DATA SCIENCE

Restaurant - Exploratory Data Analysis & Prediction Model:

An Internship Report

[Levels completed: All 3 ✔]

Intern: Kshetrimayum Galax Singh

Ref : CTI/A3/C2332

Company: Cognifyz (https://cognifyz.com/)

Domain: Data Science

Batch: Jan-Feb 2024

GitHub repo: https://github.com/galax19ksh/Restaurant-Analysis-and-Predictive-Model

Contact: +91 7005788363

Mail: galaxkshetrimayum16@gmail.com

Introduction:

I am so grateful to have undertaken a Data Science internship at Cognifyz technologies. My main role is to gather meaningful insights by conducting exploratory data analysis on the large restaurant dataset, as well as build a ML model to predict ratings. I was given 3 levels each comprising of 3 tasks. Due to my commitment to the projects, I completed all 3 levels in time. Well I took ample time to work to learn and hone my skills that involved data analysis, machine learning knowledge and application to solve problems. I did the necessary data exploration, preprocessing and various visualization methods to dive deep into finding interesting insights.

More details are listed below:

Tasks: All tasks of three levels are given in the pdf attached along with dataset.

Levels completed: All 3 ✔

Platform used: Google Colab

Libraries used: pandas, numpy, matplotlib, seaborn, scikitlearn, folium, geopanda

Data Preprocessing & Feature Engineering

- Cuisines had 9 null values. So dropped the rows
- Removed features that will inhibit model performance
- Split training data and test data in the ratio 8:2
- Some features/columns needed label encoding.

Model Training and Performance

- Used Random Forest, Decision Tree Logistic Regression algorithms to build the models
- My restaurant rating prediction model (Random Forest and Decision Tree) obtained an aggregate R2 score of 0.93

Data Analysis : Insights

(level and task wise conclusions are given in the .ipynb files or github)

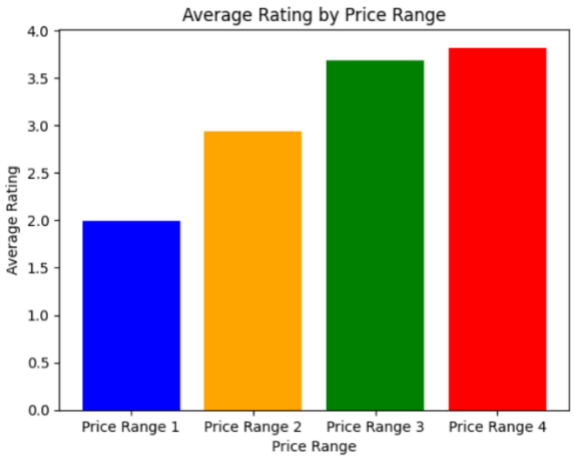
- There are many restaurants having 0 rating probably due to less popularity.
- Visualized the geospatial distribution of restaurants on the map coordinates using folium and geopanda
- Most popular restaurants come in the range of ratings 3 to 3.5.
- Expensive restaurants (higher price range) tend to have higher ratings.
- New Delhi has the highest number of restaurants.

- By country, country code “1”, probably North America has most no of restaurants.
- 'North Indian' is the most popular cuisine overall, followed by "Chinese" and "fast food".
- Restaurants having table booking facility have fairly higher average rating.
- “Sunda” is the highest rated cuisine and also has the most votes.

Conclusion:

