
# Optimising the Safety of Crossing Railroads to Aim for a Safer Rail Net

*Bachelor Thesis*


| | |
|---|---|
| Author: | H.S.J. (Hugo) Vaatstra |
| Student number: | s2434806 |
| Educational Institution: | University of Twente |
| Faculty: | Behavioural and Management Sciences |
| Programme: | Industrial Engineering and Management |
| Date: | 23-09-2022 |
| Internal supervisor: | J. (Jeroen) van Doorn |
| First External supervisor (UT): | Dr. rer. nat. D. (Daniel) Braun |
| Second External supervisor (UT): | Dr. G. (Gayane) Sedrakyan |


ProRail

UNIVERSITY OF TWENTE.

# Preface

Dear Reader,

Before you lay the thesis, 'Optimising the safety of crossing railroads to aim for a safer rail net', that is written to conclude the bachelor's degree in Industrial Engineering and Management (IEM) at the University of Twente. The research was performed at ProRail, which is located in Utrecht.

At ProRail, I have worked with many different people from several departments. During brainstorming sessions and meetings, they provided feedback or helped me by sharing their expertise. I want to thank Jeroen van Doorn, who has given me the opportunity to make this bachelor thesis. Besides, I would like to thank Eduard de Vries for providing feedback on the dashboard.

I would also like to thank my supervisor D. Braun for his excellent guidance and patience. Especially at the beginning, he helped me overcome problems and gave me helpful advice. Also, I would like to thank my second supervisor G. Sedrakyan for their feedback. My supervisors helped improve the thesis, for which I am grateful.

I would also like to thank my supportive friends at the university and my friends at home. Last but not least, I would like to thank my parents and especially my sister for their support during the process.

Enjoy reading the thesis!

**Hugo Vaatstra**
Apeldoorn, September 2022

# Management summary

This research has been conducted on behalf of ProRail, the only railway company in the Netherlands. It is responsible for maintaining, renewing, expanding, and securing the railway. The number one priority of ProRail is safety; therefore, they are always trying to optimise this.

At this moment, ProRail is not satisfied with the number of incidents and wants to decrease this number. According to ProRail, especially on and around railroad crossings, many incidents happen. Currently, around a quarter of the primary-, secondary- and almost incidents are railroad crossing-related incidents. These incidents are partly caused by the non-existence of a report/tool showing factors influencing safety on and around railroad crossings. Since this is not available, creating a report on these possible factors with an additional visualisation tool will be focused on within this thesis. Research questions have been made to solve this problem.

The Design Science Research Methodology (DSRM) has been used to create these research questions. Therefore, first, the problem of this research has been identified. After this, the objective for a solution was given: a dashboard and a report on railroad crossing incidents. Then the actual design and development of this dashboard and report are started. Several phases have been followed to create this dashboard and write the report. For every phase, an extensive literature study has been conducted that provides a theoretical framework. The first phase was gathering data from several databases and registers. This data has been cleansed using functions from Microsoft Excel and coding from Microsoft VBA. By using this data, and with some additional conducted semi-structured interviews with employees from the Competence Centre Business Application (CCBA) and from the Safety department, key metrics and metrics were determined. Within Power BI, these measures are visualised in an interactive dashboard. Besides that, additional statistical analyses are conducted to find a correlation between key metrics.

The dashboard has shown that when it is dark, percentage-wise, more incidents happen than when it is light. Besides that, it has been shown that there happen a lot of incidents at railroad crossings that have Andreas crosses when comparing it to the number of railroad crossings that have such a security type. Further, it has been shown that when the line speed of railroad crossings is higher, percentage-wise, more primary incidents happen than secondary- or almost-incidents. Lastly, the statistical analysis has shown a moderate correlation between the type of security and the category of incidents.

It is recommended to ProRail keep an eye on railroad crossings that have these characteristics. Besides that, interesting possible further research has been elaborated.

# Table of Contents

# List of Figures

# List of Tables



# List of Acronyms

| | |
|---|---|
| DSRM | Design Science Research Methodology |
| KPI | Key performance indicator |
| POC | Proof of Concept |
| ProVat | ProRail, Safety Information, Current, and Transparent |
| SAP | Systems, Applications and Products |
| SPSS | Statistical Package for Social Science |

# 1. Introduction

This chapter provides background information on ProRail and its role in the Dutch railway system. After that, information on the problem is given and the core problem is identified. The last part of this chapter elaborates on the research methodology used to solve the core problem.

## 1.1. ProRail

On the 20th of September in 1839, the first railway in the Netherlands was opened and used (Booth, 2021). After that, multiple organisations have been responsible for maintaining the railway. Since 2005, after a merging of Railinfrabeheer, Railned, and Railverkeersleiding ProRail, ProRail has been responsible for the maintenance of the railway.

ProRail is an independent Dutch government task organisation that is private; however, it is a not-for-profit organisation (ProRail, 2016). It is in charge of all 7,000 kilometres of railway in the Netherlands. Therefore, they are responsible for maintaining, renewing, expanding, and securing the railway. Besides that, it is responsible for the railway stations and the train traffic (ProRail, n.d. -a). Currently, ProRail is building new stations and renovating existing ones. This way, it hopes to increase the train traffic in the upcoming years to fulfil the increasing demand. Besides that, ProRail wants to become CO2-neutral before 2030 (ProRail, 2016).

While trying to achieve the objectives of increasing train traffic and becoming more sustainable, safety is the biggest priority. As long as ProRail exists, the number one objective of ProRail is optimising its safety on and around the rail net (ProRail, n.d. -b). To obtain the different objectives, ProRail wants to become data-driven. A data-driven organisation is an organisation which makes decisions based on data (Brynjolfsson et al., 2011).

## 1.2. Problem Identification

At this moment, ProRail is not satisfied with the number of incidents and wants to decrease this number. According to ProRail (2021a), many incidents happen on and around crossing railroads. In 2021 alone, there were a total of 1975 incidents, of which 535 were registered incidents on railroad crossings. These numbers show that 27,09% of the registered incidents happen at railroad crossings. The incidents are subcategorised into primary-, secondary-, and almost-incidents, all having different priorities to ProRail. Primary incidents have the biggest priority, followed by secondary- and almost-incidents. Appendix C gives an extensive list of the subcategories of primary-, secondary- and almost incidents.

Since ProRail wants the number of incidents at railroad crossings to decrease, the action problem within this thesis is the high number of incidents on and around crossing railroads. An action problem is a situation that is currently not how it is desired to be (Heerkens & van Winden, 2017).

To decrease the high number of incidents on and around crossing railroads, there needs to be determined what causes this problem. A problem cluster is made to create more insight, investigate the relations between all problems, and find the core problem (Heerkens & van Winden, 2017). Figure 1 shows the problem cluster for this thesis, where the arrows represent the cause-to-effect relationship. This problem cluster's last effect is 'Incidents happen on and around railroad crossings', which matches the identified action problem. Following this action problem, it is noticed that there is no current method that contributes enough to decrease the number of incidents. During a semi-structured interviews with employees from the safety department (in which was asked what current research was done on railroad crossing incidents), it became clear that ProRail has done some research on railroad crossing safety. This research showed that fewer railroad crossings (obviously) lead to fewer incidents, primarily at unsecured railroad crossings. Therefore, they are removing railroad crossings where possible. Internal documents show that the number of railroad crossings decreased by 196 in the past ten years and 1,342 railroad crossings in the past 38 years. Currently, there are 3,108 railroad crossings; however, 1,259 of these railroad crossings are only being used by freight trains. Figure 2 shows that the railroad crossing incidents decreased over ten and over 38 years. Other factors also cause this; however, removing railroad crossings seems to have an effect. However, as discussed before, ProRail stated in 2021 that still too many incidents happen at and around railroad crossings. Therefore, the current methods do not contribute enough to decreasing the incidents.

Eventually, this problem cluster shows that the action problem is caused by a limited insight and exceedance of several indicators set by ProRail. Two potential core problems can be solved to create more knowledge on these topics. These core problems are 'No report that shows what safety measures should be taken at each specific railroad crossing' and 'No tool and/or report that shows which factors influence railroad crossing incidents.'

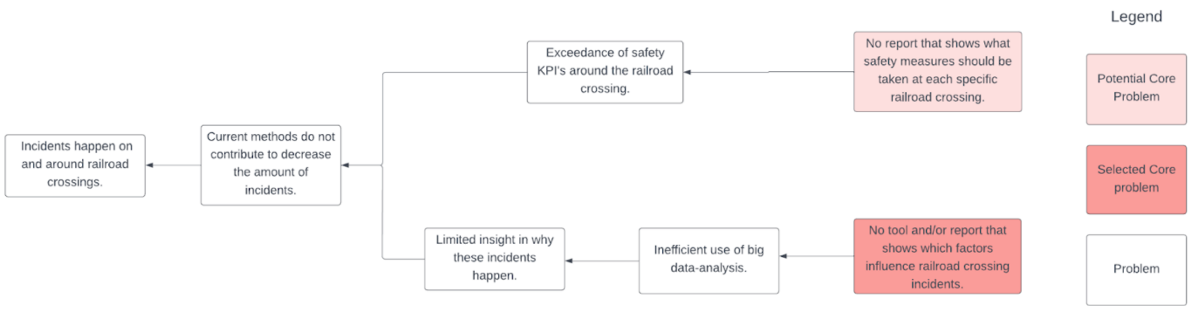


*Figure 1: Problem Cluster*

## 1.3. Core problem

As stated before, the problem cluster shows two potential core problems. According to Heerkens and Van Winden (2017), to select the best core problem, four rules of thumbs should be followed:

1. The core problem should be related to other problems and should be something that does occur in the company.
2. The core problem cannot have any direct cause itself.
3. The core problem should be influenceable.
4. The problem with the most significant impact at the lowest cost should be chosen if there are multiple core problems.

Keeping in mind the four rules of thumb, the following core problem is solved within this thesis: 'No tool and/or report that shows which factors influence railroad crossing incidents.'. This core problem is selected since this problem will most likely have the most significant impact. That is because, with the findings within this thesis, ProRail can take measures at general and crossing railroad-specific levels. Besides, this core problem seems feasible to solve within ten weeks using the theory learned during the bachelor of Industrial Engineering and Management. To solve this core problem, a dashboard will be created with an additional report that shows which factors influence railroad crossing incidents. According to Pappas and Whitman (2011), a dashboard can best be created as an artefact since it can be used to display and track specific data.

## 1.4. Reality and norm

### 1.4.1. Reality

Reality can be seen as the current situation. When incidents happen, data is gathered by a general leader, which is first transferred to Systems, Applications and Products (SAP), after which it is stored in ProVat. ProVat stands for ProRail, Safety Information, Current, and Transparent (Prorail, 2021b). The reality has been determined by using data from ProVat. Within this data, there is a distinction between the types of railroad crossing incidents. When looking at all categories of railroad crossing incidents from 1975-2021, the red line in Figure 2 is formed. However, according to ProRail (n.d. -a), collisions with crossing railroad users is the most important category. Therefore, a separate line for these incidents is figured, which gives the blue line in Figure 2. In this figure the y-axis is the number of incidents and the x-axis is the time in years. As can be seen, there is a downward trend in railroad crossing incidents. However, ProRail still wants to decrease this number substantially.

