Report : Capstone project

Topic : The Battle of Neighbourhood

Content

Introduction :

Agra is a city on the banks of the Yamuna river in the Agra district of the Indian state of Uttar Pradesh.It is 206 kilometres south of the national capital New Delhi. Agra is the fourth-most populous city in Uttar Pradesh and 24th in India.

Agra is a major tourist destination because of its many Mughal-era buildings, most notably the Taj Mahal, Agra Fort and Fatehpur Sikri, all of which are UNESCO World Heritage Sites.

As of the 2011 India census, Agra city has a population of 1,585,704, while the population of Agra cantonment is 53,053.

The urban agglomeration of Agra has a population of 1,760,285. Males constitute 53% of the population and females 47%. Agra city has an average literacy rate of 73%, below the national average of 74%. Literacy rate of males is considerably higher than that of women.

Data :

For this project we have downloaded data from kaggle which is a zomato.csv file. It contains the details of all the restaurants present all around the world.

To get the information about the restaurants present in Agra firstly we will find the restaurants present in India then we will search for all the restaurants present in Agra.

The data contains different information about each city like its locality, restaurant name, rating of restaurants, their latitude and longitude etc.

Here we will be using Foursquare API to get the information about the venue of all the restaurants in Agra.

Methodology :

Approach :

Firstly, we will collect the data from kaggle and then we will try to find the answer of the question that is lying inside the data.

Before finding the insights we need to preprocess the data which is the most important step of cleaning the data. In this we will find the unnecessary column that we don't need further and will eliminate that one.

Visualization :

Visualizing the data helps to find the insights more clear. In this after cleaning the data we will try to visualize the data which will make our work more easy and efficient.

Visualization generally means that we are trying to see the data graphically.

Data insights :

After visualization we will try to answer the following question :

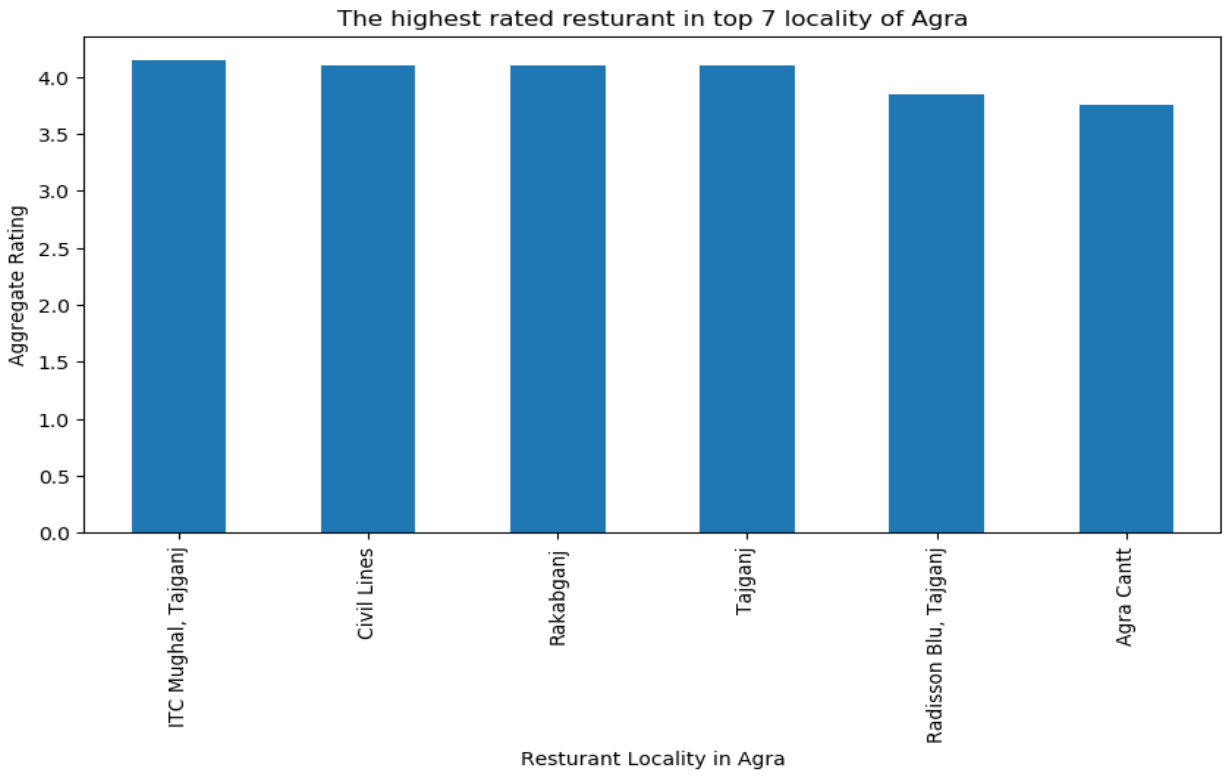
a. Which are top 7 highly rated restaurants in Agra?
b. Which are the worst 5 rated restaurants in Agra?

c. Find the locality where the highest number of restaurants are available?
d. Find the locality where the lowest number of restaurants are available?
e. Find the restaurants where best North Indian food is served?

