

Subject Code: BUAN 6337.002 – Predictive Analytics using SAS

A Group Project Report on

ZOMATO PREDICTIVE ANALYSIS

Authored By:

Vikrant Dev Rathore(VDR180000)

Krishna Darakh(KGD180000)

Mohit Jain(MXJ170006)

Mihir Raikar(MSR180000)

Nikhil Shah(NXS180034)

Yiju Tao(YXT170030)

Table of Contents

List of Figures and Tables.....	3
ACKNOWLEDGEMENT.....	4
INTRODUCTION.....	5
Data Description.....	6
Table 1: Data Description.....	6
Data Pre-Processing & Exploratory Data Analysis.....	7
Figure 1: Location of Countries.....	7
Figure 2: Location of Cities within India.....	8
Figure 3.a: Average of Aggregate Ratings for Major Cuisines.....	9
Figure 3.b: Average Aggregate Ratings for Major Cuisines.....	9
Figure 4.a: Average of votes for Major Cuisines.....	10
Figure 4.b: Average of votes for Major Cuisines.....	10
Figure 5: Average Aggregate Ratings according to number of cuisines offered by the restaurant.....	11
Figure 6: Sum of Votes for Each Rating.....	12
Figure 7: Average Price For Each Major Cuisine.....	12
Model and Analysis.....	13
.....	15
Figure 8: Linear Regression.....	15
Figure 9: LASSO.....	17
Figure 10: Forward Selection.....	24
Figure 11: Backward Selection.....	29
Figure 11: Stepwise Selection.....	33
Figure 12: Logit Model.....	35
Figure 13: Probit Model.....	36
Findings and Managerial Implications.....	37
Table 2: Results for linear models for Votes.....	37
Table 3: Results for classification of ratings.....	37
Conclusion.....	38
References.....	39
Appendix.....	40
SAS Codes.....	40

List of Figures and Tables

Table 1: Data Description	6
Figure 1: Location of Countries.....	7
Figure 2: Location of Cities within India	8
Figure 3.a: Average of Aggregate Ratings for Major Cuisines	9
Figure 3.b: Average Aggregate Ratings for Major Cuisines	9
Figure 4.a: Average of votes for Major Cuisines	10
Figure 4.b: Average of votes for Major Cuisines	10
Figure 5: Average Aggregate Ratings according to number of cuisines offered by the restaurant.....	11
Figure 6: Sum of Votes for Each Rating	12
Figure 7: Average Price For Each Major Cuisine.....	12
.....	15
Figure 8: Linear Regression	15
Figure 9: LASSO	17
Figure 10: Forward Selection	24
Figure 11: Backward Selection.....	29
Figure 11: Stepwise Selection	33
Figure 12: Logit Model	35
Figure 13: Probit Model	36
Table 2: Results for linear models for Votes	37
Table 3: Results for classification of ratings	37

ACKNOWLEDGEMENT

It has been a great opportunity to gain a lot of experience in real time projects, followed by the knowledge of how to design and analyse them. For that we would like to thank our professor, **Dr Shervin Tehrani**, who made it possible. We would like to acknowledge him for his efforts in providing us with useful information and for making the path clear.

INTRODUCTION

Background:

Zomato is an Indian restaurant search and discovery service founded in 2008 by Deepinder Goyal and Pankaj Chaddah. It currently operates in 24 countries. It provides information and reviews of restaurants, including images of menus where the restaurant does not have its own website and also online delivery.

In 2011, Zomato launched in Bengaluru, Pune, Chennai, Hyderabad and other cities. In September 2012, Zomato expanded overseas to the UAE, Sri Lanka, Qatar, UK, the Philippines, and South Africa. In April 2014, Zomato launched its services in Portugal, followed by launches in Canada, Lebanon and Ireland the same year. The acquisition of Seattle-based food portal Urbanspoon marked the firm's entry into the United States, Canada and Australia, and brought it into direct competition with Yelp, Zagat and OpenTable.

Objective:

The objective of our project or the research question which we were considering are as follows:

- **What factors drive the customers to vote for the restaurants on Zomato?**
- **How people rate a restaurant based on certain parameters?**

As we can see that Zomato takes input from the customer to rate the restaurants which can be further used by other customers. So, we are using the inputs given by the data to find the factors which affect how the customers rate a restaurant.

Data Description

The data which was given to us can be summarised as follows:

- Initially we had 9,551 Restaurants and 21 attributes.
- The total number of countries were 15 in the dataset.
- 8,649 restaurants belong to India and 899 restaurants from other 14 countries of the globe
- Hence, we chose India for our major analysis of the restaurants.

Data Fields:

Variable Name	Description
Restaurant Id	Unique identifier for each restaurant in dataset
Restaurant Name	Restaurant names
Country Code	Country code in which the restaurant is situated
City	City names in which the restaurant is situated
Address	Address of the restaurant
Locality	Locality in which the restaurant is situated
Longitude	Latitude of the restaurant is situated.
Cuisines	Different types of cuisines provided by the restaurant
Average Cost for two	Average cost for two people visiting the restaurant.
Currency	Currency in which the average cost is mentioned
Has Table booking	Binary variable for table booking.
Has Online delivery	Binary variable if the restaurant has online delivery or not.
Price range	Expected price range for two people in the restaurant
Aggregate Rating	Average rating of the restaurant
Votes	Number of votes received by the votes

Table 1: Data Description

Data Pre-Processing & Exploratory Data Analysis

- The Dataset had 9551 records, out of which there were 9 missing values in the Average Cost for Two columns and 1093 restaurants that received no votes, these rows were dropped before we began our exploratory analysis.
- The country codes were changed to the country name using the country codes file that was provided along with the dataset. The original Dataset had 75 unique cuisines which were classified into 15 primary cuisines.
- Since, India had the maximum number of restaurants (8649), it was selected for analysis and grouped into 'New Delhi' and 'Others'.
- New Delhi had 5,743 restaurants while all other remaining were in other cities of India.
- New Variables like Number of Cuisines offered by the restaurant and Major Cuisines were synthesized.



Figure 1: Location of Countries

- From Figure 1, we observe that the dataset had 15 countries, out of which New Zealand had no data. Since India had the greatest number of restaurants (8649) while all other countries combined had a 899 restaurants. As a result, India was selected for further analysis.

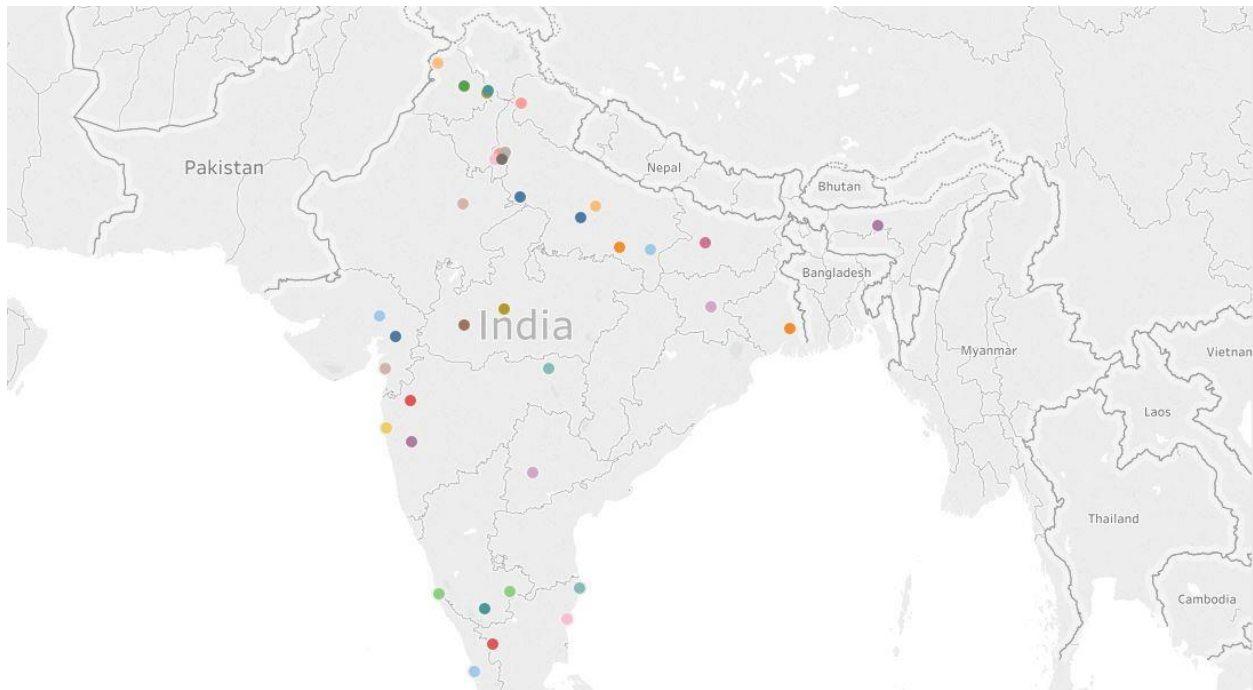


Figure 2: Location of Cities within India

- Figure 2 shows us the different location of cities within India. A total of 43 distinct cities in India were present within the dataset, however since 5,473 Restaurants were present in the capital New Delhi while the rest 3,170 were located in the other 42 cities of India, so were grouped the cities column into 'New Delhi' and 'Others'.

Ratings by Cuisine



Major Cus. Color shows details about Major Cus. Size shows average of Aggregate.rating. The marks are labeled by Major Cus.

Figure 3.a: Average of Aggregate Ratings for Major Cuisines

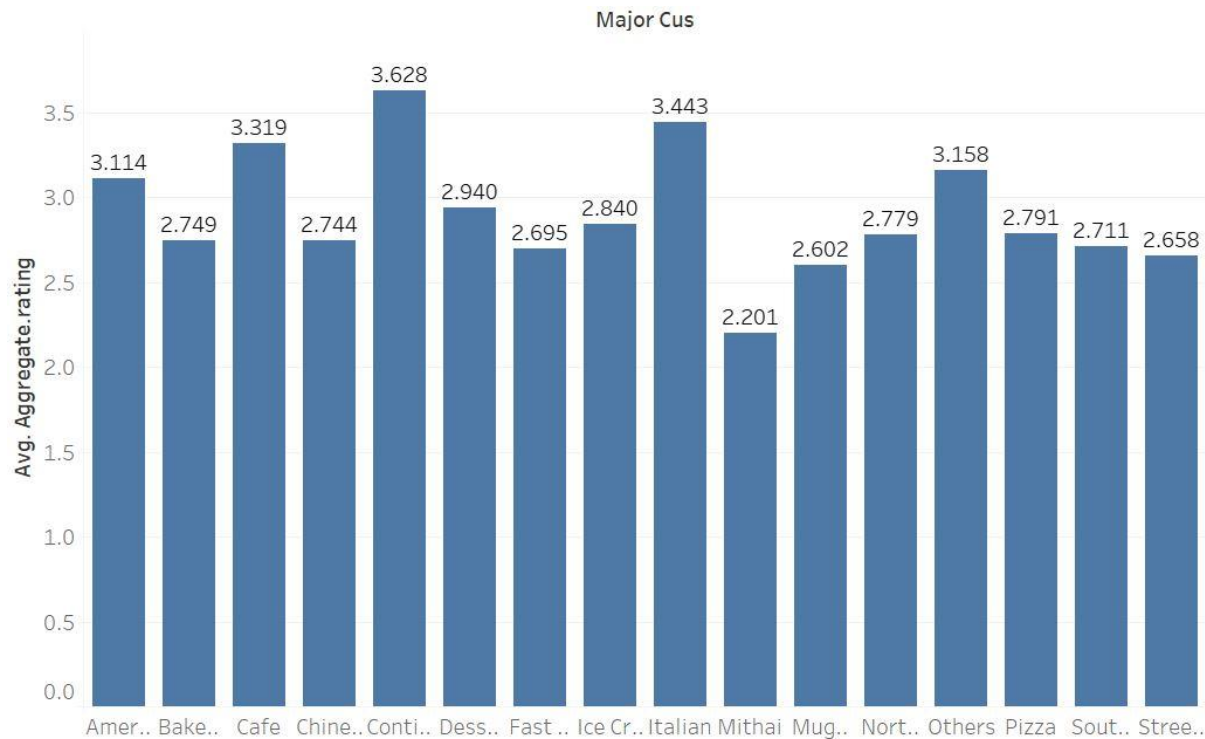


Figure 3.b: Average Aggregate Ratings for Major Cuisines

- Figure 3.a shows the average of Aggregate ratings for Major Cuisines. We see that American, Continental and Italian have higher aggregate ratings in comparison to the other cuisines which is a little surprising considering that most of the restaurants are situated in New Delhi from where the North Indian cuisine originates.
- From Figure 3.b we see that the average Aggregate ratings for Continental is 3.628 followed by Italian (3.443) and American (3.114). North Indian, the traditional food of New Delhi where the maximum restaurants are located has an average aggregate rating of 2.779.

