-: Project Name :-

Customer Churn Analysis and Prediction for a Telecom Service Provider

-krish Rathod

INDEX

1. About the Project

→Telecommunication companies operate in a highly competitive market where customer retention directly influences profitability and long-term sustainability. A recurring issue faced by these organizations is customer churn the phenomenon where customers discontinue their services in favor of competitors. Given the high cost of acquiring new subscribers compared to retaining existing ones, understanding churn behavior and predicting it in advance is a matter of strategic importance.

→This project focuses on a structured "data-driven approach" to address the churn challenge. A real-world customer dataset from a telecom service provider was analyzed and modeled. The dataset contains "7,043 customer records and 21 attributes" covering demographics (gender, senior status, dependents, partner), services subscribed (internet, phone, TV, streaming, online security, etc.), financial behavior (monthly charges, total charges, payment methods), and customer lifecycle details (tenure, contract type, billing).

→The work was executed in a “phased analytical process”, starting with “data ingestion”, “cleaning”, and “preprocessing”, followed by “exploratory data analysis (EDA)” to reveal patterns in customer behavior. Further steps included “feature engineering” to capture meaningful risk signals, and “machine learning model development” to predict churn. The project concludes with actionable insights for retention strategy, providing a decision-support framework for business stakeholders.

2. Objective and Problem Statement

A} Objective :-

→The primary objective of this project is to leverage historical customer data to **analyze factors influencing churn** and build predictive models that can classify whether a customer is at risk of leaving the company. By doing so, the organization can proactively identify high-risk customers and intervene with retention measures such as offers, discounts, or service improvements.

B} Problem Statement :-

→Customer churn erodes revenue and market share. Traditional descriptive analytics only provides retrospective views of which customers have left, without enabling proactive action. The problem addressed in this project can be articulated as:

→Given historical data about customer demographics, services, account tenure, billing preferences, and spending patterns, determine which customers are likely to churn in the near future and identify the most significant drivers of churn.”

The solution requires a two-pronged approach:

A}. Analytical Insight :–

→Understanding the key behavioral and financial attributes that correlate with churn, through exploratory and statistical analysis.

B}. Predictive Modeling : –

→Developing machine learning models that generalize these patterns and provide churn probability scores for each customer.

→Solving this problem equips the company with a strategic tool to reduce churn rate, prioritize retention resources, and improve lifetime value of its subscribers.

3. Data Ingestion

→The foundation of this project lies in the Telco Customer Churn Dataset which contains detailed customer-level information. The ingestion phase ensured that the raw dataset was successfully loaded, validated, and prepared for further analysis.

3.1 Source and Structure :-

→Source: The dataset was provided in "CSV format" named "TelcoCustomerChurnDataset.csv".

→Rows & Columns: The file contained "7,043 records" and "21 columns". Each record corresponds to a unique customer identified by a "customerID".

Schema Overview:

→Demographics:`gender`,`SeniorCit`, `Partner`, `Dependents`.

Customer Lifecycle: `tenure` (number of months with the company).

→ServicesSubscribed : - ‘PhoneService`,‘MultipleLines`, `InternetService`,`OnlineSecurity`,`OnlineBackup`,`DeviceProtection`,`TechSupport`, `StreamingTV`, `StreamingMovies`.

→ Financial Attributes :- `MonthlyCharges`, `TotalCharges`.

Target Variable: `Churn` (Yes/No).

3.2 Loading Mechanism :-

→The dataset was imported into the analytical environment using the Pandas library in Python.

→The CSV was read into a DataFrame, enabling tabular manipulation and preprocessing.

```python
df = pd.read_csv("Telco_Customer_Churn_Dataset.csv")
```

→This step formed the base for all subsequent data exploration and modeling activities.

3.3 Data Type Corrections

→Issue Identified :- The column `TotalCharges` was stored as an "object type" rather than numeric, due to embedded blank or whitespace values in the raw data.