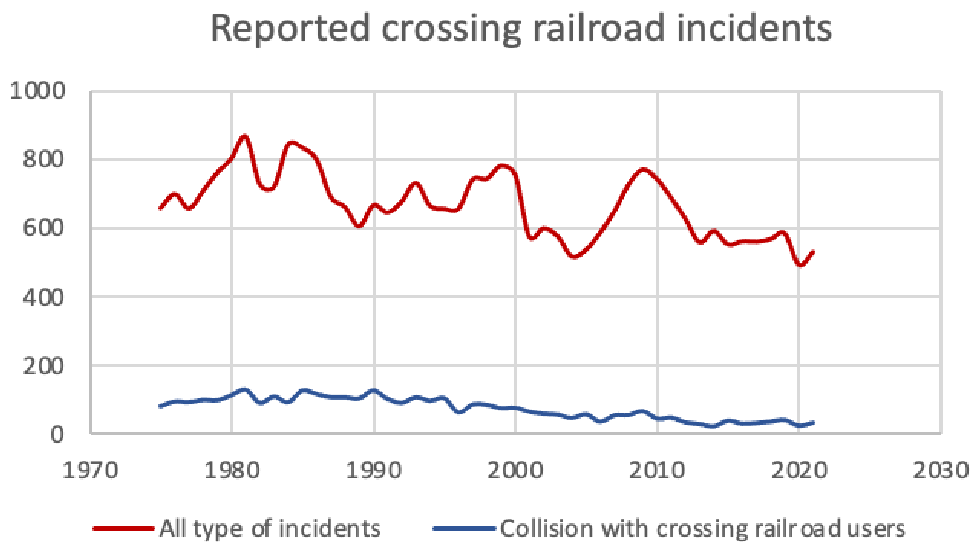


*Figure 2: Reported crossing railroad incidents*

### 1.4.2. Norm

The norm can be seen as the desired situation. This means that ProRail wants to be in this situation. For determining this norm, several sources are used. When looking at incidents in general, ProRail wants zero incidents per year. However, this is not feasible; therefore, they use the ALARP principle. This principle stands for 'As low as Reasonably Practicable' (van der Schans, 2001). This means that ProRail aims for zero incidents; however, since this is not feasible, they will do everything within reasonable limits to decrease the number of incidents as much as possible. Therefore, when looking at all the incidents at railroad crossings combined, there is no numeric norm ProRail wants to achieve.

However, as stated before, the collisions with crossing railroad users are important for ProRail. Therefore, ProRail makes norms for this every year. Last year, in 2021, this norm was 24, which was exceeded by eight collisions. For 2022, again, the norm is 24 collisions. These norms are based on numbers of past; however, no calculations are used (ProRail, 2021).

## 1.5.Research Methodology

The problem-solving approach used within this thesis is the Design Science Research Methodology (DSRM). This methodology has three objectives: consistent with prior literature, providing a nominal process model for design science research, and providing a model or template for research (Peffers et al., 2007). Since an artefact, the dashboard, is created to solve the core problem, the DSRM is chosen. The DSRM has six steps. During these phases, several knowledge problems arise. According to Heerkens & van Winden (2017), a knowledge problem is a description of the research population, the variables and, if necessary, the relations that need to be investigated. The following phases and knowledge problems are identified:

1. Problem identification and motivation

Within this phase, the problem is stated, and the current and desired situation is discussed (Peffers et al., 2007). Identifying the core problem has already been done in section 1.3. Besides that, also the current and desired situation have been reviewed.

2. Definition of the objectives for a solution

This research aims to develop a dashboard and a report that increases insight into railroad crossing incidents. Within this phase, the objectives for the solution are searched for and assessed to whether they are feasible (Peffers et al., 2007). Therefore, it is essential to know if there already are current solutions and, if so, their efficacy. This leaves the following question:

   2.1. Has there already been (quantitative) research into railroad crossing incidents?
   2.2. If research has been conducted, did ProRail use the findings, and if so, did it decrease the number of incidents on and around railroad crossings?

These questions have already been answered in section 1.2. There it has been concluded that research has been conducted. The conclusion from this research is that removing railroad crossings decreases the railroad crossings incidents; however, it does not contribute enough.

3. Design and development

During this phase of the DSRM, the artefact its desired functionalities and architecture are determined. Besides that, the artefact itself is also created during this phase (Peffers et al., 2007). This leaves the following questions:

   3.1. What steps need to be taken during data preparation?
   3.2. What key metrics should be selected to best show the relations and correlations between data of incidents?
   3.3. How to create a suitable dashboard?
   3.4. How to find a correlation between key metrics?

The findings to these research questions can be found in multiple chapters. The first research question is answered in sections 2, the second research question is answered in section 3, the third research question is answered in section 4 and the last research question in is answered section 5.

4. Demonstration

The demonstration phase includes using the artefact through experimentation, simulation, case study, proof, or other activities to solve the problem (Peffers et al., 2007). This leaves the following question:

   4.1. How to make conclusions that meet the criteria of the Safety department?
   4.2. What are the factors that influence safety on and around railroad crossings?

Both these research questions are answered in chapter 6, which discusses the findings that are discussed in previous chapters.

5. Evaluation

Within the evaluation phase, whether the artefact supports the solution to the problem is observed and measured. If not, it is possible to adjust the artefact and return to the third phase (Peffers et al., 2007). This leaves the following question:

5.1. How to evaluate the dashboard, statistical analyses and recommendations?

This research question is answered in section 7, which is the conclusion of this thesis.

6. Communication

The last phase is communicating the problem and its importance to relevant stakeholders. The relevant stakeholders within this thesis are ProRail, especially the departments of Safety and Business Application, and the University of Twente. Therefore, the opinion of relevant stakeholders on the recommendations is researched.

These six steps from the DSRM, combined with the research questions, will fill a research gap. A research gap is defined as an area in which there is missing or inadequate information (Carey et al., 2015). This research will apply the scientific literature and methods to incident data of railroad crossings at ProRail. This is something that has not been done before, and therefore, it fills the research gap.

# 2. Data Preparation

This chapter describes the preparation of the data for creating the tool. First, the theoretical framework is given, which shows what steps need to be taken. After that, in sections 2.2 and 2.3, the process of preparing the data for this thesis is explained. Therefore, the following research question is answered in this chapter: '*What steps need to be taken during data preparation?*'.

## 2.1. Theoretical framework

When decisions need to be made based on data, this data should first be prepared. Data preparation can be defined as the process of combining and structuring data (Sadiku & Musa, 2021). This means that before data can be used to solve a problem, it should be collected and cleaned (or: cleansed).

Several methods can be used when gathering data, such as surveys, interviews, focus groups, databases, and open sources (Canals, 2017). Within this thesis, especially the latter two are of importance. The first method being used is closed access databases. Databases are a well-known method of data collection. Databases are an organised data collection (Zaw et al., 2019). ProRail has several databases, which are only accessible with a password. Besides closed databases, open sources are used. According to Williams & Blum (2018), open-source intelligence (OSINT) can be defined as publicly available information collected and exploited. This method of data collection is considered acceptable in research.

Data cleaning is the process of detecting and removing errors and inconsistencies from data in order to improve the quality of a dataset. This can be done by transforming the data. This means adjusting data in such a way that it can be used for analysis. After that, it should be verified whether the transformation is done sufficiently and the data is correct and effective (Rahm & Do, 2000).

## 2.2. Data gathering

To gather the data necessary to conduct the research, several databases of ProRail are used. The first and foremost database is ProVat. This is the central system of ProRail, where it stores its information about safety incidents, safety inspections, and safety audits (ProRail, 2021b). ProVat has the information on every reported incident starting from 1975. Within this research, data of railroad crossing incidents are used between 01-01-2010 and 01-03-2022, which resulted in 7301 reported incidents. This data from ProVat consists of information on for example the time, the location, and the category of the incident.

Besides this database, also the registers that contain all the information on railroad crossings are used. These registers have information on, for example, the location, type of security, and frequency of the train. Besides that, they also contain other information such as, for example, 'an effect score: located in arc' and 'the traffic situation'. The registers are made to improve the risk profile and get a ranking of all railroad crossings in the Netherlands (ProRail, 2019). These registers solely contain information on railroad crossings and do not have information on the incidents that happen on these railroad crossings.

The last method being used is open sources; in particular, the open sources for sunset, traffic conditions, sunset and maps.

## 2.3. Data cleaning

Since the data is extracted from several sources and databases, data cleaning is necessary. The cleaning process is done in Microsoft Excel and its Microsoft programming language Visual Basic for Application.

First, the location of the incidents needed to be linked to the location of the railroad crossing. Therefore, the registers and the registered incidents in ProVat data were used. ProRail has denoted sections on the railway network by geocodes and kilometre notations. The geocode describes a specific area within the railway network. The ProVat data containing the information of the incidents consist of a specific kilometre notation and a geocode of the location of the incident. The railroad crossing registers also show the kilometre notation and the geocode of every railroad crossing; besides that, it also shows the municipality, city and street name of all the railroad crossings.

However, there are two problems when trying to link the location of the incidents registered in ProVat to the railroad crossings registered in the registers. As stated before, ProRail has divided the rail net of the Netherlands into geocodes. Figure 3, shows these geocodes for a small central part of the Netherlands. However, within one geocode, there are often multiple railroad crossings. Therefore, as shown in Figure 4, the rail net also has kilometre notation.

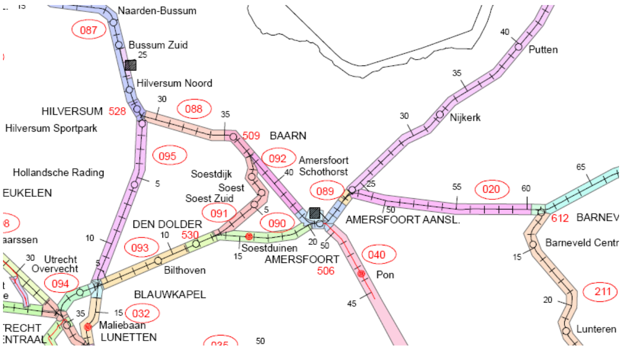


*Figure 3: Geocode for small part Netherlands (Infrasite, 2021)*

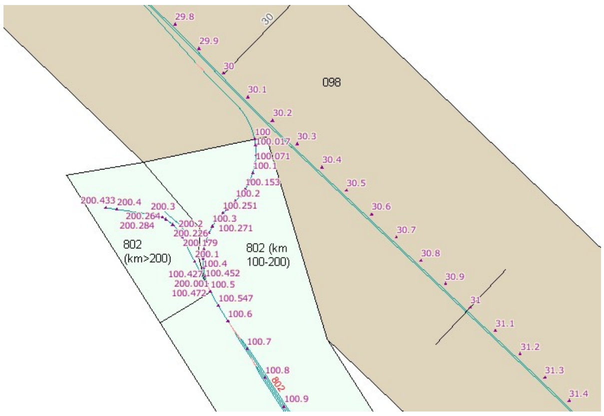

*Figure 4: Kilometre notation for one track of the Netherlands (Infrasite, 2021)*

The first problem is that a location is only specific when combining the geocode and the kilometre notation. This is because multiple railroad crossings have the same geocode or kilometre notation as explained above. Only a combination of these two is unique. Besides this problem, as stated before, incidents registered in ProVat contain a specific kilometre notation. This means that if a railroad crossing is located for example at kilometre notation 30.3, and the incident happens 10 metres past this railroad crossing, the location of the incident is notated at 30.31. This makes simply searching for the same geocode and kilometre notation impossible.

This problem cannot be solved by searching the nearest value to the kilometre notation since several geocodes can have an (approximately) similar kilometre notation. Therefore, using this selection, one must first select the geocode and search for the nearest kilometre notation. This cannot be done using (existing) Microsoft Excel functions. To solve this problem, a code is written in Visual Basic Application, which can be found in Appendix A.

This code first finds all the rows where the geocode of the railroad crossings in the register matches the geocode of the reported incident. If this is the case, the corresponding kilometre notations of the register are being used to calculate all absolute values of the difference between the kilometre notation of the incident registered in ProVat and the kilometre notation of the railroad crossing registered in the registers. The minimum value of these absolute values shows the railroad crossing where this incident happened. Using the Index Match function, the corresponding municipality, city and street are linked to the incident.

The open sources are implemented in Microsoft Excel, after which basic functions such as Mid and Search are used to get the data ready for use. Once this is done, a lot of the data cleaning can be done using Microsoft Excel functions. The Index Multiple Match function proved helpful in linking information to the incidents.

# 3. Metrics Selection

This chapter will discuss the metrics and key metrics used in the dashboard. The following research question will be answered in this section: What key metrics/metrics should be selected to best show the relations and correlations between data of incidents? The theoretical framework will elaborate on the requirements these metrics should have in section 3.2. The key metrics and metrics will be stated, and a few will be elaborated on.