Votes by Cuisine

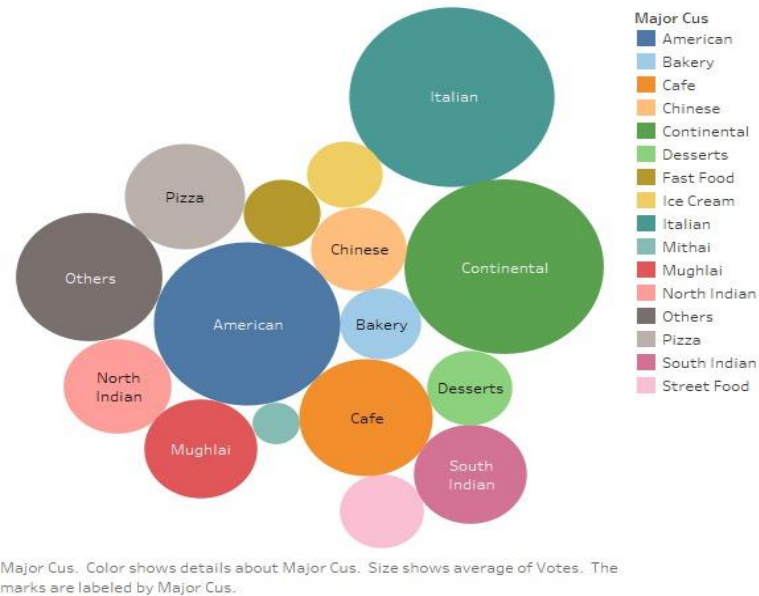


Figure 4.a: Average of votes for Major Cuisines

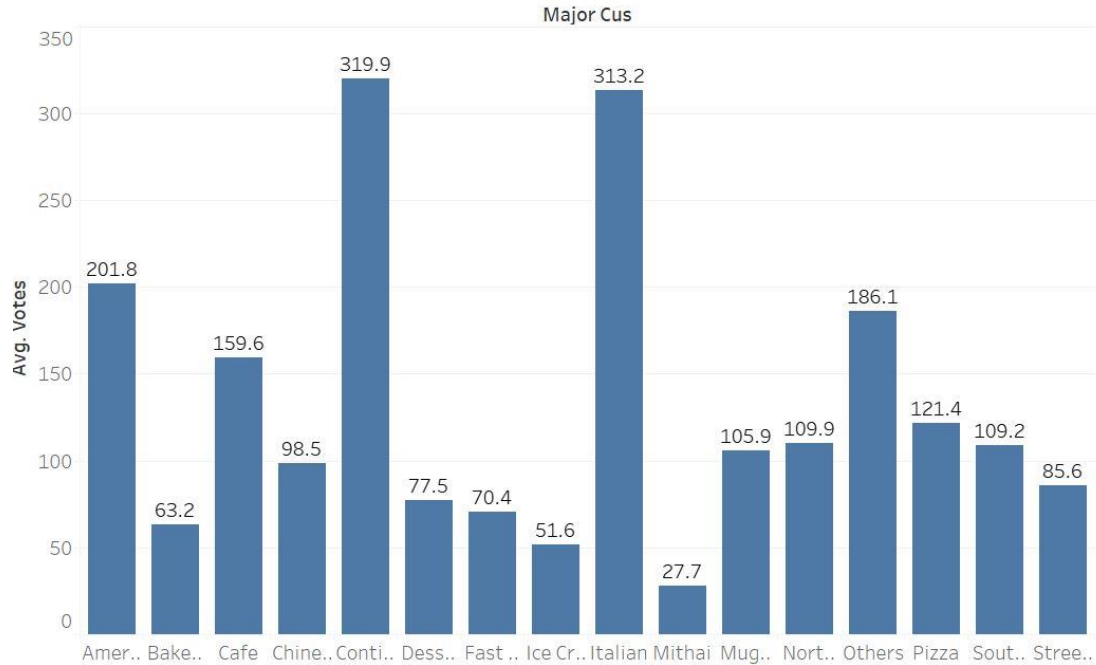


Figure 4.b: Average of votes for Major Cuisines

- Figure 4.a shows the average of Votes for Major Cuisines. We see that American, Continental and Italian have higher average votes in comparison to the other cuisines which is a little surprising considering that most of the restaurants are situated in New Delhi from where the North Indian cuisine originates.
- From Figure 4.b we see that the average votes for Continental is 319.9 followed by Italian (313.2) and American (201.8). North Indian, the traditional food of New Delhi where the maximum restaurants are located has an average vote of 109.9.

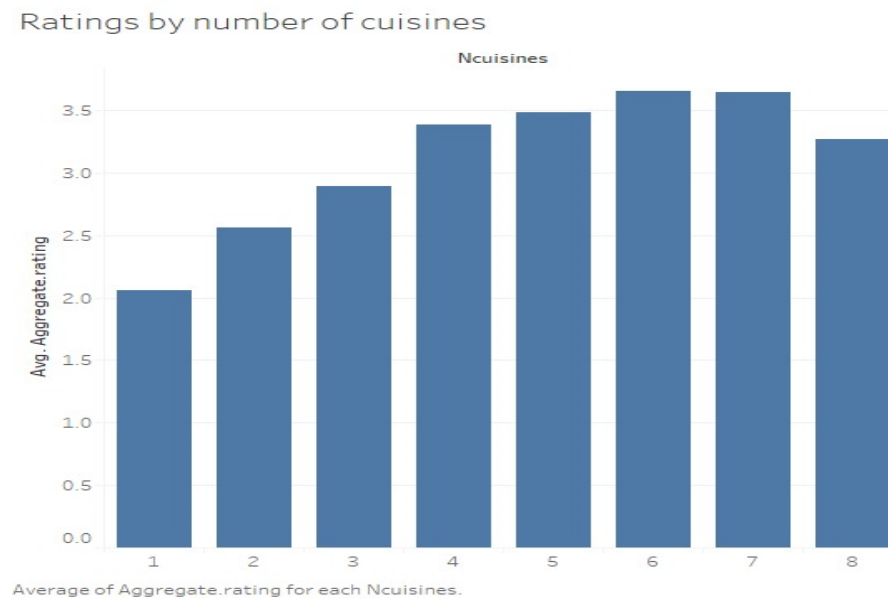


Figure 5: Average Aggregate Ratings according to number of cuisines offered by the restaurant.

- From Figure 5, we observe that the average aggregate ratings increase as the number of cuisines offered by the restaurant increase, however after 6 cuisines, the average aggregate ratings experience a drop.

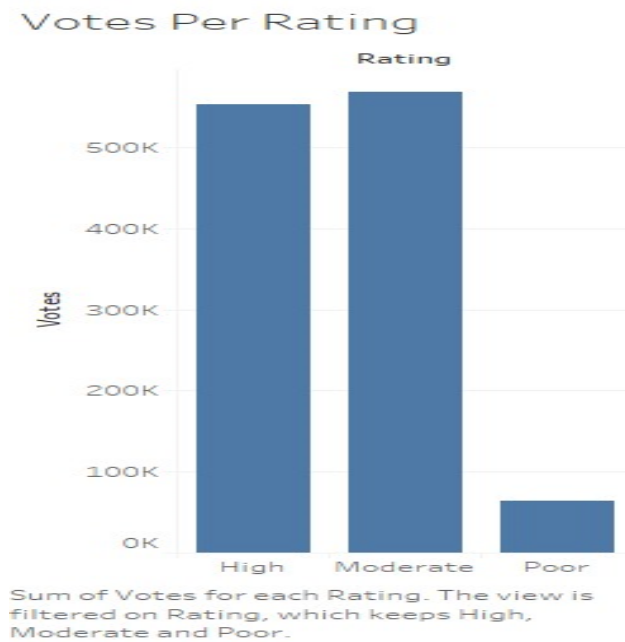
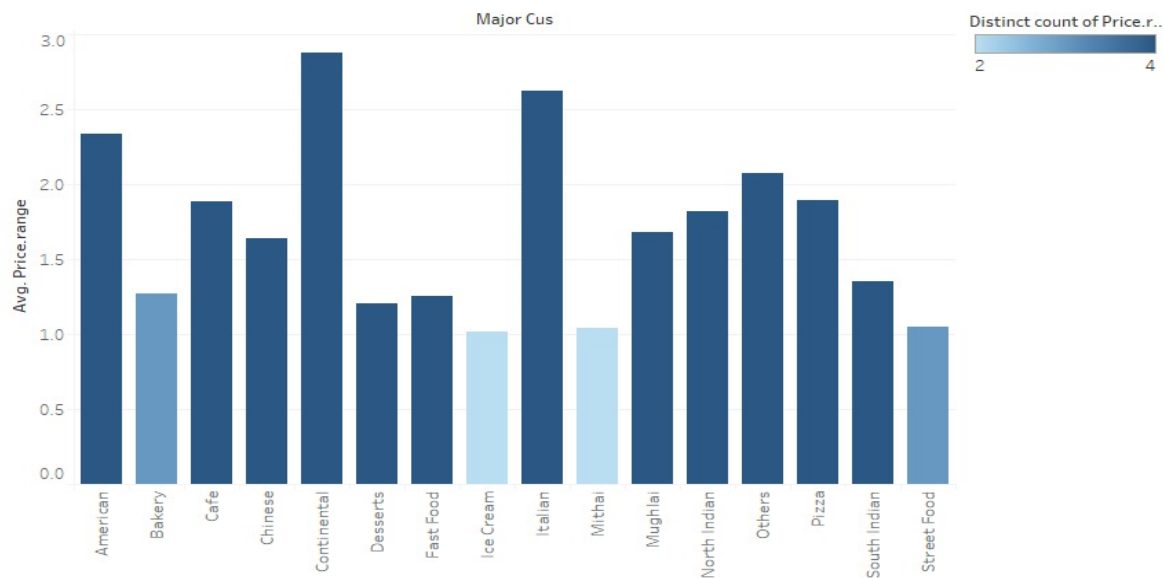


Figure 6: Sum of Votes for Each Rating

- From figure 6, we observe that High and Moderate categories have the most votes.

Average Price for 2



Average of Price.range for each Major Cus. Color shows distinct count of Price.range.

Figure 7: Average Price For Each Major Cuisine

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

- Figure 7 shows that the continental has the highest average price range followed by Italian and American. The colour coding shows the distinct count of the price range with the lightest being 2 and darkest being 4.

Model and Analysis

- We used Linear Regression, LASSO, Forward, Backward and stepwise selection for votes. For finding out how people rate a restaurant based on certain parameters, we used logit and probit model in SAS

Linear Regression:

The GLM Procedure		
Class Level Information		
Class	Levels	Values
MajorCus	16	American Bakery Cafe Chinese Continental Desserts Fast Food Ice Cream Italian Mithai Mughlai North Indian Others Pizza Street Food South Indian
City	2	New Delh Others
Has_Online_delivery	2	Yes No
Has_Table_booking	2	Yes No
Price_range	4	2 3 4 1

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-31.8909646	B 30.2848184	-1.05	0.2924
Price_range 2	84.8017998	B 30.7082289	2.76	0.0058
Price_range 3	467.0930047	B 101.9364496	4.58	<.0001
Price_range 4	240.6158144	B 219.0375843	1.10	0.2720
Price_range 1	0.0000000	B .	.	.
MajorCus American	-29.8175059	B 131.8689081	-0.23	0.8211
MajorCus Bakery	-14.7624795	B 34.2786299	-0.43	0.6667
MajorCus Cafe	10.8668961	B 36.1489382	0.30	0.7637
MajorCus Chinese	-14.6584027	B 34.8809085	-0.42	0.6743
MajorCus Continental	-22.9393130	B 68.2709752	-0.34	0.7369
MajorCus Desserts	38.7421217	B 49.6972769	0.78	0.4357
MajorCus Fast Food	14.3077491	B 34.3481825	0.42	0.6770
MajorCus Ice Cream	-21.8998503	B 44.8143073	-0.49	0.6251
MajorCus Italian	103.3281509	B 76.2321372	1.36	0.1753
MajorCus Mithai	-25.7370900	B 44.5025644	-0.58	0.5631
MajorCus Mughlai	30.1853464	B 56.3518354	0.54	0.5922
MajorCus North Indian	-20.1010493	B 30.9681489	-0.65	0.5163
MajorCus Others	10.4523476	B 33.5077813	0.31	0.7551
MajorCus Pizza	-10.5869522	B 58.3383736	-0.18	0.8560
MajorCus Street Food	-30.7282721	B 47.8577811	-0.64	0.5208
MajorCus South Indian	0.0000000	B .	.	.
City New Delh	14.3725817	B 31.5973537	0.45	0.6492
City Others	0.0000000	B .	.	.
Has_Online_delivery Yes	61.5362765	B 5.8462725	10.53	<.0001
Has_Online_delivery No	0.0000000	B .	.	.
Has_Table_booking Yes	29.7382557	B 9.7221175	3.06	0.0022
Has_Table_booking No	0.0000000	B .	.	.
ncuisines	62.1643127	B 8.0653140	7.71	<.0001
ncuisines_log	-81.0283076	B 17.9447150	-4.52	<.0001
MajorCus*Price_range American 2	-74.1636831	B 131.6749452	-0.56	0.5733
MajorCus*Price_range American 3	-245.5366443	B 168.5513107	-1.46	0.1452
MajorCus*Price_range American 4	287.4132026	B 260.2512606	1.10	0.2695
MajorCus*Price_range American 1	0.0000000	B .	.	.
MajorCus*Price_range Bakery 2	-42.5101803	B 38.4336089	-1.11	0.2687
MajorCus*Price_range Bakery 3	-428.8489007	B 119.0330236	-3.60	0.0003
MajorCus*Price_range Bakery 1	0.0000000	B .	.	.
MajorCus*Price_range Cafe 2	2.0681446	B 37.6946508	0.05	0.9562
MajorCus*Price_range Cafe 3	-272.5184685	B 106.1852194	-2.57	0.0103
MajorCus*Price_range Cafe 4	-159.1174760	B 231.5063592	-0.69	0.4919
MajorCus*Price_range Cafe 1	0.0000000	B .	.	.
MajorCus*Price_range Chinese 2	-57.9747022	B 36.1183694	-1.61	0.1085
MajorCus*Price_range Chinese 3	-251.4000809	B 104.8203691	-2.40	0.0165
MajorCus*Price_range Chinese 4	-133.2889062	B 223.2014744	-0.60	0.5504
MajorCus*Price_range Chinese 1	0.0000000	B .	.	.

