Analysis of factors affecting ratings of restaurants

Punit Lohani
x18127339
Analytical CRM
M.Sc. Data Analytics
National College of Ireland

***Abstract*— With the advancements in the technology and growth of internet, the e-commerce industry has gained a lot of attention among the people living in the country like India. With this, the online ordering of food has gained interest among people as they can order the foods and beverages of choice by sitting at their place. Although, customer goes out for having their meals but online food delivery makes their life easy. This research focuses on customer preferences i.e. the type of cuisines and location that are mostly preferred and various factors that affects the ratings of the restaurants that have tie ups with the leading food aggregator operational in India i.e. Zomato.**

Index Keywords—internet, e-commerce, online ordering, customer preferences, food aggregator.

I. INTRODUCTION

Ordering the food on the internet is different in terms of physically visiting the restaurant. Internet helps the customer to order the food by sitting at their place and enables them to interact with the seller through customer care number in case of any issues. Today, the business of online delivering the food has gained a lot of popularity and attention and it is one of the fastest growing business in the field of e commerce. Ordering the food online not only helps the customers but also it helps the business owners to expand their business. Now a days, the restaurants have made their tie ups with the leading online food delivery provider companies in order to expand their business.

Consumer preference analysis is one of the useful and powerful tool for the companies that are dealing in variety of business around the world especially in the food industry. With the effective and efficient analysis the organization can track the flow and look for the area of concerns as stated by the customer and this may help them to a great extent in improving the business and achieving the targets and goal. The world is now quite competitive with variety of organizations dealing in the same type of businesses hence, proper analysis is considered to be an important factor to survive in the market. Improving the quality of business can help the organization to retain the existing customer and in the same time attracts new customers.

II. HYPOTHESIS

The null hypothesis of the analysis is:

H_0: Aggregate rating of the restaurant is depended on the Average cost for two, votes and price range.

The main aim of the analysis would be to find out the relation of aggregate rating with the Average cost for two, votes and price range.

III. BACKGROUND OF DATASET

The dataset that has been used in the project is providing information about the restaurants of Zomato that are present worldwide. The Dataset has been downloaded from Kaggle website. The information such as restaurant name, country, city, cuisines that is being served, average cost for two, price range, aggregate rating and votes has been taken into consideration for the analysis purpose.

VARIABLES	DESCRIPTION
Country name	Contains the name of the country
City	City where the restaurants are located
Cuisines	Types of cuisine that the restaurant is offering
Average cost for two	Cost of two people on visiting the restaurant.
Price range	The range of price from least to highest
Aggregate rating	Rating received by the restaurant from 1 to 5
Votes	Number of votes casted by the people .

Fig. 1. *Description of dataset*

Country code	Country Name
1	India
30	Brazil
214	UAE
215	United Kingdom
216	United States

Fig. 2. *Description of dataset*

IV. Data Preprocessing

Data preprocessing is one of the crucial step for performing the analysis. The downloaded dataset had several columns and the columns that are needed for analyzing the dataset has been taken into consideration. For cleaning the data, R studio has been used where steps like renaming of column so as to make them understandable and removing the redundant columns has been performed. Moreover, the use of Excel has also been made for cleaning the data in detail. In the dataset only country code was mentioned, so as to understand it, the extra column containing the Country name has been introduced. The dataset contains 21 columns in total and the useful columns have been picked so as to proceed with the analysis.

V. LITERATURE SURVEY

Identification of the factors that influences most of the customers while making an order for food online has been analyzed in [1] . Additionally, the objective is to find about the preferences of customers related to the various providers that are available online. The data collection for the study has been sourced through online survey. The outputs from the study states that the price of the product influences most of the people to order online followed by ease of getting the product delivered at home. Additionally, fast food and north Indian food are widely preferred by the customers, while Zomato being the best provider in terms of service and quality to most of customers.

Various factors that contribute towards the choice of various cuisines by the Indian customers has been presented by the researcher in [2]. Online survey has been conducted in order to get the insight and the logistic regression has been employed and the method used for identifying the sample was quota sampling. The output states that with an increase in age, the consumer starts preferring Indian cuisine while the Italian dishes are mostly preferred by the youngsters. Also it has been inferred that the choice of cuisines depends from city to city and number of restaurants present.

