Knowledge graph visualization on sustainable building design

Charles Klijnman
5995272

A thesis presented for the degree of
Bachelor of Science

First supervisor: Michael Behrisch
Second supervisor: Alex Telea
Collaborator: Katharina Hecht

Dept. of Information Science

April 30th, 2023

Utrecht University
Heidelberglaan 8
3584 CS Utrecht
The Netherlands

Abstract

This project aims to address the environmental impact of buildings, which account for 40% of global energy consumption and 30% of greenhouse gas emissions. To help architects and building designers implement sustainable design strategies, a web application was developed to provide easy access to relevant data. The application enables users to analyze the dataset, extract insights, and append data to it. The report describes the research methodology, development process, challenges encountered, and project achievements. New data sources have been added to the provided dataset, which has enabled the discovery of interesting insights that were previously inaccessible. The web application successfully meets its objective of providing architects and building developers with insights into sustainable building design strategies and ecosystem services. The application offers a user-friendly interface, enabling visual exploration and evaluation of the dataset. The expert study evaluation revealed potential features and improvements that could enhance the application, most of which were successfully implemented. Additionally, it has been deployed on a server, allowing users to independently explore the dataset and discover valuable insights. However, the application has certain limitations, such as a cluttered graph when too many nodes and relationships are added. Future work includes expanding the visualizations, adding more data to the dataset, and incorporating a predictive algorithm to forecast potential design strategies. Overall, this project provides a valuable resource for architects and building designers to help reduce the environmental impact of buildings.

Keywords: knowledge graphs, sustainable buildings, design strategies, ecosystem services

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Michael Behrisch, for his valuable guidance and continuous support throughout my project. His knowledge and expertise in the field have been invaluable, and his feedback and insights were critical to the successful completion of this work. We met regularly, typically once every one or two weeks, where we reviewed progress, discussed challenges, and planned upcoming tasks.

I would also like to thank Katharina Hecht for her generous assistance and support in providing her dataset on sustainable building design. Her insight and expertise in the field were incredibly helpful, and her time spent explaining the domain and her specific needs was all-important.

Finally, I would like to thank my family and friends for their unwavering support and encouragement throughout this project. Their love and encouragement kept me motivated and inspired to work towards the successful completion of this work.

Thank you all for your contributions and support.

Contents

List of Figures

List of Tables

1 Introduction

Buildings are a significant contributor to environmental problems such as climate change and pollution. Buildings are responsible for about 40 percent of global energy consumption and 30 percent of greenhouse gas emissions. [1] For that reason, new building projects should keep sustainability in mind. Sustainable buildings are designed to minimize their environmental impact. [2] They use innovative design strategies and techniques to, for example, reduce energy consumption, minimize water usage, and reduce waste. Furthermore, urbanization is playing a role in the reduction of biodiversity, which may lead to changes in the functioning of ecosystems and their capacity to provide essential goods and services to society. [3,4] Unfortunately, there is a lack of a tool or platform for architects and building designers about the different sustainable design strategies that they can use. Therefore, I have been working closely with Katharina Hecht, a PhD candidate who is trying to bridge the gap between ecological science and the building industry. [5] Publicly available resources on this subject are not consolidated, making it challenging for architects to access them in a single place. The tool should integrate the different concepts of the domain to provide an easy-to-use implementation and make the data more intuitive and accessible. My responsibility was to create a platform based on the preliminary versions and ideas, aimed at informing building developers about the potential of integrating sustainable design strategies in their upcoming construction projects. A web application was developed to accomplish the project's goal. The application enables users to analyze the dataset and extract significant insights from it. Additionally, it provides the ability to append data to the existing dataset.

This document includes details on the research methodology utilized, along with a comprehensive breakdown of each individual phase. The report describes the achievements of specific milestones, as well as a detailed description of the development process for a self-build web application, including any challenges encountered. It also includes a data ingestion pipeline and a user manual for the application. Moreover, the report evaluates both the overall project and the effectiveness of the application.

2 Research Design

In this section, an overview of the methodology utilized throughout the course of this project will be presented. Additionally, an explanation for the employment of this method will be elucidated. Furthermore, a brief explanation of the activities undertaken in each phase of the methodology will be provided.

For this project, the CRISP-DM methodology has been employed. [6] This is an industry-standard method for data mining projects. To fit the project setup, the tasks done in each of the six major phases have been modified a little bit. A visualization of the methodology can be found in figure 1. At the start of the project, during the project setup, the main goals were made and a requirements analysis was done using the MoSCoW principle. [7] On top of that, the project planning has been visualized using a Gantt chart (Appendix C). Each one of the six phases took approximately 4 to 5 weeks to complete.

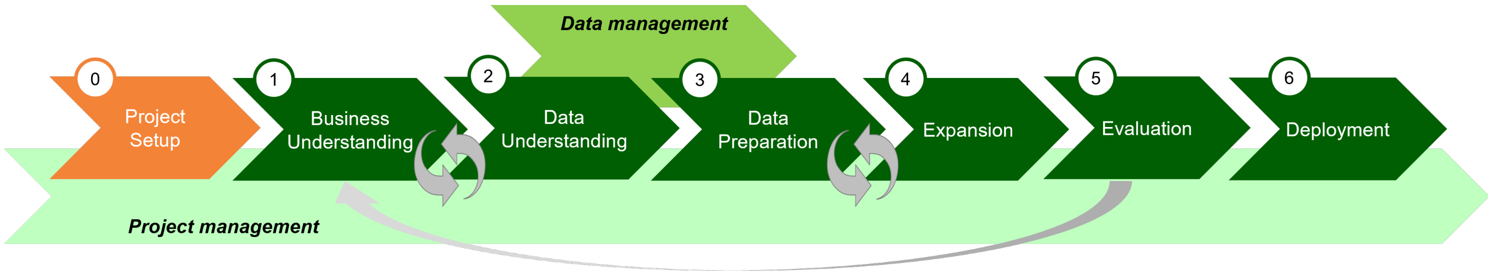


Figure 1: Modified CRISP-DM Methodology

In table 1, the outcome of the MoSCoW prioritization analysis is shown. The most important requirements are the "must-haves", which are non-negotiable needs that are mandatory requirements for this project. Additionally, there are the "should-haves", which are initiatives that are not vital but add significant value. Moreover, the "could-haves" are desirable features but not essential to meeting the expected requirements of this project. And lastly, the "won't-have" refers to the elements that are not required for this project's scope.

