

UNIVERSITY OF TWENTE

Graduation Project

**DATA VISUALISATION OF ENVIRONMENTAL RISKS IN REAL ESTATE
PROPERTIES**

Department of EEMCS

Bachelor Degree in Creative Technology

by

Alexis Philippou

2022

Abstract

Environmental risks are not taken into account when it comes to the real estate market, and neither do insurance companies. In return, this may lead to an inaccurate valuation of real estate properties or the lack of insurance coverage for damage on properties caused by natural disasters. Using GIS information, it is possible to convey environmental data through visualisations on maps and descriptive statistics, via an already established API with 23 services. Due to a lack of experience and knowledge, some wrong decisions were taken on what software to be utilised. To compromise, a combination of map objects and graphs was used to create the data visualisations based on a specified region of Cyprus selected by the user using a customised tool. Thus, the illustration of environmental insights and geodata was not accomplished to the desirable level of comprehension that was initially planned. As a result, the chosen techniques of blending map objects and graphical representations was not as successful as intended. What is recommended is the use of packages and plugins that allow the implementation of more intuitive visualisations on top of the map itself, in order to complement the physical features of the area's landscape.

Contents

Abstract	i
Contents	ii
1 Introduction	1
1.1 Context	1
1.2 Motivation	2
1.3 Research Questions	3
2 Background Research	4
2.1 Literature Review	4
2.2 State of the Art	5
3 Methodology	8
3.1 Design	8
3.2 Data Management	12
4 Realisation	15
4.1 Development of Web Application	15
4.1.1 Architecture	15
4.1.2 Adapted User Interface	16
4.2 Visualisations	17
4.2.1 Approach	17
4.2.2 Maps and Graphs	18
5 Evaluation	35
5.1 Testing	35

5.2 Results	36
6 Discussion	39
6.1 Limitations	39
6.2 Future Work	40
7 Conclusion	42

Chapter 1

Introduction

1.1 Context

Up until now, there was not enough attention paid on the potential of environmental insights being used in real estate decision making. In the information age, data is a very valuable asset when processed and visualised properly, especially in the real estate industry as the environmental impacts on properties are non-visible. With the right data, computer algorithms and programs can handle the heavy nature of environmental recordings. Data visualisation can be simply defined as a means to convert quantified data into a visual representation. With the help of data's underlying patterns and relations, exploration and analysis are possible in an easy and affordable way, using computer-generated graphics to gain information and increase understanding [1]. On the other hand, by arranging and processing empirical and heuristic information, Artificial Intelligence aims to imitate the results of human thinking and human capacities such as learning, reasoning, adaptation, and self-correction [2]. Because artificial intelligence is such a new field of technology, there aren't many well-established applications, particularly in real estate.

In the case of existing AI applications for environmental risk detection, its most common purpose is to forecast meteorological incidents based on predictions and modelling [3]. Furthermore, recent implementations of AI in the environmental industry have progressed

significantly since the appropriate hardware has been developed to tackle the intense nature of climate data. As artificial intelligence and machine learning algorithms advance, more initiatives for climate study may be explored [4]. Consequently, it has been proven beneficial utilising AI to detect and prevent environmental risks, so there should be a way to apply the same, or similar, techniques in the real estate industry.

The aim of the project is to create a graphical user interface (GUI) and visualisations on top of an existing application programming interface (API) that contains a list of services related to real estate properties such as historical risk of flooding, distance to roads and amenities, ground elevation and many more. Not only that, but it is expected to construct an integrated environmental risk model for every property in the region of interest based, more specifically on the island of Cyprus, on the environmental risks connected with each property. Finally, descriptive statistical data will be produced based on the studied material, such as the most disaster-prone regions per city.

1.2 Motivation

It's already tough to deal effectively with environmental hazards in real estate projects. The number of natural disasters has gradually increased during the past few years, one of the reasons being climate change. It is anticipated to raise the frequency and intensity of said disasters, inducing a heavier risk on real estate properties. Some of these examples include physical hazards, liability risks, and transition risks. As a result, climate change may have a substantial influence on what insurers are able to cover for their customers, since insurance companies can no longer afford to ignore or postpone dealing with the urgent effects of climate change.

One of the greater challenges is quantifying these risks based on the detected climate data in a given property, and it emphasises the need to go beyond conventional methods of analysis. Understanding and processing this information in a proper manner allows potential property investors make sound purchases. There are many aspects that need

to be taken into consideration such as location, architecture, and price, however, now is the time to incorporate climate change into the formula. With the recent development of artificial intelligence, it is our opportunity to close the gap between the unprecedented effects of environmental risks on a real estate property and the buyer. In order to enhance risk assessments, advanced analytics could be used to examine historical weather records, insured property data, and make estimates about potential climate conditions. Big data and predictive analytics can be deemed useful to model changes in the climate and dramatically expand the factors involved in risk assessments.

1.3 Research Questions

Two key questions must be researched and answered in order for the project to be completed. Since both quantitative and qualitative data will be evaluated at, hybrid research will be used to address the questions. The information will be gathered through testing, surveys, and previous research.

- How can we analyse and visualise geospatial information related to real estate properties?
- How can AI-driven data visualisations give insights for real estate?

Chapter 2

Background Research

2.1 Literature Review

Artificial Intelligence (AI) has seen its most significant advancements in the last few years, and the corporate world is not going to miss an opportunity to use this technology. Excluding the financial aspect of businesses, there are ecological issues that need to be tackled as well. Companies require particular knowledge about the consequences of climate change in order to prepare for and respond to climate-related issues. Because of the fast-changing nature of the climate, simply looking at past weather patterns is not a good way of predicting future risks [5]. There are methods to combat some of the ramifications such as using artificial intelligence for urban planning and land cover mapping, taking into account environmental factors.

Warmer average temperatures, increasing sea levels, and more frequent weather extremes are already affecting our risk landscape as a result of climate change. The average of insured losses from catastrophic weather events have more than doubled in the previous three decades. Natural disaster losses have steadily increased as a result of economic and population growth in vulnerable areas, which is intensified by a changing climate. Because most damages caused by natural disasters are not covered by insurance, many real estate properties are left vulnerable [6].

Concerning urban planning and land cover mapping, artificial neural network (ANN) has been applied to handle both tasks, processing GIS data and satellite imagery, respectively. The paper [7] proposes a way to use an environment and energy geographical information system (E-GIS) database (DB) and an ANN to plan a model on site locations for environmentally friendly urban energy supply plants. This information is used to determine the best location for an urban energy supply plant during the development stage of a green city. Modelling urban environments, predicting urban microclimate and energy demand, and assisting energy supply planning are all done with E-GIS database variables with a mesh unit. The use of artificial intelligence (AI) computing for deducing non-linear algorithms between environmental and energy planning variables, as well as the possibility for site location, follows. Finally, by providing input on the E-GIS database, visual maps, and quantitative values, and locating energy supply plant locations, the achieved result is linked to the mesh unit, which is created from the site location potential algorithms [7]. Furthermore, the goal of this research [8], is to detect changes in land cover and human population through time and space using satellite photos. Future land cover was estimated using an ANN technique. Forest, aquatic, agricultural, and built-up regions were categorised into four land cover classes with supervised and image classification using a maximum likelihood algorithm technique. Distances from roads (highways, village roads, and forest roads) and distances from streams were used as spatial variables in the ANN model, and they provide useful information regarding the effects of LULC. After five iterations, the validation was calculated by comparing the real map acquired from the current image's classification with the simulation map created using data taught by the ANN in the system [8].

2.2 State of the Art

The scope of AI applications in real estate is quite wide, ranging from personal to corporate methodologies. A few start-ups have developed predictive analytics solutions to help businesses in the real estate industry, make better estimates and preparations for extreme weather disasters. This is achieved by integrating machine learning with more traditional

methods of weather modelling, and then training their models with a combination of public, private, and customer-specific data [5].

From a personal point of view, PropertyQuants [9] enables investors to make evidence-based, data-driven, and systematic real estate decisions by applying quantitative finance and data science techniques to the global real estate market. Taking on data on a worldwide scale, it accumulates a large variety of different real estate related data from around the globe, all of which have been cleaned, standardized, and normalized for use in local settings such as neighbourhood demographics and migration patterns. There are cases where it is necessary to transform data to proprietary numerical scores, for example rating school quality and commute time. This method incorporates a mixture of cloud-based databases, geographic information systems (GIS) and artificial intelligence (AI), or to be more precise, machine learning. Their goal is to provide statistically sound insights and quantify portfolio performance, risk, and return trade-offs [9]. On the other hand, by making a digital twin of both artificial and natural environments, One Concern [10] helps visualise the impacts of severe environmental circumstances, present and future, through predictive and forward-looking models. In detail, it produces resilience measurements that can be applied across time and space, allowing investors to adjust their property offers, from selection to sale, in response to a continually changing landscape. This way, both internal and external property risks can be explored and analysed for a more solid outcome [10].

Climate insurance assists businesses in fully understanding and anticipating weather hazards, as well as protecting themselves financially. Legacy insurance companies are trying to analyse and price the different financial risks posed by climate change in real estate, which can be either too costly or impossible for many categories of climate risk [5][1]. Unexpected losses result from valuation gaps and misspent mitigation funds due to inadequate and coherent modelling that appropriately combines network relationships. Property valuation gaps refer to properties that are currently being priced erroneously because to a lack of knowledge about their genuine resilience [11]. Misspent mitigation funds occur when companies set aside expenditures for mitigation that are divided evenly among

dangers. In actuality, some dangers may be substantially more vulnerable than others, leaving businesses vulnerable to major costs [11]. When a specified event, such as a fire or a storm, occurs, parametric insurance automatically pays out a predetermined sum, ensuring payment when certain parameters are met. This eliminates the need for insurance agents to analyse and verify the losses of individual policyholders [5]. Strong data and analytics capabilities are necessary to execute successfully and make more automated, data-driven, and transparent judgments for faster and more assured pay outs [5].

Two examples are Descartes Underwriting and Kettle, both of which, focus on parametric insurance using their own unique approaches. Descartes Underwriting [12] underwrites and monitors its parametric policies with machine learning, consuming data in real-time from a number of sources such as satellite imaging, sensors, Internet-of-Things devices, radar, and sonar. They claim that their adaptable parametric solution is designed around the client's exposure and risk transfer needs. Using the client's location, loss record history, and permission to install sensors on the desired property, they design custom covers and monitor the parameter's evolution based on the predefined exposure framework [12]. On the contrary, Kettle [13] revolves around AI-powered risk modelling to boost profits using deep learning modelled pricing and proprietary neural network architecture to generate hyper accurate, hyper local wildfire risk forecasts. To forecast wildfire propensity, the deep convolutional neural network is trained on parameters that influence wildfires and builds ensembles. For wildfire prediction, over 3 petabytes of data have been explored and collected from several sources, including the moderate resolution imaging spectroradiometer (MODIS) and the light detection and ranging method (LIDAR) [13].

