CS 506 Project
Analysing the mechanics of usage of Hubway Bike Sharing services in the Boston Area

Anish Iyengar, Shreenidhi Jayaram, Abhi Vora

Problem Formulation

The city of Boston is working towards the goal of zero carbon emissions. To achieve this goal, it has been encouraging the use of services such as Hubway. Every single usage of such an emission-free travel system makes Boston and its neighborhoods a cleaner and greener community making them environmentally sustainable for generations yet to come.

Despite the benefits of such a system, Hubway bike services are not readily available outside concentrated areas of the city of Boston. This is in stark contrast with the heavy demand of such services - bikesharing service usage is reported to reach about 28 million users in 2016 compared to a mere 4.5 million in 2012. Moreover, Hubway also provides with an income based subsidy for about 24% of bikeshare services which makes it even more appealing to use for lower-income households as well. However,what is it that's stopping the city of Boston from installing more Hubway docks? Will underserved populations benefit if such services were installed in lower-income areas? What about the frequency and duration of trips tell us about the demographics of such a usage?

To address these things, our overarching question is :

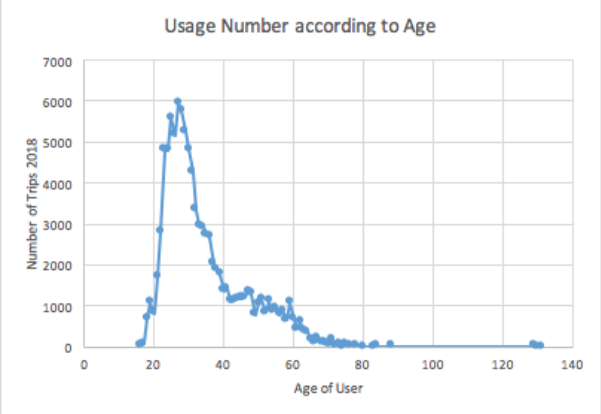
How does the demographic profile of users and their mechanics of usage contribute to the successful and purposefully resourceful use of the Hubway service?

Data Analysis

Part 1: User Demographics - Age and Gender

To get a better picture of the type of people who are using such services, we've generated a profile of users according to age and gender.

Hubway Users - A Visualization of gender and age demographics



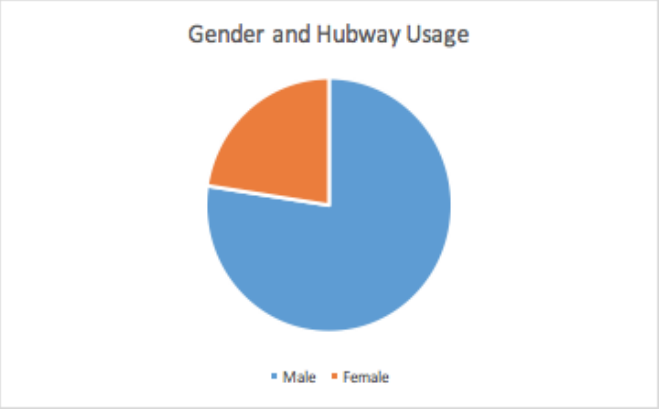


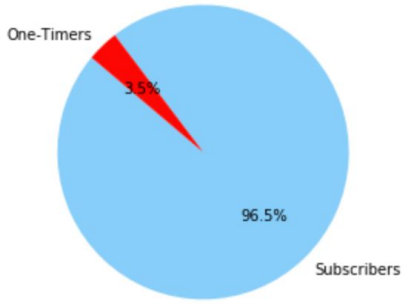
Chart 1(Left): This maps the number of trips taken in 2018(y-axis) by users of the Hubway Bikeshare system over the Age of the user(x-axis). It can be seen that the most popular age range is 27 which roughly correlates to the demographic of young working professionals. To get further insight into the gender makeup of Hubway usage, we observe **Chart 2 (Right)** which depicts the breakdown between male and female users. We are able to observe that this service is hugely popular with male users at approximately 75%.