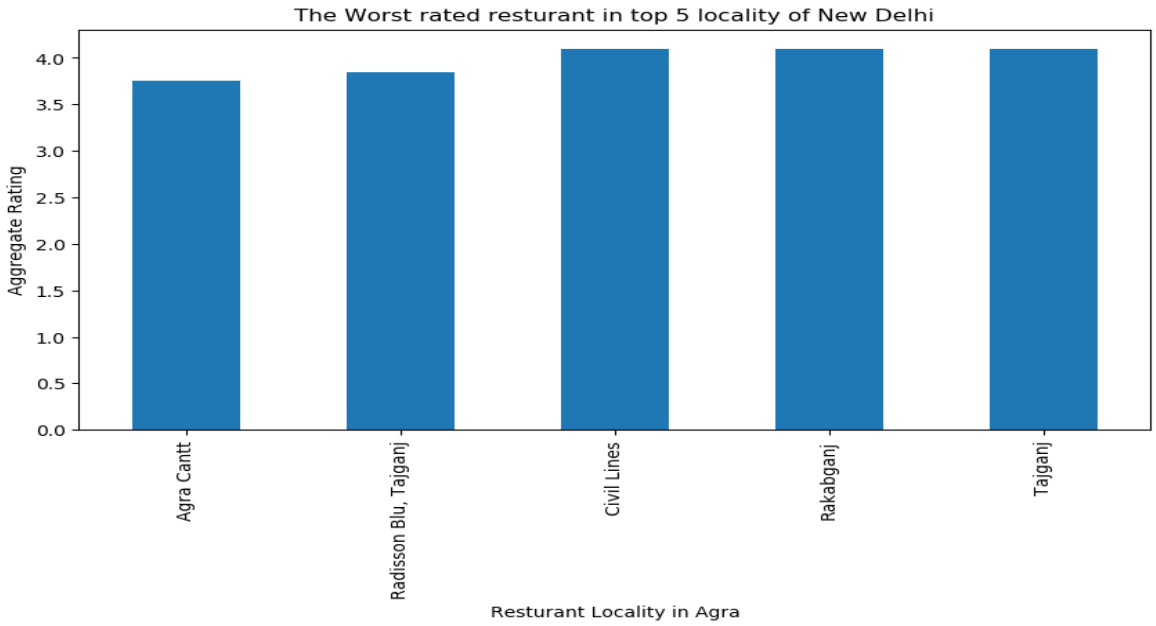
Results :

After performing all the above steps now will try to answer the questions that is asked above:

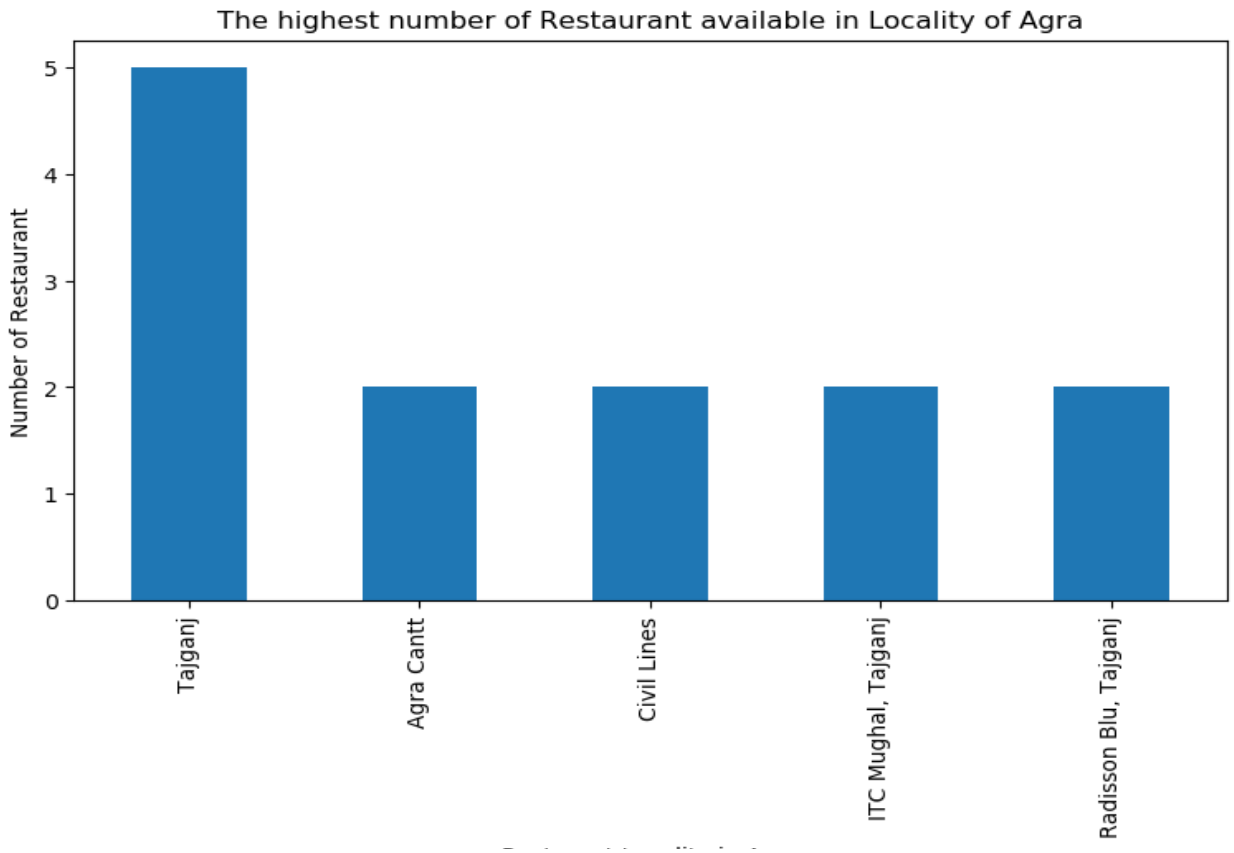
A. Following diagram shows the top 7 highly rated restaurants in Agra.



