Course Capstone

*IBM professional data science capstone*

*Purchasing a house in Ahmedabad (Gujarat, India)*

By -Maharshi Pandya

December 2019



**Introduction:**

The city of Ahmedabad (Gujarat, India), located at the centre of the state is home to many industries, branches of national and multi-national companies as well as tons of start-ups. This is due to the policy of government whose aim is to make the city a major industrial/ business friendly city. This attracts lots of people from all across the state and country for the job opportunity. This accompanied by people migrating to the city from neighbouring villages rises demand for houses.

**Business Problem:**

All the people want to buy an affordable housing for themselves. But the problem is they don’t know which area is the better one and also the price trend of that area. It’s also important to note that crime rate in that particular area as well as nearby area can also affect this decision. This decision can also be affected by mall, garden, grocery store in that area etc. So, the aim is to find ideal place for the stay which satisfies our criteria/ parameters.

**Targeted Audience:**

Targeted audiences for this project are the people looking for houses in the city as well as those who wants to invest in property. This can also be helpful for those who want some property for warehouse or non-residencia purposes.

**Data:**

Data of Ahmedabad city will be obtained from the site of city(<https://ahmedabadcity.gov.in/portal/jsp/Static_pages/demographics.jsp>).

This site also includes the education as well as education of citizen of any given word.

Data of current pricing of any place in any areas will be obtained through these two links.

1) <https://www.99acres.com/property-rates-and-price-trends-in-societies-in-ahmedabad>

2) <https://www.makaan.com/price-trends/property-rates-for-buy-in-ahmedabad>

For data abstraction Beautiful Soup will be used.