From the graphs above, it can be seen that the Bikeshare Hubway service is most popular for users of age range around 27 and demand for such a service starts declining after the age of 40. This could indicate that the service is popular among students and mostly,working professionals. From the pie chart on the left, it can be seen that the service has mostly male users.

Part 2: Subscriber/Non-Subscriber Data and Usage

It is crucial to see how users use Hubway services - whether they are regular users(subscribers) or tourists(non-subscribers).

A representation of one-timer with subscriber user breakdown of Bikeshare Services



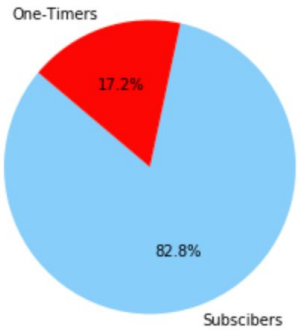


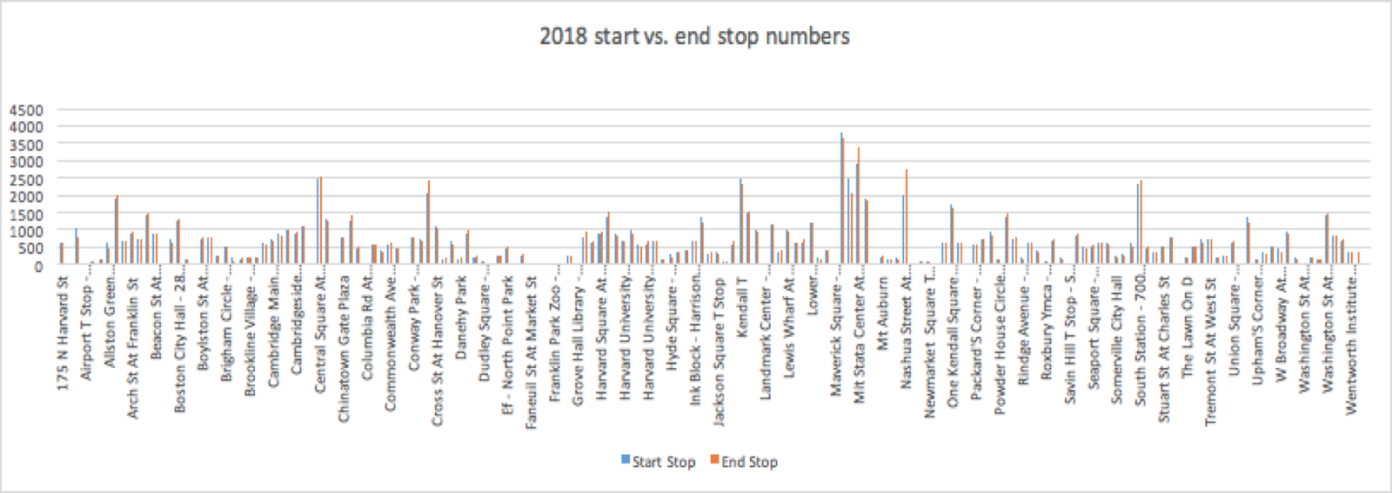
Chart 3(Left):Breakdown of percentage of total bikeshare users into one-timers and subscribers and **Chart 4 (Right)** : Breakdown of total duration of time spent on trip of one-timers and subscribers. The key observation that can be made from these pie charts is that even if one-timers make up only a 3.5% percent of hubway users, they take up almost a one-fifth time spent riding hubway. This information could actually enable hubway to make promotional pricing plans based off of hours spent riding for short-time users.