## 3.1. Theoretical framework

When creating a visual-based artefact, in this thesis, a dashboard, metrics and Key Performance Indicators (KPIs) are used (Kerzner, 2017). To know which is best for the dashboard regarding railroad crossing incidents, first, it is important to know the difference. KPIs are tools that can discover and understand potential bottlenecks and opportunities based on achieving goals (Kaganski et al., 2017). Metrics are performance measurements that monitor the progress of activities (Chiesa et al., 2009). A project (or: dashboard) can have too many KPIs. This should be lowered to a handful. However, a dashboard can have many metrics. This is because a KPI is a special kind of metric and because good metrics can support KPIs (Schiff, 2008).

This gives that a metric is a simple measurement of the progress of business activities. Therefore, a metric can be any indicator. However, before an indicator can be a KPI, certain restrictions should be kept in mind. The most important aspect is that a KPI needs to be Specific, Measurable, Achievable, Realistic, and Time-bound (SMART) (Shahin & Mahbod, 2017). Therefore, the KPI should be defined precisely and clearly. It should be measurable in a selected unit, which means it should be expressed in, for example, a fraction or percentage. Besides that, the goal of the KPI should be attainable given certain conditions. Lastly, the KPI should contribute to attaining the general objective and be given in specific time stages (Podgórski, 2015).

At this point, within ProRail, there is no documentation on target numbers, except for the collisions with railroad crossing users. This makes creating a specific KPI outcome impossible. There is elaborated on this topic in the section 'Further Research. Since the measures within this thesis cannot be SMART, it is decided that the dashboard will not exist of KPIs and metrics but rather of key metrics and metrics. According to Kotarba (2017), a key metric is an important metric; therefore, this key metric does not have to be SMART.

Within this research, two semi-structured interviews were conducted to get the optimal key metrics and metrics for this research. The first interview was with an employee of the Safety department, who specialises in railroad crossings. The second interview was with an employee of the Competence Centre Business Application. Both these semi-structured interviews started with the following question: 'What metrics and key metrics do you believe have a negative influence on the number of incidents on and around railroad crossings?'. These interviews showed that there are certain subjects in railroad crossing incidents that ProRail wants to decrease, besides the overall goal of decreasing the number of railroad crossing incidents. These subjects are used to develop key metrics and metrics. These interviews gave that, according to the respondents, there should be aimed for fewer primary incidents, fewer incidents at unsecured railroad crossings and fewer incidents in the dark. With this information, key metrics and metrics are derived to track these outcomes.

## 3.2.Formulation of the key metrics

The key metrics and metrics used for the dashboard are elaborated on within the section. These key metrics and metrics are selected during semi-structured interviews while keeping in mind the theory explained in the theoretical framework. This resulted in the thirteen measures (key metrics and metrics), as shown in Table 1.

Many of these measures do not need further elaboration. However, there are a few measures that do need some explanation. First, the key metrics will be discussed. The first key metric is *'Type of security'*. Railroad crossing can be secured or unsecured. There are approximately 1,900 secured and 350 unsecured railroad crossings (ProRail, 2017). The best-known security type is the gate with red flashing light signals; however, besides this one, there are many other types, such as, for example, fences. This key metric has two layers; the first is whether the incident happened at a secured or unsecured railroad crossing. The second layer gives the type of security. The extended list of security types, and their abbreviations, can be found in Appendix B. The second key metric is '*Category of incident'*, which again consists of two different layers. ProVat divides every railroad crossing incident into three main categories: primary-, secondary- or almost incident. These three main categories are also divided into many subcategories. All different types of categories and subcategories can be found in Appendix C. However, besides key metrics that are determined by using existing data and during the semi-structured interviews, there are also key metrics that are developed especially for this thesis. The first key metric being *'SPAD before an incident*'. A SPAD stands for a Signal Passed at Danger, which occurs when a train passes a stop signal (Punzet et al., 2018). For this key metric, all data from SPADs are extracted from ProVat between 01-01-2010 and 01-03-2022. This data compares the time and location (based on geocoding) to the time and location of all incidents. If the geocode of the SPAD and the incident were the same, and a SPAD occurred one, two, three, four, five or ten minutes before an incident, this is notated as a SPAD before an incident. The second key metric that was developed for this thesis is the '*Light situation*'. For the light situation, an open source is used. This open-source contained all sunrise and sunset data from Utrecht, a central location in Holland, between 01-01-2010 and 01-03-2022. With this data and the time of the incidents extracted from ProVat it is determined whether the registered incidents happened in the dark between sunset and sunrise, in the civil twilight or during the day when it is light.

Besides these four key metrics that need elaboration, also two metrics need some elaboration. The first metric is *'Second train'*. Within the railroad crossing registers discussed before, information such as *'Second train'* is given using an effect score. Within this register, this value is calculated using the type of railroad crossing, the number of tracks, the frequency, the speed of the train and the location in an arc. With these values, the following formula is used to calculate the value of 'second train': (Type of railroad crossing * number of tracks * frequency * speed of the train * located in arc)*35/29,5. This is a computation made by ProRail; since this dashboard is made for ProRail, these values will be used within this metric. The second metric that needs elaboration is '*Located in Arc*'. A railroad crossing can be located in an arc. This metric again is an effective score. This score goes from a minimum value of two to a maximum value of three. Within this score, two means there is no arc, and three means an arc with a radius of 0 - 500 metres. This score can also be 2.1 (which means an arc with a radius larger than 2000 metres), 2.3 (which means an arc with a radius of 1000 - 2000 metres), and 2.5 (which means an arc with a radius of 500 - 1000 metres).

| **Subject** | **Key metric/Metric** |
|---|---|
| Type of security | Key metric |
| Gender of railroad crossing user | Metric |
| Category of incident | Key metric |
| SPAD before incident | Key metric |
| Frequency of railroad crossing | Metric |
| Distance to station | Key metric |
| Line speed of railroad crossing | Key metric |
| Second train | Metric |
| Located in arc | Metric |
| Light situation during incident | Key metric |
| Location | Metric |
| Date and time | Metric |
| Angle railroad crossing to freeway | Metric |

*Table 1: All Key metrics and Metrics*

# 4. Dashboard

This chapter will discuss the dashboard. The following research question will be answered in this section: *How to create a suitable dashboard?* The theoretical framework will explain what is needed when creating a dashboard. In section 4.2, the created dashboard will be elaborated on, and there will be explained how to use it.

## 4.1. Theoretical Framework

When conclusions are drawn from a dashboard, its design is important. Within a dashboard, many different figures, maps and charts can be visualised. Several aspects should be kept in mind for creating the best possible dashboard. First of all, the usability of the dashboard needs to be taken into consideration. The second is the type of visualisation tool used. It is important to keep in mind the possibilities of the data and the limitations of the tool used to create a dashboard. For creating the dashboard, Microsoft PowerBI is used. Last, the layout is of importance when designing a dashboard (Gesing et al., 2014).

When researching the usability of a dashboard, first, it is important to know the definition of the term usability. The term usability is specified in ISO 9241-11, a multi-part standard from the International Organisation for Standardisation, and gives a framework for understanding usability (Hui et al., 2020). Usability gives the degree to which selected users can easily use a dashboard to achieve goals with effectiveness, efficiency, and satisfaction (Bevan et al., 2015). Therefore, it is important to focus on effectiveness, efficiency, and satisfaction when designing a dashboard. Effectiveness and efficiency can be ensured by choosing the right visualisation tools; satisfaction can be ensured by creating a good layout (Kirk, 2012).

When the data has been cleansed, several figures, maps and charts can be created to make the data's information, relations and correlations visible. When selecting a visualisation tool, it is important to know the purpose of the tool and whether this is possible with the selected data (Saraiya et al., 2004). Within this research, Microsoft PowerBI is used, which has several visualisation tools available, which can be found in table 2.

| Chart Type | Function/ explanation |
|---|---|
| Area charts | Emphasise the magnitude of change over time and can be used to show a trend |
| Bar and column charts | Show specific value across different categories |
| Cards | Can display single or multiple facts or data point |
| Combo charts | Combines a column and a line chart for quicker comparison of data |
| Decomposition tree | Visualises data across multiple dimensions and can drill down |
| Doughnut and Pie charts | Show relationship of parts to a whole |
| Funnel charts | Visualise a process that has stages, and items flow sequentially from one stage to the next |
| Gauche charts | Displays a single value that measures progress towards a goal |
| Key influencers chart | Displays the major contributors to a selected result or value |
| KPIs | Visual cue that shows the amount of progress made towards a goal |
| Line charts | Emphasise the overall shape of an entire series of value (over time) |
| Basic, ArcGIS, Azure, Filled and Shape map | All these types of maps have different functions but they all show data and information with locations. |
| Matrix | Table visualisation that supports a stepped layout |
| Scatter(-high density), bubble and dot plot chart | All types of scatter charts that display points at the intersection of an x and y numerical value, combining these values into single data points. |
| Slicers | Used to filter the other visuals on the page |
| Smart narrative | Adds text to reports to point out important aspects |
| Standalone image | A figure |
| Tables | Used for camping detailed data and exact values |
| Tree maps | Charts of coloured rectangles, with size representing values |
| Waterfall charts | Shows a running total as values are added or subtracted |

*Table 2: Visualisation tool types (Microsoft, 2021)*

When the visualisation tools are created, the next step is to develop the layout of the dashboard. A dashboard can have the purpose of pushing and pulling information. If a dashboard has to pull information, it is designed for the user to get information through the dashboard. On the other side, a dashboard can be meant to push information, which means a dashboard is created to catch the users' attention. This dashboard is designed to push important information to the user (Janes et al., 2013). Within this thesis, the users are the employees of the Safety department of ProRail. Because the goal of the dashboard is to push information, firstly, the dashboard should be visible without effort. Besides that, users should not need to interact with the visualisation to understand data. This is only permitted if they want extra information on a topic. Further, the dashboard should be designed so that minimal time is needed to consult the dashboard. This means that information on one topic should be in the same place. Lastly, the aesthetic factors are also important when making the dashboard layout. This means it should be kept clean and simple and only highlight important data using pre-attentive processing (Janes et al., 2013).

## 4.2. The dashboard design

The dashboard is made using an interactive data visualisation tool called Microsoft PowerBI. Figure 5 shows the dashboard. Several choices have been made to ensure the usability, type of visualisation tools and layout. The thirteenth measures discussed in chapter 3 are all displayed on one page to minimise the effort needed to understand the dashboard. Besides, the user does not need to interact with the visualisation tools to understand the data. This is only needed if the user wants to gain more knowledge on one metric and its influence on the other or to go into deeper layers. Besides that, the dashboard has a simple red layout that matches the company's colour and does not have too many different visualisation tools.

Within this dashboard, every key metric or metric is linked with each other, making it interactive. This means that if the user only wants to see information on the incidents that happened when it was dark outside, the user simply needs to press this part of the pie chart. If the user does that, every other visualisation tool change to incidents that happened during the dark. This applies to every metric within this dashboard. Figure 6 shows an example where the incidents that happened during the dark are selected.

The header contains the logo of ProRail, a name for the dashboard and a slider for the time. With this slider, the user can adjust the range of the timespan visible on the dashboard. Therefore, it can only show data in a certain timeframe. Below the header, the thirteen metrics and key metrics explained in section 3.2 are shown. First, the '*Location*' metric is displayed on a map and in a matrix with the corresponding percentages compared to the total number of incidents happening on railroad crossings. The '*SPAD before incident*' is displayed on a simple card. The '*Line speed of railroad crossing*', '*Angle railroad crossing to freeway*', and '*Second train*' are also shown in matrices with a corresponding percentage. The metrics and key metrics' *Frequency of railroad crossings*', '*Located in arc*', '*Distance to station*', and '*Date & time*' are shown as bar charts. The last visual used is the pie/doughnut chart for the '*Light situation during incidents*', '*Gender of the railroad crossing user*', '*Type of security'*, and '*Category of incident*'. The doughnut charts contain two levels, one with the main categories and type of securities and one with the subcategories and subtypes of securities.