The use of AI for environmental risk detection in real estate has been explored more in the corporate domain than in the academic. There are companies that offer AI services for investment and sale purposes in real estate taking into account many different factors, and for parametric insurance that benefit both client and service parties. They all employ an AI system with different objectives and measurements. However, they all use this technology to detect and adapt to environmental and climate changes in real estate.

Chapter 3

Methodology

3.1 Design

The design revolves around the idea of an interactive map of Cyprus where a list of services will be able to run over a selected region. This was inspired by other satellite services such as Google Maps, or any delivery system where the delivery is tracked in real-time. The idea is that the user will be able accumulate services. So the user will have the option to add multiple services to run for a selected region, allowing the user to explore and gain more insights about a property.

In Figure 3.1 we can see how the main screen could look like. This as far as it would zoom because all the regions at hand are within the borders of Cyprus.



Figure 3.1: Main Page

In Figure 3.2 the drop-down menu is shown with the choice of service focused on Vegetation. Scrolling through the list, all the services are available.



Figure 3.2: Drop-down Menu

In Figure 3.3, after the clicking the desired service, the description along with both the input and output parameters are illustrated.

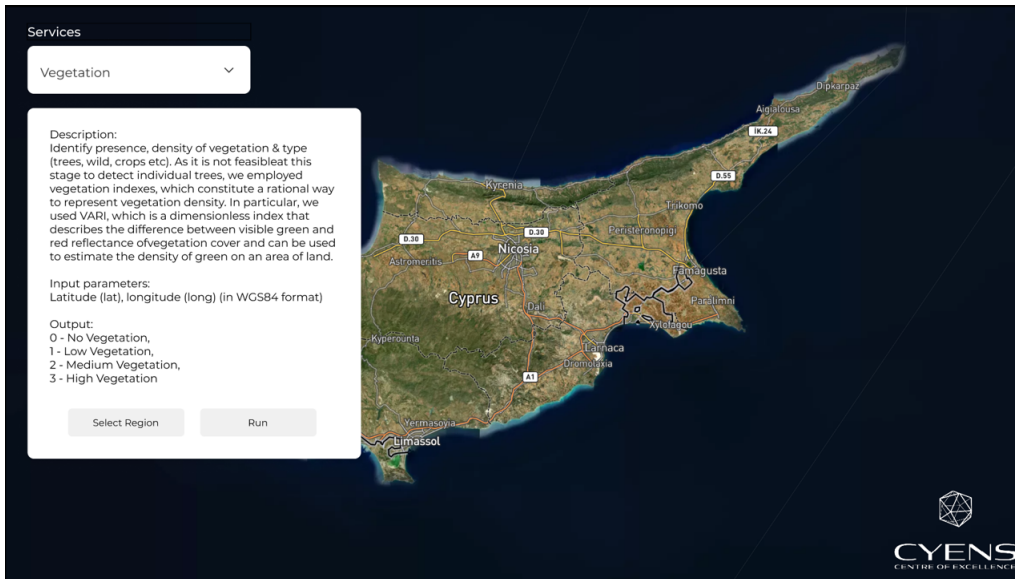


Figure 3.3: Service Info

Figure 3.4 shows a zoomed in area to make a more accurate selection. This is done using the middle click to scroll in and out of the map.



Figure 3.4: Zoomed View

In Figure 3.5 the “Select Region” button is clicked to select a specific region to run the service in. Using a simple selection tool, the user can draw a square around a property.

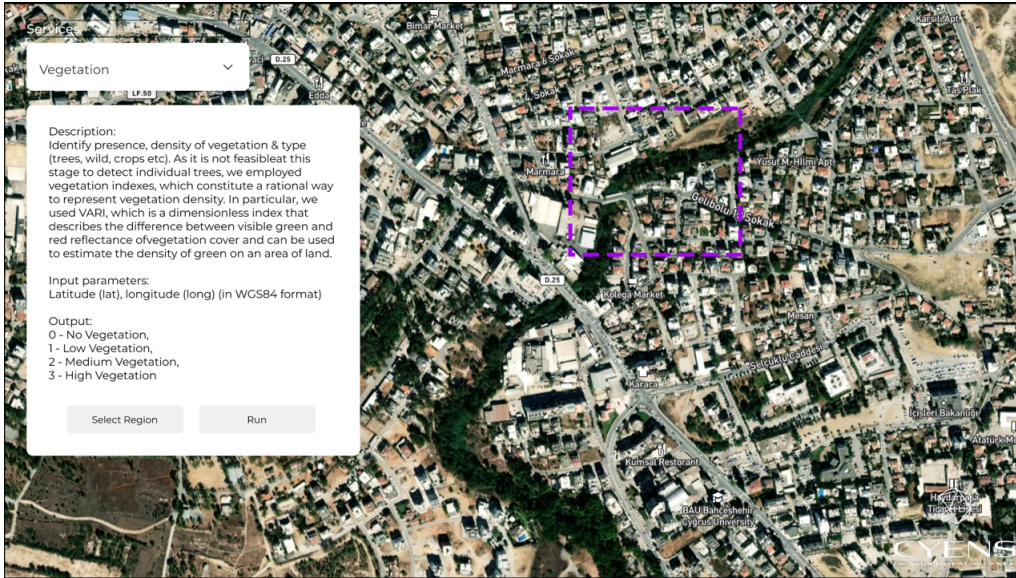


Figure 3.5: Region Selection

Finally, Figure 3.6 is the final result after the “Run” button was clicked for the selected region. The system highlights, with a vibrant colour, the vegetation detected in the specified area.



Figure 3.6: Highlighted Area

3.2 Data Management

There are 23 real estate related services in total, each with a different purpose, meaning they gather and analyse different data for a specific result. Taking into account these factors, there are a few elements that some services share. Hence, a way to categorise them in a table is given below:

Property related		Area related		
Within property	Around property	Status	Risks	Measurements
1. Detection of swimming pools inside the property	16. Proximity to roads	2. Land use change	6. Risk of landslides	11. Slope and aspect
3. Building quality	17. Proximity to the sea	8. Detection of burnt areas	7. Wildfires risk	12. Geology
4. Detection of vegetation around the property	18. Proximity to blue-flag beaches	15. Land use	9. Flooding risk	13. Precipitation
5. Subsidence risk	19. Proximity to amenities	21. Natura 2000		14. Elevation
10. Buildings' damage risk from earthquakes	22. Proximity to electricity network			23. Wind
20. Building area				

Figure 3.7: Categorisation of Services

When looking at the services as whole, there is information that could lead to a better understanding of the terrain of Cyprus. Many questions can be asked regarding the environmental effects on real estate properties, and the answers can be found by combining some these services together to produce a compound output. The following table provides multiple scenarios on environmental effects and the services needed to examine them:

Scenario	Services		
Are swimming pools more common further away from the sea?	1. Detection of swimming pools inside the property	17. Proximity to the sea	
Probability of a house having a pool with respect to its size	1. Detection of swimming pools inside the property	20. Building area	
How do swimming pools affect the level of humidity in an area?	1. Detection of swimming pools inside the property	(+) humidity information from weather stations	
Burnt areas where there could be illegal development	2. Land use change	8. Detection of burnt areas	
Buildings that could host illegal activities	2. Land use change	15. Land use	16. Proximity to Road
	20. Building area	22. Proximity to electricity network	
Are low quality buildings more prone to earthquake damage?	3. Building quality	10. Buildings' damage risk from earthquakes	
To what extent do salty winds degrade building materials?	3. Building quality	17. Proximity to the sea	
Can vegetation reduce landslides?	4. Detection of vegetation around the property	6. Risk of landslides	12. Geology
Whether wildfire correlates with precipitation?	4. Detection of vegetation around the property	7. Wildfires risk	13. Precipitation
Areas that, despite high precipitation, do not contain vegetation	4. Detection of vegetation around the property	12. Geology	13. Precipitation
	14. Elevation	23. Wind	
Does soil depend on vegetation and/or elevation?	4. Detection of vegetation around the property	12. Geology	14. Elevation
Types of land use with the most vegetation?	4. Detection of vegetation around the property	12. Geology	15. Land use
Is vegetation correlated to precipitation?	4. Detection of vegetation around the property	13. Precipitation	
How does soil influence subsidence risk?	5. Subsidence risk	12. Geology	15. Land use
Do steeper slopes cause more landslides?	6. Risk of landslides	11. Slope and aspect	
Does wind speed and direction determine wildfire spread?	7. Wildfires risk	8. Detection of burnt areas	23. Wind
Spots to remove unnecessary electricity poles, to reduce wildfires	7. Wildfires risk	22. Proximity to electricity network	
Do coastal areas have a higher flood risk?	9. Flooding risk	13. Precipitation	17. Proximity to the sea
Visualisation of micro climate and measurements for Natura 2000	11. Slope and aspect	12. Geology	13. Precipitation
	14. Elevation	21. Natura 2000	23. Wind
Do steeper slopes constitute more flooding risk?	11. Slope and aspect	13. Precipitation	14. Elevation
	17. Proximity to the sea		
How are elevation and wind correlated with precipitation?	13. Precipitation	14. Elevation	23. Wind
Areas which are far away from hospitals	15. Land use	19. Proximity to amenities	
Visualisation of Natura 2000 with nearest city/village	15. Land use (using specific category urban/rural)	21. Natura 2000	
Appropriate areas for installing wind generators	15. Land use	22. Proximity to electricity network	23. Wind
Types, length of road in side Natura 2000 areas	16. Proximity to Road	21. Natura 2000	
High-risk areas for an illegal development/risk of fire	21. Natura 2000	22. Proximity to electricity network	

Figure 3.8: Combination of Services

There are particular scenarios that of more particular importance considering the context of the project, meaning real estate. Out of a total of twenty six possible cases, ten of them stand out, in terms of the information that would be useful to know for both investments and insurance. It is useful to know the extent of the environmental risks before buying a house in a certain location. For insurance companies, this data is valuable information to assess property damage as accurately as possible.

- How do swimming pools affect the level of humidity in an area?
- Buildings that could host illegal activities
- Are low quality buildings more prone to earthquake damage?
- To what extent do salty winds degrade building materials?
- Is vegetation correlated to precipitation?
- Does wind speed and direction determine wildfire spread?
- How are elevation and wind correlated with precipitation?