It is essential to note that even if non-subscribers do not make up a big portion of number of usages(**Chart 1**) , they spend more time per usage of bike share services which denote that they may be using it for long or leisurely explorations as tourists.This can be seen as even if they make up only 3% of all bikeshare users, they take up almost about 18% of the time on their bike based on duration of each trip(**Chart 2**). Since it costs more for an individual to ride out on a single ride when compared to subscribing it for a year, Hubway might be able to capitalise on this tourist opportunity by marketing it to tourists and New England sight seekers. (We are assuming that one-time users are either tourists or people who are visiting the area for a short period of time)

Methodologies and Results

Part 3: Stations , Neighborhoods and Frequency of Usage

From our data analysis, we were able to find that there were a total of 193 Hubway stations in 2018 (an increase from ~140 in 2014) . In order to see which stations are popular in terms of start and end stop usage, we have graphed the chart below:



Mapping the popularity of all current Hubway Stations - a detailed insight

Chart 5: The chart above is a representation of number of trips(y-axis) over Hubway Stations(x-axis). Taller orange bars in the bar chart correspond to the relatively high popularity of the station as an "end" station and taller blue bars denote the the popularity of a station as a "start" station. For example, 'Nashua Street Station' is a more popular end stop than a start stop which may possibly suggest that the number of docks places for bikes could be increased whereas 'Maverick Square' is more popular of a start station which could call for an ean increase in number of bikes made available.

From the above,we can see that the most popular places of usage was around South Station,other T-stations(such as Kendall) and other places of student or professional place of residence. This correlates positively with the age data that we found above with which we justified that the students and young professionals were most likely to use hubway services. However,to be able to cluster these distinct Hubway bike stations into neighborhoods, in order to truly see if income levels of a neighborhood affect Hubway usage at all, we made use of the Google Maps API to cluster each of the stations into distinct neighborhoods as seen below :
The data is used to map out the hubway stations with each of the Boston neighborhoods. The code snippet(attached in appendix) goes through July 2017 hubway data set and searches each stations' latitude and longitude using the google maps API through reverse geodecoding and returns which neighborhood the station is located in.

Clustering Hubway Bike Stations according to neighborhood

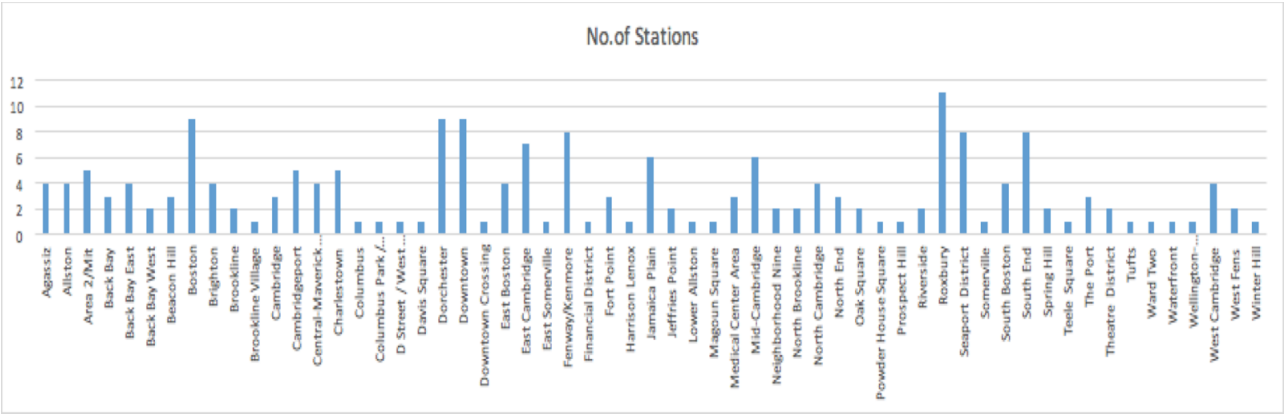


Chart 6: The above charts plot the number of Hubway Stations(y-axis) over stations' neighborhood location(x-axis). This information is vital to identify where Hubway stations are pre-dominantly placed and provide a gateway to compare further demographic indicators such as income and population of an identified area to the number of stations currently present for usage. For example,Roxbury is one of the low-income neighborhoods in the greater boston area and has one of the higher number of Hubway Stations. Does that mean that there is a positive correlation between income of an area and number of Hubway Stations? We explore this aspect further below.