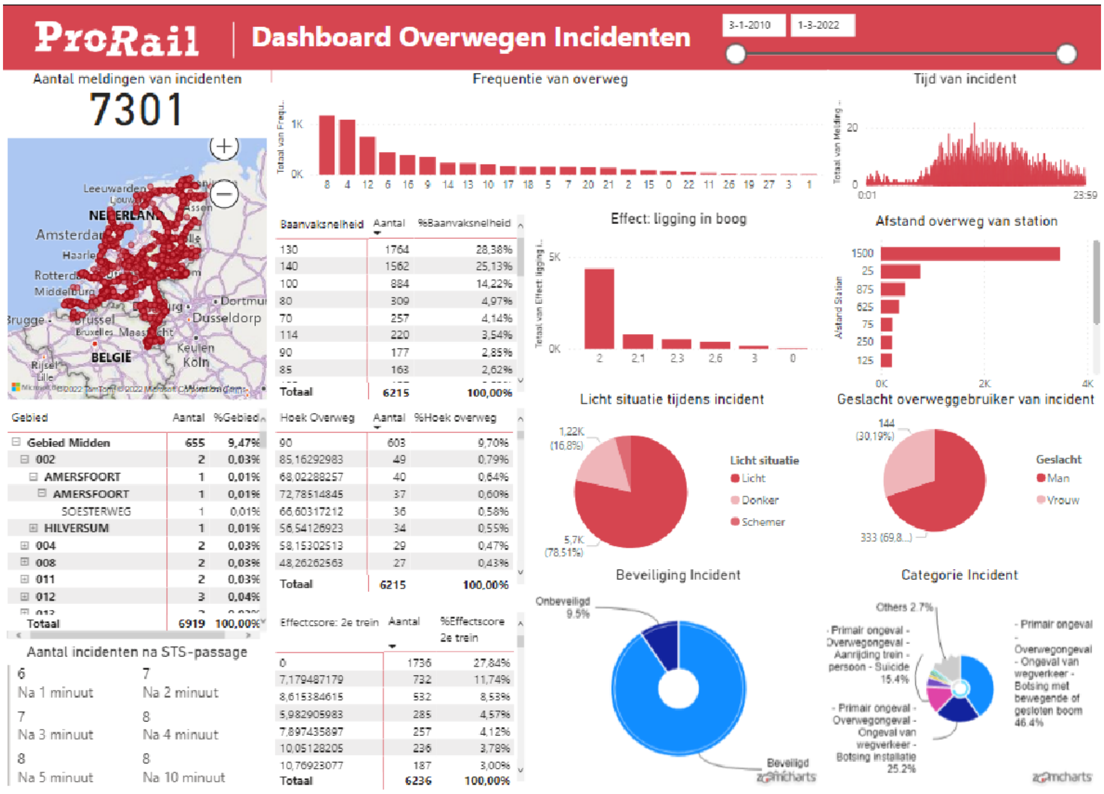


*Figure 5: Dashboard 'Railroad Crossings Incidents'*

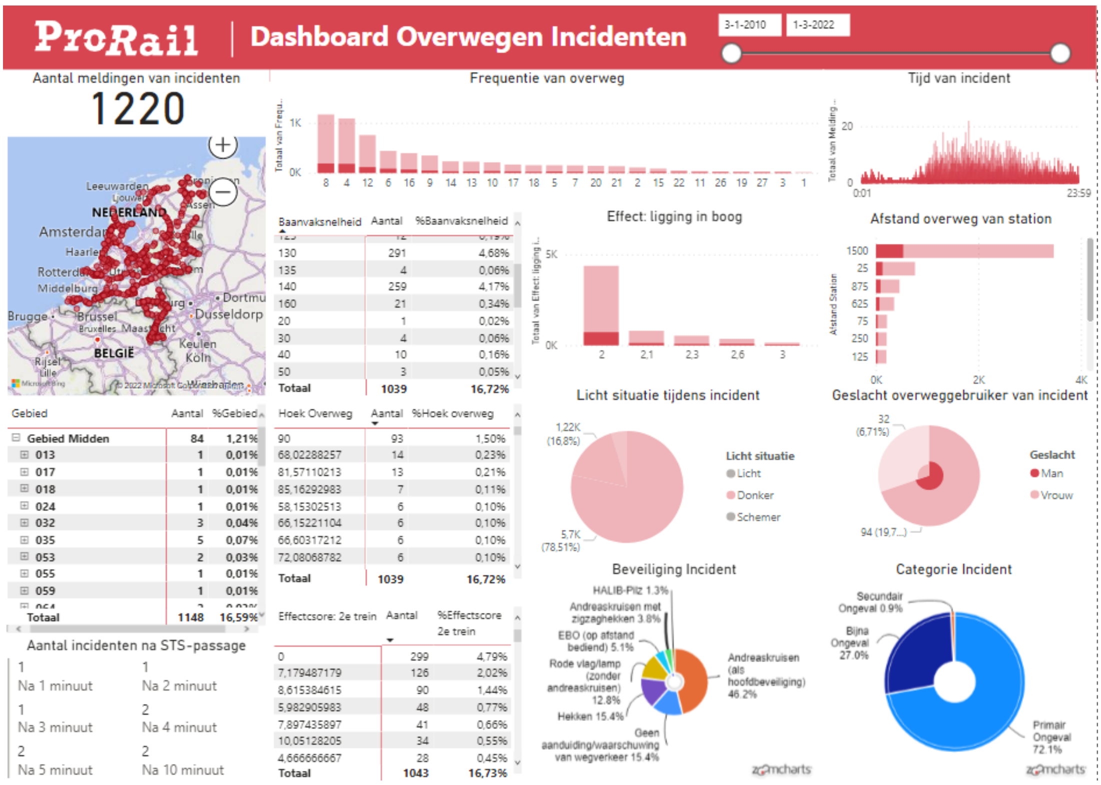


*Figure 6: Dashboard 'Railroad Crossings Incidents' with only incidents that happened during the dark selected*

# 5. Statistical Analysis

This chapter will discuss the statistical analysis. The theoretical framework will explain what correlation is and what different tests can be used. Section 5.2, there is elaborated on the correlation between the key metrics. The following research question will be answered in this section: *'How to find a correlation between key metrics?'*

## 5.1. Theoretical framework

The dashboard created shows several measures and calculations. With this dashboard, certain conclusions can be drawn, and some recommendations can be made for every measure and key metric. However, to show if patterns and trends between data exist, statistical analysis needs to be performed (Egger & Carpi, 2008). To expand this research, within this thesis, there is not looked at every single key metric on its own since this already has been done at the dashboard. However, there is analysed on whether there is a statistical correlation between the key metrics. This is something the dashboard does not show and therefore it expands this research. According to Gelderman (1998), a correlation is a relationship between two or more variables.

Several methods can be used to perform statistical analysis, such as mean, standard deviation, regression, and hypothesis testing (Gnanadesikan, 2011). The latter will be used within this thesis since it is used to define the relationship between two or more variables (Mourougan & Sethuraman, 2017). Within this method, a claim is made, which is called a hypothesis, and by a procedure, it is tested whether this claim is false (Krishnan & Idris, 2015). This procedure can be one-tailed or two-tailed, depending on the claim. With a one-tailed test, there is only one end of the sampling distribution involved, so whether it is higher or lower than the critical value. Therefore, this can be left- or right-tailed. With the two-tailed test, there is a look at both sides of the sampling distribution. The above-mentioned critical value is the score that separates the rejection region from the rest of the curve; this is based on the distribution of the test statistic and the significance level (Mourougan & Sethuraman, 2017). Figure 7 shows figures of a left-, right- and two-tailed test, the critical value and the rejection region.

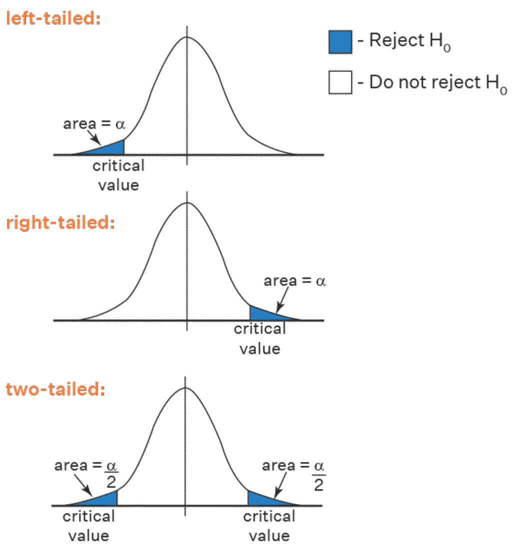


*Figure 7: Graphs of left-, right- and two-tailed test (Cuemath, 2022)*

As stated before, there are several procedures (or: tests) one can do to perform hypothesis testing. Depending on several factors, the proper test can be chosen. First, the parametric assumptions need to be decided on; after that, the number and kind of variables need to be determined. The last step before choosing the right statistical test is looking at the hypothesis that will be tested (McDonald, 2009). Figure 8 shows a flow diagram with all choices needed within this thesis to find the proper statistical tests.

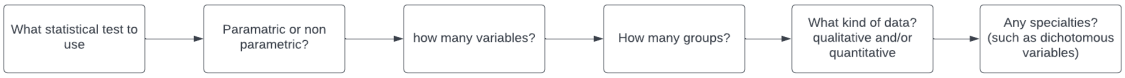


*Figure 8: Flow diagram with all choices find right statistical tests*

A statistical test can either be parametric or nonparametric. A parametric statistical test assumes that data is normally distributed. Nonparametric statistical procedures have no assumptions (Hoskin, 2012). Data can be classified under the following types: nominal, ordinal, interval, and ratio (Freelon, 2013). When data is nominal or ordinal, it is assumed to be categorical (Suparji et al., 2021). Whereas interval and ratio data can be categorized as continuous (Kaur, 2018). When data is ordinal or nominal, which applies to most of the data within this thesis, only a nonparametric test can be used (Singh et al., 2013).

The key metrics are all ordinal or nominal except for the key metrics '*Line speed of railroad crossing'* and '*Distance to statio*n', which are ratio types. Therefore, a nonparametric as well as a parametric test, can be used. However, as stated before, the data must be normally distributed using a parametric test. To test whether these key metrics are normally distributed, a Kolmogorov-Smirnov test can be used (Ghosh et al., 2016). Table 3 shows the outcome of these tests for the '*Line speed of railroad crossing*' and '*Distance to station*' key metrics. If the p-value is lower than the alpha value of 0.05, the data is not normally distributed (Hamsal et al., 2021). According to Panagiotakos (2008), the p-value is defined as the probability of getting a result that is at least as extreme as the one being observed. Whereas alpha is defined as the threshold for statistical significance (Lakens et al., 2018). Since the p-value (in table 3 'Sig.') equals .000 for both key metrics, the data from these key metrics are not normally distributed. To confirm this conclusion, another procedure can be done to test whether a dataset is normally distributed. If a dataset is normally distributed, the histogram is a bell-shaped curve with one peak (Barri, 2019). As can be seen in figure 9 and figure 10, for both key metrics, this is not the case. Therefore, it can be concluded that only nonparametric tests will be used within this thesis.

| | **Statistic** | **df** | **Sig.** |
|---|---|---|---|
| **Line Speed** | .260 | 6215 | .000 |
| **Distance to station** | .357 | 6215 | .000 |

*Table 3: Test of Normality: Kolmogorov-Smirnov test for 'Line speed of railroad crossing' and 'Distance to station'*

