Name- Shatabarta Chanda

Registration No.- 2024JULB00140

Problem Statement-

**“High Customer Churn Rate”**

**Phase 1: Dirty Data Assessment:-**

|  |  |
| --- | --- |
| Column Name | Description |
| Customer ID | Unique identifier for each customer. |
| Name | Full name of the customer. |
| Age | Age in years. |
| Gender | Gender of the customer. |
| Location | |  | | --- | |  |  |  | | --- | | City or town of residence. | |
| Signup date | |  | | --- | |  |  |  | | --- | | Date of signing up for the service. | |
| Last Interaction Date | |  | | --- | |  |  |  | | --- | | Date of last customer interaction. | |
| Service Used | Type of plan: Basic, Standard, or Premium. |
| Total Spent | |  | | --- | |  |  |  | | --- | | Total amount spent by customer. | |
| Support Tickets Raised | Number of customer service interactions. |
| Satisfaction Score | Rating from 1 (low) to 10 (high). |
| Engagement Score | |  |  |  | | --- | --- | --- | | |  | | --- | |  |  |  | | --- | | Internal score from 0–100 showing customer activity. | |  |  | | --- | |  | |
| Preferred Channel | |  | | --- | |  |  |  | | --- | | Customer’s preferred way of communication. | |
| Subscription Status | |  | | --- | |  |  |  | | --- | | ‘Active’ or ‘Cancelled’. | |
| Churned | 1 = Yes (churned), 0 = No (retained). |

**Phase 2: Data Visualization & Insights:-**

1. Customer Churn Distribution

What it shows:  
A simple count of customers who churned (1) versus those who retained (0).

Insight:  
Churned customers are fairly balanced with retained ones — a warning sign for further analysis.

2. Satisfaction Score by Churn Status

What it shows:  
Box plot comparing satisfaction levels between churned and retained customers.

Insight:  
Churned customers tend to have lower satisfaction scores, indicating a clear link between dissatisfaction and churn.

3. Average Engagement Score by Churn Status

What it shows:  
Bar plot of average engagement scores for churned vs retained.

Insight:  
Engagement is significantly lower among churned users, showing the need for better retention strategies for low-engagement users.

|  |  |  |  |
| --- | --- | --- | --- |
| Metric | Churned (1) | Retained (0) | Interpretation |
| Avg. Satisfaction | Lower | Higher | Dissatisfied users are more likely to churn. |
| Engagement Score | Lower | Higher | Disengaged customers are more prone to leave. |
| Total Spent | Slightly lower | Slightly higher | High-value customers tend to stay longer. |

**`Analyzing Conclusions`**

“Insights from Customer Churn Analysis.”

**Data Cleaning:-**

- Missing Values: Missing values in the 'Gender' column were imputed using the mode, and rows with any remaining missing values were dropped.

- Outliers: Outliers in 'Total\_Spent', 'Support\_Tickets\_Raised', 'Satisfaction\_Score', and 'Engagement\_Score' were removed using the IQR method.

- Data Types: Ensured numeric columns ('Age', 'Total\_Spent', 'Support\_Tickets\_Raised', 'Satisfaction\_Score', 'Engagement\_Score', 'Churned') were of numeric type, coercing errors to NaN and dropping resulting NaN rows.

**Feature Engineering:-**

- Date Conversion: 'Signup\_Date' and 'Last\_Interaction\_Date' were converted to datetime objects.

- Time-Based Features:

- 'Time\_Since\_Signup' was calculated as the number of days between a fixed current time and the signup date.

- 'Time\_Since\_Last\_Interaction' was calculated as the number of days between the current time and the last interaction date.

- Interaction Metrics:

- 'Interaction\_Frequency' was calculated as the number of support tickets raised per day since signup.

- 'Spending\_Per\_Interaction' was calculated as the total spent divided by the number of support tickets raised plus one.

- Binary Feature: 'Has\_Support\_Tickets' was created to indicate whether a customer raised any support tickets.

**Statistical and Correlation Analysis:-**

- Descriptive Statistics: Descriptive statistics were computed, providing measures of central tendency and dispersion for numerical features.

- Churn Correlation: The correlation matrix was computed to identify the correlation of each feature with the 'Churned' status.

**Data Visualization:-**

- Churn Correlation vs. Descriptive Statistics: A scatter plot visualizes the relationship between descriptive statistics of various features and their correlation with churn.



**Conclusion and Insights:-**

- Key Churn Drivers: Based on the correlation analysis, the features most correlated with churn are identified.

- Support\_Tickets\_Raised: Has a positive correlation (0.165) with churn, suggesting that customers who raise more support tickets are more likely to churn.

- Interaction\_Frequency: Shows a positive correlation (0.143) with churn, indicating that a higher interaction frequency might be associated with increased churn.