The data above maps out the hubway stations with each of the Boston neighborhoods. Our program searches each stations' latitude and longitude using the google maps API through reverse geo-decoding and returns which neighborhood the station is located in. From the above chart, we can see that some of the lowest income neighborhoods in Boston(see Part 3 for detailed demonstration) such as Roxbury have one of the highest number of Hubway Stations. Downtown, Seaport District and Kenmore also have a high number of hubway stations owing to financial and educational activity.

Experimental Results

Part 4: Correlation between income of neighborhood and hubway services usage

We wanted to learn if an important economic factor such as average household income of a particular region is having a impact on the number of Hubway stations present as this would help in identifying where new Hubway station could potentially be added. But on analysis, we found that there was no strong correlation between income and number of stations. In our analysis, the Coefficient was -0.10 which meant there was really weak negative correlation between the two parameters.

Correlation Analysis : Household Income vs. Number of Stations present in Boston

```
[[ 1.         -0.10018914]
 [-0.10018914  1.        ]]
```

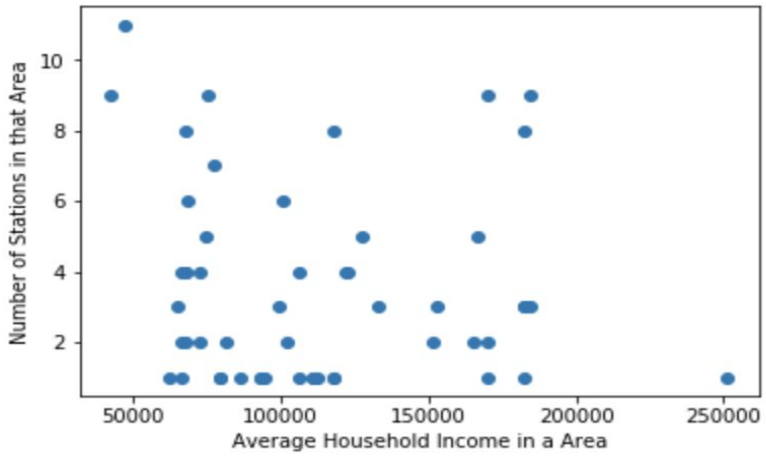


Chart 7: The above chart maps number of stations in the area(y-axis) over the average household income in the identified area(x-axis). The heavier the concentration of blue dots in a particular area, the higher the incidence of station numbers to that particular area's household income. Since the blue dots do not form a strong pattern, there is a weak negative correlation.

Part 5: Solving the "last mile Problem" - MBTA and Hubway

In order to understand the dynamics of how people use public transport and hubway systems, it is vital to analyse the prevalence of MBTA stations to Hubway Stops. This is important information as the usage of Hubway services is impacted by the presence of T-stops and other MBTA services. Interestingly enough, it's the "last mile"

that people use Hubway for when it's close yet far to reach the place of destination from the T-stop. In order to analyse that, we have a graph below that plots the prevalence of MBTA stations to Hubway Stops. Some trends that are visible are -

- In the city,MBTA Stations are more prevalent when compared to Hubway Stations. Hubway stations have a more consistent amount of stops through the city and the suburbs right outside of the greater Boston area.
- Given that there are more MBTA stations in the city, Hubway is potentially used as the next best ride to "bridge" between the T-stop and the destination as an easy and convenient way of getting around.
- It can also be seen that in places directly in the city centre(downtown,financial,Boston ity), there are a higher number of MBTA stops located per area than Hubway Stops. However, in places outside of such city hubs such as Riverside,Tufts,Kendall in Cambridge, MBTA and Hubway numbers in the same neighborhood mirror each other. This stands to show that MBTA exists in places with more established institutions whereas the Hubway services aim to connect commuters going to and from such institutions to places that might not be the busiest providing higher access to all areas otherwise not easily reachable.

