Zomato Bengaluru Restaurant Analysis

Problem:-

The basic idea of analyzing the Zomato dataset is to get a fair idea about the factors affecting the establishment of different types of restaurant at different places in Bengaluru, aggregate rating of each restaurant, Bengaluru being one such city has more than 12,000 restaurants with restaurants serving dishes from all over the world. With each day new restaurants opening the industry has'nt been saturated yet and the demand is increasing day by day. Inspite of increasing demand it however has become difficult for new restaurants to compete with established restaurants. Most of them serving the same food. Bengaluru being an IT capital of India. Most of the people here are dependent mainly on the restaurant food as they don't have time to cook for themselves.

Opportunity:-

From all the Data available, we can bring out some neat insights or conclusions such as

- Which franchise has highest number of Restaurants?
- How many Restaurants are accepting online orders?
- How many has a book table facility?
- Which location has heighest number of Restaurants?
- How many types of Restaurant types are there?
- What is the most liked Restaurant type?
- What is the Average cost for 2 persons?
- What is the most liked Dish type?

And so on…..

Tools used:-

- Python 3.6
- Numpy
- Pandas
- Matplotlib
- Data science
- Machine learning

Dataset:-

The data is accurate to that available on the zomato website until 15 March 2019. The data we have has 51717 records with 17 feautres. Each and every feature has their own importance and helps to better understand the problem statement here. We have file size of 6.7MB of csv file with both int and object types.

We have features like url, address, name, online order, book table, rating, votes, phone, location, Restaurant type, dish liked, cuisines, average cost, reviews list, menu items, and so on..

Process:-

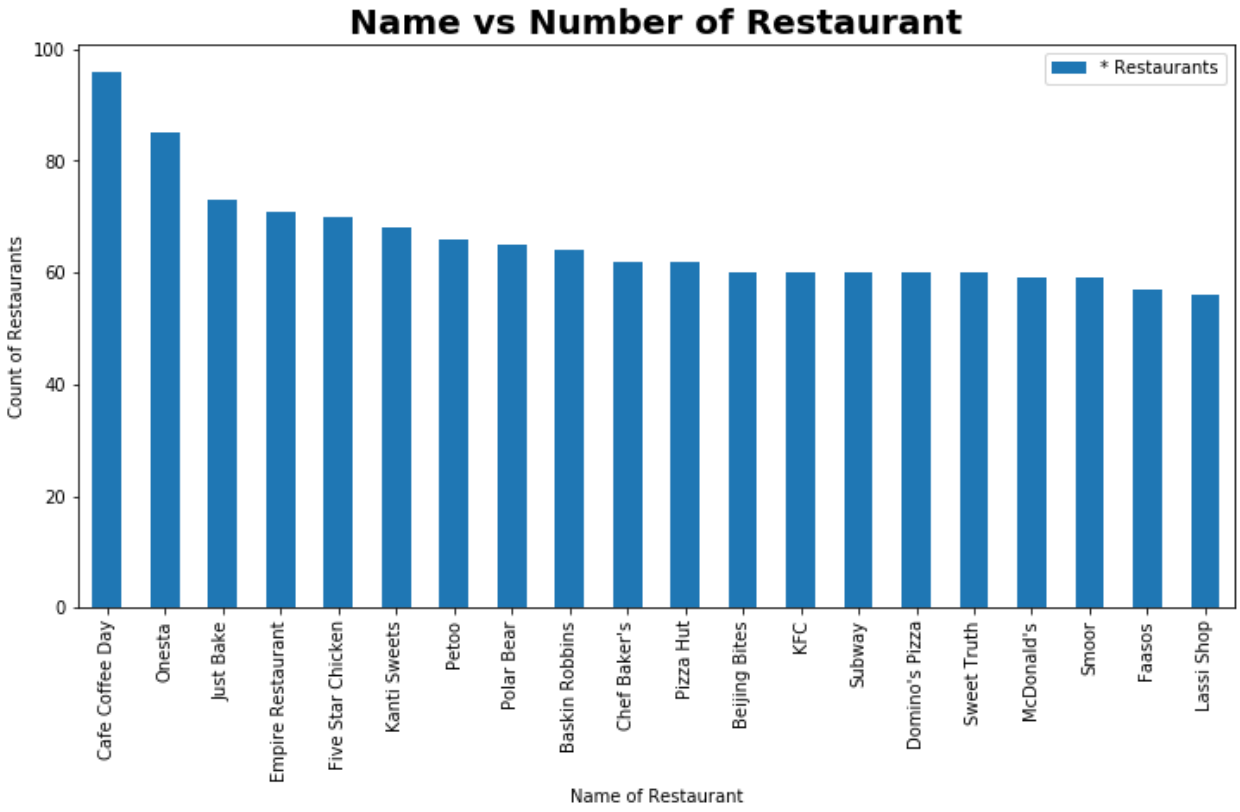
Use the tools that we have like Numpy, Pandas, Matplotlib, Seaborn and Python to get insights as asked above from the data. Lets answer each and every point one by One.

