ZOMATO RESTAURANTS ANALYSIS AND PREDICTION

GROUP MEMBERS:

Harish Kumar(D12A-27)

Shikhar Niranjan(D12A-46)

Aniket Pawar(D12A-49)

INTRODUCTION TO PROJECT

- Amidst COVID-19 pandemic, one of the most grown business companies in India has been Zomato. The Zomato founder said that the revenue of the company grew 105 per cent in FY20 as compared to FY19, while costs only grew 47 per cent in the corresponding period.
- People being not able to eat their favourite savoury foods due to strict lockdown instructions made delivery boys to go and fetch their food for the same from their loved restaurants.
- The reason for selection of the project is the same one for visualizing their data as to what aspects of this company services does attract people and thus helping indian food chain of restaurants to scale their business in the same manner.

PROBLEM DEFINITION

- The basic idea behind Zomato Restaurants Analysis and Prediction project was to get a fair idea about the factors affecting the aggregate rating of each restaurant, establishment of different types of restaurant at different places, 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 hasn't been saturated yet and the demand is increasing day by day. In Spite of increasing demand it has become difficult for new restaurants to compete with established restaurants. 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.
- With such an overwhelming demand of restaurants it has therefore become important to study the demography of a location. This kind of analysis can be done using the data, by studying different factors.

DATA WAREHOUSING AND MINING TOOLS USED

- Jupyter Notebook
- Colab Notebook
- Orange Tool
- mySQL database server

DATASET USED

https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants/

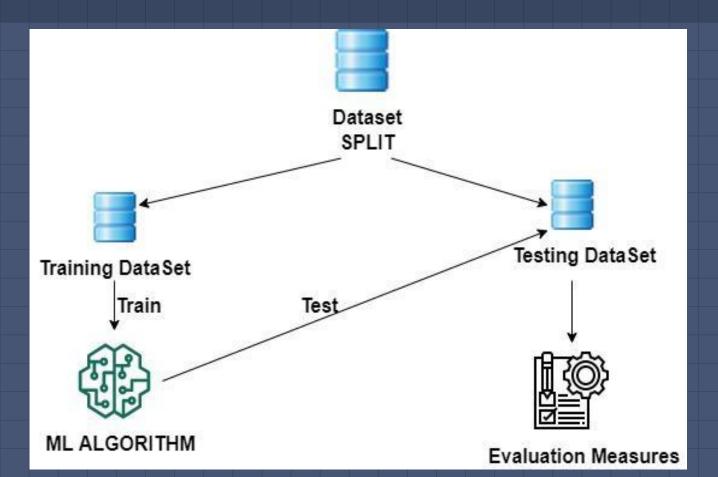
⇔ url	▲ address	▲ name	✓ online_order	✓ book_table	≜ rate	# votes
https://www.zom ato.com/bangalo re/jalsa- banashankari? context=eyJzZSI 6eyJlIjpbNTg2OT QsIjE4Mzc1NDc0I iwi	942, 21st Main Road, 2nd Stage, Banashankari, Bangalore	Jalsa	Yes	Yes	4.1/5	775
https://www.zom ato.com/bangalo re/spice- elephant- banashankari? context=eyJzZSI 6eyJlIjpbIjU4Nj k0IiwxODM	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th Block, Kathriguppe, 3rd Stage, Banashankari, Bangalore	Spice Elephant	Yes	No	4.1/5	787
https://www.zom ato.com/Sanchur roBangalore? context=eyJzZSI 6eyJlIjpbIjU4Nj k0IiwiMTgzNzU0N zQiLDU5MDkwLC	1112, Next to KIMS Medical College, 17th Cross, 2nd Stage, Banashankari, Bangalore	San Churro Cafe	Yes	No	3.8/5	918
https://www.zom ato.com/bangalo re/addhuri- udupi-bhojana- banashankari? context=eyJzZSI 6eyJlIjpbIjU4Nj k0	1st Floor, Annakuteera, 3rd Stage, Banashankari, Bangalore	Addhuri Udupi Bhojana	No	No	3.7/5	88

(51717 rows X 17 columns)

ALGORITHMS IMPLEMENTED

- Linear Regression
- Decision Tree Regression
- Random Forest Regression
- Extra Tree Regression

FLOW DIAGRAM



IMPLEMENTATION DETAILS

- Loading the dataset Load the data and import the libraries.
- Data Cleaning
 - a. Deleting redundant columns.
 - b. Renaming the columns.
 - c. Dropping duplicates.
 - d. Remove the NaN values from the dataset
 - e. Some Transformations(Changing data types of some columns)
- Regression Analysis
 - a. Linear Regression
 - b. Decision Tree Regression
 - c. Random Forest Regression
 - d. Extra Tree Regression

IMPLEMENTATION DETAILS

- Data Visualization Using plots to find relations between the features.
 - a. Restaurants delivering Online or not
 - b. Restaurants allowing table booking or not
 - c. Table booking Rate vs Rating
 - d. Best Location
 - e. Relation between Location and Rating
 - f. Restaurant Type
 - g. Gaussian Rest type and Rating
 - h. Types of Services
 - i. Relation between Type and Rating
 - i. Cost of Restaurant
 - k. No. of restaurants in a Location
 - I. Restaurant type

COMPARISON OF EVALUATION MEASURES USING VARIOUS ALGORITHMS

MODEL	R2_SCORE					
Linear Regression	0.27					
Decision Tree Regression	0.85					
Random Forest Regression	0.88					
Extra Tree Regression	0.94					