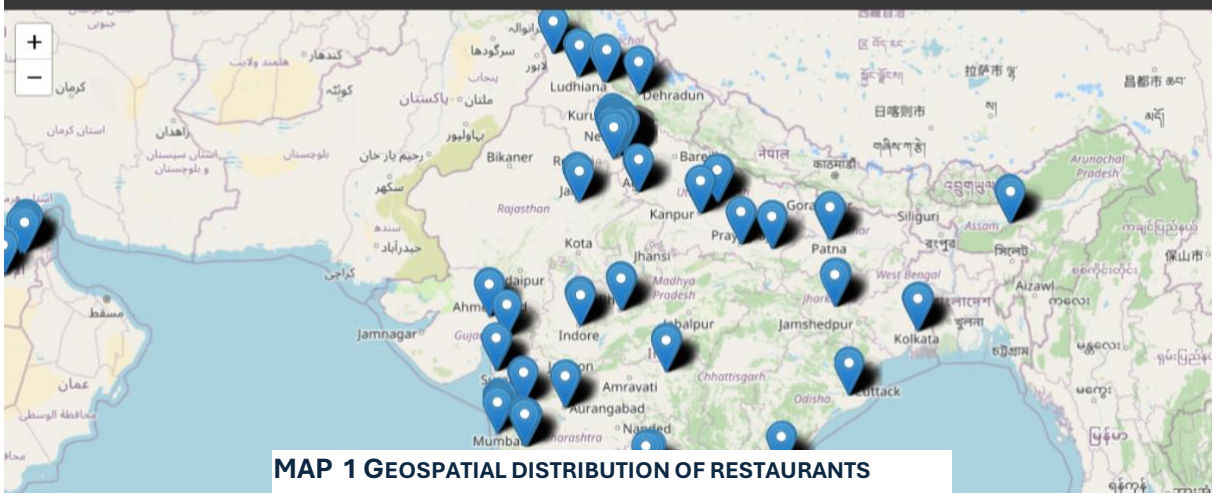
My internship at Cognifyz has been invaluable in providing practical insights into the field of Data Science. Over the course of the internship, I have gained hands-on experience in Restaurant dataset, further enhancing my skills in data analysis both in quantitative and visualization areas. I am confident that the experiences gained during this internship will greatly boost my career.

Data Visualization for references:

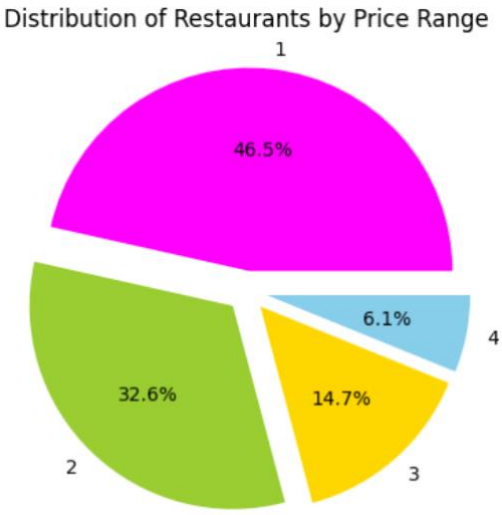


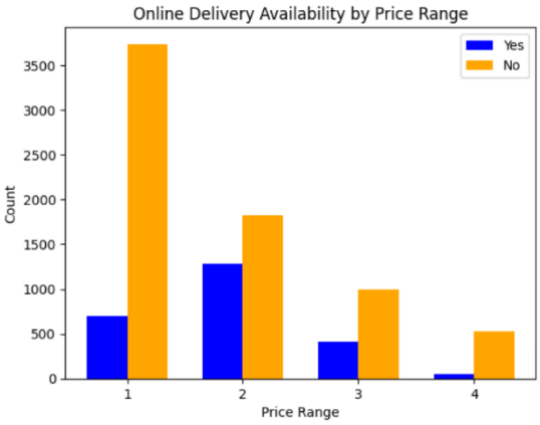
```
map_restaurants = folium.Map(location=[restaurants_gdf['geometry'].y.mean(), restaurants_gdf['geometry'].x.mean()], zoom_start=12)
restaurants_gdf.crs = 'EPSG:4326'
folium.GeoJson(restaurants_gdf).add_to(map_restaurants)

#map_restaurants.save('restaurants_map.html')
map_restaurants
```



