

## Declaration on Plagiarism

*This form must be filled in and completed by the student(s) submitting an assignment*

<b>Name:</b>	Joseph Cherian
<b>Student Number:</b>	19210222
<b>Programme:</b>	MCM1
<b>Module Code:</b>	CA682
<b>Assignment Title:</b>	Data Visualisation
<b>Submission Date:</b>	13 Dec 2019
<b>Module Coordinator:</b>	Dr Suzanne Little

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the referencing guidelines found at <http://www.dcu.ie/info/regulations/plagiarism.shtml>, <https://www4.dcu.ie/students/az/plagiarism> and/or recommended in the assignment guidelines

Name: Joseph Cherian

Date: 20-11-2019

## Project Title or Headline

# Zomato Bangalore Restaurants Analysis

Abstract (max 200 words)

*What is the question or story you are trying to tell?*

The main idea of analysing the zomato dataset is to get a fair idea about the factors affecting the establishment of different types of restaurants at different locations in Bangalore, India.[1] Most importantly it will help new restaurants in deciding their theme, menus, cuisine and cost for a particular location. It also aims at finding similarity between neighbourhoods of Bengaluru on the basis of food. These are the questions I am trying to answer: 1. What is the average cost for 2 people in rupees generally in a restaurant in Bangalore? 2. What are the types of restaurants in general in Bangalore? 3. Does online order facility impact the rating of the restaurant?

*What is the conclusion that you reached?*

I was able deduce some interesting results. The average cost for 2 people in a restaurant in Bangalore is mostly Rupees 300 indicating that people in Bangalore are generally middle-class people. Generally, the types of restaurants in Bangalore are Quick Bites followed by Casual Dining. The higher the rating of the restaurant the more likely it has online ordering. This can be inferred from the horizontal plot I have created.

## 1. Dataset [½ page]

*Where/how did you retrieve it or them*

I retrieved the data from the Kaggle website. The link for the dataset is <https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants>. The data is accurate for the records in the **Zomato** website(<https://www.zomato.com/>) until 15 March 2019.

*Describe the data - size (GB or attributes), number of rows, attributes, data types present  
What aspects (if any) of big data (volume, variety, velocity) are present in your data*

The data is 547 MB in size. The data consists of 51717 rows and 17 columns.

The dataset contains the following columns:

- 1.url (Datatype-URL): Contains the url (Universal Resource Locator) of the particular restaurant on the zomato website.
2. address (Datatype-String): Address of the restaurant in Bangalore city.
3. name (Datatype-String): Name of the restaurant
4. online\_order(Datatype-Boolean): whether the restaurant provides online ordering option or not.
5. book\_table (Datatype-Boolean): Table booking option is available at the restaurant or not.
6. rate (Datatype-String): Overall rating of the restaurant.

7. votes (Datatype-Integer): The total number of people who have rated the restaurant as of 15 March 2019.
8. phone (Datatype-String): The phone number of the restaurant.
9. location (Datatype-String): The neighbourhood in which the restaurant is located.
10. rest\_type (Datatype-String): Restaurant type
11. dish\_liked (Datatype-String): Dishes reviewers liked in the restaurant.
12. cuisines(Datatype- String): Food styles offered by the restaurant.
13. approx\_cost(for two people)(Datatype- Decimal): contains the approximate cost for a meal for two people
14. reviews\_list(Datatype- String): list of tuples containing reviews for the restaurant, each tuple consists of two values, rating and review by the customer.
15. menu\_item(Datatype-String) contains list of items available in the menu of the restaurant.
- 16.listed\_in(type)(Datatype-String): type of meal
- 17.listed\_in(city)(Datatype- String): contains the neighbourhood in which the restaurant is located.

In the data **volume** characteristic of big data is present. The data is extracted from **Zomato** website and has more than 50,000 records.

## 2. Data Exploration, Processing, Cleaning and/or Integration [½ page]

*What did you need to do to prepare the dataset(s) to create your graph/chart?*

*How did you choose the attributes to visualise?*

The unwanted columns which we are not using for our visualisation needs to be removed.