Resolution :- Applied the function

→` pd.to_numeric(DataFrame[column_name], errors='ignore')` to convert all values to numeric. Invalid entries were coerced into `NaN`, which were later treated in the cleaning step.

→Impact :- This ensured that `TotalCharges` could be correctly used in statistical computations, visualizations, and machine learning algorithms without errors.

3.4 Handling Missing Values

→During ingestion, missing values were identified primarily in `TotalCharges`.

→Approach Adopted :- Imputation with the "median value" of the column, as median is robust to skewness and outliers.

→This preserved the integrity of the dataset while avoiding biased distortions that might arise from mean imputation.

```python
df['TotalCharges']=df['TotalCharges'].fillna(df['TotalCharges'].median())
```

3.5 Target Variable Transformation

→The`Churn` column contained categorical labels (“Yes” / “No”).

→For modeling purposes, these were mapped into **binary numeric values**:

* Yes → 1
* No → 0

→This conversion was essential for supervised learning algorithms, which require numerical labels.

```python
df['Churn'] = df['Churn'].map({'Yes':1, 'No':0})
```

3.6 Identifier Management

→The column `customerID` was identified as a "non-predictive unique identifier".

→It was excluded from the feature set to avoid misleading the models while being retained in the dataset for reference purposes.

3.7 Integrity Check

After ingestion and initial preprocessing:

A.Shape Verification:-

→7,043 rows and 21 columns confirmed.

B.No Critical Missing Values:-

→All key numerical attributes imputed or corrected.

C.Target Availability:-

→`Churn` variable successfully transformed into binary form.

D.Readiness for Analysis:-

→Dataset deemed consistent and complete, ready for exploratory data analysis.

📌 Up to this stage, the dataset was “clean, consistent, and analysis-ready”. Each customer’s demographic, behavioral, and financial information was preserved in structured form, ensuring that insights generated in subsequent steps would be reliable and accurate.

4. Exploratory Data Analysis (EDA)

The exploratory data analysis aimed to identify patterns, anomalies, and relationships in the dataset that explain customer churn behavior. Below are the major insights**:**

4.1 Churn Rate Distribution

- The overall churn rate was calculated from the dataset.
- Around 26.54% of customers churned, indicating that retention is a significant business challenge.
- Visualization: A bar chart clearly highlighted class imbalance, with more customers staying than leaving.

Key takeaway: Even though the majority stayed, the churn percentage is large enough to demand proactive strategies.

4.2 Demographic Analysis

- Gender: Churn rates were almost evenly distributed between male and female customers → gender is not a strong churn driver.
- Senior Citizens: A higher churn rate was observed among senior citizens compared to younger customers.

- Dependents: Customers with dependents and partners showed lower churn rates, suggesting stronger brand loyalty in family-oriented groups.

Key takeaway: Age and household composition are subtle but relevant churn factors.

4.3 Service Subscription Patterns

- Internet Service: Fiber optic users exhibited the highest churn, followed by DSL, while customers without internet service had the lowest churn.
- Multiple Services: Customers subscribed to multiple services (phone, TV, online security, tech support) churned less frequently.
- Streaming Services: Streaming services had a mixed impact; by themselves they did not strongly predict churn, but combined with other services they reduced churn risk.

Key takeaway: Service bundling increases stickiness, while single-service users are more likely to churn.

4.4 Contract & Payment Methods

- Contract Type:
  - Month-to-month contracts had the highest churn rate.
  - One-year and two-year contracts significantly reduced churn.
- Payment Method:
  - Electronic check payments correlated with the highest churn.
  - Customers using credit cards, debit cards, or bank transfers were more stable.

Key takeaway: Longer commitments and secure payment methods reduce churn risk.

4.5 Billing & Tenure Behavior

- Monthly Charges: Customers with higher monthly charges showed a positive correlation with churn.
- Tenure: Churn decreased sharply as tenure increased — new customers were most likely to churn.
- Total Charges: Consistent with tenure, lower total charges (newer customers) were linked to higher churn.

