**Capstone Project: Coffee Joint in Mumbai**

**A. Introduction**

**A.1. Description & Discussion of the Background**

*Mumbai* (formerly called Bombay) is a densely populated city on India’s west coast. A **financial centre**, it's **India's largest city**. On the Mumbai Harbour waterfront stands the iconic *Gateway of India* stone arch, built by the British Raj in 1924. Mumbai encloses an area of **603.4 km²** and has a population of nearly **2 crores**.

Mumbai is a city with a high population and population density. Being such a crowded city leads the fast-food chain business and social sharing places in the city where the population is dense. When we think of it by the investor, we expect from them to prefer the areas or suburban places where there is a lower competition and the type of business, they want to install is less intense. If we think of the city residents, they may want to choose the regions where options are more, and prices are economical. However, it is difficult to obtain information that will guide investors in this direction, nowadays.

Me and few of my entrepreneur friends wants to open new coffee hangout joint in Mumbai city. We're interested in a location where there are least No. of coffee shops already present, so that we can maximize the profits.

Mumbai is a costly city and since I'm not much aware about the city, I decided to do a quick study of different areas in Mumbai to see if I can make a decent investment.

In this project, I will try to find an optimal location to set-up a coffee hangout place.

Specifically, this report will be targeted to stakeholders interested in opening a Coffee Hangout in Mumbai, India.

Since there are lots of coffee hangouts in Mumbai, I will try to detect locations that are not already crowded with coffee joints.

I'll be using Data Science techniques to generate different visualizations, in forms of graphs and maps, depicting the promising neighbourhoods to open a coffee join in Mumbai.

**A.2. Data Description**

Following data sources will be needed to extract/generate the required information:

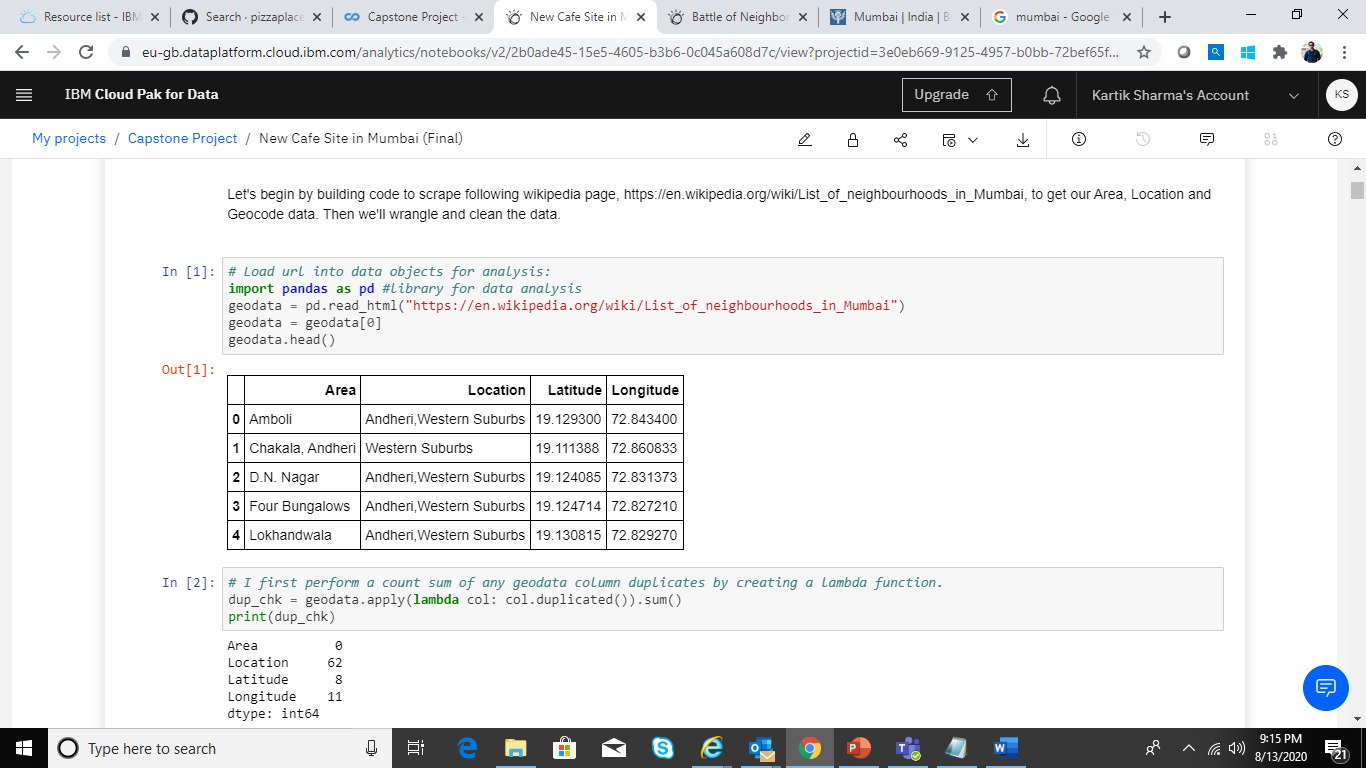
1. **Wikipedia** source for different areas and locations in Mumbai: <https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai>
2. **FourSquare API** to search for existing Coffee Joints in Mumbai.
3. Different **Python Libraries** like *Folium* for visualizing the data on Maps, *Matplotlib* for graphs, *Pandas* and *Numpy* for Data Analysis, etc.

