APPLIED DATA SCIENCE CAPSTONE

SCHOOL FINDER

PROJECT BY-

AKANKSHA YADAV

INDEX

INTRODUCTION

“Among world’s most dense cities, Mumbai stands still at two” says The UN Habitat data analyses. Mumbai has a population over 20,411,274 as reported by populationu.com.

India counts for more than 315 million students. With respect to population, in Mumbai there are up to 100 million students. Each student needs education, for which nearly every one of them attends a school. Observing the population of students in mumbai plus their need for school, they need best possible education.

School is one of the most important things in our lives! If we are not educated then we do not have the skills needed to function in everyday society. School not only helps students learn necessary concepts and skills, but it also allows students to interact with other students socially, academically, and emotionally.

School boosts confidence and teaches us to establish and maintain friendships, and helps us learn how to work together as a team, which is a primary tenet of any successful society. Without school, knowledge would not spread as quickly, and our access to new ideas and people could easily be cut off.

Mumbai has an area of about 4355 sq. km which makes it difficult to locate the best possible school for students. For students and parents, selecting best school is difficult due to large area in Mumbai and many other factors due to which every year many of them faces these issues.

Business Problem: The problem here arises that which school is preferred by your child according to your family income and child's interest. Plus which education board it is and in which area it is located are the factors a parent must consider while searching for a school for his/her child. It will be utterly problematic to search for each school and compare these factors. Also as Mumbai is big city with huge population, schools for children should be chosen in such a way that travelling is easier. Being Mumbai, monsoon season causes much problem in travelling as everyone knows. This idea can solve the problems faced by parents and students by helping them choose the best school based on their nearest proximity. Hence, this project aims to collect data of the schools of Mumbai and locate them with their characteristics.

DATA ANALYSIS

Data Collection/Gathering: The data used was the location of schools that was acquired using the **foursquare** website. To gather the data, foursquare API was used along with the foursquare credentials Client ID and Client Secret. A 'search' query was made in the **IBM Watson Studio** with **Python kernel**, so as to search the schools. Through the website, coordinates of schools in a particular location were generated and processed.

Data Preprocessing and Wrangling: Using the modules of python, only valid and usable data was selected from the JSON file generated by foursquare and data-frame was created using **'Pandas'**. Since the project required only the locations, the '**venues**' section under the '**response**' section was selected. The generated data-frame still had numerous of unwanted data inside the 'venues' section, which needed to be filtered and cleaned.

3]:

	categories	hasPerk	id	location.address	location.cc	location.city	location.country	location.crossStreet	location.distance	location.formattedAd
0	[{'id': '4bf58dd8d48988d13b941735', 'name': 'S...	False	4dc4e64b2271f270511faaf0	Sai Hill Rd.	IN	Mumbai	India	NaN	1551	[Sai Hill Rd., M 400024, Mahārāshtra,
1	[{'id': '4bf58dd8d48988d198941735', 'name': 'C...	False	5c46fb1e3092be002c901113	1st Floor, Kohinoor Education Complex, Kurla West	IN	Mumbai	India	Kurla	992	[1st Floor, Ko Education Complex, K
2	[]	False	4f9bfabfe4b0027bd481213c	NaN	IN	NaN	India	NaN	356	
3	[{'id': '4bf58dd8d48988d1ad941735', 'name': 'T...	False	5199db81498ec7ed165a1a8a	NaN	IN	NaN	India	NaN	505	
4	[{'id': '4bf58dd8d48988d13b941735', 'name': 'S...	False	55afa0ff498e688093b22bf2	Kohinoor Education Complex, Kohinoor City	IN	Mumbai	India	NaN	769	[Kohinoor Edu Complex, Kohinoor Cit

Data Filtering: There are many irrelevant columns in the dataset. So remove the irrelevant columns as it will be of no use. This is known as filtering of data. So we removed columns like categories, hasPerk, id, location.cc, referral id, etc

	name	location.address	location.city	location.distance	location.lat	location.lng
0	Nehru Nagar Municipal School	Sai Hill Rd.	Mumbai	1551	19.062062	72.877558
1	janseva motor training school	NaN	NaN	356	19.073660	72.880009
2	School of Communications & Reputation	1st Floor, Kohinoor Education Complex, Kurla West	Mumbai	992	19.075196	72.887099
3	Kohinoor Business School	NaN	NaN	505	19.076290	72.882491
4	Kohinoor Business School	Kohinoor Education Complex, Kohinoor City, Kir...	Mumbai	769	19.073995	72.870699