The first thing you should do when you have the dataset is to look for Duplicates, Missing vlaues, Outliers and Balanced/not. Check for the features that place major role in taking better decisions for our problem solving. Lets visualize each and every feature.

Which franchise has highest number of restaurants?

We have different franchises in bangalore . each franchise has many number of restaurants in bangalore. Look at the graph below. From the graph you can understand that we have more Restaurants of franchise called “Cafe Coffee Day”. It has almost 100 number of restaurants in bangalore.

The below graph says Name vs number of restaurants in bangalore. Please have a look for better understanding.

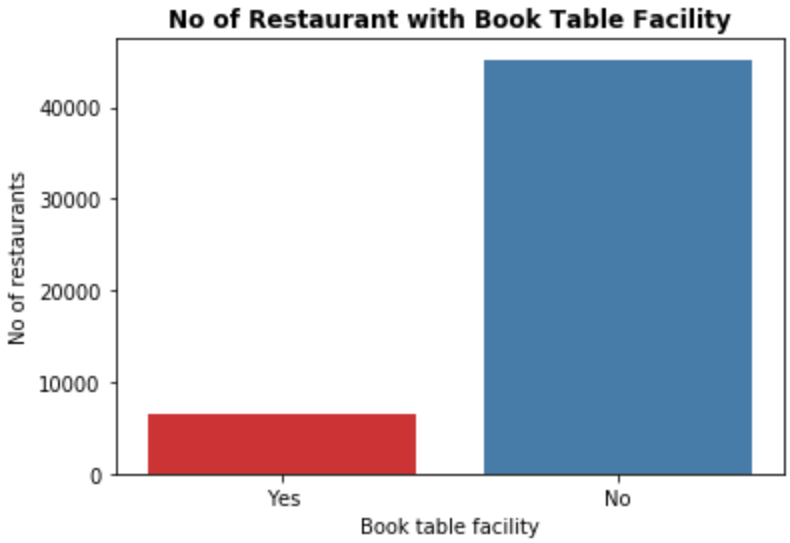


How many Restaurants are accepting online orders?

Though we have several Restaurants in bangalore, we have only few restaurants that are accepting the online orders from customers and few are not. Look at the graph below to see the number of restaurants accepting the online orders.

Book table:-

We have only few Restaurants that are accepting the book table facility. Look at the graph below to understand the Number of restaurants with book table.



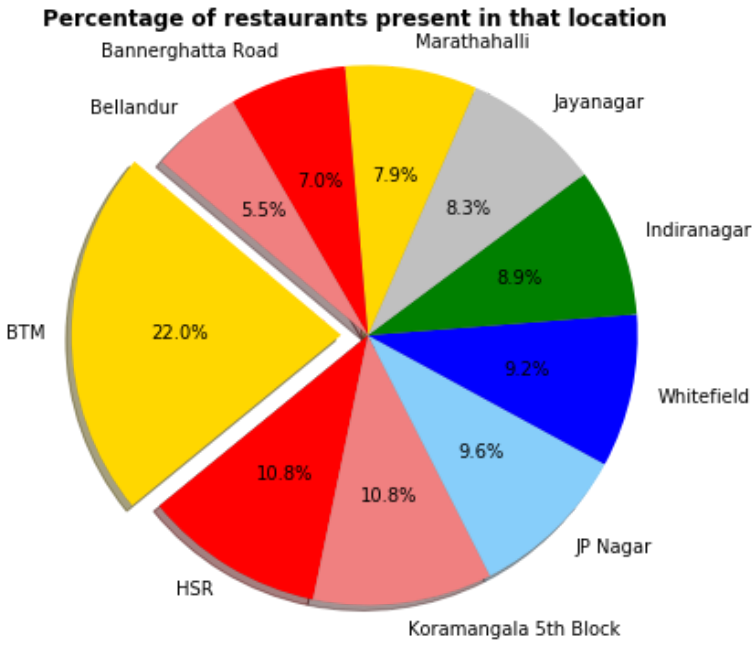
We have nearly 45268 restaurants with no book table facility.

