**BATTLE OF THE NEIGHBORHOODS**

***Schools govern Relocation to New Place- Myth or Reality: Case Study of Charlotte Area, NC***

**Introduction**

*The Battle of the Neighborhoods* is a capstone project for Coursera and IBM’s *Data Science Specialization*. This project encompasses the vision of the Data Scientist to portray the tools learned during 9 courses on *data analysis* for the certification. One of the requirements is to use Foursquare location to explore and solve the neighborhood analysis using clustering method.

**Background**

*“Schools govern Relocation to New Place- Myth or Reality: Case Study of Charlotte Area, NC”* is a project taken to understand how the schools affect the relocation to a new place/ neighborhood. Charlotte area is known for its Banking jobs and the increase of financial institutions making Charlotte and its surrounding areas, their financial headquarter.

The goal of this project is to find a neighborhood for a person relocating with his family to Charlotte, NC. A family man requires an easily accessible neighborhood with some good schools. Certainty and uncertainty of renting or owning a house revolves around a good neighborhood. Estate agents try to their potential to sell houses that are most profitable to them. Personally, having a little insight about the area and its neighborhoods can help the family plan their relocation smoothly.

This study is an attempt to find such a neighborhood for a family with school going children.

**Data Requirement**

Best approach will be to follow a *Top-Down Approach* for getting the best results. The identified data for such a study is:

* Affluent Neighborhoods in Charlotte Area,
* Real Estate Values of Charlotte Area,
* List of Schools,
* School Ranking,
* Accessibility of neighborhoods to nearby venues,
* Crime Rate, if possible

**Methodology**

As per the earlier studies, it has been established that real estate is dependent on its surroundings and proximity to important landmarks. To establish the same correlation, following methodology will be followed:

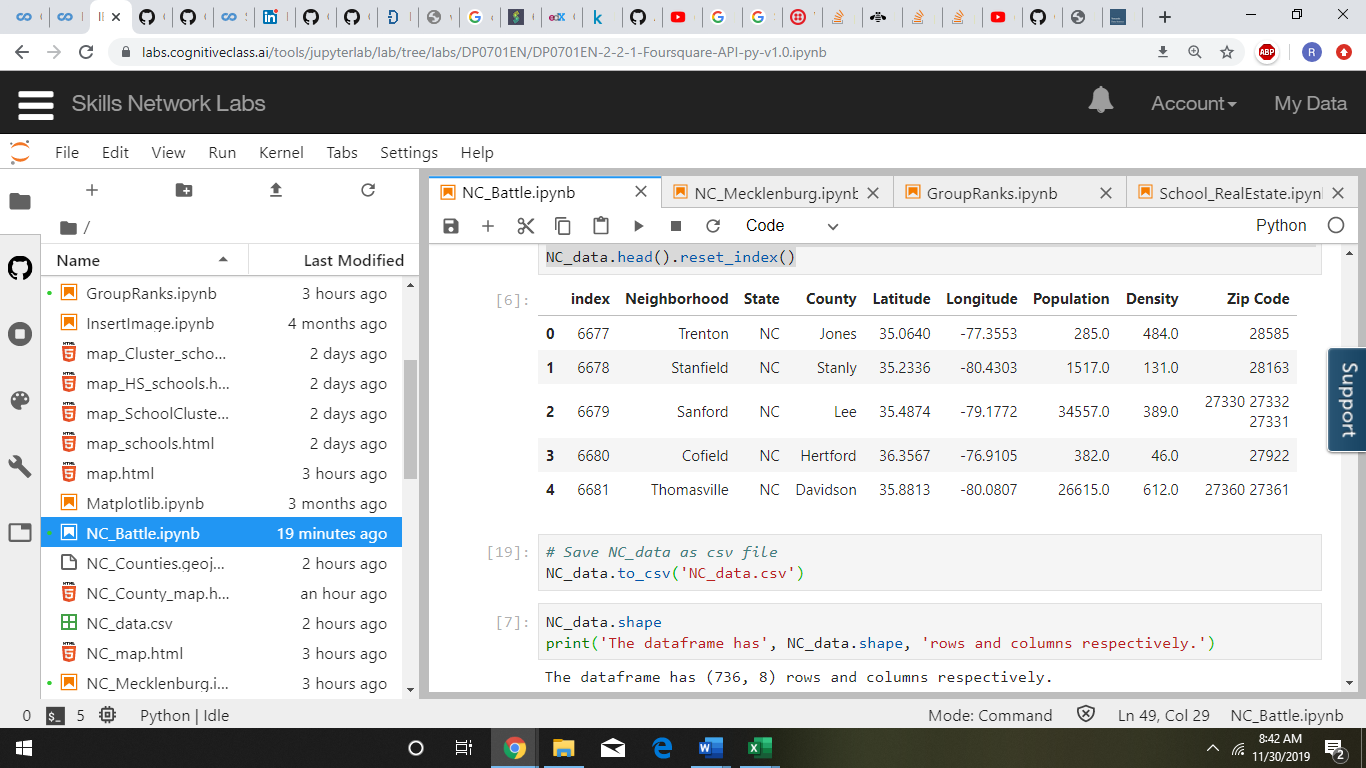
*Data Collection*- Since primary data collection is out of scope of this project, secondary data and tertiary data will be used for analysis.

Secondary data was acquired from government sites, reputable real estate sites and great schools’ rankings. In the world of internet, where most of the data is available online, following websites became the data repository for this project:

* List of Schools and School Performance Grades: http://www2.cms.k12.nc.us/cmsdepartments/accountability/Pages/PerformanceDashboardNew.aspx;
* Zip code names: https://www.zipdatamaps.com/mecklenburg-nc-county-zipcodes;
* School codes: https://nces.ed.gov/ccd/schoolsearch/school\_list.asp?Search=1&DistrictID=3702970;
* School and Neighborhood Rankings: https://www.niche.com/k12/search/best-public-elementary-schools/d/charlotte-mecklenburg-schools-nc/;
* Real estate values: https://www.zillow.com/charlotte-nc-28205/home-values/;
* Maps: https://opendata.arcgis.com/datasets/87376bdb0cb3490cbda39935626f6604\_0.geojson?where=STATE%20like%20'%25NC%25'%20AND%20CITY%20like%20'%25Huntersville%25';
* US cities data: https://simplemaps.com/data/us-cities.