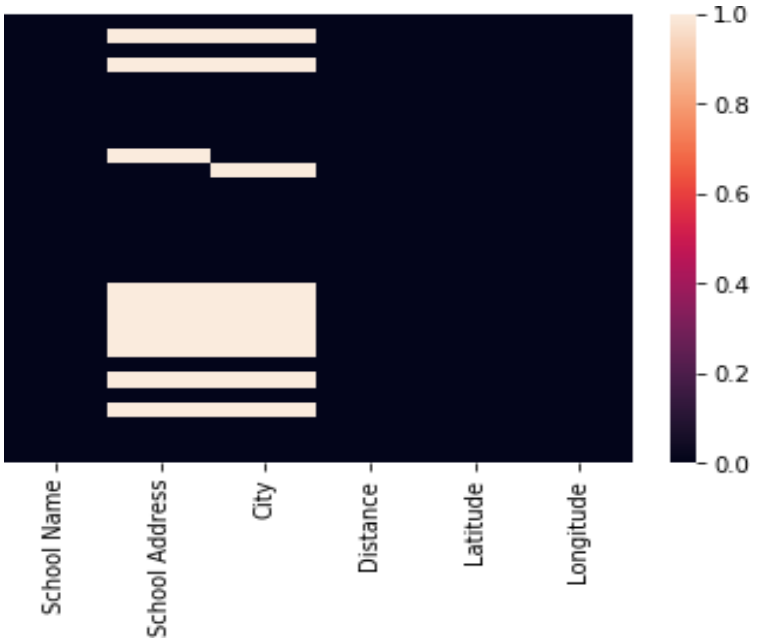
Now we will change the columns name as per our convenience.

	School Name	School Address	City	Distance	Latitude	Longitude
0	Nehru Nagar Municipal School	Sai Hill Rd.	Mumbai	1551	19.062062	72.877558
1	janseva motor training school	NaN	NaN	356	19.073660	72.880009
2	School of Communications & Reputation	1st Floor, Kohinoor Education Complex, Kurla West	Mumbai	992	19.075196	72.887099
3	Kohinoor Business School	NaN	NaN	505	19.076290	72.882491
4	Kohinoor Business School	Kohinoor Education Complex, Kohinoor City, Kir...	Mumbai	769	19.073995	72.870699

.

Data Cleaning: To fit the model, one needs to get rid of the null values. Hence, the firstly, the columns with null, none or NaN values, were identified. The cleaning of data was done by removing the columns with NaN or null values.

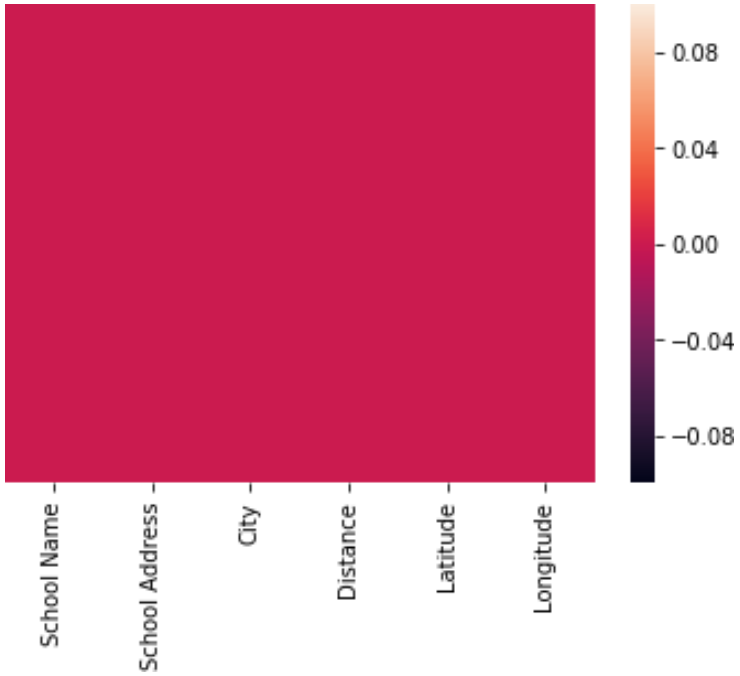
The columns were dropped keeping in mind whether they were really useful for analysis purpose or not. Heat map is generated for visualizing all the null values of the columns.



Now fixing null values for City, as we know city is Mumbai so we can replace all the null values in City column by Mumbai.

For column School Address, we will drop the schools which have address as null values as the data about their address is essential. Other columns do not have any null values.

Now regenerate heat map again with no null values for verification.



MODEL FITTING

CLUSTERING: Now we will divide schools of Mumbai on the basis of their location by making clusters of different colors.

We will use K-means clustering technique for this project. In this we will specify value of k i.e. number of clusters. K is chosen 4 here. The model was the fitted and the labels were generated in the form of array.