Which location has more number of Restaurants?

We have restaurants in several locations of bangalore. Please look at the graph below to understand the number of locations in each location of bangalore.

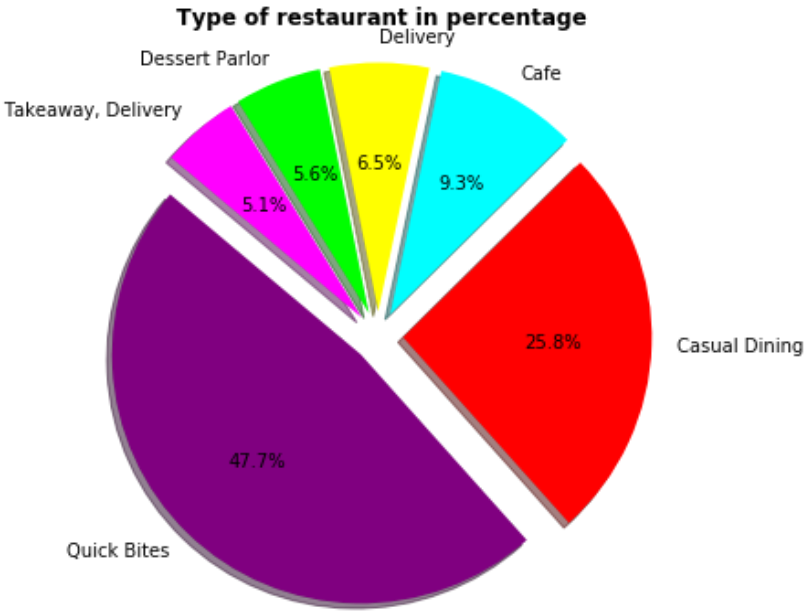
If you look at the graph we can easily figure out that BTM, a location in bangalore has more number of restaurants. After BTM, we have more number of restaurants in HSR layout from bangalore.

We have total of 93 locations in bangalore that are tied up to zomato . We have a nice pie graph below to understand the percentages of restaurants in each location.



Restaurant type:-

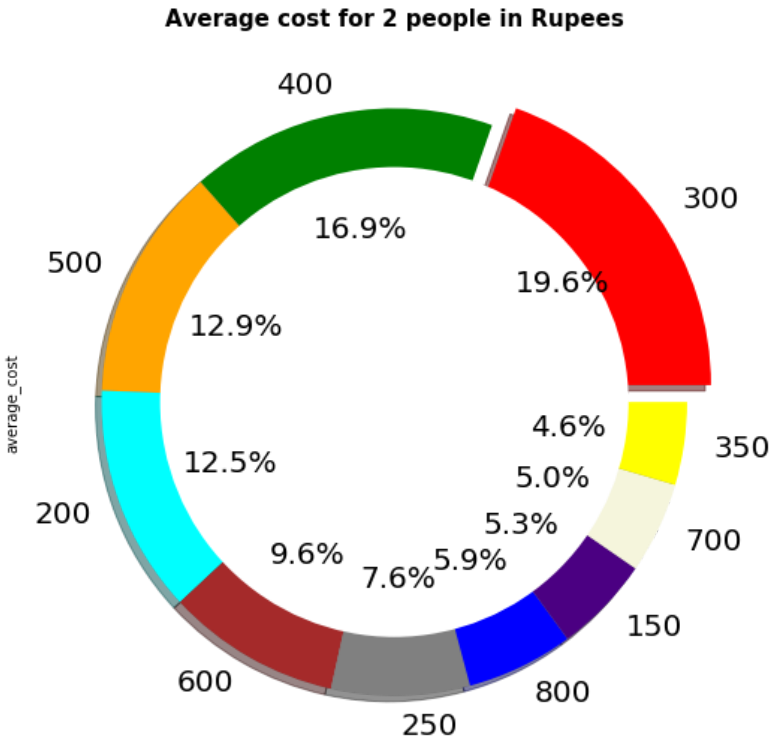
We have a feature called Restaurant type that tells about type it had. We have different types of restaurant types such as Quick bites, dining, cafe, delivery, dessert parlour,bar, pub and so on…



From the graph above you can see that we have more number of restaurants types in bangalore are in Quick bites section. After that we have more number of restaurant type is Casual Dining.

