Churn Rate Analysis and Prediction: Project Insights Document

Project	Telecom Customer Churn Analysis
Objective	Analyze historical churn drivers, profile at-risk customers, and deploy a machine learning model for future churn prediction.
Tools	SQL Server (ETL), Python/Jupyter (Random Forest Classifier), Power BI (Visualization & Reporting)
Key Insight	The highest-risk segment is **Female customers aged 50+**, primarily churning due to **competitor offers and device quality**.

---

1. Data Engineering and Modeling Overview

The foundation of the analysis was a robust ETL (Extract, Transform, Load) process and a logical data model, ensuring data integrity and usability for both descriptive analytics and predictive modeling.

A. Data Source & Preparation (SQL)

The raw customer data was loaded into an SQL Server database, structured into STG_Churn (Raw Staging) and Prod_Churn (Clean Production) tables.

- **Data Cleaning:** Null values (primarily in Churn Category and Churn Reason) were systematically handled and replaced with "None" or "Others" to maintain data completeness without biasing the model.
- **Feature Engineering:** Custom views (VW_Churn_Data, VW_Join_Data) were created to separate historical data (for model training) from newly joined customers (for prediction).

B. Power BI Data Model

Two key transformations were applied in Power BI and connected to the main data:

1. **Bucketing:** Continuous numerical variables like Age and Tenure in Months were grouped into discrete, descriptive categories (Mapping_Age_Grp, Mapping_Tenure_Grp). This improved visualization and interpretability.
2. **Unpivoting Services:** The many binary service columns (e.g., Online Security, Streaming TV) were unpivoted into a single table (Prod_Services), allowing for concise analysis of service penetration and its correlation with churn within one Matrix visual.

---

2. Descriptive Analysis: Key Churn Drivers

The Summary page dashboard provides a breakdown of historical churn by demographic, account, and service factors.

2.1. High-Risk Customer Profile (Demographics)

Insight Category	Finding	Actionable Implication
Gender Skew	**64%** of all historical churners are **Female**.	Gender-specific marketing materials may be more effective.

Age Concentration	Of the female churners, the largest group is **Age > 50** (**31.5%**).	High-value retention resources should be allocated to this specific demographic.
Conclusion	The primary target for immediate retention efforts should be **Female customers aged 50+**.	

2.2. Contract and Account Analysis

- **Contract Type:** The **Month-to-Month** contract type accounts for the highest volume of churners and the highest overall churn rate. This segment needs aggressive monitoring and immediate conversion campaigns to longer-term commitments (1-year/2-year).
- **Tenure:** Customers with a **Tenure of 6-12 Months** show a critical peak in churn rate, indicating a retention-vulnerable period shortly after the initial sign-up honeymoon phase.

2.3. Service Penetration and Quality Gaps

The Service Status Matrix highlights which services act as **"sticky" retention factors** and which are indicators of **service disappointment**.

Service Subscription Status	**Finding (Churn Rate)**	**Implication (Action)**
LOW Subscription Churn (High Churn if **NO**)	Customers *not* subscribed to **Online Security**, **Online Backup**, and **Device Protection Plan** show a churn rate often exceeding **60%**.	These are effective retention tools. Focus on upselling these exact services to at-risk segments.
HIGH Subscription Churn (High Churn if **YES**)	Customers *subscribed* to **Phone Service** and **Unlimited Data** show elevated churn rates.	This suggests a **quality or value problem** with these core services (e.g., poor signal, slow speeds, perceived high cost). Requires product team investigation.

2.4. Root Cause (Churn Reason Drill-Through)

By using the tooltip functionality to drill into the **Competitor** category, the two main historical reasons for loss were identified:

1. **Competitor made better offers.**
2. **Competitor had better devices.**

Recommendation: Retention offers must focus on value bundling (including the "sticky" services like Online Security) and superior device trade-ins or upgrades to counteract competitor influence.

---

3. Predictive Modeling and Future Intervention

The second phase of the project introduced a proactive, forward-looking element using machine learning.

A. Model Methodology (Python/Random Forest)

A Random Forest Classifier was developed and trained on the historical churn data to learn patterns in customer behavior.

Metric	Score	Interpretation
Overall Accuracy	**84%**	The model correctly classified customer status (Churned/Stayed) 84% of the time on the test data.
Stayed (Class 0) Recall	**0.94** (94%)	The model is highly effective at correctly identifying customers who will **stay** (94% correctly classified).
Churned (Class 1) Recall	**0.59** (59%)	The model correctly identified **59%** of all actual churners. This is the primary area for **model tuning** (e.g., using techniques to handle class imbalance) to improve proactive detection.
Churned (Class 1) Precision	**0.80** (80%)	When the model predicts a customer will **churn**, it is correct 80% of the time.

Confusion Matrix Breakdown

The matrix output from the notebook was [[813 49] [141 199]]:

- **True Negatives (813):** Correctly predicted 813 customers would **stay**.
- **True Positives (199):** Correctly predicted 199 customers would **churn**.
- **False Negatives (141):** Incorrectly predicted 141 customers would stay (missed churners).
- **False Positives (49):** Incorrectly predicted 49 customers would churn (false alarms).

B. Feature Importance

The model determined the most influential factors in predicting whether a customer will churn. These variables should be prioritized in marketing segmentation and data monitoring:

1. **Contract Type** (Highest influence)
2. **Tenure in Months**
3. **Monthly Charges**
4. **Internet Service Type**

C. Churn Prediction Page (Actionable Output)

The model output was loaded into the **Churn Prediction** page, which serves as a live early warning system.

- **Actionable List:** The page displays a list of **future customers** (from the joined-customer view) predicted to churn in the coming month/quarter.
- **Predicted Profile:** Selecting any predicted churner instantly generates a full profile, including their revenue, contract details, and services used.

Final Recommendations

1. **Targeted Retention Campaign:** Launch a high-priority campaign focused on **Female customers over 50** , using offers that directly address competitor advantages (device upgrades/better offers) and bundle retention-positive services (Online Security, Device Protection).
2. **Service Quality Audit:** Urgently investigate the high churn among subscribers of **Phone Services** and **Unlimited Data** to address core quality or value issues.
3. **Operationalize Predictions:** Integrate the **Churn Prediction List** into the Retention Team's workflow for weekly or daily action. Prioritize customers with the highest churn probability to maximize retention ROI.