Must	Convert existing dataset into a graph db dataset
	Method: Visual Data exploration including a Data Science pipeline.
	Evaluation: Case-study evaluation (some valuable insights); Description of insights based on visualizations
Should	Merging more (1-3) datasets
	Apply machine learning algorithms
	Multiple visualizations on dataset
	Data ingestion pipeline
Could	Interactive Dashboard
	Data ingestion frontend
	Expert study (1 expert)
Won't	Technical evaluation metrics (speed, throughput, etc)

Table 1: Project's MoSCoW prioritization

The *Business Understanding* phase includes a thorough investigation of the sustainable building design field, which is essential for interpreting and analyzing data gathered during the *Data Understanding* phase. This exploration was performed using a literature review. As additional domain-related questions emerged during the *Data Understanding*, there was some iterative movement between these two phases. In addition to comprehending the data, the data management process also implicated the 'Data Preparation' phase. This phase involved tasks such as selecting, cleaning, constructing, formatting, and merging the data. The final undertaking of this phase, which entailed transforming the data, also ensured the accomplishment of the initial requirement of converting the existing data into a graph database. Subsequently, the *Expansion* phase began, which consisted of the exploration of potential machine learning and visualization implementations. Once again, there was iterative movement between this *Expansion* phase and the preceding phase. In the *Evaluation* phase, an assessment of the project outcomes was conducted, including a review process. In the final phase, *Deployment*, some time was allocated for completing the documentation and, if applicable, deploying the solution. Detailed elaboration on each phase is presented in the subsequent sections.

3 Business Understanding

In this phase, an extensive exploration of the sustainable building domain is conducted. This literature review will explore the significance of sustainable buildings, the design strategies utilized to achieve them, and the range of ecosystem services they provide.

3.1 Sustainable Buildings

Sustainable buildings are designed and built to minimize their negative impact on the environment while maximizing their positive impact on the people who live and work in them. They achieve this through the implementation of various design strategies or techniques. [8] This is particularly crucial given that buildings account for 40% of the world's energy consumption. [9] One way to decrease building energy consumption is by improving energy efficiency in the utilization of heating, lighting, and cooling, as these consume the majority of a building's energy. [1] Moreover, the most sustainable design strategy is to conserve energy as much as possible. Passive techniques like natural or hybrid ventilation, as opposed to air conditioning, can substantially reduce primary energy usage. [10]

In addition to their energy consumption, buildings are a significant contributor to greenhouse gas emissions, accounting for approximately 30% of global emissions. [11] Furthermore, buildings today rank third among the largest consumers of fossil energy, following agriculture and industry. [12] By integrating renewable energy systems into buildings the reliance on fossil fuels can greatly decrease. For example, generating energy using solar panels has become a popular implementation in recent times. [13] Encouraging the adoption of innovative renewable technologies and strengthening the market for renewable energy will help protect the environment by cutting emissions.

3.2 Design Strategies

Besides strategies that revolve around reducing energy usage and emissions, there are tons of different other design strategies that can generate a variety of ecosystem services. Including a couple revolving around urban agriculture, which involves integrating food production with buildings. Similar to other sustainable design strategies, building integrated agriculture should be multi-functional and able to integrate well with the cultural and social system of the location. This approach provides a way to use available land without affecting the many other uses of the city and offers numerous benefits as well. For example, adding green spaces to rooftops can help mitigate the urban heat island effect. [14,15] By bringing food production closer to where it is consumed, food miles can be reduced, resulting in lower carbon emissions from transportation. [16] In addition, this approach eliminates the need to allocate land outside the city for large-scale commercial farming, which has numerous health, social, and environmental impacts. There are several other design strategies available, including Blue/Green roofs and walls, Rainwater harvesting, and Material reuse and recycling, among others. A detailed explanation of each strategy's implications won't be provided since it's beyond the scope of this project.

However, Appendix A contains a complete list of the design strategies defined by Katharina Hecht.

3.3 Ecosystem Services

Each sustainable design strategy generates an ecosystem service. [5] Ecosystem services are the benefits human derive from natural ecosystems such as clean air and water, pollination, and climate regulation. These services are essential for human well-being and are often undervalued or underappreciated. [17] According to the ecosystem service framework based on a revision of the work of the Millennium Ecosystem Assessment and Pedersen Zari, they can be divided into three different groups. [18–20] Appendix B contains a comprehensive table of all the ecosystem services and their respective groups.

- **Provisioning Services**
  These are the direct products obtained from ecosystems, such as food, timber, fiber, and medicinal plants.
- **Regulating Services**
  These are the benefits that ecosystems provide by regulating natural processes, such as climate regulation, water purification, and pollination.
- **Cultural Services**
  These services are non-material benefits that humans obtain from ecosystems, such as recreation, aesthetic enjoyment, and cultural and spiritual significance.

Although sustainable buildings can mitigate the impacts on the environment, urban form in itself cannot achieve fully socio-ecological sustainability. Sustainability is a biophysical design problem. [21] However, a design that mimics the functioning of ecosystems and utilizes synergies between mitigation and adaptation strategies in relation to climate change could be a beneficial long-term biomimetic built environment response to climate change in the context of built environment design. [19]

4 Data Understanding

This section contains a description of the dataset provided by Katharina, The data is explored through several visualizations to gain a better understanding of its structure. Data quality is also assessed to ensure that it is clean and suitable for analysis. The primary objective of this phase is to develop a comprehensive understanding of the data that will be used for subsequent analysis.

4.1 Describe and Explore

The data is formatted in a .xls file format (276kb) and contains several tables distributed over different sheets. A link to the dataset is provided in Appendix D. The first sheet, named "ES List" consists of two columns 'Group' and 'Categories'. There are three different kinds of groups as discussed in the previous section. Figure 2 displays the number of services assigned to each group. The distribution shows that Provisioning Services have the highest number of categories, and Cultural Services have the fewest. The complete list of all ecosystem services and their respective group can be found in Appendix B.

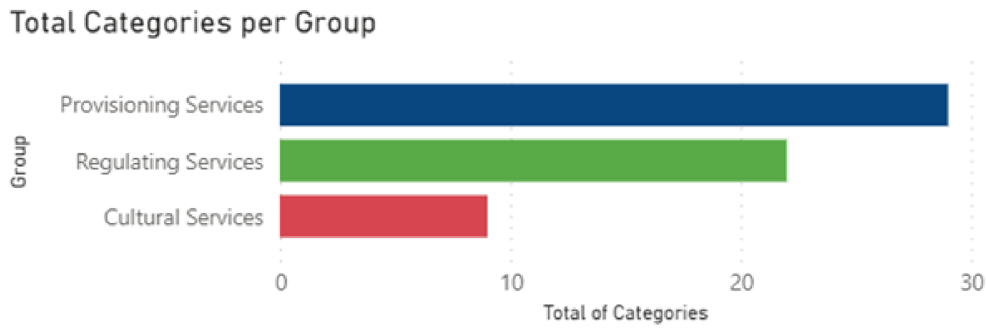


Figure 2: Categories per Ecosystem group

The second sheet, named "Database Design and ES" provides a list with design strategies (Appendix A) and the ecosystem services they generate. In the sheet, there are 187 design strategies, and each strategy has the potential to produce anywhere from 1 to 20 ecosystem services. These services are displayed in distinct columns.