- Satisfaction\_Score: Displays a negative correlation (-0.115) with churn, implying that lower satisfaction scores are linked to higher churn rates.

- Total\_Spent: Exhibits a negative correlation (-0.113) with churn, suggesting that customers with lower total spending are more prone to churn.

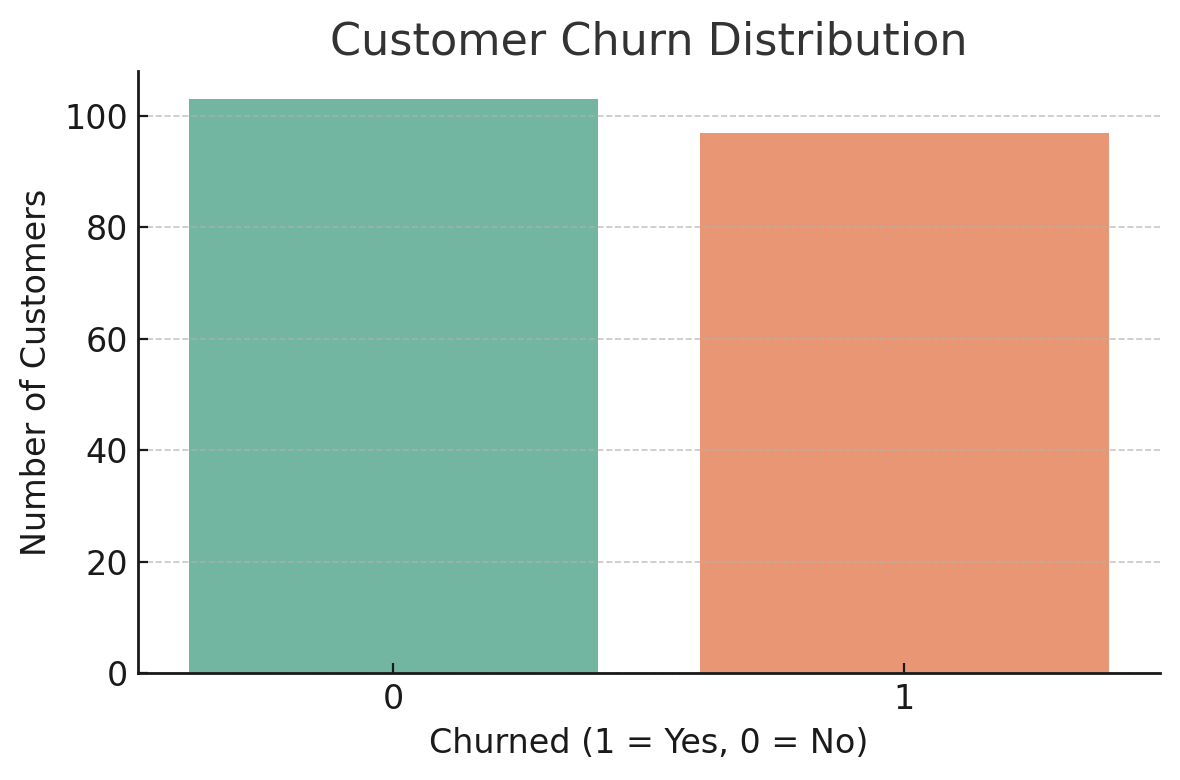
- Descriptive Statistics: The descriptive statistics provide a summary of the central tendency of the dataset.

- Age: The average age of customers is approximately 42.66 years.

- Engagement\_Score: The average engagement score is around 50.13.