Key takeaway: High-cost services drive churn for new customers, while long-tenured customers remain loyal.

4.6 Correlation Analysis

- Heatmap of numerical features revealed:
    - tenure and TotalCharges strongly correlated (logical, since longer tenure → higher total charges).
    - MonthlyCharges moderately correlated with churn.
- No evidence of multicollinearity that would disrupt predictive modeling.

✅ Summary of EDA Findings:

- Churn is concentrated in month-to-month, high-charge, electronic check users, particularly among new and senior citizen customers.
- Customers with long tenure, bundled services, and stable payment methods exhibit strong retention.

5.Power BI (Dashboard & Mobile Reporting)

5.1 Power BI Dashboard – Desktop Version

The desktop dashboard serves as the main executive reporting tool. Below is the structured breakdown:

A. KPI Cards (Top Panel)

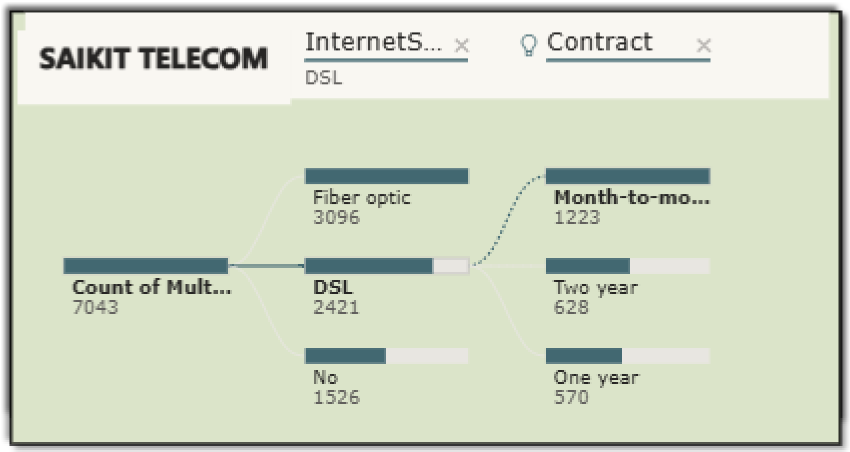
Located at the top row, these provide at-a-glance business metrics:

- Total Charges SUM (16.06M): Total revenue generated by customers.
- Churn Rate (26.54%): Critical metric indicating percentage of customers lost.
- High Risk % (27.83%): Customers at churn risk, identified via predictive attributes (tenure, contract type, payment behavior).
- Monthly Charge AVG (64.76): Typical customer billing level.
- Total Customers (7043): Overall customer base.
- Active Customers (5174): Current retained customers.

5.2 Visuals

1. Customer Segmentation by Internet Service & Contract (Tree/Bar Visual)

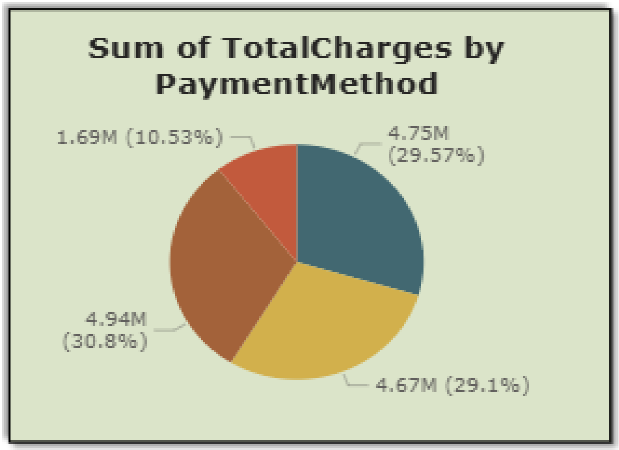
- Shows distribution: Fiber optic (3096), DSL (2421), No internet (1526).
- Contract split: Month-to-month (1223), One-year (570), Two-year (628).
- Insight: Fiber optic users are the largest segment but churn most due to higher pricing.



