

Midpoint Report

CS-410: Exploration of Data Science

Project Objective

The objective of this project is determining the malfunctional detector by analysis the data from highway dataset. By matching and ranking the error data, this project could provide the dashboard of the priority of the detectors need maintenance.

Project Approach

Our project approach has three steps goal-orientation approach: data analysis, establish infrastructure, query generation.

In the first phrase, data analysis phrase, the team will analysis the data and figure out the key variable of the dataset and find the best statistic method to determine the malfunctional detector and eliminate other possible error condition. In the second phrase, establish infrastructure, the team will cooperate and create the backend and frontend page for data analysis and user handling. In the third phrase, query generation, the team will cooperate and generate, implement, review the query that required for the dashboard function and show necessary data to support analysis result.

Team Structure

Team will working together, but for tasks beside major development, each member will take different smaller tasks to improve working efficiency.

Mingjue Wang: infrastructure development, data analysis, Response optimization.

Ronnie Song: infrastructure development, Query development.

Frank Sun: query development, data analysis, documentation.

Project Milestones

1. Dataset and data structure analysis, approach generation. – Done.
2. Infrastructure: backend development. – Done.
3. Infrastructure: frontend development.
 - a) API developing. – Done.
 - b) Frontend graph development. – In progress.
4. Query generation and review.
 - a) Query development. – Done.
 - b) Add more functions. – yes – Done.
 - i. Make frontend looks better – In Progress.
(frontend rework, display more data graph)
 - ii. More queries, totally about 10 queries. – In Progress.
(analysis more and display more)
Based on time range search speed by time range (greater, smaller, between)
Return top 20 result based on the speed range (greater, smaller, between)
Return station based on the station ID,
Return volume and travel time based on station ID and time
More basic info display
 - iii. Performance (if have free time) – Not Start Yet.
(if we have more space)
 - c) Dataset and file structure review. – In Progress.
5. Project review and summary.
 - a) Project evaluation. – Not Start Yet.
 - b) Presentation Prepare – Not Start Yet.