Project Design Report

Team Lookout:

Daniel Brewer Ernesto Martinez Orion Crocker Seth Seeman Ebele Esimai

Project Goal:

Our project integrates car traffic, crash, and weather data, revealing relationships between them. The moderate scope and complexity of the project gives us students an excellent introduction to the major aspects of data visualization: project planning, data extraction and preprocessing, visualization, and presentation.

Project Summary:

The project focuses on the exploration of relationships between crash reports, traffic information and weather data on major highways around Portland, Oregon. We hope to identify locations, times, and conditions with a higher likelihood of slowdowns or crashes. This study is beneficial in diverse ways. For the transportation system as a whole, knowing peak periods and hotspots can impact resource allocation, planning, emergency services, safety and regulation. For individual drivers, this information could improve their safety or reduce their travel times. Also, incorporating the results of this analysis may improve the richness of the "time-to-travel" feed available for drivers by providing safety information.

We plan to use a combination of data for this project from sources: PORTAL for traffic information, ODOT for crash reports and traffic incident reports, and Dark Sky for weather information. Data preparation - cleaning, wrangling, and analysis - will be done using Pandas and some python scripts for portability. We'll use Tableau to find interesting relationships and visualizations.

Team Roles:

We will incorporate multiple team members into each task so we can learn together as we do our first data science project. We will designate a leader for each task that will be responsible for making sure the task is progressing in a timely manner and gets completed on time. Task Leaders are listed in the Task sheet.

Daniel Brewer will work as team coordinator. Orion Crocker will work on understanding our data sources and extracting the raw data. Seth Seeman will use his SQL expertise to join the raw data into a more searchable database. Ebele Esimai has the strongest ideas for visualizations on our team, and she will do analysis and reformatting of the SQL. Ernesto Martinez has the most Tableau experience of anyone on the team and will lead us in importing the final data and creating visualizations.

Task estimates:

- **Collect Data** (2 weeks)
    - At least 3 different sources, ~ 3 people
- **Data Sanitation** (2 weeks)
    - At least 3 different sources, ~ 3 people
- **Database Implementation** (2 weeks)
    - At least 3 different sources, ~ 3 people
- **Data Processing and Analysis** (3 weeks)
    - Everyone can query the data and make observations
- **Finalized Data Transfer** (1 week)
    - Everyone should export the data and import to tableau on their own system
- **Develop Visualizations** (3 weeks)
    - Everyone should learn to use tableau and create visualizations
- **Make Final Presentation** (2 weeks)
    - We will divide the presentation into sections that each team member will create slides for and present to the class - rough outline:
        - Introduction/Summary of Project and Task Overview
        - Data collection, processing
        - Analysis and Statistical Review
        - How we used the visualization tool (Tableau)
        - Visualizations and final conclusions