Exploring Schools in a Neighborhood

IBM - Coursera Capstone

Introduction

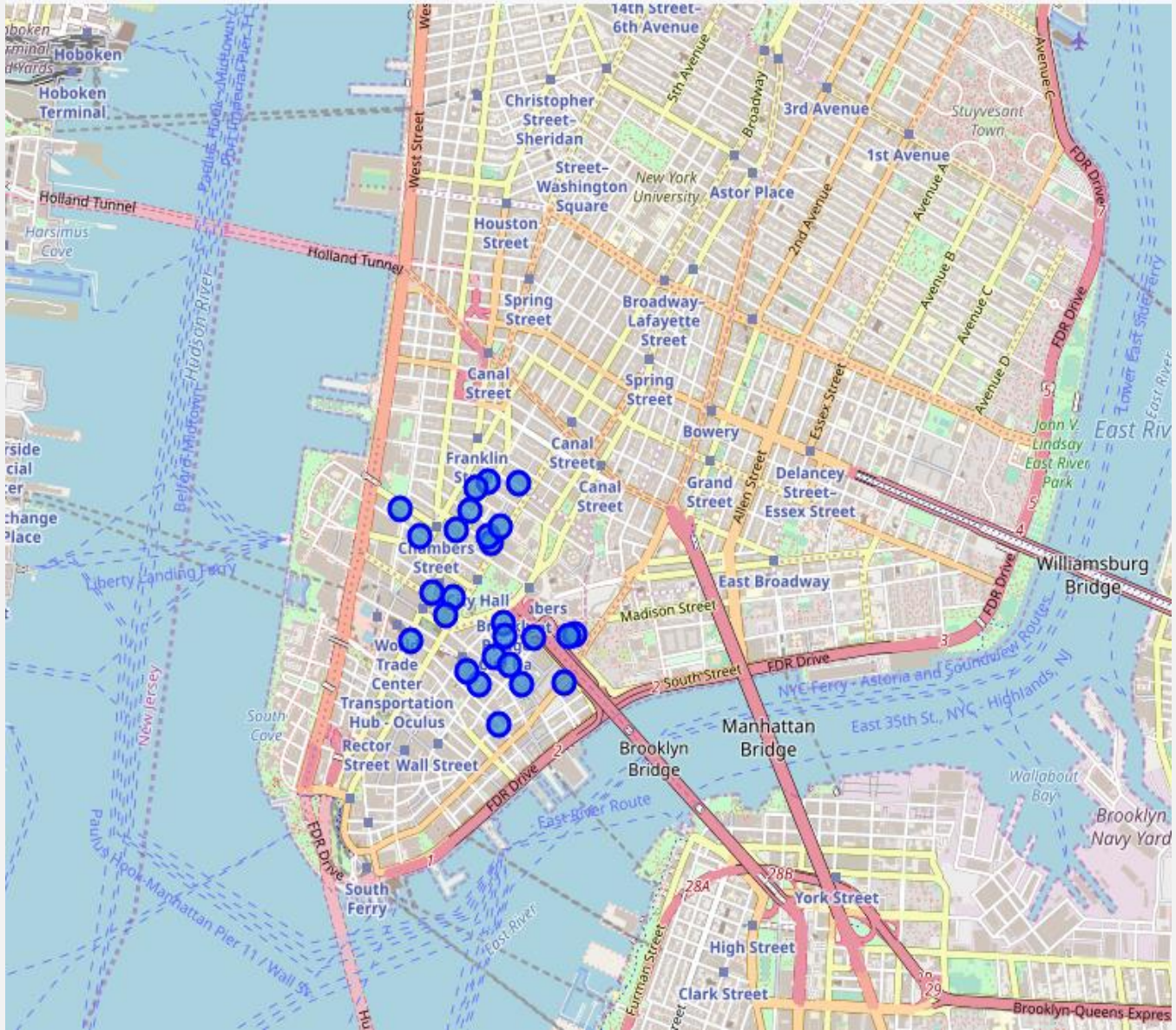
- Getting into the right institution means better knowledge and better outgrowth in career
- Distance and Quality of education provided by schools are vital parameters
- Longer the distance, longer is the travel time and larger the transportation costs
- Effective data model required to predict the best schools in a neighbourhood.

