IBM Coursera Capstone Final Report

1. Business problem

Public schools ratings is one of the main concern for home buyers. Most parents consider schools ratings in their house search and prefer to buy a house in neighborhoods which are part of good school districts. Real estate investors, on the other hand, appreciate that schools with higher ratings positively impact the house pricing, rental opportunity, and revenue generation. This makes the first time home buyers to compete with real estate investors for houses in school districts of their choice. Unfortunately, many first time home buyers have been forced out of the market recently because they are usually priced out by real estate investors and they eventually stop their house searching as they are not aware of other schools with their desired points of interests nearby them.

Our client is non-profit counseling organization in Durham, North Carolina that helps first time home buyers. This organization is interested in providing insightful information to home buyers about schools and other venues of interest such as malls, shops, restaurants etc. nearby each school so that they have a better understanding of other available options and make an informed decision. Unfortunately, schools rating are not usually listed in map or geo providers such as Google Maps or Foursquare. So the **business problem** is: Where are the schools located in city of Durham and what are their ratings and top venues nearby them?

2. Data

The data for this project was acquired from two different sources:

1. Foursquare was used to retrieve the different venues around schools and their respective categories.

2. City-Data.com was used to retrieve the list of schools and their rating which is available here: http://www.city-data.com/schools-dirs/schools-NC.html

Figure 1. City-Data database listing schoocls in North Carolina

3. Methodology and results

The list of schools in North Carolina was parsed from City-Data.com.

```
#URL of the web page that has all schools rating.
url = 'http://www.city-data.com/schools-dirs/schools-NC.html'
```

```
#Parse the 7th table on the web page and convert "-" values in Rationg to NaN.
table = pd.read_html(url, na_values = ['-'])[7]
table.head(7)
```

]:

	Name	Address	Rating	Grade span
0	A G Cox Middle	2657 Church Street, Winterville	50.0	6-8
1	A H Snipes Academy of Arts/Des	2150 Chestnut Street, Kings Grant	9.0	KG-5
2	A L Brown High	415 E 1st Street, Concord	59.0	9-12
3	A L Stanback Middle	3700 NC 86 South, Hillsborough	69.0	6-8
4	A T Allen Elementary	3939 Abilene Rd, Concord	69.0	KG-5
5	Aberdeen Elementary	503 Sandhills Blvd N, Aberdeen	21.0	3-5
6	Aberdeen Primary	310 Keyser St, Aberdeen	NaN	PK-2

```
#The city-data.com has other 25 pages listing the schools. We create a loop to parse all and save them.
url_core = 'http://www.city-data.com/schools-dirs/schools-NC'
i = 1
while (i < 26):
    url_pages = url_core + str(i) + '.html'
    table = table.append(pd.read_html(url_pages, na_values = ['-'])[7])
    i = i + 1
    print(url_pages)
```

Figure 2. Parsing list of schools and rating

Then schools outside Durham and also schools without rating were removed. Then the coordinates of schools were retrieved using geocode and the index was reset to prepare the working dataframe.

	Name	Address	Rating	Grade span	Latitude	Longitude
0	Bethesda Elementary	2009 S Miami Boulevard, Durham, NC	18.0	KG-5	35.9407	-78.8378
1	Brogden Middle	1001 Leon Street, Durham, NC	37.0	6-8	36.0245	-78.9093
2	Central Park School For Child	724 Foster Street, Durham, NC	74.0	KG-5	36.0046	-78.901
3	Chewning Middle	5001 Red Mill Road, Durham, NC	21.0	8	36.0717	-78.8235
4	Club Boulevard Elementary	400 W Club Boulevard, Durham, NC	22.0	KG-5	36.0192	-78.8966
5	Durham School of the Arts	401 North Duke Street, Durham, NC	75.0	6-12	36.0025	-78.9064
6	Durham's Performance Learning Center	401 North Driver Street, Durham, NC	8.0	8-12	35.9896	-78.8795
7	E K Powe Elementary	913 9th Street, Durham, NC	15.0	KG-5	36.0133	-78.9222
8	Easley Elementary	302 Lebanon Circle, Durham, NC	74.0	KG-5	36.0754	-78.9269
9	Eastway Elementary	610 North Alston Avenue, Durham, NC	11.0	KG-5	35.9912	-78.885
10	Eno Valley Elementary	117 Milton Road, Durham, NC	25.0	KG-5	36.0854	-78.913

Figure 3. Dataframe includes schools names, address, rating, grade span, and coordinates.