The main focus on how the variety of features that are offered by the restaurants impacts the overall ratings and review has been discussed in [3]. The data has been scraped from an online website and focuses on the Andaman and Nicobar region. R and logistic regression technique has been used. The outcome from the dataset stated that restaurants providing home delivery and parking facility tends to have good rating and review as compared to other restaurant and concludes that if the restaurant provides better services and features to customers, more likely they get good reviews and ratings.

This research [4] focuses on the satisfaction of the customer and the importance and effects of factors like quality of food, price and environment that helps in satisfying the customers. The collection of data has been made through the fast food court. SPSS and structure equation model technique has been implemented. The result shows that the quality of food plays a crucial role and has the positive impact in the satisfaction of customer followed by price and the overall environment and in order to retain and satisfy the customer the restaurant must majorly focus on these factors.

The work by the researcher [5] basically focuses on the impact of average score given online by the customers and its relation with the quality of the product. The data has been collected from Amazon and the useful analysis along with the mathematical models has been applied. The result shows that the average rating casted by the customer like 3 shows that the almost equal number of customers liked and disliked the product. Sometimes this may mislead the purchase decision. The average score can reveal the quality if more and more consumers present their views online.

The insight about the customers and their preference towards choosing the restaurant to eat has been discussed in [6]. A factor that mainly contributes towards decision making includes location, reputation, the food quality it is offering along with the food type. Also it has been identified that the selection of restaurant depends on other factors like the age of the customer as the preference changes over a period of time, any special occasion, prior information and knowledge about the restaurants etc.

The study [7] presents a crucial information about the ordering experience of the customers that can help in the operation of food services online. Factors like trust on the website, satisfaction and loyalty are included. After collection of data, SEM applied to present the outcome of the research. The result shows that the there is a significant relation between the quality of service and satisfaction among the customers along with the relation between the trust on the website and satisfaction of customer. In addition, delivery of food also leads to the positivity and results in satisfying the customers.

VI. Research Methodology

The objective of the project is to first analyze the preferences of the customers. Since the majority of restaurants are in India so it has been taken into consideration. Apart from this, it has been identified that in India which cities are having maximum number of restaurants. So it has been inferred that top 3 cities are New Delhi, Gurgaon and Noida. Then in order to get the clear understanding, further the data has been divided into number of restaurants present in these cities. Further, analysis

involves that what factors affects the aggregate rating of the restaurants. For this, data has been divided into dependent and independent variables. Our dependent variable involves Aggregate rating and the independent variables involve average cost for two, votes, and price range. In order to proceed further, finding the correlation was the first step where correlation matrix has been drawn so as to find the correlation among these variables. For plotting the correlation matrix, R code has been used. Furthermore, to analyze better multiple regression would be applied so that the variables that affects the aggregate rating can be found. With this, we can get the understanding that which among the three factors affects the aggregate rating of the restaurants. R code and visualization tools like Data wrapper and Power BI has been used in the project for visualizing the data so as to make them understandable and easy to interpret.

VII. IMPLEMENTATION OF TECHNIQUES

The histogram plot of the data represents the normal distribution of data. Therefore, the further processing of the data can be done by implementing the technique like multiple regression and correlations in order to find the dependency between the dependent and independent variables.

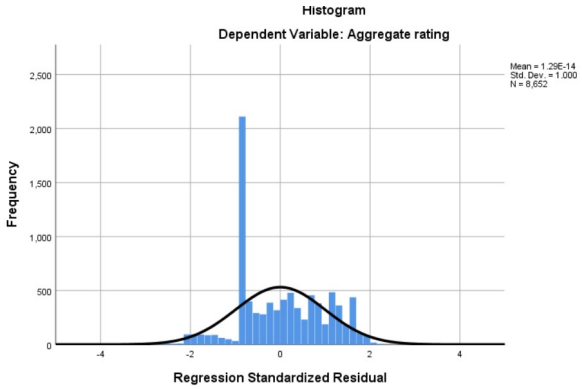

Fig. 3. *Histogram plot of the Dataset*

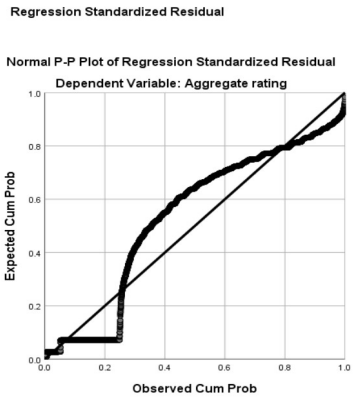

Fig. 4. *Normal P-P plot of the Dataset*

Fig. 5.