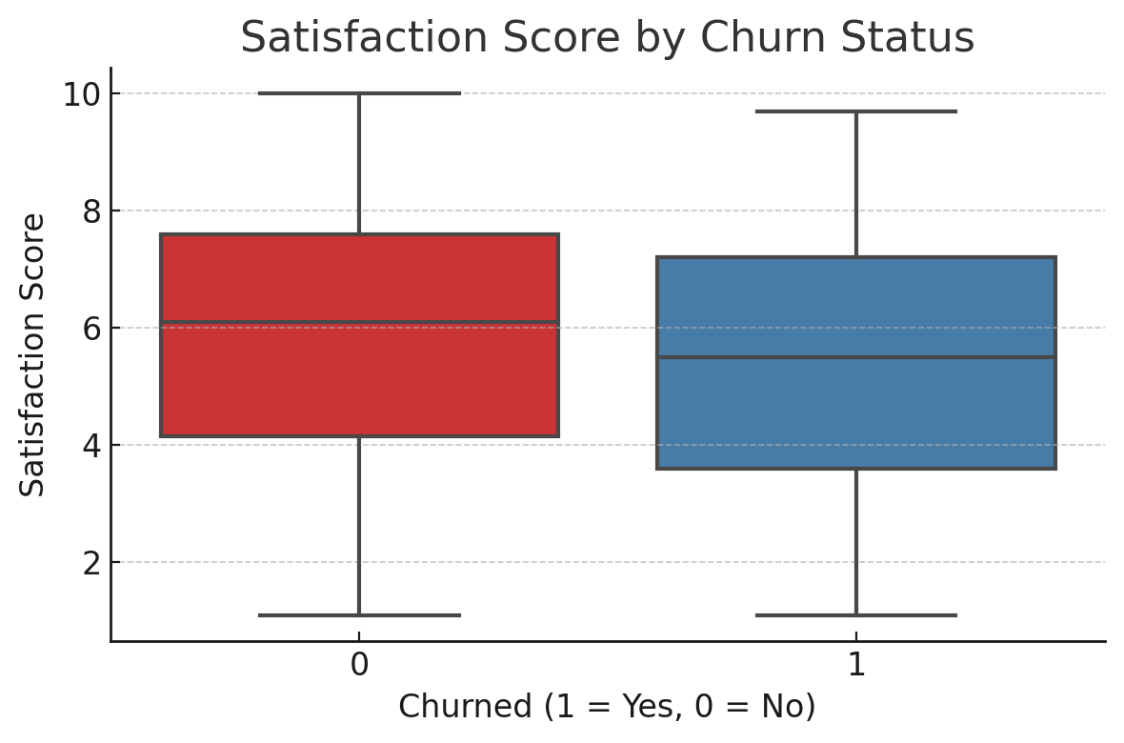
- Time\_Since\_Signup: Customers have been signed up for an average of about 1039.79 days.

- Time\_Since\_Last\_Interaction: The average time since the last interaction is approximately 180.38 days.



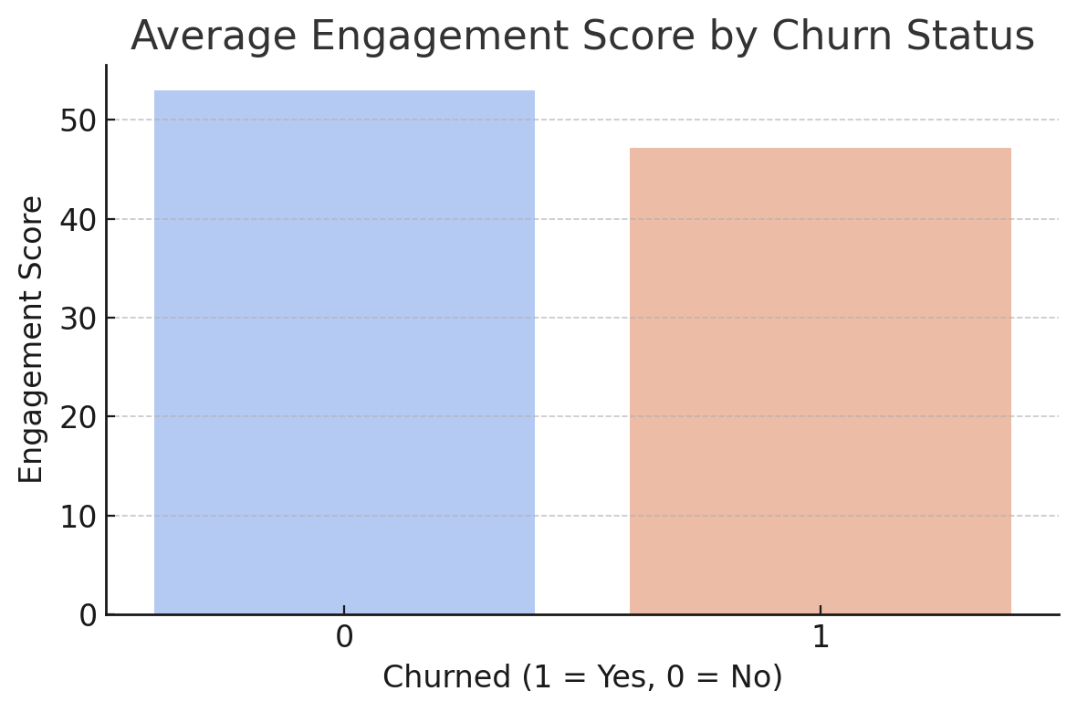
**Customer Churn Distribution:-**

* This bar chart shows a fairly even split between customers who **churned** (1) and those who **retained** (0).
* Interpretation: A **high churn rate** signals an urgent need to investigate customer dissatisfaction and engagement levels.



**Satisfaction Score by Churn Status:-**

* This box plot shows that customers who **churned** have **lower satisfaction scores**, often below the median for retained customers.
* Interpretation: Improving satisfaction could significantly reduce churn.



**Average Engagement Score by Churn Status:-**

* This bar graph clearly shows **higher engagement among retained customers**.
* Interpretation: Keeping customers **actively engaged** is key to **reducing churn**.