Average cost:-

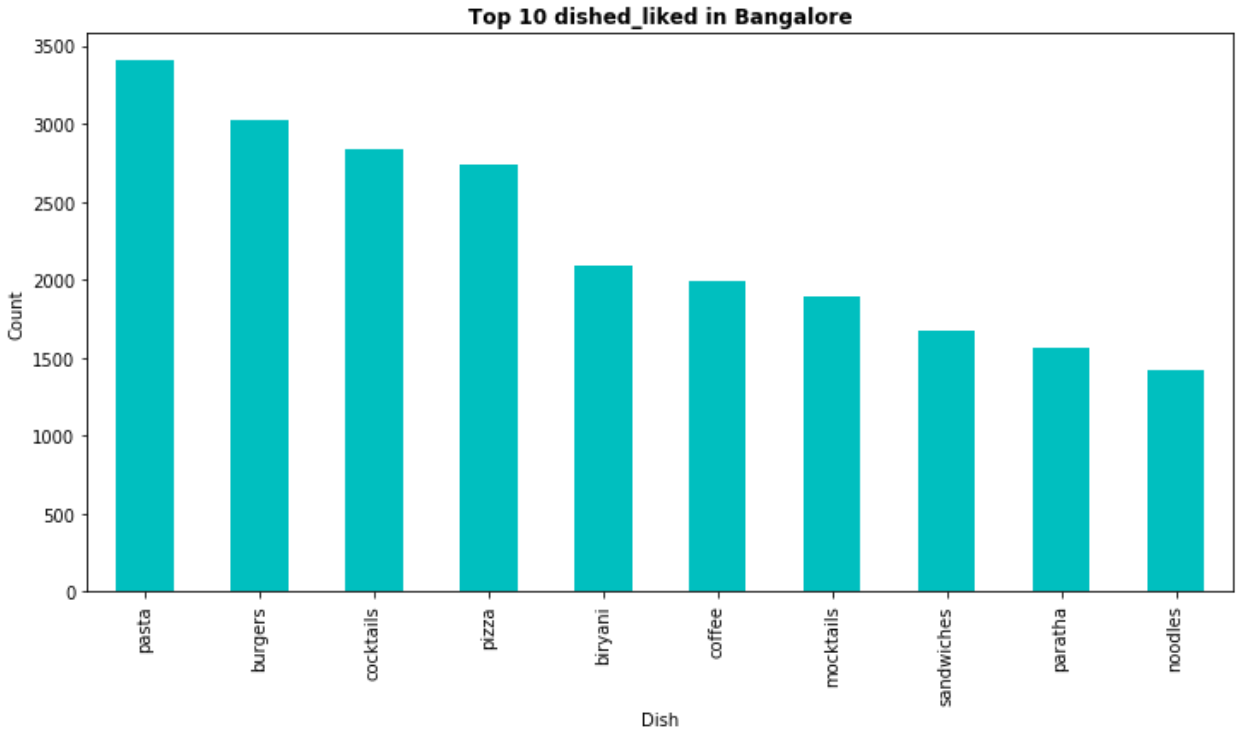
We have different restaurants with different cost for 2 persons in bangalore. The average cost for 2 persons ranges from 300 to 4000. Look at the graph below. From the graph you can understand that we have 19.6% of restaurants with the average cost for 2 persons is 300 rupees.



Dishes:-

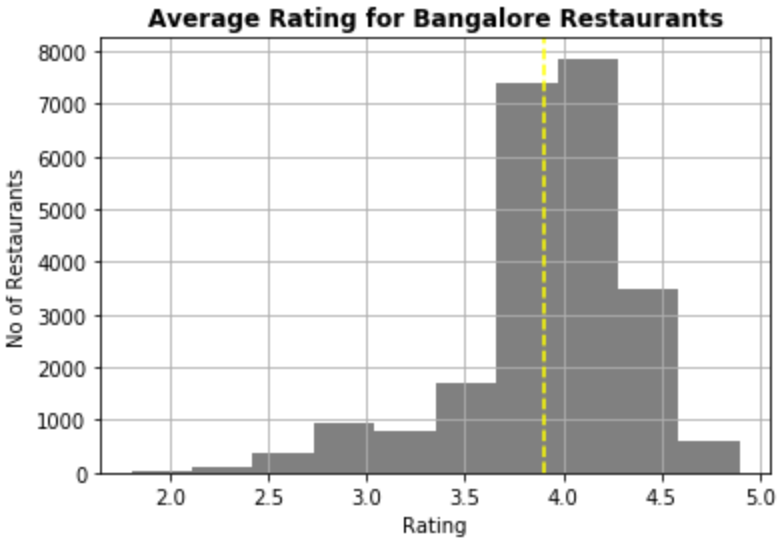
We have different types of Dishes from different Restaurants all over tha bangalore. From the graph below we can see that most liked dish in bangalore is Pasta. In second place we have burgers. We also have cocktails, pizza, biryani sequentially in the row.