- Areas which are far away from hospitals
- Visualisation of Natura 2000 with nearest city/village
- Appropriate areas for installing wind generators

Chapter 4

Realisation

4.1 Development of Web Application

4.1.1 Architecture

The Python-based web development framework, Flask, was used to create the web application because it helps projects integrate elements used in web design through HTML, CSS, and JavaScript code. Since Python was being used, the available mapping service, called Folium, was used to construct the maps and the map objects, later used in the visualisations. Each service is allocated a specific map and an HTML page as an extension to the main index page. In addition, one other plugin was used to plot the graphs called Plotly, which allows data sets to be graphed under specific format for each type of diagram.

The data source is a static JSON file that contains all of the data that is used to represent each service from the API. Since we didn't have direct access to the API, a dataset was created using random data following the JSON responses given by the API, indirectly emulating its use. Each recorded coordinate on the Cypriot map has a separate JSON object within the structure of the data. Since the presented data will cover only a specific region, the data is removed from the general JSON file and written on a new one by extracting only the JSON objects in which the points inside the provided region are stored. This will also allow the integration of the actual API, if the web application is ever needed

outside the scope of the current project.

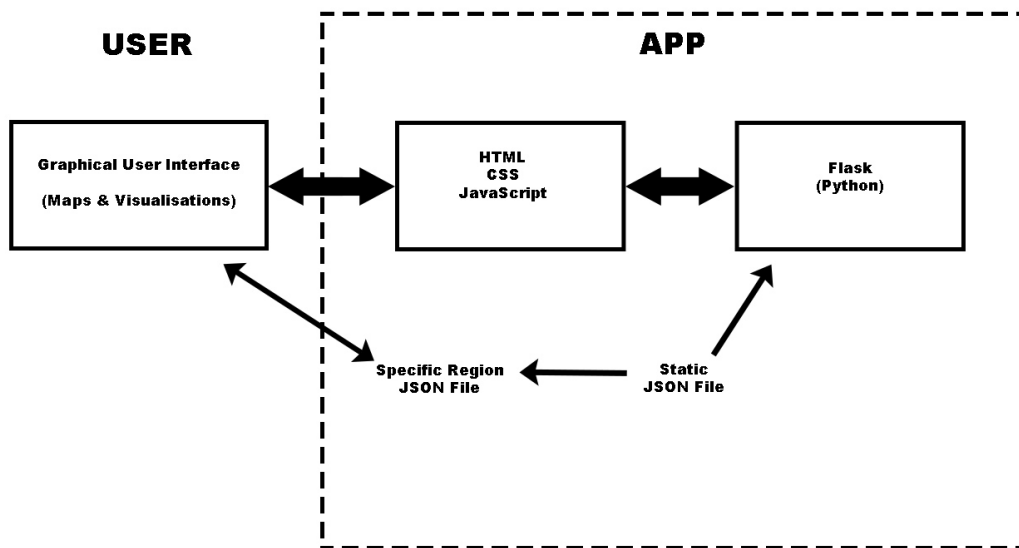


Figure 4.1: Application Architecture Diagram

4.1.2 Adapted User Interface

There are obvious differences between the user interface's design from the project's ideation phase and the version that was actually executed. This change resulted from the need to adapt to the restrictions presented by the chosen software used to develop the application. Just like any project, it is an iterative process of development and optimisation as new tools are used and new problems arise. The software that was used in the end is called Folium, which is designed to incorporate interactive maps in Python projects. Compared to the corresponding JavaScript library, called Leaflet, Folium lacks in the areas of features and customisability, however it still allows data to be rendered on web pages and plotted on interactive maps using markers and circles.

Even though some aspects of the functionality behind the user interaction, there are some other elements that remained the same such as the position of the dropdown menu on the top left corner and the use of a map as the main feature of the interface. One of the

aforementioned changes revolves around the way the user would select a specific region. In the implemented version, the use of a plugin provided by the Folium library called Measure Control. Eventually, allowing the user to select a specific region from the map by drawing up a polygon which encapsulates the filtered data. The way this was done involved one further step, using a customised JavaScript script to store the filtered data into a JSON file as a list. By default, the polygon's perimeter and area are also provided.

4.2 Visualisations

4.2.1 Approach

The data visualisations are broken down into two components: the map objects and the graph plots. The map objects were made using the Folium library, which usually displays all the information about any geographic place in Cyprus all at once. This may be led to an overwhelming experience since too much information is shown all at once without even interacting with the map. Hence, a design choice was made to hide the data rather than displaying it immediately on the map. With the alternate approach, a user can use map control, mentioned above, to view details about a particular area in which they are interested in.

The second component of the visualisations are the graphs depicting various characteristics of the specified data that the user has selected using the measure control. The graphs were created utilising the Plotly library which was chosen to display the data in a graphical manner to accompany the, rather lacking, map objects. To be more precise, this library can dynamically filter and model the data that is being visualised on the map centred on the region selection.

There are few types of graphs that are prominent throughout the services, one of which being heatmaps. Thematic maps, another name for heatmaps, are mostly used in risk management to evaluate flood and landslide risks [14], as well as wildfire risk in a given area [15]. Since there are multiple kinds of risks, heatmaps were used to visualise them in graph format. Another type of graph that was used were histograms that decrease the

complexity of geospatial data by subdividing numeric values in categorical bins, which are the bars used in histograms [16].

4.2.2 Maps and Graphs

Detection of swimming pools inside the property:

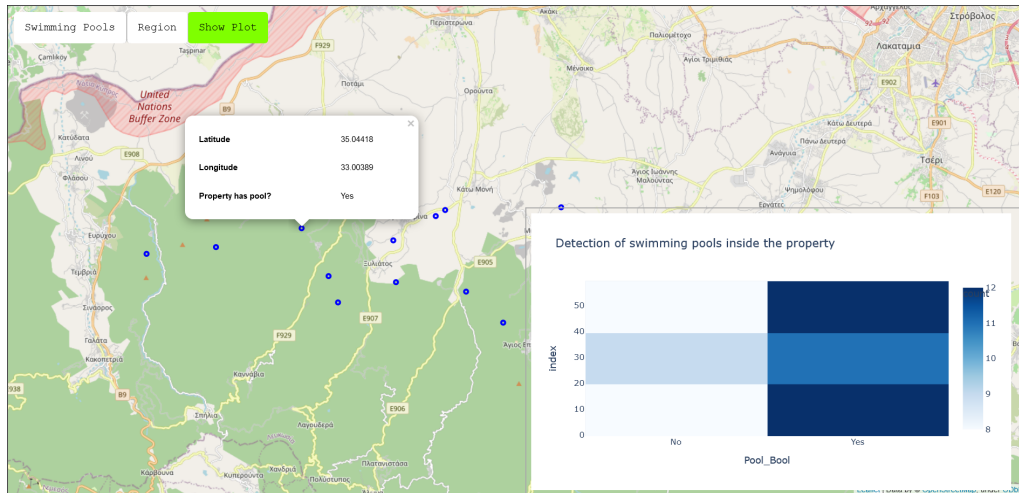


Figure 4.2: Detection of swimming pools inside the property

On the map, only the properties with a swimming pool are shown using the colour blue, as the given JSON response is either a 0 or 1. Information is also given in written format after clicking on the specific point that reveals a popup. The heatmap illustrates the number of properties with and without swimming pools within the selected region. Again the JSON response indicates whether the property has a pool or not in a list created by region which then Plotly interprets as the index.

Building Quality:

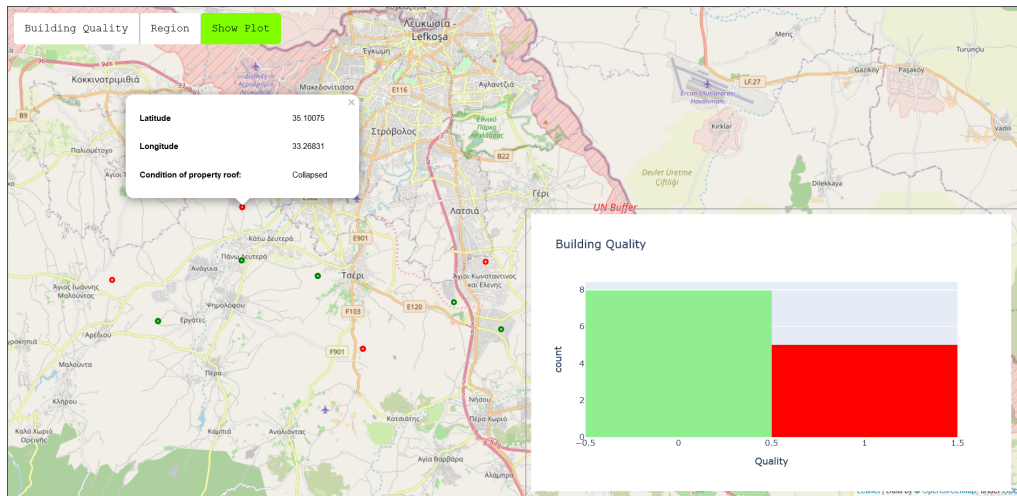


Figure 4.3: Building Quality

Once again, the JSON response is either 0 or 1, so the building quality is conveyed through coloured points, green being the properties without collapsed roofs, red indicating a collapsed roof. The same colours are, hence, used in the bar chart which sums up the total count for each roof status within the indicated region.

