CS 238 Project Proposal

Ingrid Fan - ifan@stanford.edu
Madison Hall - mhall38@stanford.edu
Anna Yang - ayang7@stanford.edu

October 5, 2018

Problem To Be Solved

The goal of our project is to determine the optimal public transportation paths in San Francisco in order to maximize city planning efficiency through the analysis of Uber data sets.

Uber is a ride-hailing giant that allows customers to book drivers using their own cars, fulfilling over 40 million rides monthly and capturing 77% of the US ride-hailing market. Uber has rapidly transformed the gig economy and has fundamentally changed the way that cities think about public transportation. Although Uber and other ride-sharing services do promote carpooling and are a step towards alleviating traffic congestion, many commuters are opting towards rideshare instead of using public transportation due to long commute times and inconveniently located stops/ stations. According to a study by the UC Davis Institute of Transportation Studies, public transit has seen a 6% decline in use by commuters in major cities after the development of ride-hailing companies such as Uber and Lyft (Clewlow)[1].

Specifically, we will be looking at the problem of traffic congestion in San Francisco and how we can use the public ride-hailing data from Uber to better improve existing public transportation infrastructure. Our findings could be extended to locating optimal locations for public transportation stops and routing optimal paths that get commuters from point a to point b in the shortest amount of time with the fewest number of connections.

Uncertainty

Our main uncertainty is the optimal paths for public transportation. We want to find the probability of each point on our map being the commuter's end-point and be able to optimize the likeliest paths based on the variables of the commuter's starting location and the time of the day. In summary, the uncertainty we want to reduce for each commuter trip a priori is: Given a starting point and the time of day, where is this person going and how can public transportation get this person there in the most efficient way?

Decisions

The decisions we will be making include where to place ideal public transit stops, which stops should connect to each other, and at what times public transit will be in the most use.

[1] Clewlow, Regina R. and Gouri S. Mishra (2017) Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States. Institute of Transportation Studies, University of California, Davis, Research Report UCD-ITS-RR-17-07

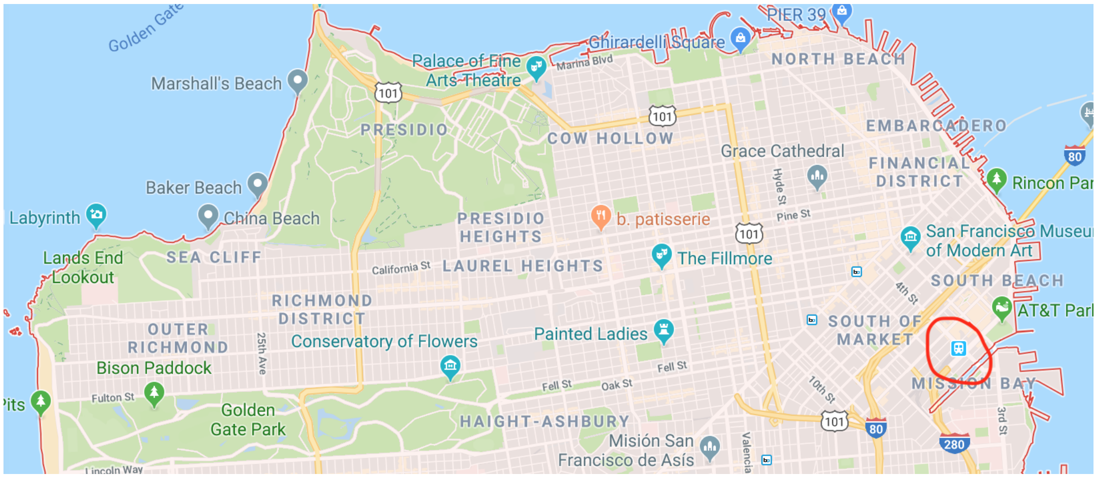

Figure 1: Proposed boundary of San Francisco to be covered by our model. The Caltrain station, a likely start and end trip point, is circled in red.

Possible Variables to Consider

1. Time & Day of the Week - the number of people trying to get from location to location varies based on time/day of the week. This affects the frequency of public transportation cycles. The time it takes to cover areas with heavier traffic congestion will also need to be considered.

2. Traffic Congestion - particular locations could be more difficult and could take more time to reach depending on congestion. These are the areas that are important to be covered by public transportation.

3. Boundaries - Besides drawing boundaries of the city to define which parts of the bay are covered by our public transit system, there are also key locations already in place that need special consideration, such as the Caltrain station.

Solution Approach

We will use logistic regression since we want to map coordinates (continuous values). In this context, coordinates means the coordinates where trips are started and where trips are ended.

Training for Optimal Public Transportation Paths

Using public Uber drop-off and pick-up data and public traffic data from Google and other map service companies, we want train our model to optimize public transportation paths. Given a starting point and a time, we want to determine what the most likely end point is.