

CS 488P/588 Data Management in the Cloud

Fall 2021 Quarter

Freeway: Data Set and Queries

This dataset uses a portion of the “Multimodal Test Data Set.” The Freeway Data is provided for a section of I-205 NB for a 2-month test period. Documentation will be posted on D2L in Content → Project:

- Freeway_Data_Description.pdf – an overview of the context
- Freeway_Data_Documentation.pdf – detailed documentation
- TravelTimeCalculation.pdf – how to calculate travel times

Data source

Link:

<https://drive.google.com/drive/folders/1drDpuALouiUxtWlOwqCRW5jyoqQfmc8b?usp=sharing>

You have to be coming from pdx.edu to access this link, so you will need to be on campus or VPN to the campus network to access it.

The data you will be using is the files: freeway_loopdata.csv, freeway_detectors.csv, freeway_stations.csv, highways.csv. (The “oneday” and “oneweek” versions are subsets of these files.) Note that the documentation mentions other data files; these other files are not used for this project.

Size: 700 MB

Analysis Questions

These questions are to be answered over the dataset.

- 1) *Count low speeds and high speeds:* Find the number of speeds < 5 mph and > 80 mph in the data set.
- 2) *Volume:* Find the total volume for the station Foster NB for Sept 15, 2011.

- 3) *Single-Day Station Travel Times*: Find travel time for station Foster NB for 5-minute intervals for Sept 15, 2011. Report travel time in seconds.
- 4) *Peak Period Travel Times (station)*: Find the average travel time for 7-9AM and 4-6PM on September 22, 2011 for station Foster NB. Report travel time in seconds. [Give an answer for each period.]
- 5) *Peak Period Travel Times (freeway)*: Find the average travel time for 7-9AM and 4-6PM on September 22, 2011 for the I-205 NB freeway. Report travel time in minutes. [Give an answer for each period. When computing a travel time over multiple segments, you need to compute the time per segment and add them up. Don't try to add up distances and divide into speed.]
- 6) *Route Finding*: Find a route from Johnson Creek to Columbia Blvd on I-205 NB using the upstream and downstream fields. [Your query should work in a database where there is data from multiple highways.]