We can delete the columns **url**, **address** and **phone** as these contain the contact details of the restaurants in Bangalore which is not required for data visualisation.

**menu\_item** column name can be deleted as it contains a lot of empty records. Also, the dataset contains attributes like **rest\_type**, **cuisines**, **listed\_in(type)** and **dish\_liked** which give us an idea about what the restaurants are offering.

**reviews\_list** attribute also can be deleted as it is pure text and we have features like **rate** and **votes** that can give us necessary information.

We have 2 attributes **location** and **listed\_in(city)** that give us similar information according to the description in Kaggle. Here as there are 0 null values for the attribute **listed\_in(city)** and there are 21 null values for the **location** attribute, we delete the location feature.

## Data Cleaning

The **rate** attribute is a string as it contains the character '/'. We can remove the '/5' from each entry in the column, changing the datatype of the attribute to float.

The following attributes are renamed:

approx\_cost(for two people) : avg\_cost\_2

listed\_in(type) : meal\_type  
listed\_in(city) : Area

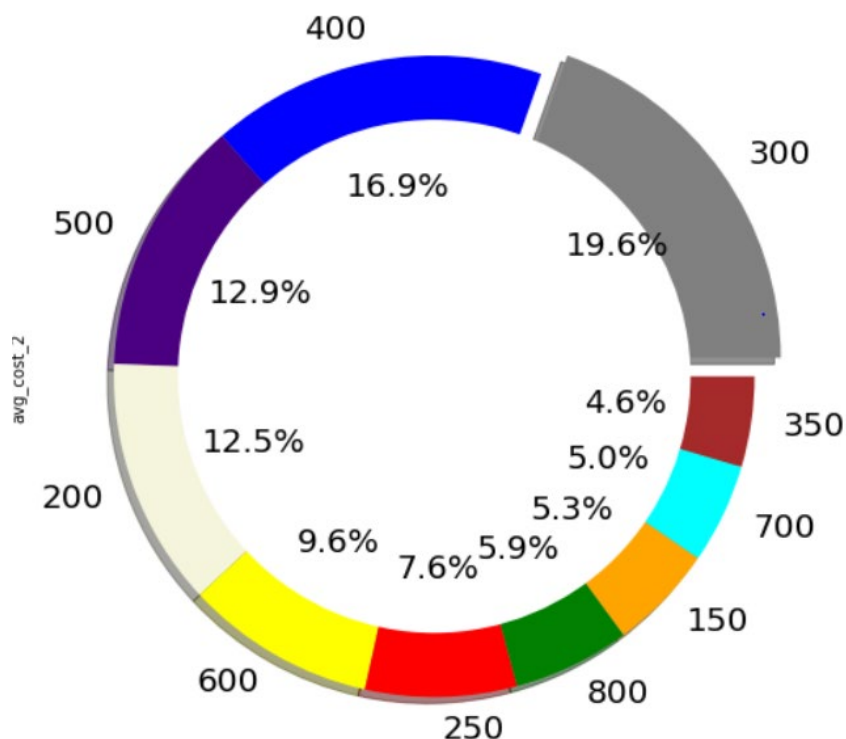
Some records have NEW and '-' values for the rate attribute. I have changed these NEW values and '-' to nan (not a number) values. I have filled all the missing data for the **rate** and **rest\_type** attributes in the Zomato dataset using nan value.

I chose the attributes to visualise from the questions I was trying to answer. The attributes I chose were **avg\_cost\_2**, **rest\_type**, **online\_order** and **rate**.