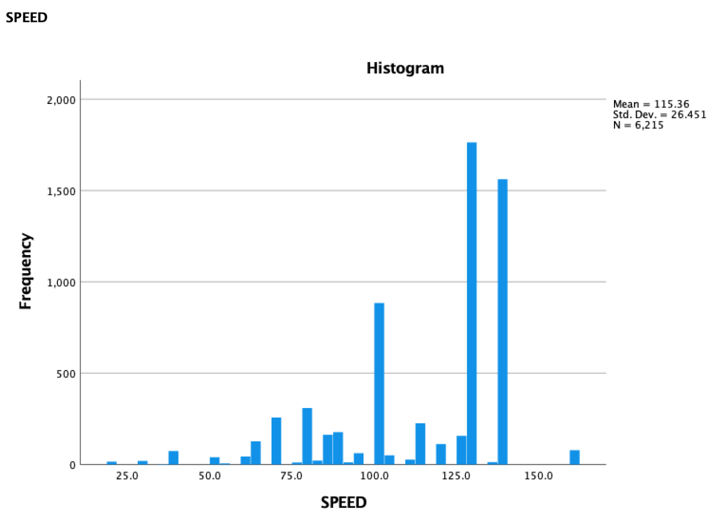


*Figure 9: Histogram of 'Line speed railroad crossings'*

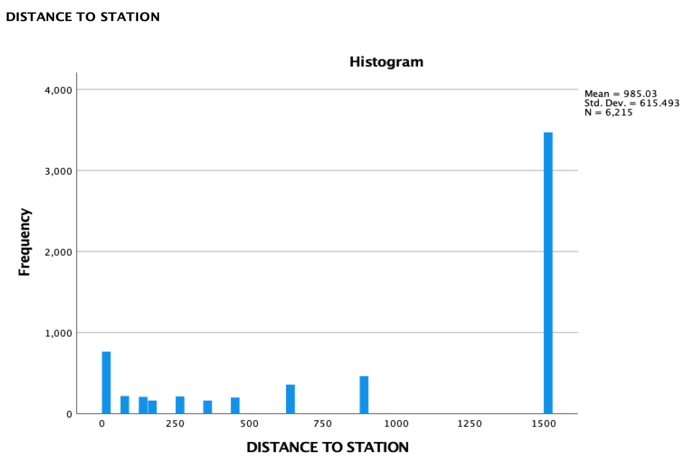


*Figure 10: Histogram of 'Distance to station'*

Since two variables are involved in testing whether they correlate, a few tests remain optionable. Since the dataset consists of quantitative and qualitative data, the following tests will be used: Kruskal-Wallis test, Chi-square test of independence, Fisher's exact test and Spearman's rank correlation (Nayak & Hazra, 2011). According to Kothari (2007), the chi-square test of independence can be used to find if there exists a correlation between non-numeric variables. When using the chi-square test, there are a few requirements for the data. First, the observations must be collected randomly, and the sample must be independent. Besides, the sample must have at least 50 observations, and none of the different groups can have too few observations (Nihan, 2020). A Chi-square test compares the observed values with the expected values to test if they are significantly different. The Chi-square value increases as the difference between the observed and expected values increases (Kothari, 2007). In Statistical Package for Social Science (SPSS), the null hypothesis, which claims no association between groups, is rejected if the 'Asymptotic Significance' value is lower than the alpha value (Gajanova et al., 2019). Within this thesis, an alpha level of 0.05 is chosen.

Phi can be used to measure the strength of the association between the two groups. Phi can only be used in a 2x2 contingency table. If the table is more extensive, Cramer's V can be used to express the strength of the association (Daneshpazooh et al., 2006). Phi and Cramer's V have a value between zero and one, and the higher the value, the stronger the relationship. Table 4 shows the different stages of association strengths (Akoghlu, 2018).

| Phi and Cramer's V value | Interpretation |
|---|---|
| 0 - 0.05 | No or very weak |
| >0.05 - 0.10 | Weak |
| >0.10 - 0.15 | Moderate |
| >0.15 - 0.25 | Strong |
| >0.25 | Very strong |

*Table 4: Interpretation of Phi and Cramer's V values*

However, there is an exception for using the Chi-square test. If more than 20% of the expected cell counts are less than five, Fisher's exact test should be used (Nowacki, 2017). The Fisher's exact test runs an exact procedure, whereas the Chi-square test runs an approximation (Kim, 2017). However, the interpretation of the results works the same as a Chi-square test.

Besides the Chi-square test and Fisher's exact test, Spearman's rank correlation will also be used. This test can be used with ordinal data (Schober et al., 2018). The Spearman correlation coefficient can take on a value between minus one and one, where the minus one gives a negative correlation, zero no correlation and one a positive correlation. The absolute value of this number gives the strength of the correlation (Zou et al., 2003). Table 5 again shows different stages of association strengths (Dancey & Reidy, 2007).

| Spearman value | Interpretation |
|---|---|
| 0 - 0.19 | No or negligible |
| 0.20-0.29 | Weak |
| 0.30-0.39 | Moderate |
| 0.40-0.69 | Strong |
| >0.69 | Very strong |

*Table 5: Interpretation of Spearman values*

These three tests cover most of the key metrics and can show whether there is an association between them. However, no test is discussed that can show a correlation between a continuous and a categorical variable. When testing a continuous and a categorical variable, several tests can be used, such as the Kruskal Wallis test and the point biserial correlation test (Field, 2013). The point biserial correlation test includes a dichotomous (e.g. gender) variable (Ovwigho, 2013). Within this thesis, one key metric is dichotomous, and therefore the point biserial correlation test will be used. Just as Spearman's rank correlation test, the point biserial correlation test can take values between minus one and one (Brown, 2001). No point biserial correlation coefficient interpretations value can be found; however, according to Varma (2006), a value higher than 0.15 means that there is a correlation. This test covers one key metric but cannot be used by the other categorical key metrics since they exist in more than two groups. Therefore, the Kruskal Wallis test could be used; however, this test shows if there is a significant difference between the two variables (Chan & Walmsley, 1997). Therefore, this test does provide information on the correlation between the two variables. Since no test shows a correlation between a continuous and categorical variable, under the conditions of this thesis, it is decided not to test this.

## 5.2.Statistical tests

In the theoretical framework, all tests necessary to test whether there is a correlation between the key metrics are discussed. Within this thesis, the Chi-square test, Fisher's exact test and Spearman's rank correlation test are used when looking at the conditions on when to use what test; table 6 shows the test between every two variables.

| Key Metrics | Test |
|---|---|
| Type of security - Category of incident | Chi-Square test with Cramer's V |
| Type of security - Distance to station | No test |
| Type of security - Line speed | No test |
| Type of security - Light situation | Chi-Square test with Cramer's V |
| Type of security - SPAD before an incident | Fisher's exact test with Cramer's V |
| Category of incident - Distance to station | No test |
| Category of incident - Line speed | No test |
| Category of incident - Light situation | Chi-Square test with Cramer's V |
| Category of incident - SPAD before an incident | Fisher's exact test with Cramer's V |
| Distance to station - Line speed | Spearman's rank correlation |
| Distance to station - Light situation | Point Biserial correlation test |
| Distance to station - SPAD before an incident | No test |
| Line speed - Light situation | No test |
| Line speed - SPAD before an incident | Point Biserial correlation test |
| Light situation - SPAD before an incident | Fisher's exact test with Cramer's V |

*Table 6: Correlation test between every key metric*

For conducting these tests, SPSS is used. SPSS is software that can perform statistical tests (Bala, 2016). Therefore, no calculations are provided since the software does this. The outcome of all tests can be found in Table 7. The first layer is used for the key metrics with two layers (‘*Category of incident’* and *‘Type of security’*). Again, if the ‘Asymptotic Significance’ is lower than 0.05, there is assumed to be a correlation between the two variables.

| **Key Metric** | | **Type of security** | **Category of incident** | **Distance to station** | **Line speed** | **Light situation** | **SPAD before an incident** |
|---|---|---|---|---|---|---|---|
| **Type of security** | p-value | X | <0.001 | No test | No test | 0.04 | 0.642 |
| | Cramer’s V | X | 0.176 | No test | No test | 0.043 | 0.009 |
| **Category of Incident** | p-value | <0.001 | X | No test | No test | 0.038 | 0.715 |
| | Cramer’s V | 0.176 | X | No test | No test | 0.029 | 0.010 |
| **Distance to station** | Significant at 0.05 | No test | No test | X | Yes | No test | No |
| | Correlation Coefficient | No test | No test | X | 0.331 | No test | 0 |
| **Line speed** | Significant at 0.05 | No test | No test | Yes | X | No test | No |
| | Correlation Coefficient | No test | No test | 0.331 | X | No test | -0.10 |
| **Light Situation** | p-value | 0.04 | 0.0038 | No test | X | | 0.178 |
| | Cramer’s V | 0.43 | 0.029 | No test | No test | X | 0.016 |
| **SPAD before an incident** | p-value or significant at 0.05 | 0.0642 | 0.715 | No | No | 0.178 | X |
| | Cramer’s V or Correlation coefficient | 0.009 | 0.010 | 0 | -0.10 | 0.016 | X |

*Table 7: All correlation between key metrics*

As can be seen, there are several key metrics that have a p-value which is lower than 0.05. This gives that there is a correlation between the following key metrics: ‘*Type of security*’ - ‘*Category of incident*’, ‘*Type of security’* - ‘*Light situation*’, ‘*Category of incident’* - *‘Light situation’*, and ‘*Distance to station’* - ‘*Line speed’*. This gives that there is coherence between these two key metrics. A more in-depth analysis of the correlations can be found in the next section.

# 6. Results

This thesis has provided a detailed explanation of gathering and cleansing data, selecting key metrics and metrics, creating a dashboard and doing additional statistical analyses. Within this chapter, results will be discussed, conclusions will be drawn, and recommendations will be made. Therefore, within this chapter, the following research questions will be answered: *'How to make the right conclusions that meet the criteria of the Safety department?'* and *'What are the factors that influence the safety on and around railroad crossings?'*. First, a small theoretical framework will provide information on the process of concluding, after which the results of the dashboard and statistical analysis will be discussed.

## 6.1. Dashboard

A dashboard can have many different objectives (Kawamoto & Mathers, 2007). This thesis aims to show information on all selected key metrics and metrics on one dashboard. To draw results from this dashboard, the percentage of incidents per key metric per category is compared to the percentage of this category registered in the railroad crossing registers. If this percentage is significantly different, it is concluded that ProRail should further monitor these characteristics at railroad crossing incidents. Due to time limitations, it is decided to exclude all metrics for this section.

The first key metric being discussed is *'Line speed'*. According to Schumann (2016), *'Line speed'* is the maximum speed a track can support. Therefore, the maximum speed of the trains on tracks with a high-speed line will be higher. According to Wegman & Aarts (2006), speed is assumed to be one of the fundamental risk factors. Figure 11 shows the difference between the percentage of incidents at every possible line speed and the percentage of the total number of railroad crossings having this line speed. In this figure, the y-axis is the percentage difference and the x-axis are all line speeds categories. There can be concluded that there is no big difference between those since only the line speed of 100 kilometres per hour has a difference of more than 5%. However, under these circumstances, no real trend or other concerning observations can be made.