B. Following diagram shows the worst 5 rated restaurants in Agra.

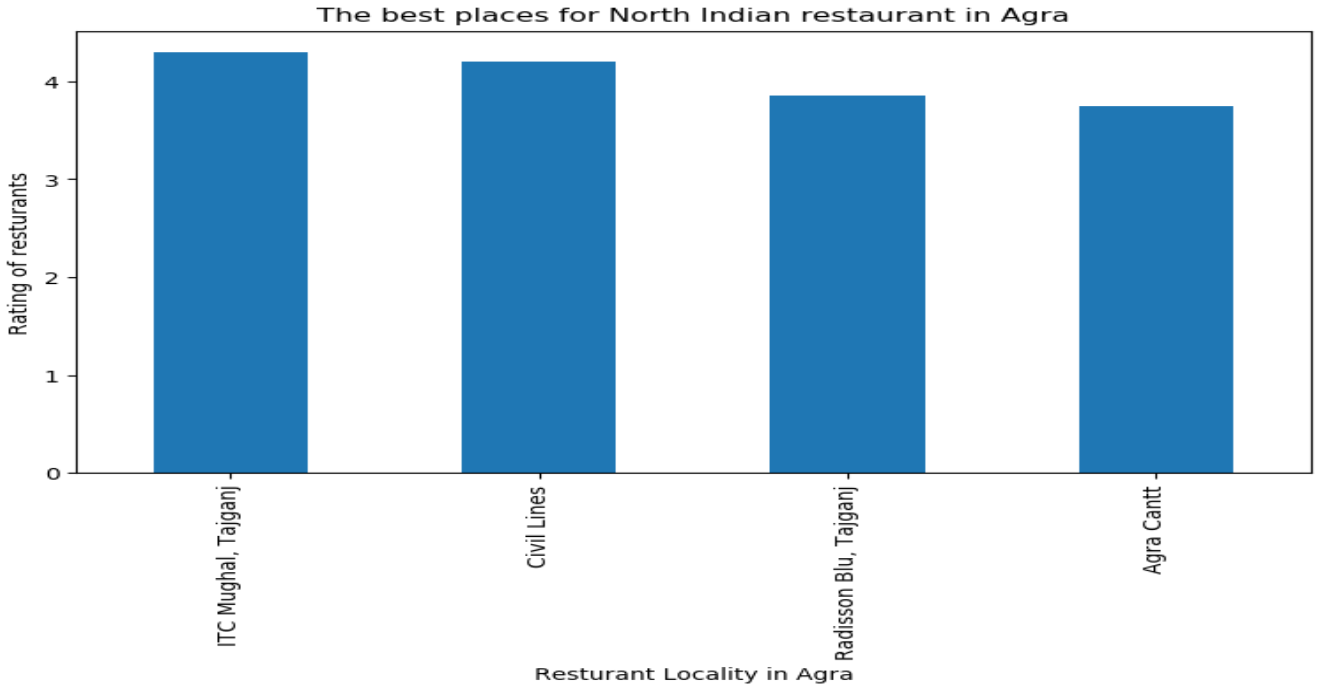


C. Following are the localities where the highest number of restaurants are available.



D. Following are the localities where the lowest number of restaurants are available.

E. Following are the restaurants where best North Indian food is served.



Discussion :

Here we are using the K-means clustering algorithm to find the cluster of restaurants that are located in a locality. K-means clustering is one of the simplest and popular unsupervised machine learning algorithms.

Using the k-means we found the different cluster which helps to answer the insights that are present inside the data.

Even though we don't have a sufficient number of restaurants in Agra, we managed to find the best insights from that itself.

Conclusion :

In this study is that relationship between the different entities that are present inside the data. In this report we managed to find out the best restaurants, worst restaurants, highest/lowest number of restaurants in a locality , best North Indian food restaurant in Agra etc.

Hence, this concludes this report.