Optimizing Coffee Shop Commutes

Vee Nguyen, Tung-Ho Kang, Christopher Trinh

BU CS 504

Introduction

Our idea revolved around coffee shops and transportation. Boston is a city full of busy commuters, resulting in an often hectic morning commute. We wanted to explore those commutes that revolves around stopping at a coffee shop to get one's morning caffeine before heading to the train station to continue their commute to their workplace -- a narrative that we all know much too well. Many commuters struggle to optimize their morning routine in order to both energize and prepare themselves while still managing to get to work on time, and we challenged ourselves to try to optimize this situation.

We decided to tackle this problem by optimizing the current placement of coffee shops, whether that involves moving existing ones or adding new ones, along the MBTA train lines. More specifically, we will retrieve data focused on just the Green Line to limit the scope of our calculations. By determining optimal locations of coffee shops based on MBTA train station locations, we hope to theoretically minimize travelling time in Boston during the morning commute for those who need to get from a coffee shop to the nearest train station, which would hopefully decrease both traffic and additional environmental pollution from motor vehicles.

Assumptions

The "peak morning commute times" that we mentioned earlier no doubt vary for many parts of Boston, but for the sake of convenience, we will be operating under the assumption that peak morning commute times are any time from 6am to noon.

In order to assign priority to parts of the Green Line, we will put more focus on stations that are more busy than others during peak morning commute times. We assumed that the more departures a station has, the more people waiting to get on at that stop, which would be an effective measure of how busy that station is.

Another important assumption we made is that commuters take Uber rides from the coffee shop to the nearest train station, so we could provide an estimate for how much Uber travel times could decrease if coffee shops were moved closer to the nearest train station using the k-means algorithm. In order to further optimize the locations of the coffee shops to lean towards fitting more busy train locations, we added a weight to each train station.

Methodology

Our first course of action was to identify which MBTA stations along the Green Line were most busy, in order to determine which stations should be closer to a coffee shop; we did this by assigning each station a weight based on number of daily departures from that station. We first get the list of arrival schedules for each station and aggregate all the arrivals for the station, an example data point would be ['South Station', lat, lon, 300], where 300 is the number of arrivals for the station, lat as latitude and lon as longitude. In order to feed in the correct weight for the k-means algorithm, one data point is generated if a station has 100 arrivals. Using the example above, ['South Station', lat, lon, 300] would generate 3 data points. Since k-means would move the clusters closer to a location where more data points are present, the existing coffee shop location would move towards the stations with more data points. The primary data sets we used are as follows:

- Boston coffee shops coordinates
- MBTA Green Line station coordinates
- 2018 Uber mean travel times

In order to calculate the new optimized placement of coffee shops along the Green Line, we had to do the following:

1. Calculate each coffee shop's distance from their nearest train station
2. Calculate the average distance (in meters) from a coffee shop to their nearest train station
3. Perform the k-means algorithm to calculate new coffee shop locations, and then pick the new closest location to each station
4. Calculate how much more efficient the new coffee shop locations are in terms of Uber mean travel time.

Data Sets

- *coffee.py* to get the coffee shops from the Yelp Fusion API.
- *stations.py* to get the MBTA Green Line train station location data along with their closest coffee shops from the MBTA API.
- *uber.py* to get the mean Uber travel times of 2018 sorted by hour of day, provided they fall under our "peak morning commute times", from the Uber Movement API.
- *getMBTA.py* to get data from the MBTA API and calculate each coffee shop's distance from their closest train station.
- *getMean.py* to get the average distance (in meters) that a coffee shop was from their closest train station.
- *getKMeans.py* to perform the k-means algorithm and calculate new locations of each of our coffee shops.
- *kmeansDist.py* to finds the closest distance for a station from the list of generated k-means and get the new distance for each coffee shop, averaging it over the number of stations.
- *savings.py* to calculate how much more efficient the new coffee shop locations are in terms of Uber travel times.
- We also wrote additional scripts to create GeoJSON files necessary to create our visualizations, which are in the project3/scripts folder in our subdirectory.