The third sheet, "examples and designs," contains a list of example buildings along with the design strategies that have been implemented in them. Additional details such as the country, city, and street are used to provide the location of each building. In Figure 3, the buildings are shown on a map. The map and Figure 2 were created using Microsoft PowerBI. [22] Furthermore, the sheet also provides links to the website or article where each building is discussed, as well as a separate link to a photo of the building.

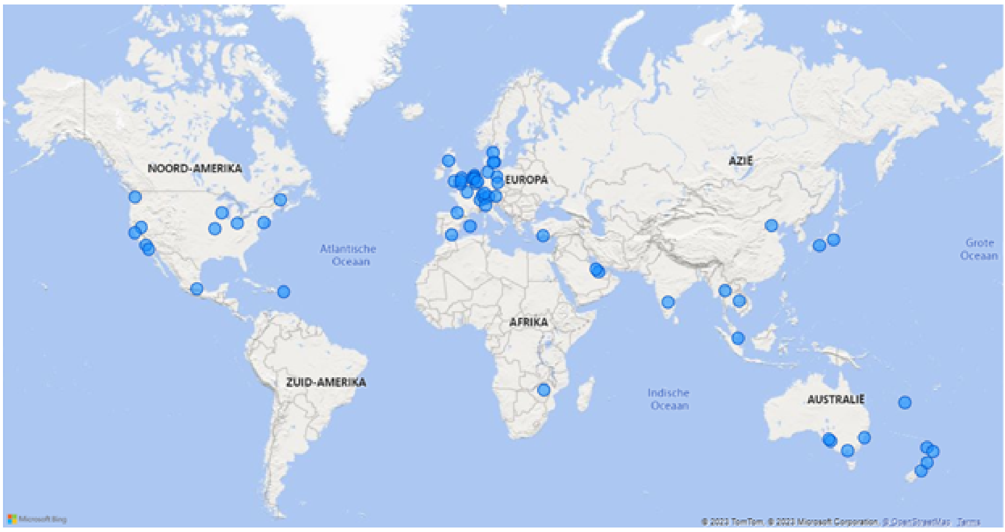


Figure 3: Map of Buildings

The visualization shows that a majority of the buildings in the dataset are situated in Europe, while North America and the Pacific Islands also have a decent number of buildings. In contrast, there are only a few buildings located in Africa and Asia, and none in South America.

In addition to the three sheets that were utilized as data input for the application, there is another sheet included in the dataset. This fourth sheet contains ecosystem indicators and their corresponding measuring methods. Furthermore, the dataset also includes several other sheets that contain minor notes or links to other data sources. However, due to the project's scope, this information was not included and left out of the analysis.

4.2 Verify Data Quality

The cleanliness of a dataset is a crucial factor that affects the quality of the analysis and insights drawn from it. In this dataset, some values were missing, particularly in the building's location (city/street), which could affect the accuracy of the results. Additionally, there were instances of controversial naming of design strategies and ecosystem services between different data sheets, which may cause confusion and inconsistencies in the interpretation of the data. Therefore, it is essential to ensure that the dataset is thoroughly checked for completeness and consistency before using it for any analysis or decision-making processes. The actual data cleaning process was conducted during the subsequent phase.

4.3 Knowledge Graph

The dataset was found to be quite cluttered and difficult to explore in its current state, with limited ability to establish connections between different tables. Due to the complexity of the data, alternative storage approaches were explored, and this led to the decision of using a graph database. Graph databases are particularly suitable for use cases that require high scalability and performance in handling interrelated data. Unlike relational databases, where data is stored in tables, graph databases store data using nodes to represent objects and edges to depict relationships between them. [23] This structure allows for faster performance, making it ideal for situations where speed is crucial. Additionally, graph databases are more flexible, enabling you to add or modify relationships between nodes without altering the entire database schema. Moreover, graph databases enable the execution of complex queries, revealing previously undiscovered relationships and patterns within your data. This capability leads to deeper insights and better decision-making. [24] Neo4j was selected as the preferred graph database management system, given its widespread use and reputation. [25]

4.4 Collect Additional Data Sources

During the data understanding phase, I also searched for additional data sources to supplement the existing data, which could potentially reveal new opportunities for analysis or research that were previously unavailable. Since the dataset was already linked to the climate classification map of Wladimir Köppen [26], I decided to include this data source as well, as the climate could explain why specific design strategies are used in certain locations and not in others. There was a text file available that contained latitude and longitude data, along with the climate zone in which each location falls. Initially, I considered adding this information as a property to each building node. However, I realized that even places within the same climate zone could have significant differences that could affect the effectiveness of a design strategy, such as cultural differences or variations in humidity, wind, and photovoltaic power. To address this, I began searching for datasets that contained this information on a global scale and stumbled upon maps that could potentially be incorporated into my project. I found the following interesting data sources:

- **Climate Zones**
  https://ggis.un-igrac.org/layers/igrac:other_climate_2007_koppen_geiger
- **Average Annual Relative Humidity**
  https://sage.nelson.wisc.edu/data-and-models/atlas-of-the-biosphere/mapping-the-biosphere/ecosystems/average-annual-relative-humidity/
- **Global Photovoltaic Power Potential**
  https://globalsolaratlas.info/
- **Mean Wind Power Density**
  https://globalwindatlas.info/

5 Data Preparation

The data preparation phase is a crucial step in the data analysis process that involves cleaning, constructing, and transforming raw data into a structured and formatted dataset suitable for analysis. Additionally, during this phase, the data was converted into a graph database. The data preparation phase is critical because it ensures that the data is accurate, consistent, and complete, and it sets the foundation for meaningful and reliable insights to be derived from the data analysis. [6]

5.1 Data Cleaning

In this phase, the data is first cleaned. There were some instances in which the value of the link was not an actual link but some sort of path to an image on a computer, these were cells were cleared. Furthermore, there were some building instances that had different empty property values. It was decided to remain these cells empty as they would need manual input as these cells characterized information about the buildings such as city, street, and a link to the building. Once the data was cleaned, it had to be transformed into a structured format suitable for importing it to Neo4j, luckily most of the data was already quite structured. However, some column names have been slightly adjusted for better comprehension of the data. After that, the three sheets that contained relevant information for this project scope were extracted and transformed into three separate CSV files using a Python script (Appendix E).

5.2 Data schema model

A data schema model has been created to visually represent the database architecture. The nodes in the model are represented by circles, while the relationships are indicated by arrows, similar to a graph database. The initial sketch of the data schema model is shown in Figure 4. It depicts four nodes and three relationships.