### 3. Visualisation [½-1 page]

*Screenshot or image of visualisation*

#### **Average cost for 2 people in Rupees**



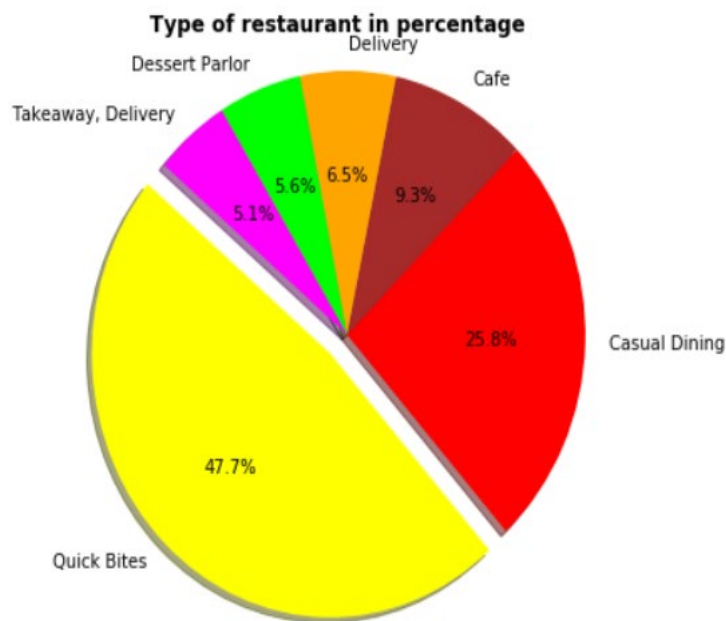
*Explain your choice of chart or graph type - what relationship or data type are you showing?*

I have used a Donut graph here. I am trying to visualise the Average cost for 2 people in rupees in a restaurant in Bangalore. Hence it is appropriate to use a donut graph as I can clearly see what is the Average cost for 2 people in rupees in a restaurant in Bangalore mostly which is 300 rupees here. I am showing the percentage for the top 10 value counts for the average cost for 2 people in Rupees.

*Design choices - justify your use of colour, shapes, marks, layout, structure, font, labels*

I have used different colours grey, blue, indigo, beige, yellow, red, green, orange, cyan and brown for the top 10 value counts for the average cost for 2 people in Rupees to identify the various value counts clearly. I have given 0.1 for the first value of the explode variable to distinguish the average cost of 2 people in a restaurant in Bangalore in rupees generally from the remaining values of the average cost of 2 people in rupees depicted in the donut graph.

*Screenshot or image of visualisation*



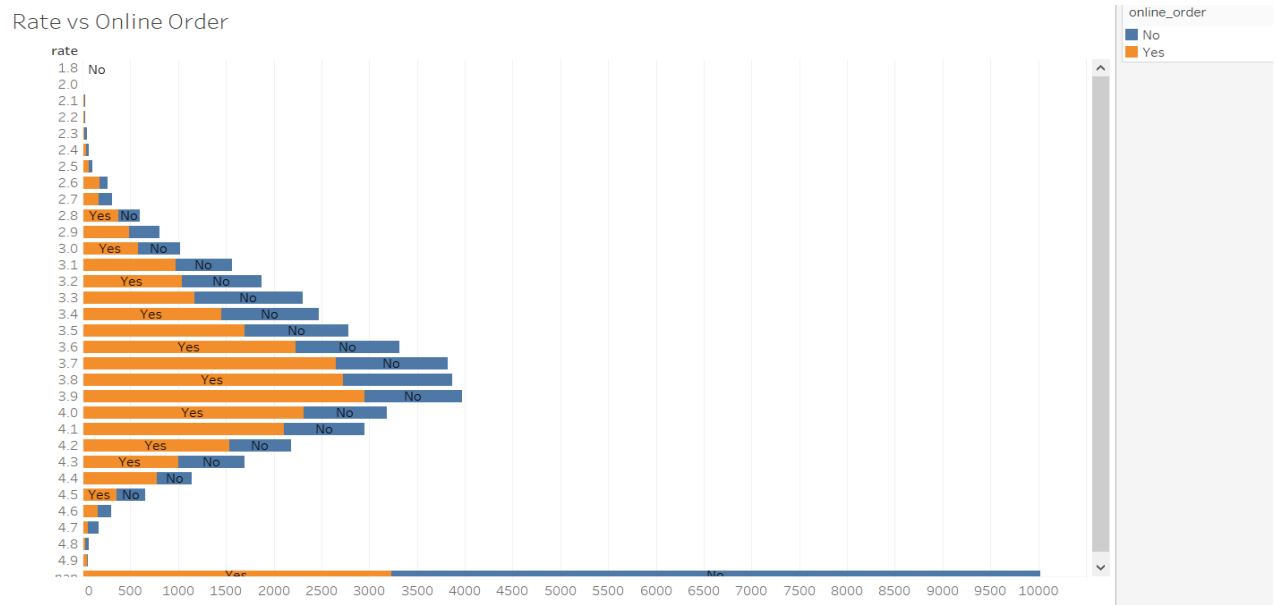
*Explain your choice of chart or graph type - what relationship or data type are you showing?*