'

The Last Mile Problem : Placement of Hubway Stations vs MBTA stops

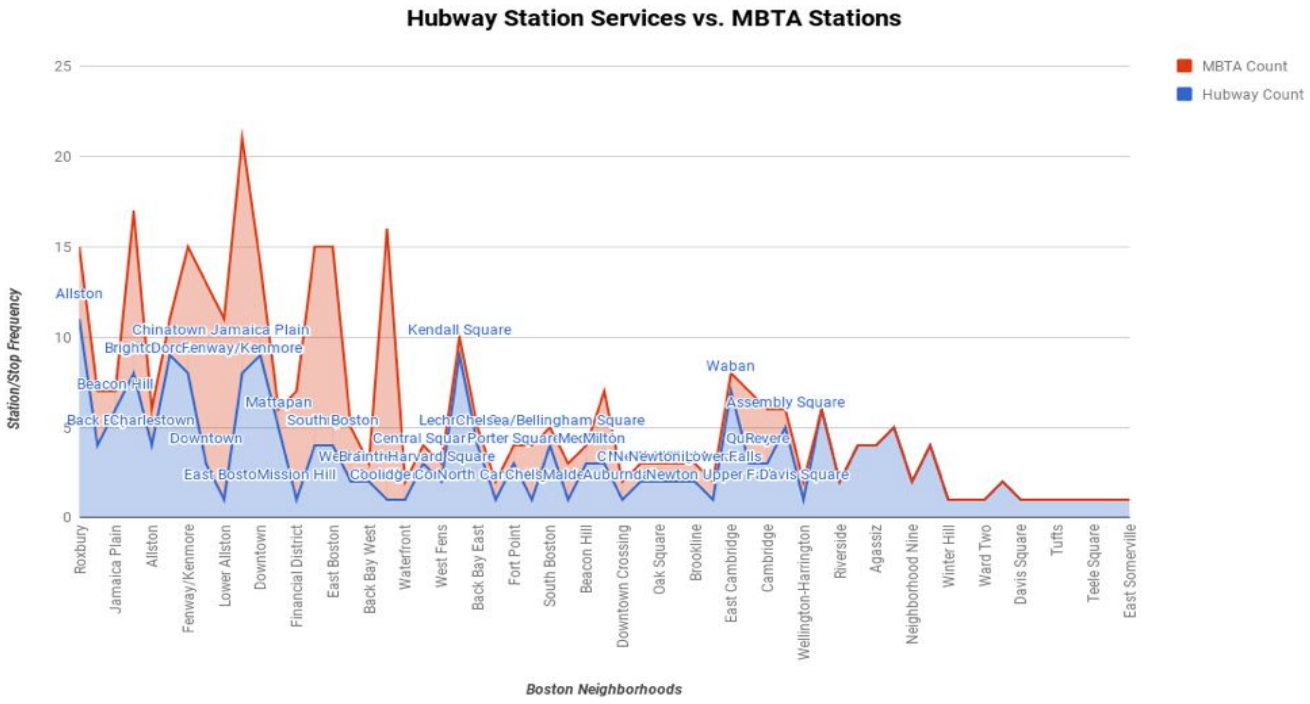


Chart 8: The above graph plots the frequency of identified start and end stations(y-axis) on identified Boston Neighborhoods(x-axis). The orange peaks indicate the number of MBTA stops in the particular area whereas the blue peaks indicate the number of Hubway stops . For example. Fenway/Kenmore has a high Orange Peak combined with an almost non-existant blue peak indicating that the ratio of MBTA stops:Hubway Stations in the area is really high, depicting a huge disparity. Whereas, stops outside the city lie Agassiz, Winter Hill,Tufts have almost the same number of MBTA and Hubway Stops.

Part 6 : Trends between the percentage change of number of stations and percentage change between total number of trips between years 2011 and 2016

As we expected to follow up in the future from the last report, we wanted to see how the percentage change in number of trips correlated to percentage change in numbet of stations. To predict the success of Hubway and the

variability of the data caused by different factors such as weather and looking at the adoption curve, we map percentage change of number of stations to the number of total trips.