2. Total Charges by Payment Method (Pie Chart)

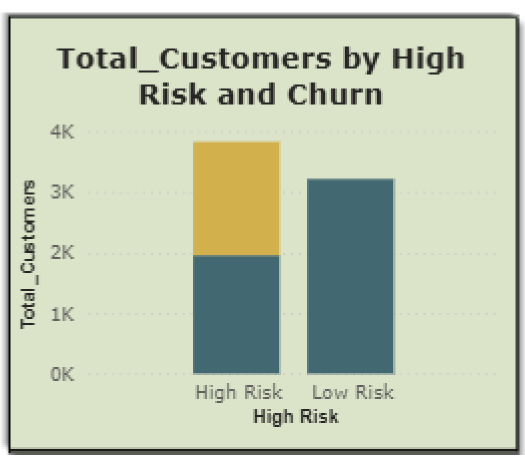
- Payment split across Credit Card, Bank Transfer, Electronic Check, etc.
- Largest share ~30.8% and 29.5% in two categories.
- Insight: Customers using electronic check show higher churn risk – consider incentivizing digital autopay.

- Share : 30.8% is Electronic check, 29.57% is Bank transfer , 29.1% is Credit Card & 10.53% is M Check.



3. Total Customers by Risk & Churn (Clustered Column Chart)

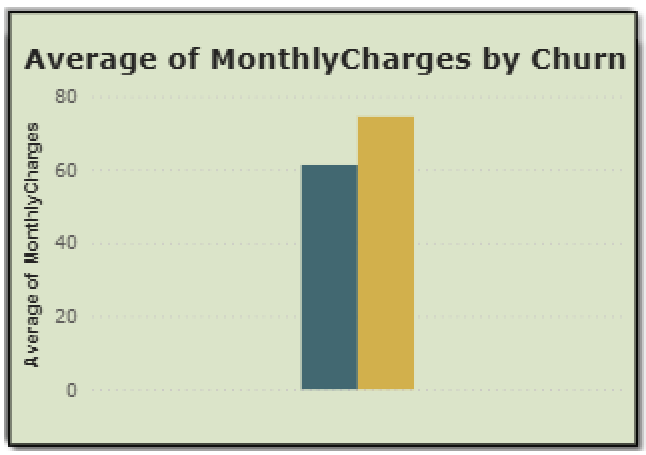
- Splits customers into high risk churners vs low risk non-churners.
- Highlights that churn is concentrated among high-risk customers.



- The blue portion denote churn 0 means not leaving and yellow portion denotes that persion in churn 1 means already leave hue by risk

4. Average Monthly Charges by Churn (Bar Chart)

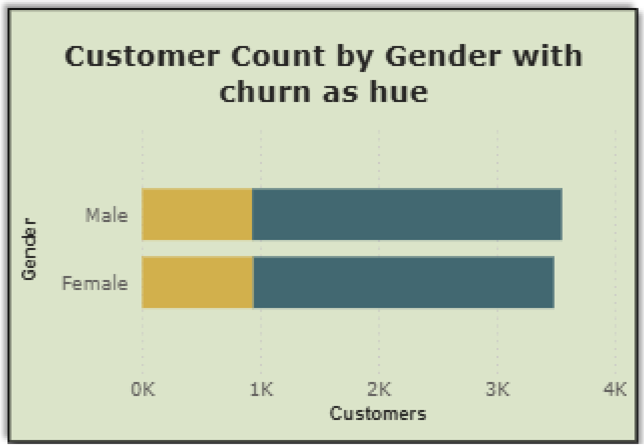
- Compares churned vs non-churned customers.
- Insight: Churned customers pay ~15–20% higher bills.



- The blue portion is churn 0 that average of monthly charges is $61.27 whereas the yellow portion is churn 1 that average of monthly charges is $74.44 which makes the diffrance of $13.17

5. Customer Count by Gender with Churn (Bar Chart)

- Splits male vs female churn.
- Insight: No significant gender bias, churn evenly distributed.

