E-COMMERCE PLATFORM CUSTOMER SATISFACTION:

MARKETING ANALYTICS REPORT

Report by: Vidhi Desai

Executive Summary:

In a recent quarterly analysis, an ecommerce platform realized that very few customers are satisfied with the products and the services associated with it. The company decided to run Marketing analytics with a primary objective of determining which factors significantly influence customer satisfaction. By understanding the variables and their relationships with satisfaction, the company can tailor its strategies to better meet customer needs and expectations.

The dependent variable was customer satisfaction, while the independent variables included age, gender, purchase amount, subscription status, shipping type, and discount application. A logistic regression analysis was employed as an analytical tool to assess the impact of various factors on customer satisfaction. The key findings indicate that factors such as age, gender, purchase amount, subscription status, shipping type, and discount application significantly influence customer satisfaction. Notably, applying discounts has the most substantial impact, followed by gender (with female customers being more likely to be satisfied), choosing express shipping and age.

The business implications of the findings suggest focusing on leveraging discounts, targeting female customers, promoting express shipping, and catering to older customers through specific offers and advertisements. By addressing these areas, the company can enhance customer satisfaction, leading to improved retention and acquisition rates.

Introduction:

- **Business Issue:**
  An e-commerce platform is trying to find out if there is any relationship between various controllable (Purchase amount, discount offers, shipping service, etc.) and uncontrollable (Age, gender, season of purchase, etc.) factors and customer satisfaction, as indicated by ratings from 1 to 5.
- **Analysis Goal:**
  Identifying factors that affect customer satisfaction so that the company's strategies and offerings can be tailored effectively, ultimately leading to improved customer experiences and loyalty.
- **The significance of the Analysis:**
  Some of the variables may be more significant than others in building the customer ratings. Understanding the key drivers of satisfaction is crucial for the firm's success in a highly competitive market.

Data:

- The company has collected data of **3900 customers** who have provided the review ratings which can be utilized to understand the **customer satisfaction level**.
- It is supported by the additional data like **customer profile** (Age, gender, subscription status) and **purchase specifics** (Purchase amount, discount application, shipping type).

Here is a screenshot of the data that is used for analysis:

CustomerID	Age	Gender	Female	Purchase_USD	Customer_Rating	Satisfaction	Satisfaction_Dummy	Subscription_Status	Taken_Subscription	ShippingType	DiscountApplied
1	55	Male	0	53	3.1	Dissatisfied	0	Yes	1	Express	Yes
2	19	Male	0	64	3.1	Dissatisfied	0	Yes	1	Express	Yes
3	50	Male	0	73	3.1	Dissatisfied	0	Yes	1	Free Shipping	Yes
4	21	Male	0	90	3.5	Satisfied	1	Yes	1	Next Day Air	Yes
5	45	Female	1	49	2.7	Dissatisfied	0	Yes	1	Free Shipping	Yes
6	46	Male	0	20	2.9	Dissatisfied	0	Yes	1	Standard	Yes
7	63	Male	0	85	3.2	Dissatisfied	0	Yes	1	Free Shipping	Yes
8	27	Male	0	34	3.2	Dissatisfied	0	Yes	1	Free Shipping	Yes
9	26	Male	0	97	2.6	Dissatisfied	0	Yes	1	Express	Yes
10	57	Male	0	31	4.8	Satisfied	1	Yes	1	2-Day Shipping	Yes
11	53	Female	1	34	4.1	Satisfied	1	Yes	1	Store Pickup	Yes
12	30	Male	0	68	4.9	Satisfied	1	Yes	1	Store Pickup	Yes
13	61	Male	0	72	4.5	Satisfied	1	Yes	1	Express	Yes
14	65	Male	0	51	4.7	Satisfied	1	Yes	1	Express	Yes
15	64	Male	0	53	4.7	Satisfied	1	Yes	1	Free Shipping	Yes
16	64	Male	0	81	2.8	Dissatisfied	0	Yes	1	Store Pickup	Yes
17	25	Female	1	36	4.1	Satisfied	1	Yes	1	Next Day Air	Yes
18	53	Male	0	38	4.7	Satisfied	1	Yes	1	2-Day Shipping	Yes

Y Variable – Dependent Variable:

Customer Satisfaction:

- The customer has rated the products and overall purchasing experience from 0 to 5, with 0 being the lowest and 5 being the highest.
- Based on the ratings, the customer satisfaction has been classified as Satisfied (More than 4 ratings) and Neutral or Dissatisfied (Less than or equal to 4 ratings)

X Variables - Determinants:

- Demographic – Age, gender
- Purchase Specifics – Purchase amount, Discount application, Type of shipping, Payment method
- Others – Subscription status, Season of purchase

Data Directory: (D = Discrete and C = Continuous)

Variable	Definition	Type
Age	Customer's age in completed years	C
Female (Gender)	Customer's gender; 1 = Female, 0 = Male	D
Purchase_USD	Total amount of Purchase (In USD)	C
Applied_Discount	Whether discount or promocode was applied or not. 1 = Yes, 0 = No	D
Express_Shipping	Type of shipping used for delivering the order; 1 = Express, 0 = Standard	D
Taken_Subscription	Whether a customer has taken a subscription or not; 1 = Yes, 0 = No	D
Satisfaction	Whether a customer rating fits into the satisfaction criteria or not; 1 = Satisfied, 0 = Neutral or unsatisfied	D

Note: All the discrete variables have been converted to dummy variables by assigning the value 1 or 0 to indicate the presence or absence of some categorical effect that may be expected to shift the outcome.

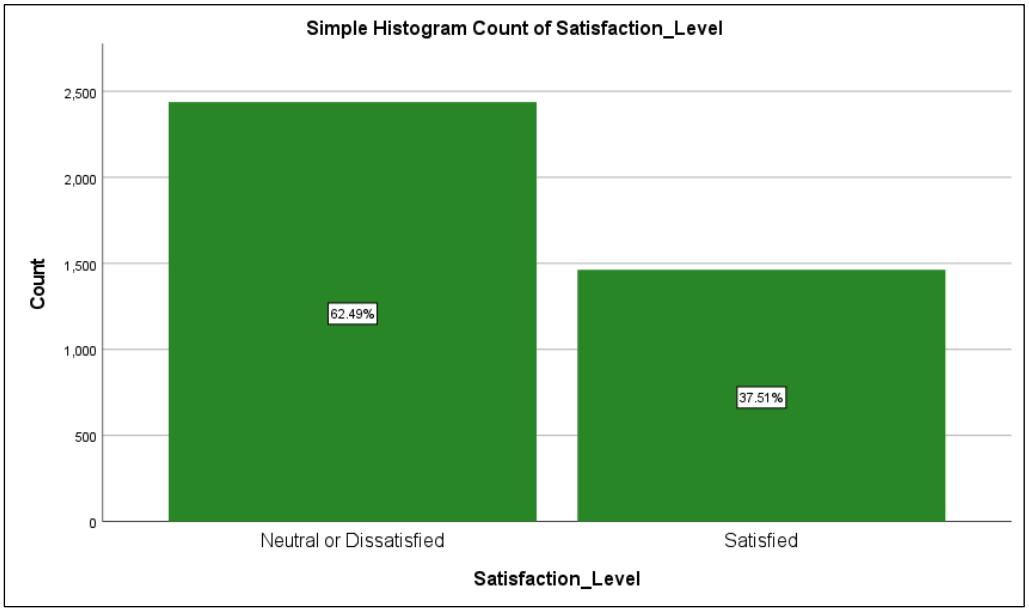
Descriptive Statistics (Discrete Variables):

In this section, we will interpret the descriptive statistics by displaying the frequency tables and the histogram showing how each X variable affects the Y variable (Customer ratings) individually.

1. Satisfaction Level: 0 = No, 1 = Yes

Satisfaction

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	2437	62.5	62.5	62.5
	1	1463	37.5	37.5	100.0
	Total	3900	100.0	100.0	