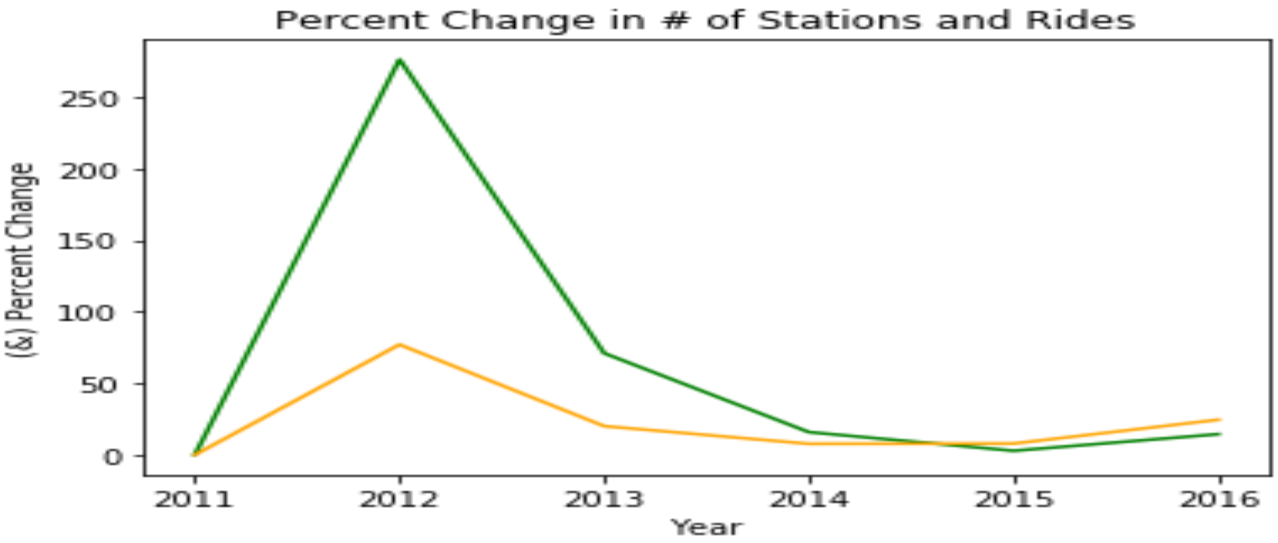


Chart 10: The above graph plots the percent change number of trips taken(y-axis) and percent change number of stations taken(y-axis)over the year(x-axis). 2014 to 2016 had a decrease in the change of rides from year to year and this anomaly can be attributed to the low rides at the first 3-4 months of 2015 due to a bad winter.

The percent change of the number of stations is represented by the orange line while the percent change of rides per year is represented by the green line. While the Hubway system was debuted in 2011, one can notice the spark increase in rides between 2011 and 2012. Between 2012 and 2015, the percent change of additional rides actually decreases. As for the percent change of additional stations, the number fluctuates with an initial spike and then a gradual decrease leading into an increase. Firstly, the number of stations is likely to flatten out because the area covered by each stations starts to overlap in the city, so each additional bike station reduces the efficacy of the system and profitability of the business. Between 2014 to 2016, there was a decrease in percent change of rides per year but an increase in number of stations. These years display the idea that weather can affect ridership. An interesting trend however is the percent change of rides per year. The curve represents the effects of factors such as weather on the Bike sharing system. Analyzing the frequency of rides and number stations through their percent changes from year to year, Hubway optimization towards its sustainability initiative and profitization for its business fluctuates directly due to external, uncontrollable factors.

Analysis on monthly trends shows an anomaly in the years of 2015 and 2016. To further explain the anomaly between years 2015 and 2016, we observe the graph below -