To proceed with the analysis, Independent and dependent variables has been taken into account and these are:

Independent variables:

- Price range: The range of price from least to highest
- Average cost for two: Cost of two people on visiting the restaurant
- Votes: Number of votes casted by the people

Dependent variable:

- Aggregate Rating: Rating received by the restaurant from 1 to 5

The correlation matrix between the dependent and the independent variables is presented below:

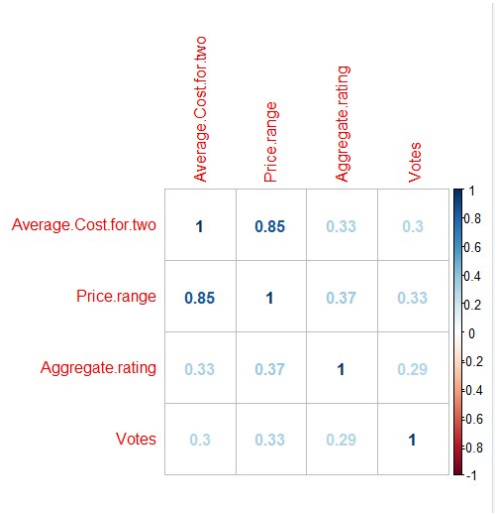

Fig. 6. *Correlation matrix for dependent and independent variables*

Correlations

		Aggregate rating	Average Cost for two	Price range	Votes
Pearson Correlation	Aggregate rating	1.000	.344	.409	.288
	Average Cost for two	.344	1.000	.838	.282
	Price range	.409	.838	1.000	.311
	Votes	.288	.282	.311	1.000

Fig. 7. *Output of Correlation values from SPSS*

The above mentioned correlation matrix represents that the correlation between the Aggregate rating and Average cost for two is 33% while between the Aggregate rating and Price range is 37%. Moreover, between Aggregate rating and votes the correlation is of 29%. This shows that out of these three the aggregate rating majorly depends on the Price range and Average cost for two.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1.350	.037		36.181	.000
	Average Cost for two	-2.079E-5	.000	-.008	-.464	.643
	Price range	.639	.032	.361	20.205	.000
	Votes	.001	.000	.178	17.518	.000

a. Dependent Variable: Aggregate rating

Fig. 8. *Coefficients table generated from SPSS output*

The above mentioned analysis has been done from SPSS output by performing the multiple regression. The significance value of price range and votes is 0.000 which is less than 0.05. This shows that only these two variables contributes statistically significant in explaining the dependent variable.

On referring the standardized coefficients Beta value, it can be seen that price range is having the value i.e. 0.361 as compared to other variables. This shows that it shows its highest contribution in explaining the aggregate ratings of the restaurants.

- **Top 5 countries having maximum number of restaurants**

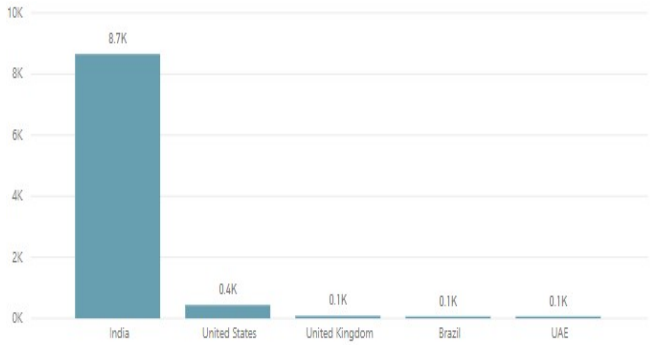

Fig. 9. *Number of restaurants by country*

The bar graph represents the top 5 countries in which the zomato has its major restaurants. India is the leading country having maximum number of restaurants i.e. 8653 while the United States is the second largest country with 435 restaurants followed by United Kingdom, UAE and Brazil.