**B. Methodology**

For start, I identified the below tasks to

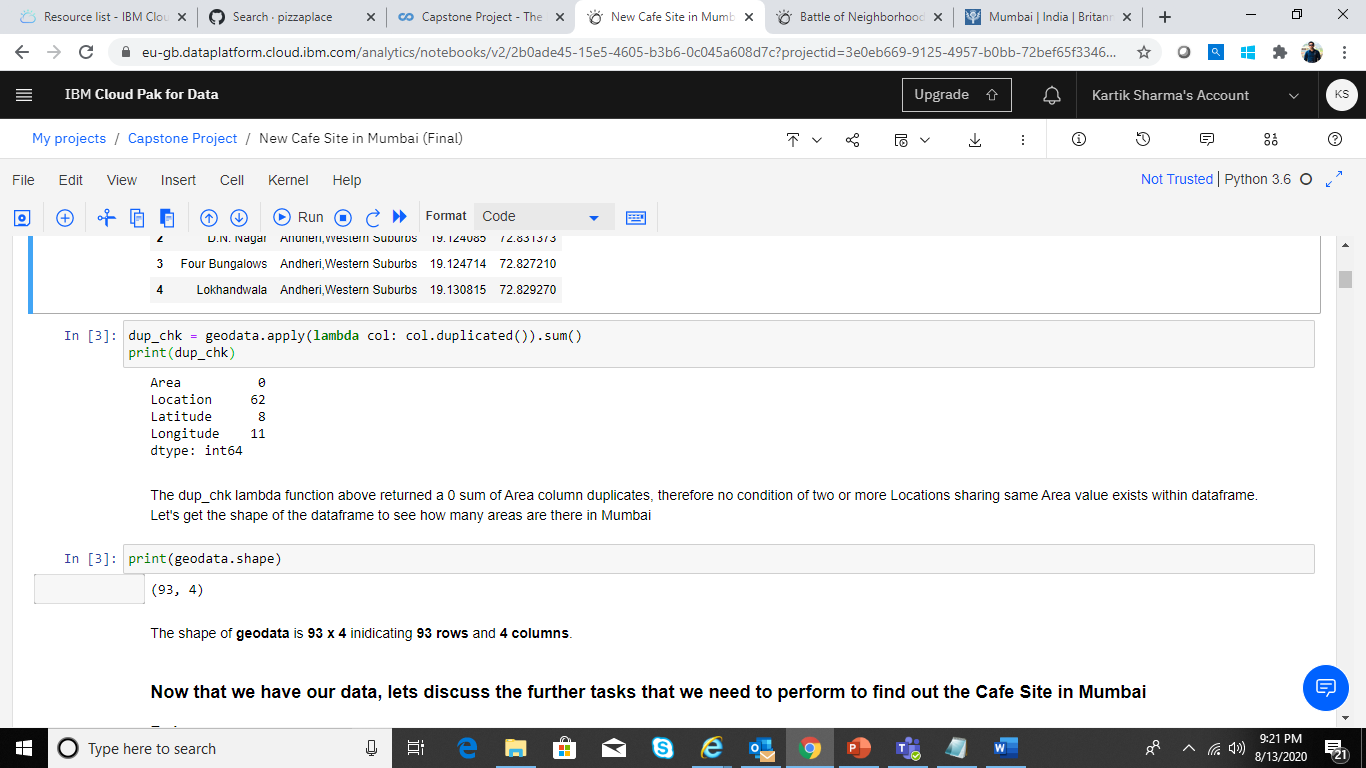
1. Scrape Mumbai postal code information from [Wikipedia](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai) website.
2. Load the data into a dataframe where data is wrangled, cleaned and transformed into a final dataframe.
3. Populate the dataframe with Mumbai Area, Locations and Latitude and Longitude information.

To scrape the Wikipedia website, I imported pandas dataframe and created the dataframe **geodata**:



This gave me the list of all Areas, Locations and geocoordinates for Mumbai City, as shown above.

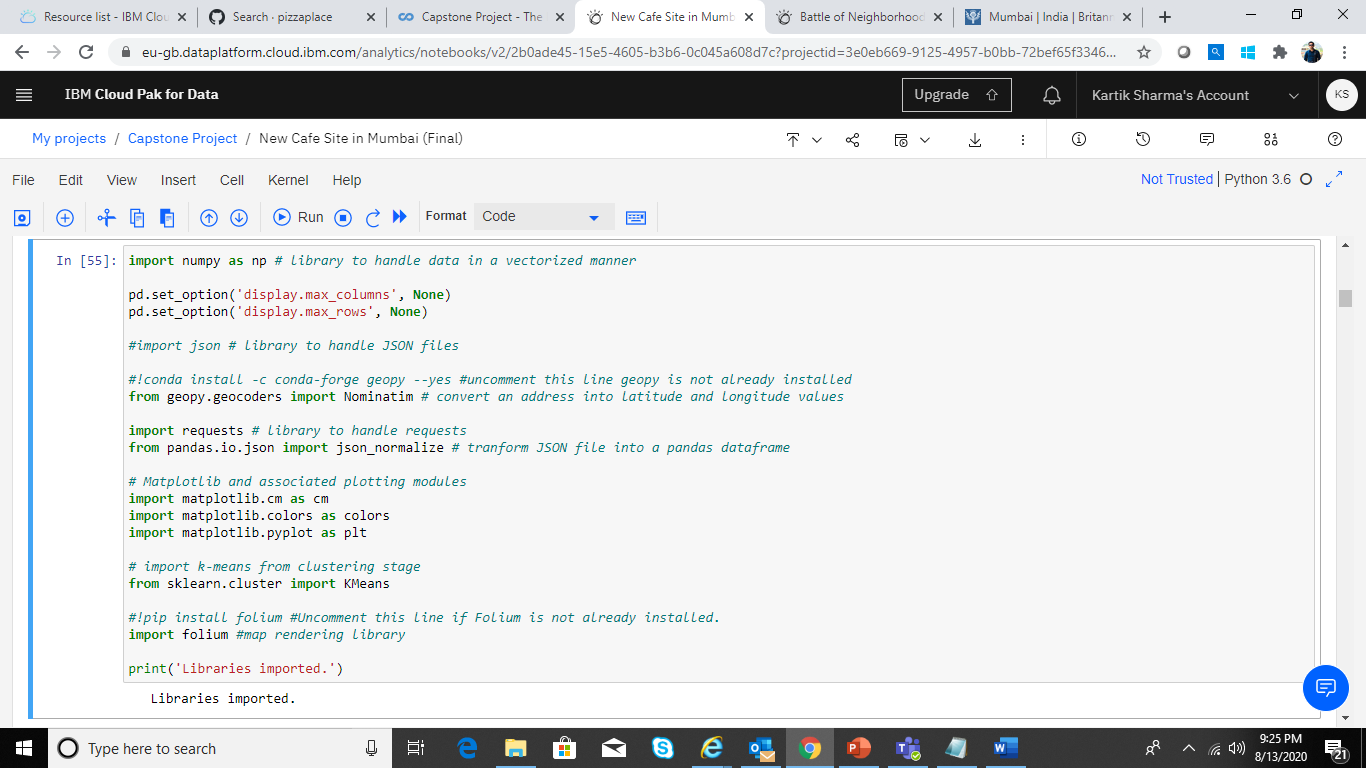
Then, I performed a check on the data to ensure that there are no duplicate Area names in the dataset, following which I also, noted the shape of the data frame:



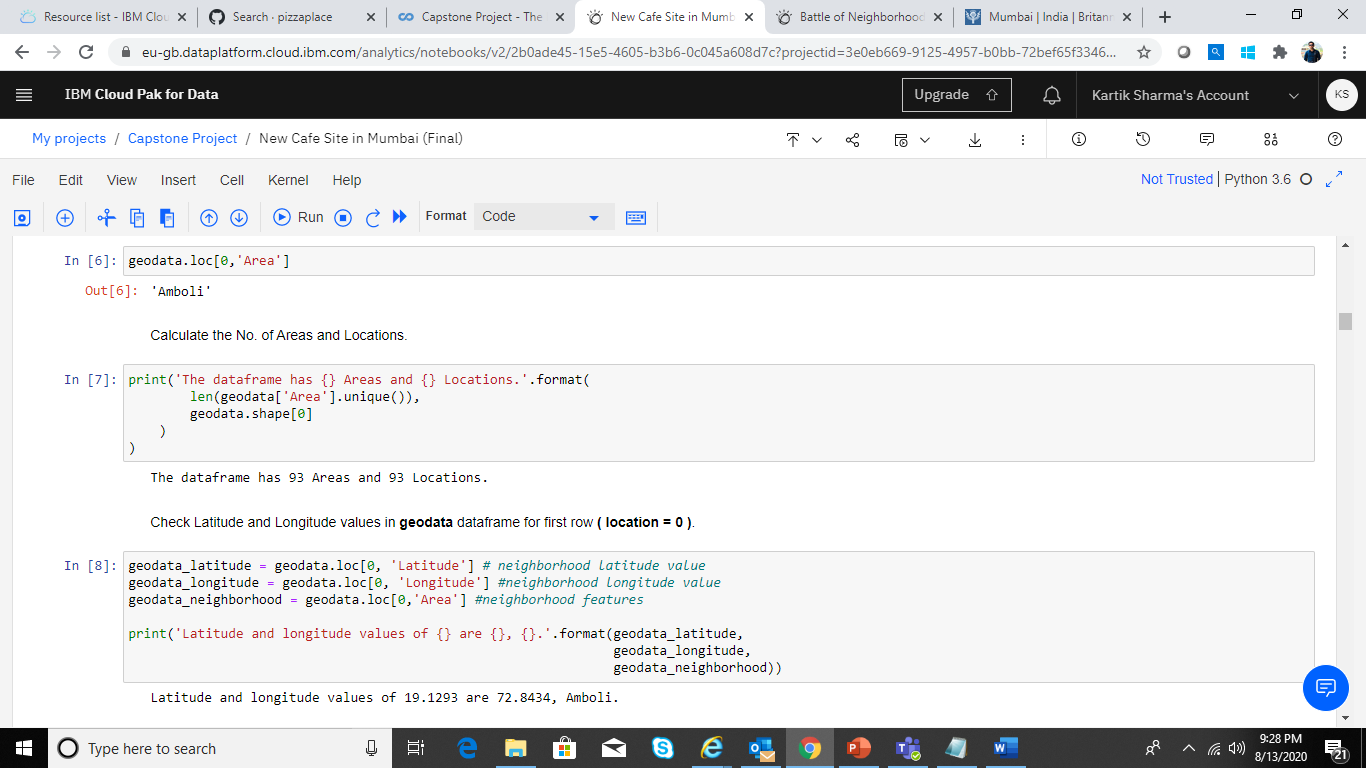
Once this task was completed and I validated my dataset, I identified the below tasks to perform in order to achieve my goal:

1. Validate the Data further
2. Setup FourSquare Parameters
3. Search for Coffee Places in API Response
4. Explore Neighbourhoods in Mumbai
5. Visualize the Areas with Least No. of Coffee Shops

In order to continue further, imported the below Python Libraries:

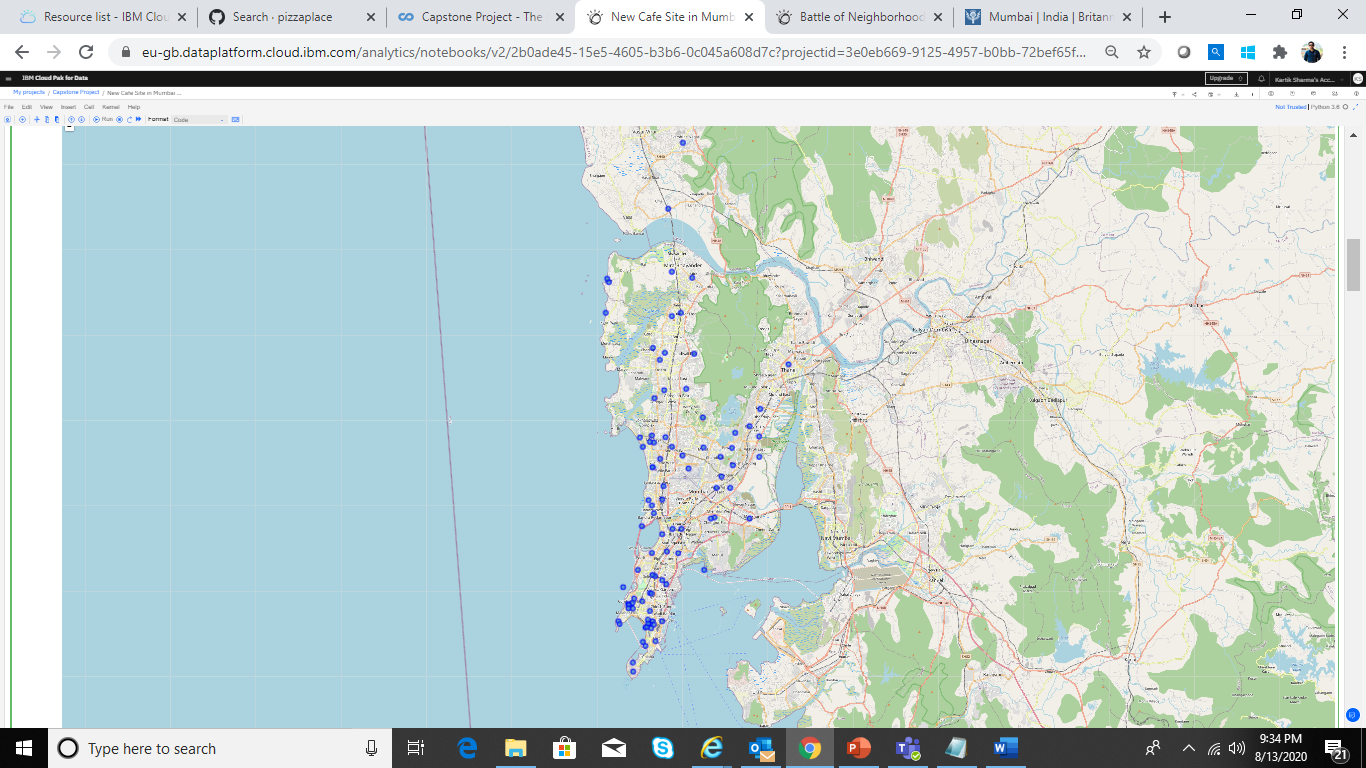


I validated the data in my dataframe further using the first row of the dataframe and calculating it’s geocodes:

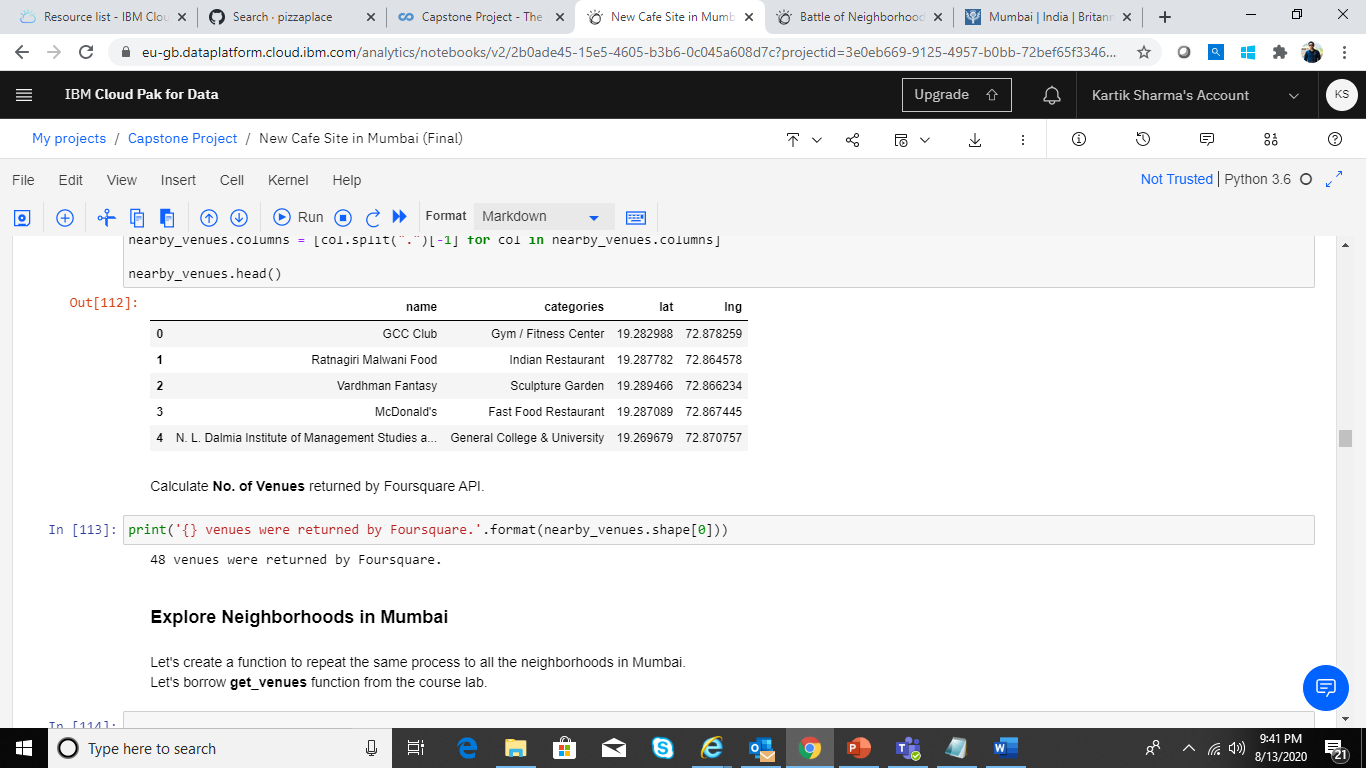


Next, I setup my FourSquare API credentials using my Client ID and Client Secret ID.

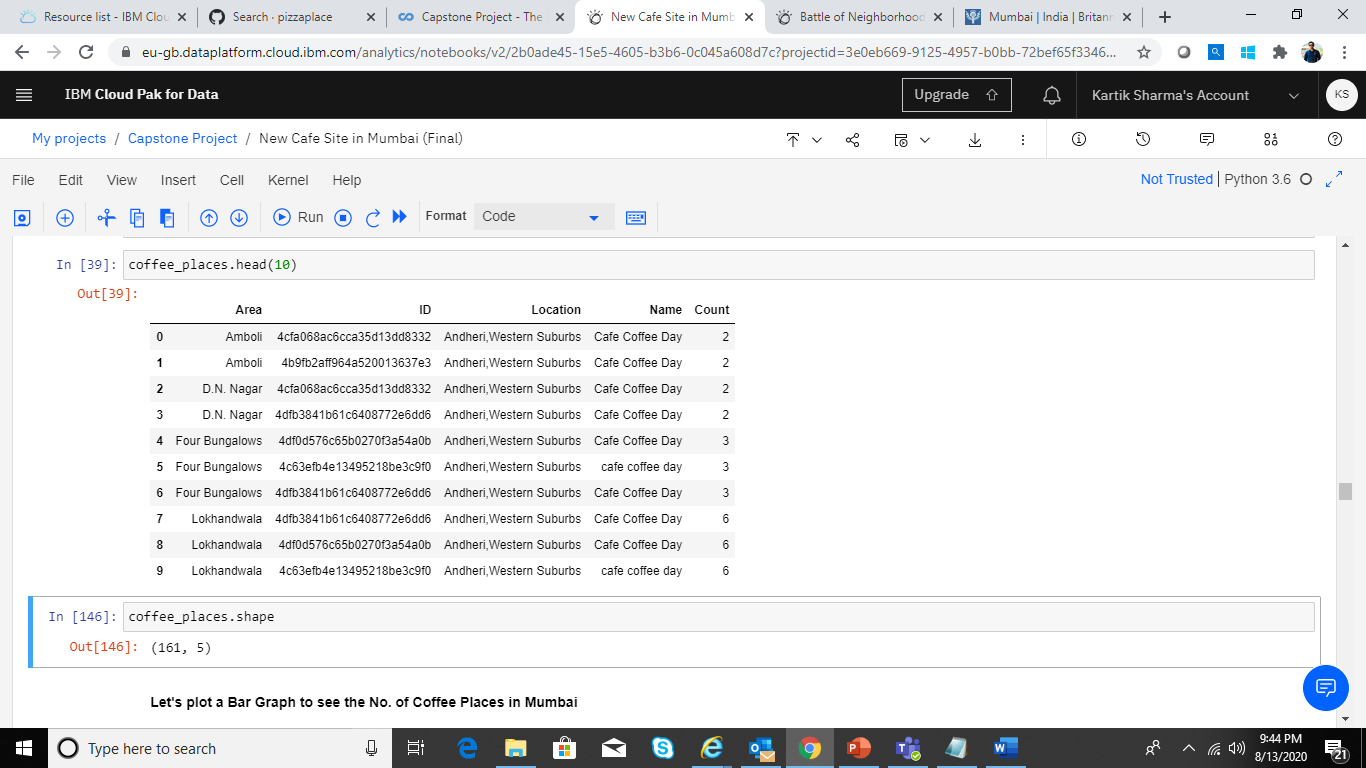
I used python Folium library to visualise geographic details of Mumbai city and its suburban areas and I created a map of Mumbai with areas superimposed on top. I used latitude and longitude values to get the visual as below:



