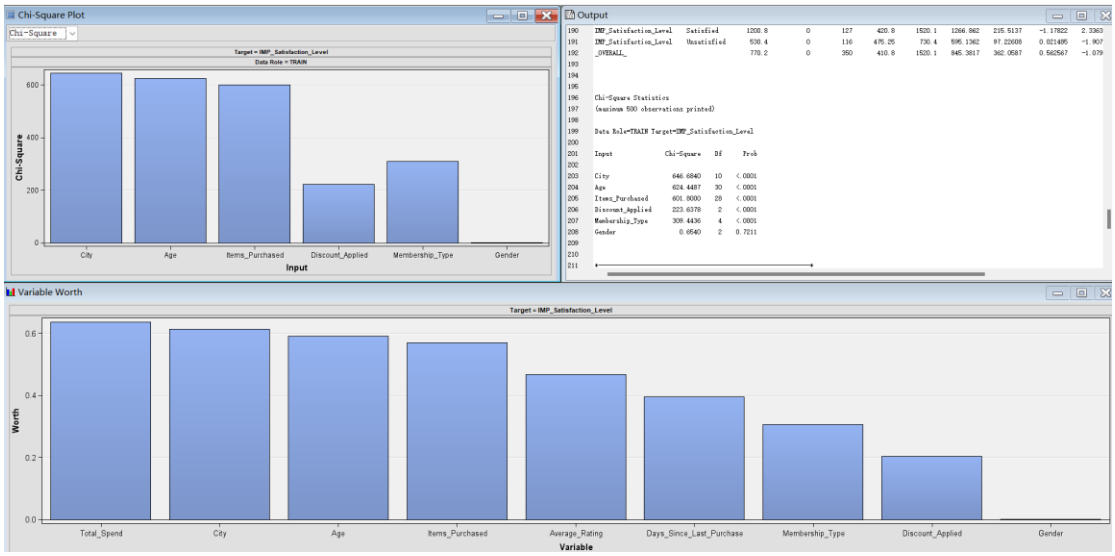
Results and Analysis

Explore Data Analysis:

The result of StatExplore:

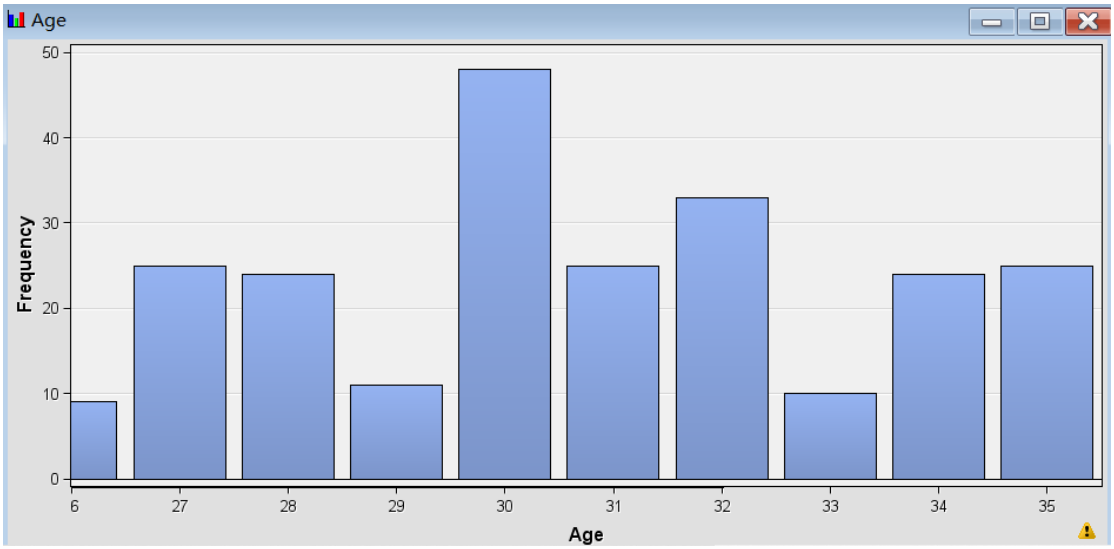


The explore for variable of age:

The age of individuals in the data set is concentrated around 30 years old.

In terms of satisfaction, the most common feedback was "satisfied."