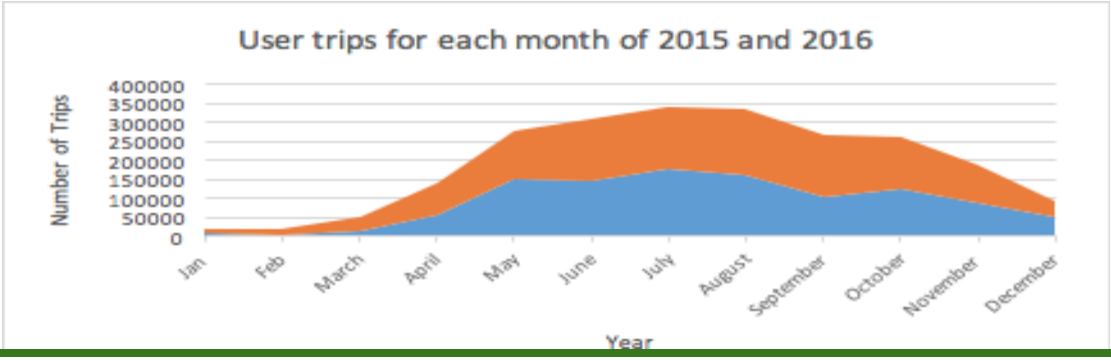


Chart 11: The above graph plots the number of trips taken(y-axis) over the year(x-axis). 2015 has much lesser rides than 2016 and this anomaly can be attributed to the low rides at the first 3-4 months of 2015 due to a bad winter.

The reason we have a change from 14% to 3% between years 2015 and 2016 can be attributed to the heavy winter and bad weather of 2015 . This trend is highly visibile because for the months january match, there is very low rider-ship for 2015 whereas for 2016, this is not the case. Hence, we are able to see that the reason for the dip between years 2015 and 2016 does not fit our general positive correlation prediction model because of the anomaly - winter. Our hypothesis is based off the fact that heavier winters cause low usage of Hubway Services due to practical problems in rider-ship.

From the above, two important things stand out, First, we look at the adoption curve or the "S-curve" of saturation.That is, what the adoption rate of Hubway so far was and what we expect it to be in the future. Our Given curve looks like this:

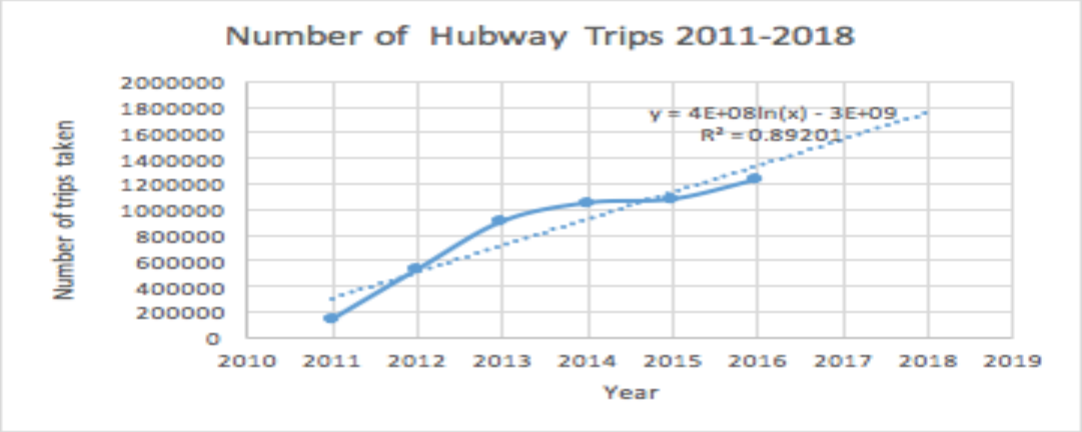


Chart 11: The above graph attempts to plot the adoption curve or "s-curve" for hubway bikes between years 2010 and 2018. We have used regression to approximate the data for 2018. We plot number of trips taken(y-axis) against year (X-axis)

From the given graph we can notice a couple things -

1) We plot the adoption curve which plots the number of trips taken between years 2011 and 2016 and we use a logarithmic regression model to extrapolate to estimate the total possible number of trips for 2018.
2) We use a logarithmic model because our data increases up till a point and then starts levels out. This regression model is preferred over polynomial or moving average beause our data set does not huge fluctuations of ups and downs.
3) Our R-squared value is 0.89 which shows a pretty strong correlation between the years and number of trips over years which further affirms that there is an increasing number of use of rides over the years although the rate of change between the years seem to be levelling down and decreasing according to the previous chart.

