Stage 1: Detailed Project Description (Project Proposal)

1. Project Title

US Railway Safety and Operational Data Analysis Platform

2. Project Summary:

The United States boasts one of the largest freight railroad networks globally. According to the American Association of Railroads (AAR), freight railroads transport approximately 1.6 billion tons of raw materials and finished goods annually.

Beyond moving vast amounts of cargo, the railroad network also generates a wealth of data. The AAR and the Federal Railroad Administration (FRA) compile and publish several valuable datasets on their websites. For instance, the FRA releases data on railroad equipment accidents, encompassing all types of incidents across the US rail network where total monetary damages exceed a certain threshold. Additionally, traffic data for each railroad company is accessible for download from the FRA website. The Surface Transportation Board (STB) and AAR also provide publicly available data on railroad companies' operational details, including economic factors and shipment information.

The primary objective of this project is to integrate these publicly available datasets into an exhaustive database. This will enable users to examine railroad safety through various accident characteristics across the entire US network and by individual railroad companies, analyze railroad economic and operational factors, and compare safety or operational performance with different railroad companies. The main beneficiaries of our application include the railroad companies themselves, consumers requiring product shipment via rail, and researchers focused on railway safety and operations.

3. Application Description:

The envisioned application aims to provide a comprehensive platform for displaying and analyzing safety and operational data pertinent to the U.S. railroad industry, catering both to railroad companies and customers involved in goods transportation.

Enhancing the user experience, the application includes advanced search and filtering capabilities, allowing for precise data retrieval based on criteria like date range, location, type of accident, and railway company. It also boasts trend analysis and reporting tools that automatically generate reports and charts, highlighting safety or operational efficiency trends and facilitating comparative analyses among different railway companies.

This application seeks to fill a gap not adequately covered by existing resources, such as the Federal Railroad Administration's (FRA) website, by integrating and expanding upon datasets from the FRA, American Association of Railroads (AAR), and Surface Transportation Board (STB). The goal is to offer an unparalleled resource for in-depth railroad data analysis, thus addressing the need for a holistic view of the U.S. railroad sector's safety, operational efficiency, and financial standing.

4. Creative Component:

Since our database has information for specific accident locations, we can create a map view of the US and present accidents on it. Moreover, we can set a tooltip that shows information of accidents(time, location, etc.) when users hover their cursor over these locations. Users can switch between map view and table view.

5. Application Usefulness:

The purpose of this application is to show safety and operational information in detail for railroad companies and customers who plan to make goods for rail transportation in the US.

The basic functions of this application are to insert, delete, update, and search for information under various filters.

- Interactive Map View: An interactive map view that displays the locations of railway accidents, showing detailed information (such as time, location, type of accident, losses incurred, etc.) when users hover their cursor over specific spots. This not only provides an intuitive way to view data but also makes it easier for users to understand the geographical distribution of accidents. Users can switch between map view and table view which displays filtered information as a standard table.
- Advanced Search and Filtering Features: This allows users to search and filter based on multiple criteria (such as date range, location, type of accident, railway company, etc.), enabling them to find accidents or operational data that meet specific conditions.
- Trend Analysis and Reporting: Provide automatically generated trend analysis reports and charts to help users identify patterns and trends in safety or operational efficiency. For instance, it could show changes in the frequency of accidents over a specific period or comparative analyses between different railway companies. This could be achieved via stored procedures.
- We will have a user database that stores users' ID, password, user type(customers or company administrator), and their search history. Search history will store the result link after users' specific search and display the search word in the website's search bar.

The only partly similar website is the FRA official web which just has a dashboard on it that displays some railroad safety information. But besides the datasets generated from FRA, we also plan to include and combine other datasets from AAR and STB. Thus, this application should be unique in railway transportation.

6. Data Realness:

Four data sources are currently potentially considered in the project:

1: FRA Rail Equipment Accidents (REA) (6180.54): This dataset provides detailed information on railroad accidents. The raw datasets can be downloaded from https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx. It includes approximately 145 columns (degree 145) and around 2,000 rows for each year of data (cardinality ~2,000). Data cleaning will be conducted post-download. The team aims to utilize

accident data spanning from 2014-2023, totaling 10 years, with an overall dataset cardinality of around 20,000.

