**Name : Raghavendra J P**

**SRN : PES1201802162**

**Email : raghavendrajp5132@gmail.com**

**Part 1: Data Set Identification and Analysis**

**Dataset name: Online Food Delivery Preferences-Bangalore region**

**Introduction**

The dataset is chosen from Kaggle.com which contains most of the user information, delivery location, food preference of the user, user ratings for many attributes, delivery details and food quality etc. Total of 55 columns and 388 rows are available for exploring.

**Source :** https://www.kaggle.com/benroshan/online-food-delivery-preferencesbangalore-region

**Columns in dataset**

|  |  |
| --- | --- |
| Column name | Description |
| Age | # Age of Consumer |
| Gender | # Gender of Consumer |
| Marital Status | # Marital Status of Consumer |
| Occupation | # Job or Occupation status of Consumer |
| Monthly Income | # Income bracket of consumer |
| Educational Qualifications | # Education of consumer |
| Family size | # Number of family members/ friends living with |
| Latitude | # latitude of residence |
| Longitude | # longitude of residence |
| Pin code | # pincode of residence in Bangalore |
| Medium (p1) | #medium of order(Preference 1) |
| Medium (p2) | #medium of order(Preference 2) |
| Meal (p1) | #type of meal order(Preference 1) |
| Meal (p2) | # Type of meal order(Preference 2) |
| Preference (p1) | # Cuisine (Preference 1) |
| Preference (p2) | # Cuisine (Preference 2) |
| Ease and convenient | # Ease and convenience of online ordering |
| Time saving | # Does it save time ? |
| More restaurant choices | # More restaurant choice influences |
| Easy Payment option | # Payment option influences |
| More Offers and Discount | # Offers and discount influences |
| Good Food quality | # Food quality influences |
| Good Tracking system | # Tracking system influences |
| Self-Cooking | # Self cooking causes not purchasing |
| Health Concern | # Health concern causes not purchasing |
| Late Delivery | # Later Delivery causes not purchasing |
| Poor Hygiene | # Poor Hygiene causes not purchasing |
| Bad past experience | # Past experiences causes not purchasing |
| Unavailability | # Unavailability causes not purchasing |
| Unaffordable | # Un-affordability causes not purchasing |
| Long delivery time | # Long delivery causes cancellation |
| Delay of delivery person getting assigned | # Delay of delivery person assigned causes cancellation |
| Delay of delivery person picking up food | # Delay of delivery person waiting at restaurant causes cancellation |
| Wrong order delivered | # Previous wrong order causes cancellation |
| Missing item | # Missing item in order causes cancellation |
| Order placed by mistake | # Placed order by mistake causes cancellation |
| Influence of time | # Is there any influence in time ? |
| Order Time | # When do you order? |
| Maximum wait time | # How long can you wait ? |
| Residence in busy location | # Residence in busy location |
| Google Maps Accuracy | # My location in google maps is accurate |
| Good Road Condition | # My residence area road condition is good |
| Low quantity low time | # low quantity low delivery time |
| Delivery person ability | # Delivery person ability depends on time of delivery |
| Influence of rating | # Is there any influence in restaurant rating ? |
| Less Delivery time | # Importance of Less delivery time |
| High Quality of package | # Importance of Quality of package |
| Number of calls | # Importance of Number of calls made by delivery captain |
| Politeness | # Importance of Politeness of delivery captain |
| Freshness | # Importance of Freshness of food |
| Temperature | # Importance of Temperature of food |
| Good Taste | # Importance of taste |
| Good Quantity | # Importance of Quantity in food |
| Output | #Will I purchase again ? |
| Reviews | # Feedback on online delivery services |

**Analysis of Contents in Dataset**

The dataset contains a lot of attributes that can represent most of the necessary data needed for exploring food-delivery.

I don’t think any of the attribute is missing that can cause us problems in understanding the data. This is the best dataset I have found that is latest uploaded on 8/11/2020 and only 9 people have downloaded so the data analysis that will be done on this dataset will not have plagiarism throughout the whole internet.

The data for each individual customer history, data of the delivery associate could have been given in separate files with unique key to represent each of them. This would help us expand our viewing on the complete online food delivery environment.

The dataset is free from missing and invalid values so there is no need for cleaning of the dataset.

For attributes containing unique values less than 10 which are represented in string format I will be assigning numbers for each unique value that will help visualizing and finding correlation using machine learning algorithms. Example for Monthly Income the data is a string and also does contain the exact value but an estimate of min to max only. This will be solved by dividing into classes.

There are many customers who have not given any reviews. Review attribute is a sentence so we need sentiment analysis and text classification with natural language processing to classify them as positive, neutral or negative. The Pin code attribute is of no use since the dataset has latitude and longitude. Other than that, there are no limitations in the dataset since it has more than enough attributes that can give a complete visualization from all the possible direction a person can think.

There is only single file completing all the attributes combined together.

The data does not contain deep personal information of an individual that can track that user and all the data is given by the customer which he can prefer to give or prefer not to so the data is completely ethical for using it for exploration.

The dataset contains information of latitude and longitude that can be used in a beautiful manner for plotting a map of Bangalore with several marks.

I will be using a mix of tableau and python for visualization and analysis.