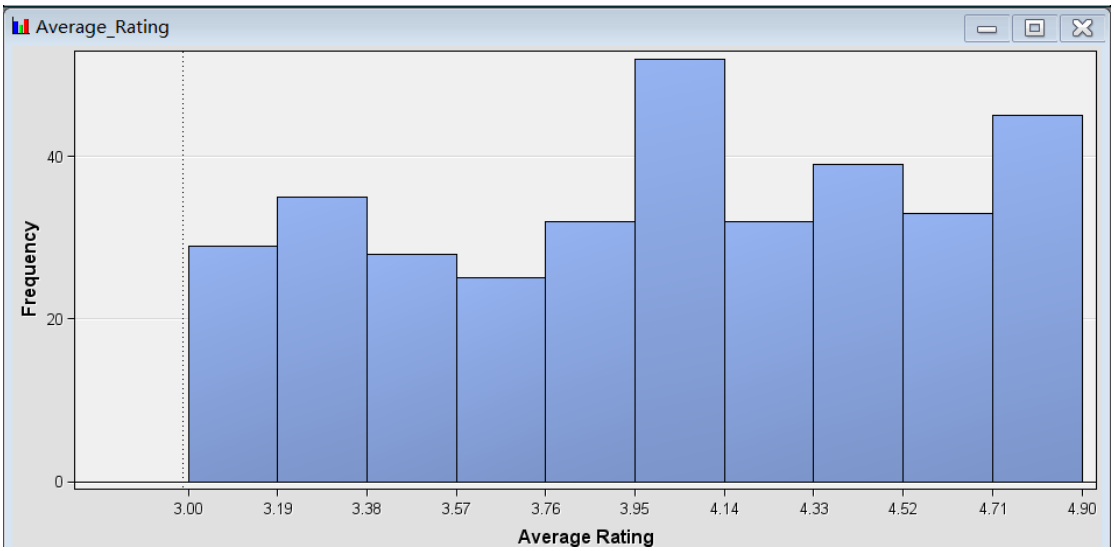
Obs #	Variable Name	Label	Type	Percent Missing	Minimum	Maximum	Mean	Number of Levels	Mode ...	Mode
1	IMP_Satisfaction_Level	Imputed: Satisfaction Level	CLASS	0	.	.	.	3	36.28571	SATISFIED
2	Age		VAR	0	26	43	33.59714	.	.	
3	Customer_ID	Customer ID	VAR	0	101	450	275.5	.	.	



The explore for variable of Average Rating:

The minimum value of the average rating is 3, the maximum value is 4.9, the mean is approximately 4.19, and the mode is 4.90. This indicates that most customers' ratings tend to be on the higher side, with greater satisfaction.

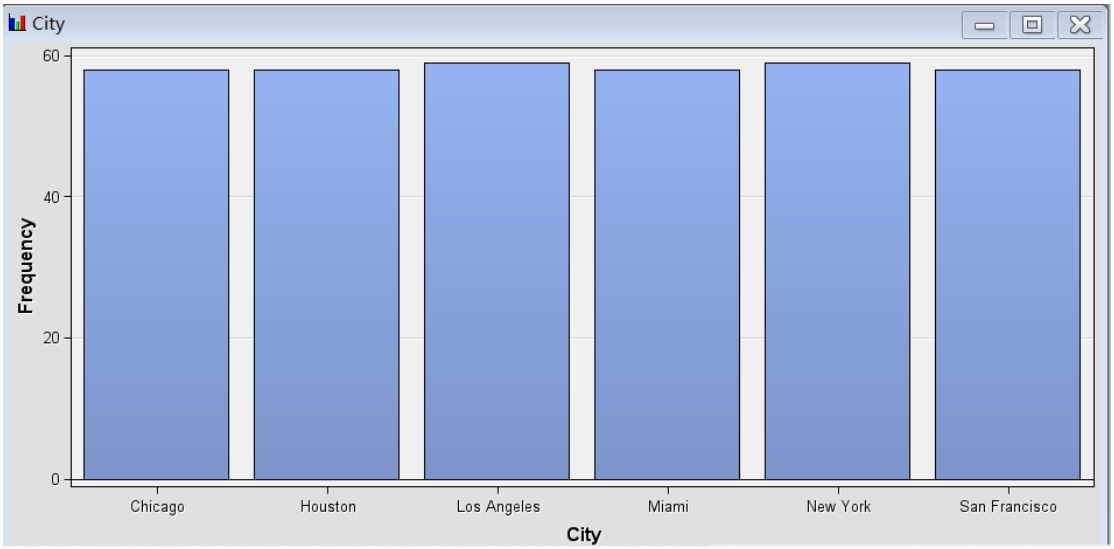
Obs #	Variable Name	Label	Type	Percen...	Minimum	Maximum	Mean	Numbe...	Mode
1	IMP_Satisfaction_Level	Imputed: Satisfaction Level	CLASS	0	.	.	.	3	36.285
2	Average_Rating	Average Rating	VAR	0	3	4.9	4.019143	.	
3	Customer_ID	Customer ID	VAR	0	101	450	275.5	.	