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-31.8909646	B 30.2848184	-1.05	0.2924
Price_range 2	84.8017998	B 30.7082289	2.76	0.0058
Price_range 3	467.0930047	B 101.9364496	4.58	<.0001
Price_range 4	240.6158144	B 219.0375843	1.10	0.2720
Price_range 1	0.0000000	B -	-	-
MajorCus American	-29.8175059	B 131.8689081	-0.23	0.8211
MajorCus Bakery	-14.7624795	B 34.2786299	-0.43	0.6667
MajorCus Cafe	10.8668961	B 36.1489382	0.30	0.7637
MajorCus Chinese	-14.6584027	B 34.8809085	-0.42	0.6743
MajorCus Continental	-22.9393130	B 68.2709752	-0.34	0.7369
MajorCus Desserts	38.7421217	B 49.6972769	0.78	0.4357
MajorCus Fast Food	14.3077491	B 34.3481825	0.42	0.6770
MajorCus Ice Cream	-21.8998503	B 44.8143073	-0.49	0.6251
MajorCus Italian	103.3281509	B 76.2321372	1.36	0.1753
MajorCus Mithai	-25.7370900	B 44.5025644	-0.58	0.5631
MajorCus Mughlai	30.1853464	B 56.3518354	0.54	0.5922
MajorCus North Indian	-20.1010493	B 30.9681489	-0.65	0.5163
MajorCus Others	10.4523476	B 33.5077813	0.31	0.7551
MajorCus Pizza	-10.5869522	B 58.3383736	-0.18	0.8560
MajorCus Street Food	-30.7282721	B 47.8577811	-0.64	0.5208
MajorCus South Indian	0.0000000	B -	-	-
City New Delh	14.3725817	B 31.5973537	0.45	0.6492
City Others	0.0000000	B -	-	-
Has_Online_delivery Yes	61.5362765	B 5.8462725	10.53	<.0001
Has_Online_delivery No	0.0000000	B -	-	-
Has_Table_booking Yes	29.7382557	B 9.7221175	3.06	0.0022
Has_Table_booking No	0.0000000	B -	-	-
ncuisines	62.1643127	8.0653140	7.71	<.0001
ncuisines_log	-81.0283076	17.9447150	-4.52	<.0001
MajorCus*Price_range American 2	-74.1636831	B 131.6749452	-0.56	0.5733
MajorCus*Price_range American 3	-245.5366443	B 168.5513107	-1.46	0.1452
MajorCus*Price_range American 4	287.4132026	B 260.2512606	1.10	0.2695
MajorCus*Price_range American 1	0.0000000	B -	-	-
MajorCus*Price_range Bakery 2	-42.5101803	B 38.4336089	-1.11	0.2687
MajorCus*Price_range Bakery 3	-428.8489007	B 119.0390236	-3.60	0.0003
MajorCus*Price_range Bakery 1	0.0000000	B -	-	-
MajorCus*Price_range Cafe 2	2.0681446	B 37.6946508	0.05	0.9562
MajorCus*Price_range Cafe 3	-272.5184685	B 106.1852194	-2.57	0.0103
MajorCus*Price_range Cafe 4	-159.1174760	B 231.5063592	-0.69	0.4919
MajorCus*Price_range Cafe 1	0.0000000	B -	-	-
MajorCus*Price_range Chinese 2	-57.9747022	B 36.1183694	-1.61	0.1085
MajorCus*Price_range Chinese 3	-251.4008089	B 104.8203691	-2.40	0.0165
MajorCus*Price_range Chinese 4	-133.2889062	B 223.2014744	-0.60	0.5504
MajorCus*Price_range Chinese 1	0.0000000	B -	-	-

MajorCus*City American New Delh	-9.8243682	B 50.4831607	-0.19	0.8457
MajorCus*City American Others	0.0000000	B -	-	-
MajorCus*City Bakery New Delh	-1.4283496	B 37.5014333	-0.04	0.9696
MajorCus*City Bakery Others	0.0000000	B -	-	-
MajorCus*City Cafe New Delh	-17.5106656	B 37.0332509	-0.47	0.6363
MajorCus*City Cafe Others	0.0000000	B -	-	-
MajorCus*City Chinese New Delh	-18.5603857	B 36.6504119	-0.51	0.6126
MajorCus*City Chinese Others	0.0000000	B -	-	-
MajorCus*City Continental New Delh	-89.6803026	B 43.3052958	-2.07	0.0384
MajorCus*City Continental Others	0.0000000	B -	-	-
MajorCus*City Desserts New Delh	-61.8169396	B 54.6548949	-1.13	0.2581
MajorCus*City Desserts Others	0.0000000	B -	-	-
MajorCus*City Fast Food New Delh	-32.5494684	B 37.2703233	-0.87	0.3825
MajorCus*City Fast Food Others	0.0000000	B -	-	-
MajorCus*City Ice Cream New Delh	3.5185075	B 50.7443788	0.07	0.9447
MajorCus*City Ice Cream Others	0.0000000	B -	-	-
MajorCus*City Italian New Delh	-7.4761634	B 47.1507095	-0.16	0.8740
MajorCus*City Italian Others	0.0000000	B -	-	-
MajorCus*City Mithai New Delh	-8.8009357	B 48.7529881	-0.18	0.8567
MajorCus*City Mithai Others	0.0000000	B -	-	-
MajorCus*City Mughlai New Delh	-38.5949928	B 52.6380415	-0.73	0.4635
MajorCus*City Mughlai Others	0.0000000	B -	-	-
MajorCus*City North Indian New Delh	-19.1574304	B 32.8723615	-0.58	0.5801
MajorCus*City North Indian Others	0.0000000	B -	-	-
MajorCus*City Others New Delh	-44.8868105	B 35.0546319	-1.28	0.2004
MajorCus*City Others Others	0.0000000	B -	-	-
MajorCus*City Pizza New Delh	-38.2789994	B 46.2889002	-0.83	0.4083
MajorCus*City Pizza Others	0.0000000	B -	-	-
MajorCus*City Street Food New Delh	47.5395168	B 51.7920335	0.92	0.3587
MajorCus*City Street Food Others	0.0000000	B -	-	-
MajorCus*City South Indian New Delh	0.0000000	B -	-	-
MajorCus*City South Indian Others	0.0000000	B -	-	-

R-Square	Coeff Var	Root MSE	Votes Mean
0.234749	177.4773	217.9460	122.8022

Figure 8: Linear Regression

- From Figure 8, we observe that price has an important effect on votes, as the price increases, the votes also increase. Votes increase by the biggest margin for price range 3.
- The votes decrease when the cuisine is North Indian. The interaction between Street Food and Price range 2 have a positive impact on votes.
- If online delivery is available, then voting is bound to increase.
- We have taken log of cuisines to show that very large number of cuisines offered by a restaurant can lead to a fall in votes. Votes with the number of cuisines offered increase up to a certain extent and then decrease.
- If table booking is available, then the votes increase.
- We also observe if that if the city is New Delhi, there is a decline in votes.
- The R-square is very low(0.23), we will use LASSO, forward , backward and best subset selection in order to better explain the variance in the model.

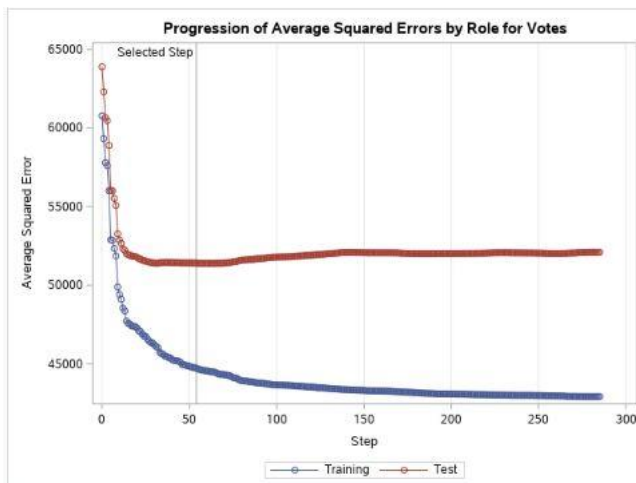
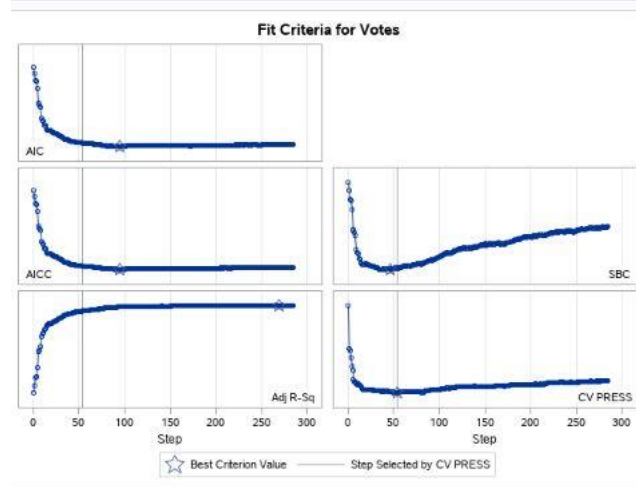
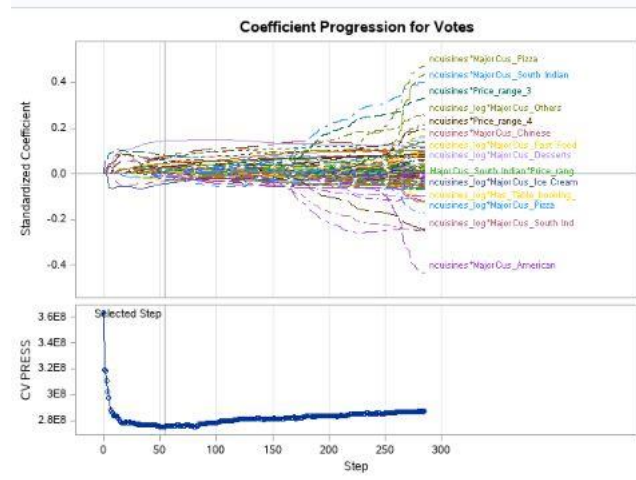
LASSO:

The GLMSELECT Procedure	
Data Set	WORK.ZOMATO_TRAINING
Test Data Set	WORK.ZOMATO_TEST
Dependent Variable	Votes
Selection Method	LASSO
Stop Criterion	None
Choose Criterion	Cross Validation
Cross Validation Method	Split
Cross Validation Fold	5
Effect Hierarchy Enforced	None

Observation Profile for Analysis Data	
Number of Observations Read	5981
Number of Observations Used	5977
Number of Observations Used for Training	5977

Observation Profile for Test Data	
Number of Observations Read	1495
Number of Observations Used	1491

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS



BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

Root MSE	212.35091
Dependent Mean	122.69834
R-Square	0.2641
Adj R-Sq	0.2582
AIC	70080
AICC	70081
SBC	64429
ASE (Train)	44723
ASE (Test)	51405
CV PRESS	275021172

Parameter Estimates		
Parameter	DF	Estimate
Intercept	1	86.641759
Price_range_1	1	-21.796978
Price_range_3	1	5.993878
MajorCus_American*Price_range_4	1	188.685505
MajorCus_Bakery*Price_range_3	1	-85.801654
MajorCus_Cafe*Price_range_2	1	10.453361
MajorCus_Cafe*Price_range_3	1	-2.712683
MajorCus_Continental*Price_range_3	1	25.068307
MajorCus_Continental*Price_range_4	1	65.997999
MajorCus_Desserts*Price_range_3	1	-37.190825
MajorCus_Italian*Price_range_3	1	80.608659
MajorCus_Italian*Price_range_4	1	-18.532092
MajorCus_North Indian*Price_range_1	1	-5.068000
MajorCus_North Indian*Price_range_2	1	-19.620385
MajorCus_North Indian*Price_range_4	1	2.228266
MajorCus_Others*Price_range_2	1	26.018843
MajorCus_Pizza*Price_range_3	1	178.182507
MajorCus_South Indian*Price_range_3	1	182.418841
MajorCus_Street Food*Price_range_2	1	75.003782
City_Others*Price_range_3	1	5.429686
MajorCus_Continental*City_Others	1	20.452054
MajorCus_Others*City_Others	1	6.482069
MajorCus_Street Food*City_New Delh	1	3.274310

Has_Online_delivery_No	1	-6.787561
Has_Online_delivery_No*Price_range_1	1	-16.438438
Has_Online_delivery_No*Price_range_3	1	60.398286
Has_Online_delivery_Yes*Price_range_4	1	236.479368
MajorCus_Cafe*Has_Online_delivery_Yes	1	27.183810
MajorCus_Chinese*Has_Online_delivery_No	1	-1.506727
MajorCus_Continental*Has_Online_delivery_Yes	1	25.867679
MajorCus_Italian*Has_Online_delivery_No	1	40.363956
MajorCus_Pizza*Has_Online_delivery_Yes	1	25.273445
MajorCus_South Indian*Has_Online_delivery_Yes	1	42.138325
MajorCus_Street Food*Has_Online_delivery_Yes	1	24.979651
Has_Table_booking_No*Price_range_3	1	69.353512
Has_Table_booking_No*Price_range_4	1	92.413816
MajorCus_American*Has_Table_booking_Yes	1	72.805654
MajorCus_Italian*Has_Table_booking_Yes	1	87.338286
MajorCus_Others*Has_Table_booking_Yes	1	46.014390
City_Others*Has_Table_booking_Yes	1	50.467180
Has_Online_delivery_Yes*Has_Table_booking_Yes	1	155.436968
ncuisines*Price_range_1	1	-5.424482
ncuisines*Price_range_3	1	18.679749
ncuisines*Price_range_4	1	25.475098
ncuisines*Has_Online_delivery_Yes	1	5.005358
ncuisines_log*MajorCus_Cafe	1	31.118553
ncuisines_log*MajorCus_Continental	1	33.243635
ncuisines_log*MajorCus_North Indian	1	-5.914380
ncuisines*ncuisines_	1	8.836236

Figure 9: LASSO

- From the model, we can see that the LASSO has done the selection steps 285 times to find the best effects on votes.
- The R -square value has improved to 0.26 in comparison to the Linear regression Model.
- The AIC is 70080
- In LASSO, people tend to vote for high priced restaurant, the result is same as linear regression but for North Indian, Italian and dessert, high priced restaurants tend to get less votes than lower priced range restaurant.
- If online delivery is available, then voting is bound to increase.
- The R-square is still very low(0.26), we will use forward , backward and best subset selection in order to better explain the variance in the model.