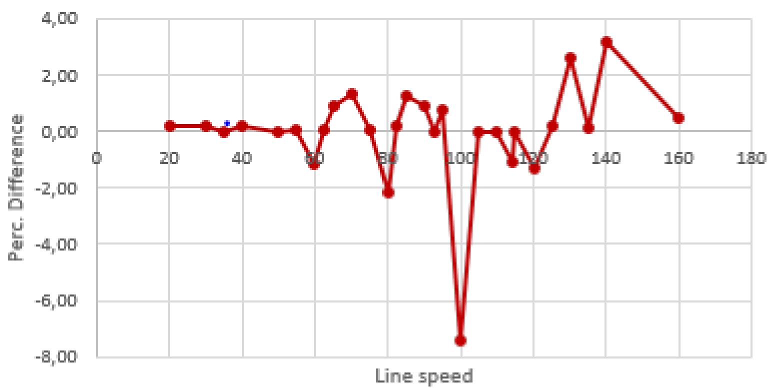


*Figure 11: Percentage difference key metric 'Line speed of railroad crossing' incident dashboard and registers*

The second key metric being discussed is *'Light situation'*. As discussed, the light situation is divided into light, dark, and civil twilight. For this key metric it is more difficult to show a significant percentage difference. The open source used for the dashboard shows that on the 1st of July (the middle of the year), it is 07:18 hours dark, the civil twilight is 1:16 hours, and it is 15:26 hours light. Assuming these are the average numbers, one could show the percentage difference. However, in the Netherlands, the trains do not drive all night (Nielsen et al., 2003). The trains stop at around 1 A.M. and start at around 05:30 A.M., leaving an average of only 02:48 of darkness. These assumptions give the percentages shown in table 8.

| Light situation | Dashboard (%) | Part of the day (without 1 A.M. – 5:30 A.M.) |
|---|---|---|
| Light | 78,51% | 64,3% |
| Dark | 16,8% | 11,67% |
| Civil twilight | 4,68% | 5,28% |

*Table 8: Percentage difference key metric 'Light situation' incident dashboard and registers*

However, the assumption of train traffic is not considered within these numbers. In the night and civil twilight, there is far less train traffic when comparing it to the train traffic when it is light. An assumption of 0.75 times the regular train traffic at night (a low assumption) gives 8,75% of the day is dark. Therefore, twice as many incidents happen when it is dark. Therefore, there is concluded that light situations do have an impact on the number of incidents. However, it should be kept in mind that not every factor (such as, for example, street lighting) is taken into consideration.

The third key metric being discussed is *'SPAD before an incident'*. There were only a handful of SPADs before the incidents at railroad crossings, approximately 0,08% of the registered incidents between 01-01-2010 and 01-03-2022. Based on these numbers, it can easily be stated that no conclusions can be drawn under these circumstances.

The fourth key metric being discussed is *'Distance to station'*. This key metric is divided into ten categories between 25 and 1500 metres. Figure 12 shows the difference in the percentage of incidents at every distance and the percentage of the total number of railroad crossings at this distance to the station. In the figure, the y-axis is the percentage difference and the x-axis are all distances to station categories. Again, under these circumstances, no conclusions can be drawn. The only relatively big difference is within the range of 1500 metres, which does not show anything since more than 65% of all railroad crossings have a distance to a station of 1500 metres. Therefore, a more considerable percentage difference is expected.

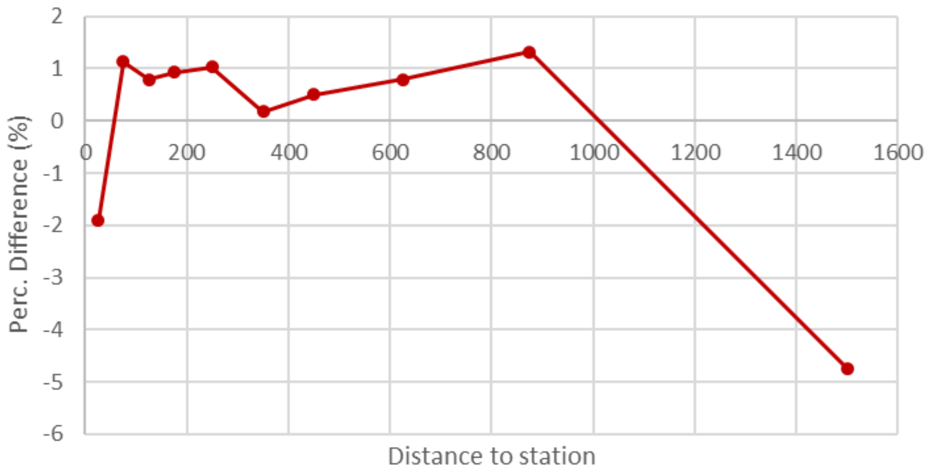


*Figure 12: Percentage difference key metric 'Distance to station' incident dashboard and registers*

The fifth key metric being discussed is *'Type of security'*, which consists of two layers. The first layer consists of the categories secured and unsecured. When looking at these categories, it can be seen that most incidents happen at secured railroad crossings. Especially when comparing it to the number of unsecured railroad crossings. Table 9 shows the percentage of secured and unsecured railroad crossings and the percentage of incidents at secured and unsecured railroad crossings. However, compared to the number of unsecured railroad crossings, this lower percentage of incidents at unsecured railroad crossings can be reasoned by the fact that of these railroad crossings, a lot is not used that much. Of the 3,108 railroad crossings, 1,259 are only used by freight trains. Therefore, an extra column is added with the percentages of railroad crossings, excluding the ones only used by freight trains. These percentages do not show a worrying difference; therefore, under the circumstances of this thesis, it is concluded that this first layer is not a factor that influences the safety on and around railroad crossings.

| Security type First layer | Dashboard (%) | Percentage of all railroad crossings | Percentage of railroad crossings (except the ones only used by freight trains) |
|---|---|---|---|
| Secured | 90,5% | 61,4% | 83,5% |
| Unsecured | 9,5% | 38,6% | 16,5% |

*Table 9: Percentage difference key metric 'Security type' first layer incident dashboard and registers*

As can be seen in Appendix B, there are many different security types for railroad crossings. Due to time restrictions, within this analysis, the three types with the most incidents, secured railroad crossings as well as unsecured railroad crossings, will be analysed to see if this influences the number of incidents. The dashboard shows that of the incidents at a secured railroad crossing, 77,2% happened at railroad crossings with an AHOB security, 15% with a mini AHOB and 2,5% with a sidewalk AOB. For the unsecured railroad crossings, 45,2% had Andreas crosses, 17,0% had fences, and 13,7% had a red light/flag. Figures 13 and 14 show the difference in the percentage of incidents at these three security types and the percentage of the total number of railroad crossings having this security type. There has been made a distinction between secured railroad crossings and unsecured railroad crossings.

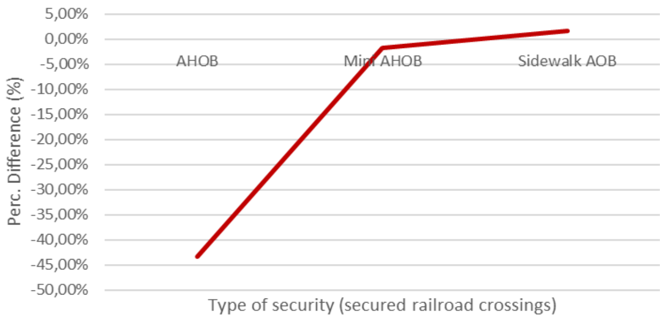


*Figure 13: Percentage difference key metric 'Type of security' incident dashboard and register secured railroad crossings*

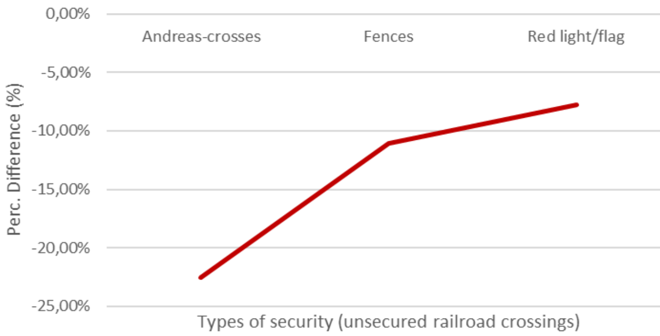


*Figure 14: Percentage difference key metric 'Type of security' incident dashboard and register unsecured railroad crossings*

In these figures, the y-axis is the percentage difference and the x-axis are the three most common security types. Both cases show there is one big difference in percentage. Figure 11 shows that in percentage terms, more incidents happen at an AHOB security type than secured railroad crossings with an AHOB security type. This significant difference can be explained because the railroad crossings that are only used by freight trains do not have AHOB security. Since there are a total of 1,259 railroad crossings used by freight trains, this impacts the percentage. On the other hand, figure 12 shows that many more incidents happen at Andreas-crosses security type than there are unsecured railroad crossings with Andreas-crosses as security type. This can be worrying since no explanation can be found even after a semi-structured interview. Therefore, this is something ProRail should keep an eye on.

The last key metric being discussed is *'Category of incident'*. This is the only key metric being discussed which is not classified as a characteristic of a railroad crossing. Therefore, this does not influence the safety of and around railroad crossings. However, as stated in section 1, ProRail wants to reduce all categories of incidents, but primary incidents have the biggest priority. Therefore, an analysis is made which shows every key metric and how it influences the primary incidents. This way, ProRail can know what categories of the key metrics mainly cause primary incidents. Table 10 shows, per category of a key metric, what percentages of the incidents are primary incidents. The key metric *'SPAD before an incident'* is excluded from this analysis since no primary incidents can be linked to this key metric.

| **Key metric** | **Perc. Primary incidents** | **Key metric** | **Perc. Primary incidents** | **Key metric** | **Perc. Primary incidents** |
|---|---|---|---|---|---|
| **Line speed** | | *92,5* | 100% | **Distance to station** | |
| *20* | 81,2% | *95* | 80,6% | *25* | 64,4% |
| *30* | 90% | *100* | 73,6% | *75* | 73,9% |
| *35* | 50% | *105* | 70,0% | *125* | 74,9% |
| *40* | 63,5% | *110* | 70,4% | *175* | 72,2% |
| *50* | 87,5% | *114* | 77,7% | *250* | 75,5% |
| *55* | 0% | *115* | 100% | *350* | 71,6% |
| *60* | 54,5% | *120* | 61,1% | *450* | 77,9% |
| *62,50* | 85,7% | *125* | 86% | *625* | 70,7% |
| *65* | 84,2% | *130* | 73,4% | *875* | 74,0% |
| *70* | 73,2% | *140* | 80,3% | *1500* | 77,8% |
| *75* | 63,6% | *160* | 88,5% | **Type of security** | |
| *80* | 60,2% | **Light situation** | | *Secured* | 77,2% |
| *82,5* | 63,6% | *Dark* | 70% | *Unsecured* | 61,9% |
| *85* | 62,6% | *Light* | 72,1% | | |
| *90* | 76,3% | *Civil twilight* | 74,5% | | |

*Table 10: Percentages primary incidents per category per key metric*

The table shows no outstanding results, except for some cells in line speed. Therefore, further analysis of this key metric is done. The high and low values of percentage primary incidents are mostly caused by a lack of observations in that category. Often there are less than ten observations on those categories. However, something else can be concluded besides these high and low values. There seems to be a (vague) trend that shows that the faster the line speed, the more primary incidents happen. Figure 15 shows the percentage of primary incidents (y-axis) per line speed category (x-axis), and as the blue dotted line shows, there is a higher percentage of primary incidents when the line speed is higher.

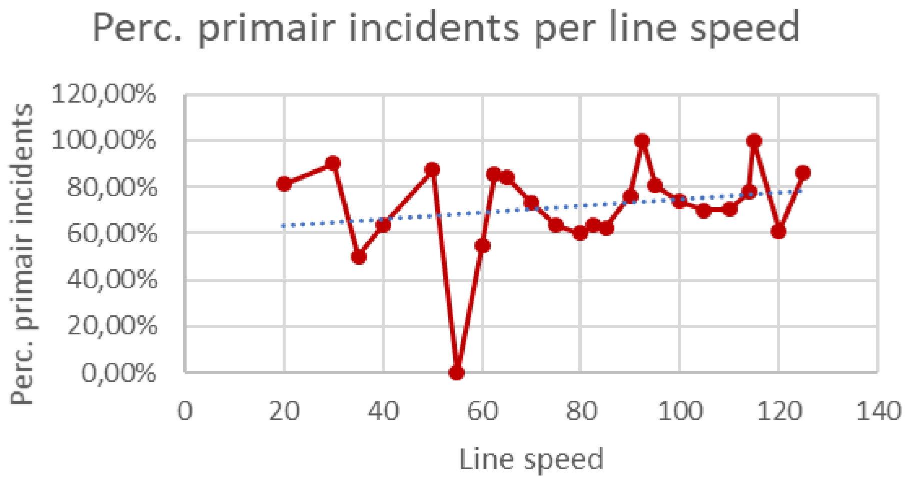


*Figure 15: Perc. Primary incidents per line speed*

## 6.2. Statistical analysis

In section 5.2, statistical tests are conducted on all key metrics. These tests show that there are several key metrics that show a correlation. Within this section of the thesis, these correlations will be elaborated on.