Detection of vegetation around the property:

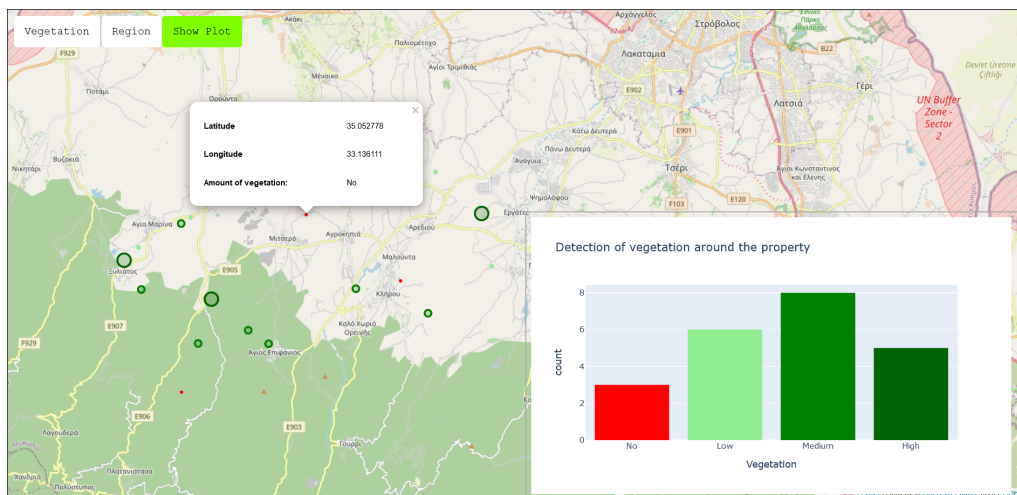


Figure 4.4: Detection of vegetation around the property

Here the level of vegetation is presented with both colour and size. In this case, the JSON

response is an integer from 0-3, where 0 is the representation of no vegetation and the 3 signifies the highest level of vegetation. So, red represents the lack of vegetation, and from light to dark green allocate to low, medium, and high amount of vegetation, respectively. The histogram also uses the same colours, and on the y-axis the number of properties for each level are recorded.

Subsidence Risk:

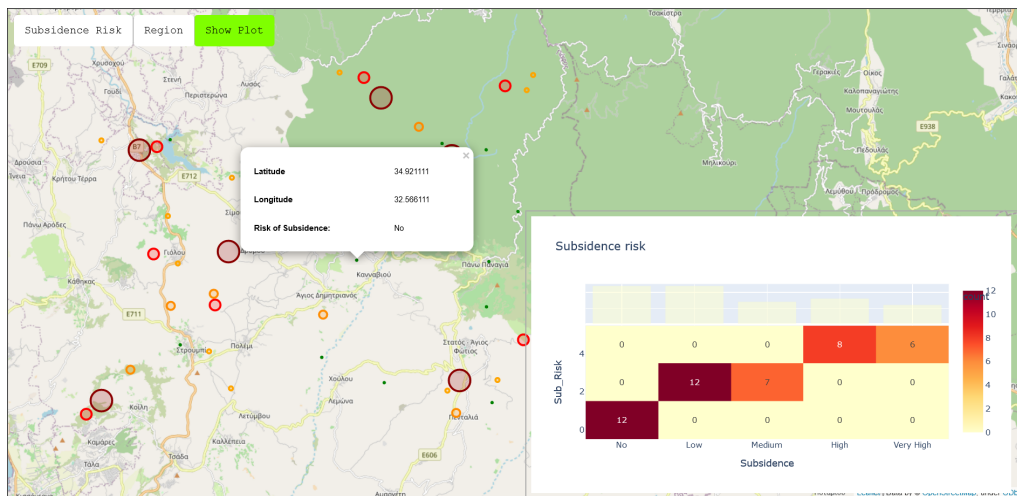


Figure 4.5: Subsidence Risk

The property points on the map illustrate the risk of subsidence through colour and size coordination. Green areas indicate that that specific property has no subsidence risk, which is also evident from its small size. Similarly, the reddest and largest points indicate the highest risk. This is because the JSON response is, similarly as the previous service, an integer from 1 to 5. The density heatmap shows how many properties inside the specified region are allocated to each risk level and a histogram is added to convey the same data through height difference. The number of properties for each level and risk is conveyed through the use of red and its shades.

Buildings' damage risk from earthquakes:

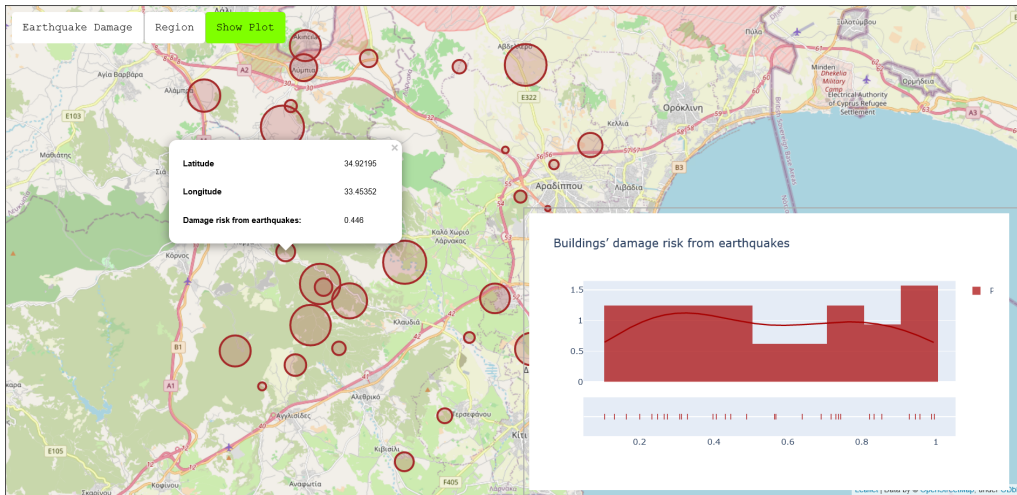


Figure 4.6: Buildings' damage risk from earthquakes

The risk level of the potential damaged caused by an earthquake is visualised on the map with the allocation of size and opacity according to the JSON response. The response is a number from 0 to 1 which is the probability of damage occurring. The larger and the more opaque the circle is the higher the probability of damage. The graph illustrates the probability density of the selected area, with the help of a curve to show the distribution.

Building Area:

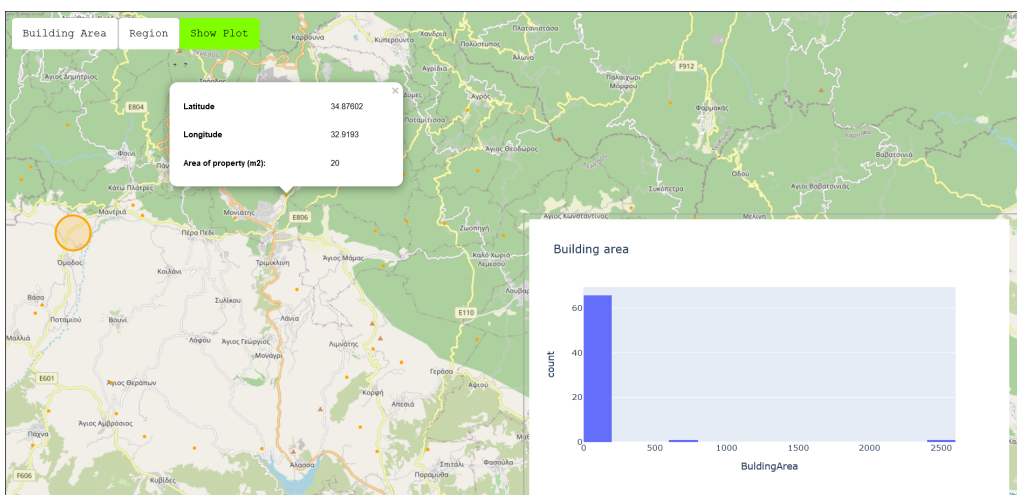


Figure 4.7: Building Area

The building area is represented through the circle's size of the coordinate point, where the larger the circle is, the larger the area of the property's building is as well. The JSON response is basically just a number which is also the area. The histogram conveys that most properties have a building area around 20m².

Proximity to roads:

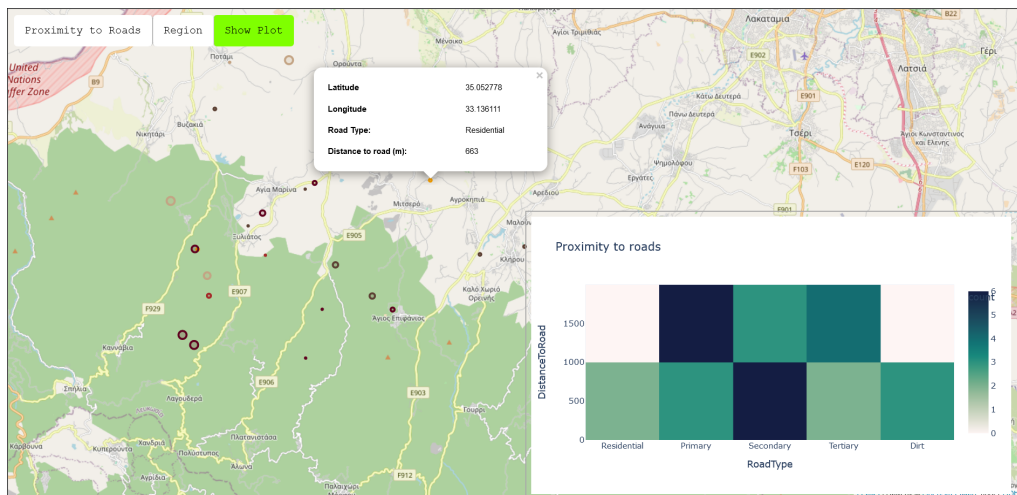


Figure 4.8: Proximity to roads

Each point, representing a property, has a certain distance from a different type of road. According to the JSON response, the given values are for the road types and the distance to that road type within a 2km radius, so 2000m. The colours indicate the type of road and the radius of the circle show the distance. A heatmap is used to show how many of each type there are and how far they are from their allocated property inside the defined region.

Proximity to the sea:



Figure 4.9: Proximity to the sea

Based on the JSON response, the value that is given is that of distance from the sea for each property is measured in meters. Each property is represented by a circle with fixed size and light blue colour where the popup states the exact distance to the sea. The histogram illustrates the number of properties for different ranges in distance. E.g. the selected property is 3479m away from the sea, so it's placed in the >3000 section of the x-axis.