The schools were plotted over the map of Durham, showing schools as circles with different colors depending on their rating.
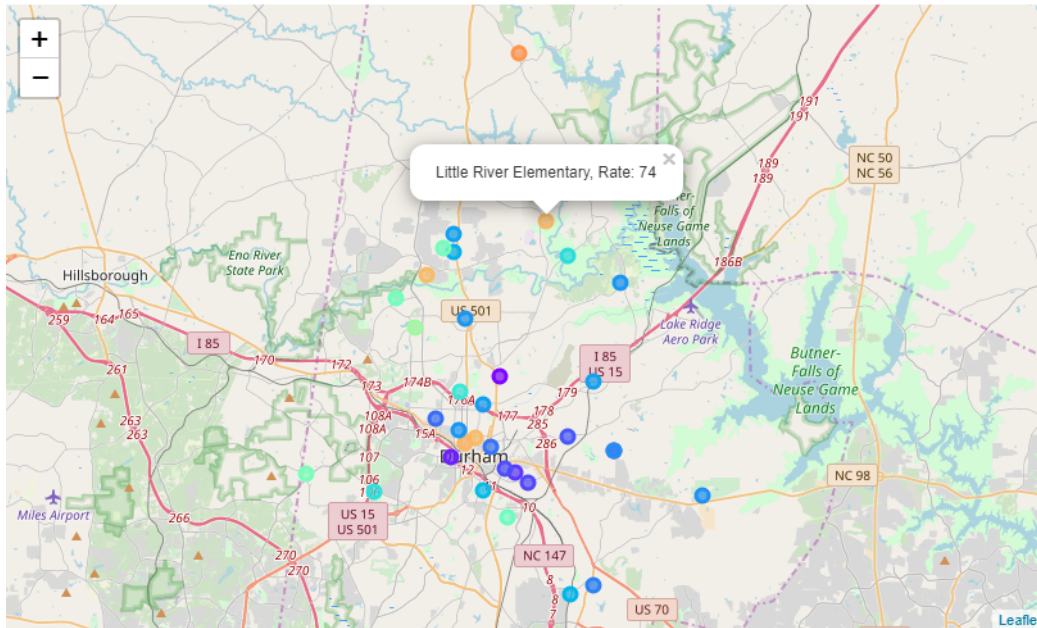

Figure 4. Schools of Durham with pop up box showing schools name and rating

Then we used **KMEANS** to cluster the schools into 5 classes only based on their rating.
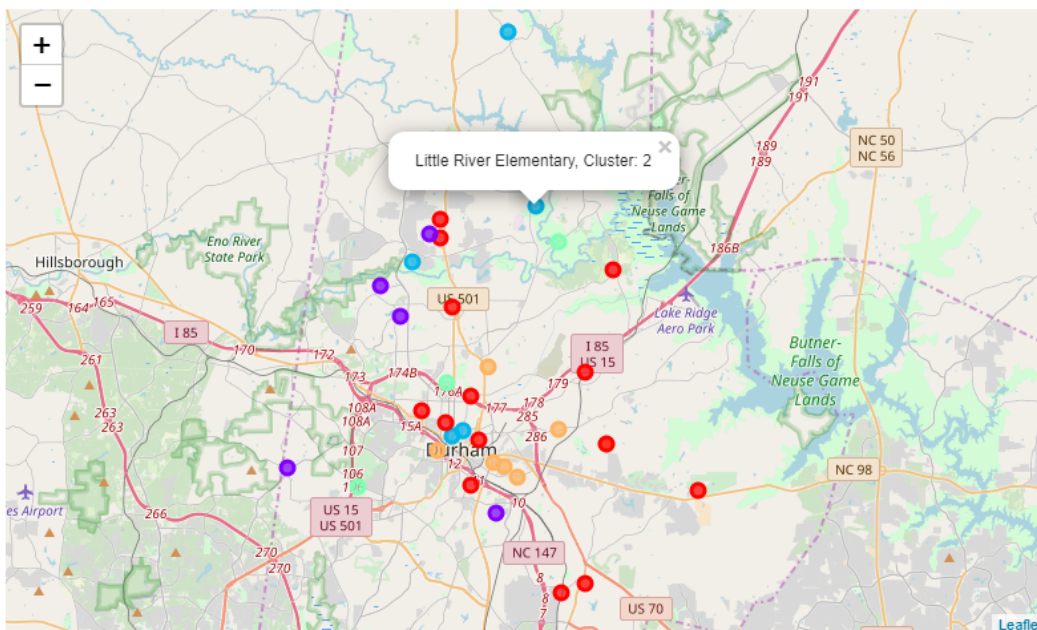

