

# Spring 2018 | Final Projects

## Data Mining in Public Transportation -- Take BART as an Example

Name: Libao Jin

Email: [ljin1@uwyo.edu](mailto:ljin1@uwyo.edu)

### Motivation/Abstract

In metropolitan areas, different kind of public transportsations such as buses, taxis, metros, etc., play an essential role in people's daily life. However, it is not as efficient as we expected sometimes. Hence, this project mainly looks into the public transportation, i.e., BART, to make some useful suggestion.

### Background

BART, short for "Bay Area Rapid Transit", is the transit system severing the San Francisco Bay Area in California. BART operates six routes, 46 stations, and 112 miles of track. It serves an average weekday ridership of 423,000 people, making it the fifth-busiest rapid transit system in the United States.

The dataset contains daily information on BART ridership for a period covering all of 2016 and part of 2017. Unlike some other rapid transit system datasets, this data includes movements between specific stations (there are just over 2000 station-to-station combinations).

### Project Summary & Major Tasks

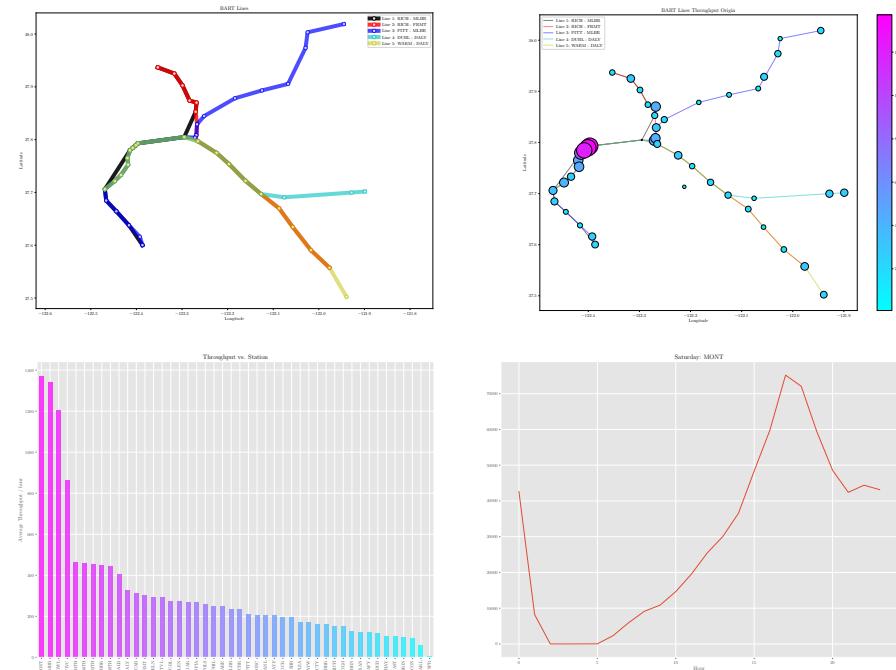
This project aims to answer the following questions:

- Which BART station is the busiest?
- What is the least popular BART route?
- When is the best time to go to SF if you want to find a seat?
- Which day of the week is the busiest?
- How many people take the BART late at night?
- Does the BART ever stop in a station without anyone going off and on?

### Limitations & Challenges

- The features of the datasets are too less to perform machine learning algorithms.
- Due to time limitation, unable to visualize routes on the geo basemap.

### Results



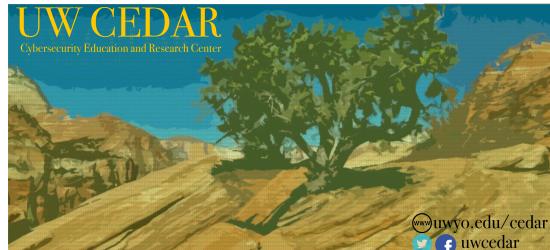
### Conclusions & Future Work

Conclusions:

- The busiest station is Montgomery St.
- The least popular BART route would be Richmond – Fremont.
- The best time to go to SF if you want to find a seat is around 11:00 AM.
- The busiest day of the week is Saturday.

Future Work:

- Use D3.js to build a interactive interface which enables us to understand the data better by the animation.



# UNIVERSITY OF WYOMING



This poster made possible by:  
UWYO College of Engineering & Applied Science, UWYO  
Computer Science Department, and UWYO CEDAR.

