown above.

3.1 Timeline RASFF data

This visual shows the timeline of data from RASFF, with filters this timeline can be used to collect important data bout if there are changes per season in certain groups of products, if there are changes over time in general in certain aspects, and to see if the volume of tests has changed of RASFF.

3.2 Table RASFF Data

This table investigates the mix of products and hazards that is mentioned within the RASFF database. This Table takes the product group and lists the hazards it can have behind it. After that it lists the frequency of violations per hazard per product group, as well as the risk factor that is connected to this value. The graph shows both the risk value and frequency over the total database import (3 years) and the last year (365 days) to see if in short term changes are visible per product group or hazard. This can show important trends within the data based on time-related events.

3.2 Number of Tests by TAB

This visual is a tabbed visual looking into ordering the total data within groups to analyze a layer deeper. This visual has 4 tabs that can be used to show the data, these options area:

Hazard
Processing
Product category
Origin

Using filters and the mentioned tabbed column chart that shows the different groups of data, different combinations can be made to gain proper insight into relations such as the sorts of products together with origin country, or the effect of way of processing on hazards.

Appendix F:
BPMN 2.0 model of the Eurofins FSS consultancy process