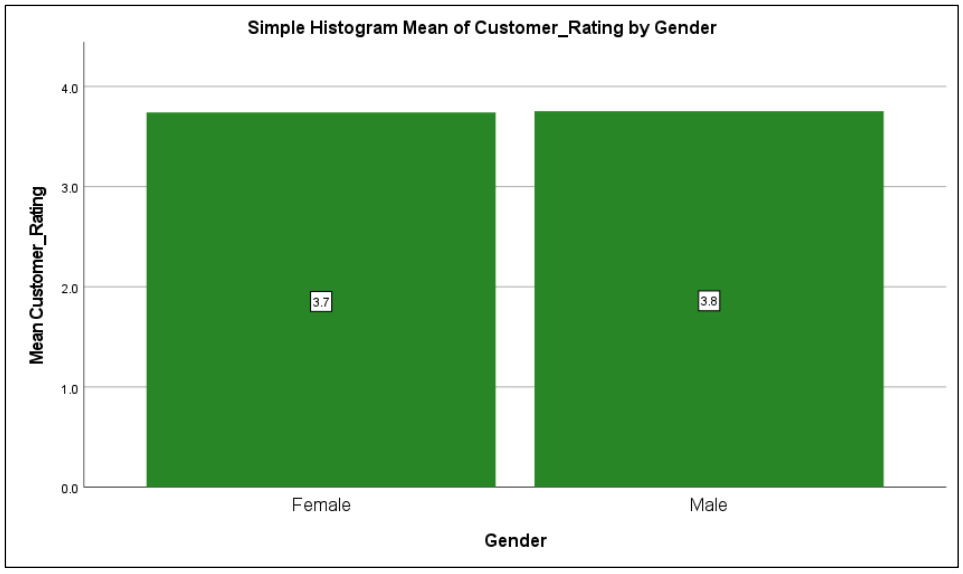
- As the frequency table and the histogram show, only 37.5% of the total customers i.e. 1463 people are satisfied, while other 62.5% i.e. 2437 people are neutral or dissatisfied.

2. Gender: 0 = Male, 1 = Female

Female

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	2652	68.0	68.0	68.0
	1	1248	32.0	32.0	100.0
	Total	3900	100.0	100.0	

- The majority of the respondents are male, constituting 68% of the total customers, while only 32% customers are Female.



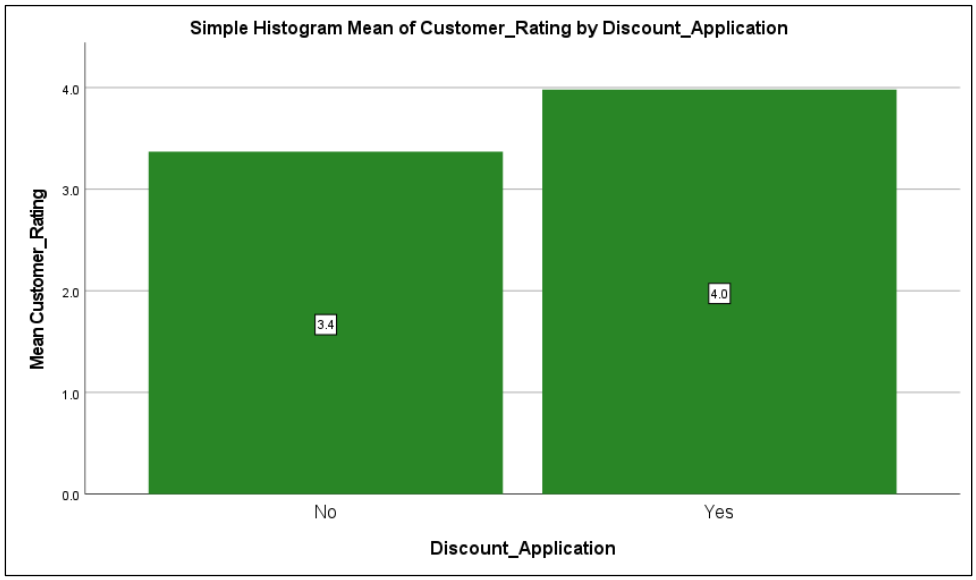
- The histogram shows the average customer ratings (1 to 5) by gender. It appears that the average customer ratings are almost equal for both Male and Female.

3. Discount Application: 0 = No, 1 = Yes

Applied_Discount

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	1475	37.8	37.8	37.8
	1	2425	62.2	62.2	100.0
	Total	3900	100.0	100.0	

- The data suggests that offering discounts is a prevalent practice, since approximately 62% of transactions involve discount applications. There are still 1475 out of 3900 customers (Approx 38%) who made purchases without applying the discounts or promo offers.