-1.00

- 0.75

- 0.50

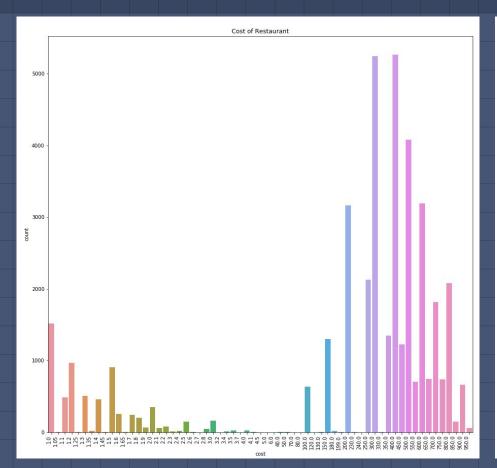
- 0.25

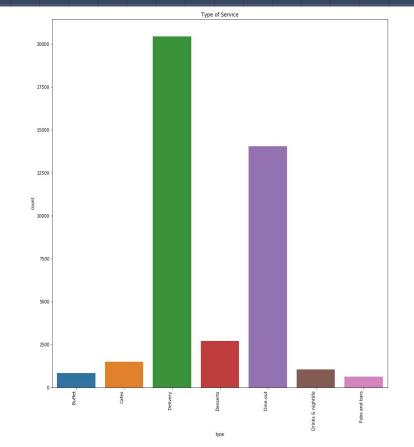
- 0.00

RESULTS (SCREENSHOTS)

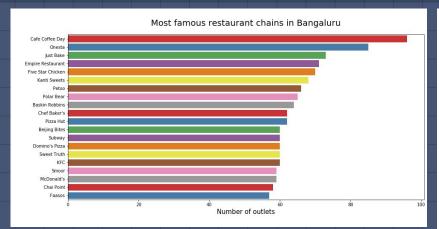
address -	- 1	0.62	0.14	0.016	-0.027	-0.045	0.4	0.015	0.12	0.00082	0.45	-0.022	0.076	0.28
name -	0.62	1	0.19	-0.04	-0.04	-0.09	0.3	0.0076	0.19	-0.068	0.29	-0.079	0.09	0.2
online_order	0.14	0.19	1	-0.055	-0.066	-0.13	0.056	0.11	0.019	-0.15	0.043	-0.4	0.23	0.047
book_table ·	0.016	-0.04	-0.055	1	-0.38	-0.38	-0.053	-0.054	-0.19	0.22	-0.11	0.043	-0.1	-0.025
rate -	-0.027	-0.04	-0.066	-0.38	1	0.53	0.044	0.083	0.15	-0.061	0.13	0.052	0.024	0.019
votes -	-0.045	-0.09	-0.13	-0.38	0.53	1	0.035	-0.034	0.15	0.034	0.15	0.067	0.03	0.008
location -	0.4	0.3	0.056	-0.053	0.044	0.035	1	0.037	0.096	-0.025	0.3	0.014	0.041	0.26
rest_type -	0.015	0.0076	0.11	-0.054	0.083	-0.034	0.037	1	0.026	-0.23	0.026	-0.0086	0.028	0.022
cuisines -	0.12	0.19	0.019	-0.19	0.15	0.15	0.096	0.026	1	-0.031	0.11	0.0059	0.053	0.065
cost -	0.00082	-0.068	-0.15	0.22	-0.061	0.034	-0.025	-0.23	-0.031	1	0.0033	0.066	-0.12	-0.017
reviews_list -	0.45	0.29	0.043	-0.11	0.13	0.15	0.3	0.026	0.11	0.0033	1	0.075	0.06	0.61
menu_item ·	-0.022	-0.079	-0.4	0.043	0.052	0.067	0.014	-0.0086	0.0059	0.066	0.075	1	-0.11	0.12
type -	0.076	0.09	0.23	-0.1	0.024	0.03	0.041	0.028	0.053	-0.12	0.06	-0.11	1	0.028
city -	0.28	0.2	0.047	-0.025	0.019	0.008	0.26	0.022	0.065	-0.017	0.61	0.12	0.028	1
	address -	name -	online_order -	book_table -	rate -	votes -	location -	rest_type -	cuisines -	cost -	reviews_list -	menu_item -	- phe -	aty -

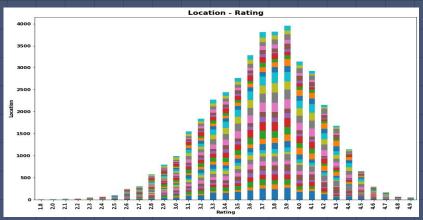
RESULTS (SCREENSHOTS)

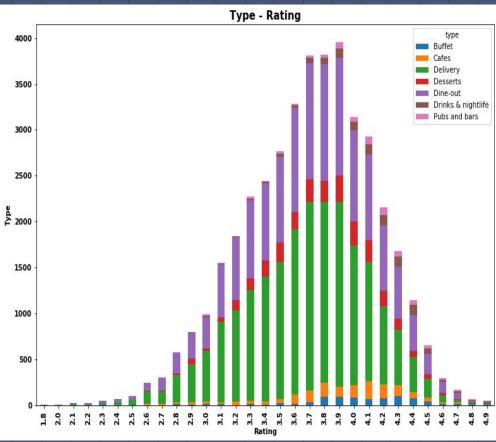




RESULTS (SCREENSHOTS)







CONCLUSION

Thus, we have effectively analysed the dataset and visualized according to different parameters and predicted the rating of restaurants based on those parameters. As a result, we have successfully studied various Data Warehousing and Mining concepts and applied algorithms to our problem statement Zomato Restaurants Analysis and Prediction on their dataset.

REFERENCES

- https://towardsdatascience.com/
- https://www.kaggle.com/
- https://scikit-learn.org/stable/
- https://matplotlib.org/
- https://seaborn.pydata.org/
- https://pandas.pydata.org/