The explore for variable of City:

Since each city appears roughly equally frequently in the data set, this means it makes sense to compare satisfaction levels across these cities as there will be no bias caused by the sample size.

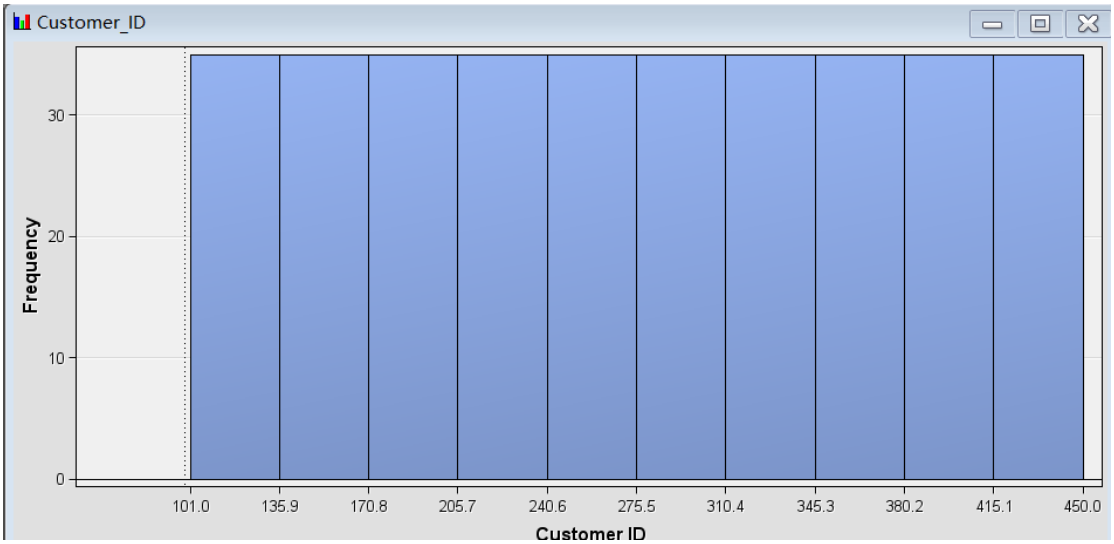
Obs #	Variable Name	Label	Type	Percen...	Minimum	Maximum	Mean	Numbe...	Mode ...	Mode
1	City		CLASS	0	.	.	.	6	16.85714	LOS ANGELES
2	IMP_Satisfaction_Level	Imputed: ...	CLASS	0	.	.	.	3	36.28571	SATISFIED
3	Customer_ID	Custome...	VAR	0	101	450	275.5	.	.	



The explore for variable of Customer ID:

Each customer number occurs very closely, which means the data set may be evenly distributed, or there may be only one record for each customer.

Obs #	Variable Name	Label	Type	Pe...	Minimum	Maximum	Mean	Numbe...	Mode ...	Mode
1	IMP_Satisfaction_Level	Imputed: Satisf...	CLASS	0	.	.	.	3	36.28571	SATISFIED
2	Customer_ID	Customer ID	VAR	0	101	450	275.5	.	.	