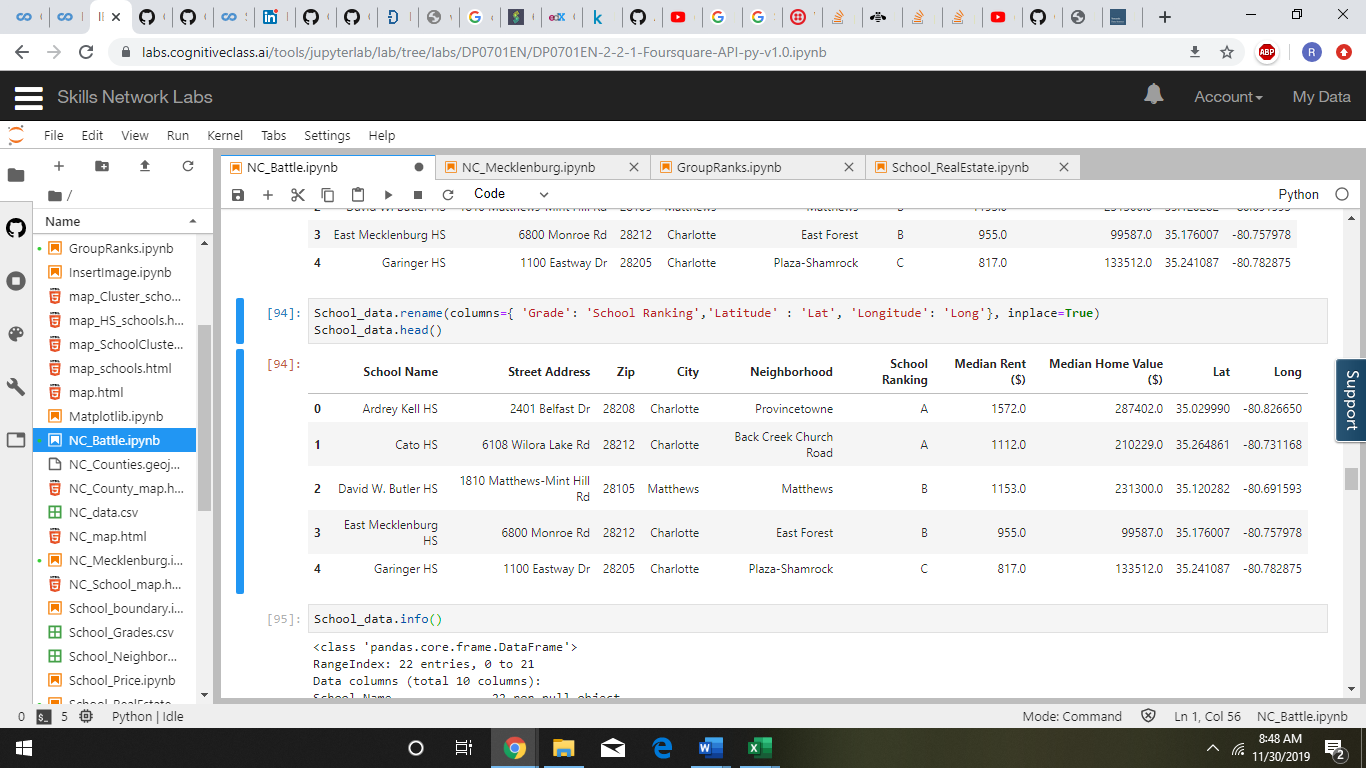
*Data Wrangling*- To get an insight of the data, data cleaning is one of the most important tasks. As per the requirement, data was filtered from the available sources. Main two tables used in the study are:

**Table 1: North Carolina Information**

Source: https://simplemaps.com/data/us-cities

**Table 1** is a cleaned data with required information on Neighborhood, County, Population, Density and Zip Code.

**Table 2: Charlotte, NC School Information**

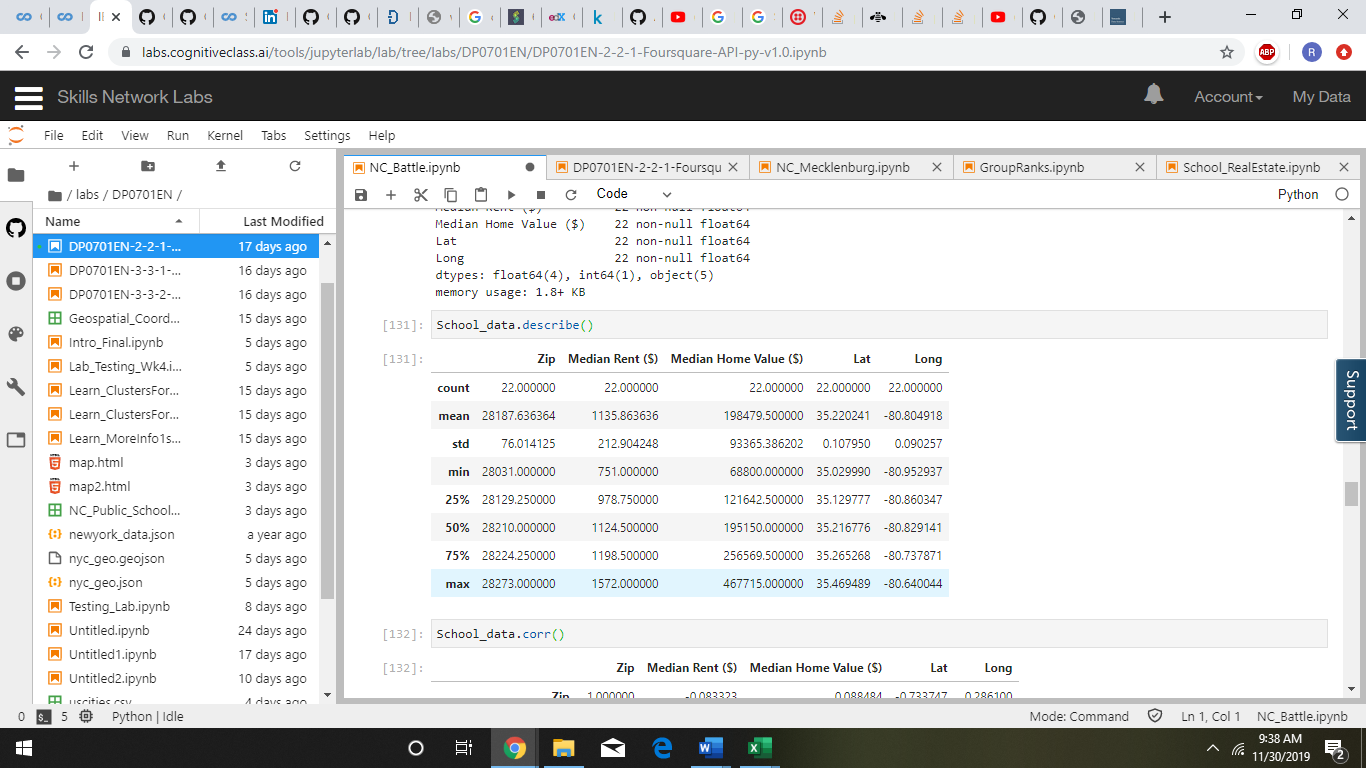


**Table 2** contains the clean data showing School Name, City, School Ranking, Median Rent ($) and Median Home Value ($) for the Neighborhoods.