Forward Selection:

The GLMSELECT Procedure

Data Set	WORK.ZOMATO_TRAINING
Test Data Set	WORK.ZOMATO_TEST
Dependent Variable	Votes
Selection Method	Forward
Select Criterion	C(p)
Stop Criterion	C(p)
Effect Hierarchy Enforced	Single

Observation Profile for Analysis Data

Number of Observations Read	5081
Number of Observations Used	5077
Number of Observations Used for Training	5077

Observation Profile for Test Data

Number of Observations Read	1495
Number of Observations Used	1491

Class Level Information

Class	Levels	Values
MajorCus	16	American Bakery Cafe Chinese Continental Desserts Fast Food Ice Cream Italian Mithai Mughlai North Indian Others Pizza South Indian Street Food
City	2 *	New Delh Others
Has_Online_delivery	2 *	No Yes
Has_Table_booking	2 *	No Yes
Price_range	4 *	1 2 3 4
* Associated Parameters Split		

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

The GLMSELECT Procedure

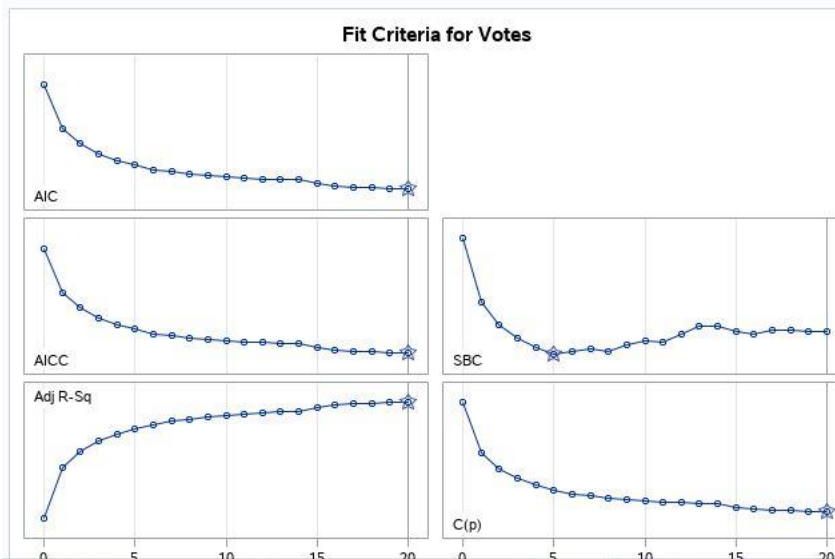
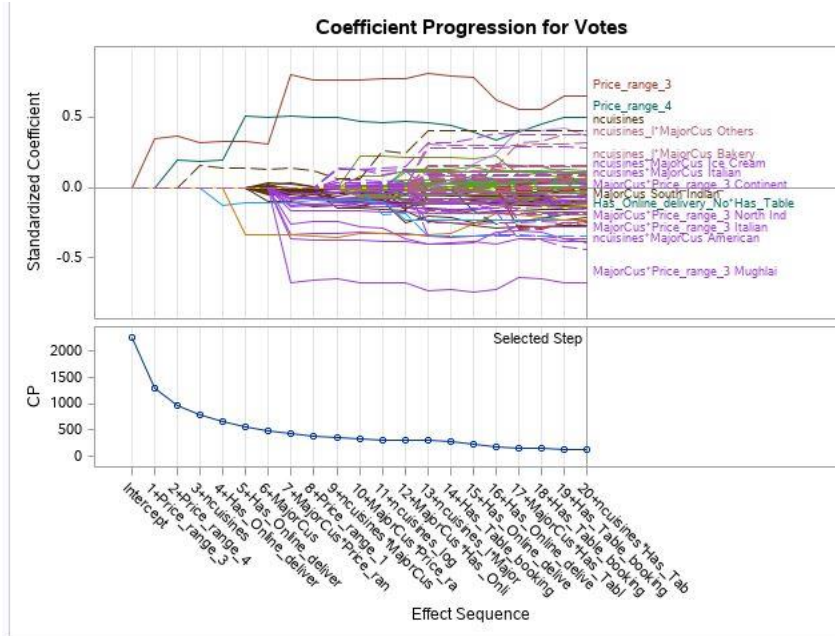
Forward Selection Summary						
Step	Effect Entered	Number Effects In	Number Params In	CP	ASE	Test ASE
0	Intercept	1	1	2268.1663	60777.4876	63888.5477
1	Price_range_3	2	2	1290.0690	53551.1596	56036.5642
2	Price_range_4	3	3	972.7389	51196.7111	53753.5413
3	ncuisines	4	4	784.2051	49791.8915	52076.0479
4	Has_Online_delivery_No	5	5	654.3176	48819.4746	51169.0496
5	Has_Online_delivery_No*Price_range_4	6	6	557.4723	48090.8813	51032.4411
6	MajorCus	7	21	477.1219	47277.0589	50700.3594
7	MajorCus*Price_range_3	8	33	434.9059	46788.8507	51109.1370
8	Price_range_1	9	34	395.3641	46482.5524	50884.2013
9	ncuisines*MajorCus	10	49	365.2665	46039.3749	50644.4075
10	MajorCus*Price_range_4	11	59	338.8403	45697.1449	51695.0727
11	ncuisines_log	12	60	319.8738	45542.5574	51690.0032
12	MajorCus*Has_Online_delivery_No	13	75	304.0348	45204.5829	51698.5739
13	ncuisines_I*MajorCus	14	90	296.0837	44624.7682	51464.0778
14	Has_Table_booking_Yes	15	91	290.3098	44807.4485	51672.3542
15	Has_Online_delivery_No*Has_Table_booking_Yes	16	92	222.7683	44354.9356	51979.6712
16	Has_Online_delivery_No*Price_range_3	17	93	181.9226	44038.8097	52770.4616
17	MajorCus*Has_Table_booking_Yes	18	103	161.0346	43737.3389	52803.6626
18	Has_Table_booking_Yes*Price_range_4	19	104	154.3013	43672.9481	52671.4303
19	Has_Table_booking_Yes*Price_range_3	20	105	136.8499	43529.5313	52285.3652
20	ncuisines*Has_Table_booking_Yes	21	106	134.4076*	43496.7780	52193.2116

* Optimal Value of Criterion

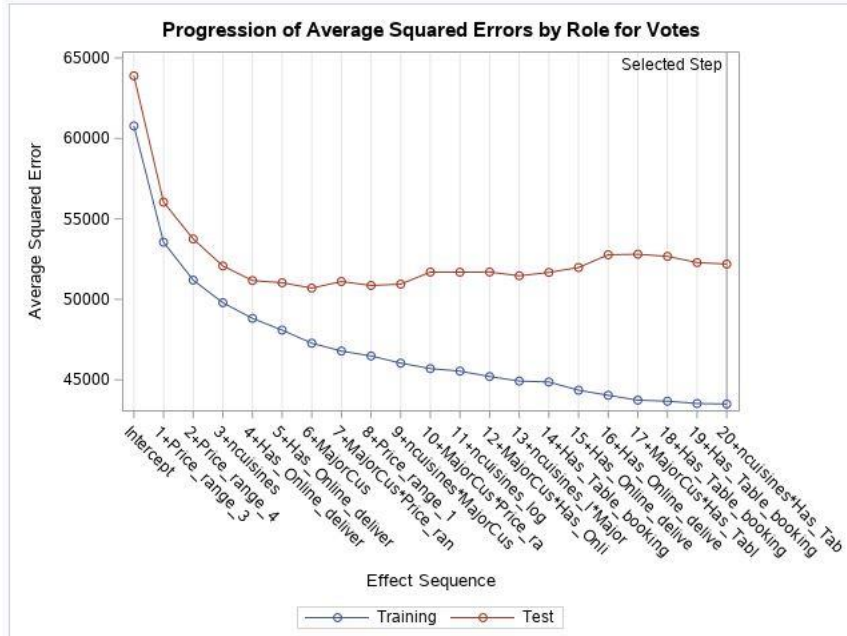
Selection stopped at a local minimum of the C(p) criterion.

Stop Details				
Candidate For	Effect	Candidate C(p)		Compare C(p)
Entry	City_Others	134.9338	>	134.4076

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS



BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS



BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > t
Intercept	1	148.908561	86.369471	1.72	0.0847
Price_range_1	1	-51.469530	7.455290	-6.90	<.0001
Price_range_3	1	462.667554	120.787976	3.83	0.0001
Price_range_4	1	575.155524	137.409975	4.19	<.0001
MajorCus American	1	-34.941508	120.744719	-0.29	0.7723
MajorCus Bakery	1	-80.971982	98.613747	-0.82	0.5364
MajorCus Cafe	1	-94.942178	94.711492	-1.00	0.3162
MajorCus Chinese	1	-123.414327	94.071024	-1.31	0.1896
MajorCus Continental	1	-123.537802	99.525899	-1.24	0.2146
MajorCus Desserts	1	8.372418	130.242222	0.06	0.9487
MajorCus Fast Food	1	-0.220602	100.196128	-0.00	0.9982
MajorCus Ice Cream	1	-207.563848	160.890573	-1.29	0.1971
MajorCus Italian	1	-143.034921	103.284192	-1.38	0.1661
MajorCus Mithai	1	36.769054	139.842556	0.26	0.7926
MajorCus Mughlai	1	15.325153	135.408035	0.11	0.9099
MajorCus North Indian	1	-112.959012	87.485357	-1.29	0.1967
MajorCus Others	1	-11.919088	90.507624	-0.13	0.8952
MajorCus Pizza	1	-424.058172	180.945891	-2.34	0.0191
MajorCus South Indian	1	-178.080639	115.411435	-1.54	0.1229
MajorCus Street Food	0	0	.	.	.
MajorCus*Price_range_3 American	1	-409.964851	139.883556	-2.93	0.0034
MajorCus*Price_range_3 Bakery	1	-597.949402	141.131915	-4.24	<.0001
MajorCus*Price_range_3 Cafe	1	-452.941516	122.996395	-3.68	0.0002
MajorCus*Price_range_3 Chinese	1	-310.630857	122.826685	-2.53	0.0115
MajorCus*Price_range_3 Continental	1	-145.534779	126.758772	-1.15	0.2510
MajorCus*Price_range_3 Desserts	1	-593.121265	181.211040	-3.27	0.0011
MajorCus*Price_range_3 Fast Food	1	-155.290889	154.996377	-1.00	0.3164
MajorCus*Price_range_3 Italian	1	-277.568200	127.123975	-2.18	0.0290
MajorCus*Price_range_3 Mughlai	1	-469.840636	138.487028	-3.39	0.0007
MajorCus*Price_range_3 North Indian	1	-296.326852	119.779325	-2.47	0.0134
MajorCus*Price_range_3 Others	1	-361.131460	121.458542	-2.97	0.0030
MajorCus*Price_range_3 Pizza	1	-147.683988	152.259133	-0.97	0.3321
MajorCus*Price_range_3 South Indian	0	0	.	.	.
MajorCus*Price_range_4 American	1	104.769057	167.743588	0.62	0.5323
MajorCus*Price_range_4 Cafe	1	-278.286232	158.193488	-1.76	0.0786
MajorCus*Price_range_4 Chinese	1	-71.224748	140.894873	-0.51	0.6132
MajorCus*Price_range_4 Continental	1	179.706350	140.434608	1.28	0.2007
MajorCus*Price_range_4 Desserts	1	-354.557168	250.267125	-1.42	0.1566
MajorCus*Price_range_4 Fast Food	1	18.082585	248.938203	0.07	0.9421
MajorCus*Price_range_4 Italian	1	-330.845932	146.564131	-2.26	0.0240
MajorCus*Price_range_4 Mughlai	1	-220.730609	166.343173	-1.33	0.1846
MajorCus*Price_range_4 North Indian	1	-19.507774	132.897806	-0.15	0.8833
MajorCus*Price_range_4 Others	1	-135.369527	134.477246	-1.01	0.3141
MajorCus*Price_range_4 Pizza	0	0	.	.	.
Has_Online_delivery No	1	-120.577516	49.877787	-2.42	0.0157

Root MSE	210.43314
Dependent Mean	122.89834
R-Square	0.2843
Adj R-Sq	0.2715
AIC	70028
AICC	70032
BIC	64054
C(p)	134.40761
SBC	64759
ASE (Train)	43497
ASE (Test)	52193

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

Has_Online_delivery_No*Price_range_3	1	98.426136	23.107017	4.28	<.0001
Has_Online_delivery_No*Price_range_4	1	-281.894279	47.919472	-5.88	<.0001
MajorCus*Has_Online_delivery_No American	1	116.220545	70.171624	1.66	0.0977
MajorCus*Has_Online_delivery_No Bakery	1	82.862289	55.251756	1.50	0.1337
MajorCus*Has_Online_delivery_No Cafe	1	48.239802	55.234630	0.87	0.3825
MajorCus*Has_Online_delivery_No Chinese	1	73.272378	53.902550	1.36	0.1741
MajorCus*Has_Online_delivery_No Continental	1	-50.697334	63.724719	-0.80	0.4263
MajorCus*Has_Online_delivery_No Desserts	1	55.823947	71.168112	0.78	0.4328
MajorCus*Has_Online_delivery_No Fast Food	1	111.254470	54.201717	2.05	0.0402
MajorCus*Has_Online_delivery_No Ice Cream	1	80.006922	64.207867	1.25	0.2128
MajorCus*Has_Online_delivery_No Italian	1	209.635250	65.767575	3.19	0.0014
MajorCus*Has_Online_delivery_No Mithai	1	-11.111247	105.540663	-0.11	0.9162
MajorCus*Has_Online_delivery_No Mughlai	1	79.500593	66.048807	1.20	0.2288
MajorCus*Has_Online_delivery_No North Indian	1	92.088094	51.099334	1.80	0.0716
MajorCus*Has_Online_delivery_No Others	1	72.698899	53.118801	1.37	0.1712
MajorCus*Has_Online_delivery_No Pizza	1	9.245537	63.788507	0.14	0.8848
MajorCus*Has_Online_delivery_No South Indian	1	5.404988	63.576868	0.09	0.9323
MajorCus*Has_Online_delivery_No Street Food	0	0	.	.	.
Has_Table_booking_Yes	1	259.671543	110.173821	2.36	0.0185
Has_Table_booking_Yes*Price_range_3	1	-111.501094	24.542621	-4.54	<.0001
Has_Table_booking_Yes*Price_range_4	1	-164.590254	35.800736	-4.80	<.0001
MajorCus*Has_Table_booking_Yes American	1	117.995357	135.603949	0.87	0.3843
MajorCus*Has_Table_booking_Yes Cafe	1	-90.841520	115.969950	-0.78	0.4335
MajorCus*Has_Table_booking_Yes Chinese	1	-130.978115	112.198881	-1.17	0.2431
MajorCus*Has_Table_booking_Yes Continental	1	-191.872678	114.782412	-1.67	0.0947
MajorCus*Has_Table_booking_Yes Fast Food	1	-231.266082	168.207277	-1.37	0.1692
MajorCus*Has_Table_booking_Yes Italian	1	93.225705	117.199925	0.80	0.4264
MajorCus*Has_Table_booking_Yes Mughlai	1	-88.412261	133.722112	-0.66	0.5085
MajorCus*Has_Table_booking_Yes North Indian	1	-117.805400	109.230570	-1.08	0.2809
MajorCus*Has_Table_booking_Yes Others	1	0.923707	111.500381	0.01	0.9934
MajorCus*Has_Table_booking_Yes Pizza	1	-187.872143	148.761005	-1.26	0.2067
MajorCus*Has_Table_booking_Yes South Indian	0	0	.	.	.
Has_Online_delivery_No*Has_Table_booking_Yes	1	-173.150359	23.563666	-7.35	<.0001
ncuisines	1	90.073886	71.386071	1.28	0.2071
ncuisines*MajorCus American	1	-237.014465	143.722607	-1.65	0.0992
ncuisines*MajorCus Bakery	1	-86.470197	84.578137	-1.02	0.3086
ncuisines*MajorCus Cafe	1	10.883700	79.743676	0.14	0.8914
ncuisines*MajorCus Chinese	1	5.994808	80.322571	0.07	0.9405
ncuisines*MajorCus Continental	1	-31.844404	82.164091	-0.39	0.6983
ncuisines*MajorCus Desserts	1	-94.534697	114.716119	-0.82	0.4099
ncuisines*MajorCus Fast Food	1	-125.782629	87.777264	-1.43	0.1519
ncuisines*MajorCus Ice Cream	1	87.067353	156.466183	0.56	0.5779
ncuisines*MajorCus Italian	1	12.126001	87.796611	0.14	0.8902
ncuisines*MajorCus Mithai	1	-75.156883	84.941947	-0.88	0.3763
ncuisines*MajorCus Mughlai	1	-132.334413	128.983326	-1.03	0.3049
ncuisines*MajorCus North Indian	1	-20.744881	72.690245	-0.29	0.7754
ncuisines*MajorCus Others	1	-119.066103	75.972107	-1.57	0.1171
ncuisines*MajorCus Pizza	1	295.967536	178.116661	1.66	0.0986