Have a look at the graph below.



Rating:-

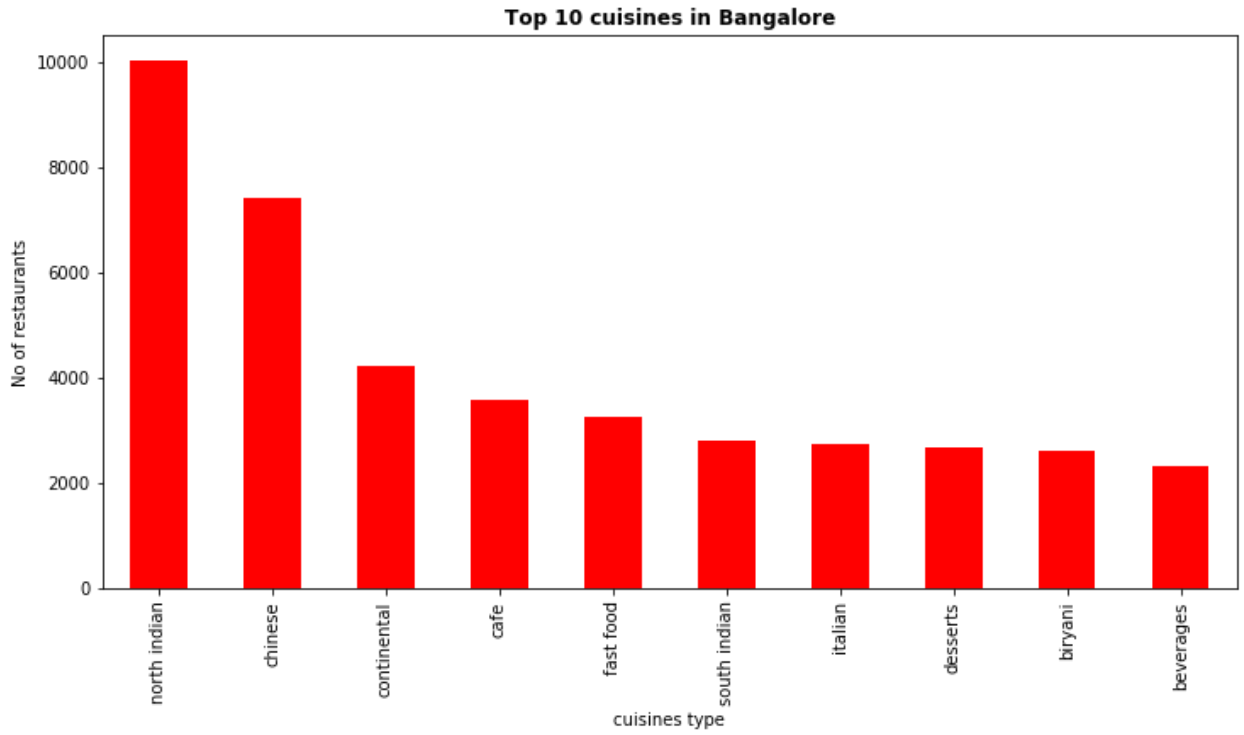
We have different Ratings that ranges from 1 to 5. From the graph below you can see that we have Average rating per restaurant in Bangalore is 3.9



Cuisines:-

We have different types of cuisines all over the bangalore from different restaurants. There are cuisines like North indian, Chinese, continental, Cafe, Fast food and so on.