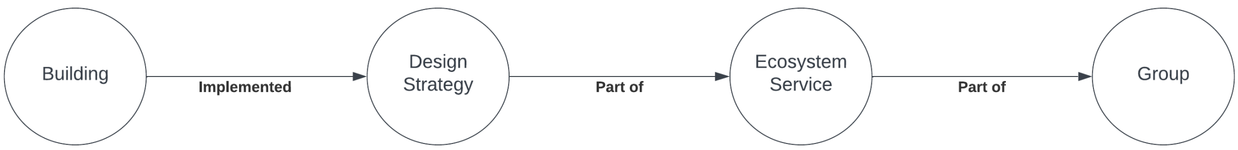


Figure 4: Draft data scheme model

The building was used as the basis (starting point) of this schema model, with the remaining elements added from left to right. Following a discussion with Katharina, it was decided that the group node could be excluded and incorporated as a property of the Ecosystem Services node. Furthermore, Katharina recommended renaming the relationship between design strategy and ecosystem services from 'Part of' to 'Generated', as it better reflects the intended meaning of the connection. This change was implemented accordingly. Figure 5 presents the completed data schema model.

Figure 5: Final data scheme model

While the data scheme model shown in Figure 5 only scratches the surface of the possibilities offered by this approach, the knowledge graph was still expanded to include a new data source even within the project's limited timeframe. More details regarding this expansion will be discussed in section 6.

5.3 Importing Data into Neo4j

Once the data schema had been modeled, the next step was to import the data into a graph database. To accomplish this, I had to familiarize myself with Cypher, a comprehensive query language that supports knowledge graph analysis in Neo4j. The task of importing the data into Neo4j was significant and it took me about four weeks to complete. During this time, I familiarized myself with knowledge graphs, Neo4j as a graph database, and the Cypher query language. Importing the data marked the completion of the first must-have requirement according to the MoSCoW principle (Table 1), which was a significant milestone in the project. To elaborate, the data was imported into Neo4j AuraDB, a fully managed cloud graph database service. [27] Neo4j AuraDB is used to securely and reliably store the data, allowing the project to focus more on developing graph applications and less on database administration.

Upon importing the data into Neo4j Aura and analyzing it in Neo4j Bloom, I observed that certain ecosystem services and design strategies lacked relationships, despite their presence in the Excel sheets. As discussed in Section 4, this was attributed to conflicting naming conventions. With the assistance of Katharina, this matter was rectified, resulting in an increase in the number of relationships from approximately 750 to over 1300. This extra data-cleaning process was crucial, as it prevented the loss of valuable insights.

It is important to remember that the project is using the free plan of Neo4j Aura, which means that the database will be paused if there is no activity for over 72 hours. To avoid this, it is necessary to regularly send a query to the database to keep it alive. This functionality has to be implemented in the application.

6 Expansion

This section outlines the expansion phase, which involved exploring and implementing potential machine learning and visualization techniques, as well as a process description of the development of the web application. Please find the link to the code along with installation instructions, user manual, and a data ingestion pipeline in Appendix D.

6.1 Exploring visualizations

The application would of course have a knowledge graph. However, upon realizing that adding the climate and environmental data as properties to individual buildings would only allow for viewing one building's data at a time within a knowledge graph (as discussed in subsection 4.4), I recognized the need for a more comprehensive solution. As a result, I decided to develop a map with various layers of information that the user could toggle between, including data on humidity, wind, photovoltaic power, and climate zones.

6.2 Development

With the aim of allowing users to explore the current data set and expand upon it, I began developing the idea of a knowledge graph that would enable users to add nodes and show relationships. Additionally, I envisioned a map on the side that would interactively display the nodes (with the label building) that were on the knowledge graph. This would allow users to search for a particular design strategy, view the buildings that have this strategy implemented, and see their locations at a glance. For instance, architects could easily determine the feasibility of implementing this design strategy based on the location of their project. The application is designed for desktop use only, and thus not mobile-friendly.

6.2.1 Start of development

The application was developed using React, an open-source JavaScript library, despite having no prior experience with either React or Javascript. It was a challenging undertaking as the application had to be built from scratch. I began by searching for informative tutorial videos on building React applications, which helped me grasp the key concepts and the React way of working. I followed a detailed video that explained these concepts very well and used its skeleton to build the initial structure of the application. With the skeleton in place, I created several different scenes for the application. The link to the tutorial video can be found in the Appendix.D. After noticing that the tutorial video contained scenes with data tables, I considered adding this feature to the application. I believed this would be a useful addition, as it provides a familiar way to view the data and serves as an additional visualization tool for the user. At this stage, my application consisted of a sidebar that allowed users to navigate between different scenes, and a light/dark mode switch. Although the light/dark mode switch was not a necessary feature for the application, its implementation provided me with an opportunity to gain further knowledge of React hooks like useState and useContext. These hooks have been utilized multiple times throughout the application to facilitate interactivity.

6.2.2 Base Map

After setting up the application structure, the next step was to develop the initial view, which was the map. To achieve this, the Leaflet library was utilized. The instructions from its website (leafletjs.com) on integrating the Leaflet library were straightforward and it didn't take much time to set up the base map. Furthermore, the map data was obtained through the OpenStreetMap API. After this, I inserted mock data into the application to create markers on the map. From there, I proceeded to style the markers and their associated pop-ups. Additionally, I implemented a marker clustering feature to improve the map's performance and visual clarity.

6.2.3 Neo4j driver

Since the application's core functionality revolves around analyzing and expanding data stored in Neo4j AuraDB, it was necessary to create a Neo4j driver component to enable communication with the database and execute queries. The neo4j-driver library was utilized for this task, which was straightforward to implement. To ensure that the driver component could be used in every scene within the application, it was necessary to implement it as a React hook using the useContext hook. This allowed for easy access and sharing of the driver component across different components within the application.

6.2.4 Forms

Once the Neo4j driver component was functional, attention was turned towards enabling users to modify the queries sent to the database. Given the importance of expanding the data, forms were created at the outset to allow users to add data. For the forms, Formik was utilized as it is a popular library that simplifies form management. The forms were designed with input fields to represent each property of a node, while for the relationships of the node, a multi-select dropdown was included to enable users to easily add related nodes to the nodes being added. In order to maintain the accuracy of available options in the multi-select, they were fetched from the database every time the form scene was loaded. This feature was implemented to address the issue of users potentially not knowing all the possible options beforehand, which could make filling in a text field difficult and lead to errors. Once the data entry forms were functional, forms for deleting data were also implemented. The addition of deleting forms was a desired feature expressed by Katharina during the expert study. This is especially useful in cases where data with incorrect values is accidentally added, as it can be easily removed. To prevent the accidental deletion of multiple nodes and relationships, the select dropdown used for deleting nodes was implemented with the multi-property set to false. This means that only one node can be deleted per executed form. However, the select dropdown type itself remained the same. Since there are only three types of nodes, it was determined that three distinct form scenes would be created, one for each node type. Each scene includes two forms: a green form for adding nodes and a red form for deleting nodes.