Revisiting the Problem Formulation:

At the start of this report, we posed a few questions that we hoped to answer through the report. Let's revisit the questions again and see how far we've come -

Q: However,what is it that's stopping the city of Boston from installing more Hubway docks?
A: From the above data, we know that the city of Boston increased its Hubway bike placements from about 130 to 183 in a span of about 4 years. We also realised that even if Hubway placement was popular over the city, more Hubway stops in the city could be placed to solve the "last mile problem" as identified.

Q:Will underserved populations benefit if such services were installed in lower-income areas?
A: Contrary as to what was expected from **Chart 6** , a quick correlation mapping of household income vs neighborhood area showed us that there were no strong links between identified income levels and Hubway Usage. However, Hubway services will most benefit those who are young working professionals, with some bit of closeness to an MBTA T-stops. Hubway could also make a more profitable revenue model to benefit off the higher amount of time spent using Hubway services by tourist one-time users in the future.

Q:How fast is the adoption rate and what do we expect in the future?
A: Our analysis from Part 6 shows that we have had a pretty solid adoption curve where the use of Hubway services grew proportionally with the number of stations implemented year by year. Our strong R-squared value confirms this trend.

Q*:Going forward, Key takeaway for Hubway to increase it's ridership would be*
A: From **Chart 10**, we can observe that rider-ship increases during the summer months. This information, coupled with information from **Chart 4** shows Hubway can be more profitable if it priced its Hubway according to number of hours spent. This is because from Chart 10 we are able to see that more riders ride Hubway and from Chart 4 we are able to see that tourists spend more time on the bike. Therefore, putting these two together - Summer Tourists charged per hour might likely bring in higher profits for Hubway if that is their long term goal.

Conclusions

In conclusion, we are able to notice a couple things. First and foremost is the fact that Hubway stations "mirror" MBTA stops in a lot of places outside the city which is able to provide better and cheaper access to places that would otherwise have to be reached by car or other means of transportation. However, from Part 3, we saw that Hubway in fact has a lot more number of frequency of trips to and from city centre areas. Therefore, in addition to the MBTA lines that already exist in Chinatown,Fenway/Kenmore, it might actually be more beneficial to install more Hubway Docks in the city centre in relation to the MBTA stops. Places like Chinatown and Fenway seem to have the highest disparity as far as the number of T stops and Hubway docs are concerned.

Hubway might be able to capitalise on marketing their services to tourists and short time visitors as they are more likely to spend a higher amount of duration riding the Bikeshare service than a regular subscriber would. Additionally, given the huge gender disparity that exists between the usage of hubway between males and females, Hubway could also play a bigger role in promoting the service for a higher female consumer base.

Given the frequency and neighborhood data, it can be seen that lower income areas such as Roxbury do have a higher number of Hubway stations(as far as places located outside the city are concerned) . However, as we

saw above in part 5, it can be seen that there is no direct correlation that exists between income of a particular neighborhood and hubway station number and usage.

Finally, We are able to see the huge spike in increase of the users between its flagship year of 2011 and 2012. After that the percentage change has been decreasing but has an increasing number of adopters over the years. Which is a great sign of adoption as we see that it almost reaches an S-curve with user increasing but reaching almost stable levels which is a healthy trend going in to the future.

References:

Hubway Data Sets : https://s3.amazonaws.com/hubway-data/index.html
Census: https://www.census.gov/quickfacts/fact/table/bostoncitymassachusetts/PST045216

Additional References:

Hubway Information for confirming trends:
https://www.boston.com/news/business/2015/03/03/whos-riding-the-hubway-this-winter
https://www.thehubway.com/about
https://www.cityofboston.gov/images_documents/2012%20Boston%20Bikes%20Annual%20Summary%20reduced_tcm3-38305.pdf

Income Data:
https://statisticalatlas.com/place/Massachusetts/Boston/Household-Income

Github link:
https://github.com/abhivora1/CS506-Bikesharing_Project