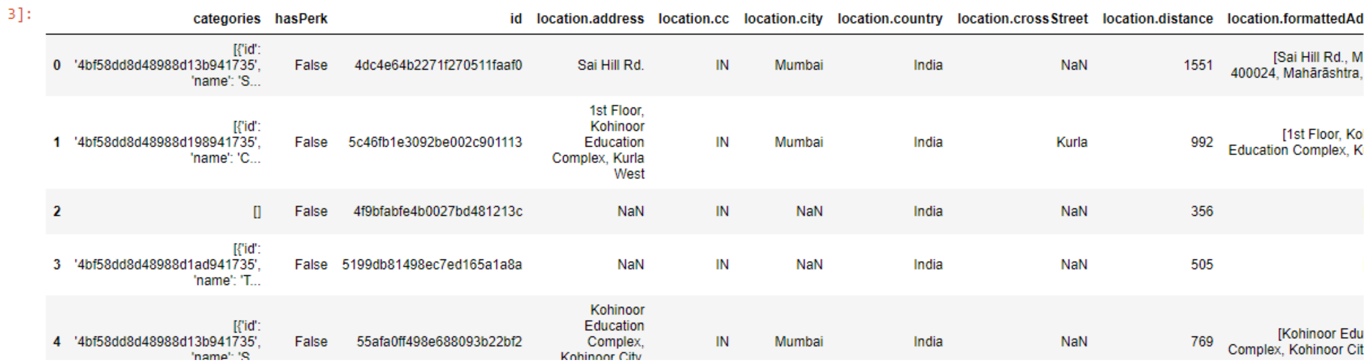
DATA ANALYSIS

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.

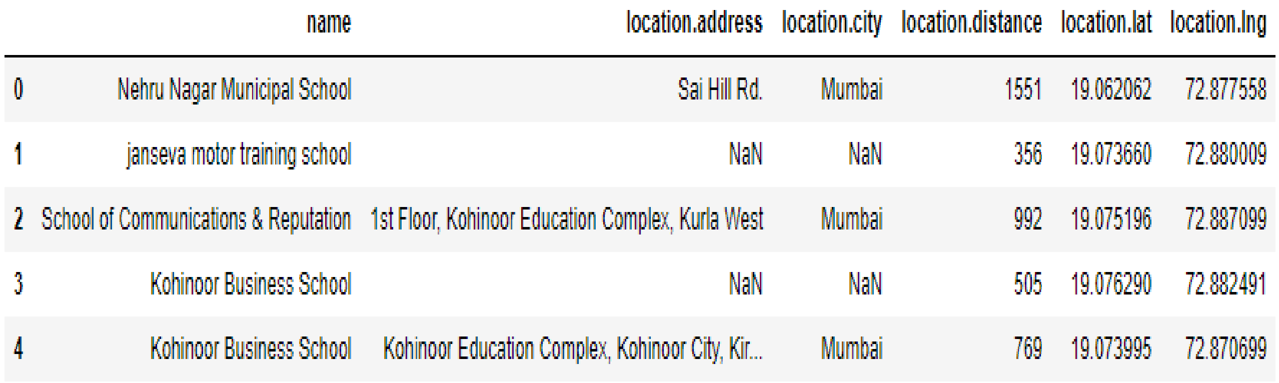
Data Preprocessing and Wrangling: Using the modules of python, only valid and usable data was selected from the JSON file generated by foursquare and data-frame was created using **‘Pandas’**. Since the project required only the locations, the ‘**venues**’ section under the ‘**response**’ section was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



3]:

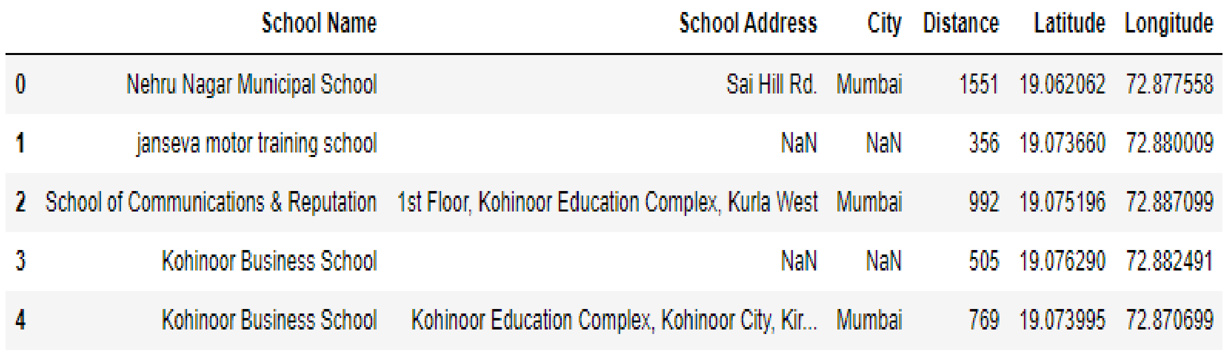
	categories	hasPerk	id	location.address	location.cc	location.city	location.country	location.crossStreet	location.distance	location.formattedAd
0	[{'id': '4bf58dd8d48988d13b941735', 'name': 'S...	False	4dc4e64b2271f270511faaf0	Sai Hill Rd.	IN	Mumbai	India	NaN	1551	[Sai Hill Rd., M 400024, Mahārāshtra,
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.