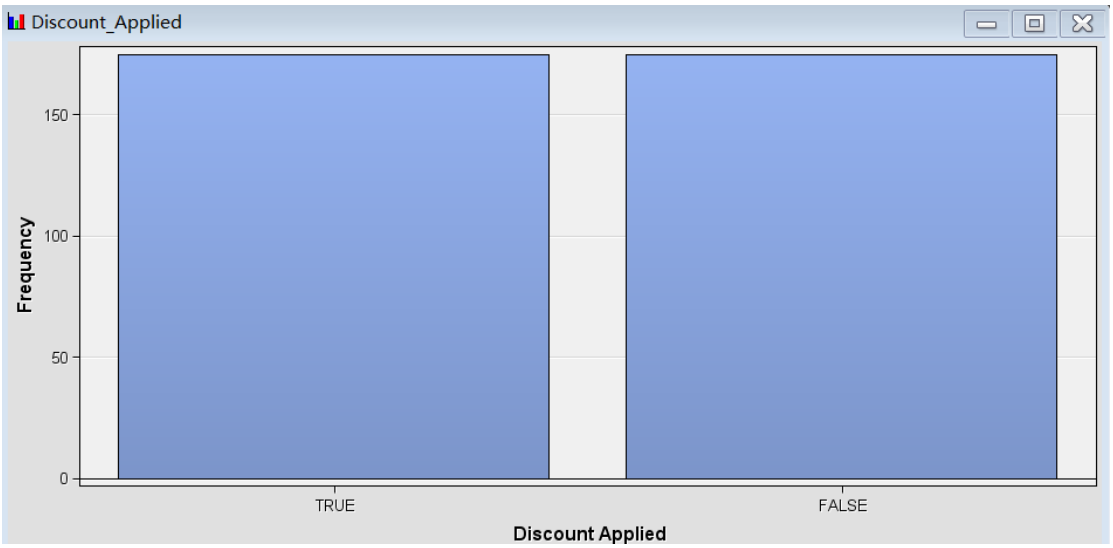


The explore for variable of Days_Since_Last_Purchase:

Obs #	Variable Name	Label	Type	Percen...	Minimum	Maximum	Mean	Numbe...
1	IMP_Satisfaction_Level	Imputed: Satisfaction Level	CLASS	0	.	.		3
2	Customer_ID	Customer ID	VAR	0	101	450	275.5	.
3	Days_Since_Last_Purchase	Days Since Last Purchase	VAR	0	9	63	26.58857	.

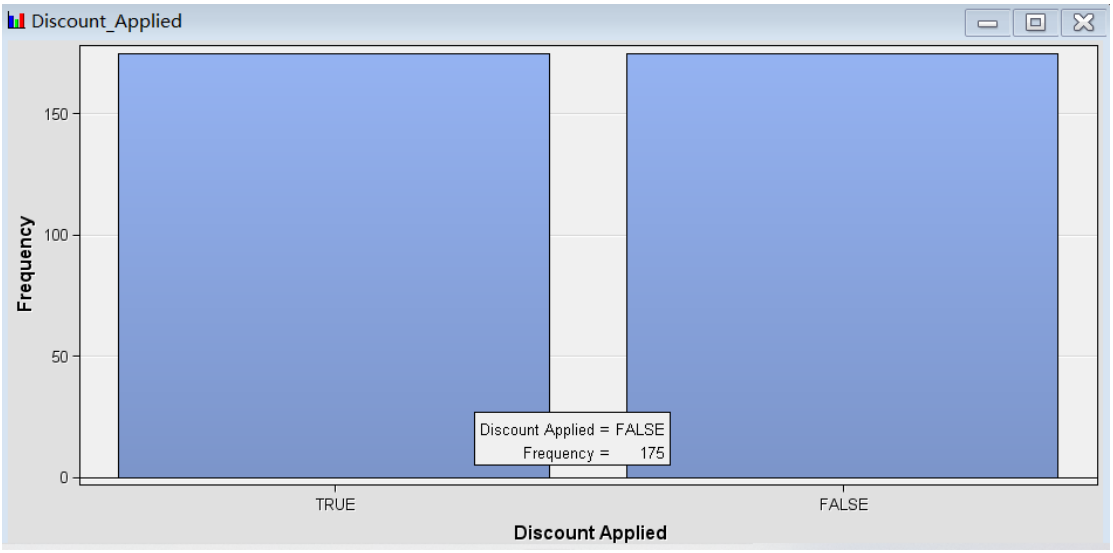
The explore for variable of Days_Since_Last_Purchase:

#	Variable Name	Label	Type	Percen...	Minimum	Maximum	Mean	Numbe...	Mode ...	
1	Discount_Applied	Discount Applied	CLASS	0	.	.		2	50	FA
2	IMP_Satisfaction_Level	Imputed: Satisfaction Level	CLASS	0	.	.		3	36.28571	SA
3	Customer_ID	Customer ID	VAR	0	101	450	275.5	.	.	