MAP 1 GEOSPATIAL DISTRIBUTION OF RESTAURANTS





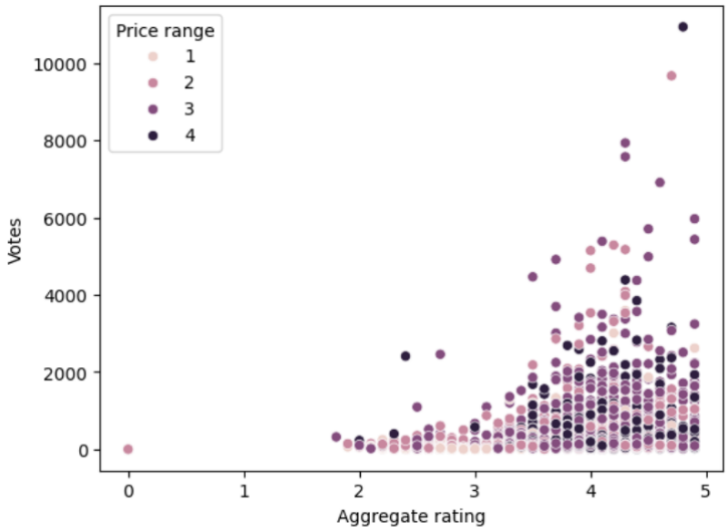
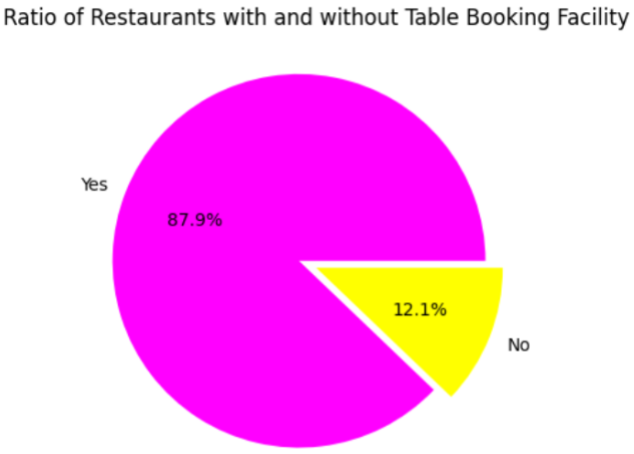
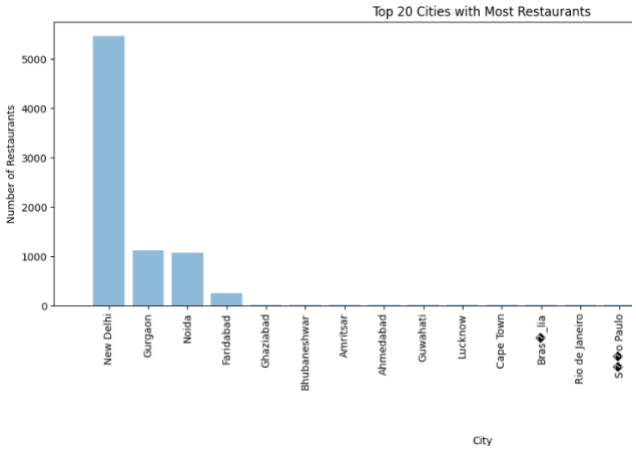
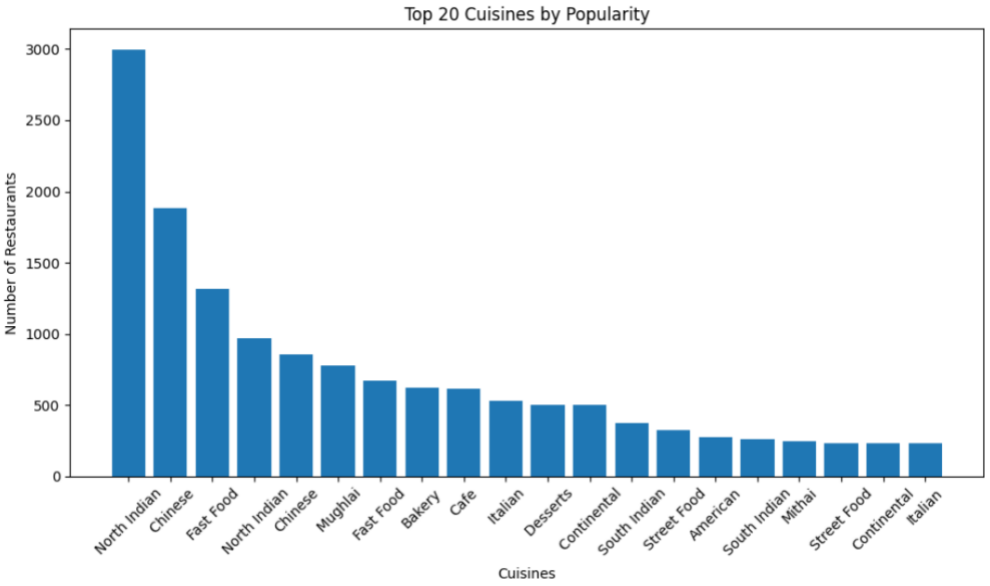
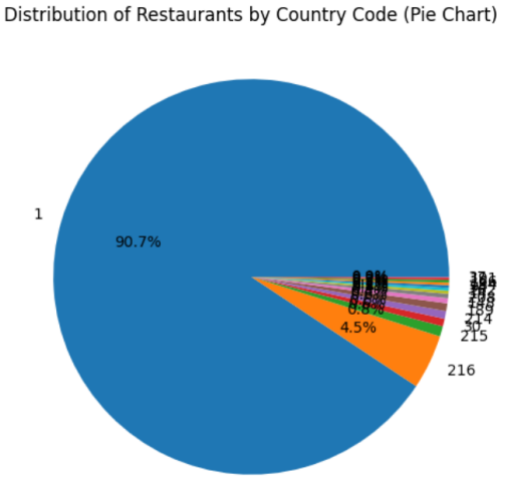


