Project Rijkswaterstaat

TU Delft – CIEM6302 Advanced Data Science for Traffic and Transport Engineering

Group 2 Members:

Heisuke Miyoshi	5733693
Klaas Sicking	4948416
Martijn Stok	5070740
Martin van Andel	4577566
Yixin Sun	5715210

Project description

The project from Rijkswaterstaat is about improving the average response time of road inspectors. The main goal is to build a model for which the average response time is less than 18 minutes. This 18 minutes is important because this is the minimum requirement of the model. This can be achieved by optimizing the locations of road inspectors that are spread over the network. Something to take into account is the social impact, because the response time is average and some locations can be very far away from the closest road inspector and this cannot happen. This is the case the model needs to be changed even though the average is sufficient. What is also important to note is that the safety is improved when the response time is lower. Also the delays that arise from incidents can be reduced by a lower response time. Based on the information above a main research and multiple sub research questions are formulated.

Research questions

Main research question:

At which locations do the road inspectors need to be placed in order to optimize the response times?

Sub research questions:

1. What methodologies can be employed to address the primary research question effectively?
2. Are there geographical areas with a history of higher incident occurrence rates, and how should they be prioritized?
3. What is the optimal number of road inspectors required to meet response time objectives?
4. To what extent do date and time factors affect the deployment of road inspectors?
5. What ethical considerations should guide the deployment and operations of road inspectors?

Datasets used

- Csv file with incidents provides the location and the duration of incidents.
- Road network map shows the research area and provides the characteristics for each road.

Tech stack

The tech stack is split up into three categories. In the first category is the tech that is used in the first phase of the project. The second category are the methods that are tested to see if they can get an average response time of less than 18 minutes and this is part of the sub research questions. The last category is the programming that is needed for the methods and that includes the location of the repository.

Data preparation:

- Data Cleaning
- Visualization tools
- Python

Methods used:

- Path finding
- K-means Clustering (method gave insufficiënt results)

Programming:

- GitHub
- Python
- Gurobi
- Create subsections to reduce calculation time of notebook

Project backlog and priorities

In order to get a better idea of the different tasks that need to be completed for this project and the time it will take, a project backlog is provided here. For each of the different smaller tasks an estimation of the time necessary to complete it is given as well. Note that these times are a rough estimation and are subject to change because many of these tasks are new to us. If during the project new tasks come up or the estimated completion times change this project backlog will be updated.

Table 1: Priorities

Activity	Time (days)	Priority
Collecting the data used in the project	0.5	High
Cleaning the data	5	High
Visualizing the data	10	Middle

Creating a path finding model from the road inspectors to the incident location	10	High
Find the path travel time and average speed	2	High
Making an optimization model using the clustering method	15	Low
Making an optimization model using path finding method	15	High
Sensitivity analyses and calibration of the parameters of the optimisation	5	High
Validate the model with the other part of the dataset	4	High
Ethical consideration	1	Middle
Create a Dashboard for data story	7	Middle

Sprints

The following 8 tables are the 8 sprint backlogs.

Sprint 2	Data cleaning and Visualizing	Time	2023/09/12 to 2023/09/19 (one week)

Activity		Time	Priority	Response	Criteria
Data clean		1/2 day	High	Klaas	Remove all data from the table that does not conform to common sense or is irrelevant to the research.
Visualization	Heat map	2 days	Middle	Martin van Andel	Create a few templates of how a heatmap could look like.
	Bar chart	2 days	Middle	Sun Yixin	Bar charts can show the average number of incidents on weekdays at different hours.
	Heat map - COROP	2 days	Middle	Heisuke Miyoshi	The heat map can show the number of incidents per COROP zone
Work on the highway network		3 days	High	Klaas	Reduce the size of the highway network dataset and think how we can combine the incident data and the road network
Prepare the Machine learning(Clustering)		1 days	Low	Martijn Stok	Understand data and find a way to cluster the data

Sprint 3	Data cleaning and Visualizing	Time	2023/09/19 to 2023/09/26 (one week)

Activity		Time	Priority	Response	Criteria
Visualization	Heat map	2 days	Middle	Martin van Andel	The heatmap can show the amount of incidents happening on the roads.
	Bar chart	2 days	Middle	Sun Yixin	A concise interactive interface allows bar charts to be called up as needed instead of being displayed all at once.
	Heat map -	2 days	Middle	Heisuke	The heat map can show the

	COROP			Miyoshi	number of incidents per COROP zone
Select potential candidates for the optimization model		3 days	High	Klaas	Reduce the size of the highway network dataset
Writing code for clustering		1.5 days	Low	Martijn Stok	Write the full code for the clustering and make a figure of all incidents for each cluster