Data Acquisition and Wrangling

- Location data for New York schools retrieved from Foursquare API
- Latitude and Longitude determined by Geocode package
- Irrelevant columns and Nan values removed from dataset
- Total shape of the final table was 25 rows and 7 columns.
- Generated user data list with home location and rating as columns.

Schools in New York

- Map representing schools in New York City
- Folium package used for this representation
- Schools located nearby to each other within distance variations of 0.8 Kms
- This was evident to apply the clustering technique to get the places as per the user requirement



Data Analysis and Modelling

- Data in Maps were categorical and had less Numerical values
- EDA and classification techniques were out of scope
- Clustering techniques used since data is unlabelled
- Also, to split data in the form of different clusters based on the category of schools selected
- One-hot encoding used to get each category as columns
- K-Means Modelling chosen for determining Euclidean distance between each school category

Model Evaluation

User List

	Home Location	School Rating
0	Barclay	8
1	William St	9
2	Greenwich St	7

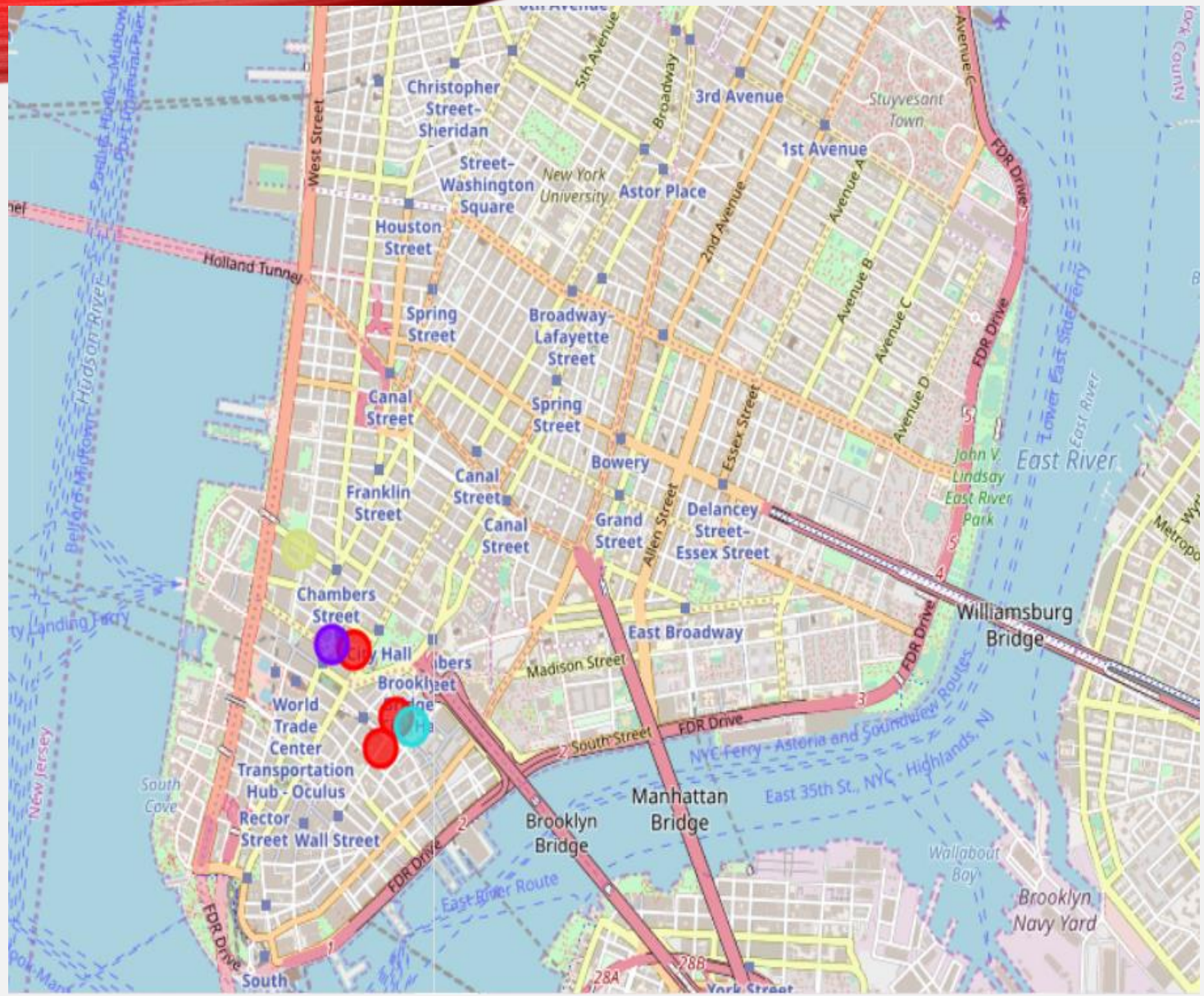
School List

	Category	Cluster Labels	id	location.address	location.distance	location.lat	location.lng	name
0	Building	0	4a78d55cf964a5208be61fe3	15 Barclay	194.0	40.712634	-74.008312	NYU School of Professional Studies
1	College Academic Building	0	4bbfa098f8219c748850b010	163 William St	276.0	40.710245	-74.005968	Seidenberg School Of CSIS
9	High School	2	4c64536211c4a593040ce911	156 William St	318.0	40.709950	-74.005116	Hawthorne Country Day School
10	School	3	4a6eff92f964a5202cd51fe3	292 Greenwich St	593.0	40.716253	-74.011296	PS 234 - Independence School
18	Nursery School	1	4fcf8f59e4b02b4665bbea22	6 Barclay St	295.0	40.712818	-74.009513	Barclay Street School
23	Office	0	53cd804e498ed3afa72314a8	123 William St	401.0	40.709171	-74.006828	Library Journal/ School Library Journal/ Junio...

User Input list with the required parameters

- *Table generated by the Data Model based on the user inputs*
- *Home location from user list matched with the final table*
- *Categories listed include only schools.*