Proximity to blue-flag beaches:

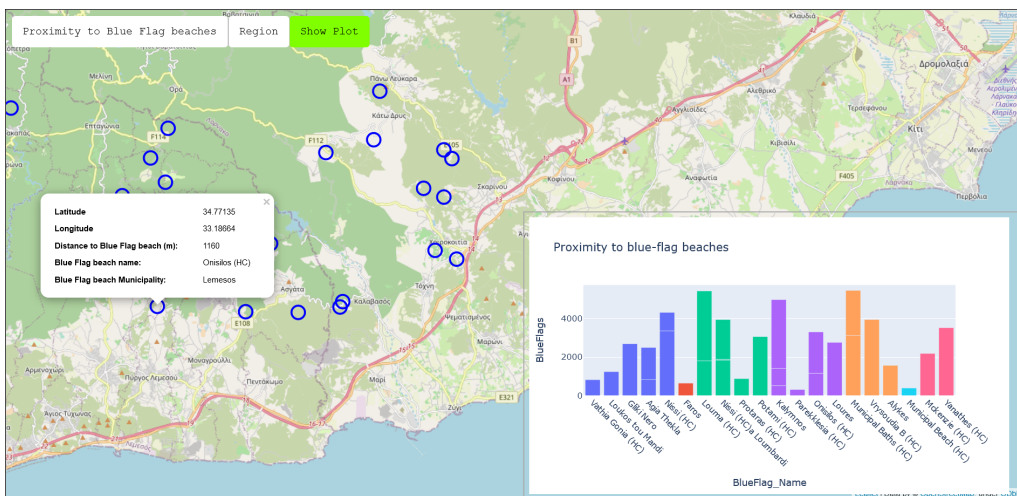


Figure 4.10: Proximity to blue-flag beaches

Other than the distance from blue flag beaches for each property represented in meters, the JSON response also provides data on the name of the blue flag beach and the municipality it is located in. Each property is represented by a blue circle with fixed size, much like the previous service on seas. The histogram presents the distance from which each property is away from different Blue Flag beaches and every colour is a different municipality. E.g. the selected property is 1160m away from Onisilos (HandiCap) beach in the municipality of Lemesos.

Proximity to amenities:

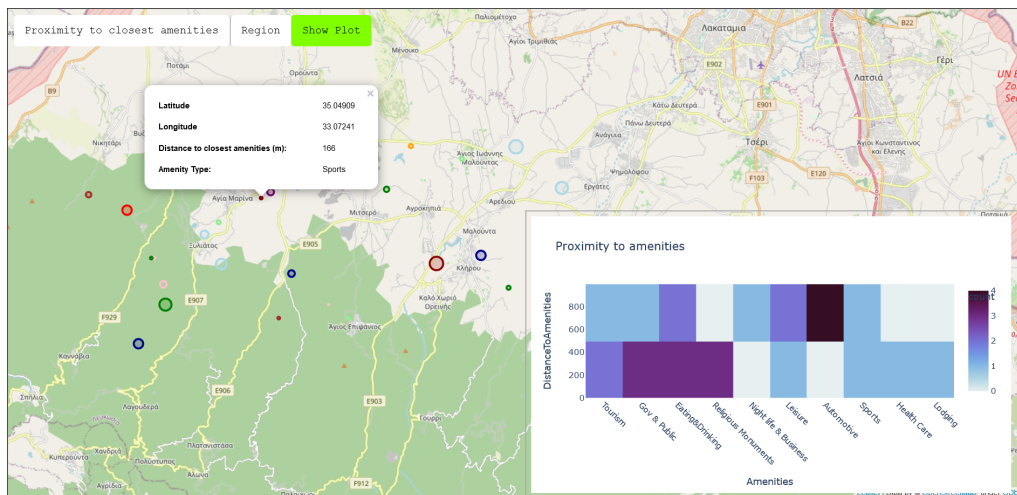


Figure 4.11: Proximity to amenities

On the map there are different colours that are allocated to different types of amenities, which are derived from the JSON response providing information on distance (within 1 km radius) and the category of the amenity. The radius of the circle is a representation of the distance to a specific amenity, while the colours refer to the category. Clicking on the points allows the popup to reveal more specifically the distance in meters and the amenity category. The heatmap shows the distance from the property to a certain amenity and colours are used to indicate the number of properties that measure their distance from the corresponding amenity. E.g. the selected property is 166m away from a sporting amenity.

Proximity to electricity network:

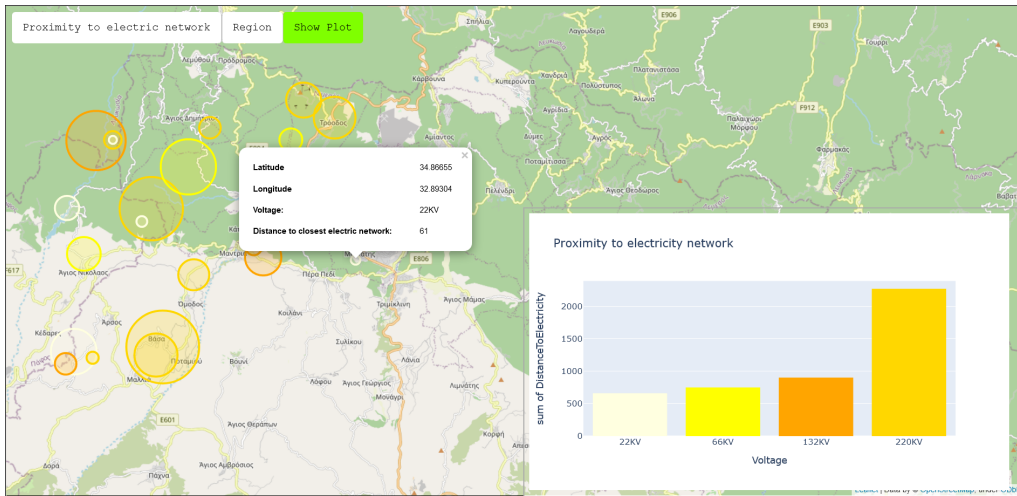


Figure 4.12: Proximity to electricity network

Present on the map are points showing how far the property is away from an electricity network. The JSON response also declares the voltage category for each electricity network which is resembled using colour intensities. Similarly, the histogram is intended to show the distance to the categorised electricity from each property in the selected region, however the way Plotly handles the data, it provides a total sum of the values representing distance that for each allocated category instead.

Detection of burnt areas:

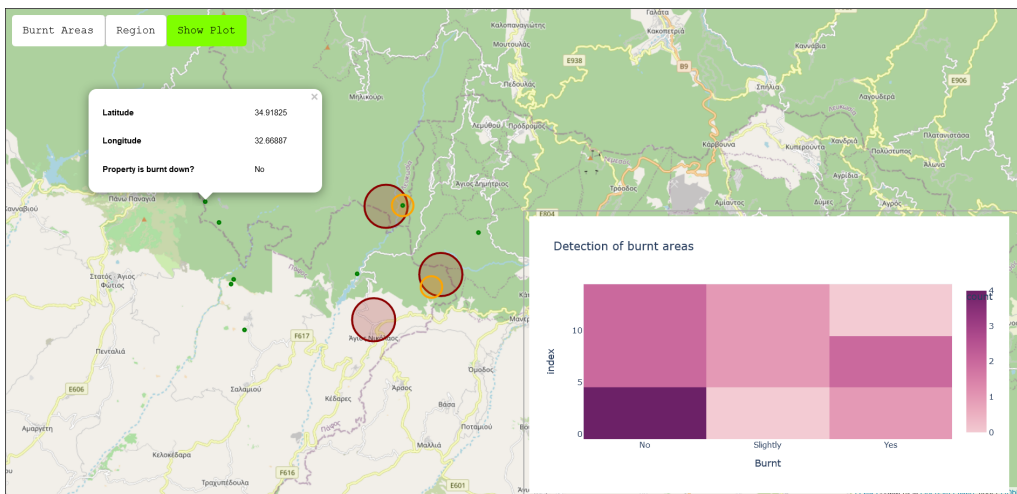


Figure 4.13: Detection of burnt areas

Three categorisations exist for this service, where the JSON response either declares a property as not burnt so it has a value of 0, slightly burnt so 0.5 or actually burnt which is a 1 . On the map, this is shown through the use of colours and size, in order to show the severity of the situation. So green spots illustrate a property that is not burnt, the slightly larger orange circle resemble slightly burnt properties and, finally, the largest red circle explicitly show that the property is burnt. The heatmap shows the state of each property within the selected area as well as the number of properties from the list of properties created when the region was specified, shown as the index.

Land use:

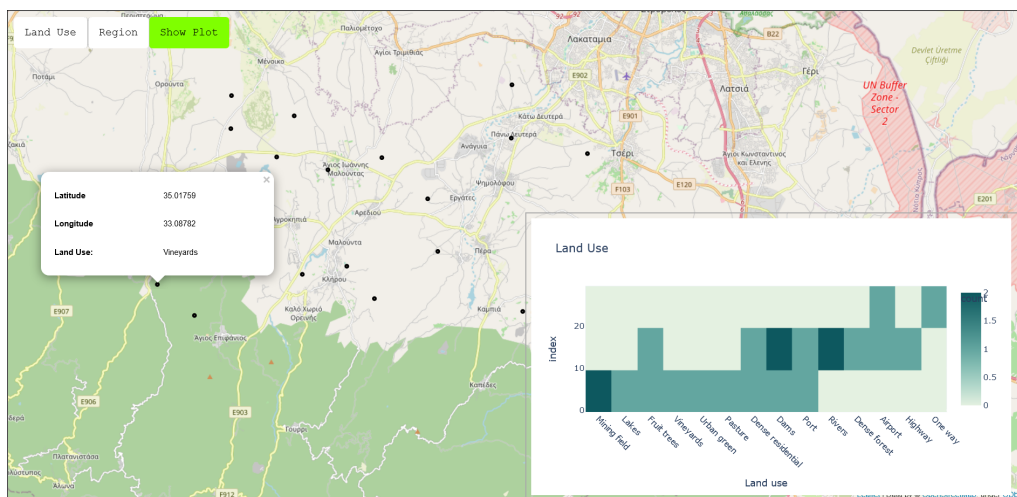


Figure 4.14: Land use

Each point resembles a specific use of the property, corresponding to the JSON response which is a piece of text labelling the classification of the type of land use. To see this, the coordinate points are clickable, so a popup provides the land use for each property and its given classification. The heatmap illustrates the type of use for each property within the selected region. E.g. the selected property is a vineyard alongside other properties within this area.