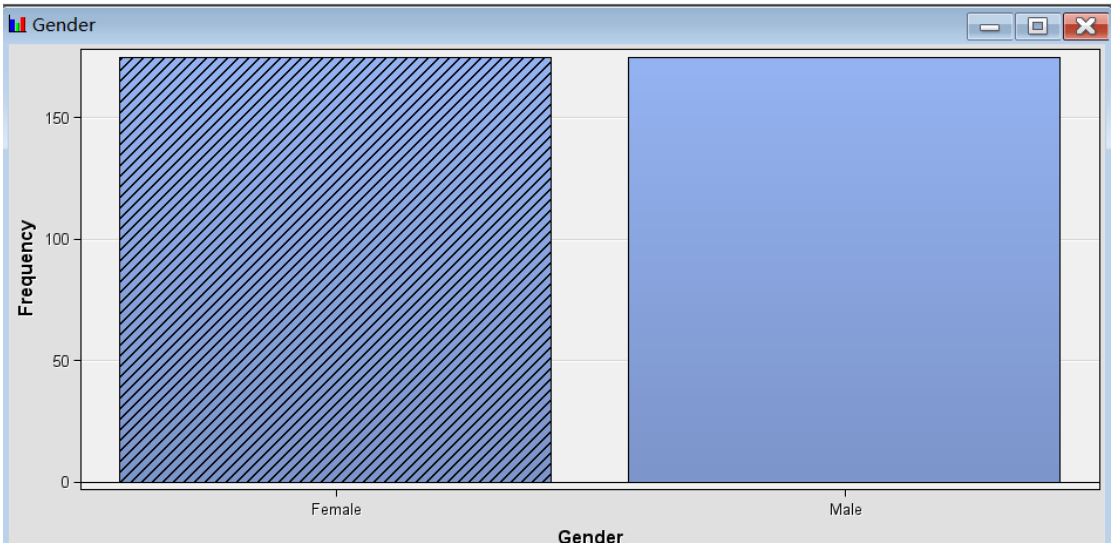
The explore for variable of Discount_Applied:

Obs #	Variable Name	Label	Type	Percen...	Minimum	Maximum	Mean	Numbe...	Mode ...	Mode
1	Discount_Applied	Discount ...	CLASS	0	.	.	.	2	50	FALSE
2	IMP_Satisfaction_Level	Imputed: ...	CLASS	0	.	.	.	3	36.28571	SATISFIED
3	Customer_ID	Customer...	VAR	0	101	450	275.5	.	.	



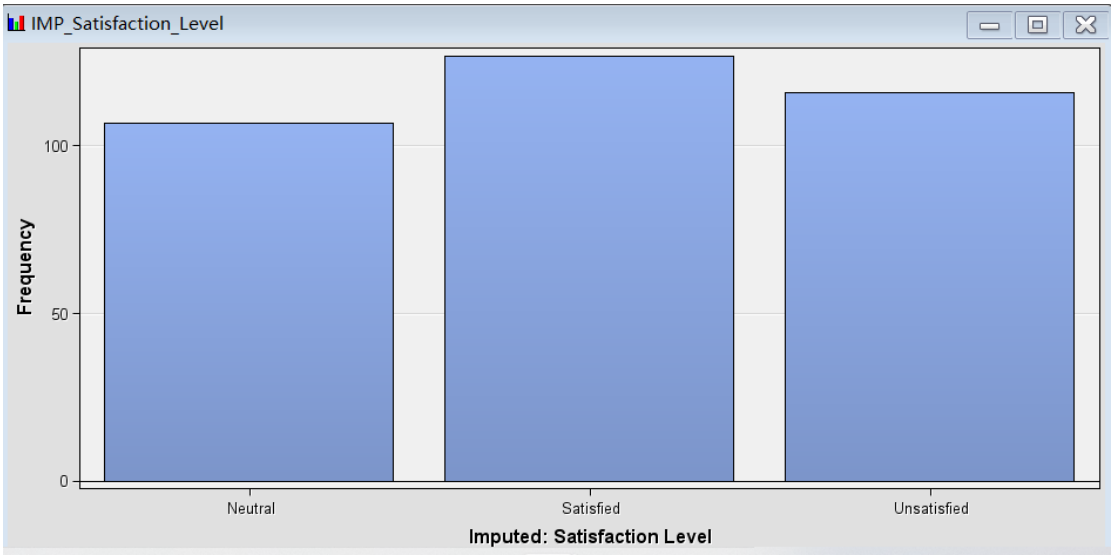
The explore for variable of Gender:

Obs #	Variable Name	Label	Type	Perc...	Minimum	Maximum	Mean	Numbe...	Mod...
1	Gender		CLASS	0	.	.	.	2	50
2	IMP_Satisfaction_Level	Imputed: Satisfaction Level	CLASS	0	.	.	.	3	36.28...
3	Customer_ID	Customer ID	VAR	0	101	450	275.5	.	.