6.2.5 Knowledge Graph

The development of the knowledge graph is what took most of the time to develop. As I spent time experimenting with different implementations for visualizations. The process of retrieving data from the Neo4j driver varied for each implementation. Initially, the D3.js library was used, which successfully displayed multiple nodes based on the query. However, customizing the visualization and adding interactive features proved to be challenging. Consequently, the neovis.js library was discovered, which offered simpler styling options and preconfigured click-events functionalities. Despite spending over 16 hours configuring the connection to neo4j auraDB and displaying the visualization, many of the expected styling and interactive features were not functional, leading to feelings of discouragement. As a last resort, the vis.js library was explored, particularly the vis network visualization, which the neovis.js library was built upon. While it required building all the features from scratch, with some persistence, nodes were successfully rendered on the screen. Implementing relationships and additional nodes presented further difficulties, requiring the creation of a useState variable to manage nodes and relationships, ensuring no duplication or errors in the knowledge graph display. Fortunately, the knowledge graph was ultimately functional, with a form enabling query changes to be sent to the database. Clicking on a node displayed a tooltip revealing the node's properties and values, while keybindings were implemented to display relationships from or to a selected node.

Each type of node has been assigned a specific color to represent it. This has been done to improve the readability of the knowledge graph. Building nodes are colored orange, design strategies are colored yellow, and ecosystem services are colored either red (for cultural services), blue (for provisioning services), or green (for regulating services).

6.2.6 Map layers and markers

Despite the data sources described in subsection 4.4 appearing straightforward to add, it turned out to be more difficult than anticipated. As the photovoltaic power data was only available as a Geotiff file, which is not supported as a map layer in Leaflet. Therefore, converting it to GeoJSON was necessary. Despite experimenting with the rasterio and GDAL Python libraries, these attempts were unsuccessful. After several days of searching for alternative solutions to visualize the data on the map, the climate zones layer was added instead. Fortunately, this data source was available in GeoJSON format, albeit in an outdated version, which required some small adjustments before it could be added to the map as a layer. Retrieving building data from Neo4j AuraDB and displaying it on the map proved to be a challenging task. Firstly, latitude and longitude information had to be assigned to each building node. To accomplish this, the neo4j spatial values known as "points" were utilized as they were specifically designed to store geometrical data. The nominatim.OpenStreetMap API was employed to determine the geographical location of each building node based on the country, city, and street properties. However, some buildings were missing information, such as city or street, which resulted in some buildings sharing the same location. While not an ideal scenario, mainly due to missing values, this issue was deemed acceptable. To prevent this problem from growing bigger, I made sure that each important property related to the location was marked as required in the data input forms. Adjusting the data obtained from executing a query that returns all buildings and formatting it correctly for both the markers and popups to display on the map proved to be somewhat challenging. To prevent links from interfering with the styling of the pop-up, any values starting with 'http' were transformed into a href with the value as the URL and the text displayed as a link. This was necessary as some links were quite lengthy.

6.2.7 Multi-view

With both the knowledge graph and map functioning properly, the next step was to develop the multi-view feature. This feature would include the existing knowledge graph, accompanied by a map displaying markers that change based on the building nodes visualized in the knowledge graph. In order for the interactivity to function, the useState variable that stored the graph's nodes and edges had to be replaced with a useContext hook, allowing for easy input of this information into the interactive map. Additionally, the getInteractiveMarkers component was created to retrieve all building data and filter the nodes based on the useContext hook. By incorporating the multi-view feature, the knowledge graph is enriched with an extra data source, allowing the climate zone for every building in the graph to be viewed on the map.

6.2.8 Architecture

A simplified architecture of the application is shown in Figure 6. It consists of the main interactions of the user and the different scenes that are handled by the application. Furthermore, the application itself uses the Neo4j driver component to connect to the Neo4j AuraDB. The primary libraries utilized are displayed in italics and accompanied by a small image.

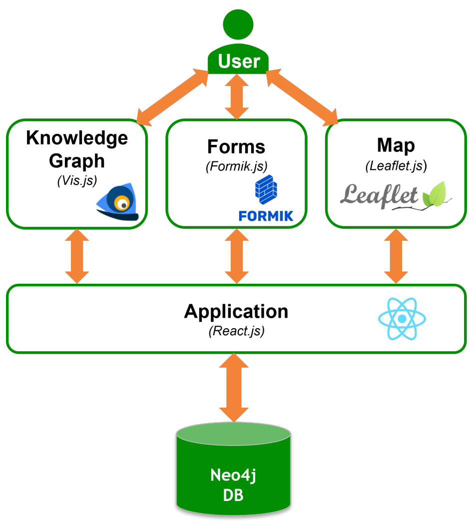
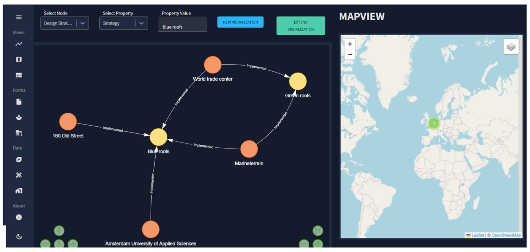


Figure 6: Application Architecture

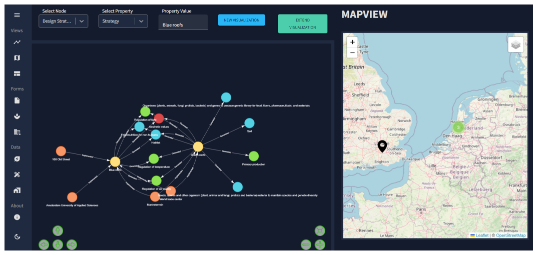
7 Evaluation

7.1 Case study evaluation

There are endless insights to be gained from the application. One way to explore the data is to add specific design strategies and extend the nodes with buildings that have implemented them. Figure 7a displays two added design strategies, Blue and Green roofs, along with the four buildings that have implemented one or both of them. Two buildings have both strategies implemented, while the other two have only blue roofs and no green roofs. The map on the right-hand side shows that these strategies are currently only used in Europe in the dataset. Zooming in (Figure 7b) reveals that three of the buildings are located in the Netherlands and one in London. Figure 7b also displays the generated ecosystem services. Notably, all the ecosystem services generated from blue roofs are also generated by green roofs, but green roofs generate four additional ecosystem services. This example demonstrates that a wealth of information can be obtained from the application with just a few clicks.



(a) Look up design strategy



(b) Extend with Ecosystem Services

Figure 7: Case study

Architects around the world are likely to explore the dataset in a different way. They would activate the climate zone layer on the map, and search for buildings in the same climate zone as their upcoming project. Let's say a Japanese architect is looking for sustainable solutions to integrate into his future project based in Tokyo. The climate in this region based on Köppen's classification is Cfa [26], which is the same climate as in Sydney and parts of China. They would then add these buildings to the knowledge graph and extend the nodes to observe the implemented design strategies (see figure 8).

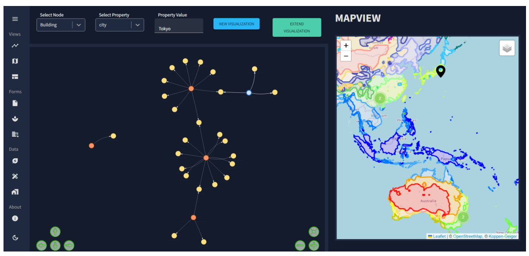


Figure 8: Add buildings with identical climate zones to graph

This method enables architects to identify popular (having many relationships to buildings) and less popular strategies at a glance, allowing them to choose the appropriate design strategy for their building project. In addition, architects can also look for a specific ecosystem service in the graph and discover design strategies that would generate that service.