Database Documents

In order to get a sense of we utilized our database, some example database documents are listed below, with the rest of our database utilized similarly.

- getMBTA
```
{ "_id" : ObjectId("5cb7ee3d437e8170aebb705e"),
          "STATION" : "Englewood Avenue",
          "NAME" : "place-engav",
          "ID" : "70235",
          "ARRIVAL_COUNT" : "99",
          "LONGITUDE" : "42.336971",
          "LATITUDE" : "-71.14566" }
```
- kMeans
```
{ "_id" : ObjectId("5cb7edfb437e816ba452e277"),
```

```
"latitude" : -71.230728,
"longitude" : 42.325943 }
```

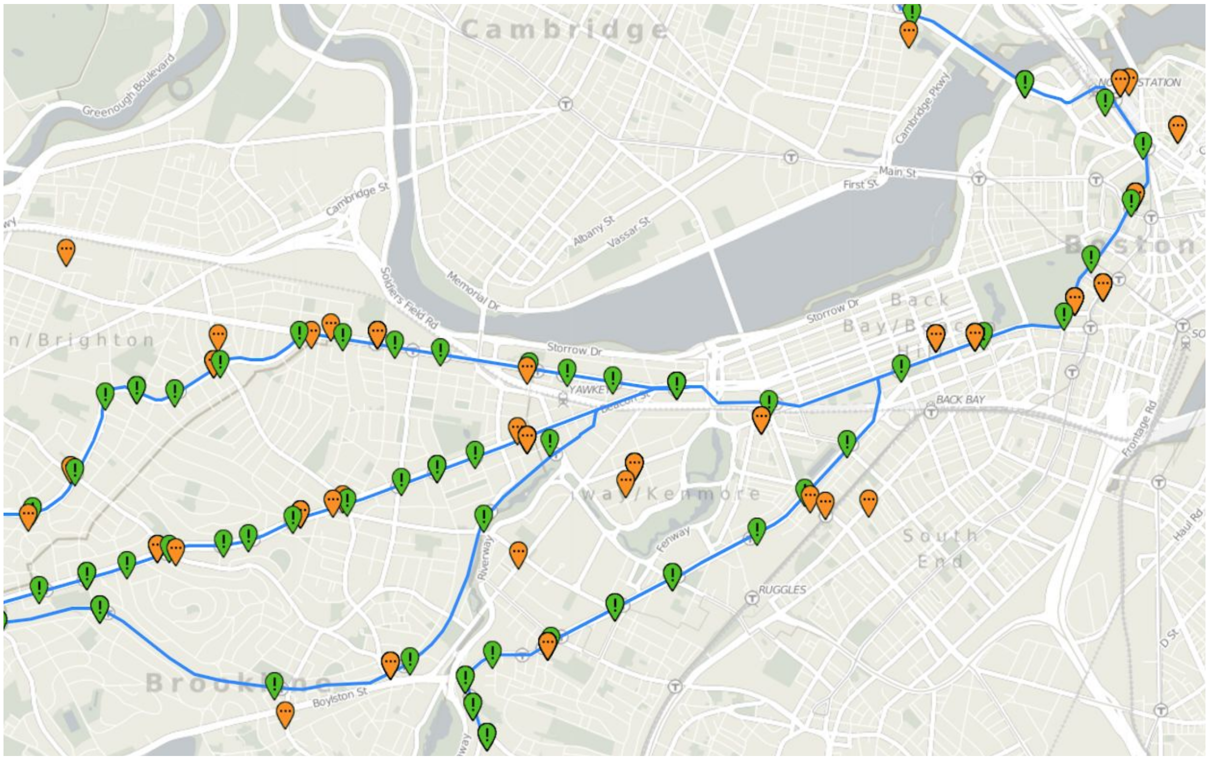

Figure 1. A map depicting the original coffee shop locations (orange) and the new coffee shop locations proposed by our project (green).