**Data at a Glance- Exploratory Data Analysis**

From the available data, it seems that instead of top down approach, bottom’s up approach will be followed to check the correlation of the data. Data Visualization will play an important aspect for the inspection and at a glance overview of the available data.

**Table 3: Statistical Information on Real Estate of Charlotte, NC**



**Table 3** gives an idea about the Maximum rent and minimum rent prevalent in the neighborhoods of Charlotte and Maximum and Minimum Home Values in the neighborhoods of Charlotte.

Minimum median rent is $ 751, which is far below the State Median Rent of $ 1500 and Maximum median rent is $ 1572, which is just a little over the State value.

**Fig. 1: Range of Median Home Value ($) and Median Home Rent ($) in Charlotte, NC**

|  |  |
| --- | --- |
| **Range of Median Home Value ($)** | **Range of Median Rent ($)** |

Figure 1 give a glance of the range of Home Value and Rent prevalent in the area.

As they as, any information is appreciated. So does the above information. It just tells us the range of home values and home rent in the Charlotte, NC area. This information might help in assessing and realizing that this area is within our pocket or not but for greater details, we will have to dig deeper. For further analysis, there is need for more diagnostic analysis.

**Data Analysis- Statistical Analysis**

Any statistical inference will be of great help for finding the desired results. In order to obtain valuable information, there is need to group the schools according to school ranking obtained from greatschools.com.

**Grouping**- The data was grouped, on School Ranking, to have an idea whether our assumption of schools playing a role in real estate is of worth or not. The result was a collaborative information of the school ranking.

**Fig. 2: Grouped Median Home Value ($) and Median Rent ($) based on School Ranking**

|  |  |
| --- | --- |
|  |  |

Figure 2 gives a collaborative information that the school ranking plays a vital role in the real estate, high Home Values and Rent in the neighborhoods with good school ranking. People are willing to pay more to live in the neighborhoods with good school ranking.

Now, after a view about the price variation, it is important to locate these schools in the Charlotte area.

**Mapping**- Using geolocator, map of the North Carolina state was created.

|  |
| --- |
|  |

**Map 1: North Carolina Map**

**Table 4: Dataframe.shape**

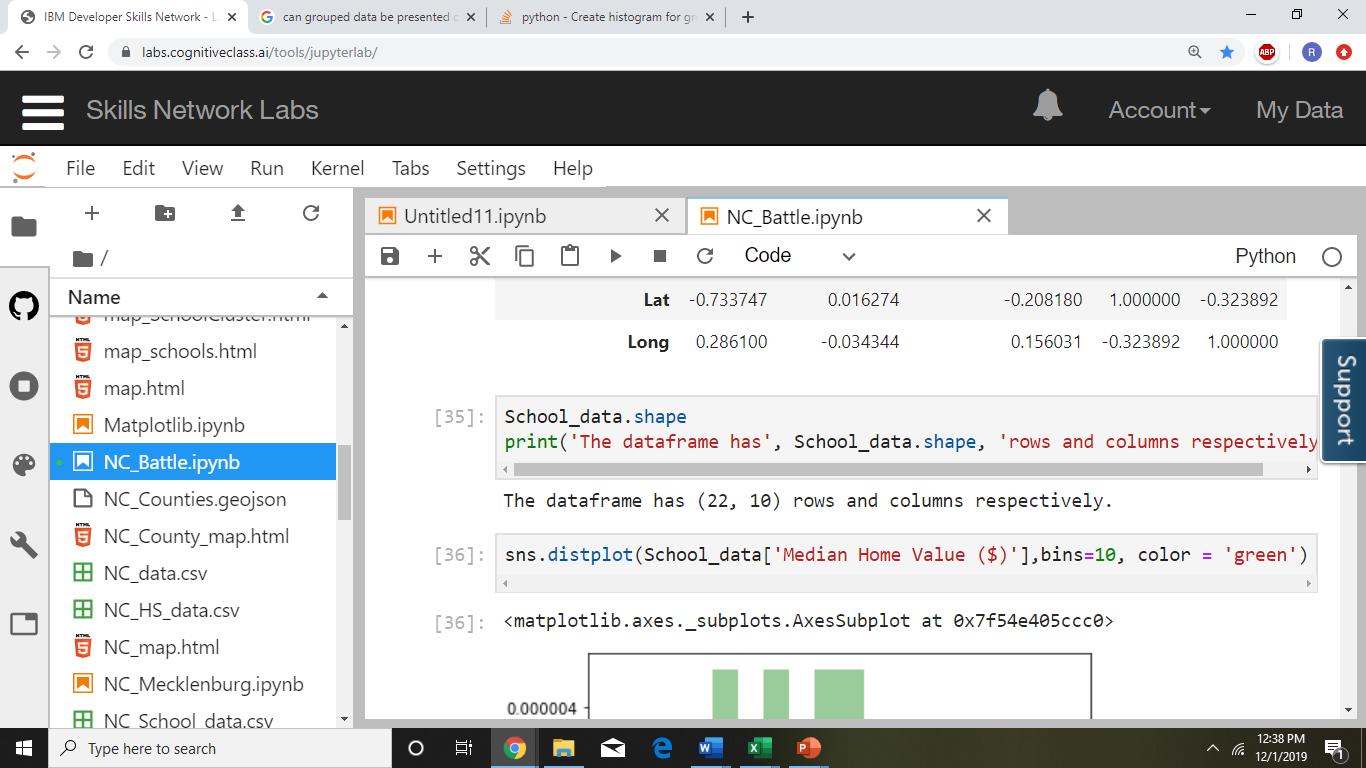


Table 4 tells that there are 22 rows, implying that there are 21 high schools that will be considered for the analysis to find best neighborhoods.