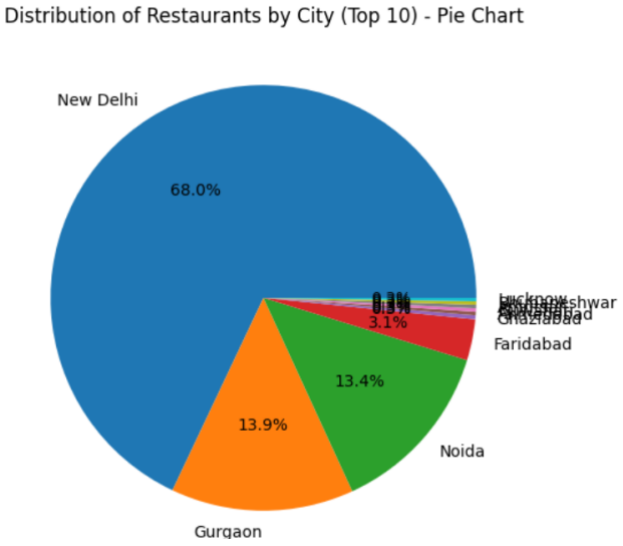
FIGURE 1 RELATION BETWEEN RATING, PRICE AND VOTES

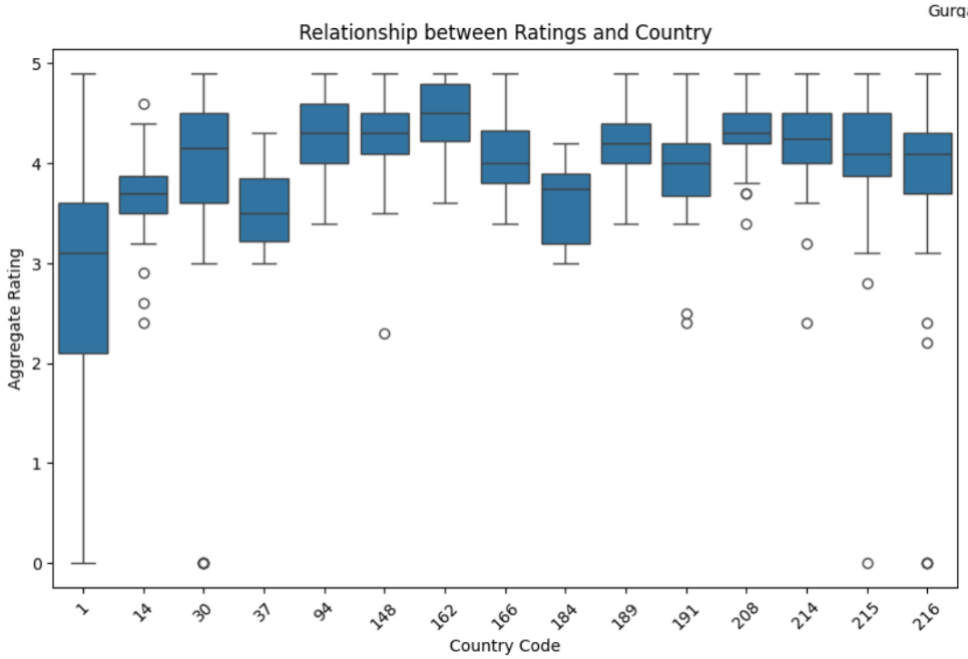


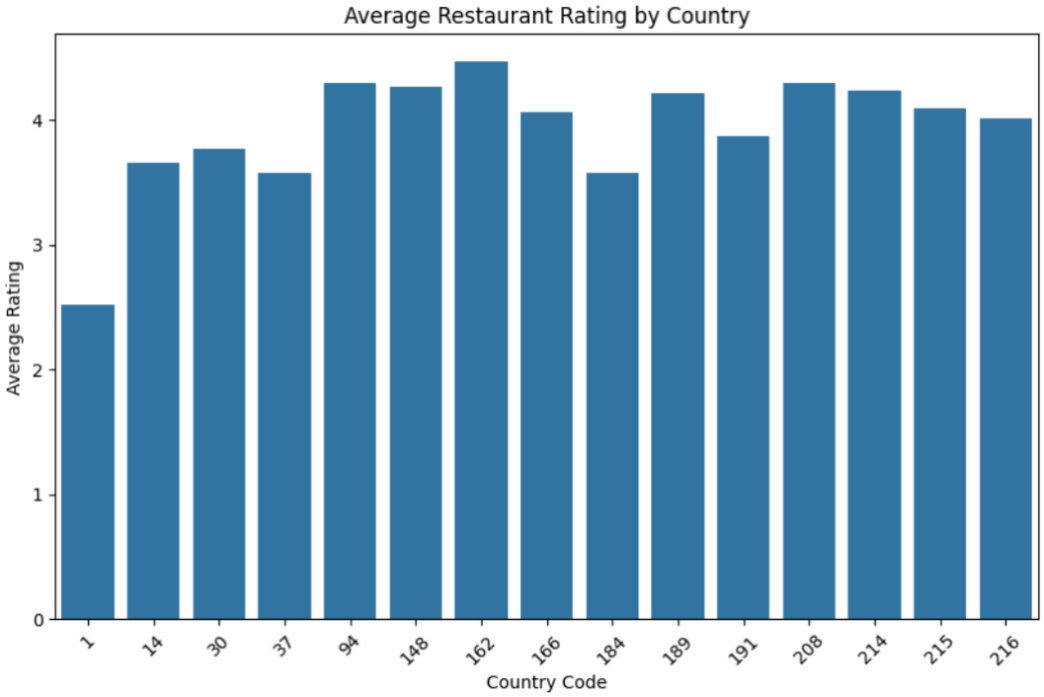


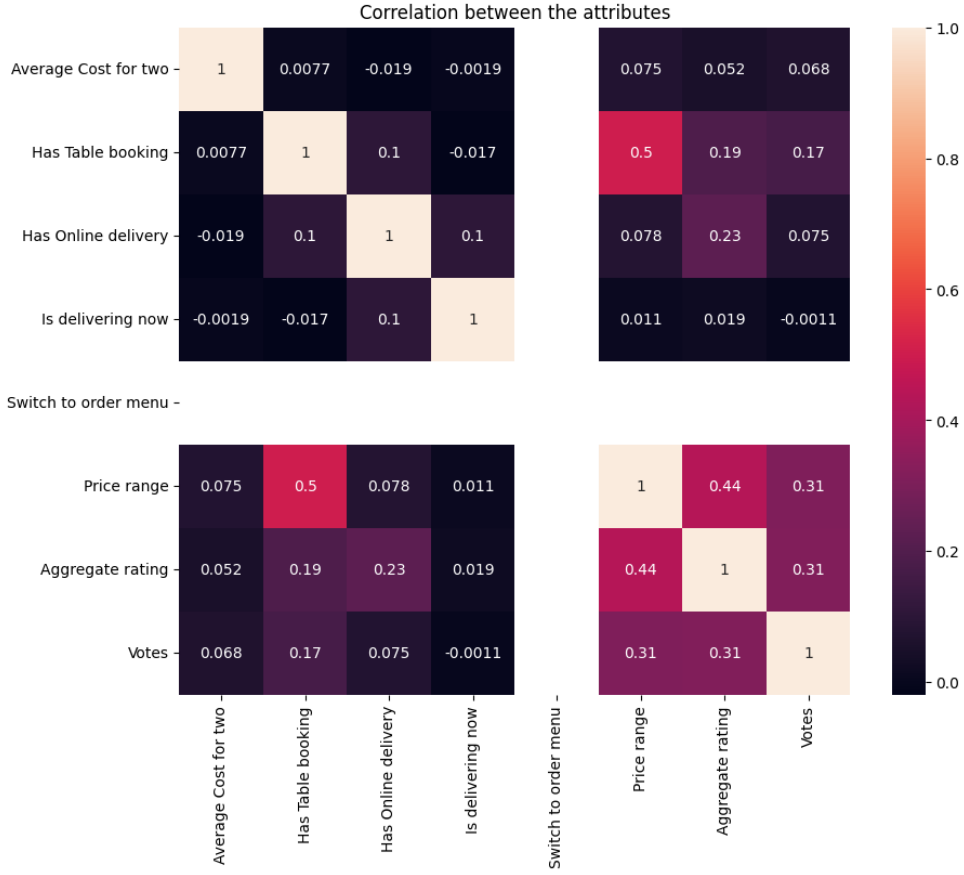

Top 20 Cuisines by Popularity
Number of Restaurants
3000
2500
2000
1500
1000
500
0
North Indian
Chinese
Fast Food
North Indian
Chinese
Mughlai
Fast Food
Bakery
Cafe
Italian
Desserts
Continental
South Indian
Street Food
American
South Indian
Mithai
Street Food