As can be seen in table 7, there are four key metrics that show a correlation. The key metrics *'Type of security'* - *'Light situation'* and *'Category of incident'* - *'Light situation'* do show some correlation; however, their Cramer's V value (respectively 0.048 and 0.039) appears to be non-existent when looking at the table 4. Therefore, under the circumstances of this thesis, the correlations between these key metrics are considered to be non-worrying for ProRail.

The key metrics *'Type of security'* and *'Category of incident'* also show a correlation; however, the Cramer's V value is 0.178. According to table 4, this value makes the correlation have moderate strength. Therefore, this key metric can be worrying and is something ProRail should keep an eye on. In section 6.1, Table 5, there has been taken a look at the percentage of primary incidents per security type. This did not show any tangible results; however, there does appear to be a correlation between the two key metric.

The last key metric that showed correlation are *'Distance to station'* and *'Line speed'*. The correlation coefficient is 0.331, which makes it a moderate relationship when looking at Table 5. However, since these are both characteristics of a railroad crossing, this correlation does not give much insight. This key metric only states that (for example) there happen to be many incidents far from a station with a higher line speed. This can easily be explained by the fact that there is a high chance that the line speed of railroad crossings far from stations is higher than close to stations due to safety measures.

# 7. Conclusion

In this chapter, an answer to the core problem is provided. Therefore, the following research question is answered within this section: *'How to evaluate the dashboard, statistical analyses and recommendations?'*. Therefore, first, a conclusion with additional recommendations is given. Besides that, this chapter discusses the limitations of this research as well as interesting further research.

## 7.1. Conclusions and recommendations

This bachelor thesis aims to solve the core problem 'No tool and/or report that shows which factors influence railroad crossing incidents.', which has been identified in section 1.3. Solving this core problem will eventually lead to fewer incidents at and around railroad crossings. To find these factors, the research included gathering and cleansing incident data, conducting semi-structured interviews, performing a literature review and a statistical analysis, and creating a visualisation tool.

This research consists of many methods since a lot of different factors can cause incidents. This thesis consisted of several phases to find these factors while dealing with the limitation discussed in the next section. First, the incident data gathered from ProVat and the railroad crossings registered is cleansed in Microsoft Excel/VBA, after which key metrics and metrics are developed. These key metrics and metrics are visualised in an interactive dashboard. Besides that, a statistical analysis of the key metrics to find a correlation has been conducted.

Due to several limitations, it cannot be stated with full certainty what factors influence the safety on and around railroad crossings. However, the results of this thesis (discussed in section 6) show that when it is dark, percentage-wise, more incidents happen than when it is light. Besides that, it has been shown that there are many incidents at railroad crossings with Andreas crosses when comparing it to the number of railroad crossings with such a security type. Next to this analysis on the numbers of railroad crossing incidents, an analysis on the effect of key metrics on primary incidents has been done. This has shown that when the line speed is higher, percentage-wise, more primary incidents happen than secondary- or almost incidents. Lastly, a statistical analysis has shown a moderate correlation between the type of security and the category of incidents. This could not be confirmed in the analysis of the primary incidents. All these factors can influence the safety of and around the railroad crossings. Therefore, advice to ProRail is given to track railroad crossings with these characteristics.

## 7.2. Limitations

As stated before, the research had several limitations: time restriction, data availability, the data quality, the accuracy of the respondent and its possible emotional strain with ProRail.

Since the research needs to be performed in ten weeks, in-depth investigation of each source and every subject is impossible. Therefore, some implementations and key metrics may be somewhat limited.

Besides this, there are also limitations on the quantitative part of this research. The data availability and, especially, the poor quality of data will cause limitations. Not all data is easy to access and some data is simply not stored (e.g. train numbers). Therefore, not all intended research could be conducted. Besides that, the quality of the data sometimes was somewhat poor. Information is sometimes stored in multiple databases and is incomplete. This can cause serious misrepresentations of data.

Also, there are several limitations concerning the qualitative part of this research. First of all, the accuracy of the respondent can be a limitation. Since there will be reliance on the knowledge of the respondent, everything he states can be used. However, if he remembers something incorrectly or has a possible emotional strain with ProRail, false information can be given. To prevent this from happening as much as possible, multiple team members will be interviewed. Lastly, this thesis was written as a proof of concept (POC) for ProRail. Therefore, not all possible resources, such as multiple employees and money, were available during this research.

## 7.3. Further research

Now, some parts of the research will be revisited to show what could be researched more and interesting topics for further research will be discussed.

The current research has been conducted with the limitations discussed in section 7.2. If the research could be conducted in more than ten weeks, other limitations could be minimal. Some data was not possible within ten weeks, which influenced the research. Therefore, to improve this thesis, it would be interesting to get access to this data and analyse it. Besides that, it would be possible to enlarge the statistical analysis and create a bigger and more detailed dashboard.

However, the most interesting topic for possible further research would be looking at the metrics and key metrics. Besides trying to gather more data and coming up with more key metrics, it would also be interesting to research the current ones further. As stated in section 3.1, KPIs should be SMART. Therefore, they should be achievable, and since ProRail does not set goals for this kind of KPI, this is impossible. This is why, within this thesis, there are only key metrics. For further research, it would be interesting to interview employees of multiple departments, especially heads of departments, to try to set these goals. This adjustment will make the key metrics trackable and SMART, which means that they can become KPIs.

Therefore, the databases used need to be linked to the dashboard. This would be a big and challenging task but would help keep track of the key metrics (and in the future KPIs). Besides this, making the dashboard in real-time would be interesting.

The last possible further research discussed is making a predictive tool that can statistically predict how many incidents will happen when crossing a railroad with certain characteristics. Therefore, multiple linear regression modelling can be used (Alqatawna et al., 2021)

# Bibliography


Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish journal of emergency medicine*, *18*(3), 91-93.

Alqatawna, A., Rivas Álvarez, A. M., & García-Moreno, S. S. C. (2021). Comparison of Multivariate Regression Models and Artificial Neural Networks for Prediction Highway Traffic Accidents in Spain: A Case Study. Transportation Research Procedia, 58, 277–284. https://doi.org/10.1016/j.trpro.2021.11.038

Bala, J. (2016). Contribution of SPSS in Social Sciences Research. *International Journal of Advanced Research in Computer Science*, *7*(6).

Barri, M. A. (2019). A simulation showing the role of central limit theorem in handling non normal distributions. *American Journal of Educational Research*, *7*(8), 591-598.

Bevan, N., Carter, J., Harker, S. (2015). ISO 9241-11 Revised: What Have We Learnt About Usability Since 1998?. In: Kurosu, M. (eds) Human-Computer Interaction: Design and Evaluation. HCI 2015. Lecture Notes in Computer Science(), vol 9169. Springer, Cham. https://doi.org/10.1007/978-3-319-20901-2_13

Booth, J. (2021). Valuing Culture and Heritage Capital: A framework towards informing decision making. *Cultural Trends*, *30*(5), 460–462. https://doi.org/10.1080/09548963.2021.1985936

Brown, J. D. (2001). Point-biserial correlation coefficients. *Statistics*, *5*(3), 12-6.

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data driven decisionmaking affect firm performance?. *Available at SSRN 1819486*.

Canals, L. (2017). *Instruments for Gathering Data*. Research-publishing. net. La Grange des Noyes, 25110 Voillans, France.

Carey, T., Yon, A., Beadles, C., & Wines, R. (2015). Identifying and prioritizing research gaps.

Chan, Y., & Walmsley, R. P. (1997). Learning and understanding the Kruskal-Wallis one way analysis-of-variance-by-ranks test for differences among three or more independent groups. *Physical therapy*, *77*(12), 1755-1761.

Chiesa, V., Frattini, F., Lazzarotti, V., & Manzini, R. (2009). Performance measurement in R&D: exploring the interplay between measurement objectives, dimensions of performance and contextual factors. *R&d Management*, *39*(5), 487-519.

Z Test - Formula, Definition, Examples, Types. (n.d.). Cuemath. Retrieved September 22, 2022, from https://www.cuemath.com/data/z-test/

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology*. Pearson education.

Daneshpazhooh, M., Mostofizadeh G, M., Behjati, J., Akhyani, M., & Robati, R. M. (2006). Anti-thyroid peroxidase antibody and vitiligo: a controlled study. *BMC dermatology*, *6*(1), 1-5.

Egger, A. E., & Carpi, A. (2008). Data: analysis and interpretation. *Visionlearning*, (1).

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Freelon, D. (2013). ReCal OIR: ordinal, interval, and ratio intercoder reliability as a web service. *International journal of internet science*, *8*(1).

Gajanova, L., Nadanyiova, M., & Moravcikova, D. (2019). The use of demo figureic and psycho figureic segmentation to creating marketing strategy of brand loyalty. *Scientific annals of economics and business*, *66*(1), 65-84.

Gelderman, M. (1998). The relation between user satisfaction, usage of information systems and performance. *Information & management*, *34*(1), 11-18.
Geocode (2021, March 17). *Geocode*. Infrasite. Retrieved September 22, 2022, from https://www.infrasite.nl/glossary/geocode/?gdpr=accept
Gesing, S., Atkinson, M., Filgueira, R., Taylor, I., Jones, A., Stankovski, V., ... & Kacsuk, P. (2014, November). Workflows in a dashboard: a new generation of usability. In *2014 9th Workshop on Workflows in Support of Large-Scale Science* (pp. 82-93). IEEE.
Ghosh, S., Roy, M. K., & Biswas, S. C. (2016). Determination of the best fit probability distribution for monthly rainfall data in Bangladesh. *American Journal of Mathematics and Statistics*, *6*(4), 170-174.
Gnanadesikan, R. (2011). *Methods for statistical data analysis of multivariate observations*. John Wiley & Sons.
Harris, M., & Tayler, B. (2019). Don't let metrics undermine your business. *Harvard Business Review*, *97*(5), 63-69.
Heerkens, H., & Winden, V. A. (2017). Solving Managerial Problems Systematically (Routledge-Noordhoff International Editions) (1st ed.). Routledge
Hoskin, T. (2012). Parametric and nonparametric: Demystifying the terms. In *Mayo Clinic* (Vol. 5, No. 1, pp. 1-5).
Hui, L. T., Hoe, L. S., & Ismail, H. (2020). Children Video Puzzle Interaction Framework Based on Multi-Touch Technology: An ISO 9241-11: 2018 Approach. *International Journal of Advanced Research in Education and Society*, *2*(2), 55-69.
Janes, A., Sillitti, A., & Succi, G. (2013). Effective dashboard design. Cutter IT Journal, 26(1), 17-24.
Kaganski, S., Majak, J., Karjust, K., & Toompalu, S. (2017). Implementation of key performance indicators selection model as part of the Enterprise Analysis Model. *Procedia Cirp*, *63*, 283-288.
Kaur, P., Stoltzfus, J., & Yellapu, V. (2018). Descriptive statistics. *International Journal of Academic Medicine*, *4*(1), 60.
Kawamoto, T., & Mathers, B. (2007). Key success factors for a performance dashboard. *Information Management*, *17*(7), 20.
Kerzner, H. (2017). *Project management metrics, KPIs, and dashboards: a guide to measuring and monitoring project performance*. John Wiley & Sons.
Kim, H. Y. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative dentistry & endodontics*, *42*(2), 152-155.
Kirk, A. (2012). *Data Visualization: a successful design process*. Packt publishing LTD.
Kothari, C.R. (2007). Quantitative techniques. New Delhi, UBS Publishers Ltd.
Krishnan, S. A. R. A. S., & Idris, N. O. R. A. I. N. I. (2015). An overview of students' learning problems in hypothesis testing. *Jurnal Pendidikan Malaysia*, *40*(2), 193-196.
Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., Apps, M. A., Argamon, S. E., ... & Zwaan, R. A. (2018). Justify your alpha. *Nature human behaviour*, *2*(3), 168-171.
McDonald, J. H. (2009). *Handbook of biological statistics* (Vol. 2, pp. 6-59). Baltimore, MD: sparky house publishing.
Microsoft. (2021, November 9). Visualization types in Power BI - Power BI. Microsoft Docs. Retrieved April 20, 2022, from https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-types-for-reports-and-q-and-a
Mourougan, S., & Sethuraman, K. (2017). Hypothesis development and testing. *IOSR Journal of Business and Management (IOSR-JBM)*, *9*(5), 34-40.