During the case study analysis, it occurred that it would have been more convenient if the climate zones were incorporated as a property of the buildings. This would eliminate the need for the user to search for the climate zones themselves, making the process even more straightforward.

7.2 Expert study on application

In the expert study conducted with Katharina Hecht, I demonstrated the application and received positive feedback on what I had delivered. Ms. Hecht was quite impressed with the application's features and functionalities. However, during the demo, we also discussed some additional features that would be beneficial to have in the final release of the application.

One example was the need for a way to delete data from the database. This feature was not available at that time, but it was recognized as an important feature that needed to be implemented in the future. Additionally, for the graph visualization, we discussed the need to have the option to choose which relationship type of a node will be added to the graph. At that time, the only option was a double click event that extended the node for each relationship type. In the final version of the application, two keybindings were added, one for relationships to the node and one for relationships from the node. Furthermore, the dropdowns in the graph view could have been made interactive to display only the properties available for the selected node in the first dropdown. Moreover, the property value field could be transformed into a dropdown menu, allowing users to search for possible property values within the options provided, rather than manually entering a value that may not match any of the existing nodes.

Due to the limited time available for the project at the time of the expert study, some features could not be implemented. However, these features will be discussed in the limitation and future work section of the report. Overall, the expert study was helpful in identifying areas where improvements could be made to the application and allowed me to receive valuable feedback from a subject matter expert.

7.3 Project Evaluation

I am pleased to say that the project has been completed successfully, and the must-haves that were set at the beginning have been achieved. When taking a second look at the MoSCoW prioritization made during the project setup, most of the should- and could-haves have been completed as well (see Figure 2). Apart from the won't-have goal, the only unfulfilled objective was the application of machine learning algorithms. However, the potential implications of this have been thoroughly considered and will be further elaborated in the section on future work. Furthermore, the planning made at the start of the project was well-structured and thorough, and I am impressed by the level of commitment shown in sticking to the original plan throughout the development process.

Must	Convert existing dataset into a graph db dataset	✓
	Method: Visual Data exploration including a Data Science pipeline.	✓
	Evaluation: Case-study evaluation (some valuable insights); Description of insights based on visualizations	✓
Should	Merging more (1-3) datasets	✓
	Apply machine learning algorithms	✗
	Multiple visualizations on dataset	✓
	Data ingestion pipeline	✓
Could	Interactive Dashboard	✓
	Data ingestion frontend	✓
	Expert study (1 expert)	✓
Won't	Technical evaluation metrics (speed, throughput, etc)	✗

Table 2: Review of MoSCoW

One of the most significant achievements of the project was gaining knowledge on the topic of sustainable design strategies, knowledge graphs, and in particular, Neo4j and its query language called Cypher. Additionally, I have learned how to use Git to push changes to a GitLab repository and work with React, Javascript, and several libraries, which were essential for the development of the project.

However, if I could start over, I would have taken more time for the development of the application. The planning only allotted two weeks for development, but the actual development process took over a month, which was quite stressful in the final stages. The main reason for this was that developing such a large-scale application was not initially anticipated at the outset of the project. I also feel that I could have made some sketches and mockups before coding the application, along with a clear list of features to be implemented in the application. Starting from nothing but an idea made the development process more challenging, and it would have been helpful to have a more detailed plan in place.

Despite these challenges, I believe that the overall project went well, and the final product meets the objectives set out in the initial plan. In conclusion, the project was a valuable learning experience, and I look forward to applying the knowledge and skills gained in future projects.

8 Deployment

The last phase of the project is also the briefest one. To deploy the application, a server was required. Due to Michael Behrisch's illness during the final phase of my project, it was challenging to deploy the application on his server as originally planned. Hence, I had to look for another solution. I searched for web hosting providers and purchased a domain (charlesklijnman.nl). Before uploading the application to the server, I created a build folder using the following command.

```
npm run build
```

This build folder contained all the necessary files and assets for the application to run. Uploading the build folder to the server using FileZilla enabled the application to be hosted and accessible online. There was an internal service error occurring when reloading a page, To resolve the error, a .htaccess file was added to the build folder. In case the application is no longer hosted on the server, please contact either the author of this paper or Michael Behrisch for further guidance, as it may potentially run on a different server in the future. Lastly, login credentials can be obtained from Appendix D.

9 Conclusion

The web application developed as the final output of the project successfully fulfills its objective of providing architects and building developers with insights into sustainable building design strategies and the associated ecosystem services. Most of the objectives outlined in the MoSCoW analysis have been successfully accomplished. The project comprehensively covers the domain of sustainable building design, and the provided dataset from Katharina Hecht has been effectively integrated into a graph database. The data has been enriched by incorporating climate zones and building locations from other sources. The application has facilitated visual exploration and evaluation of the dataset, leading to endless meaningful insights. The expert study evaluation revealed several potential features and improvements that could enhance the application, and most of these were successfully implemented. Additionally, the application offers a user-friendly interface for data input, as well as a feature to remove data from the dataset. Finally, the application has been deployed on a server, allowing users to independently explore the dataset and discover valuable insights.

9.1 Limitations

The application has certain limitations. Adding too many nodes and relationships to the knowledge graph can lead to a cluttered graph, making it difficult to interpret. As observed in the case study evaluation, including climate zones as a property to each building node could have facilitated finding relationships between buildings of the same climate zone. Furthermore, the form of the graph view, could be made even more user friendly, by making the dropdowns change on the selected options.

9.2 Future work

Throughout the application development process, several potential features have been identified that could further improve the application. These recommendations for future work include expanding the visualizations, such as additional map layers featuring information about photovoltaic power or wind speed. In terms of data processing, the dataset needs to be expanded to gain more insights, which can be achieved by manually adding more data. Additionally, a predictive algorithm could be incorporated to forecast potential design strategies based on the building's location. For instance, the algorithm could analyze buildings with similar climates, wind, and humidity to predict potential design strategies.

References

[1] A. M. Omer, "Energy, environment and sustainable development," *Renewable and sustainable energy reviews*, vol. 12, no. 9, pp. 2265–2300, 2008.

[2] U. S. Building and C. Initiative, *The Kyoto protocol, the clean development mechanism, and the building and construction sector: A report for the UNEP sustainable buildings and construction initiative.* UNEP/Earthprint, 2008.

[3] B. J. Cardinale, J. E. Duffy, A. Gonzalez, D. U. Hooper, C. Perrings, P. Venail, A. Narwani, G. M. Mace, D. Tilman, D. A. Wardle, *et al.*, "Biodiversity loss and its impact on humanity," *Nature*, vol. 486, no. 7401, pp. 59–67, 2012.