The explore for variable of Satisfication_Level:

Obs #	Variable Name	Label	Type	Percen...	Minimum	Maximum	Mean	Numbe...	Mode ...	Mode
1	IMP_Satisfaction_Level	Imputed: ...	CLASS	0	.	.	.	3	36.28571	SATISFIED
2	Customer_ID	Custome...	VAR	0	101	450	275.5	.	.	

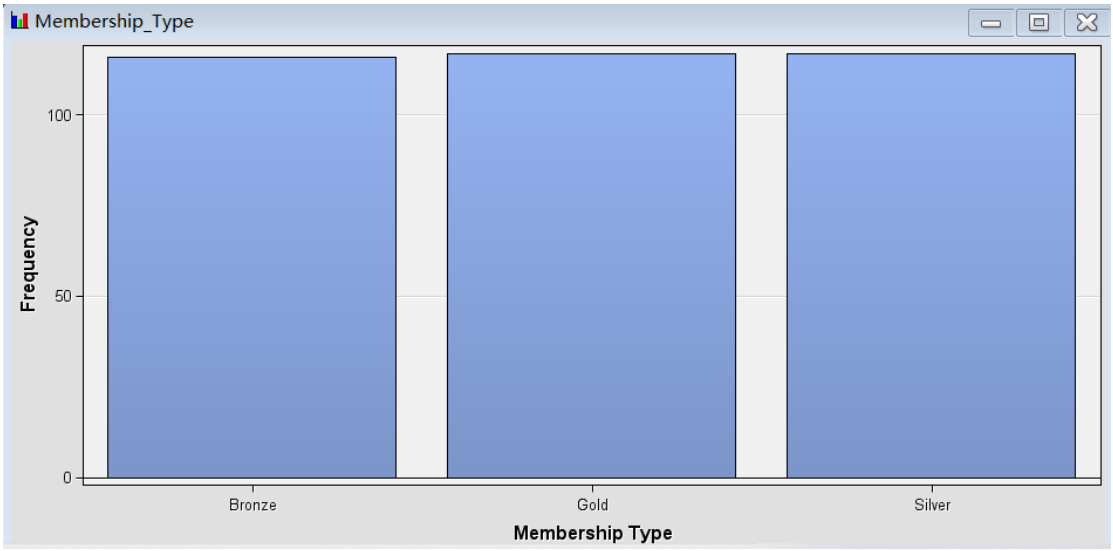


The explore for variable of Items_Purchased:

Obs #	Variable Name	Label	Type	Percen...	Minimum	Maximum	Mean	Numbe...	Mode ...	Mode
1	IMP_Satisfaction_Level	Imputed: S...	CLASS	0	.	.	.	3	36.28571	SATISFIED
2	Customer_ID	Customer ID	VAR	0	101	450	275.5	.	.	
3	Items_Purchased	Items Purc...	VAR	0	7	21	12.6	.	.	

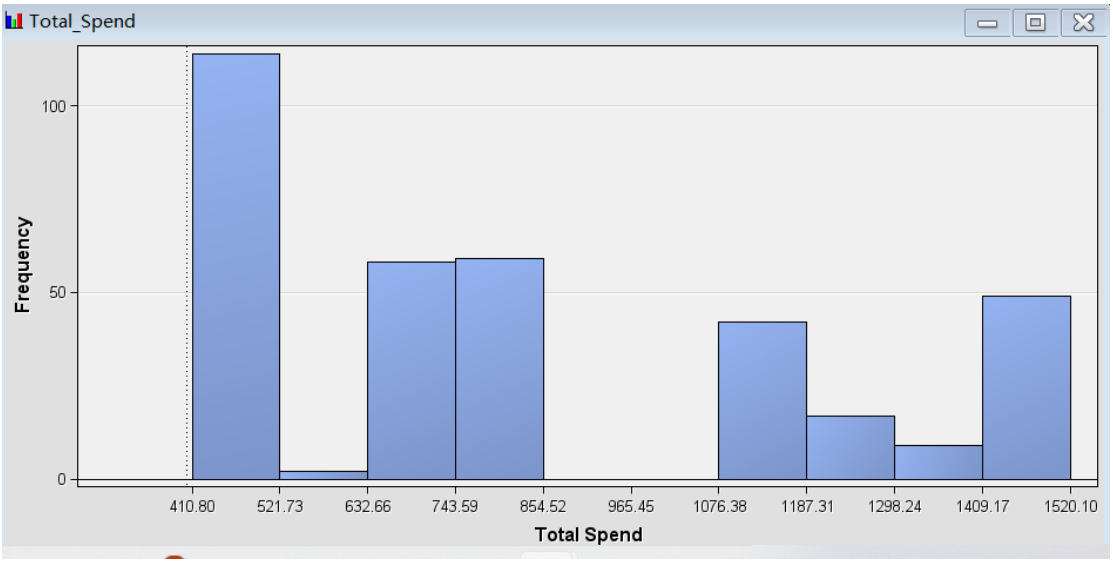
The explore for variable of Membership_Type:

Obs #	Variable Name	Label	Type	Percen...	Minimum	Maximum	Mean	Numbe...	Mode ...	Mode
1	IMP_Satisfaction_Level	Imputed: Satis...	CLASS	0	.	.	.	3	36.28571	SATISFIED
2	Membership_Type	Membership T...	CLASS	0	.	.	.	3	33.42857	GOLD
3	Customer_ID	Customer ID	VAR	0	101	450	275.5	.	.	



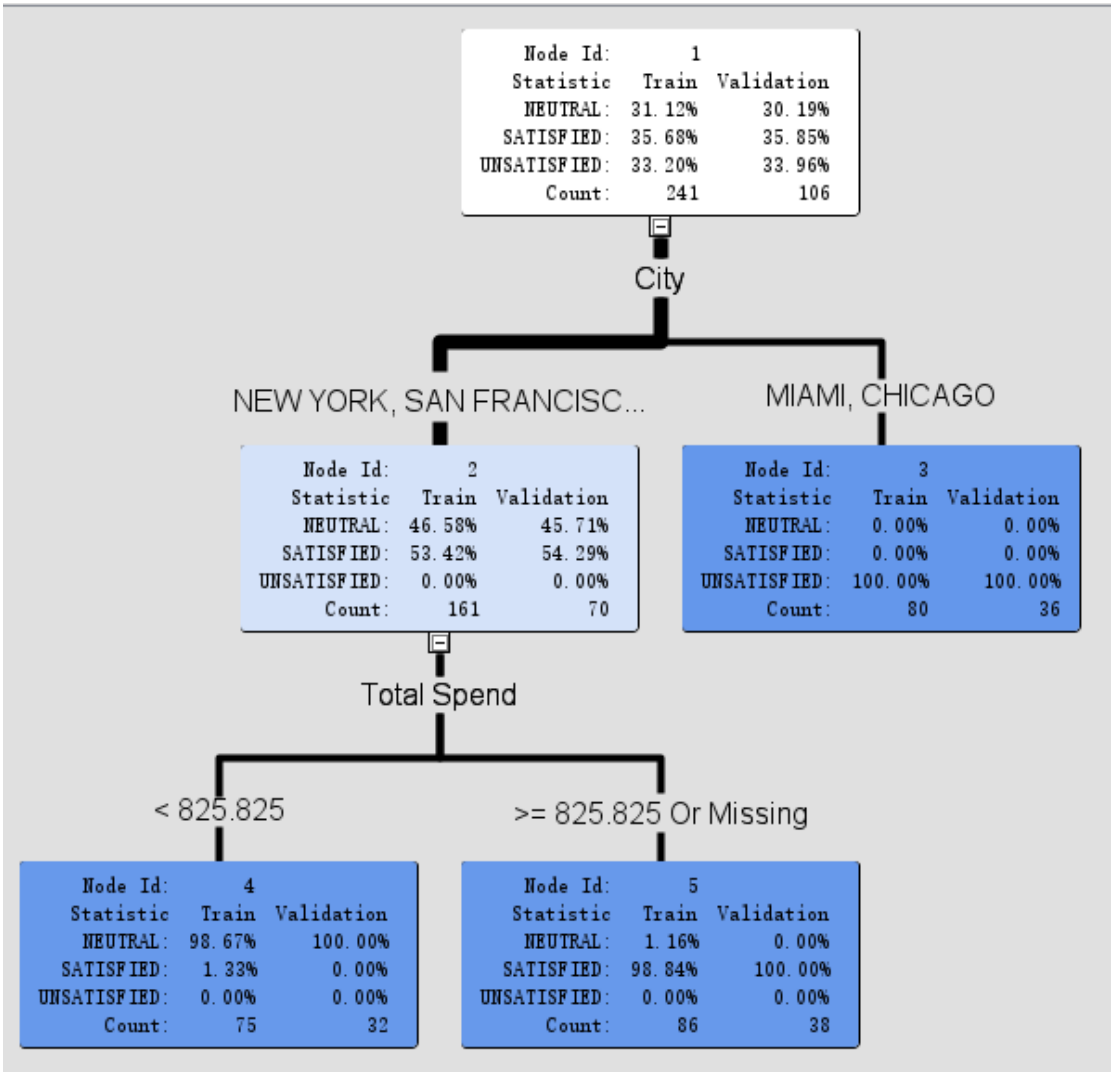
The explore for variable of Total_Spend:

Obs #	Variable Name	Label	Type	Percen...	Minimum	Maximum	Mean	Numbe...	Mode ...
1	IMP_Satisfaction_Level	Imputed: Satisfaction Level	CLASS	0	.	.	.	3	36.28571
2	Customer_ID	Customer ID	VAR	0	101	450	275.5	.	.
3	Total_Spend	Total Spend	VAR	0	410.8	1520.1	845.3817	.	.



Decision Tress:

Run decision tree node, the results are as follows:



Analyze customer behavior:

The impact of city on satisfaction:

The first branch of the decision tree is the city-based variable, which shows that the city where the user is located has a significant impact on their satisfaction. For example, users from Chicago and Miami were completely dissatisfied, while users from New York and Los Angeles were highly satisfied.

The impact of total consumption on satisfaction:

After the city variable, the decision tree considers the total consumption of the user. Users who spend less than 825.825 tend to show extremely high satisfaction or a neutral attitude, while users who spend more than this threshold are completely satisfied.

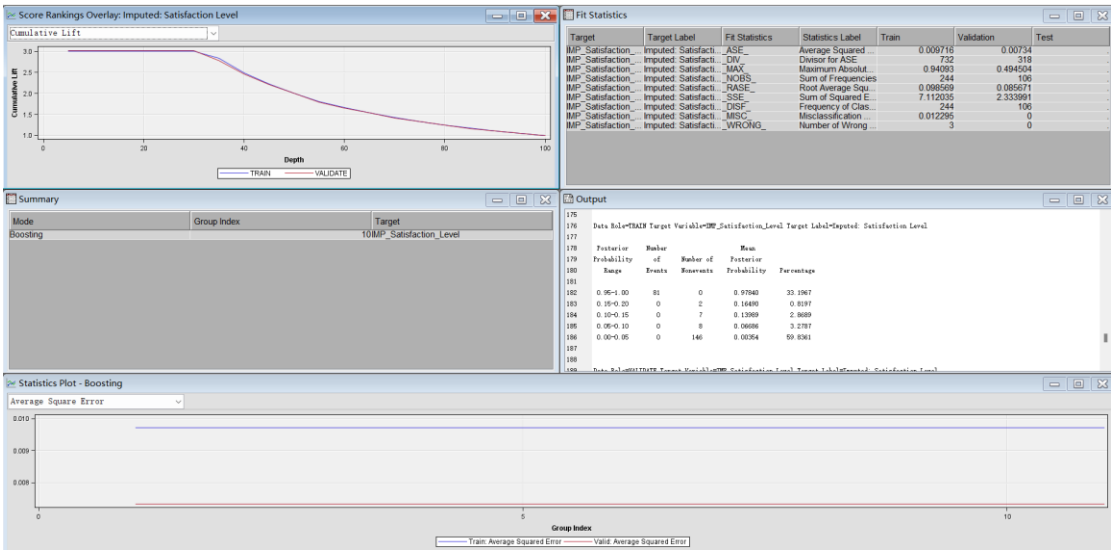
User behavior details:

Among low-consumption users, almost all are satisfied or neutral with the service, which may indicate that price has a great impact on the satisfaction of this group of users.

Ensemble Methods:

Results of Ensemble Methods:

Cumulative Lift Chart: This chart shows that the lifting value decreases as the depth increases, indicating that the previous model processing results are better.
Average Square Error Plot: The two curves in the chart represent the training set (Train) and the validation set (Validation) respectively. Ideally, the two curves should be very close, which means that the model's performance on the unknown data is similar to that on the training data, without overfitting.



The cumulative lift values in the graph decrease with depth, which generally means that the model is more predictive of the front-end percentage of customers (i.e., the customers the model thinks are most likely to be satisfied) and less predictive of the back-end.
The chart shows two curves, representing the average squared error on the training set (Train) and the validation set (Validation). The two curves are very close and the error is

low, which shows that the model's prediction of customer satisfaction maintains consistency on new data and shows good generalization ability without over-fitting problems.