Continental
Italian
Cuisines


Distribution of Restaurants by Country Code (Pie Chart)
1
90.7%
4.5%
215
216


Distribution of Restaurants by City (Top 10) - Pie Chart
New Delhi
68.0%
3.1%
Faridabad
13.4%
Noida
13.9%
Gurgaon


Relationship between Ratings and Country
Aggregate Rating
5
4
3
2
1
0
1
14
30
37
94
148
162
166
184
189
191
208
214
215
216
Country Code


Average Restaurant Rating by Country
Average Rating
0
1
2
3
4
1
14
30
37
94
148
162
166
184
189
191
208
214
215
216
Country Code


Correlation between the attributes
Average Cost for two
Has Table booking
Has Online delivery
Is delivering now
Switch to order menu
Price range
Aggregate rating
Votes
1 0.0077 -0.019 -0.0019 0.075 0.052 0.068
0.0077 1 0.1 -0.017 0.5 0.19 0.17
-0.019 0.1 1 0.1 0.078 0.23 0.075
-0.0019 -0.017 0.1 1 0.011 0.019 -0.0011
0.075 0.5 0.078 0.011 1 0.44 0.31
0.052 0.19 0.23 0.019 0.44 1 0.31
0.068 0.17 0.075 -0.0011 0.31 0.31 1
1.0
0.8
0.6
0.4
0.2
0.0