Customer Churn Prediction Analysis Report

1. Statistical Analysis & Insights

This analysis investigates customer churn behavior using feature transformations, statistical tests, and dimensionality reduction techniques.

Feature Transformation Summary:

- Label Encoding: Applied to ordinal categorical variables to convert them into numerical values whilepreserving order.

- One-Hot Encoding: Used for nominal categorical variables to create binary columns, ensuring the modeldoesn't assume any ordinal relationship.

- Sin-Cos Transformation: Applied to cyclic features (like day/month) to preserve continuity in periodic data.- MinMax Scaling: Normalized numerical values to the [0,1] range to ensure consistency in magnitude across features.

Dimensionality Reduction:

- PCA (Principal Component Analysis):

- Reduced feature dimensionality from original size to 10 components.

- Retained over 95% of the original variance, ensuring minimal information loss.

- Helped in visualizing clusters and identifying key feature patterns in churn behavior.

Statistical Testing:

- T-Test: Compared means of a continuous variable across two groups (churned vs. retained). Significantdifferences indicate feature relevance.

- Chi-Square Test: Assessed independence between categorical features and churn outcome. Highchi-square values imply stronger associations.

- ANOVA (F-One Way Test): Evaluated mean differences across more than two groups (e.g., customersegments). Used to detect variation in churn patterns among multiple categories.

2. Enhanced Visualizations

Churn Trends & Feature Importance:

- CountPlot: Shows distribution of churn across categorical features like customer score counts

- BarPlot: Used to show average churn rate by features including numeric and categorical

- Violin Plot: Displays churn probability density over numerical features avg_transaction_value, avg_time spent

- PairPlot: Illustrates pairwise relationships between key numerical features,: , avg_time spent,points in wallet, transaction.

- Plotly Dashboards: Interactive visualizations that highlight churn rates, show correlations, and offer interactivity.
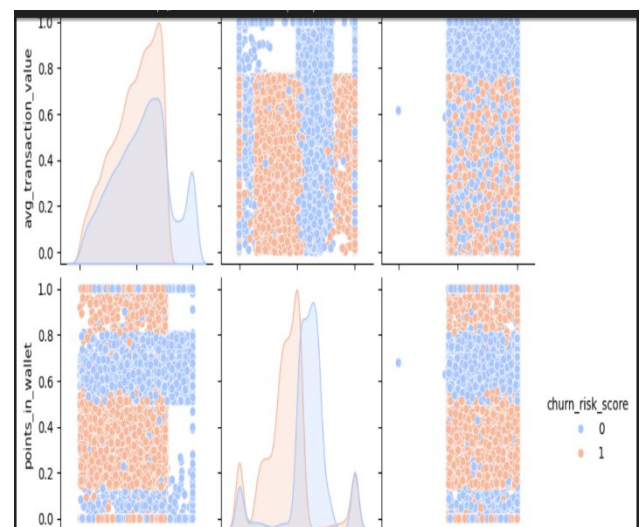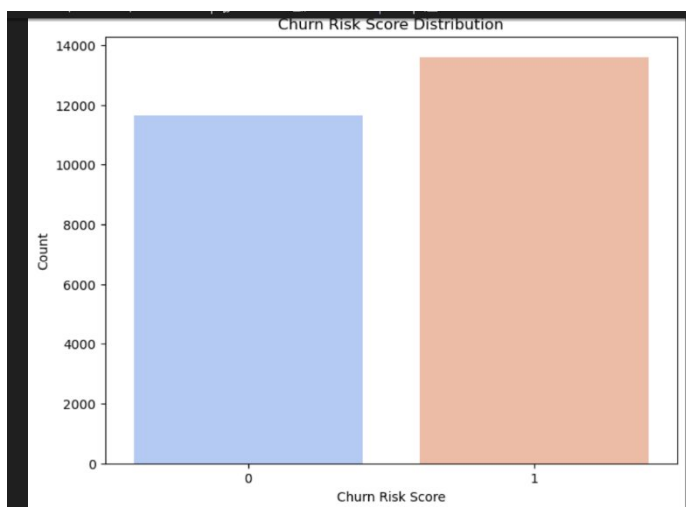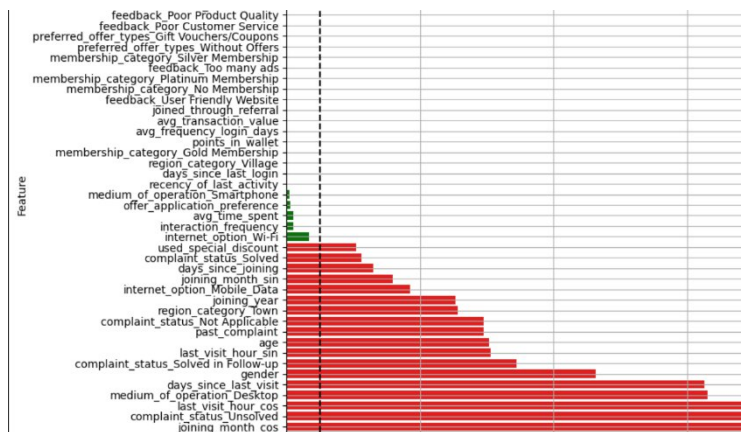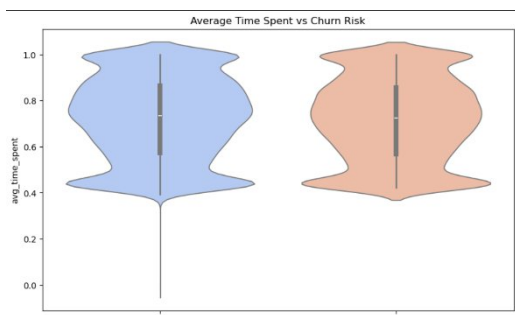
3. Feature Engineering Summary

Feature Engineering Summary:

Feature	Technique	Description	Expected Impact
['gender', 'joined_through_referral', 'used_special_discount', 'offer_application_preference', 'past_complaint']	Label Encoding	Encoded ordinal contract types	Helps capture duration-churn relationship
['region_category', 'membership_category', 'medium_of_operation', 'internet_option', 'complaint_status', 'feedback', 'preferred_offer_types']	One-Hot Encoding	Binary indicator for internet type	Categorical vars encoding to be used in prediction
joining_month_sin **joining_month_cos**	Sin-Cos Transform	Capture seasonality or periodic effects	Captures cyclical behavior like promotional cycles
Numeric data and columns	MinMax Scaling	Normalized charge values	Reduces feature dominance due to scale
PCA_1 to PCA_10	PCA Components	Orthogonal features capturing variance	Improves model performance and reduces overfitting