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

ncuisines*MajorCus Pizza	1	295.967536	178.116861	1.66	0.0966
ncuisines*MajorCus South Indian	1	190.131386	101.359847	1.88	0.0607
ncuisines*MajorCus Street Food	0	0	-	-	-
ncuisines*Has_Table_booking_Yes	1	15.886124	7.545983	2.10	0.0355
ncuisines_log	1	-167.819061	144.864725	-1.16	0.2487
ncuisines_*MajorCus American	1	595.821520	367.852579	1.62	0.1053
ncuisines_*MajorCus Bakery	1	223.812126	169.745315	1.32	0.1874
ncuisines_*MajorCus Cafe	1	64.314686	163.992189	0.39	0.6049
ncuisines_*MajorCus Chinese	1	-2.888581	162.561672	-0.02	0.9858
ncuisines_*MajorCus Continental	1	186.247197	187.449901	0.99	0.3205
ncuisines_*MajorCus Desserts	1	193.340529	225.807407	0.86	0.3919
ncuisines_*MajorCus Fast Food	1	230.288982	172.417802	1.34	0.1818
ncuisines_*MajorCus Ice Cream	1	-95.588233	279.322147	-0.34	0.7322
ncuisines_*MajorCus Italian	1	-31.152932	196.665145	-0.16	0.8741
ncuisines_*MajorCus Mithai	1	150.894698	180.904213	0.83	0.4043
ncuisines_*MajorCus Mughlai	1	285.628694	244.005949	1.17	0.2418
ncuisines_*MajorCus North Indian	1	28.648057	148.321083	0.19	0.8468
ncuisines_*MajorCus Others	1	288.306559	155.547896	1.85	0.0639
ncuisines_*MajorCus Pizza	1	-275.567788	308.043414	-0.89	0.3711
ncuisines_*MajorCus South Indian	1	-376.706208	201.378628	-1.87	0.0614
ncuisines_*MajorCus Street Food	0	0	-	-	-

Figure 10: Forward Selection

- We observe that the R-square improves to 0.28 in the case of Forward Selection which is better than both linear regression and LASSO.
- The AIC is 70028 which is lower than that of LASSO, so a better fit than LASSO can be implied.
- In Forward selection also shows when the level of price increase, the possibility of getting votes increase.
- If online delivery is available, then voting is bound to increase but people tend to not give votes for price range 3 and 4 if they provide online delivery. This is different from using LASSO.
- The R-square is still very low(0.28), we will use backward and best subset selection in order to better explain the variance in the model.

Backward Selection:

The GLMSELECT Procedure		
Data Set	WORK.ZOMATO_TRAINING	
Test Data Set	WORK.ZOMATO_TEST	
Dependent Variable	Votes	
Selection Method	Backward	
Select Criterion	C(p)	
Stop Criterion	C(p)	
Effect Hierarchy Enforced	Single	

Observation Profile for Analysis Data		
Number of Observations Read	5981	
Number of Observations Used	5977	
Number of Observations Used for Training	5977	

Observation Profile for Test Data		
Number of Observations Read	1495	
Number of Observations Used	1491	

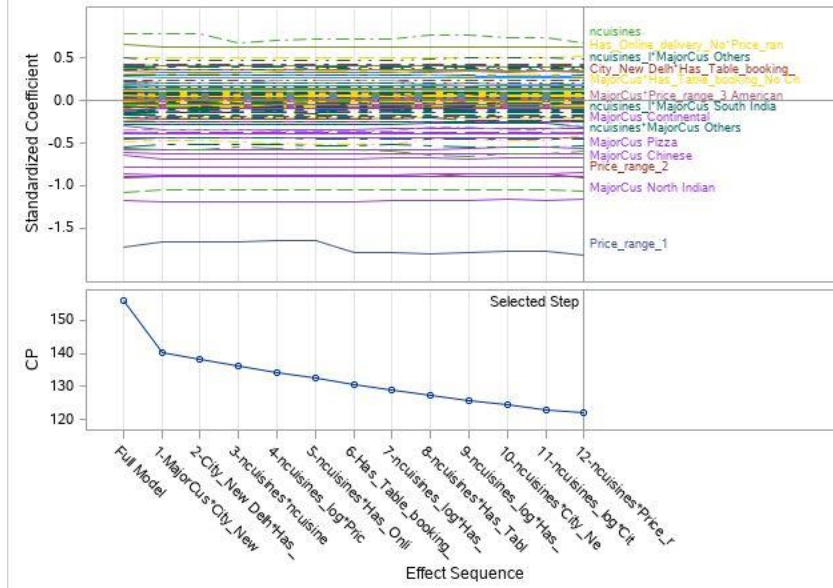
Class Level Information		
Class	Levels	Values
MajorCus	18	American Bakery Cafe Chinese Continental Desserts Fast Food Ice Cream Italian Mithai Mughlai North Indian Others Pizza South Indian Street Food
City	2 *	New Delh Others
Has_Online_delivery	2 *	No Yes
Has_Table_booking	2 *	No Yes
Price_range	4 *	1 2 3 4
* Associated Parameters Split		

Dimensions		
Number of Effects	29	
Number of Effects after Splits	83	
Number of Parameters	265	

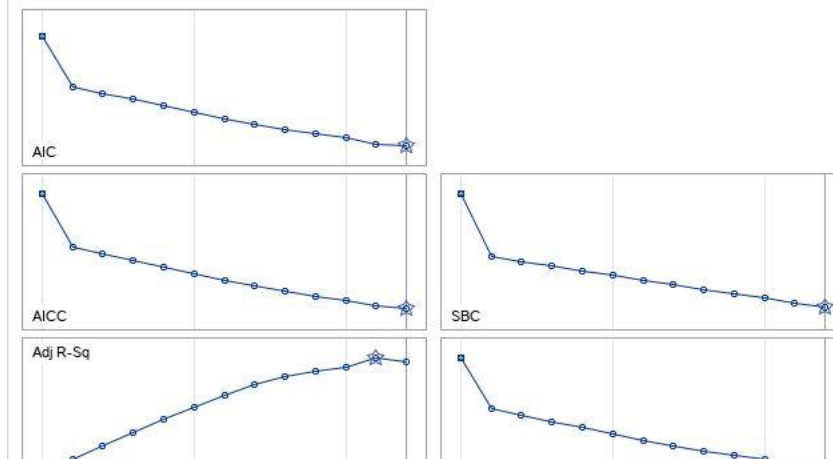
BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

The GLMSELECT Procedure						
Backward Selection Summary						
Step	Effect Removed	Number Effects In	Number Parms In	CP	A SE	Test A SE
0		83	156	156.0000	42918.6727	52096.7243
	ncuisines_jog*Has_Table_booking_Yes	82	156	156.0000	42918.6727	52096.7243
	ncuisines_jog*Has_Online_delivery_Yes	81	156	156.0000	42918.6727	52096.7243
	ncuisines_jog*City_Others	80	156	156.0000	42918.6727	52096.7243
	ncuisines_jog*Price_range_4	79	156	156.0000	42918.6727	52096.7243
	ncuisines*Has_Table_booking_Yes	78	156	156.0000	42918.6727	52096.7243
	ncuisines*Has_Online_delivery_Yes	77	156	156.0000	42918.6727	52096.7243
	ncuisines*City_Others	76	156	156.0000	42918.6727	52096.7243
	ncuisines*Price_range_4	75	156	156.0000	42918.6727	52096.7243
	Has_Online_delivery_Yes*Has_Table_booking_Yes	74	156	156.0000	42918.6727	52096.7243
	Has_Online_delivery_Yes*Has_Table_booking_No	73	156	156.0000	42918.6727	52096.7243
	Has_Online_delivery_No*Has_Table_booking_Yes	72	156	156.0000	42918.6727	52096.7243
	City_Others*Has_Table_booking_Yes	71	156	156.0000	42918.6727	52096.7243
	City_Others*Has_Table_booking_No	70	156	156.0000	42918.6727	52096.7243
	City_NewDelhi*Has_Table_booking_Yes	69	156	156.0000	42918.6727	52096.7243
	MajorCus*Has_Table_booking_Yes	68	156	156.0000	42918.6727	52096.7243
	Has_Table_booking_Yes*Price_range_4	67	156	156.0000	42918.6727	52096.7243
	Has_Table_booking_Yes*Price_range_3	66	156	156.0000	42918.6727	52096.7243
	Has_Table_booking_Yes*Price_range_2	65	156	156.0000	42918.6727	52096.7243
	Has_Table_booking_Yes*Price_range_1	64	156	156.0000	42918.6727	52096.7243
	Has_Table_booking_No*Price_range_4	63	156	156.0000	42918.6727	52096.7243
	Has_Table_booking_Yes	62	156	156.0000	42918.6727	52096.7243
	City_Others*Has_Online_delivery_Yes	61	156	156.0000	42918.6727	52096.7243
	City_Others*Has_Online_delivery_No	60	156	156.0000	42918.6727	52096.7243
	City_NewDelhi*Has_Online_delivery_Yes	59	156	156.0000	42918.6727	52096.7243
	MajorCus*Has_Online_delivery_Yes	58	156	156.0000	42918.6727	52096.7243
	Has_Online_delivery_Yes*Price_range_4	57	156	156.0000	42918.6727	52096.7243
	Has_Online_delivery_Yes*Price_range_3	56	156	156.0000	42918.6727	52096.7243
	Has_Online_delivery_Yes*Price_range_2	55	156	156.0000	42918.6727	52096.7243
	Has_Online_delivery_Yes*Price_range_1	54	156	156.0000	42918.6727	52096.7243
	Has_Online_delivery_No*Price_range_4	53	156	156.0000	42918.6727	52096.7243
	Has_Online_delivery_Yes	52	156	156.0000	42918.6727	52096.7243
	MajorCus*City_Others	51	156	156.0000	42918.6727	52096.7243
	City_Others*Price_range_4	50	156	156.0000	42918.6727	52096.7243
	City_Others*Price_range_3	49	156	156.0000	42918.6727	52096.7243
	City_Others*Price_range_2	48	156	156.0000	42918.6727	52096.7243
	City_Others*Price_range_1	47	156	156.0000	42918.6727	52096.7243
	City_NewDelhi*Price_range_4	46	156	156.0000	42918.6727	52096.7243
	City_Others	45	156	156.0000	42918.6727	52096.7243
	MajorCus*Price_range_4	44	156	156.0000	42918.6727	52096.7243
	Price_range_4	43	156	156.0000	42918.6727	52096.7243
1	MajorCus*City_NewDelhi	42	141	140.2432	43023.6888	52006.8531
2	City_NewDelhi*Has_Online_delivery_No	41	140	138.2524	43023.7570	52011.6061
3	ncuisines*ncuisines	40	139	136.2741	43023.9166	52017.4190
4	ncuisines_jog*Price_range_3	39	138	134.2964	43024.0615	52030.5943
5	ncuisines*Has_Online_delivery_No	38	137	132.3767	43024.6735	52038.2388
6	Has_Table_booking_No*Price_range_1	37	136	130.4909	43025.5149	52044.4257
7	ncuisines_jog*Has_Online_delivery_No	36	135	128.7845	43027.6798	52000.2609
8	ncuisines*Has_Table_booking_No	35	134	127.1451	43030.3360	52020.0310
9	ncuisines_jog*Has_Table_booking_No	34	133	125.7667	43034.9219	52031.0073
10	ncuisines*City_NewDelhi	33	132	124.4513	43039.9692	52004.6855
11	ncuisines_jog*City_NewDelhi	32	131	122.7895	43042.3154	52027.2965
12	ncuisines*Price_range_3	31	130	122.1191*	43052.2658	51916.2206
* Optimal Value of Criterion						

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS Coefficient Progression for Votes



Fit Criteria for Votes



BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

Root MSE	209.78435
Dependent Mean	122.65634
R-Square	0.2916
Adj R-Sq	0.2760
AIC	70015
AICC	70021
BIC	64044
C(p)	122.11905
SBC	64906
A SE (Train)	43052
A SE (Test)	51916