Sprint 4	Prepare the material for Midcheck	Time	2023/09/26 to 2023/09/29

Activity		Time	Priority	Response	Criteria
Visualization	Heat map	2 days	Middle	Martin van Andel	The heatmap is interactive with the option to select the type of incidents shown.
	Bar chart	2 hours	Middle	Sun Yixin	A clear explanation shows the bar chart and how it works
	Heat map - COROP	2 days	Middle	Heisuke Miyoshi	The heat map can show the number of incidents per COROP zone
Gather data from clustering		1.5 days	Low	Martijn Stok	Data gathered for each cluster with regard to number of incident points, average distances (with approximations) and explaining results
Ethics		1 day	Middle	Martin van Andel	A document with several considerations on multiple ethical aspects.
Integrate all the works into a single notebook and update the Github wiki		1 day	High	Sun Yixin	Create a notebook with ample explanations and considerations on ethics.
Path finding model integration		3 days	High	Klaas	finding shortest path between two arbitrary points on the highway network

Sprint 5	Optimization	Time	2023/10/02 to 2023/10/09

Activity	Time	Priority	Response	Criteria
Finalized code for clustering for centroids and starting code for medoids	1 days	Low	Martijn Stok	A working code of clustering with centroids and make a start with cod for medoids
Building optimisation model for inspectors' location	2 days	High	Martin van Andel	A base mathematical model which can determine the optimal locations of road inspectors
Working on the streamlit for result visualization	2 days	Middle	Sun Yixin	Having a deep understanding of Streamlit's features and applications
Connect the incident data to the highway network	3 days	High	Klaas	make sure that the incidents location can be used for path finding
Building optimisation model for inspectors' location	1 day	High	Heisuke Miyoshi	Obtaining optimal locations of road inspectors

Sprint 6	Optimization	Time	2023/10/10 to 2023/10/13

Activity	Time	Priority	Response	Criteria
Work on code for clustering with medoids	1 days	Low	Martijn Stok	Get a code that works for the clustering method with medoids
Update optimisation model and work on combining results	2 days	High	Martin van Andel	Correct optimisation model and create a way to combine optimised results
Creating Dashboard	3 days	Middle	Sun Yixin	Showing the visualization result in streamlit for the Data part
Data preparation and model building	4 days	High	Klaas	Deliver the data for the optimization model
Optimisation model update and visualisation of	1 day	High	Heisuke Miyoshi	Improving the optimisation model and visualising the optimal solution on maps

solutions				

Sprint 7	Validation	Time	2023/10/16 to 2023/10/20

Activity	Time	Priority	Response	Criteria
-	-	-	Martijn Stok	-
Improve and combine existing model	4 days	High	Martin van Andel	Combine the code for creating the cost matrix and the optimisation model and ensure that it can run in a reasonable time frame
Creating Dashboard and creating the weight matrix	2 days	Middle	Sun Yixin	Creating the weight matix for finding 120 inspectors location and preparing the streamlit for final presentation
Creating Cost matrix	4 days	High	Klaas	Deliver the data for the optimization model
Optimisation model update and visualisation of solutions	5 days	High	Heisuke Miyoshi	Improving the optimisation model and visualising the optimal solution on maps

Sprint 8	Validation	Time	2023/10/23 to 2023/10/27

Activity	Time	Priority	Response	Criteria
-	-	-	Martijn Stok	-
Train the model and prepare validation	3 days	High	Martin van Andel	Get the necessary training data and create code to validate the results
Creating Dashboard and find way to upload large file to Github	2 days	Middle	Sun Yixin	Create a new streamlit structure and use Github Desktop skillful

Sensitivity analysis	3 days	High	Klaas	Creating Sensitivity analysis model
Preparing for presentation	3 days	High	Heisuke Miyoshi	Complete preparing the slides and practice presentation

Sprint 9	Final check	Time	2023/10/30 to 2023/11/03

Activity	Time	Priority	Response	Criteria
Streamlit Introduction page and run 1 validation script	3 hours	Middle	Martijn Stok	Write the chapters of the introduction page and finalize backlog
Write streamlit explanations	2 days	Middle	Martin van Andel	Provide text and figures for the streamlit
Creating Dashboard and update Github Wiki	2 days	Middle	Sun Yixin	Complete the streamlit and updat the sprint into Github wiki
Get Sensitivity analysis results	1/2 day	High	Klaas	Combine all different results from the sensitivity analysis
Finalize all notebooks	3 days	High	Klaas	Combine all code files into proper notebooks
Creating Dashboard	5 days	Middle	Heisuke Miyoshi	Complete streamlit dashboard