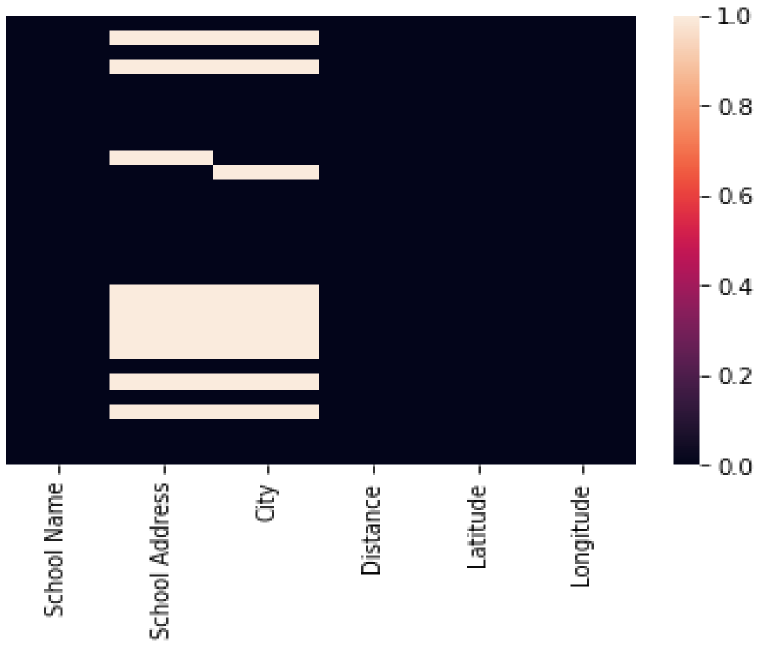


	School Name	School Address	City	Distance	Latitude	Longitude
0	Nehru Nagar Municipal School	Sai Hill Rd.	Mumbai	1551	19.062062	72.877558
1	janseva motor training school	NaN	NaN	356	19.073660	72.880009
2	School of Communications & Reputation	1st Floor, Kohinoor Education Complex, Kurla West	Mumbai	992	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.

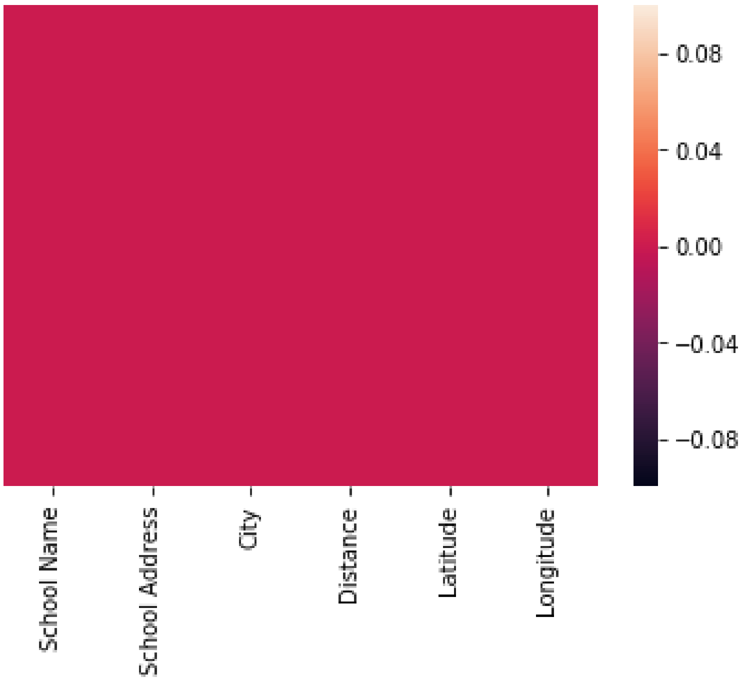
The columns were dropped keeping in mind whether they were really useful for analysis purpose or not. Heat map is generated for visualizing all the null values of the columns.



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.


School Name
School Address
City
Distance
Latitude
Longitude