Swiftly Data Analytics Project

OSCM 230

November 10, 2016

Due Date: The project is due on 12/2 to Knight Hall 455 by 5 pm.

What to turn in: A 8-10 page summary (not including code and these page limits are not strict) of your findings. You should express your findings using whatever plots, graphs and statistics you deem necessary. Attach all of your code as an Appendix to the write up.

1 Project Description

Swiftly has partnered with a major car retailer to study mobility patters. The ultimate goal is to understand how people make mobility choices in San Francisco when presented with a variety of travel mode options including : bus (transit), uber, walking, and biking. Your investigation should consider (but is not limited to) the following set of general questions:

- How does travel time and travel distance affect mode choice?
- How does the price of uber affect mode choice?
- How do incomes, population densities, and SLC scores affect mode choice?
- How does the time of day/day of week impact mode choice?

Visit **Swiftly Home(this is a link)** or **TechCrunch (link)** to get more information on the app. You could even download the app and play around with it here in Saint Louis (it works, but there are not real time arrivals). I would also highly recommend revisiting the Swiftly lecture.

2 The Data

See the video posted on Blackboard titled Data Explanation for a full description of all the .csv files and their respective columns. **Note that the number of rows in mode_chosen.csv is more than the number of rows in search_extra_info.csv file so be careful in combining this information.**

The reason for this is that we simply did not have data on these searches.

The Swiftly mobile application collects many important data attributes that can be used to better understand urban mobility. Similar to the way Google leverages online searches to determine trends and new products people desire, Swiftly can leverage trip searches to determine where people are traveling and the modes of transportation they tend to use.

At the core of this data set are the individual trip searches made by the Swiftly users. We define a trip search to be when a user enters a start and end location within the app. For each trip search, Swiftly stores an anonymized user ID, the origin and destination locations of the trip search, the travel options that were presented to the user (public transit, rideshare, biking, walking, etc.) along with the various attributes of each search (day and time, price, travel time, etc.), and the travel options that were clicked within the app. We restrict ourselves to searches that have at least one click, of which there over 50,000. In this way, we can study how users make choices when presented with a wide selection of travel modes with varying prices and travel times. This data set covers the following travel modes:

- Walking, biking
- Transit: all bay area agencies but primarily focuses on SFMTA Muni, BART, and AC Transit, and Caltrain.
- Rideshare: UberX, UberXL, UberSELECT, UberBLACK, UberACCESS, and UberTAXI.

Additionally, 80% of Swiftly users save their home and work locations in the app's shortcut feature. Using census data, we can match each user's home location to an income level (See next section for how to define high/low income). You might want to juxtapose the travel behaviors of high and low income users. Secondly, we are able to assign Smart Location Calculator (SLC) scores to the home locations of users and the origin location of each trip. The SLC score can be viewed as a proxy for the relative availability of public transportation options in a particular area. Using this metric, we partition the Bay Area into three neighborhood types: urban, semi-urban, and suburban and we study how mode choice varies across these three areas. We discuss the SLC score in more detail in the next section.

2.1 SLC Score and Income Neighborhood Classifications

At a high level, urban neighborhoods should be viewed as densely populated areas with a rich collection of public transportation options at the disposal of its residents. At the other end of the spectrum, suburban neighborhoods have low population densities and a fairly sparse array of public transportation options.

Finally, as the name suggests, semi-urban neighborhoods fall somewhere in the middle of the urban and suburban neighborhood types in terms of population density and availability of public transportation options.

To be more specific, we classify each neighborhood based on its SLC Score; a metric developed by the EPA which boils down the many feature that characterize a neighborhood into a single number. Underlying the SLC Score is a model that predicts the vehicle miles traveled by residents of a particular area using key demographic data such as population density and household income as well as factors that capture the availability of public transit such as the proximity to nearest bus stop and various standardized public transit accessibility scores. You can categorize neighborhoods as follows:

- Urban - SLC scores between 75 - 100, this group represents high density neighborhoods that typically have good access to public transit.
- Semi-urban - SLC scores between 60 - 75, this group represents medium density neighborhoods that typically have fair access to public transit.
- Suburban - SLC scores between 40 - 60, this group represents low density neighborhoods that typically have poor access to public transit.

The method for distinguishing high and low income neighborhoods was based on a recent memorandum by the California department of housing and community development. In this report, they define low income neighborhoods to be those that fall below 80% of the Area Median Income (AMI). Seeing that the AMI for the Bay Area is around $95-100K, we can define high income neighborhoods to be those with median household incomes above $75K and low income neighborhoods to be those with a median household income below $75K.

3 Grading

The projects will be graded using the following rubric

Analysis (10 pts)

- Are the observations insightful and interesting?
- Are plots and figures used affectively to convey the findings of the report?
- Did the report go above and beyond the recommended questions (bullet points in first section) in the analysis of data? I'm not looking for anything mind-blowing, but I want you to do a bit of additional exploration.

Writing (8 pts)

- Are the main findings of the report conveyed clearly?
- Is the report organized and presented cleanly?

Code (2 pts.)

- Is the code organized and commented?