- The average customer ratings by application of discounts suggest that customers who applied discounts are more satisfied than those who did not.

4. Shipping Type: 0 = Standard, 1 = Express

Express_Shipping

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	1938	49.7	49.7	49.7
	1	1962	50.3	50.3	100.0
	Total	3900	100.0	100.0	

- It appears that the Shipping type is balanced between both the modes, with Express shipping (50.3%) weighing a little more than Standard shipping (49.7%)

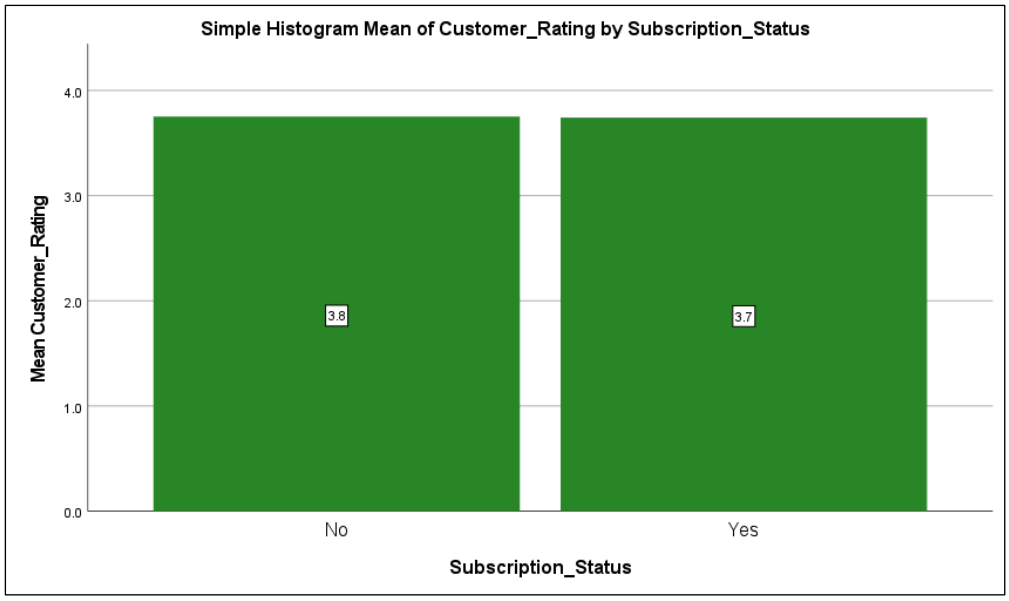
- It is evident from the histogram that the customers who opted for Express shipping for the product delivery tend to be more satisfied than the ones who chose standard shipping.

5. Subscription Status: 0 = No, 1 = Yes

Taken_Subscription

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	2847	73.0	73.0	73.0
	1	1053	27.0	27.0	100.0
	Total	3900	100.0	100.0	

- From the total of 3900 customers, only 1053 i.e. 27% have taken the online subscription of the e-commerce platform. Rest 73% do not have the subscription plan.



- The above histogram indicates that the customers who have taken the platform subscription are less satisfied than those who does not have the subscription.

Descriptive Statistics (Continuous Variables):

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Age	3900	18	70	35.60	16.600
Purchase_USD	3900	20	100	59.76	23.685
Valid N (listwise)	3900				

- The ages of customers range from 18 to 70 years, with an average age of approximately 36 years, indicating a fairly wide age distribution.
- The customers have made purchases worth as high as $100 to as low as $20. The average purchase amount is $59.76.