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > t
Intercept	1	1155.236388	237.312267	4.87	<.0001
Price_range_1	1	-900.383744	222.085324	-4.05	<.0001
Price_range_2	1	-465.403226	203.853596	-2.28	0.0225
Price_range_3	1	-221.406420	164.647286	-1.34	0.1788
MajorCus American	1	-171.231965	257.697019	-0.66	0.5064
MajorCus Bakery	1	-641.330120	169.685305	-4.44	<.0001
MajorCus Cafe	1	-664.901684	259.506662	-3.33	0.0009
MajorCus Chinese	1	-665.697057	242.800579	-2.74	0.0061
MajorCus Continental	1	-472.213723	241.893346	-1.95	0.0510
MajorCus Desserts	1	-785.294113	329.245317	-2.32	0.0201
MajorCus Fast Food	1	-517.402588	345.636715	-1.50	0.1345
MajorCus Ice Cream	1	-369.650611	268.331030	-1.38	0.1684
MajorCus Italian	1	-721.427435	246.278440	-2.93	0.0034
MajorCus Mithai	1	-103.218332	197.617574	-0.52	0.6015
MajorCus Mughlai	1	-632.623096	278.422668	-2.27	0.0231
MajorCus North Indian	1	-600.488661	235.763406	-2.55	0.0109
MajorCus Others	1	-516.490453	237.537067	-2.17	0.0297
MajorCus Pizza	1	-972.917783	283.431049	-3.43	0.0008
MajorCus South Indian	1	-350.864520	140.796285	-2.49	0.0127
MajorCus Street Food	0	0	-	-	-
MajorCus*Price_range_1 American	1	422.705630	353.430979	1.20	0.2317
MajorCus*Price_range_1 Bakery	1	786.078364	162.968483	4.82	<.0001
MajorCus*Price_range_1 Cafe	1	719.294592	235.535997	3.05	0.0023
MajorCus*Price_range_1 Chinese	1	470.134114	222.065368	2.12	0.0343
MajorCus*Price_range_1 Continental	1	210.074676	229.117145	0.92	0.3592
MajorCus*Price_range_1 Desserts	1	789.948105	304.690132	2.59	0.0095
MajorCus*Price_range_1 Fast Food	1	341.429551	302.234465	1.13	0.2587
MajorCus*Price_range_1 Ice Cream	1	161.466085	227.466814	0.71	0.4778
MajorCus*Price_range_1 Italian	1	808.777116	234.784600	3.44	0.0006
MajorCus*Price_range_1 Mithai	1	154.709134	117.084218	1.32	0.1864
MajorCus*Price_range_1 Mughlai	1	564.435639	240.133172	2.35	0.0188
MajorCus*Price_range_1 North Indian	1	409.874159	216.704482	1.89	0.0586
MajorCus*Price_range_1 Others	1	516.205336	218.421303	2.36	0.0181
MajorCus*Price_range_1 Pizza	1	303.434363	176.682752	1.72	0.0860
MajorCus*Price_range_1 South Indian	1	165.672169	86.806626	1.91	0.0564
MajorCus*Price_range_1 Street Food	0	0	-	-	-
MajorCus*Price_range_2 American	1	81.934843	226.959940	0.36	0.7181
MajorCus*Price_range_2 Bakery	1	594.371485	144.302345	4.12	<.0001
MajorCus*Price_range_2 Cafe	1	543.593608	221.659626	2.45	0.0142
MajorCus*Price_range_2 Chinese	1	266.260394	207.742512	1.28	0.2000
MajorCus*Price_range_2 Continental	1	47.456669	207.975292	0.23	0.8194
MajorCus*Price_range_2 Desserts	1	704.338472	299.424483	2.35	0.0187
MajorCus*Price_range_2 Fast Food	1	149.045756	292.904051	0.51	0.6109
MajorCus*Price_range_2 Ice Cream	0	0	-	-	-
MajorCus*Price_range_2 Italian	1	498.989612	213.085690	2.34	0.0192
MajorCus*Price_range_2 Mithai	0	0	-	-	-
MajorCus*Price_range_2 Mughlai	1	411.461645	225.929575	1.82	0.0698
MajorCus*Price_range_2 North Indian	1	212.182163	201.978659	1.05	0.2935

MajorCus*Price_range_2 Others	1	377.790037	204.105252	1.85	0.0642
MajorCus*Price_range_2 Pizza	1	141.522181	153.269943	0.92	0.3559
MajorCus*Price_range_2 South Indian	0	0	-	-	-
MajorCus*Price_range_2 Street Food	0	0	-	-	-
MajorCus*Price_range_3 American	1	-304.214866	183.523510	-1.66	0.0974
MajorCus*Price_range_3 Bakery	0	0	-	-	-
MajorCus*Price_range_3 Cafe	1	89.354106	187.386694	0.48	0.6335
MajorCus*Price_range_3 Chinese	1	-24.115218	169.469303	-0.14	0.8868
MajorCus*Price_range_3 Continental	1	-114.660782	166.085331	-0.69	0.4900
MajorCus*Price_range_3 Desserts	1	20.488322	299.420619	0.07	0.9454
MajorCus*Price_range_3 Fast Food	1	-28.927303	283.649930	-0.10	0.9188
MajorCus*Price_range_3 Italian	1	254.293449	170.530072	1.49	0.1360
MajorCus*Price_range_3 Mughlai	1	-60.184050	190.523648	-0.32	0.7521
MajorCus*Price_range_3 North Indian	1	-67.911949	162.314724	-0.42	0.6757
MajorCus*Price_range_3 Others	1	-4.598252	163.562691	-0.03	0.9778
MajorCus*Price_range_3 Pizza	0	0	-	-	-
MajorCus*Price_range_3 South Indian	0	0	-	-	-
City_New Delhi	1	-6.539940	29.394178	-0.22	0.8239
City_New Delhi*Price_range_1	1	-105.985474	33.804319	-3.14	0.0017
City_New Delhi*Price_range_2	1	-111.456320	33.100786	-3.37	0.0008
City_New Delhi*Price_range_3	1	-66.635238	32.704021	-2.04	0.0416
Has_Online_delivery_No	1	-562.221935	67.441517	-8.34	<.0001
Has_Online_delivery_No*Price_range_1	1	263.080549	49.966990	5.27	<.0001
Has_Online_delivery_No*Price_range_2	1	263.485243	48.343671	5.45	<.0001
Has_Online_delivery_No*Price_range_3	1	371.771055	45.985639	8.06	<.0001
MajorCus*Has_Online_delivery_No American	1	105.119801	71.569068	1.47	0.1419
MajorCus*Has_Online_delivery_No Bakery	1	76.172233	55.254000	1.38	0.1681
MajorCus*Has_Online_delivery_No Cafe	1	41.975471	56.003831	0.75	0.4538
MajorCus*Has_Online_delivery_No Chinese	1	67.179058	54.445595	1.23	0.2173
MajorCus*Has_Online_delivery_No Continental	1	-58.652778	64.356809	-0.91	0.3622
MajorCus*Has_Online_delivery_No Desserts	1	42.308948	71.195375	0.59	0.5524
MajorCus*Has_Online_delivery_No Fast Food	1	104.110861	54.491227	1.91	0.0561
MajorCus*Has_Online_delivery_No Ice Cream	1	74.962791	64.317403	1.17	0.2437
MajorCus*Has_Online_delivery_No Italian	1	191.557917	67.761547	2.83	0.0047
MajorCus*Has_Online_delivery_No Mithai	1	22.228679	105.762149	0.21	0.8335
MajorCus*Has_Online_delivery_No Mughlai	1	76.131257	67.435324	1.13	0.2590
MajorCus*Has_Online_delivery_No North Indian	1	83.221045	51.906671	1.60	0.1089
MajorCus*Has_Online_delivery_No Others	1	64.357059	54.107784	1.56	0.1190
MajorCus*Has_Online_delivery_No Pizza	1	6.578733	65.578964	0.10	0.9201
MajorCus*Has_Online_delivery_No South Indian	1	7.661612	65.090331	0.12	0.9063
MajorCus*Has_Online_delivery_No Street Food	0	0	-	-	-
Has_Table_booking_No	1	-160.983592	115.566659	-1.39	0.1637
Has_Table_booking_No*Price_range_2	1	-200.913807	37.744996	-5.32	<.0001
Has_Table_booking_No*Price_range_3	1	-74.697271	35.177300	-2.12	0.0338
MajorCus*Has_Table_booking_No American	1	-133.138286	136.453236	-0.98	0.3293
MajorCus*Has_Table_booking_No Bakery	0	0	-	-	-
MajorCus*Has_Table_booking_No Cafe	1	89.891362	116.046886	0.80	0.5482
MajorCus*Has_Table_booking_No Chinese	1	106.367308	112.542644	0.95	0.3448
MajorCus*Has_Table_booking_No Continental	1	156.847358	114.914220	1.36	0.1723
MajorCus*Has_Table_booking_No Desserts	0	0	-	-	-
MajorCus*Has_Table_booking_No Fast Food	1	167.229525	168.356524	1.11	0.2661
MajorCus*Has_Table_booking_No Ice Cream	0	0	-	-	-
MajorCus*Has_Table_booking_No Italian	1	-109.737010	117.361716	-0.94	0.3498
MajorCus*Has_Table_booking_No Mithai	0	0	-	-	-
MajorCus*Has_Table_booking_No Mughlai	1	73.431450	133.776917	0.55	0.5831
MajorCus*Has_Table_booking_No North Indian	1	96.108975	109.442175	0.88	0.3799
MajorCus*Has_Table_booking_No Others	1	-15.974299	111.696675	-0.14	0.8863
MajorCus*Has_Table_booking_No Pizza	1	226.761323	148.796640	1.52	0.1278
MajorCus*Has_Table_booking_No South Indian	0	0	-	-	-
MajorCus*Has_Table_booking_No Street Food	0	0	-	-	-
City_New Delhi*Has_Table_booking_No	1	115.670339	22.284755	5.19	<.0001
Has_Online_delivery_No*Has_Table_booking_No	1	184.272560	23.873526	7.72	<.0001

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

City_New Delhi*Has_Table_booking_No	1	115.670339	22.284755	5.19	<.0001
Has_Online_delivery_No*Has_Table_booking_No	1	164.272580	23.873526	7.72	<.0001
ncuisine*	1	149.616302	78.370216	1.96	0.0501
ncuisine*Price_range_1	1	-67.247650	29.746625	-2.26	0.0238
ncuisine*Price_range_2	1	-75.082995	24.294409	-3.09	0.0020
ncuisine*MajorCus American	1	263.826464	156.952667	-1.81	0.0706
ncuisine*MajorCus Bakery	1	-78.533544	84.481429	-0.93	0.3526
ncuisine*MajorCus Cafe	1	0.547780	81.728489	0.01	0.9947
ncuisine*MajorCus Chinese	1	-12.942245	81.336428	-0.16	0.8736
ncuisine*MajorCus Continental	1	-57.860042	84.957488	-0.68	0.4959
ncuisine*MajorCus Desserts	1	-107.552318	115.020378	-0.94	0.3498
ncuisine*MajorCus Fast Food	1	-124.860521	87.781695	-1.42	0.1550
ncuisine*MajorCus Ice Cream	1	96.696122	156.378358	0.62	0.5364
ncuisine*MajorCus Italian	1	-10.535483	90.393064	-0.12	0.9072
ncuisine*MajorCus Mithai	1	-76.257636	87.373909	-0.87	0.3828
ncuisine*MajorCus Mughlai	1	-121.831428	130.194003	-0.94	0.3494
ncuisine*MajorCus North Indian	1	-21.905310	74.355909	-0.29	0.7683
ncuisine*MajorCus Others	1	-138.687088	77.395255	-1.79	0.0732
ncuisine*MajorCus Pizza	1	315.042319	180.392721	1.75	0.0808
ncuisine*MajorCus South Indian	1	198.408998	101.187168	1.96	0.0499
ncuisine*MajorCus Street Food	0	0	-	-	-
ncuisine_log	1	293.103582	157.537380	-1.86	0.0629
ncuisine_log*Price_range_1	1	122.586217	65.854927	1.88	0.0627
ncuisine_log*Price_range_2	1	154.470484	59.725435	2.59	0.0097
ncuisine_log*MajorCus American	1	753.850174	414.889157	1.82	0.0693
ncuisine_log*MajorCus Bakery	1	222.934087	169.586609	1.31	0.1887
ncuisine_log*MajorCus Cafe	1	94.415685	188.006968	0.56	0.5742
ncuisine_log*MajorCus Chinese	1	46.925081	164.147056	0.29	0.7750
ncuisine_log*MajorCus Continental	1	259.096109	193.195635	1.34	0.1799
ncuisine_log*MajorCus Desserts	1	225.903372	226.035655	1.00	0.3176
ncuisine_log*MajorCus Fast Food	1	243.663444	172.566372	1.41	0.1580
ncuisine_log*MajorCus Ice Cream	1	-97.111930	279.278910	-0.35	0.7281
ncuisine_log*MajorCus Italian	1	45.642127	203.156252	0.22	0.8222
ncuisine_log*MajorCus Mithai	1	166.866831	183.028001	0.91	0.3620
ncuisine_log*MajorCus Mughlai	1	261.219658	248.496568	1.05	0.2932
ncuisine_log*MajorCus North Indian	1	47.950320	151.248944	0.32	0.7510
ncuisine_log*MajorCus Others	1	339.930757	158.167734	2.15	0.0317
ncuisine_log*MajorCus Pizza	1	-309.716546	313.545110	-0.99	0.3233
ncuisine_log*MajorCus South Indian	1	-363.309991	200.969190	-1.91	0.0565
ncuisine_log*MajorCus Street Food	0	0	-	-	-

Figure 11: Backward Selection

- In case of backward selection, we observe that the R square is 0.27 which is lower than that of forward selection.
- The AIC of backward selection is better than forward and LASSO with a value of 70015 which implies a better fit.
- In Backward selection, price range 4 restaurant have higher possibility of getting votes.
- If online delivery is available, then voting is bound to increase.
- The R-square is lower than forward selection (0.28), we will use stepwise selection and best subset selection in order to better explain the variance in the model.

Stepwise Selection:

The GLMSELECT Procedure		
Data Set	WORK.ZOMATO_TRAINING	
Test Data Set	WORK.ZOMATO_TEST	
Dependent Variable	Votes	
Selection Method	Backward	
Select Criterion	C(p)	
Stop Criterion	C(p)	
Effect Hierarchy Enforced	Single	

Observation Profile for Analyze Data	
Number of Observations Read	5981
Number of Observations Used	5977
Number of Observations Used for Training	5977

Observation Profile for Test Data	
Number of Observations Read	1495
Number of Observations Used	1491

Class Level Information		
Class	Levels	Values
MajorCuis	16	American Bakery Cafe Chinese Continental Desserts Fast Food Ice Cream Italian Mithai Mughlai North Indian Others Pizza South Indian Street Food
City	2 *	New Delhi Others
Has_Online_delivery	2 *	No Yes
Has_Table_booking	2 *	No Yes
Price_range	4 *	1 2 3 4

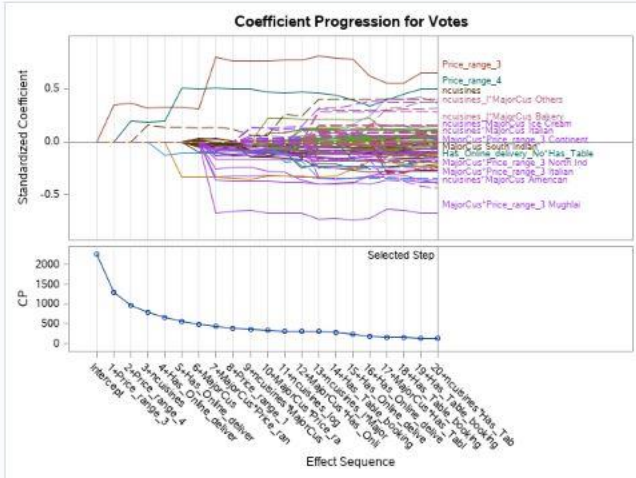
* Associated Parameters Split		
Dimensions		
Number of Effects	29	
Number of Effects after Splits	83	
Number of Parameters	265	

