Churn Prediction - Feature Engineering & Analysis Report

1. Data Analysis Report

Insights from Advanced Feature Analysis:

- Chi-Square Test:
  - Selected top 10 features based on statistical association with Churn.
  - Effective for categorical or non-negative numeric data.

- Correlation Analysis:
  - Removed features with high correlation (> 0.9) to reduce redundancy.
  - Heatmap used to visualize feature interactions.

- ANOVA (F-Test):
  - Evaluated numerical features with f_classif.
  - Selected features with high F-Score and P-Value < 0.05.

- RFE (Recursive Feature Elimination):
  - Used Logistic Regression to identify top 10 predictive features.

- Mutual Information:
  - Captured non-linear relationships between features and Churn.

- Final Feature Selection:
  - Chose features appearing in 3 or more methods.

2. Enhanced Visualizations

- Correlation Heatmap:
  - Showed numerical feature relationships and correlation with Churn.

Churn Prediction - Feature Engineering & Analysis Report

- Feature-Churn Association:

  - Identified key drivers of churn based on new features.

- Suggested Interactive Dashboards:

  - Display churn distribution by service category.

  - Visualize spend_trend and total_services.

  - Segment customers using engineered flags.

3. Feature Engineering Summary

Feature	Description	Impact
has_protection	Any security/support service active	Engagement in safety features
has_streaming	Streaming TV or Movies active	Interest in entertainment services
TotalPaid	MonthlyCharges × tenure	Customer lifetime value
avg_monthly_spend	TotalCharges / (tenure + 0.01)	Historical average spend
spend_trend	MonthlyCharges / avg_monthly_spend	Detects increasing or decreasing
total_services	Count of all subscribed services	Overall service engagement
service_adoption_rate	total_services / (tenure + 1)	Speed of service adoption