I have used a Pie chart as it is the most suitable chart or graph type to clearly identify the percentage of the different restaurant types in Bangalore clearly.

*Design choices - justify your use of colour, shapes, marks, layout, structure, font, labels*

I have used different colours yellow, red, brown, orange, light green and purple to identify the percentage of the top 6 restaurant types in Bangalore. I have given 0.1 for the first value of the explode variable to distinguish the highest percentage for the restaurant type in Bangalore from the remaining restaurant types depicted in the pie chart.

## Screenshot or image of visualisation



*Explain your choice of chart or graph type - what relationship or data type are you showing?*

I have used a horizontal plot to identify the relationship between the count of the number of restaurants providing online order or no online order for a particular rating. I have used horizontal plot here as it is easy to identify the number of restaurants providing online order and no online order when compared to pie chart or tree maps.

*Design choices - justify your use of colour, shapes, marks, layout, structure, font, labels*

I have used orange and blue colour to identify the number of restaurants providing online order or no online order for a particular rating. I have used labels for the online order field to find out the number of restaurants providing online order and no online order.

*Any interactivity or animation and how it helps answer your question*

I have used interactivity in the horizontal plot I have created in tableau for Rate vs Online Order. It helps answer the question as I can view the number of records that have online order and no online order for a particular rating.

*List of tools or libraries used*

Tools used- Tableau, Python 3

Libraries used- Numpy, Pandas, Matplotlib and Seaborn API in Python

#### 4. Conclusion [½ page]

*Critically analyse the outcome of your visualisation.*

##### 1st Visualisation

The top 10 values for the average cost for 2 people (in rupees) is between 150 to 800. The average cost for 2 people (in rupees) in a restaurant in Bangalore is mostly 300 Rupees (19.6%). These statistics indicate that people in Bangalore are mostly middle-class people.

##### 2nd Visualisation

The most frequently visited restaurant type in Bangalore are Quick Bites followed by Casual Dining. In Bangalore most of the people are software employees who are very busy hence they would prefer to have quick meals rather than whole meals.

##### 3rd Visualisation

Majority of restaurants have rating 3.3 to 4.1, no matter whether you order online or not. The higher the rating of the restaurant the more likely it has online ordering. It makes sense because many software employees stay in Bangalore and they tend to order a lot of food online.[1]

*Were there aspects that you think could be improved upon?*

The Rate vs Online Order horizontal plot could have been made more interactive

*Were there effects or functionality that you were technically unable to achieve?*

Animation

#### References

*Include any citation of the dataset*

Dataset link-- <https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants>

*Include links to any tutorial or example that contributed significantly to your work*

1. <https://www.marktechpost.com/2019/07/13/exploratory-data-analysis-tutorial-analyzing-the-food-culture-of-bangalore/?cn-reloaded=1&cn-reloaded=1>
2. <https://www.youtube.com/watch?v=DIgsN2-SHWo>

*Include any articles or web resources supporting your design choices*

# References

- [1] "Mark\_Tech\_Post," [Online]. Available:  
<https://www.marktechpost.com/2019/07/13/exploratory-data-analysis-tutorial-analyzing-the-food-culture-of-bangalore/?cn-reloaded=1&cn-reloaded=1>.
- [2] "Kaggle," [Online]. Available: <https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants>.