- **Top 3 cities having maximum number of restaurants**

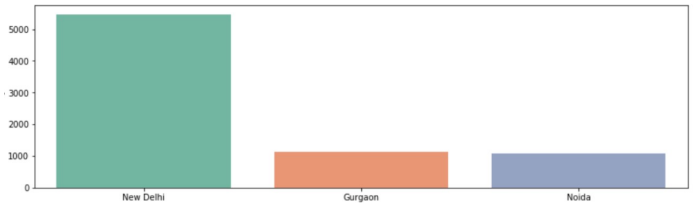

Fig. 10. *Major cities and number of restaurnats*

As India is the leading country having maximum number of restaurants, so the focus has been laid on the cities in India that are having maximum number of registered restaurants. Here, we have taken top 3 cities to get an insight about the existence of restaurants. As can be clearly observed by the graph, majority of registered restaurants are in New Delhi that are 5473, followed by Gurgaon and Noida with 1118 and 1080 restaurants respectively. So the focus will be on these major cities.

- **Comparison between price range and number of restaurants**

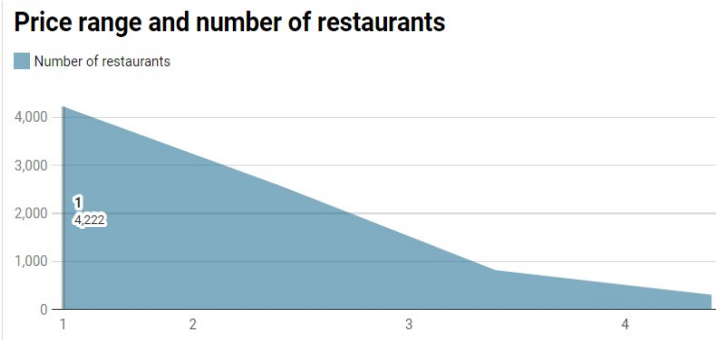

Fig. 11. *Comparison of price range and number of restaurants*

The above graph represents the information about the number of restaurants and the price range. Basically the price range from 1 to 4 indicates the range of restaurants from cheapest to highest price. As can be seen that the majority of restaurants are having the price range of 1 that means the price ranges between 50 INR to 450 INR. While the minimum number of restaurants are having their price range of 4 which greater than 2000 INR and less than 8000 INR. This shows that more than 50% of registered restaurants are having their price range of 1.

- **Comparison of cuisines preferences**

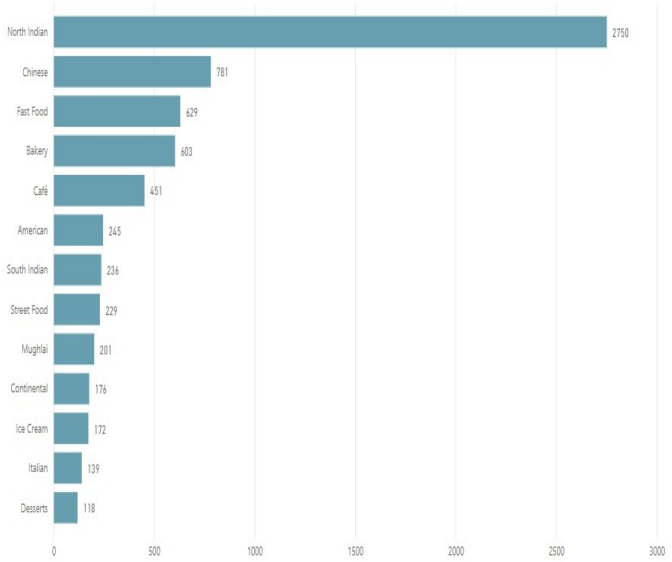

Fig. 12. *Popularity of cuisines*

The above graph depicts the popularity of cuisines among the people. It can be seen that the mostly all the restaurants are serving the North Indian food and have the highest number of counts which shows the majority of people prefers to have North Indian food when they visits the restaurants. Apart from this, other cuisines like Chinese, Fast Food, Mughlai, American and South Indian are preferred after the North India. This states the craze of North Indian food among the people.