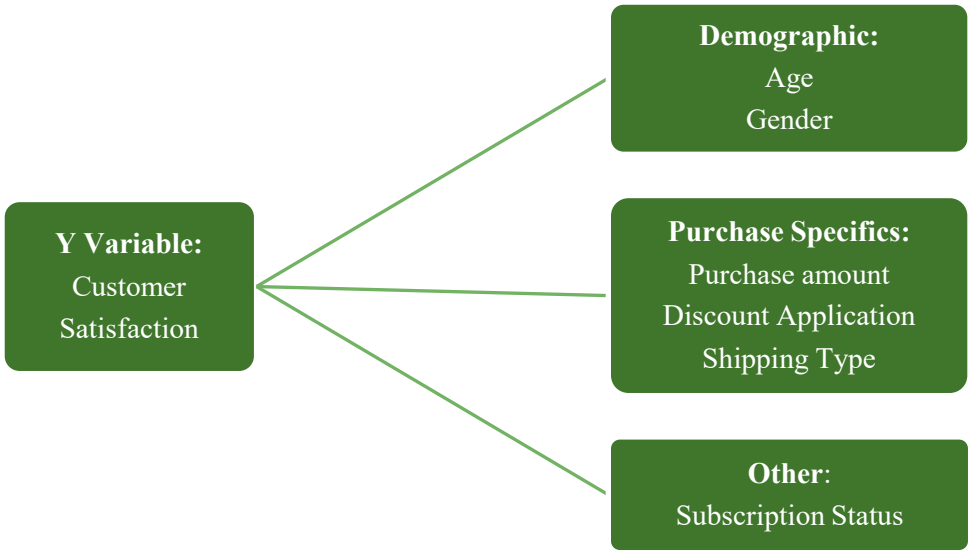
Methodology for Analysis:

Y**: Variable of Interest: Customer Satisfaction:* ***Discrete
X**: Determinants: Mainly* ***continuous *or discrete which can be converted to dummy variables*

Because of the above scenario, the ideal methodology for analyzing the relationship will be **Logistic Regression**, which calculates the probability of the outcome or event and classify the observations into one of the outcome groups as Satisfied or Neutral/Dissatisfied customers (denoted by 0 or 1).

Conceptual Model:

Based on the factors that are significant and strongly affect the Y variable, we can present the conceptual model in a following way:



Logistic Regression Model:

$$ln\frac{Prob(Personal\ Loan=1)}{1-Prob(Personal\ Loan=1)} = b0 + b1*Age + b2*Female\ (Gender) + b3*Purchase\ amount + b4*Subscription\ taken + b5*Express\ Shipping + b6*Discount\ application$$

Logistic Regression Equation:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Age	.049	.003	328.276	1	<.001	1.051
	Female	1.433	.155	85.852	1	<.001	4.193
	Purchase_USD	.005	.002	8.209	1	.004	1.005
	Taken_Subscription	-1.053	.100	110.884	1	<.001	.349
	Express_Shipping	.540	.090	36.015	1	<.001	1.716
	Applied_Discount	4.026	.166	587.002	1	<.001	56.057
	Constant	-5.979	.237	634.619	1	<.001	.003

a. Variable(s) entered on step 1: Age, Female, Purchase_USD, Taken_Subscription, Express_Shipping, Applied_Discount.

Using the given values in the above table, we can write the Logistic Regression equation:

$$ln\frac{Prob(Personal\ Loan=1)}{1-Prob(Personal\ Loan=1)} = -5.979 + 0.049*Age + 1.433*Female\ (Gender) + 0.005*Purchase\ amount - 1.053*Subscription\ taken + 0.540*Express\ Shipping + 4.026*Discount\ application$$

Significance and Interpretation of the Variables:

- Null Hypothesis: H0: The particular X variable does not affect the customer satisfaction.
- Alternative Hypothesis: H1: The particular X variable affects the customer satisfaction.

- The significance for all the chosen factors is derived using p-value < significance 0.05.
  If p-value < 0.05, reject the null hypothesis and hence the effect on Y is significant.
  If p-value > 0.05, accept the null hypothesis and hence the effect on Y is not significant.

- The variables can be interpreted using the signs of Beta(B) and values of the x. e^b i.e., Exp(B) is the multiplicative factor by which the odds of belonging to class 1 increase when the value of x is increased by one unit.

Variable	Significance	Interpretation
Age	Significant	If age increases by one unit, the log odds of customer being satisfied increases by a multiplicative factor of 1.051.
Female (Gender)	Significant	The log odds that a female customer is satisfied relative to a male customer is 4.193.
Purchase Amount	Significant	If purchase amount increases by one USD, the log odds of customer being satisfied increases by a multiplicative factor of 1.005.
Subscription Taken	Significant	If a customer has taken subscription, it is less likely that he/she will be satisfied relative to a customer who has not taken subscription.
Express Shipping	Significant	The odds that a customer who has chosen express shipping will be satisfied relative to a customer who has chosen standard shipping is 1.716.
Discount Application	Significant	The odds that a customer who has applied a discount will be satisfied relative to a customer who has not applied a discount is 56.057.