Natura 2000:

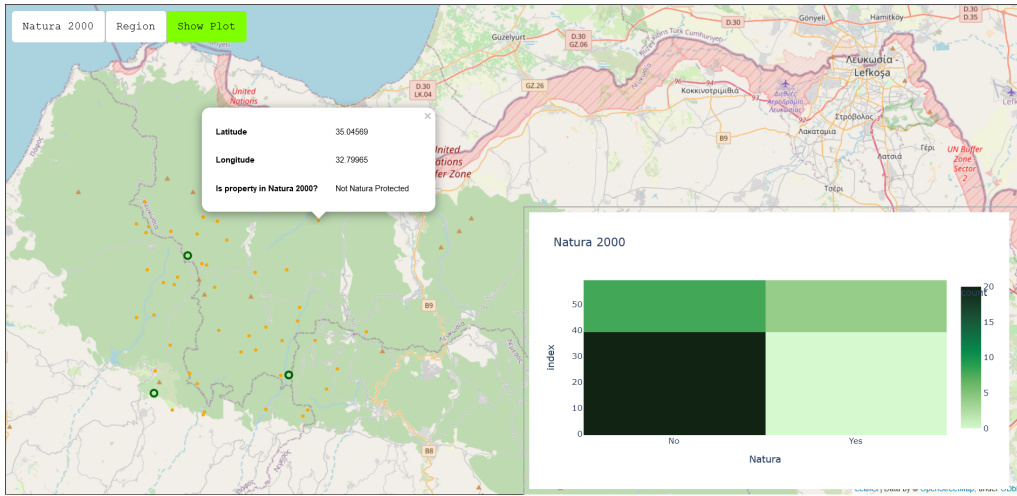


Figure 4.15: Natura 2000

The green circles correspond to a property under Natura 2000 protection, and the orange spots correspond to the properties that are not protected. Hence, the JSON response, that this is established on, has values that are either 1 or 0. The heatmap shows the locations of the Natura 2000 protected properties using the index which basically represents the order of the list of coordinates that are present in the selected region.

Risk of landslides:

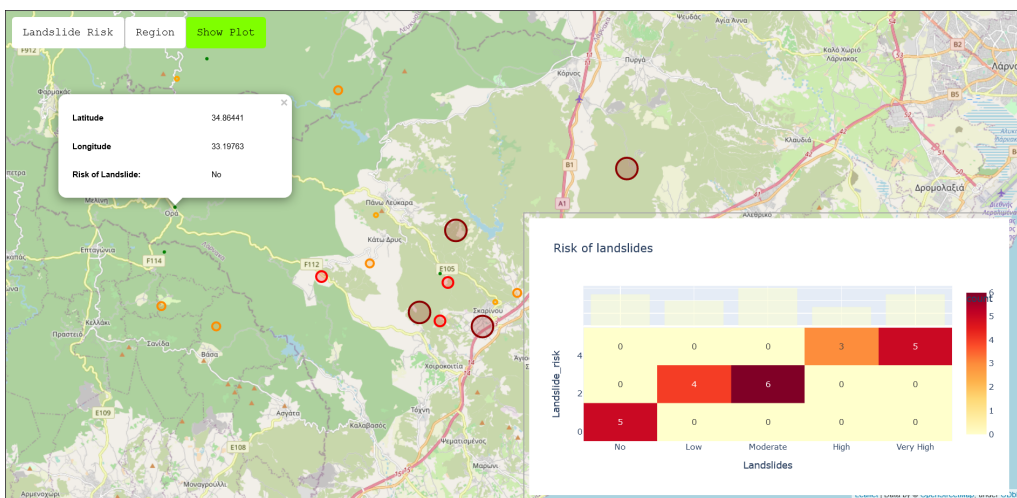


Figure 4.16: Risk of landslides

On the map there are circle markers with different colours ranging from green to red, where green signifies that there is no landslide risk and red signifies a very high risk. These allocations were derived from the JSON response of this service assigning an integer from 1 to 5 in order to show level of risk in an ascending order. On the density heatmap it is illustrated how many properties inside the indicated area are under different levels of risk using different intensities of the same colour. The histogram on top shows the same data through difference of height as well.

Wildfire Risk:

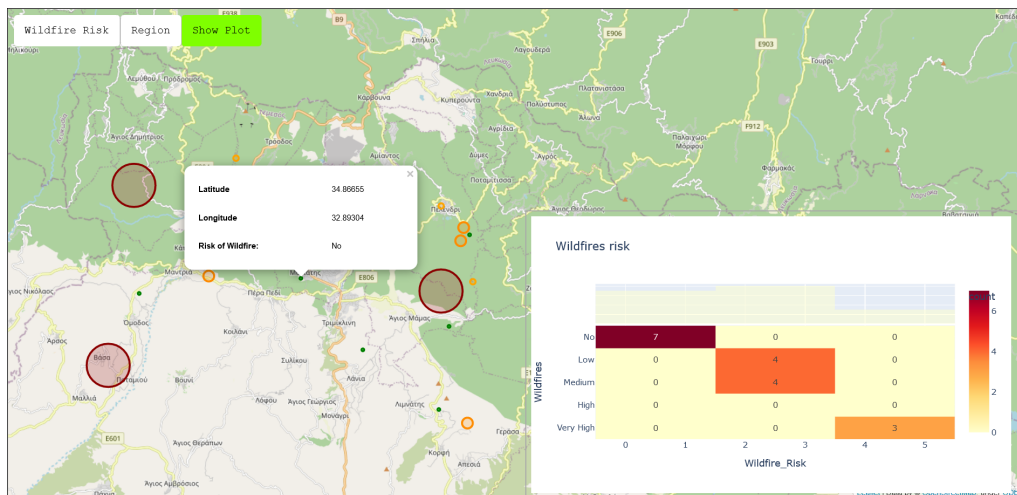


Figure 4.17: Wildfire Risk

The map is covered with colour coded circles of different sizes depending on the risk of a wildfire, which are derived from the values 1-5 given by the JSON response. The green colour conveys that there is no risk of wildfire and the gradual increase from orange to the darkest red convey a low to a very high risk of wildfire, respectfully. The density heatmap presents the number of properties for each risk allocation with the use of a shaded colour group. The histogram is added above to show the difference through height.

Flooding Risk:

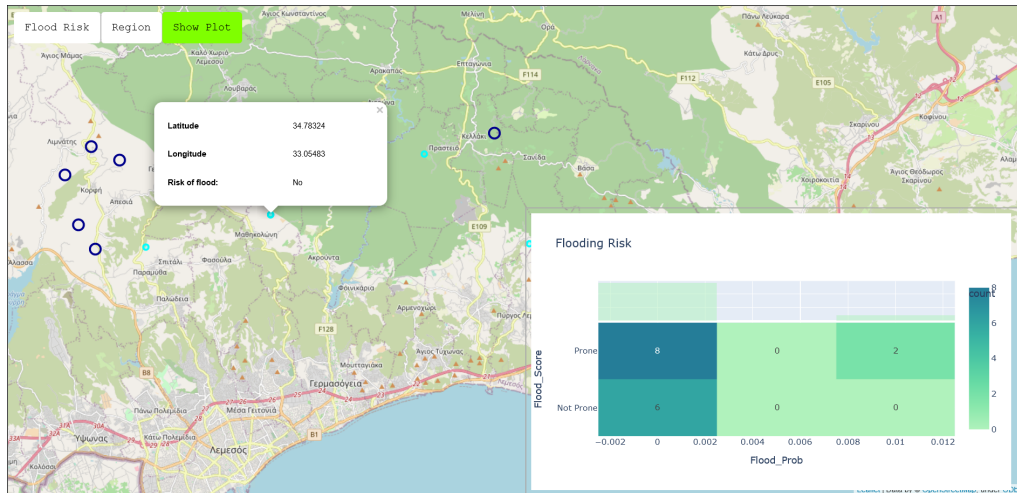


Figure 4.18: Flooding Risk

On the map the dark blue circles represent the properties, and the popup reveals if it is either at flood risk or not, originating from the JSON response data. Not only that but, the density heatmap is used to demonstrate the number of properties that are allocated at each property's probability of experiencing a flood, consisting of 1/20 chance of occurrence, 1/100 and 1/500. The colour concentration allocates the total count for each probability combination.

Slope and Aspect:

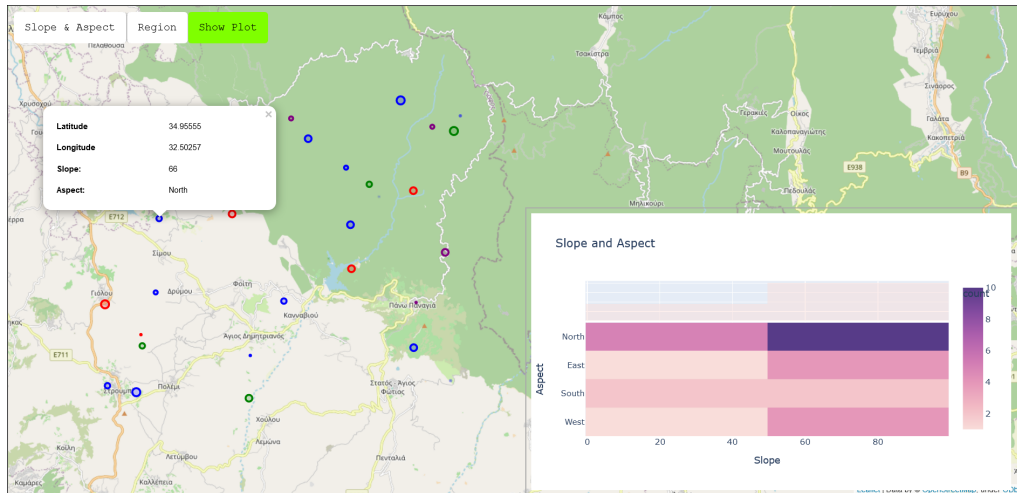


Figure 4.19: Slope and Aspect

The map is used to show the properties' slope and aspect, channelling the data from the JSON response. The slope takes values from 0-90 degrees and the aspect takes values from the corresponding angles of compass directions, meaning the 0 and 360 degrees are North, 90 degrees mean East, 180 degrees refer to the South, and finally at the 270 degree mark there is the West. The size of the circles correspond to the slope size and the colour is based on the aspect categorisation and the popup reveals the slope value and the aspect direction. The density heatmap shows the number of properties with all combination of aspect and slope present in the selected area. E.g. this point has a slope of 66 degrees facing North.