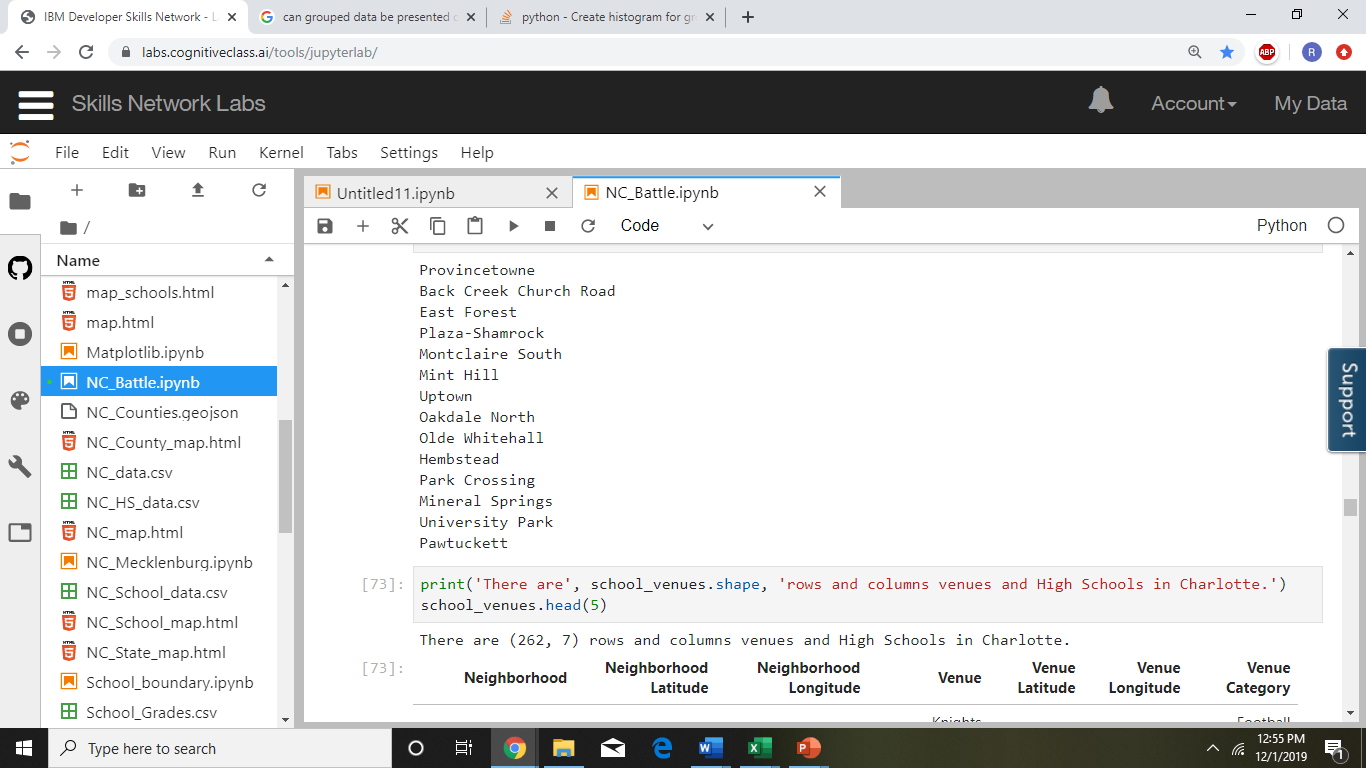
With the map of North Carolina map, Charlotte area along with the High Schools in this region was sliced so that only the area of interest can be viewed.

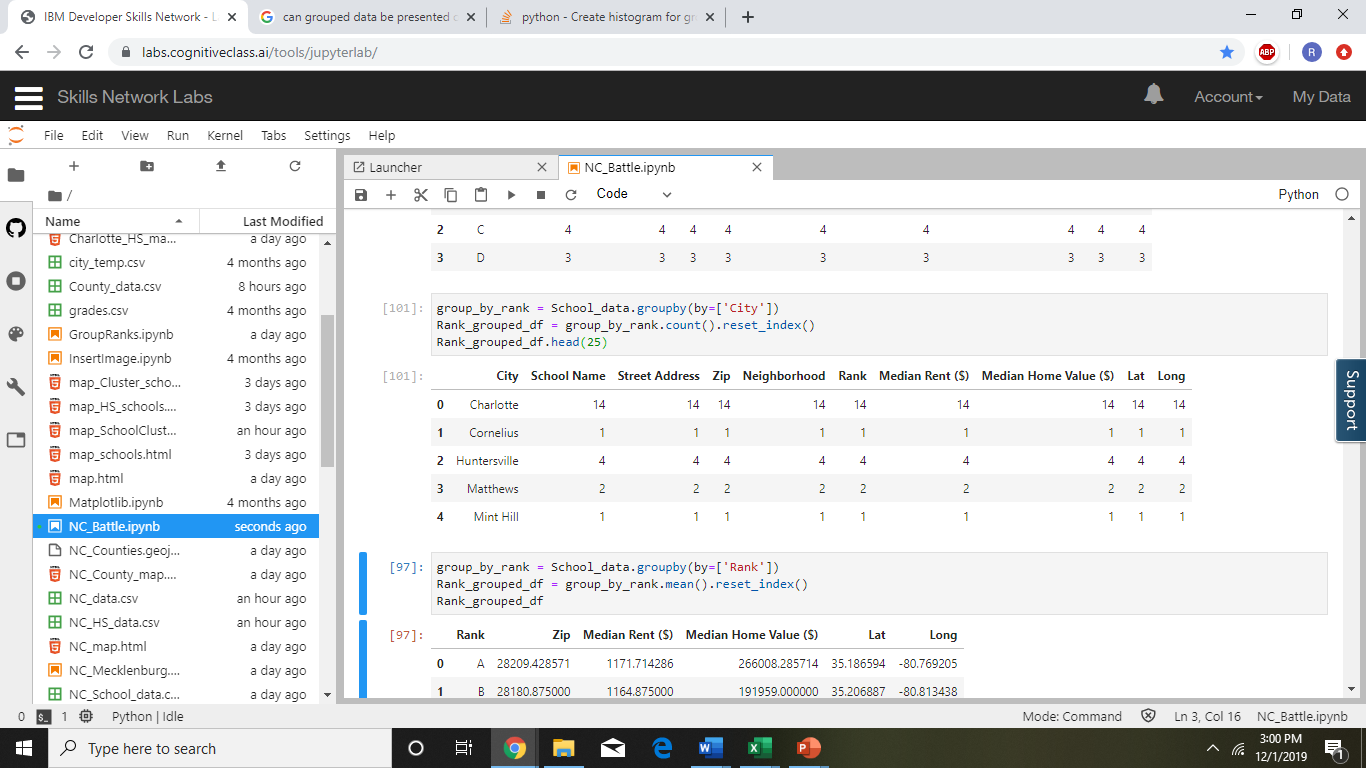
**Map 2: Charlotte High School Map**

|  |
| --- |
|  |

**Foursquare Venue Analysis**- Using the foursquare venue, it was observed that these schools are concentrated in 14 neighborhoods, namely:

**Fig. 3: Foursquare Analysis on Neighborhoods and Schools**





These neighborhoods are in these cities with schools concentrated more in Charlotte City

These neighborhoods have different characterisitics other than just the targeted high schools. We further analysed to see how many venues were there in each neighborhood.