- **Comparison of Aggregate rating and its count**

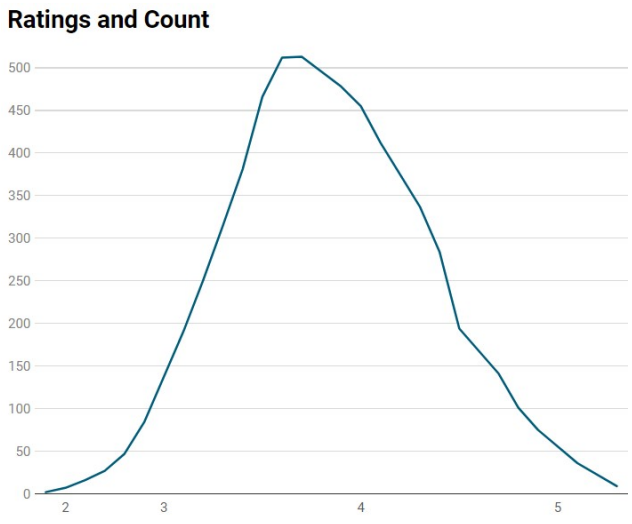

Fig. 13. *Aggregate rating and counts*

The mentioned graph represents the comparison between the aggregate rating and the corresponding counts. The trend shows that the majority of restaurants are having their aggregate rating in the bracket of 2.9 to 3.8.

- **Online delivery restaurants and their ratings**

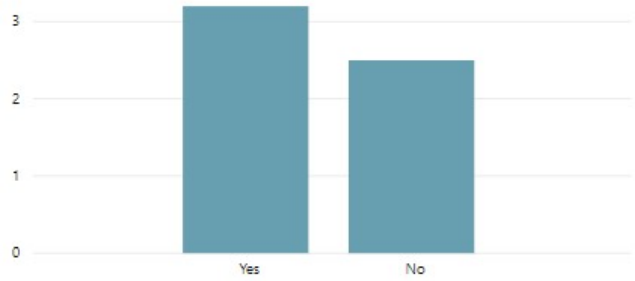

Fig. 14. *Comparison of restaurants having online delivery and their ratings*

Graph is representing the information about the restaurants that are having and not having online delivery facility. It can be inferred that those who are having online delivery facility have higher average rating than the one who don't have this facility.

- **Restaurants with table booking and their ratings**

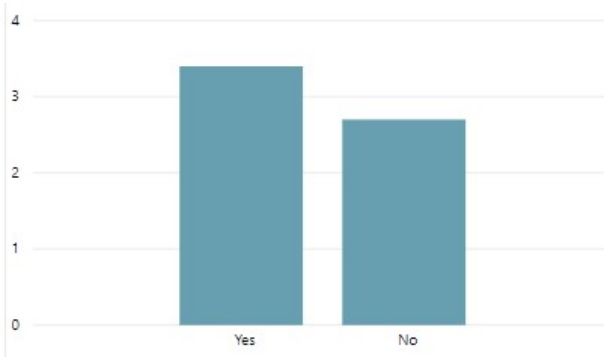

Fig.13. Comparison of restaurants having table booking and their ratings

The above graph elucidates the information about the restaurants that are providing table booking facility to the customer and their average rating. It can be inferred that the restaurants that have table booking facility have slightly

higher average rating as compared to those whose done provide this facility.

- **Comparison of rating texts and votes**

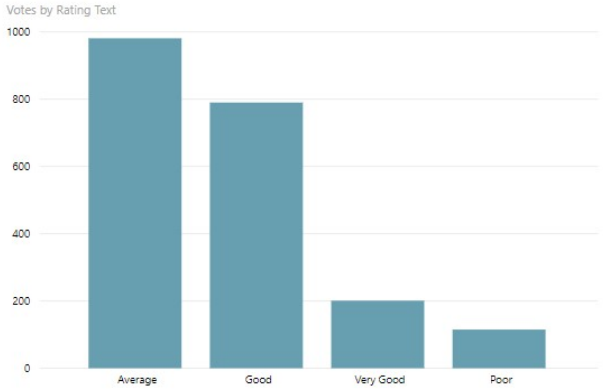

Fig.14. Comparison of rating texts and votes

The graph represents the information between the rating text and the number of votes for the restaurants that are providing online delivery to the customer. It can be seen that the majority of restaurant are rated Average by the people who means their rating is between the range of 2.5 to 3.4. There is also a significant correlation between the rating text and the votes. It can be inferred that the restaurant that provides home delivery are rated slightly higher.