Geology:

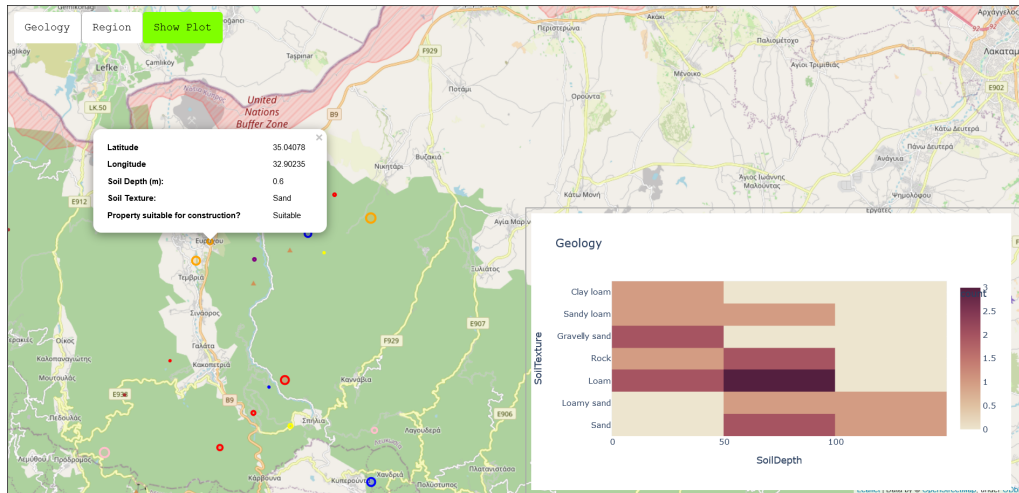


Figure 4.20: Geology

On the map each point is allocated a colour that represents a different kind of soil texture. The circle radius indicates the soil depth in m and the popup reveals if the property is suitable for construction, basing this information is on the JSON response. Using a heatmap the soil textures can be separated based on soil depth regions and their frequency with the given area using colour intensities. E.g. at the chosen property the soil texture is sand, the soil depth is 0.6m and it is suitable to build on.

Precipitation:

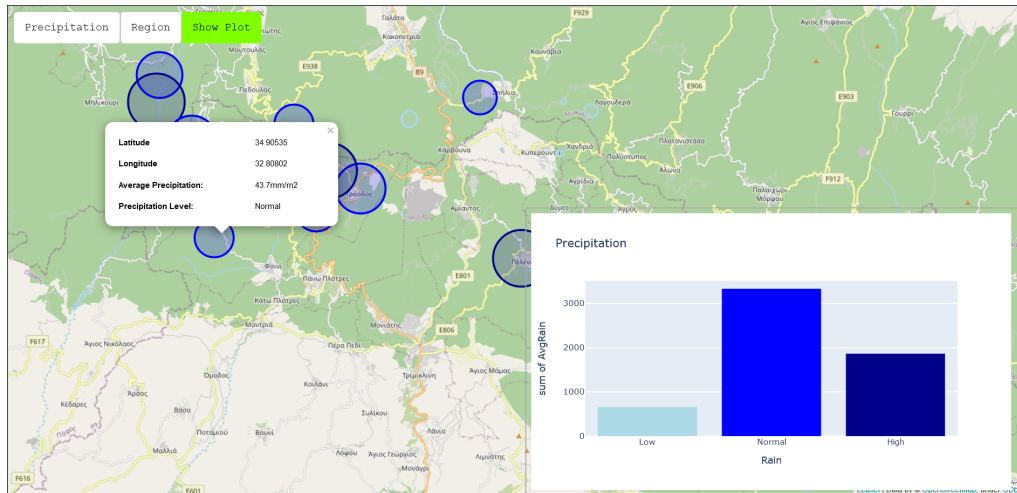


Figure 4.21: Precipitation

The average precipitation is shown on the map using the radius of the circle as reference, declared by the JSON response taking in the average value of precipitation per year, considering 39 years of data in mm per metre squared. Although there is not further information of which years each average is centred on. The precipitation level is illustrated through three different shades of blues from light to dark. The histogram sums up the total precipitation from each level with the same colours as in the map. Initially, the idea was to show the amount of rain on each property in the selected region, however the way Plotly handles the data, it provides a total sum of the precipitation values for each allocated category instead. The categories are obtained from having an average over 600mm/m² which is considered high, between 350 and 600 mm/m² is said to be normal and anything under 350mm/m² is seen as low levels of precipitation. E.g. this particular property has an average precipitation of 43.7mm/m² and a normal precipitation level.

Elevation:

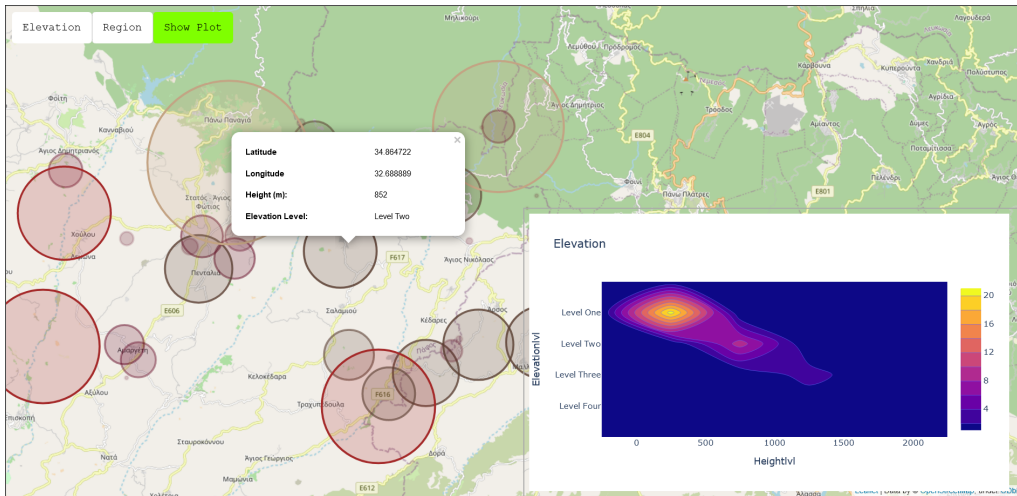


Figure 4.22: Elevation

The elevation is visible on the map through circles of different sizes, depending on the height in m, and the elevating level is conveyed using colour shades of the same blue colour group. These values are denoted by the JSON response of the height in metres and the elevation level which are categorised in four levels. A density contour graph shows the height of the location around the corresponding elevation level. E.g. in this case the property location is at a height of 852m at the second level of elevation.

Wind:

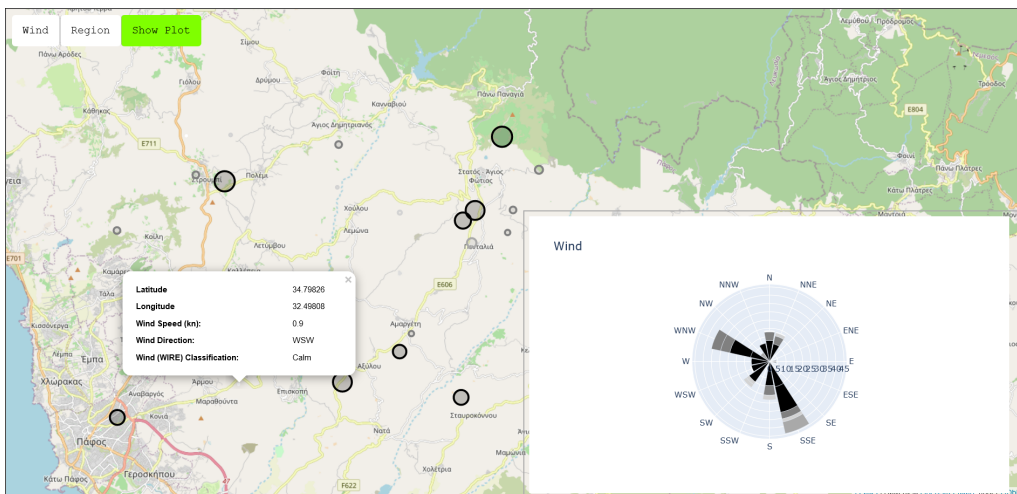


Figure 4.23: Wind

On the map the speed of the wind is illustrated by the circle radius and the colour is a representation of each property's wind class and utilising the popup, the wind direction is also displayed. The JSON response also delivers a classification for the wind strength according to the speed of wind in that point. The polar bar chart is divided into a section based composition, much like on a compass, for the wind direction. Again the size of the slice describes the speed, and the colour is the classification of the wind for that selected property within the specified region. E.g. the chosen area has a wind speed of 0.9knots pointing WSW falling under the calm wind class.

Chapter 5

Evaluation

5.1 Testing

The project's evaluation phase was carried out in the form of usability testing, taking into account the nature and status of the project. The researcher basically asks a participant to complete tasks using the created user interface during an allocated usability testing session. The researcher monitors the participant's behaviour and receives feedback while the testers complete each task. The major objective of the testing is the user interface; it is not to test the user. As a result, usability testing makes it possible to spot issues with the web application's design and identify areas for improvement, all while learning from the actions and experiences of the intended users. Finally, the results are crucial for determining whether the overall effort was successful or not.

Setting up a usability test to conduct, testers were given a description of the goal of the project in general and the test's objectives. Not only that but, each user agreed to keep all details regarding the project remain confidential and, from our part, we ensure them that any personal data collected on the testers will also stay confidential. The gathered information consists of how well the potential users can understand the data visualisations, how challenging it is to navigate through the user interface, and finally their own opinion expressed based on their experience as a whole. Moreover, it was also crucial to consider

the testers' varied backgrounds since the project's main stakeholders should eventually be regular people. The inclusion of individuals from non-technical fields helped create objectivity in the end results because they are not expected to have highly developed technical skills and understanding, or any other background of a similar nature.

The tasks that the participants had to complete were based on possible scenarios the web application may be used in a real life situation. These scenarios cover a wide range of different cases using different map objects and graphs to illustrate the appropriate data:

- Locate an area with high levels of vegetation
- Identify the city with the highest precipitation
- Identify the coastal city with the highest flood risk
- Identify the most common amenities in Larnaca
- Identify the property with the largest building area
- Locate the Natura 2000 protected property nearest to Lemesos
- Locate an area suitable to install wind generators

For each of the scenarios there were three questions to answer, first being “How comprehensible is the map visualisation?”. Similarly, the second question is “How comprehensible is the graph representation?”. Both questions had a Likert scale from one to five, “Not at all” and “Very Comprehensible”, respectively. Finally, the third question was “What would you change in this service?”, where the participants could write down something new they would rather see in the visualisation.

5.2 Results

In general, there was a variety of participants ranging from architecture students and primary school teachers to more tech oriented interests, such as computer programmers and engineers. Something that was evident was the fact that individuals with more technical backgrounds grasped the concept of the application quicker than the non-technical

participants. Not only that but, being in the presence of the evaluator did not interfere with the responses by the users since the testing session was transparent and constructive criticism was welcome. The presence of the evaluator also helped participants understand the context of the project more concretely, as well as clearing up any of the questions that arose throughout the testing.