Nayak, B. K., & Hazra, A. (2011). How to choose the right statistical test?. *Indian journal of ophthalmology*, *59*(2), 85.

Nielsen, J. C., Lundén, R., Johansson, A., & Vernersson, T. (2003). Train-track interaction and mechanisms of irregular wear on wheel and rail surfaces. Vehicle System Dynamics, 40(1-3), 3-54

Nihan, S. T. (2020). Karl Pearsonâ€™ s chi-square tests. Educational Research and Reviews, 15(9), 575-580.

Nowacki, A. (2017). Chi-square and Fisher's exact tests. Cleveland Clinic journal of medicine, 84(2), e20.

Ovwigho, B. O. (2013). Empirical demonstration of techniques for computing the discrimination power of a dichotomous item response test. IOSR Journal of research and Method in Education, 3(2), 12-17.

Panagiotakos, D. B. (2008). The value of p-value in biomedical research. The open cardiovascular medicine journal, 2, 97.

Pappas, L., & Whitman, L. (2011, July). Riding the technology wave: Effective dashboard data visualization. In Symposium on Human Interface (pp. 249-258). Springer, Berlin, Heidelberg.

Peffers, K., Tuunanen, T, Rothenberger, M.A. & Chatterjee, S (2007). A Design Science Research Methodology for Information Systems Research, Journal of Management Information Systems, 24:3, 45-77, DOI: 10.2753/MIS0742-1222240302

Podgórski, D. (2015). Measuring operational performance of OSH management system–A demonstration of AHP-based selection of leading key performance indicators. Safety science, 73, 146-166.

ProRail. (2016, April 18). Veelgestelde vraag: Hoe wordt ProRail aangestuurd? Retrieved April 14, 2022, from https://www.prorail.nl/nieuws/veelgestelde-vraag-hoe-wordt-prorail-aangestuurd

ProRail. (2017, August 16). Wat doen Edo, Wilo en Ali-B bij het spoor. Retrieved August 17,2022, from https://www.prorail.nl/nieuws/wat-doen-edo-wilo-en-ali-b-bij-het spoor

ProRail. (2019, May). Gedragscode. https://www.prorail.nl/siteassets/homepage/over ons/documenten/gedragscode-prorail.pdf

ProRail. (2021a, April 13). Aantal gevaarlijke situaties moet omlaag. Retrieved April 7, 2022, from https://www.prorail.nl/nieuws/aantal-ongevallen-en-gevaarlijke-situaties-op het-spoor-moet-omlaag

ProRail. (2021b, August). ProRail halfjaarrapportage 2021 Deel A. https://open pilot.overheid.nl/repository/ronl-3c8292f9-4c13-4801-9d0f 48fb279c763a/1/pdf/halfjaarrapportage-prorail-deel-a.pdf

ProRail. (n.d. -a). Over ons. Retrieved April 7, 2022, from https://www.prorail.nl/over ons#:%7E:text=Dit%20is%20ProRail%20Over%20ons&text=We%20verdelen%20a s%20onafhankelijke%20partij,%C3%A9n%20met%20oog%20voor%20toekomst.

ProRail. (n.d. -b). Veiligheid. Retrieved August 9, 2022, from https://www.prorail.nl/veiligheid#:%7E:text=Op%20en%20rond%20het%20spoor% 0Veiligheid&text=Door%20overwegen%20veiliger%20te%20maken,op%20en%20r nd%20het%20spoor.

Punzet, L., Pignata, S., & Rose, J. (2018). Error types and potential mitigation strategies in Signal Passed at Danger (SPAD) events in an Australian rail organisation. Safety science, 110, 89-99.
Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. IEEE Data Eng.Bull., 23(4), 3-13.
Sadiku, M.N.O., Musa, S.M. (2021). Business Intelligence. In: A Primer on Multiple Intelligences. Springer, Cham. https://doi.org/10.1007/978-3-030-77584-1_14

Saraiya, P., North, C., & Duca, K. (2004, October). An evaluation of microarray visualization tools for biological insight. In IEEE Symposium on Information Visualization (pp. 1-8). IEEE.
Shahin, A., & Mahbod, M. A. (2007). Prioritization of key performance indicators: An integration of analytical hierarchy process and goal setting. International Journal of Productivity and Performance Management.
Schiff, C. (2008). Three things you should know about dashboards. Information Management, 18(6), 29.
Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. Anesthesia & Analgesia, 126(5), 1763-1768.
Schumann, T. (2016). Increase of capacity on the shinkansen high-speed line using virtual coupling.
Singh, N. U., Roy, A., Tripathi, A. K., Test, C. S., Test, B., Test, O. S. K. S., ... & Kendall’s, W. (2013). Non parametric tests: hands on SPSS. ICAR Research Complex for NEH Region, Umiam, Meghalaya.
Suparji, S., Nugroho, H. S. W., & Martiningsih, W. (2021). Tips for Distinguishing Nominal and Ordinal Scale Data. Aloha International Journal of Multidisciplinary Advancement (AIJMU), 1(6), 133-135.
van der Schans, E. (2001, October). Hoeveel mag een risico kosten. NVVK info. https://docplayer.nl/176514457-Hoeveel-mag-een-risico-kosten.html
Varma, S. (2006). Preliminary item statistics using point-biserial correlation and p-values. Educational Data Systems Inc. 16(07).
Wegman, F., & Aarts, L. (2006). Advancing sustainable safety. SWOV, Leidschendam.
Williams, H. J., & Blum, I. (2018). Defining second generation open source intelligence (OSINT) for the defense enterprise. Rand Corporation.
Zaw, T. M., Thant, M., & Bezzateev, S. V. (2019, June). Database security with AES encryption, elliptic curve encryption and signature. In 2019 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF) (pp. 1-6). IEEE.
Zhou, H., & He, Y. (2018). Comparative Study of OKR and KPI. In 2018 International Conference On E-Commerce And Contemporary EconomicDevelopment (Eced 2018), DEStech Transactions on Economics Business and Management.
Zou, K. H., Tuncali, K., & Silverman, S. G. (2003). Correlation and simple linear regression. Radiology, 227(3), 617-628.

# Appendices

## Appendix A – Code

```
Sub FindLocation()
Dim i As Integer
Dim j As Integer
Dim sal As Range
Set sal = Range("H2:H7302")
For j = 2 To 7302
For i = 2 To 1850
If Cells(j, 2).Value = Cells(i, 3).Value Then
Cells(i, 8).Value = Abs(Cells(j, 1).Value - Cells(i, 4).Value)
Cells(7500, 8).Value = Application.WorksheetFunction.Min(sal)
Else
End If
Next i
Cells(j, 9).Value = Application.Index(Range("G2:G1850"), Application.Match(Cells(7500, 8).Value, Range("H2:H7302"), 0), 0)
Range("H:H").ClearContents
Next j
End Sub
```

*Figure 16: Microsoft VBA code written to find location of incidents*

This is the code to find the location as explained in section 2.2. Therefore, Microsoft VBA has been used. Within this coding program, it is needed to dim variables before they can be used. Within this coding two loops are used, a range is set and several Worksheet Functions are used.

## Appendix B – Types of security

There are different types of secured and unsecured railroad crossings. The following list provides every type. This list is copied from: [https://www.prorail.nl/nieuws/wat-doen-edo-wilo-en-ali-b-bij-het-spoor](https://www.prorail.nl/nieuws/wat-doen-edo-wilo-en-ali-b-bij-het-spoor)

Secured level crossings:

1. EBO > Electrically operated level crossing barriers, remotely or locally operated
2. ADOB > Automatic double level crossing barriers, which close both halves of the road
3. AHOB > Automatic half level crossing barriers
4. Footpath AOB > Automatic crossing barriers at stations, these close off the footpath that provides access to a platform. Without red lights, but usually with a fence under the tree.
5. AKI > Automatic flashing light installation. Two automatic flashing lights and an alarm bell. In the AKI-AHOB conversion program completed in 2007, most were converted to (Mini-)AHOB.
6. HAHOB > Semi-automatic half level crossing barriers: closing is done with a push button, opening is automatic.
7. HBHOB > Manually operated half level crossing barriers. This is no longer applied.
8. HAKI > Semi-automatic flashing light
9. installation HBKI > Manual flashing light
10. installation WIDO > Warning installation for service
11. crossings WILO (with country gates) > Warning installation for national crossings. A single automatic flashing light without an alarm bell.
12. HAVIO > Semi-automatic traffic light installation for level crossings (became HALI)
13. AVIO > Automatic crossing traffic light installation (became ALI)
14. ALI > Automatic light installation (was AVIO)
15. HALI > Semi-automatic light installation (became HAVIO)
16. ALIB > Automatic light installation with Boom
17. HALIB > Half automatic light installation with Boom
18. Mini Ahob > Level crossing with automatic half level crossing barriers for narrow level crossings
19. ARW 5/2 (work crossings only)
20. VKL > Traffic lights at a level crossing
21. MBO > DMechanically operated level crossing barriers. These are barriers that are closed and opened by hand instead of a motor, but this type is no longer used.

Unsecured level crossings:

1. Zig-zag/folding gates
2. Andrew's crosses with zigzag gates
3. Gates
4. Gates with Andrew 's crosses Folding gates MeBeLa
5. Andrew 's crosses (as main protection)
6. Locomotive sign with train instruction
7. Andrew's crosses with train instruction
8. Red flag/lamp (without Andrew's crosses)
9. Train whistles (without Andrew's crosses)
10. Level crossings without indication /road traffic warning

## Appendix C – Categories and subcategories of incidents

Railroad crossings incidents are categorised in primary, secondary and almost incidents. These types have subcategories which are shown below.

Primary incident:
Primary incident – railroad crossing incident
Primary incident – railroad crossing incident – Collusion train – person
Primary incident – railroad crossing incident – Collusion train – person – incident
Primary incident – railroad crossing incident – Collusion train – person – incident with confinement
Primary incident – railroad crossing incident – Collusion train – person – suicide
Primary incident – railroad crossing incident – clash train – object
Primary incident – railroad crossing incident – clash train – traffic
Primary incident – railroad crossing incident – clash train – traffic – incident
Primary incident – railroad crossing incident – clash train – traffic – incident caused by barrier
Primary incident – railroad crossing incident – clash train – traffic – incident with confinement
Primary incident – railroad crossing incident – clash train – traffic - suicide
Primary incident – railroad crossing incident – traffic incident
Primary incident – railroad crossing incident – traffic incident – clash catenary
Primary incident – railroad crossing incident – traffic incident – clash installation
Primary incident – railroad crossing incident – traffic incident - Clash with moving or closed barrier
Primary incident – railroad crossing incident – traffic incident – vehicle hit flooring

Secondary-incident:
Secondary incident – railroad crossing incident – incident of traffic – clash with moving or closed barrier
Secondary incident – railroad crossing incident – incident of traffic – vehicle hit flooring
Secondary incident – railroad crossing incident – incident of traffic – clash installation
Secondary-incident – railroad crossing incident – clash train – object
Secondary-incident – railroad crossing incident – clash train – traffic - incident
Secondary-incident – railroad crossing incident – incident of traffic
Secondary-incident – railroad crossing – almost-suicide

Almost-incident:
Almost-incident – railroad crossing
Almost-incident – railroad crossing – almost-collusion train – person
Almost-incident – railroad crossing – almost-clash train – traffic
Almost-incident – railroad crossing – almost-suicide