**Methodology:**

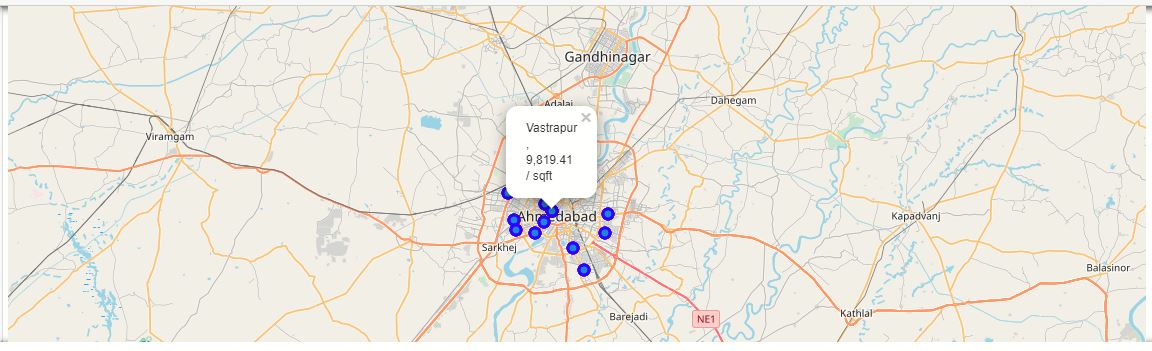
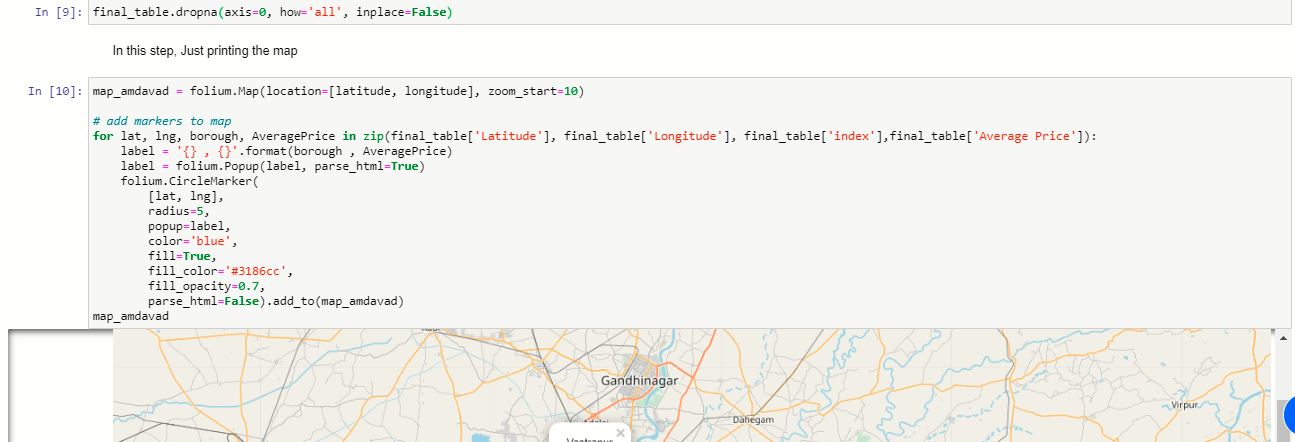
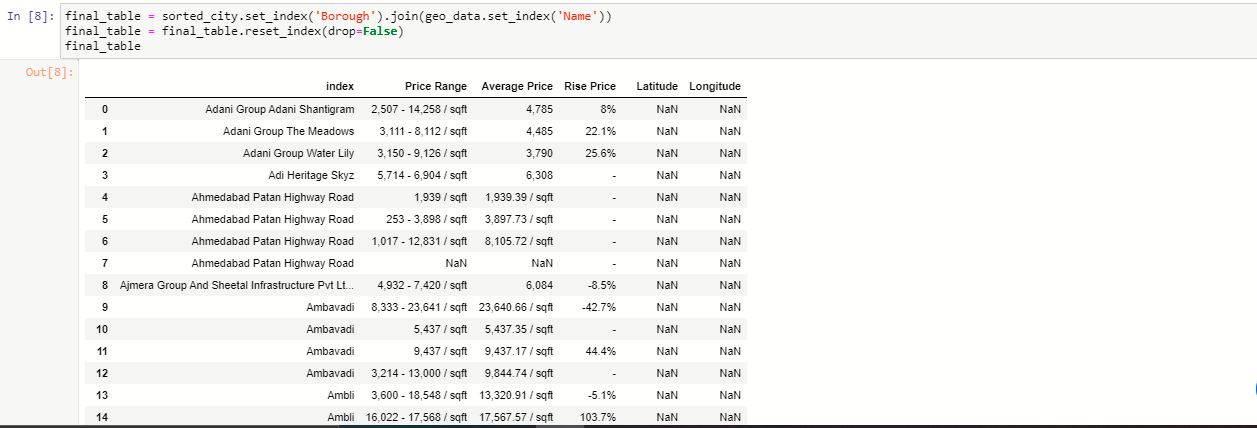
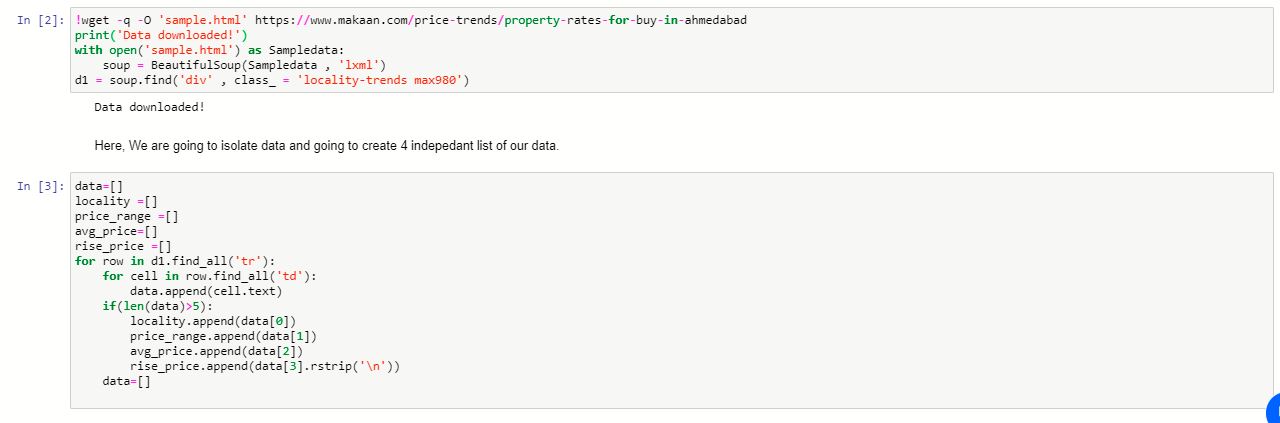
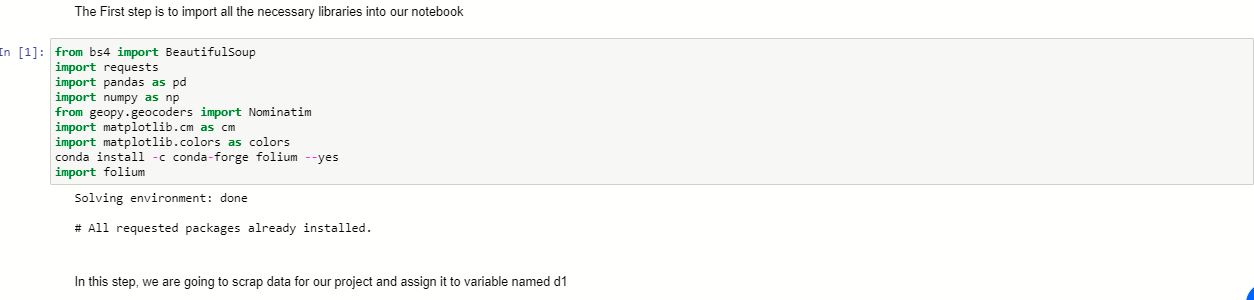
Here, first of all I initialised all the variables and then used web scraping the data from link 2. Next stage was data wrapping. i.e. cleaning the data and removing the any not assigned value from the link 2 as the link 1 giving some sort of unknown error (server got disconnected). After cleaning the data, I searched for the for the file with the ward wise location of Ahmedabad city. This took me some time but I found it after some time. Then I merged both the table. Then I found location of Ahmedabad using Nominatim and plotted the map around that central location.

**Problem Faced:**

The first problem I faced was in web-scrapping. Initially I tried link 1 for the web-scrapping but I kept getting error as mentioned above, so after some time I switched to link 2 which was completely normal and moved onto the next stage.

The second problem I faced was while finding the location of all the places mentioned on the data frame. To find the location of all the places, I tried using Nominatim in the loop which gave me two error: network not reachable and geocode not available. I posted the question on coursera discussion and stack overflow but did not get any answer for 2 days, so I moved on to finding the csv or excel file in which I can find latitude and longitude of all the places. But there is not a single file that I found. And after hours of googling I found a geojson file. I converted it into csv file using online converter and then I moved on to the project

**Snapshot:**



**Conclusion:**

Hence, I concluded my project on Ahmedabad city housing analysis.