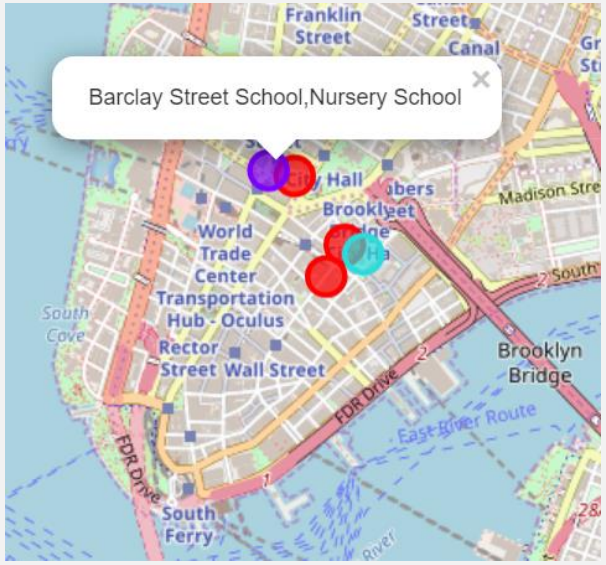
Final User Output in Folium



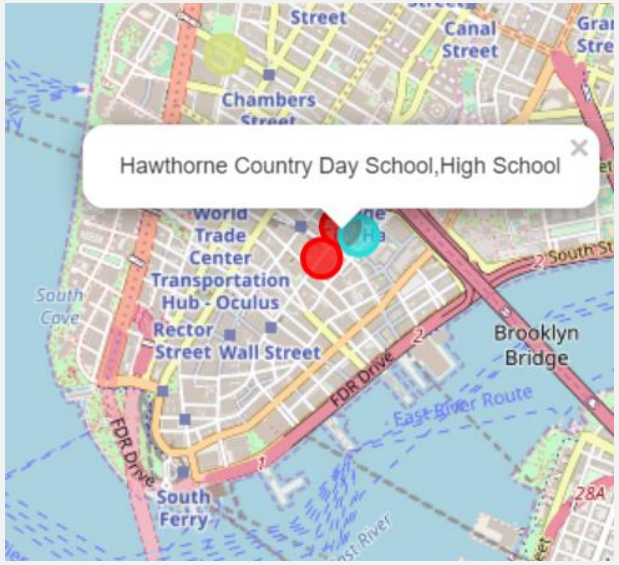
- Map represents set of schools in the User's locality
- Each colour encodes to one cluster type.
- This model reduces the complexity of looking at multiple schools at a time
- Provides better readability to user due to popups and colour encoding

K-Means Data Model

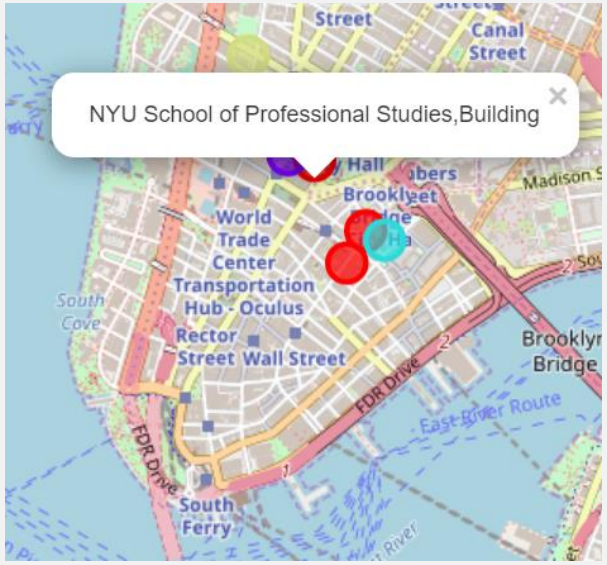
- K-Means Model generated for 4 different types of Clusters.



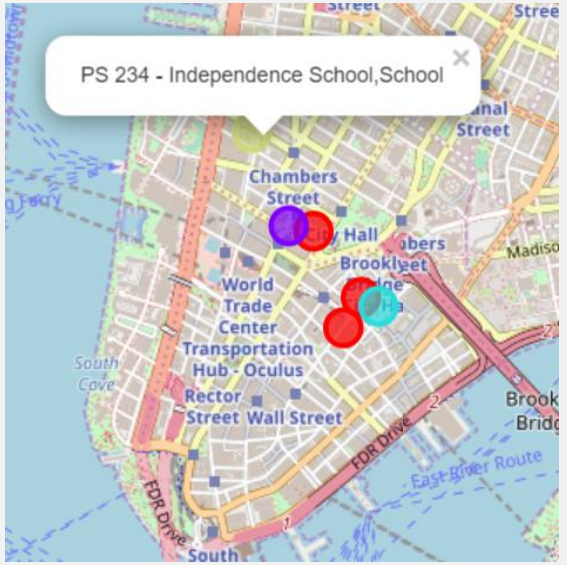
Cluster 1 – Nursery School



Cluster 2 – High School



Cluster 3 – Others



Cluster 4 – School

Conclusion and Future Scope

- Analysed and predicted the various schools that would be preferred by the User in a neighbourhood
- This model would provide a better insight of the options ,people have within their locality
- Model can be enhanced to determine the schools based on likes/rating to provide more efficient results
- With the ratings, the model can provide the exact school for the User

Presentation

By

Smruthy