**Portland’s Public Transportation – Applied Data Science Series**

I. Background

TriMet provides publicly available data for bus, rail, cab/lift, and WES vehicles. In this series, we will focus only on rail vehicles. The data is comprised mainly of 3 topics: vehicle, ridership, and performance. Monthly performance metrics are provided from 07/2008 to 03/2018. Annual ridership metrics are available from fiscal year 2000 to 2017 (TriMet’s fiscal year begins and ends in July). Vehicle data is provided for all rail vehicles for time frames spanning one hour, day, week, month, and 11 months. For a first run, we suggest loading and exploring the one hour data which is only 34KB. The longer time frames contain the exact same information (columns), only more rows. This will facilitate a quick set up and troubleshooting if needed.

Optional data sets are provided for correlation studies relating to 911 calls, housing, and weather.

II. Data Organization

The performance and ridership data each have their own csv file.

The vehicle data is broken into several relational tables. This significantly reduces the memory required to save the individual files. The events tables contains events logged by each vehicle such as location coordinates and delay info. The trips tables contain trip specific information such as route info. Trip data can be merged with events by using the tripID and vehicle\_id columns together. The messages table contains a dictionary between messageCode in the main table and the corresponding messages displayed on vehicle signs and the meta data applies to all data.

III. Data Sources

All data sources are provided below. Please use caution when downloading the raw vehicle data and heed the pop-up warning of how many rows you are attempting to query. One month of data can be roughly 30 million rows and 5-8GB in size! If multiple individuals attempt to download data of this size, you will crash the host server and that’s not nice! **Keep your queries small by using the filters provided in the web interface to limit the data to only what you need.**

1. TriMet vehicle data – <https://viewer.db4iot.com/trimet> (SEE THE WARNING ABOVE)

2. TriMet performance and ridership data – <https://trimet.org/about/performance.htm>

3. Housing data - <https://www.redfin.com/blog/data-center>

IV. Resources

1. Vehicle data column dictionary – <https://developer.trimet.org/ws_docs/vehicle_locations_ws.shtml>