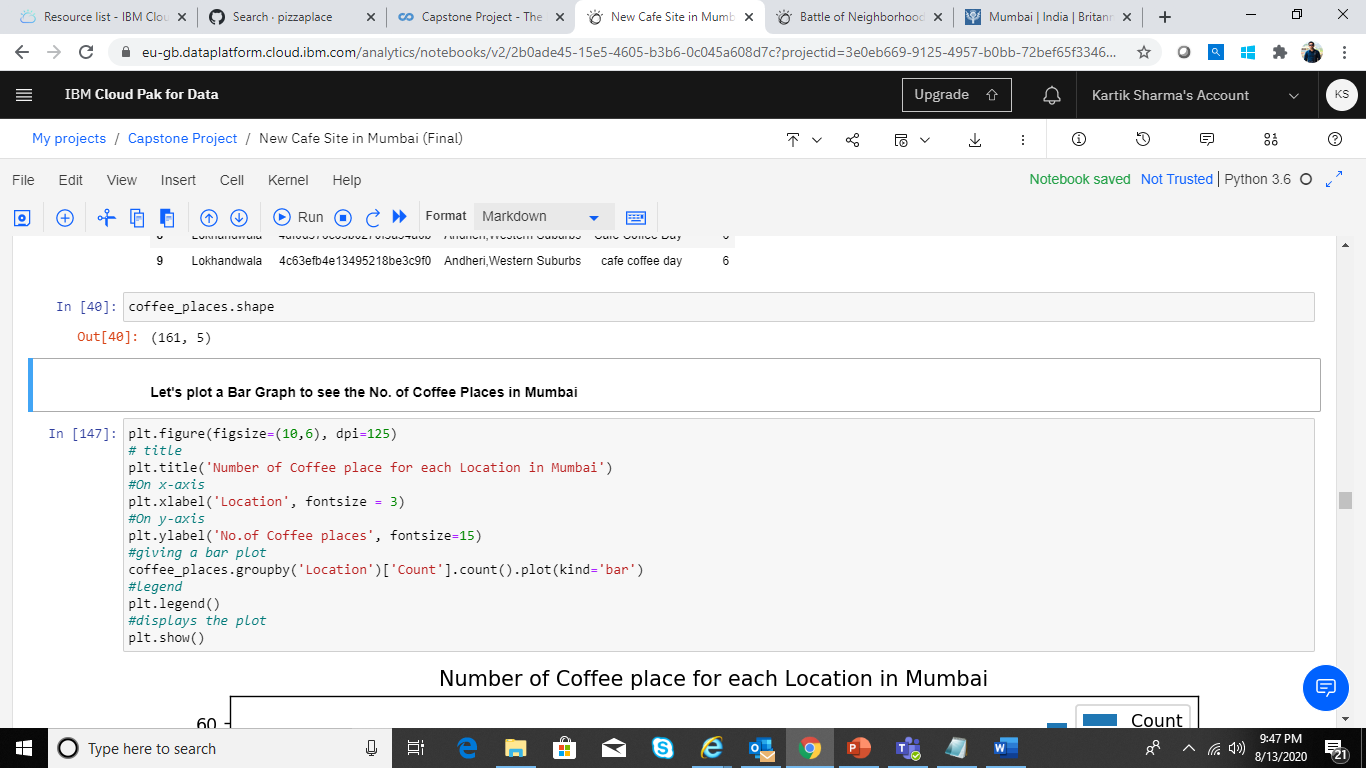
Next, I used the FourSquare API to explore the areas and locations in Mumbai city. Particularly, I searched a radius of 3 Kms around **Mira Road** in Mumbai to identify different categories of places identified by the API call:



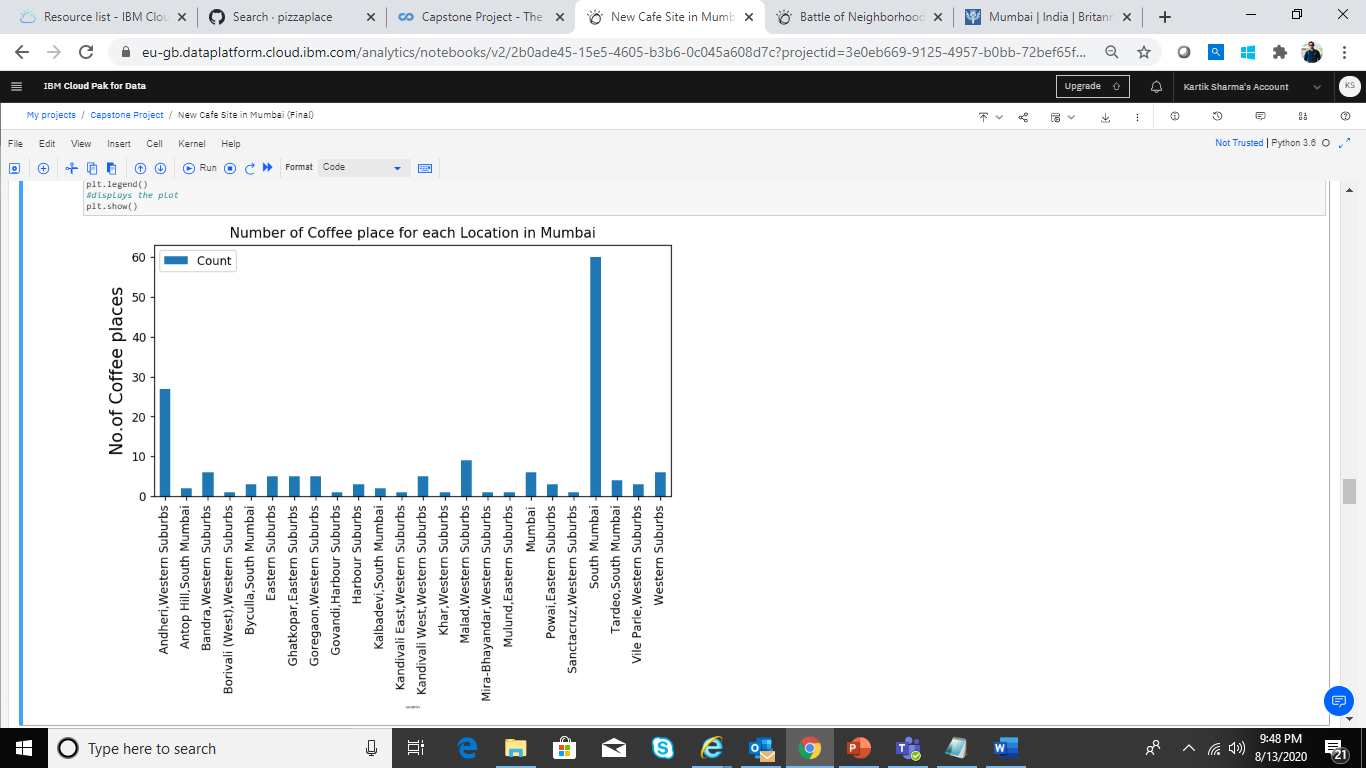
Next, I used FourSquare API calls to generate a list of all coffee places listed in Mumbai. Below are few of them:



A total of 161 Coffee joints were identified by the FourAquare API in Mumbai:

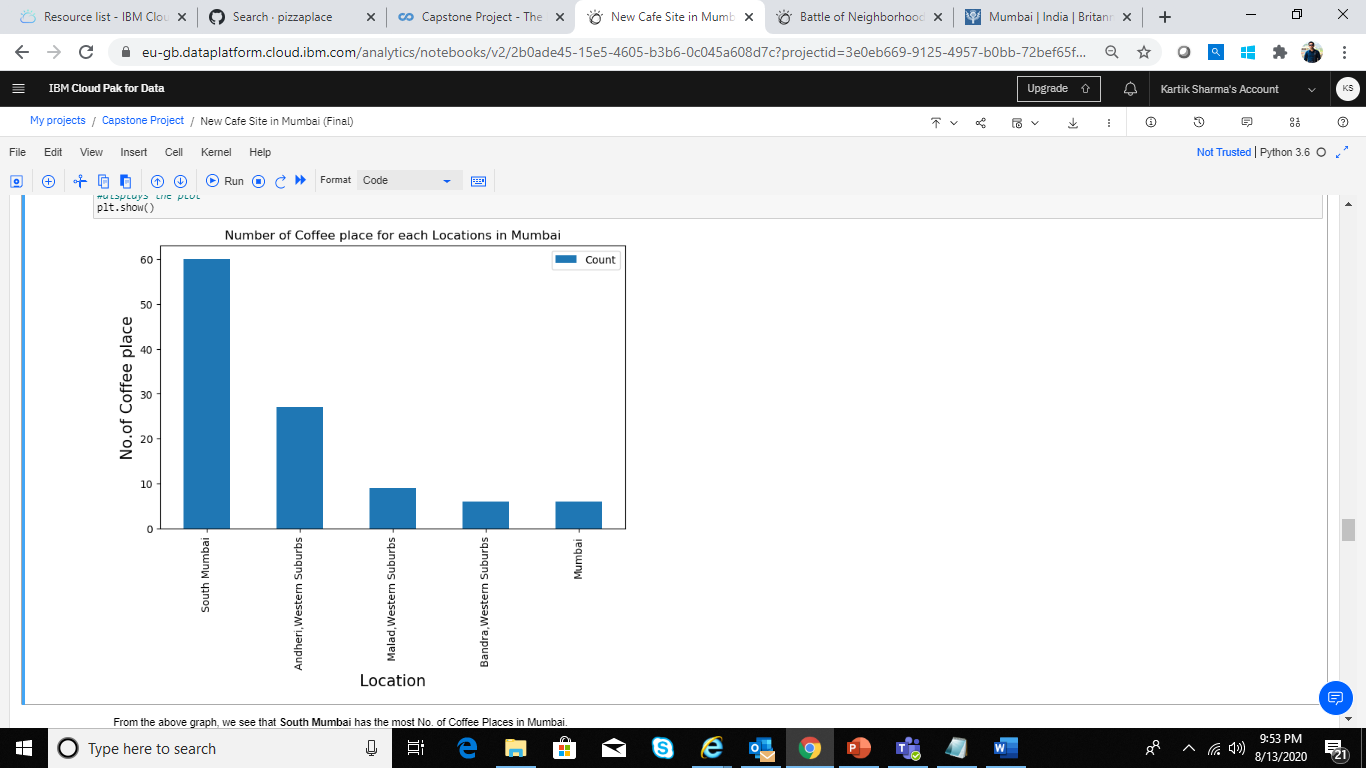


Based on different locations in Mumbai, I plotted a simple graph to identify which zones has a greater number of coffee places:

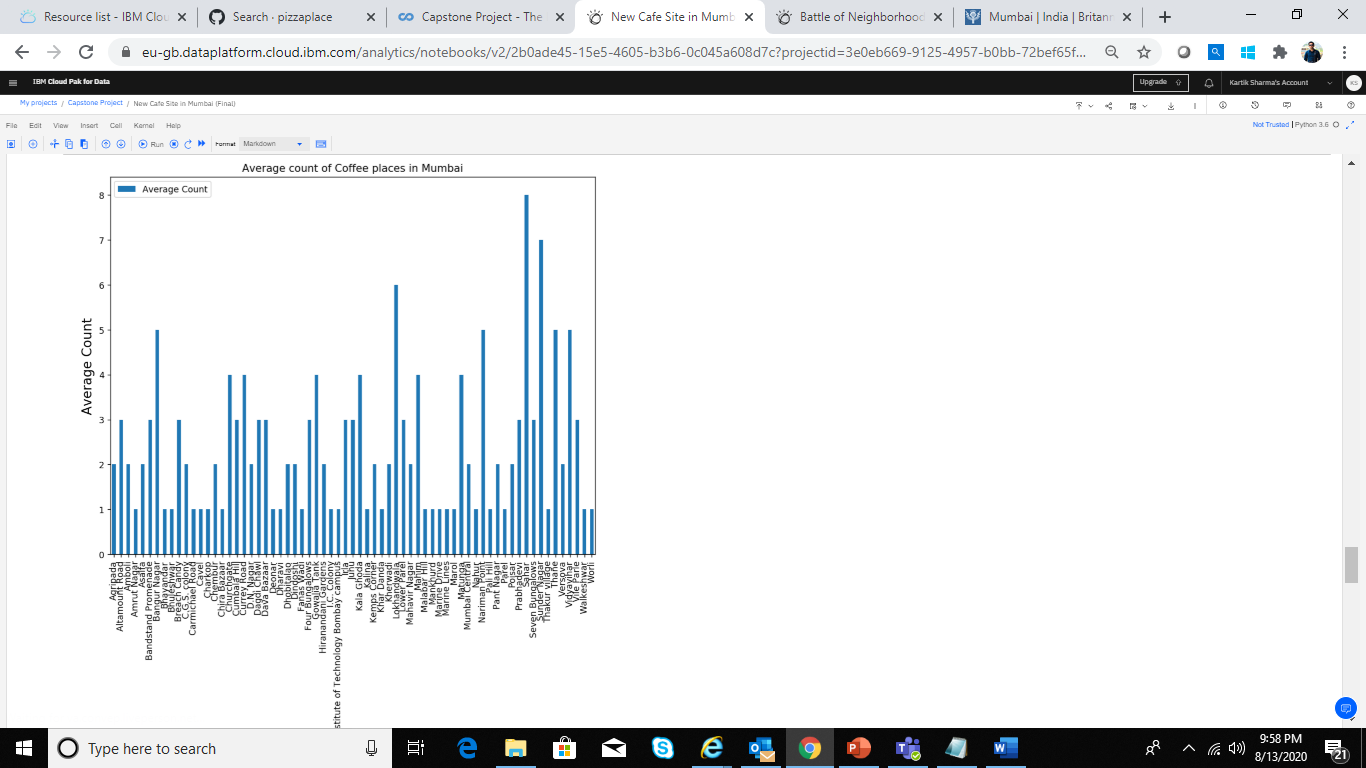


From the above graph, it was evident that **South Mumbai** has a greater number of Coffee shops followed by **Andheri, Western Suburbs**.

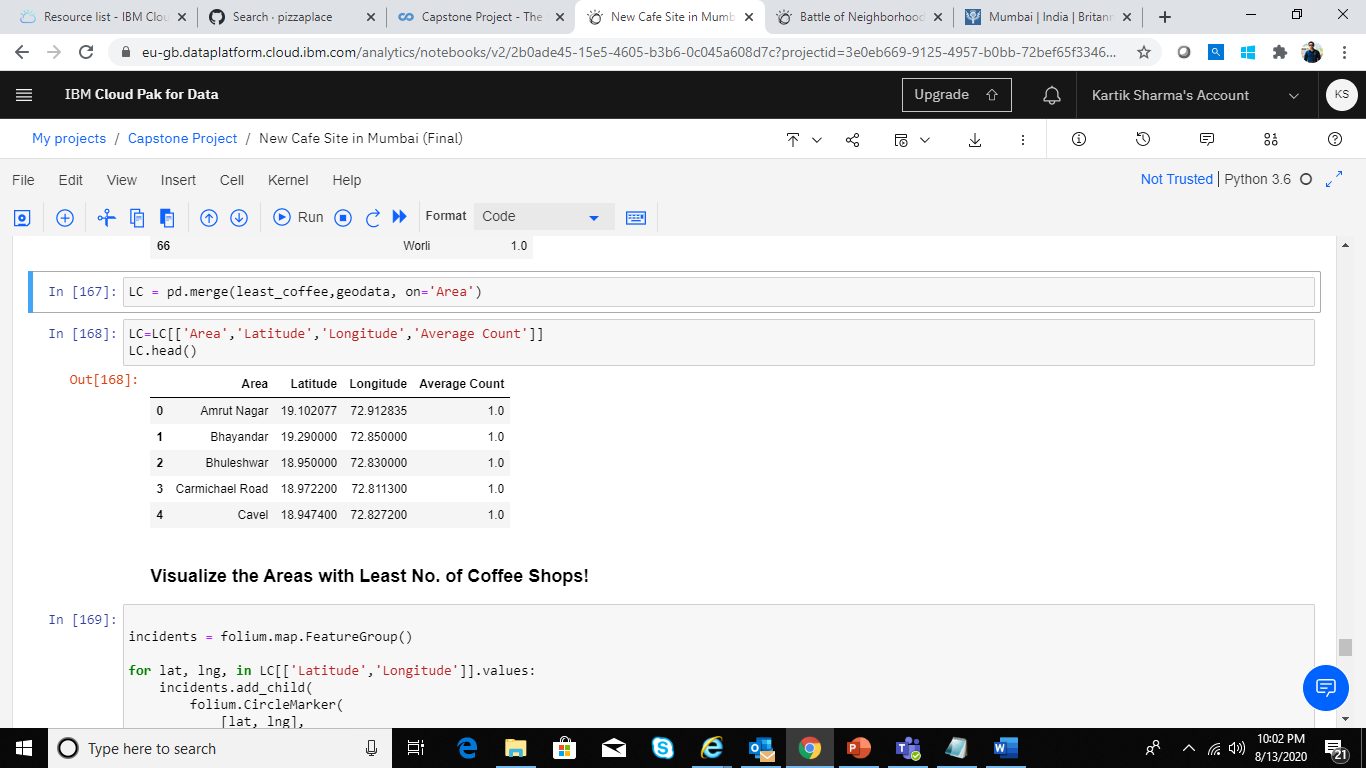
Below graph shows the Top 5 Zones having maximum number of Coffee Shops.



Drilling down further, it was found that on an average **Sahar** has the greatest number of coffee shops in Mumbai:

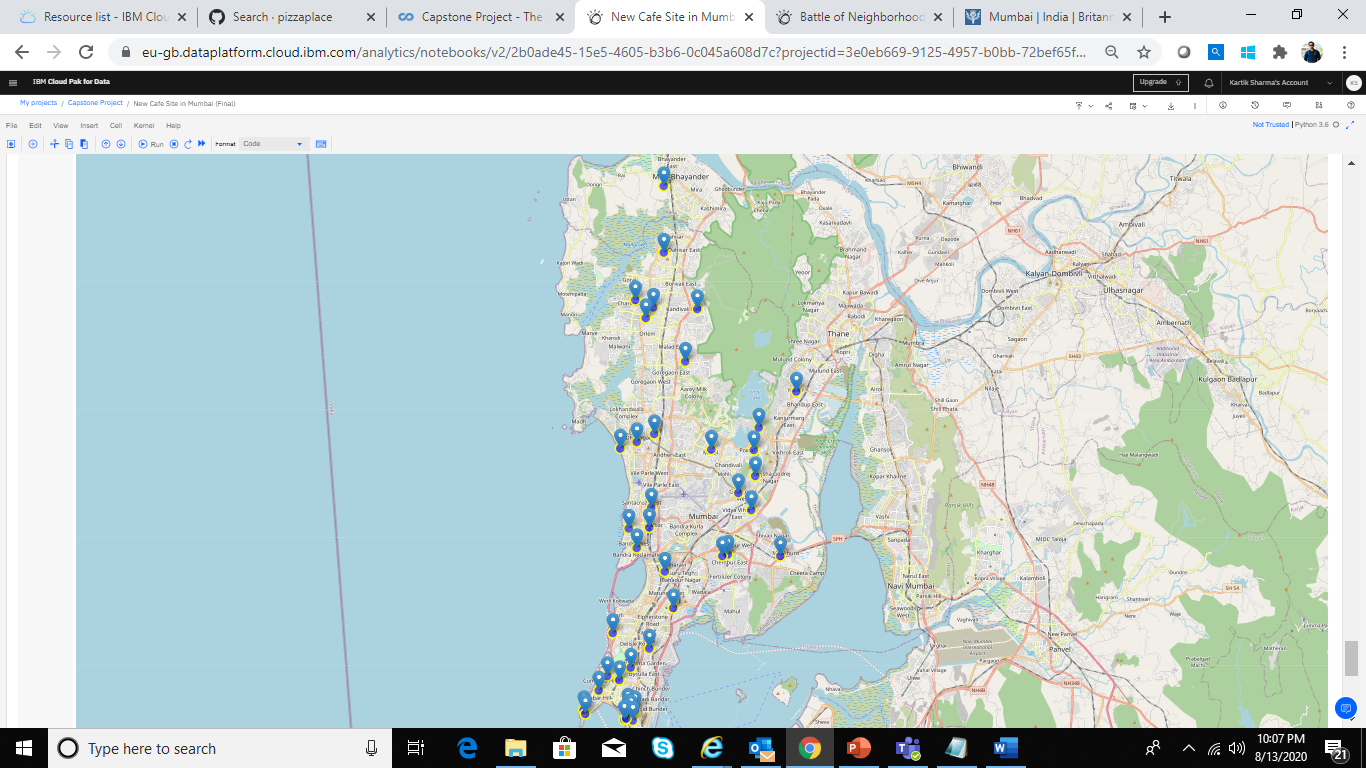


Moving on, I identified the areas having the least average of coffee shops in Mumbai:



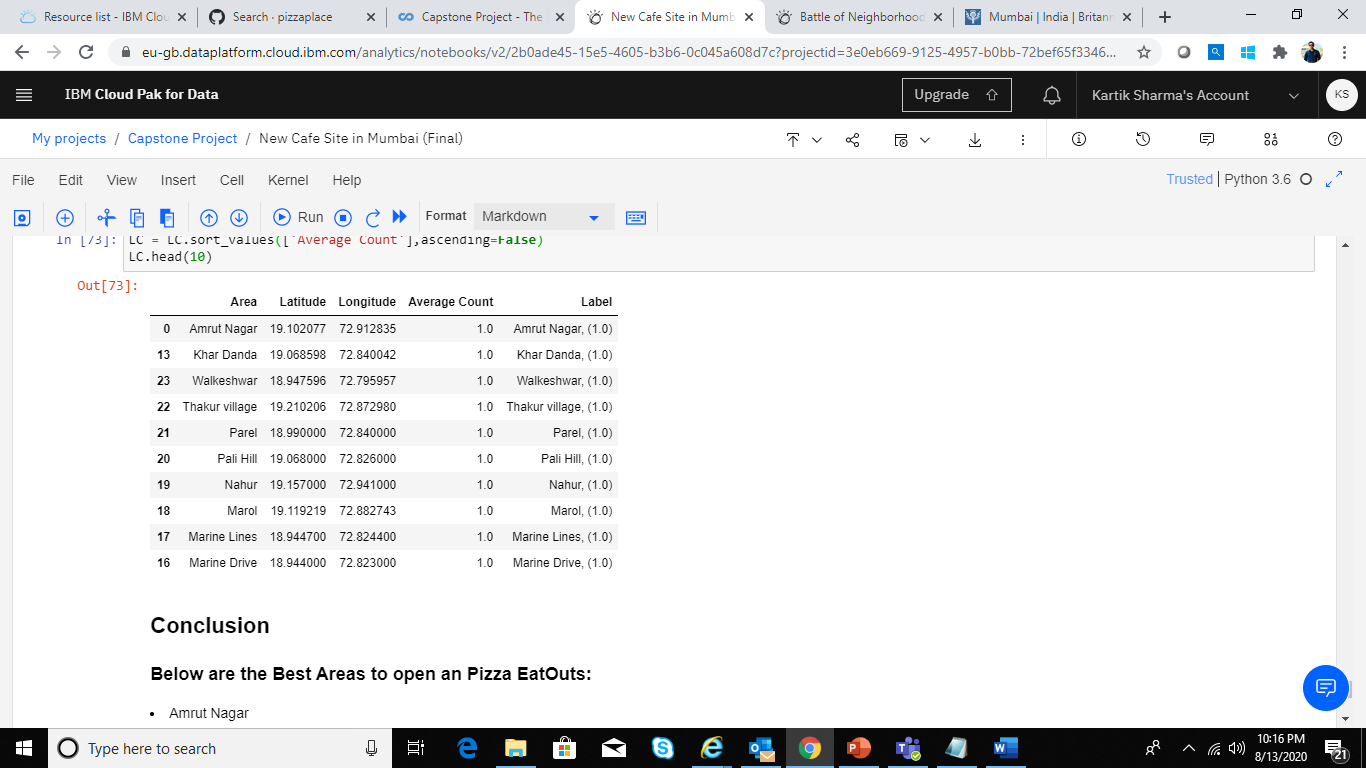
To identify an area to invest in coffee shop, I used the above areas which have least average of coffee shops.

Finally, visualizing these areas on map, using Folium, I was able to clearly locate the spots for a good coffee joint where people would come in for a nice coffee:



**C. Results**

Using Data Science techniques, I was able to identify the below areas in which I can look forward to invest in a coffee shop as these areas have the least number of coffee joints:



As we can see below are the list of areas which has less coffee joints and it would be good for business, if we consider these areas to open new coffee hangout.

1. Amrut Nagar
2. Khar Danda
3. Walkeshwar
4. Thakur Village
5. Parel
6. Pali Hill
7. Nahur
8. Marol
9. Marine Lines
10. Marine Drive

**Limitations:**

The Count is highly dependent on Foursquare API details.

There might be some other coffee hangouts which are not listed in foursquare database.

**D. Discussion**

Like I mentioned earlier in my problem statement, Mumbai is the largest Indian city which happens to be the costliest with a high population density. Mumbai is known for its variety of food joints for all price ranges.

Probably, it would be best for me to analyse further on the areas where I need to invest to get the best possible profit. For example, which areas cost more to setup a joint compared to others and what best fits my budget.

**E. Conclusion**

Going through the data science course and having a hands-on this project, I am better familiar with data science techniques and methodology. I’ll try to analyse further into my project and see if I can better my analysis and come up with best possible investment areas