Moreover, the initial assignment of Folium to produce the map visualisations was deemed rather unsuccessful for a few outcomes. Taking into account that the circular points on the map, representing the coordinates within the restricted area chosen by the user, were dependent on the type of data they would show. Since the only aspect that could be changed was the radius of the circles, some of the data illustrations were successful such as the Wind service where the circle markers matched well with the unique type of graph used in that specific scenario. However, some other types of services didn't really hit the mark such as the services that have to do with distances from a given property. Even between services with these kind of similar aspects, there were still some elements that they did not share which made the difference between a comprehensible and an incomprehensible visualisation. For instance, the Proximity to Amenities service differs from other proximity services because the points are allocated based on the type of service and not the distance. So there are services that take values which are represented through the circle areas and colour.

Similar issues apply to the Plotly graph visualisations since only specific examples were comprehensible enough, such as the vegetation and wind service. To be more precise, there was feedback gathered expressing a negative reaction to the graph representations of the third scenario, "Identify the coastal city with the highest flood risk", and fourth scenario, "Identify the most common amenities in Larnaca". The problem lied in the heatmap graphs, and the participants that shared this opinion were those with less technical experience []. Perhaps the responses would have been different if the heatmap could have been applied on the map itself successfully. Additional comments included the confusion with some of the graph axis, particularly the ones with the "index" label, meaning the list order the coordinates are stored in.

Other than the visualisations, using the measure control tool to draw a polygon, for specific area selections on the map, was complicated at first but the testers got used to it after one or two uses. However, most users did indicate that the whole procedure to select a specific region was tedious. While the testing session was still taking place, the participants were asked orally if they would have figured out how to select a region on their own. Some of the participants replied that they could have actually figured it out without any external help, while others said that it would be more time consuming and confusing. The difference between the testers that gave opposing responses was once again based on their own profession and the underlying skills surrounding their professional activities. This leads to the final part of the testing session where participants were asked if they would have used this application in the case they wanted to enter the real estate market. Most stated that they would not use the present version of the application, but they would if it was completely finished properly in terms of both functionality and appearance.

Chapter 6

Discussion

6.1 Limitations

Throughout the course of the project, a number of limitations have been discovered. Going further along in the development process, the tools that were chosen initially for the project's implementation presented technical restrictions. As mentioned in the Realisation chapter, Folium and Plotly were eventually used to compensate for the minimal map features provided by Folium. Another issue that arose was the inability to communicate dynamic data between Flask and the web languages which seemed difficult to accomplish within the given timeframe. As a result, there are some altercations that had to occur in order to decrease the chances of running into an error while navigating through the graphical user interface.

Furthermore, as there was no specific drawing tool available for use with the selected programming tools, which made it even more challenging for the user to specify an area on the map. Although a solution was found with the appropriate resources, more time was spent modifying an existing plugin within the Folium library, aforementioned above, called the Measure Control plugin. Not only that but, as stated in the Evaluation, this method consisted of too many steps to be repeated each time the user wanted to select a new region to work in.

Finally, rather than using real-world data during the implementation phase of the project, due to not having access to the API, dummy data was employed for evaluation and testing purposes. This resulted in random data points which didn't match the visible landscape of the map, for example it is somewhat clear which areas have vegetation or not b simply looking at the map.

6.2 Future Work

Regarding the potential of this project, there are changes that can be made in order to improve the application as whole. First of all, with the use of Leaflet, the JavaScript equivalent of Folium, more detailed visualisation can be produced without the need to combine two or more different elements using different methods. In addition, this will open the door to use other external tools to enhance the data visualisations even further. One example is the utilisation of Mapbox, another JavaScript library, which allows the integration of geo data on the map itself and not be limited to singular coordinate points, as explained in a previous chapter. Currently, all data about the services comes from a static JSON file, but with the integration of official data into a database, the data required to run the services can be updated dynamically. In return, this will make the use of Mapbox even better as it uses databases to run the data and not JSON files.

Excluding the technical aspects of the project, there are multiple initiatives that can use this application. The two main candidates are in the real estate market, of course, and the insurance industry. This web application, or the future working version of it, can bridge the gap between client and service provider. For instance, environmental disasters have the potential to harm many properties, and insurance companies don't fully consider the underlying dangers of any kind of natural disaster, making some property destructions "uncoverable". Even though it is outside the scope of this project, behind the product's accessibility, meaning who has actual access to and use of the software, lies an ethical conundrum. In order to actually inform the clients on the environmental data surrounding

their already owned property, or even if they're trying to invest in a new one, the web application would have to be accessed publicly and not only by private corporations. Moreover, should we push for the project's scope to extend beyond Cyprus's borders if doing so would garner additional support and funding? Well, the answer will be revealed in the future as the project grows larger.

Chapter 7

Conclusion

When correctly processed and visualised, geodata may be a very significant tool, particularly in the real estate sector because environmental risks are sometimes invisible before the disaster occurs. The heavy nature of environmental records can be processed by computer algorithms and programmes with the correct data and illustrated by applying data visualisation, which can transform quantitative data into a visual representation. Using computer-generated images, it is simple and economical to explore and analyse data's underlying relationships and patterns in order to learn more and improve comprehension about environmental risks. More efforts for climate study may be investigated as artificial intelligence and machine learning algorithms progress. As a result, it has been demonstrated to be advantageous to use AI to identify and mitigate environmental concerns, hence this is a straightforward method to implement analogous methods in the real estate business.

The project's goal was to build a GUI and visualisations on top of an application programming interface (API) that already exists and contains a list of services relevant to real estate properties, but things turned out differently. Quantifying these hazards based on the climate data found in a particular property was one of the most difficult tasks, considering the constraints that emerged along the way.

Revisiting the first research question, analysis and visualisation of geospatial information related to real estate properties can be done using geodata and satellite imagery. Then, these numbers can be quantified and represented in numerical or category form to be seen visually. However, the execution did not follow the initial design concept but still managed to represent the data by combining two types of data representation. This leads to the second research question revolving around AI-driven data visualisations and how they give insights for real estate depending on the interpretation of the said data visualisations. Instead of using rows and columns filled with numbers that don't make relative sense, except the developers involved, more detailed and descriptive methods can be used which are also based on context. The overall result may be considered unsuccessful as the full potential of the project has not been reached within the allocated time period.

Bibliography

- [1] Qi Li. “Overview of Data Visualization”. In: *Embodying Data: Chinese Aesthetics, Interactive Visualization and Gaming Technologies*. Singapore: Springer Singapore, 2020, pp. 17–47. ISBN: 978-981-15-5069-0. DOI: 10.1007/978-981-15-5069-0_2. URL: https://doi.org/10.1007/978-981-15-5069-0_2.
- [2] Prof Grewal. “A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering”. In: *IOSR Journal of Computer Engineering* 16 (Jan. 2014), pp. 09–13. DOI: 10.9790/0661-16210913.
- [3] Federica Zennaro et al. “Exploring machine learning potential for climate change risk assessment”. In: *Earth-Science Reviews* 220 (2021), p. 103752. ISSN: 0012-8252. DOI: <https://doi.org/10.1016/j.earscirev.2021.103752>. URL: <https://www.sciencedirect.com/science/article/pii/S0012825221002531>.
- [4] Chris Huntingford et al. “Machine learning and artificial intelligence to aid climate change research and preparedness”. In: *Environmental Research Letters* 14.12 (Nov. 2019), p. 124007. DOI: 10.1088/1748-9326/ab4e55. URL: <https://doi.org/10.1088/1748-9326/ab4e55>.
- [5] Rob Toews. *These Are The Startups Applying AI To Tackle Climate Change*. Forbes. URL: <https://www.forbes.com/sites/robtoews/2021/06/20/these-are-the-startups-applying-ai-to-tackle-climate-change/>.
- [6] *Mitigating climate risk | Swiss Re*. www.swissre.com. URL: <https://www.swissre.com/risk-knowledge/mitigating-climate-risk.html> (visited on 03/27/2022).
- [7] In-Ae Yeo and Jurng-Jae Yee. “A proposal for a site location planning model of environmentally friendly urban energy supply plants using an environment and

- energy geographical information system (E-GIS) database (DB) and an artificial neural network (ANN)”. In: *Applied Energy* 119 (Apr. 2014), pp. 99–117. DOI: 10.1016/j.apenergy.2013.12.060. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0306261913010696> (visited on 11/27/2021).
- [8] Ender Buğday and Seda Erkan Buğday. “MODELING AND SIMULATING LAND USE/COVER CHANGE USING ARTIFICIAL NEURAL NETWORK FROM REMOTELY SENSING DATA”. In: *CERNE* 25 (June 2019), pp. 246–254. DOI: 10.1590/01047760201925022634. (Visited on 04/17/2020).
- [9] *PropertyQuants | Bringing Quantitative Investment Strategies to Global Real Estate*. propertyquants.com. URL: https://propertyquants.com/?utm_source=google&utm_medium=cpc&utm_campaign=P-Real%20Estate%20Data&utm_term=ai%20and%20real%20estate&utm_content=&gclid=Cj0KCQiAmpyRBhC-ARIsABs2EAoqYF6yQfUQeR08cbSzs0Dst4JQCvYupcDj01yCJGtQywS9Lr7VekIaAmXjEALw_wcB#approach.
- [10] *Commercial Real Estate*. One Concern. URL: <https://oneconcern.com/en/industries/commercial-real-estate/> (visited on 03/27/2022).
- [11] Andrew Dlugolecki. “Climate Change and the Insurance Sector”. In: *The Geneva Papers on Risk and Insurance - Issues and Practice* 33 (Dec. 2007), pp. 71–90. DOI: 10.1057/palgrave.gpp.2510152.
- [12] *Parametric Insurance*. Descartes. URL: <https://www.descartesunderwriting.com/technologies/parametric-insurance/> (visited on 03/27/2022).
- [13] *Reinsurance Solutions*. ourkettle.com. URL: <https://ourkettle.com/reinsurance-solutions/> (visited on 03/27/2022).
- [14] M. Khojah M. A. Erskine and A. E. McDaniel. “Location selection using heat maps: Relative advantage, task-technology fit, and decision-making performance”. In: *Computers in Human Behavior* 101 (Dec. 2019), pp. 151–162. DOI: 10.1016/j.chb.2019.07.014.
- [15] M. DeBoer. “Understanding the Heat Map”. In: *Cartographic Perspectives* 80 (Nov. 2015), pp. 39–43. DOI: 10.14714/cp80.1314.

- [16] J. Zhang and S. You. “Supporting Web-Based Visual Exploration of Large-Scale Raster Geospatial Data Using Binned Min-Max Quadtree”. In: *Lecture Notes in Computer Science* (2010), pp. 379–396. DOI: [10.1007/978-3-642-13818-8_27](https://doi.org/10.1007/978-3-642-13818-8_27).