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

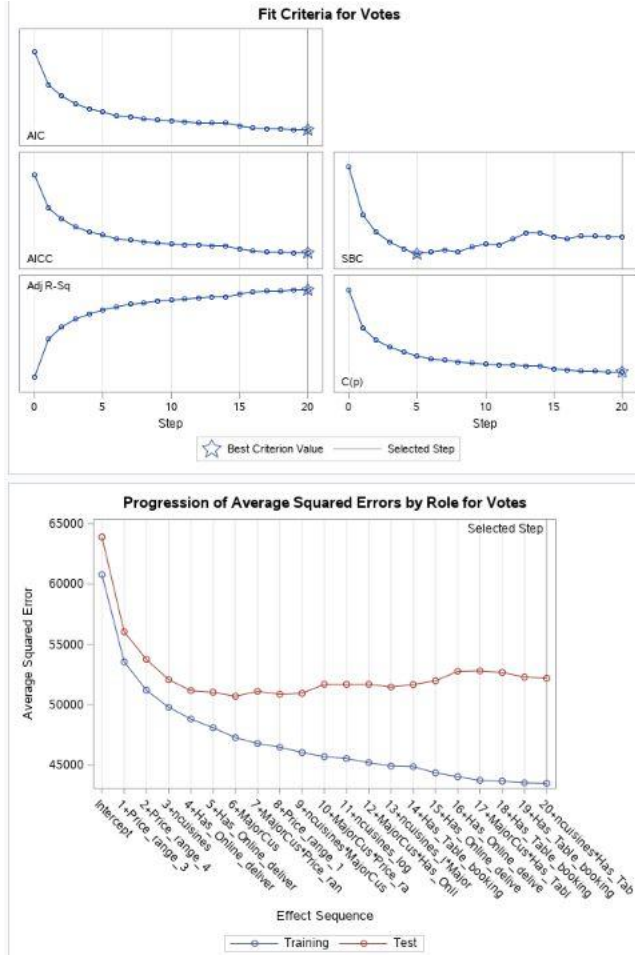
Stepwise Selection Summary							
Step	Effect Entered	Effect Removed	Number Effects In	Number Params In	CP	A SE	Test A SE
0	Intercept		1	1	2266.1863	60777.4876	63868.5477
1	Price_range_3		2	2	1290.0690	53551.1566	56036.5642
2	Price_range_4		3	3	972.7389	51196.7111	53753.5413
3	ncut=Ies		4	4	784.2051	49791.8915	52076.0479
4	Has_Online_delivery_No		5	5	854.3178	48819.4748	51169.0486
5	Has_Online_delivery_No Price_range_4		6	6	557.4723	40960.6813	51032.4411
6	MajorCus		7	21	477.1219	47277.0589	50700.3594
7	MajorCus Price_range_3		8	33	434.9069	46786.8507	51108.1370
8	Price_range_1		9	34	395.3641	46482.5524	50684.2913
9	ncut=Ies MajorCus		10	49	365.2565	46039.3749	50044.4075
10	MajorCus Price_range_4		11	59	336.8403	45687.1449	51695.0727
11	ncut=Ies_log		12	60	319.6738	45542.5574	51690.0032
12	MajorCus Has_Online_delivery_No		13	75	304.0348	45204.5829	51696.5738
13	ncut=Ies_MajorCus		14	90	296.0837	44924.7862	51484.0778
14	Has_Table_booking_Yes		15	91	290.3098	44867.4485	51672.3542
15	Has_Online_delivery_No Has_Table_booking_Yes		16	92	222.7983	43354.9056	51979.6712
16	Has_Online_delivery_No Price_range_3		17	93	181.9226	44038.8097	52770.4916
17	MajorCus Has_Table_booking_Yes		18	103	161.0348	43737.3389	52803.6626
18	Has_Table_booking_Yes Price_range_4		19	104	154.3013	43672.9481	52671.4303
19	Has_Table_booking_Yes Price_range_3		20	105	136.8499	43529.5313	52285.3952
20	ncut=Ies Has_Table_booking_Yes		21	106	134.4076*	43496.7780	52193.2116

Selection stopped at a local minimum of the $C(p)$ criterion.

Stop Details			
Candidate For	Effect	Candidate C(p)	Compare C(p)
Entry	City_Others	134.9338	> 134.4076
Removal	nousines*Has_Table_booking_Yes	136.8499	> 134.4076



BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS



Dependent Mean	122.66634
R-Square	0.2843
Adj R-Sq	0.2715
AIC	70028
AICC	70032
BIC	64054
C(p)	134.40761
SBC	64759
ΔSE (Train)	43497
ΔSE (Test)	52193

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > t
Intercept	1	148.906561	86.369471	1.72	0.0847
Price_range_1	1	51.466630	7.455290	-6.90	<.0001
Price_range_3	1	462.667554	120.787976	3.83	0.0001
Price_range_4	1	575.155624	137.409975	4.19	<.0001
MajorCus American	1	-34.941508	120.744719	-0.29	0.7723
MajorCus Bakery	1	-60.971962	96.813747	-0.62	0.5364
MajorCus Cafe	1	-94.942178	94.711492	-1.00	0.3162
MajorCus Chinese	1	-123.414327	94.071024	-1.31	0.1898
MajorCus Continental	1	-123.537602	99.525669	-1.24	0.2146
MajorCus Desserts	1	8.372418	130.242222	0.06	0.9487
MajorCus Fast Food	1	-0.220602	100.196126	-0.00	0.9962
MajorCus Ice Cream	1	-207.563848	160.890573	-1.29	0.1971
MajorCus Italian	1	-143.034921	103.284192	-1.38	0.1661
MajorCus Mithai	1	36.769054	139.842596	0.28	0.7828
MajorCus Mughlai	1	15.325153	135.406035	0.11	0.9099
MajorCus North Indian	1	-112.956012	87.485357	-1.29	0.1967
MajorCus Others	1	-11.915086	90.507624	-0.13	0.8952
MajorCus Pizza	1	-424.058172	180.945691	-2.34	0.0191
MajorCus South Indian	1	-178.080639	115.411435	-1.54	0.1229
MajorCus Street Food	0	0	-	-	-
MajorCus*Price_range_3 American	1	-409.964851	139.883596	-2.93	0.0034
MajorCus*Price_range_3 Bakery	1	-567.948402	141.131915	-4.24	<.0001
MajorCus*Price_range_3 Cafe	1	-452.941516	122.996395	-3.68	0.0002
MajorCus*Price_range_3 Chinese	1	-310.630857	122.826685	-2.53	0.0115
MajorCus*Price_range_3 Continental	1	-145.534779	126.758772	-1.15	0.2510
MajorCus*Price_range_3 Desserts	1	-563.121265	181.211040	-3.27	0.0011
MajorCus*Price_range_3 Fast Food	1	-155.290689	154.996377	-1.00	0.3164
MajorCus*Price_range_3 Italian	1	-277.566200	127.123975	-2.18	0.0290
MajorCus*Price_range_3 Mughlai	1	-469.840636	138.487026	-3.39	0.0007
MajorCus*Price_range_3 North Indian	1	-296.326852	119.775325	-2.47	0.0134
MajorCus*Price_range_3 Others	1	-361.131460	121.458542	-2.97	0.0030
MajorCus*Price_range_3 Pizza	1	-147.683988	152.258133	-0.97	0.3321
MajorCus*Price_range_3 South Indian	0	0	-	-	-
MajorCus*Price_range_4 American	1	104.766057	167.743598	0.62	0.5323
MajorCus*Price_range_4 Cafe	1	-278.286232	158.193486	-1.76	0.0796
MajorCus*Price_range_4 Chinese	1	-71.224748	140.894873	-0.51	0.6132
MajorCus*Price_range_4 Continental	1	179.706350	140.434608	1.28	0.2007
MajorCus*Price_range_4 Desserts	1	-354.557169	250.267125	-1.42	0.1568
MajorCus*Price_range_4 Fast Food	1	18.062585	248.936203	0.07	0.9421
MajorCus*Price_range_4 Italian	1	-330.845932	146.564131	-2.26	0.0240
MajorCus*Price_range_4 Mughlai	1	-220.730609	166.343173	-1.33	0.1646
MajorCus*Price_range_4 North Indian	1	-19.507774	132.897806	-0.15	0.8833
MajorCus*Price_range_4 Others	1	-135.386527	134.477246	-1.01	0.3141
MajorCus*Price_range_4 Pizza	0	0	-	-	-
Has_Online_delivery_No	1	-120.577516	49.877787	-2.42	0.0157
Has_Online_delivery_No*Price_range_3	1	98.426136	23.107017	4.26	<.0001
Has_Online_delivery_No*Price_range_4	1	-261.894279	47.919472	-5.66	<.0001
MajorCus*Has_Online_delivery_No American	1	116.220545	70.171624	1.66	0.0977
MajorCus*Has_Online_delivery_No Bakery	1	82.862269	55.251796	1.50	0.1337
MajorCus*Has_Online_delivery_No Cafe	1	48.239602	55.234630	0.87	0.3825
MajorCus*Has_Online_delivery_No Chinese	1	73.272378	53.902550	1.36	0.1741

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

MajorCus*Has_Online_delivery_No Chinese	1	73.272378	53.902550	1.36	0.1741
MajorCus*Has_Online_delivery_No Continental	1	-50.697334	63.724719	-0.80	0.4283
MajorCus*Has_Online_delivery_No Desserts	1	55.823947	71.168112	0.78	0.4328
MajorCus*Has_Online_delivery_No Fast Food	1	111.254470	54.201717	2.05	0.0402
MajorCus*Has_Online_delivery_No Ice Cream	1	80.006922	64.207867	1.25	0.2128
MajorCus*Has_Online_delivery_No Italian	1	209.635250	65.767575	3.19	0.0014
MajorCus*Has_Online_delivery_No Mithai	1	-11.111247	105.540663	-0.11	0.9162
MajorCus*Has_Online_delivery_No Mughlai	1	79.500593	66.046807	1.20	0.2286
MajorCus*Has_Online_delivery_No North Indian	1	92.088094	51.095334	1.80	0.0716
MajorCus*Has_Online_delivery_No Others	1	72.696699	53.119601	1.37	0.1712
MajorCus*Has_Online_delivery_No Pizza	1	9.245537	63.786507	0.14	0.8848
MajorCus*Has_Online_delivery_No South Indian	1	5.404988	63.576668	0.09	0.9323
MajorCus*Has_Online_delivery_No Street Food	0	0	-	-	-
Has_Table_booking_Yes	1	259.671543	110.173821	2.36	0.0185
Has_Table_booking_Yes*Price_range_3	1	-111.501094	24.542621	-4.54	<.0001
Has_Table_booking_Yes*Price_range_4	1	-164.580254	35.800736	-4.60	<.0001
MajorCus*Has_Table_booking_Yes American	1	117.995357	135.803949	0.87	0.3843
MajorCus*Has_Table_booking_Yes Cafe	1	-60.841520	115.968650	-0.78	0.4335
MajorCus*Has_Table_booking_Yes Chinese	1	-130.978115	112.196881	-1.17	0.2431
MajorCus*Has_Table_booking_Yes Continental	1	-191.872678	114.782412	-1.67	0.0947
MajorCus*Has_Table_booking_Yes Fast Food	1	231.256082	168.207277	-1.37	0.1692
MajorCus*Has_Table_booking_Yes Italian	1	93.225705	117.199625	0.80	0.4264
MajorCus*Has_Table_booking_Yes Mughlai	1	-88.412261	133.722112	-0.66	0.5085
MajorCus*Has_Table_booking_Yes North Indian	1	-117.805400	109.230570	-1.08	0.2809
MajorCus*Has_Table_booking_Yes Others	1	0.923707	111.500381	0.01	0.9634
MajorCus*Has_Table_booking_Yes Pizza	1	-187.872143	148.761005	-1.26	0.2067
MajorCus*Has_Table_booking_Yes South Indian	0	0	-	-	-
Has_Online_delivery_No*Has_Table_booking_Yes	1	-173.150359	23.563666	-7.35	<.0001
ncuisinees	1	90.073698	71.386071	1.26	0.2071
ncuisinees*MajorCus American	1	237.014465	143.722607	-1.65	0.0992
ncuisinees*MajorCus Bakery	1	-86.470197	84.578137	-1.02	0.3086
ncuisinees*MajorCus Cafe	1	10.883700	79.743676	0.14	0.8914
ncuisinees*MajorCus Chinese	1	5.994808	80.322571	0.07	0.9405
ncuisinees*MajorCus Continental	1	-31.844404	82.164091	-0.39	0.6963
ncuisinees*MajorCus Desserts	1	-84.534697	114.716119	-0.82	0.4099
ncuisinees*MajorCus Fast Food	1	-125.782629	87.777264	-1.43	0.1519
ncuisinees*MajorCus Ice Cream	1	87.067353	156.466183	0.56	0.5779
ncuisinees*MajorCus Italian	1	12.126001	87.796611	0.14	0.8902
ncuisinees*MajorCus Mithai	1	-75.156683	84.941947	-0.88	0.3783
ncuisinees*MajorCus Mughlai	1	-132.334413	128.983326	-1.03	0.3049
ncuisinees*MajorCus North Indian	1	-20.744681	72.890245	-0.29	0.7754
ncuisinees*MajorCus Others	1	-119.086103	75.972107	-1.57	0.1171
ncuisinees*MajorCus Pizza	1	295.967536	178.116661	1.66	0.0666
ncuisinees*MajorCus South Indian	1	190.131366	101.356647	1.88	0.0607
ncuisinees*MajorCus Street Food	0	0	-	-	-
ncuisinees*Has_Table_booking_Yes	1	15.866124	7.545983	2.10	0.0355
ncuisinees_log	1	-167.819061	144.864725	-1.16	0.2467
ncuisinees_1*MajorCus American	1	595.821520	367.852579	1.62	0.1053
ncuisinees_1*MajorCus Bakery	1	223.812126	169.745315	1.32	0.1874
ncuisinees_1*MajorCus Cafe	1	64.314966	163.992189	0.39	0.6949
ncuisinees_1*MajorCus Chinese	1	-2.889561	162.561672	-0.02	0.9659
ncuisinees_1*MajorCus Continental	1	186.247197	167.449901	0.99	0.3205
ncuisinees_1*MajorCus Desserts	1	193.340529	225.807407	0.86	0.3919
ncuisinees_1*MajorCus Fast Food	1	230.266982	172.417802	1.34	0.1818
ncuisinees_1*MajorCus Ice Cream	1	-95.588233	279.322147	-0.34	0.7322
ncuisinees_1*MajorCus Italian	1	-31.152932	196.665145	-0.16	0.8741
ncuisinees_1*MajorCus Mithai	1	150.894698	180.904213	0.83	0.4043
ncuisinees_1*MajorCus Mughlai	1	285.628694	244.005949	1.17	0.2418
ncuisinees_1*MajorCus North Indian	1	26.848057	148.321063	0.19	0.8488
ncuisinees_1*MajorCus Others	1	268.306559	155.547666	1.65	0.0639
ncuisinees_1*MajorCus Pizza	1	275.567786	308.043414	-0.89	0.3711
ncuisinees_1*MajorCus South Indian	1	-376.706208	201.378628	-1.87	0.0614
ncuisinees_1*MajorCus Street Food	0	0	-	-	-

Figure 11: Stepwise Selection

- The stepwise selection has R square value of 0.27 which is higher than that of LASSO however lower than the Forward selection.
- In terms of fit, the AIC is 70028 which is equivalent to the forward selection model.
- The price range result showing the same pattern as before, when the price range is low especially in range 1, the possibility of getting votes is low.
- If online delivery is available, then voting is bound to increase.
- Forward selection has the highest R square for now.