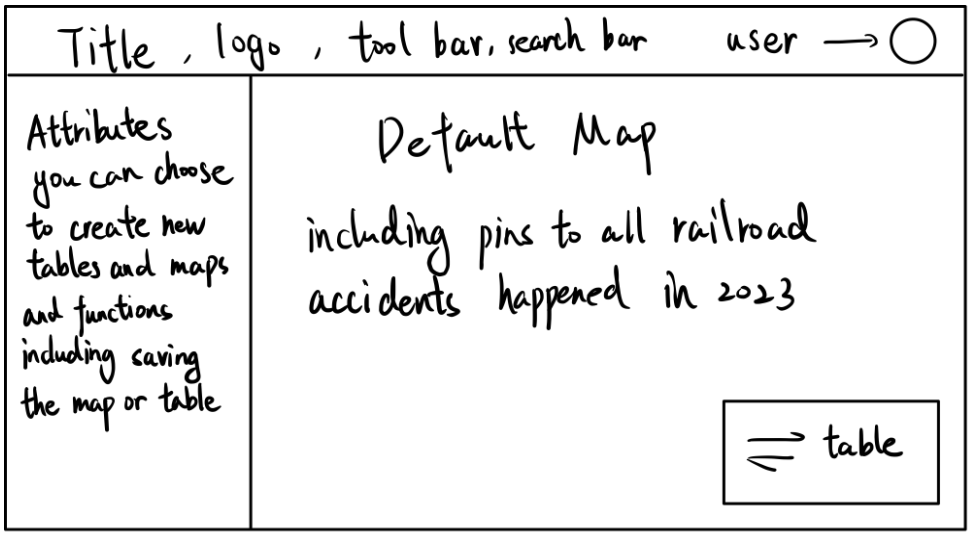
2: FRA Operational Data (6180.55): The raw datasets can be downloaded from https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx. This dataset contains comprehensive information on railroad traffic, featuring a degree of 26 and about 9,000 rows (cardinality ~9,000) for each year's data.

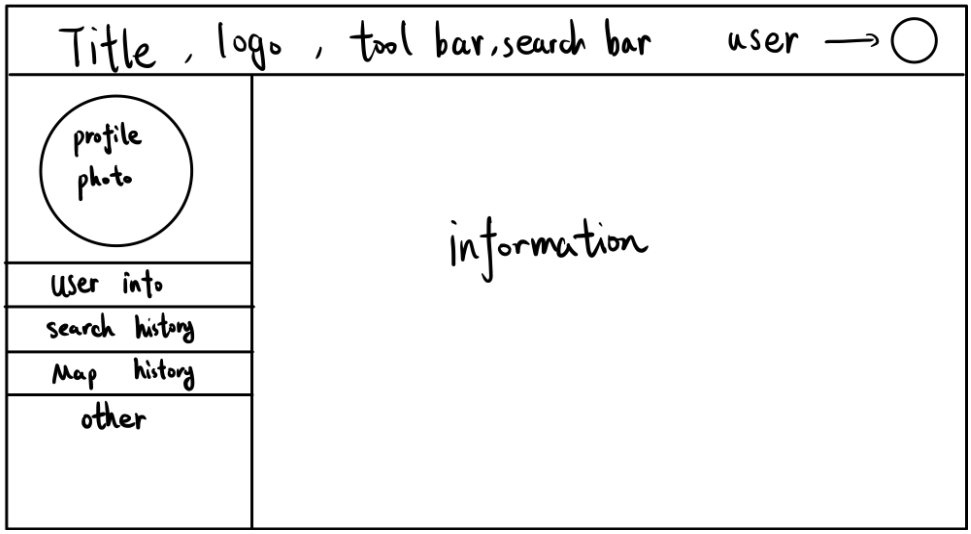
3: AAR Railroad Classification Table: This dataset provides classification details for each railroad company and will serve as a reference table. It has a degree of 9 and a cardinality of 539.

4: (Potential Use) STB Public Sample Waybill Data: The raw datasets can be downloaded from Surface Transportation Board (stb.gov). This includes detailed information for each shipment invoiced by a railroad company, with yearly data cardinality ranging between 400,000-700,000 and the raw dataset having a degree of 62.

Note: It's important to mention that some or all of the datasets described will undergo a data-cleaning process. The figures provided are based on the raw datasets.

7. **Website description and functionality that your offers.**





Project work distribution:

Web design: Aoyang Li, Zhuoang Tao, Xinhao Liu,, Haiyang Qin

Dataset collection: Xinhao Liu, Aoyang Li,, Zhuoang Tao, Haiyang Qin

Data cleaning: Xinhao Liu, Zhuoang Tao, Haiyang Qin, Aoyang Li

Report writing: Xinhao Liu, Haiyang Qin, Zhuoang Tao, Aoyang Li