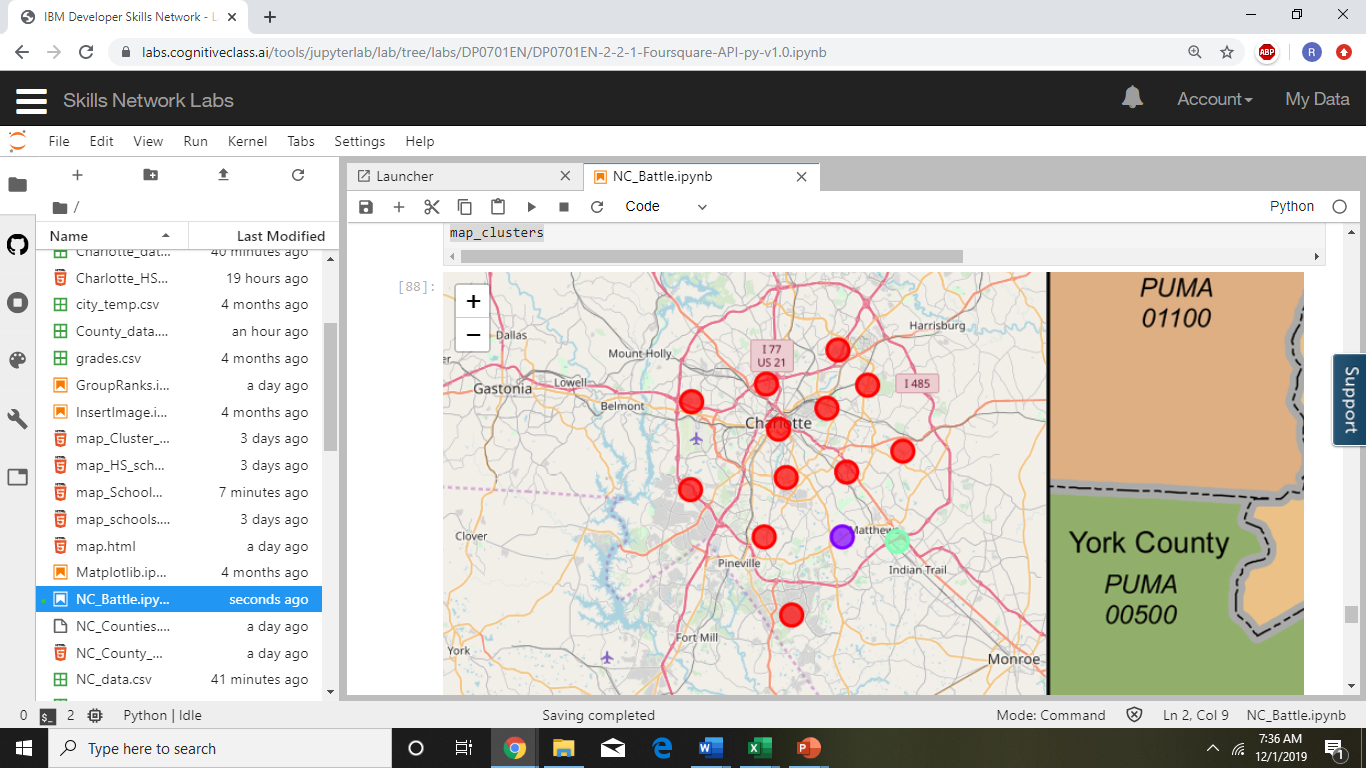
**Fig. 4: Foursquare Venues of 14 Charlotte Neighborhoods**

|  |  |
| --- | --- |
|  |  |

It is observed that there is huge gap between Uptown and the rest of the neighborhoods. Uptown is the Central Business District (CBD) of Charlotte. It has the main business area, sports arena, cultural activity area alongwith many food joints and residential area. Mint Hill is a small town in the suburbs of Charlotte which shows the least number of venues.

**KClustering**- Finally, after mrging two data sets, clustering was done with kclustering with 3 clusters to identify the best possible neighborhood. These clusters are shown in the following map:

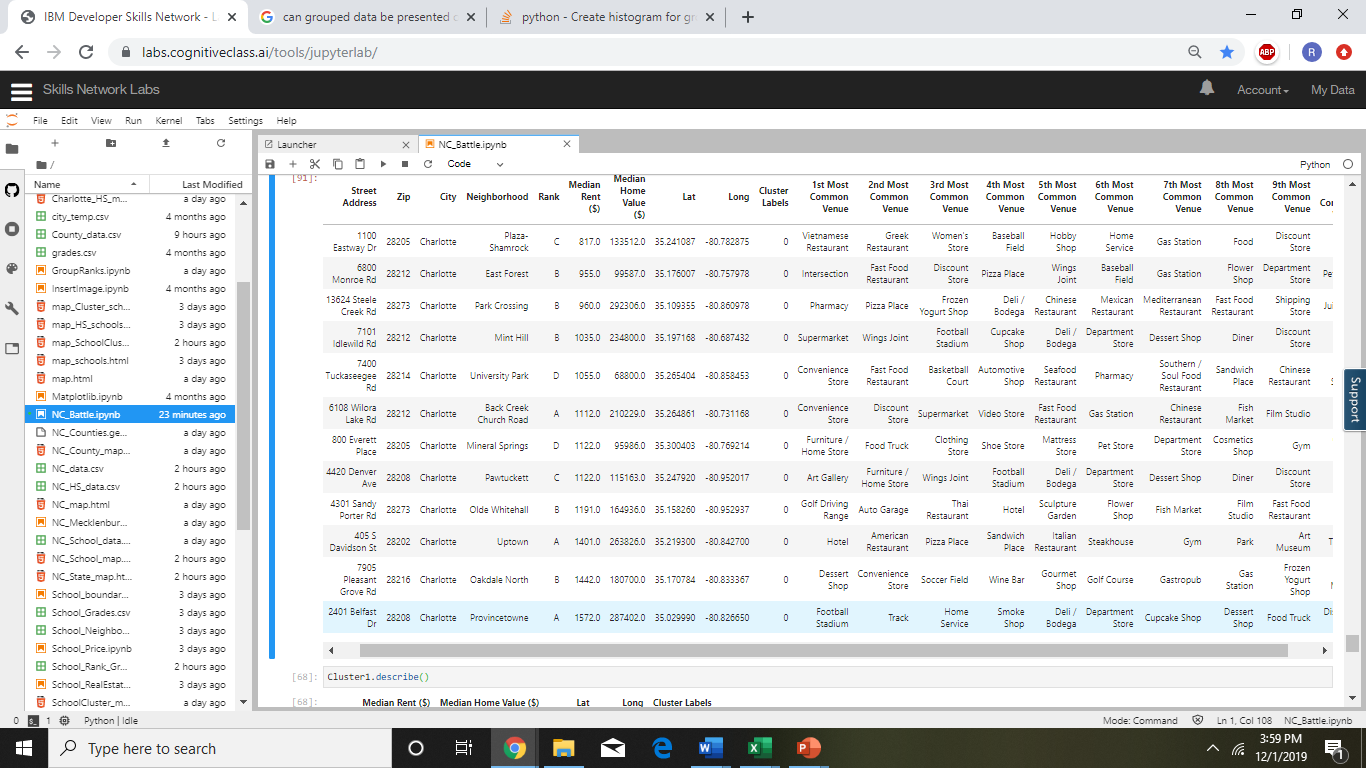
**Map 3: Clusters of Neighborhoods**

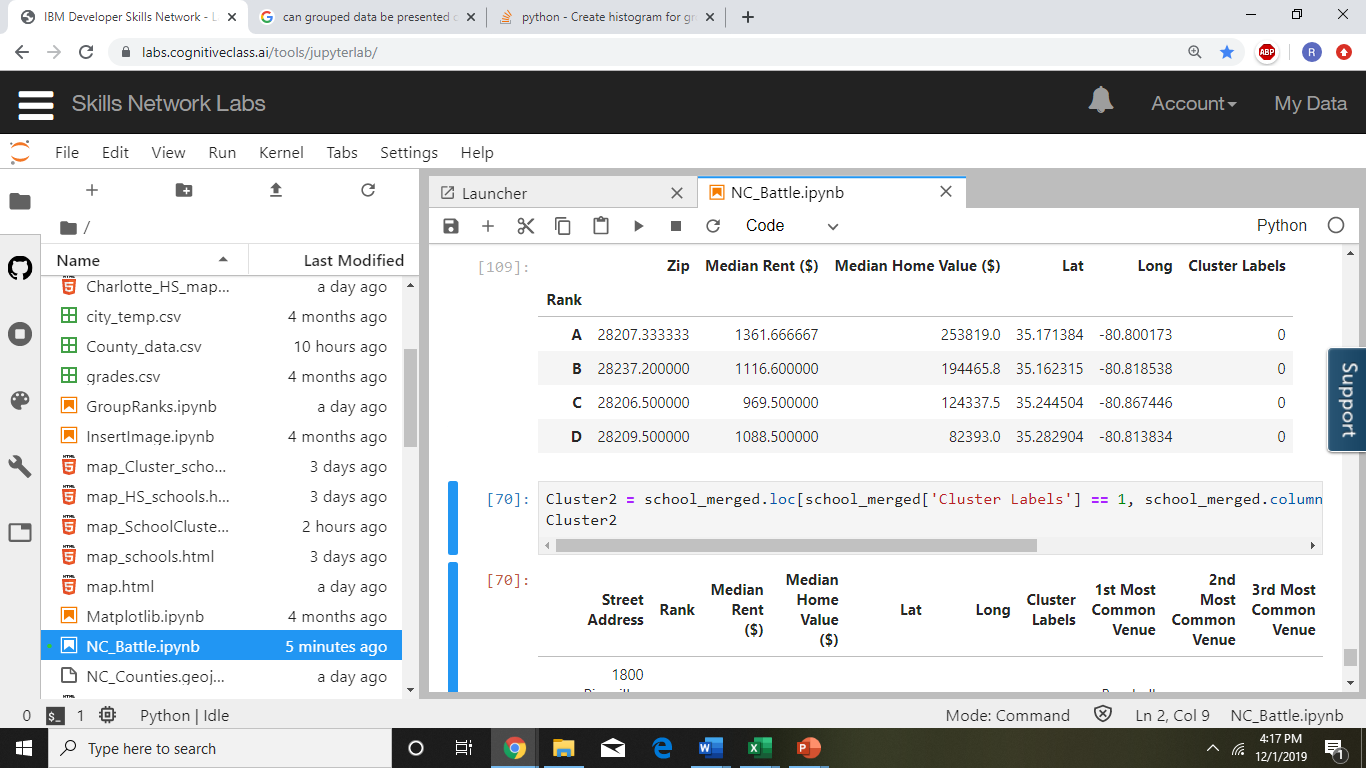