- **Comparison of rating and number of votes**

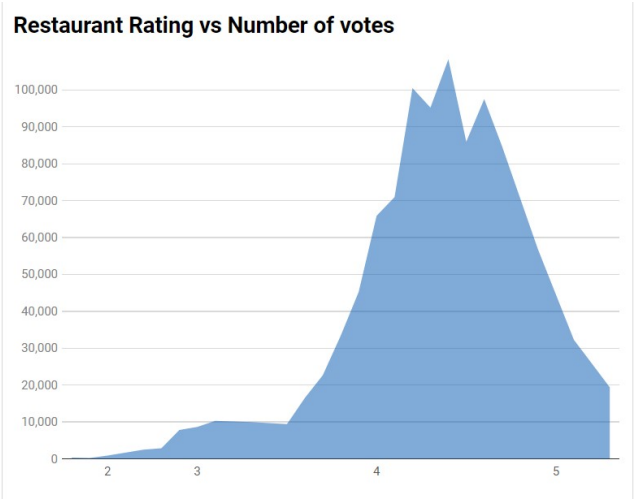

Fig.15. Comparison between how the rating varies with the number of votes

The above graph is representing the information about the no. of votes and ratings of the restaurants. It can be inferred from the graph that the restaurants that are having lesser number of votes are also having less rating and as the number of votes are increasing, the increase can be seen in the rating as well. Restaurants that are having rating between 3.5 to 4 are having more number of votes corresponding to them. While rating greater than 4 comparatively have lesser number of votes.

VIII. RESULTS

By observing the result that has been obtained, the hypothesis is accepted that the aggregate rating is dependent on the price range followed by average cost for two and the number of votes casted. The correlation matrix states that there is a high correlation between the aggregate rating and the price range.

After analyzing the other features that may affect the aggregate ratings, some other interesting results were found like, the restaurants that are providing the online delivery to the customer have slightly better aggregate rating. Addition to this, the feature of table booking may also impact the aggregate rating to some extent.

IX. CONCLUSION

With the help of visualizing the data, the preferences of customers have been identified like the type of cuisines they prefer and the cities where maximum of number of registered restaurants are located. In addition, the factors have been identified that affects the ratings of the restaurants. The above analysis can give an insight to the leading food operating aggregator by looking at the trend of consumer preferences and the localities that are mostly preferred. In this way, they can provide the consultancy to the new restaurants that are facing difficulties and in this way restaurants can improve and expand their business. While on the other side, the leading food operating aggregators can retain their tie ups with these restaurants, which in return helps them as well to increase their revenue and expansion of business.

REFERENCES

[1] S. Rathore and M. Chaudhary, "Consumer's Perception on Online Food Ordering", *International journal of managment and business studies*, vol. 8, no. 4, 2019. [Accessed 23 July 2019].

[2] V. Tharavath, D. Gupta and S. Gunasekar, "Factors Influencing Choice of Cuisines While Indian Consumers Eat Out", *International Conference on Data Management, Analytics and Innovation (ICDMAI)*, 2017. [Accessed 23 July 2019].

[3] S. Gunasekar, A. Nand and S. Sudhakar, "Impact of restaurant's features on the reviews and ratings", *International Journal of Pure and Applied Mathematics*, vol. 119, no. 1311-8080, 2018. [Accessed 23 July 2019].

[4] J. Hanaysha, "Testing the effects of food quality, price fairness, and physical environment on customer satisfaction in fast food restaurant industry", *Journal of Asian Business Strategy*, vol. 6, no. 2, pp. 31-40, 2016. Available: 10.18488/journal.1006/2016.6.2/1006.2.31.40.

[5] Z. Yanchun and Z. Wei, "Can Average Score of Online Reviews Reveal Product's Real Quality?", *2012 International Conference on Management of e-Commerce and e-Government*, no. 978-0-7695-4853-1, 2012. Available: 10.1109/ICMeCG.2012.94 [Accessed 23 July 2019].

[6] F. Cullen, "Factors Influencing Restaurant Selection in Dublin", *Journal of Foodservice Business Research*, vol. 7, no. 2, pp. 53-85, 2005. Available: 10.1300/j369v07n02_05.

[7] Z. Kedah, Y. Ismail, A. Haque and S. Ahmed, "Key Success Factors of Online Food Ordering Services: An Empirical Study", vol. 50, no. 2, 2015. [Accessed 23 July 2019].

APPENDIX

Visualization of the correlation matrix and data processing has been performed in R. SPSS has been used for analyzing the data. Related sources along with the project activity log has been published on the below link

https://github.com/punitglb/CRM_Project