Figure 4, Schools clustered based on their rating

Then we used **FOURSQUARE API** to explore the top 100 venues within 1 km of each school and create a dataframe with the name of top 10 categories close to each school.

	School	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bethesda Elementary	Fast Food Restaurant	Pawn Shop	Moving Target	Lingerie Store	Basketball Court	Fried Chicken Joint	Park	Japanese Restaurant	Construction & Landscaping	Diner
1	Brogden Middle	Clothing Store	Pizza Place	Shoe Store	Rental Car Location	Lingerie Store	Jewelry Store	Pharmacy	Gym / Fitness Center	Park	American Restaurant
2	Central Park School For Child	Bar	Hotel	Pizza Place	Gastropub	American Restaurant	Mexican Restaurant	Italian Restaurant	Brewery	Pub	Deli / Bodega
3	Chewning Middle	Garden	Zoo Exhibit	Ethiopian Restaurant	Flower Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Farm	Fabric Shop	Exhibit
4	Club Boulevard Elementary	Convenience Store	Sandwich Place	Park	Fast Food Restaurant	Chinese Restaurant	Peruvian Restaurant	Theme Park Ride / Attraction	Big Box Store	Dog Run	Food Truck
5	Durham School of the Arts	Bar	Hotel	Pizza Place	Brewery	American Restaurant	Mexican Restaurant	Gastropub	Italian Restaurant	Seafood Restaurant	Bakery
6	Durham's Performance Learning Center	Convenience Store	Fast Food Restaurant	Grocery Store	Gas Station	Hardware Store	Bakery	Park	Discount Store	Café	Food
7	E K Powe Elementary	Café	Record Shop	Bakery	Gastropub	Sandwich Place	Pizza Place	Shipping Store	Bookstore	Yoga Studio	Ice Cream Shop
8	Easley Elementary	Office	Candy Store	Intersection	Trail	Fast Food Restaurant	Farmers Market	Farm	Fabric Shop	Exhibit	Event Space
9	Eastway Elementary	Coffee Shop	Brewery	Building	Café	Food Truck	Garden	Spa	Gas Station	Tattoo Parlor	Park
10	Eno Valley Elementary	Mexican Restaurant	Video Store	Pizza Place	Food & Drink Shop	Basketball Court	Discount Store	Chinese Restaurant	Pharmacy	Sandwich Place	Gym