[4] M. E. Assessment, *Ecosystems and human well-being: a framework for assessment.* 2003.

[5] M. Pedersen Zari and K. Hecht, "Biomimicry for regenerative built environments: Mapping design strategies for producing ecosystem services," *Biomimetics*, vol. 5, no. 2, p. 18, 2020.

[6] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, vol. 1, pp. 29–39, Manchester, 2000.

[7] K. Waters, "Prioritization using moscow," *Agile Planning*, vol. 12, p. 31, 2009.

[8] N. Lechner, *Heating, cooling, lighting: Sustainable design methods for architects.* John wiley & sons, 2014.

[9] P. Nejat, F. Jomehzadeh, M. M. Taheri, M. Gohari, and M. Z. A. Majid, "A global review of energy consumption, co2 emissions and policy in the residential sector (with an overview of the top ten co2 emitting countries)," *Renewable and sustainable energy reviews*, vol. 43, pp. 843–862, 2015.

[10] A. Chel and G. Kaushik, "Renewable energy technologies for sustainable development of energy efficient building," *Alexandria engineering journal*, vol. 57, no. 2, pp. 655–669, 2018.

[11] T. Ibn-Mohammed, R. Greenough, S. Taylor, L. Ozawa-Meida, and A. Acquaye, "Operational vs. embodied emissions in buildings—a review of current trends," *Energy and Buildings*, vol. 66, pp. 232–245, 2013.

[12] C. Zou, Q. Zhao, G. Zhang, and B. Xiong, "Energy revolution: From a fossil energy era to a new energy era," *Natural Gas Industry B*, vol. 3, no. 1, pp. 1–11, 2016.

[13] N. Kannan and D. Vakeesan, "Solar energy for future world:-a review," *Renewable and sustainable energy reviews*, vol. 62, pp. 1092–1105, 2016.

[14] N. H. Wong and C. Yu, "Study of green areas and urban heat island in a tropical city," *Habitat international*, vol. 29, no. 3, pp. 547–558, 2005.

[15] J. Yang, Q. Yu, and P. Gong, "Quantifying air pollution removal by green roofs in chicago," *Atmospheric environment*, vol. 42, no. 31, pp. 7266–7273, 2008.

[16] L. Y. Astee and N. T. Kishnani, "Building integrated agriculture: Utilising rooftops for sustainable food crop cultivation in singapore," *Journal of Green Building*, vol. 5, no. 2, pp. 105–113, 2010.

[17] M. E. Assessment, "Millennium ecosystem assessment," 2001.

[18] M. E. Assessment, *Ecosystems and human well-being: wetlands and water.* World Resources Institute, 2005.

[19] M. P. Zari, *Regenerative urban design and ecosystem biomimicry.* Routledge, 2018.

[20] R. Haines-Young and M. Potschin-Young, "Revision of the common international classification for ecosystem services (cices v5. 1): a policy brief," *One Ecosystem*, vol. 3, p. e27108, 2018.

[21] J. Birkeland, *Net-positive design and sustainable urban development.* Routledge, 2020.

[22] "Microsoft power bi," *https://powerbi.microsoft.com/*.

[23] S. Batra and C. Tyagi, "Comparative analysis of relational and graph databases," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 2, no. 2, pp. 509–512, 2012.

[24] I. Robinson, J. Webber, and E. Eifrem, *Graph databases: new opportunities for connected data.* " O'Reilly Media, Inc.", 2015.

[25] A. Vukotic, N. Watt, T. Abedrabbo, D. Fox, and J. Partner, *Neo4j in action*, vol. 22. Manning Shelter Island, 2015.

[26] M. Kottek, J. Grieser, C. Beck, B. Rudolf, and F. Rubel, "World map of the köppen-geiger climate classification updated," 2006.

[27] "Neo4j," *https://neo4j.com/*.

Appendices

Appendix A - List of Design Strategies

This list has been obtained by the dataset provided by Katharina Hecht.

Aeroponic systems
Agroforestry and agrosilvopasture
Airdrop irrigation systems
Aquaponic systems
Beehives
Biofilters
Biomass energy collection
Biophilic design
Bioremediation
Black/greywater treatment
Green roofs
Blue roofs
Botanical garden
Burning hydrogen
Cisterns
Coated glass
Community gardens
Compact construction
Compaction reduction
Companion planting
Composting
Cradle to cradle design
Cross ventilation
Delayed drainage systems
Design for deconstruction
Downspout strategies
Drip irrigation systems
Dynamic/aerodynamic architecture
Ecological analogue design
Green facades
Elevated constructions
Extensive green roof
Façade cladding
Fit form to function
Floating contructions
Fog harvesting
Food forests
Green corridors
Green façade
Green wall system
Greenhouse agriculture
Heat-cold storage installation
Hedgerows consisting of edible plants
Herb gardens
High albedo materials
Hydro turbines
Hydrogen burning system
Hydronic radiant system
Hydroponics system
Increase of native plants
Increase species diversity
Indoor (food) gardens
Infiltration systems for excess rain water
Insect attracting structures
Insulation material
Integration of retaining walls
Increased biodiversity
Intensive green roof
Intercropping
Interlocking panels
Introduction of keystone species in property
Keep open/pervious surfaces
Leakage control
Light filtering for temperature keeping
Living Plant Constructions (Baubotanik)
Location planning with little sun and wind
Low albedo materials
Medicinal gardens
Micro livestock
Microbial fuel cells
Moss wall
Noise reducing technologies
Organic farming
Orientation sound barriers
Permaculture
Permeable paving system
Photovoltaic cells integration
Photovoltaic panels
Pile foundations
Placed on columns or stilts
Planter green wall
Pocket garden/park
Private gardens
Rainwater Cooling System
Rainwater harvesting systems
Reduction of light pollution
Replace equipment or install water saving devices
Retaining wall system
Revegetation
Rivers or streams, including re-meandering, reopening Blue corridors
Roof ponds
Sacrificial ground floors
Seismic architecture
Semi-intensive green roof

Separation of waste streams
Smart irrigation
Smart roofs
Soakwells
Soil amendment
Soil filtration
Soil regeneration
Soil washing
Solar water heaters
Solid walls
Source separation of wastewater
Stack ventilation
Stepping stone habitats
Structural bracing strategies
Subsistence farming
Sunscreens
Sustainably sourced materials
Texture and form based sound barriers
Thermal desorption
Thermal energy storage system
Topsoil management
Trellis/fence farms
Urban mining
Urban orchards
Urban (rooftop) farming
Use of biodegradable materials
Use of carbon/GHG storing building materials
Use of daylight
Use of locally sourced materials
Use of mulch
Use of pre-existing vegetation
Use of readily available materials
Use of recycled materials
Use of traditional techniques
Vegetable gardens
Vegetated grid pave
Vegetated pergola
Vegetation cover increase
Vegetation for insulation
Vertical farming
Vertical mobile garden
Waste management
Water cooling systems
Water efficient systems
Water less systems
Water source heat pump
Water storage
Wind barriers
Wind turbines
Xeriscaping