Challenges

We had a fair amount of difficulty obtaining information and performing our calculations and analysis, but the component we struggled the most with was obtaining our data from the MBTA. Due to how limited the MBTA API was for our purposes, we could only perform two separate queries that contained the information that we needed: (1) a query that listed information about every Green Line stop, including its coordinates and stop ID; (2) a query that gave the time and stop ID of every departure on a given day. However, we needed a data set that contained both the necessary stop information and also every departure from that same stop on a given day. We ended up joining both these queries together and extracting a combined data set, which extracted every departure that matched a stop ID and placed it with the coordinates of the corresponding stop information data entry.

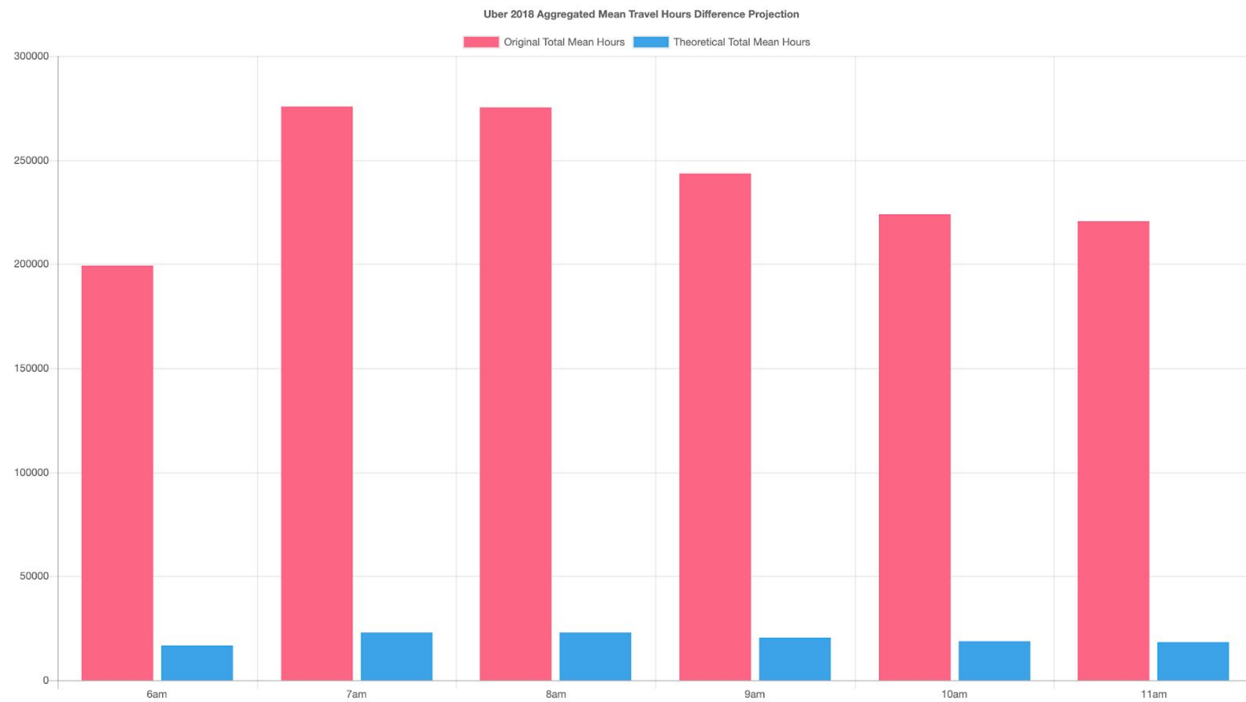

Figure 2. A bar graph depicting the original aggregate mean Uber hours versus the new theoretical mean Uber hours with the coffee shop relocations.