The dataframe is shown below:

	School Name	School Address	City	Distance	Latitude	Longitude	Labels
0	Nehru Nagar Municipal School	Sai Hill Rd.	Mumbai	1551	19.062062	72.877558	0
2	School of Communications & Reputation	1st Floor, Kohinoor Education Complex, Kurla West	Mumbai	992	19.075196	72.887099	0
4	Kohinoor Business School	Kohinoor Education Complex, Kohinoor City, Kir...	Mumbai	769	19.073995	72.870699	0
5	Richard Ivey School of Business, India	Regus Trade Centre,	Mumbai	788	19.072643	72.871102	0
6	karthika high school	new hall road	mumbai	755	19.079910	72.883575	0
7	Cardinal Gracias High School	Ali Yawar Jung Marg, Bandra (East)	Mumbai	3363	19.068198	72.846814	1
8	Babu Driving School	India	Mumbai	859	19.082269	72.882462	0
10	Michael High School	Kurla, Mumbai, Maharashtra	Mumbai	986	19.083023	72.883425	0
11	Kartika High School	Kurla	Mumbai	990	19.082386	72.884256	0
12	Our Lady of Perpetual Succour (OLPS) High School	St. Anthony's Rd, Chembur (East)	Mumbai	3945	19.051014	72.904300	0
13	GEI School	Kurla Bus Depot	Mumbai	949	19.067577	72.879110	0
14	St Mary's High School	Kalina	Mumbai	1143	19.076634	72.866851	0
15	Ecole Mondiale World School	Gulmohar Cross, Road No. 9, JVPD Scheme	Mumbai	6151	19.112641	72.833930	3
16	Dharavi Transit Camp Municipal School	Mahatma Gandhi Road	Mumbai	4167	19.042748	72.859493	2
17	The ITA School of Performing Arts Pvt. Ltd	Unit No.2, Ground Floor, Techniplex 2, S.V.Roa...	Mumbai	1161	19.071781	72.867607	0

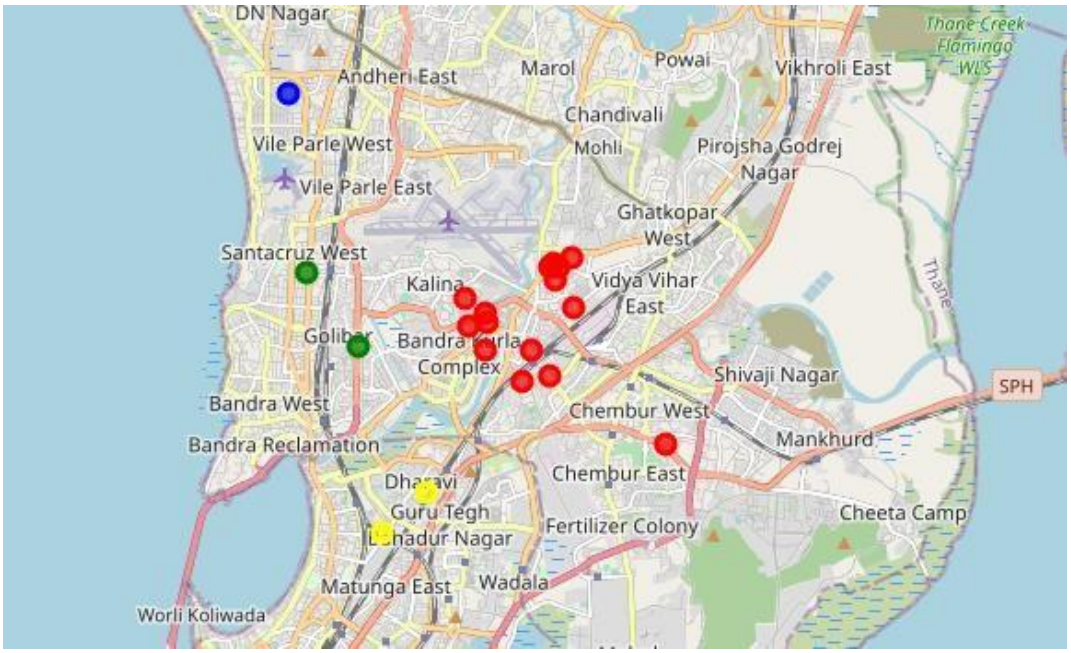
Color code: The color codes used in this project are shown below:

Cluster	**Color Code**
0	Red
1	Green
2	Yellow
3	Blue

DATA VISUALIZATION

All the 4 clusters were visualized on a map centered on Mumbai. The color coding was applied while visualizing for differentiating between the clusters.

Visualizing target cluster: Now we visualize the schools on the basis of cluster which is located on the basis of latitude and longitude of each school. There are four colors: Red, Yellow, Blue and Green.



Observation: Hence parents must choose a school from the cluster near to their home by comparing other listed facilities provided. This would make their task easier for selecting best school for their children.

Project URL:
https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/a948d563-4679-4e03-9a09-4f9a71665604/view?access_token=7101ef3e8c697760247f43c03d606f25740c8b47f7f8c271af8d785295a0128c