Figure 5. Top 10 venues near each school

Then rating of each school was added to the dataframe containing frequency of each venue category presence near each school. Using **KMEANS**, schools were clustered this time not only based on their rating, but also the category of top 10 venues near them.
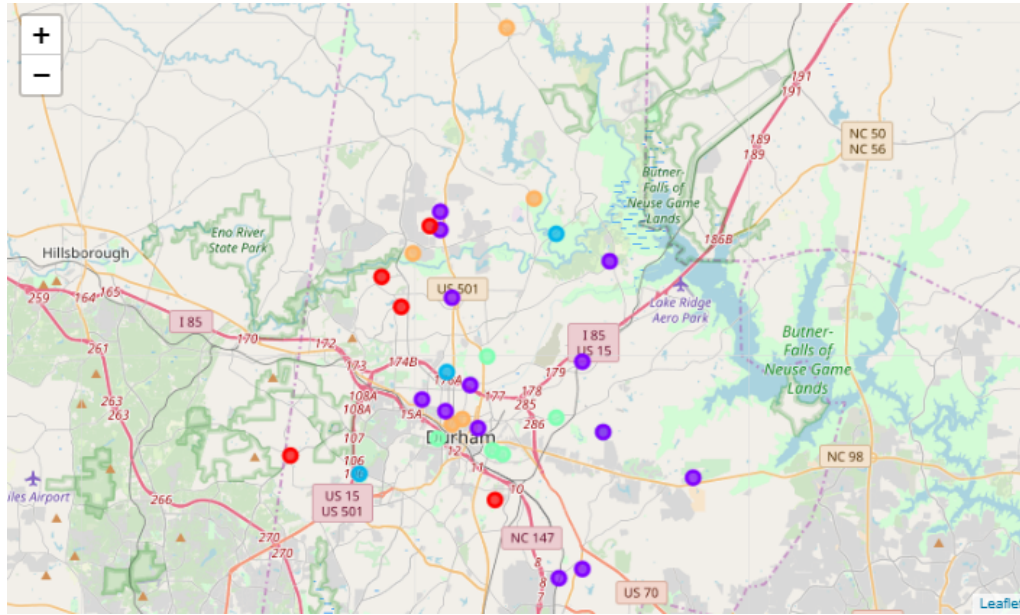
Figure 6. Schools clustered based on their rating and top ten venues close to them.

Then a final table was prepared for the customer listing schools name, grade span, address, rating, cluster label, and all 10 most common venues around them.

	School	Grade Span	Address	Rating	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bethesda Elementary	KG-5	2009 S Miami Boulevard, Durham, NC	18	1	Fast Food Restaurant	Pawn Shop	Moving Target	Lingerie Store	Basketball Court	Fried Chicken Joint	Park	Japanese Restaurant	Construction & Landscaping	Diner
1	Brogden Middle	6-8	1001 Leon Street, Durham, NC	37	2	Clothing Store	Pizza Place	Shoe Store	Rental Car Location	Lingerie Store	Jewelry Store	Pharmacy	Gym / Fitness Center	Park	American Restaurant
2	Central Park School For Child	KG-5	724 Foster Street, Durham, NC	74	4	Bar	Hotel	Pizza Place	Gastropub	American Restaurant	Mexican Restaurant	Italian Restaurant	Brewery	Pub	Deli / Bodega
3	Chewning Middle	8	5001 Red Mill Road, Durham, NC	21	1	Garden	Zoo Exhibit	Ethiopian Restaurant	Flower Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Farm	Fabric Shop	Exhibit
4	Club Boulevard Elementary	KG-5	400 W Club Boulevard, Durham, NC	22	1	Convenience Store	Sandwich Place	Park	Fast Food Restaurant	Chinese Restaurant	Peruvian Restaurant	Theme Park Ride / Attraction	Big Box Store	Dog Run	Food Truck
5	Durham School of the Arts	6-12	401 North Duke Street, Durham, NC	75	4	Bar	Hotel	Pizza Place	Brewery	American Restaurant	Mexican Restaurant	Gastropub	Italian Restaurant	Seafood Restaurant	Bakery

Figure 7. The final table showing schools name, grade span, address, rating, cluster label, and all 10 most common venues around them.

4. Discussion

After clustering schools based on only their rating and also other top venues around them, we were able to separate Durham schools dataset into 5 distinct clusters. A potential home buyer is able to refer to the final table and see the schools and their rating, address and top venues around them and also see what schools are within the same cluster.

5. Conclusion

We were able to provide a table showing similar schools in terms of rating and other top venues around them so that potential home buyers would have a better perspective about the area where they could start their house search which is more aligned with their criteria.