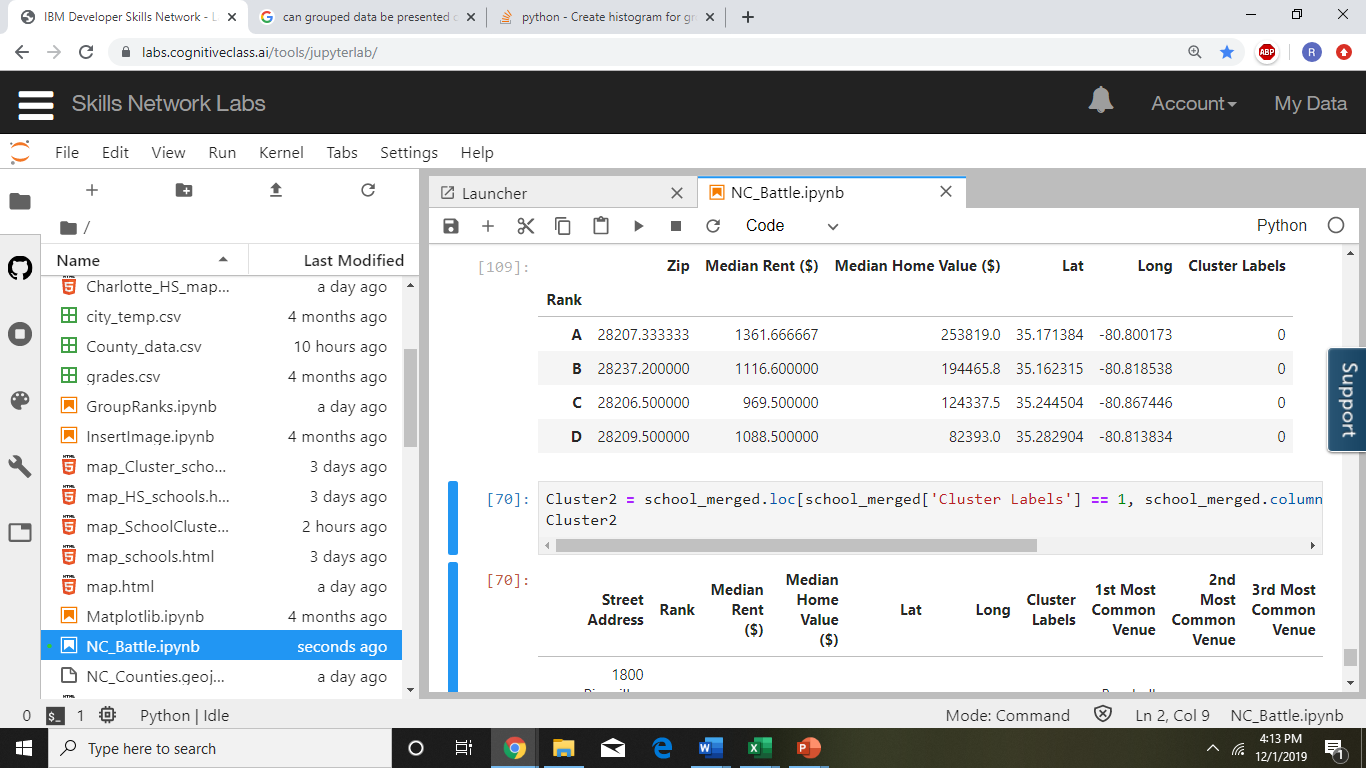
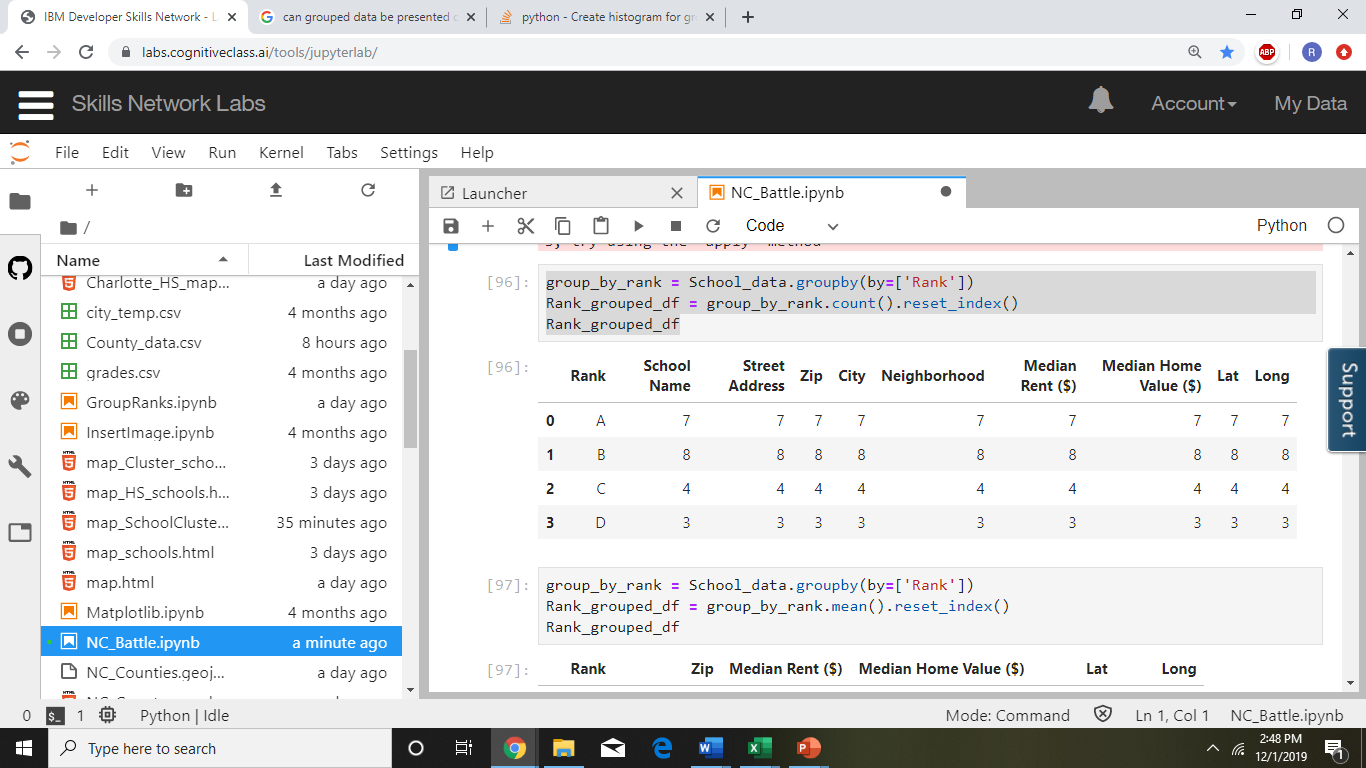


