Customer Churn Analysis Report for Telecom Provider

**Project Objective:**Analyze customer data to identify key drivers of churn and recommend actionable retention strategies.

**1. Datasets Used:**

* **Customer Records** (15,000 rows): CustomerID, PlanType, MonthlyBill, Tenure, Churn, Complaints
* **Customer Interactions** (20,000 rows): CustomerID, InteractionDate, InteractionType, IssueResolved
* **Demographic Data** (15,000 rows): CustomerID, Age, Region, DeviceType

**2. Data Preparation:**

* Merged datasets on CustomerID
* Handled missing values:  
  + Complaints: filled with median (1.0)
  + Region: filled with mode
  + Interaction data is missing due to many customers not contacting support
* No negative values in MonthlyBill due to generation constraints
* MonthlyBill outliers (~0.9%) observed but not treated as they appeared genuine

**Feature Engineering:**

* ComplaintsGroup (None, Low, Medium, High)
* MonthlyBillGroup (Low, Mid, High, Very High)
* TenureGroup (<6M, 6-12M, 1-2Y, 2-5Y, 5Y+)
* UnresolvedComplaints (TotalComplaints \* (1 - IssueResolvedRate))
* BillToTenure Ratio

**3. Churn Driver Identification (Chi-Square Tests):**

| **Feature** | **P-value** | **Significance** |
| --- | --- | --- |
| MonthlyBillGroup | 3.70e-239 | Strong |
| ComplaintsGroup | 4.71e-166 | Strong |
| TenureGroup | 0.000 | Strong |
| UnresolvedComplaints | 5.67e-44 | Strong |

**Insights:**

* Customers with **High Monthly bills** and **High Complaints** are highly likely to churn
* Churn rate is **51% for customers with <6 months tenure**
* Engineered feature **UnresolvedComplaints** showed a strong correlation with churn

**4. Predictive Modeling (Logistic Regression with SMOTE):**

* **Precision**: 0.40
* **Recall**: 0.55
* **F1-Score**: 0.46
* **ROC AUC**: 0.84

**Confusion Matrix:** [[3370 510]

[ 278 342]]

The model is capable of detecting churners with reasonable accuracy and excellent AUC, especially after balancing.

**5. Recommendations:**

* **Target high-paying, short-tenure customers** with retention offers
* **Resolve support issues faster** to lower the UnresolvedComplaint risk
* **Monitor and act on customers with multiple complaints** quickly
* **Enhance onboarding for new customers (<6 months tenure)**

**Conclusion:**Churn is concentrated in customers with high complaints, unresolved issues, and short tenure. The model and statistical analysis together identified actionable churn drivers with confidence.

**Tools Used:** Python (pandas, seaborn, scikit-learn, statsmodels), Faker for data generation, chi-square test, logistic regression with SMOTE

**Institute:** Skilled Score

**Company:** Kangroo Ventures

**Intern:** Adil Naeem

Thanks for reading.