Table 3: List of Design Strategies

Appendix B - Table of Ecosystem Service

Provisioning Services	Regulating Services	Cultural Services
Food/nutrition for humans	Regulation of temperature	Cultural diversity
Food/nutrition for non-humans	Regulation of flood events	Recreation
Biochemicals for medicine and pesticides	Protection against harmful cosmic radiation	Cultural heritage and historical values
Biological fertilizers	Regulation of noise	Spiritual and artistic inspiration
Fresh water (<3,000 Mg/L TDS)	Pest and disease regulation	Educational values and knowledge
Brackish water (3,000-10,000 Mg/L TDS)	Control of invasive species and other natural hazards	Sense of place
Saline water (>10,000 Mg/L TDS)	Regulation of wind	Aesthetic values
Brine water (groundwater >35,000 Mg/L TDS)	Regulation of humidity, ventilation and transpiration	Relaxation and psychological wellbeing
Reclaimed and recycled water	Regulation of drought	Social relations
Fresh air	Regulation and attenuation of seismic activity	
Raw materials (timber, fiber, stone, minerals, ores)	Attenuation of erosion and mass movement	
Recycled materials	Fire protection and moderation	
Reused materials	Regulation of weathering processes	
Biomass energy	Regulation of water quality	
Active solar energy	Recycling of materials	
Wind energy	Regulation of air quality	
Fresh water as energy source	Regulation of soil quality	
Marine water as energy source	Regulation of smell	
Mineral substances as energy source	Pollination and seed dispersal	
Geothermal energy	Soil formation	
Passive solar energy	Regulation of biogeochemical cycles (nutrient cycling and storage)	
Hydrogen energy	Primary production	
Human body heat	Regulation of light	
Seeds, spores and other organism (plant, animal and fungi, protists and bacteria) material to maintain species and genetic diversity		
Organisms (plants, animals, fungi, protists, bacteria) and genes to produce genetic library for food, fibers, pharmaceuticals, and materials		
Decoration/adornment		
Habitat		
Soil		
Electrical heat		

Table 4: Ecosystem services per group

Appendix C - Gantt Chart of Project Planning

Link to Gantt chart

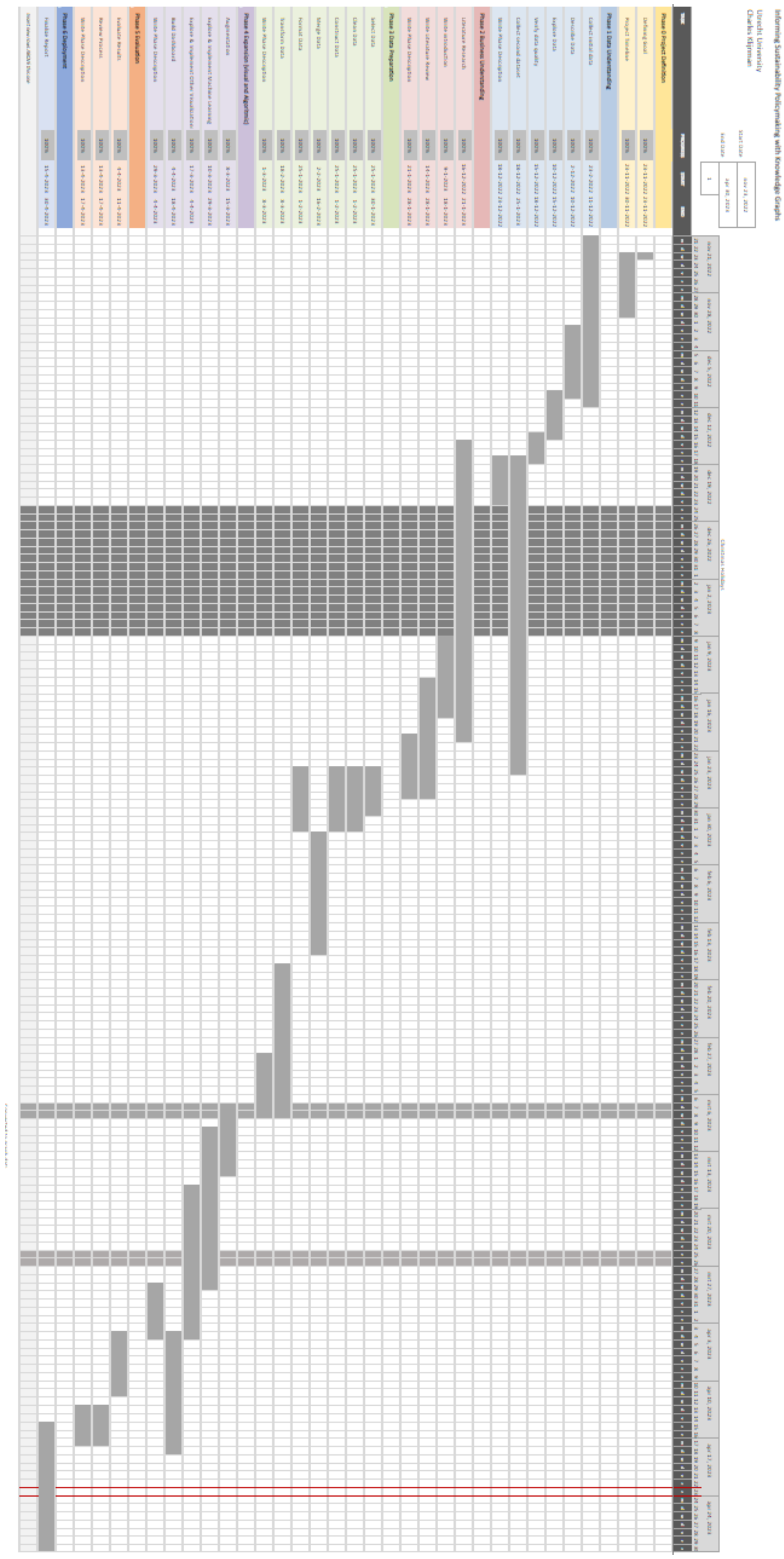


Figure 9: Gantt Chart

Appendix D - Link to code and web application

Credentials and link to web application:
Username: CharlesKlijnman
Password: 30-4-2023
Url: https//charlesklijnman.nl

Link to code on Gitlab:
https://git.science.uu.nl/vig/bscprojects/sustainability-graph-data-exploration

Link to main tutorial video:
https://www.youtube.com/watch?v=wYpCWwD1oz0t=2200sab_channel=EdRoh

Appendix E - Code fragments

Python code to transform Excel data into CSV:

```
import pandas as pd

data = pd.read_excel('./CleanedData.xlsx', sheet_name=None)

# loop through the sheets and save as csv
for sheet_name, df in data.items():
    df.to_csv(f'{sheet_name}.csv')
```