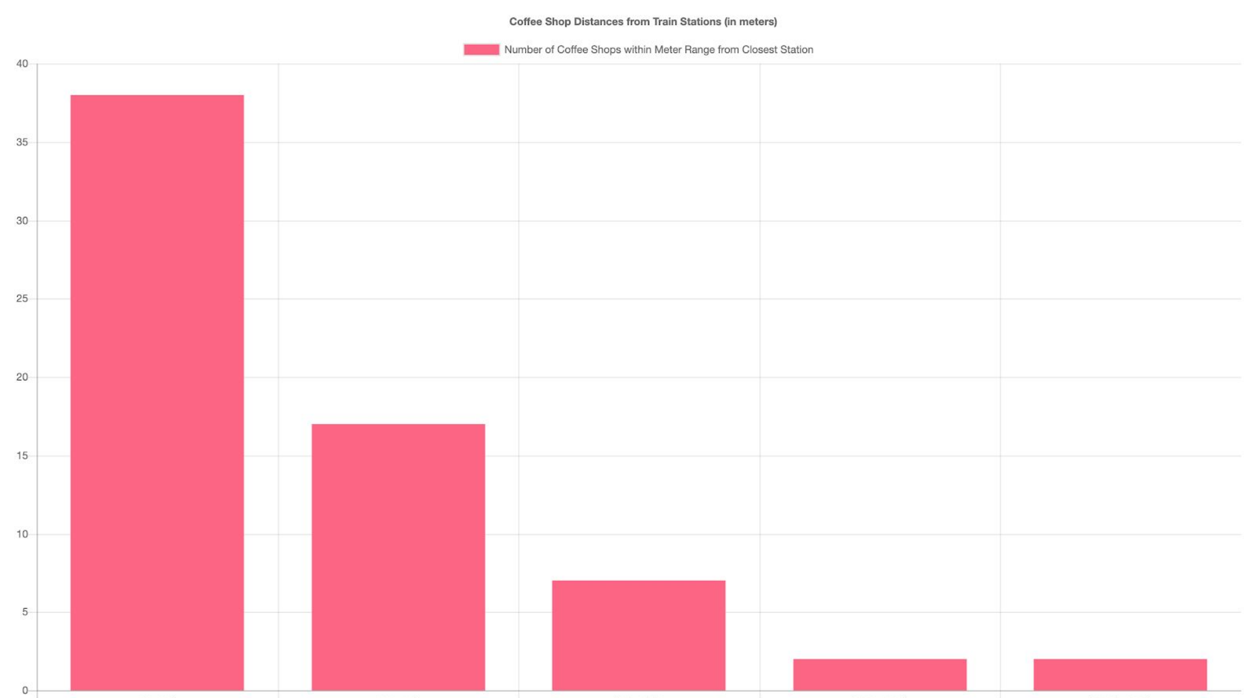

Figure 3. A histogram depicting the number of coffee shops that fell within a certain distance to the nearest train station, according to the bins on the X-axis.

Conclusion

After running k-means on the data set we generated, we calculated the average distance from the coffee shops to train stations. We then compared the result to the original average distance and found a significant 91.65% decrease in the average distance. To put this result in context, we gathered the aggregated Uber hourly mean travel time of a weekday in Boston. Again assuming that people travel from a coffee shop to the nearest train station using Uber, we can conclude that Uber cars would be driving 91.65% less during the morning weekday commute in Boston. During 2018, Uber cars spent a total of ~199,109 hours driving at 6-7am; with these proposed coffee shop locations, they would instead spend ~16,625 hours driving during 6-7am.

Future Improvements

In the future, we would like to have a more comprehensive data set. For example, it would be useful if we could find a data set that includes traveling intentions or more precise locations of people riding with Uber, as that would allow us to be more conclusive about how many hours people spend riding from a coffee shop to a train station via Uber, and thus how many hours would be saved with the proposed coffee shop locations. Our hope is to be able to gather enough data to be as precise as possible about when people travel to and from coffee shops and thus draw more accurate conclusions.