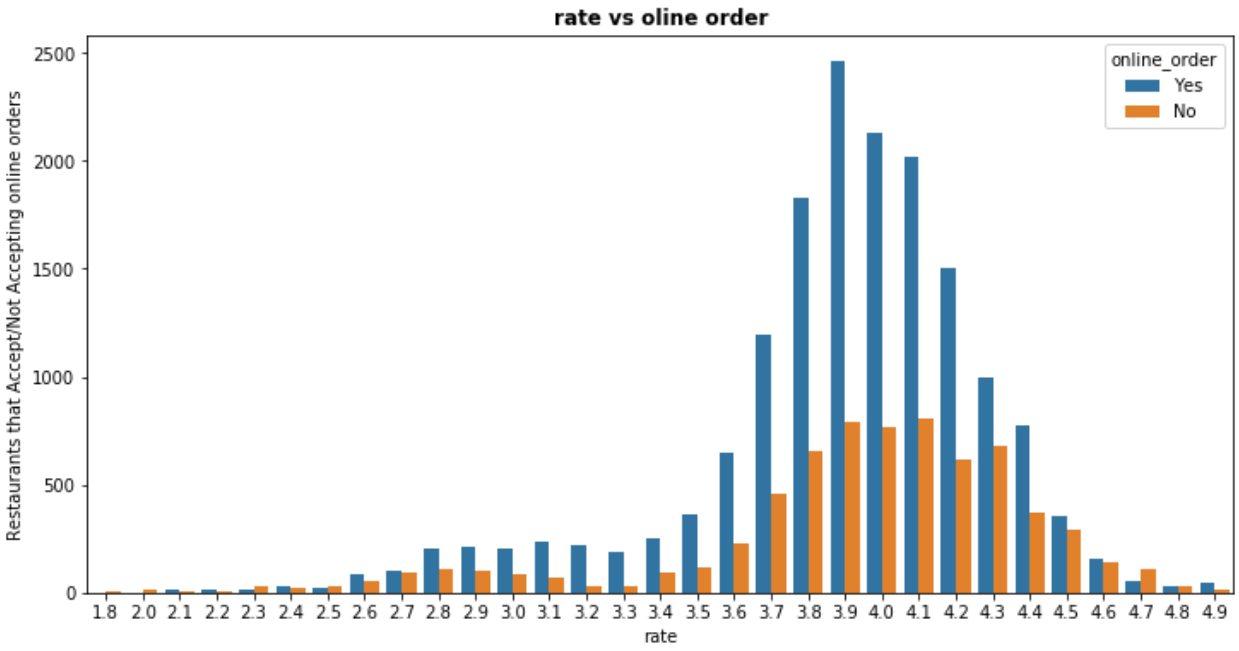
Look at the graph below to understand the the number of restaurant with different types of cuisines. The top 10 cuisines are shown below in the graph.



Rating vs Online order:-

We have a graph below that shows the number of restaurants in particular rating that accepting the online orders and that are not.

For example:
From the graph below, if you look at the graph at rating 3.9 , you can observe that we have nearly 2500 restaurants that accepting online orders, and we also have number of restaurants that do not accept online order, but have 3.9 rating, might be just becuase of food, ambience, service, and so on..



Now lets go to the exciting machine learning model part to predict the rating of the restaurant based on the features like Online order, Book table, Rate, Votes, Location, Restaurant type, Cuisines,and Dish liked along with Average cost.

As we have Categorical variables like location, Restaurant type, cuisines , and dish, lets turn them into numerical data using **Label encoding**.

Split the Data in to X and Y variables for model building. After that split the whole dataset to Train and Test sets for model building.

We also have values ranging from 1 in rating to average cost 4000 rupees. Hence use **standardisation** to bring the values under one scale.

We applied 4 models and we got accuracy as below

- Linear regression - 21.07% accuracy
- Random forest - 89.45% accuracy
- Ridge regression - 21.07% accuracy
- Lasso regression - 18.05% accuracy

We got pretty low accuracy for Ridge, linear regression and Lasso regression. We got low accuracy, because we just used label encoding. Now Lets try to increase the accuracy using One-Hot encoding. So that there is chance to increase accuracy.

Now you can see below

- Ridge Regression - 88.45% accuracy
- Random Forest Regression - 90.65% accuracy

Lets fit the Random Forest model and see the output that our model predicted for us.

Index number	**Actual rating**	**Model Predicted Rating**
35957	4.0	4.0
4975	4.0	4.0
21830	3.9	3.9
11982	3.7	3.7
2597	3.8	3.8
35155	4.5	4.5
820	4.0	3.8
25002	3.5	3.6
8342	4.4	4.4

Thank you.