These clusters were calliberated in python to provide us with clusters that can help in narrowing our search for neighborhoods with good ranking schools.

Cluster 1:

**Table 5: Cluster 1 details**

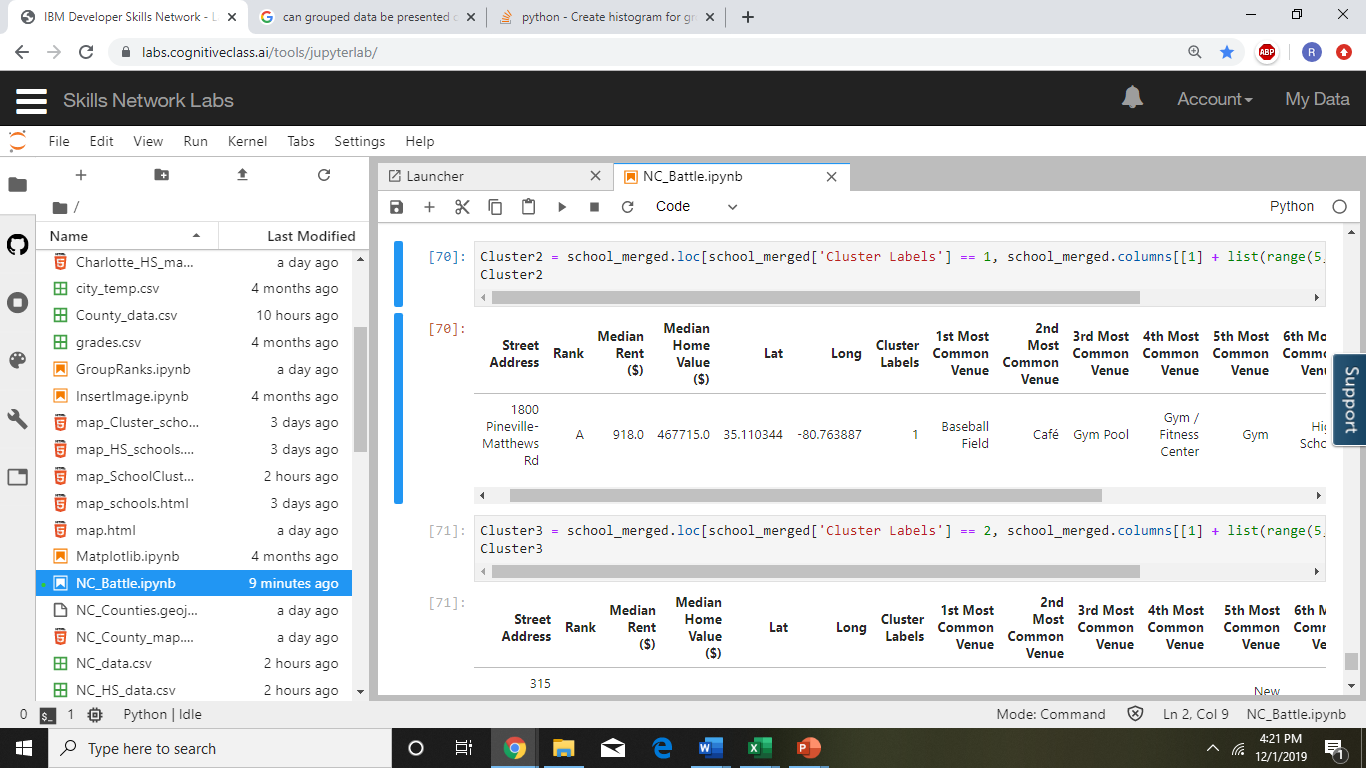




Cluster 1 has 12 schools having all ranks of high schools. Home values for Rank A schools is the maximum and the least are for schools with Rank D.

Cluster 2:

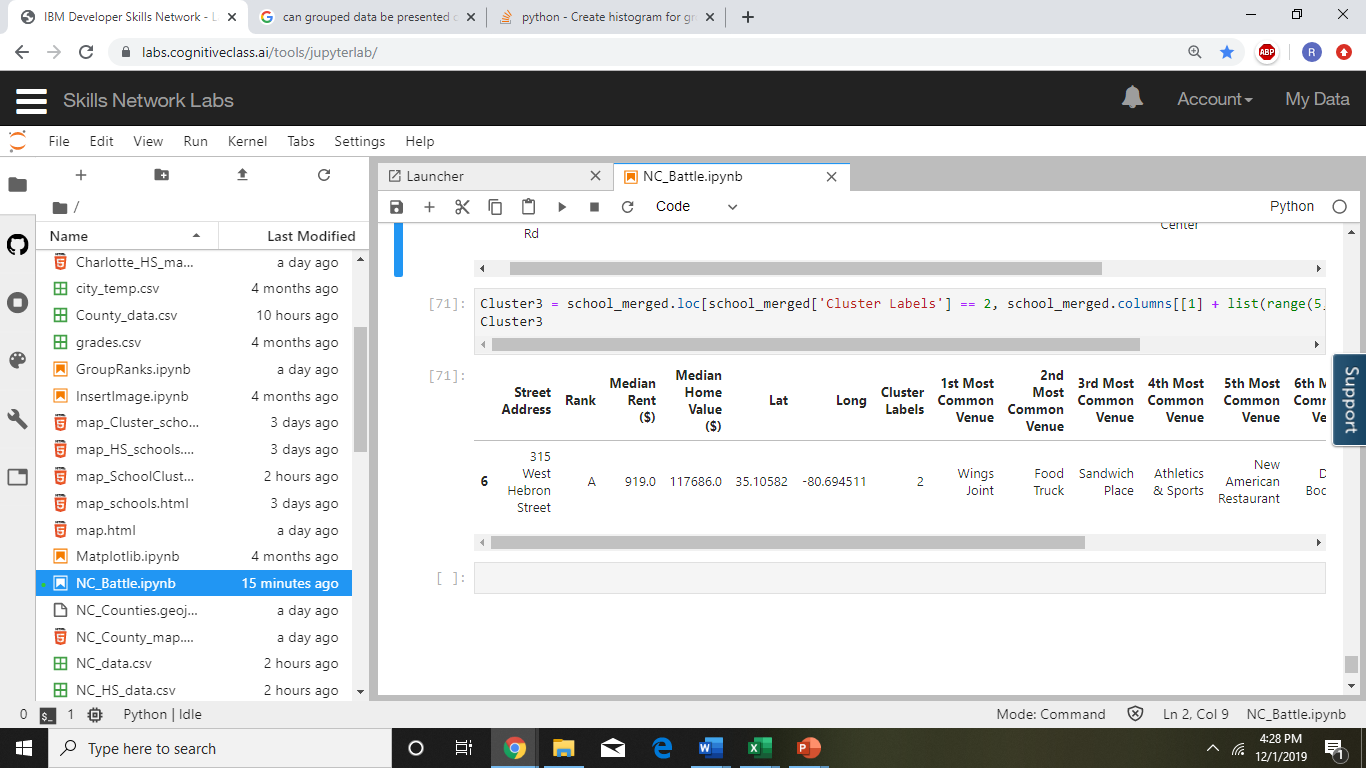
**Table 6: Cluster 2 details**



Cluster 2 consists of school with Rank A. On deep analysis, it is found that this school, Providence High School is 10 on 10 school and have received Red Ribbon awards from the President. This area has other amenities for extra curricular activities like Gym, baseball field, etc for overall development of students. Home values in this neighborhood are exhorbitantly high as compared to the whole Charlotte area whereas the rent is low. This can be attributed to the fact that there are many rental communities in this area vying for the families looking for high ranked school.

Cluster 3:

**Table 7: Cluster 3 details**



Cluster 3 also consists of school with Rank A. On deep analysis, it is found that this school, Harper Middle College High is a school offering College credits with aggregated Rank A. This area has a wide age variation and many food joints and near it. Surprisingly, the home values are below average.

**Discussion**

The whole project to find a neighborhood that would be fair for any family to relocate to Charlotte area was an interesting topic. Battling of neighborhoods would have given ample opportunity to explore the neighborhoods to find a perfect fit. But, unfortunately this could not be achieved as k-clustering randomly segregated the clusters without taking other parameters into account. Only Foursquare venue analysis cannot give enough feedback for proper analysis that could have helped finalize a neighborhood.

**Recommendations**

Based on the study, research and analysis done so far on the battle of the neighborhoods, I would recommend taking other inputs such as crime rate rather than just the location factor to the school for deciding on any neighborhood. Though Schools play an important role but they should not only be the only criteria. But there were definite findings that better ranked neighborhoods have high home values and rent.

|  |  |
| --- | --- |
|  |  |

**Conclusion**

Battle of the neighborhoods has given ample opportunity to exhibit all the tools learnt during 9 courses. This project report along with presentation gives a real-life scenario for developing as a data scientist. In my opinion, the process of digging the data and data wrangling was the biggest and most important part of this project which took lot of time and patience. Only data analysis is not enough, literature review is also one of the important parts that gives insight of the area that is being analyzed.