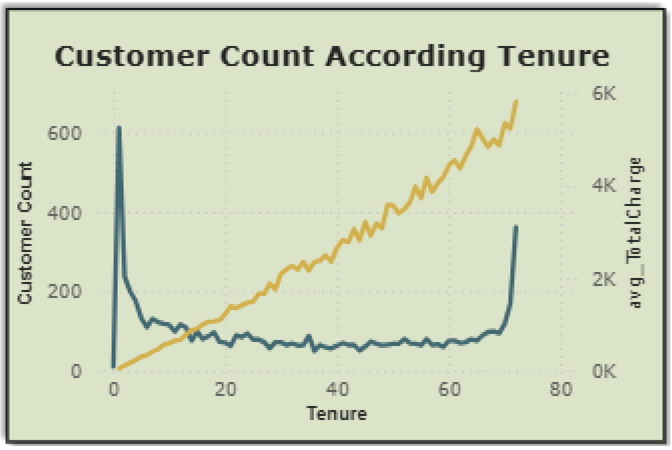


- This chart shows the distribution of gender hue by churn where in Male gender churn 0 customer count is 930 and churn 1 customer count is 2625. in Female gender churn 0 customer count is 939 and churn 1 customer count is 2549. Which shows no gender bias

6. Customer Count by Tenure (Line Chart)

- Shows customer survival curve vs average charges.
- Spikes at beginning (new customers drop early), then stabilizes.
- Long-term customers (>60 months) show loyalty.

- This chart shows the count of customers along with average total charges on the basis of tenure



5.3. Power BI Dashboard – Mobile Version

The mobile dashboard was designed for executives on-the-go, ensuring accessibility without losing core insights.

Mobile Layout Strategy

- Top Section (KPIs Only): Total Charges, Churn Rate, High Risk %, Avg Monthly Charge, Total Customers, Active Customers.
- Middle Section (Key Drivers): Internet Service & Contract (bar), Payment Method (pie).
- Bottom Section (Churn Factors): Risk vs Churn (bar), Monthly Charges by Churn (bar), Gender split, Tenure curve.

Benefits of Mobile Version

- Ensures field executives can track churn trends instantly.
- Provides quick health check of revenue and churn.
- Prioritizes vertical scroll flow – KPIs → Drivers → Factors → Retention.

6. Conclusion

This telecom customer analysis project allowed me to apply end-to-end data analytics — from Python-based exploratory data analysis (EDA) to the development of interactive dashboards in Power BI. By combining statistical insights with visual storytelling, I was able to identify the **key drivers of customer churn, revenue distribution, and customer behavior**.

The findings reveal that:

- Around **26.54% of customers have churned**, with churn being more common among **month-to-month contract holders** and **customers paying through electronic checks**.
- **Fiber optic customers**, although the largest revenue segment, are at higher risk of churn due to comparatively higher charges.
- Early-tenure customers (within the first year) show a **higher probability of leaving**, emphasizing the importance of proactive engagement.
- High monthly charges strongly correlate with churn, signaling that pricing and perceived value are crucial to customer satisfaction.

Through this project, I demonstrated not only technical data processing and visualization skills but also the ability to **translate raw data into actionable business recommendations**.

📌 Recommendations to Reduce Churn

1. **Contract Incentives:** Encourage customers to move from month-to-month contracts to **long-term plans** by offering discounts, loyalty rewards, or bundled services.
2. **Payment Stability:** Transition customers away from electronic check payments by providing cashback offers or extra benefits for **digital auto-pay methods**.
3. **Pricing Strategy:** Review and optimize **fiber optic plan pricing** to improve competitiveness while maintaining profitability.
4. **Onboarding Programs:** Create strong **early engagement strategies** for new customers (first 12 months), such as welcome offers, personalized communication, and proactive customer support.
5. **Risk Monitoring with Predictive Models:** Continuously track high-risk customers using predictive analytics and intervene with **retention campaigns** before churn occurs.