Importance of the Variables:

Based on the above table (Considering Exp(B) values as absolute values), we can arrange the X variables in descending order of importance (Exponentiation of the coefficients B):

> Applied_Discount (56.057) > Female (4.193) > Express_Shipping (1.716) > Age (1.051) > Purchase_USD (1.005) > Taken_Subscription (0.349)

From this, we can say that Discount application, gender, shipping type and age are the most important factors in determining the customer satisfaction.

Model Fit:

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	3160.134[a]	.401	.547

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

- The above table provides the values of two types of R-square – Cox & Snell R^2 and Nagelkerke R^2. The R^2 values for this model are 0.401 and 0.547 i.e., 40.1% and 54.7%, which is relatively low. Higher the coefficient, more accurately the model can explain the relationship between X and Y.
- Based on the R^2 values, we can say that 40.1% (Cox & Snell $R^{2)}$ and 54.7% (Nagelkerke $R^{2)}$ variance in satisfaction is explained by variance in the chosen X variables.

Classification Table (Actual v/s Predicted results):

Satisfaction * Predicted group Crosstabulation

			Predicted group 0	Predicted group 1	Total
Satisfaction	0	Count	1950	487	2437
		% within Satisfaction	80.0%	20.0%	100.0%
	1	Count	226	1237	1463
		% within Satisfaction	15.4%	84.6%	100.0%
Total		Count	2176	1724	3900
		% within Satisfaction	55.8%	44.2%	100.0%

Note:
The cut-off value 0.37 yielded the best results instead of the default value of 0.50, and it is close to the probability of a customer being satisfied i.e., 0.375 (1463 out of 3900). Hence, 0.37 has been used as a cut-off value here. By lowering the cut-off value, the model can effectively be made more sensitive in classifying observations correctly.

Interpretation:

- To determine whether the model predicts the satisfaction level accurately, this crosstabulation compares the predicted value to the actual value.
- The values highlighted in green are the correctly predicted values, while the ones highlighted in orange are the incorrectly predicted values.
- Out of 1463 customers who were satisfied, 1237 are predicted correctly, which makes up 84.6% of the total customers.
- From 2437 neutral or dissatisfied customers, only 20% i.e., 487 people were misclassified, all the other 1950 people were predicted correctly.

Hit Ratio = Sum of values predicted correctly / Sum of predicted values
= (1950+1237)/3900
= 0.797 ~ **79.7%**

The Hit ratio implies that 79.7% of the instances in the dataset were correctly classified by the logistic regression model. This suggests that the model could accurately predict the outcomes for a majority of the instances. Accordingly, we can conclude that this model is moderately fit for the test, but it can be improved.

Business Implications:

According to the analysis, most customers will be satisfied if we focus on the top affecting factors. The business should focus on improving the customer ratings and ultimately satisfaction level strategically.

The combination of targeted marketing, improved shipping options, and attractive discounts will likely lead to higher satisfaction levels among different customer segments. It will eventually result in a better retention rate of existing customers as well as a higher rate of customer acquisition.

The following **promotional strategies** can be charted out for getting the desired results:

1. **Leverage Discounts:**
   - Develop *loyalty programs* that provide regular discounts or special offers to repeat customers.
   - Offer *limited-time discounts* to create sense of urgency and increase satisfaction and drive sales.

2. **Focus on Female Customers:**
   - *Offer special discounts*, exclusive products, or services catering to women's preferences.
   - *Enhance engagement with female customers* through social media, personalized emails, and events.

3. **Promote Express Shipping:**
   - Provide *free or discounted express shipping* for orders over a certain amount to encourage its use.
   - Offer *complementary products* for customers who opts for express shipping.

4. **Target Older Customers:**
   - *Billboard advertisements* highlighting products and services that are likely to appeal to the older customers.
   - Offer *special discounts* for senior customers.

Conclusion:

The logistic regression analysis has provided valuable insights into the factors influencing customer satisfaction. By leveraging these insights, the company can refine its strategies to better meet customer expectations, ultimately improving customer experiences and fostering greater loyalty.