Presentation - Group 2

Final Project - Ride, Bike or Walk?

Selected Topic

We chose to analyze car, pedestrian, and bike traffic data collected in Ramsey County Minnesota from June 2015 to September 2018.

Why we selected this topic

We are interested in how weather, day of the week, time of year and holidays affected commuters' preferred method of transportation.

Description of Data Sources

Our car, pedestrian and bike data came from the Minnesota Department of Transportation.

Our weather data came from OpenWeatherMap.

The original datasets can be found at the links below:

https://archive.ics.uci.edu/ml/datasets/Metro+Interstate+Traffic+Volume#

https://www.dot.state.mn.us/bike-ped-counting/reports.html

Questions we hope to answer with the data

- How does commuter behavior change given the day of the week and time of year?
- How do weather conditions affect commuter behavior?
- Can we predict non-vehicle (bikers and walkers) traffic on a given day assuming vehicle traffic and weather conditions?

Data Exploration

Most of the data exploration and cleaning was done in SQL. We first dealt with duplicate datetime entries in the raw_vehicle_traffic table by dropping the textual description columns that did not provide any hard data, then averaging any disparate weather readings for a given datetime. We found that not all hours for a given holiday date had the the holiday listed, so we had to assure that the holidays were correctly labeled when aggregating on date.

In the raw_bike_pedestrian_traffic tabled, columns containing metadata such as type of bike path or technology used to track traffic were dropped. The dataset contained data for several counties in Minnesota, so we filtered to only Ramsey County where the vehicle traffic was recorded. The weather columns contained some null values, so we dropped those columns, since the vehicle traffic also contained weather data with no nulls. Each date in this table had an entry for pedestrian traffic and an entry for bike traffic, so we summed those values to get a total non-vehicle traffic value for each date.

Analysis Phase

Once the datasets had been cleaned and joined our team was able to produce initial visualizations using Seaborn. A variety of visualizations including bar charts, line graphs, scatter plots, swarm graphs, and linear regression plots, allowed us to identify outliers and decide what inputs were crucial and which ones just created noise. As a second step new visualizations were created in Tableau to include in the final dashboard as well as make some of the visualizations interactive.