Logit Model:

The LOGISTIC Procedure

Model Information		
Data Set	WORK.ZOMATO_SAMPLED	
Response Variable	rating	
Number of Response Levels	3	
Weight Variable	Selected	Selection Indicator
Model	generalized logit	
Optimization Technique	Newton-Raphson	

Number of Observations Read	7476
Number of Observations Used	5977
Sum of Weights Read	5981
Sum of Weights Used	5977

Response Profile			
Ordered Value	rating	Total Frequency	Total Weight
1	High	597	597.0000
2	Moderate	3364	3364.0000
3	Poor	2016	2016.0000

Logits modeled use rating='Poor' as the reference category.

Analysis Variable : Match
Mean
0.6481605

Frequency Percent Row Pct Col Pct	Table of rating by I_rating				
	rating	I_rating(Into: rating)			
		High	Moderate	Poor	Total
	High	35	116	1	152
		2.34	7.76	0.07	10.17
		23.03	76.32	0.66	
		68.63	12.06	0.21	
	Moderate	15	665	212	892
		1.00	44.48	14.18	59.67
		1.68	74.55	23.77	
		29.41	69.13	43.98	
	Poor	1	181	269	451
		0.07	12.11	17.99	30.17
		0.22	40.13	59.65	
		1.96	18.81	55.81	
	Total	51	962	482	1495
		3.41	64.35	32.24	100.00

Figure 12: Logit Model

- From the confusion Matrix in Figure 12 we observe that High is correctly predicted 23.03% times, Moderate is correctly predicted 74.55% times while poor is predicted correctly 59.5% times.
- The overall accuracy of the model is 64.81%.

Probit Model:

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

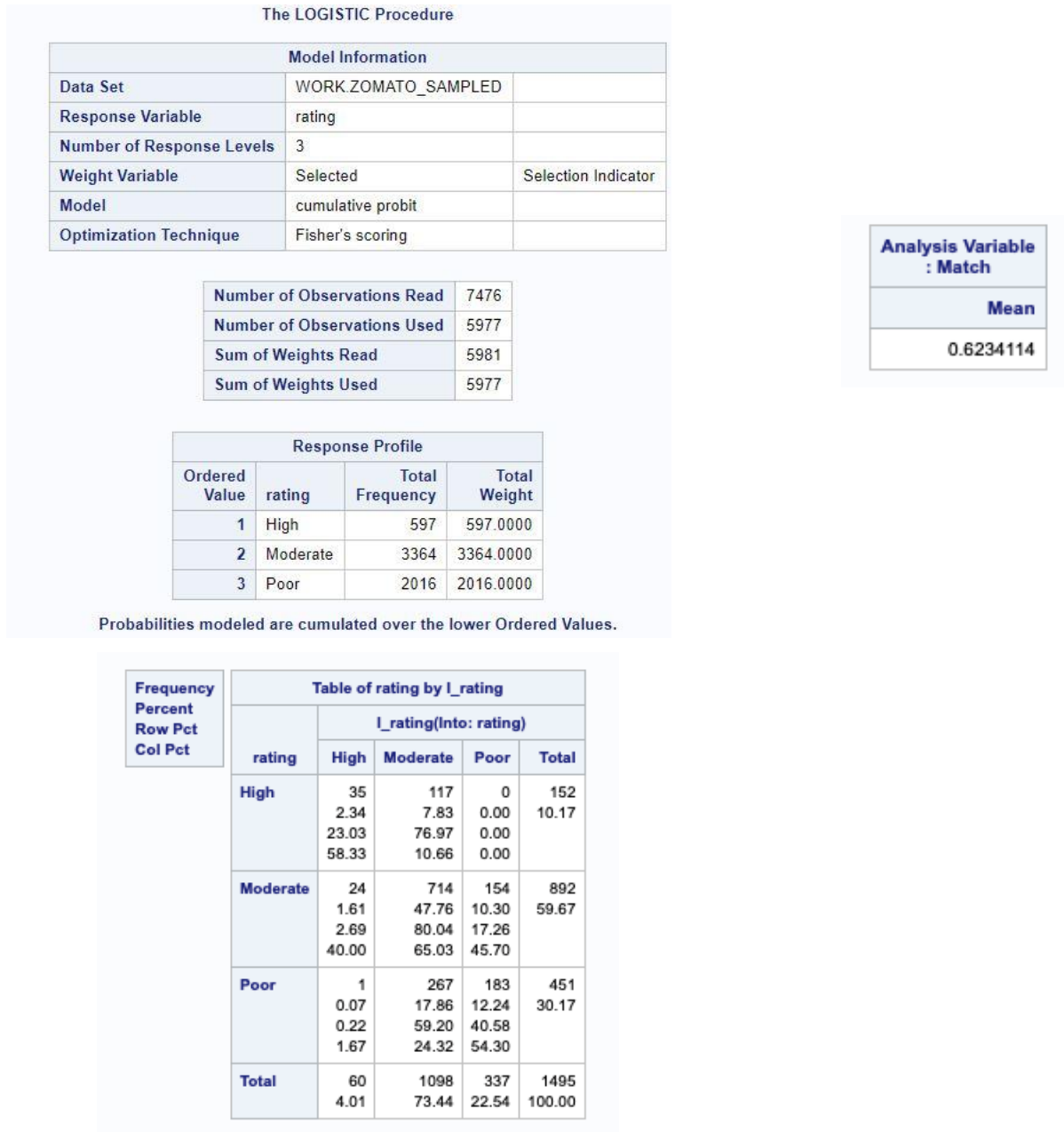


Figure 13:Probit Model

- From the confusion Matrix in Figure 13 we observe that High is correctly predicted 23.03% times, Moderate is correctly predicted 80.04% times while poor is predicted correctly 40.58.% times.
- The overall accuracy of the model is 62.34%.

Findings and Managerial Implications

Model	R-Sq	AIC
Linear Regression	0.23	-
LASSO	0.26	70080
Forward	0.28	70028
Backward	0.27	70015
Stepwise	0.27	70028

Table 2: Results for linear models for Votes

Model	%correct High	%correct Moderate	%correct Low	Overall Accuracy
Logit	23.03	74.55	59.65	64.8%
Probit	23.03	80.04	40.58	62.32%

Table 3: Results for classification of ratings

Based on the descriptive analysis, opening a Italian or continental restaurants would get more votes and higher rating compare to other cuisine. Italian and continental restaurants are fall at price range 4 restaurants.

Moreover, Forward selection is the best model among the five models. Based on Forward selection model, if the restaurants wants to get more voting. First, the higher priced restaurants, the possibility of getting votes is higher. Second, online delivery service will increase the possibility of getting more votes, but not work on the price rang 4 restaurants. If The luxury restaurants focus on getting more votes, they do not need to spend money on provide online delivery. Third, providing table booking increase the possibility of getting more votes but this also not work on restaurant in price range 3 and 4. This result shows if the restaurants are in lower price range, they should provide table booking in order to increase the possibility of getting more votes.

In Summary, based on the dataset, open a Italian or continental restaurants would have the highest possibility of getting higher rating and more votes and since they are high price range restaurants, the owners do not need to spend money on providing delivery service and table booking.

Conclusion

90 % of the people buying decisions are based on online review. Having many good ratings is a big opportunity to drive more traffic to your business (Vermaat, 2018). Only have good rating is not enough to attract more customers, you have to have huge quantity reviews to make people believe in restaurant consistency.

From Zomato dataset, we found out the price range affect the possibility for people to rate the restaurant. Restaurants owners should adjust their service based on their food price to get more rating. For the price range 1, people do not bother or feel obligate to rate the restaurants but for the restaurant people spend a lot of money on, they expect more than price range 1 restaurants so they feel more obligate to rate the restaurants. However, they can provide online delivery and table booking to increase the possibility of getting more reviews. Second, we found out in price range 4 restaurants, people do not care about if they have online delivery. Customer service and restaurant atmosphere are a huge factors for people paying more money, people do not order take out at luxury restaurants which lead to the result luxury restaurants do not need to hire people to do delivery.

There are a lot of factors affect restaurants rating and the motivation to rate. The dataset does not include many other factors. According to Wakefield (2004) who identify five important factors for restaurants which are customers interaction, technological investments, capacity, quality of physical facilities. Satisfied customer service increases the possibility of rating them online (Vermaat, 2018).in the future dataset, these factors can be added in the survey to come up with a better outcome.

References

1. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani(2009). An Introduction to Statistical Learning. Springer
2. Zomato Restaurants Data retrieved from <https://www.kaggle.com/shrutihehta/zomato-restaurants-data>
3. Budhwar, K. (2004). Understand the success factors for independent restaurants in Delhi.
4. Vermaat, D. (2018). 11 tips to encourage customers to leave reviews for your business online. Retrieved from loopyloyalty: <https://blog.loopyloyalty.com>
5. Wakefield, K. (2004). Research Reveals Five Most Important Success Factors For Pizza. Retrieved from Pizza marketing Quarterly.

Appendix

SAS Codes

```
LIBNAME PRO 'C:\Users\vdr180000\Desktop';

/* This imports the csv dataset into SAS. */
PROC IMPORT OUT= zomato_in
    DATAFILE= "C:\Users\vdr180000\Desktop\zomato_in.csv"
    DBMS=CSV REPLACE;
    GETNAMES=YES;
    DATAROW=2;
RUN;
/* generating the working dataset in Work library */
data zomato;
    set zomato_in;
run;

/* Defining log */
data zomato;
    set zomato_in;
    ncuisines_log=log(ncuisines);
run;

/*splitting*/
proc surveyselect data=zomato out=zomato_sampled outall samprate=0.8 seed=2;
run;
data zomato_training zomato_test;
set zomato_sampled;
if selected then output zomato_training; /* Tell SAS that only keep the 80% selected one in
sample. The rest
will be in test data */
else output zomato_test;
run;

/*linear regression*/
ods graphics on;
proc glm data =zomato_sampled ;
class MajorCus(ref='South Indian') City(ref='Others') Has_Online_delivery(ref='No')
Has_Table_booking(ref='No') Price_range(ref='1');
model votes = price_range MajorCus City Has_Online_delivery Has_Table_booking ncuisines
ncuisines_log price_range*MajorCus MajorCus*City /solution;
run;
```


ods graphics off;

```
/*ASE in train vs. test data */
/* Stepwise with LASSO regression with AIC as criteria */
proc glmselect data=zomato_training testdata=zomato_test seed = 2 plots=all;
class MajorCus City(split) Has_Online_delivery(split) Has_Table_booking(split)
Price_range(split);
model Votes=
Price_range|MajorCus|City|Has_Online_delivery|Has_Table_booking|ncuisines|ncuisines_log
@2
/selection=lasso(choose=cv stop=none) hierarchy=single cvmethod=split(5) showpvalues;
performance buildsscp=incremental;
run;
```

```
/*ASE in train vs. test data */
/* Selection with forward,backward and stepwise */
proc glmselect data=zomato_training testdata=zomato_test seed = 2 plots=all;
class MajorCus City(split) Has_Online_delivery(split) Has_Table_booking(split)
Price_range(split);
model Votes=
Price_range|MajorCus|City|Has_Online_delivery|Has_Table_booking|ncuisines|ncuisines_log
@2
/selection=forward(select=cp) hierarchy=single showpvalues;
performance buildsscp=incremental;
run;
```

```
proc glmselect data=zomato_training testdata=zomato_test seed = 2 plots=all;
class MajorCus City(split) Has_Online_delivery(split) Has_Table_booking(split)
Price_range(split);
model Votes=
Price_range|MajorCus|City|Has_Online_delivery|Has_Table_booking|ncuisines|ncuisines_log
@2
/selection=backward(select=cp) hierarchy=single showpvalues;
performance buildsscp=incremental;
run;
```

```
proc glmselect data=zomato_training testdata=zomato_test seed = 2 plots=all;
class MajorCus City(split) Has_Online_delivery(split) Has_Table_booking(split)
Price_range(split);
model Votes=
Price_range|MajorCus|City|Has_Online_delivery|Has_Table_booking|ncuisines|ncuisines_log
@2
/selection=stepwise(select=cp) hierarchy=single showpvalues;
performance buildsscp=incremental;
```

```
run;
```

```
/*logistic regression*/
proc logistic data=zomato_sampled;
class MajorCus(ref='South Indian') City(ref='Others') Has_Online_delivery(ref='No')
Has_Table_booking(ref='No') Price_range(ref='1');
model rating(event='High') = Price_range City MajorCus Has_Online_delivery
Has_Table_booking ncuisines ncuisines_log Votes/ link=glogit;
weight selected;
output out=preds predprobs=individual; /*only training sample is used for estimation, since
selected = 0 for test sample */
score data=zomato_test out=zomatoPred;
run;
proc freq data=zomatoPred;
    table rating*I_rating / out=CellCounts;
    run;
data CellCounts;
    set CellCounts;
    Match=0;
    if rating=I_rating then Match=1;
    run;
proc means data=CellCounts mean;
    freq count;
    var Match;
    run;
```

```
/*probit model*/
proc logistic data=zomato_sampled;
class MajorCus(ref='South Indian') City(ref='Others') Has_Online_delivery(ref='No')
Has_Table_booking(ref='No') Price_range(ref='1');
model rating(event='High') = Price_range City MajorCus Has_Online_delivery
Has_Table_booking ncuisines ncuisines_log Votes/ link=probit;
weight selected;
output out=preds1 predprobs=individual; /*only training sample is used for estimation, since
selected = 0 for test sample */
score data=zomato_test out=zomatoPred1;
run;
proc freq data=zomatoPred1;
    table rating*I_rating / out=CellCounts1;
    run;
data CellCounts1;
    set CellCounts1;
    Match=0;
    if rating=I_rating then Match=1;
```

BUAN 6337.002 – GROUP PROJECT – ZOMATO PREDICTIVE ANALYSIS

```
run;  
proc means data=CellCounts1 mean;  
  freq count;  
  var Match;  
run;
```