Customer Churn Analysis

Overview

This exploratory data analysis (EDA) was conducted on a customer churn dataset comprising 7,043 customer records from a telecommunications company. The dataset includes 21 attributes, such as customer demographics (e.g., gender, senior citizen status), service subscriptions (e.g., phone, internet, streaming services), billing details (e.g., monthly charges, total charges, payment method), and **churn status (Yes/No).** The objective was to identify patterns and factors contributing to customer churn, providing actionable insights for retention strategies. The analysis involved thorough data cleaning, statistical inspections, and visualizations, with a focus on understanding churn behavior through key attributes like payment methods.

Data Preparation and Cleaning

- **Initial Inspection**:
  - The dataset was loaded using Pandas and inspected with **df.info()** and df.head() to understand its structure.
  - No missing values were found (df.isnull().sum().sum() = 0), ensuring data completeness.
  - No duplicate customer IDs (df['customerID'].duplicated().sum() = 0) or rows (df.duplicated().sum() = 0) were detected, confirming data integrity.
- **Data Type Correction**:
  - The TotalCharges column was initially an object type due to blank values in records with zero tenure. These blanks were replaced with 0 (df['TotalCharges'].replace(" ", 0)), and the column was converted to float (df['TotalCharges'].astype("float")) for numerical analysis.
- **Feature Transformation**:
  - The SeniorCitizen column, originally binary (0/1), was converted to categorical values ("No"/"Yes") using a custom function (conv) to enhance interpretability in visualizations and analysis.
- **Data Quality**:
  - Post-cleaning, the dataset consisted of 2 float64 columns (MonthlyCharges, TotalCharges), 2 int64 columns (tenure, SeniorCitizen before transformation), and 17 object columns (categorical variables like gender, PaymentMethod, Churn).
  - Memory usage was approximately 1.1 MB, indicating efficient handling for EDA.

Statistical Insights

- **Descriptive Statistics** (df.describe()):
  - **Tenure**: Mean of 32.37 months, standard deviation of 24.56 months, ranging from 0 to 72 months. The median tenure was 29 months, with 25% of customers having tenure below 9 months and 75% below 55 months.

  - **Monthly Charges**: Mean of $64.76, standard deviation of $30.09, ranging from $18.25 to $118.75. The median was $70.35, indicating a slight skew toward higher charges.
  - **Total Charges**: Mean of $2,279.73, standard deviation of $2,266.79, ranging from $0 to $8,684.80. The median was $1,394.55, reflecting a wide distribution due to varying tenure and service plans.
  - **Senior Citizen**: Approximately 16.21% of customers were senior citizens (mean of 0.162147 before transformation).
- **Churn Distribution**:
  - The Churn column indicated that 26.54% of customers churned (1,869 out of 7,043), while 73.46% remained (5,174 out of 7,043), highlighting an imbalanced dataset that may require attention in predictive modeling.

Key Visualizations and Findings

- **Churn by Payment Method**:
  - A count plot was generated using Seaborn (sns.countplot) to analyze churn across payment methods: Electronic check, Mailed check, Bank transfer (automatic), and Credit card (automatic).
  - **Raw Counts** (from bar labels):
    - **Electronic check**: 1,295 non-churned (55.01%), 1,076 churned (44.99%).
    - **Mailed check**: 1,304 non-churned (81.30%), 308 churned (18.70%).
    - **Bank transfer (automatic)**: 1,285 non-churned (83.23%), 258 churned (16.77%).
    - **Credit card (automatic)**: 1,290 non-churned (84.70%), 233 churned (15.30%).
  - **Churn Rates** (calculated as percentage of churned customers per payment method):
    - **Electronic check**: 44.99% churn rate, significantly higher than others, indicating a strong association with churn.
    - **Mailed check**: 18.70% churn rate, moderate but notable.
    - **Bank transfer (automatic)**: 16.77% churn rate, relatively low.
    - **Credit card (automatic)**: 15.30% churn rate, the lowest among all methods.
  - **Insight**: Customers using electronic checks are nearly three times more likely to churn compared to those using automatic payment methods (bank transfer or credit card). This suggests potential issues with the electronic check process, such as transaction delays, user experience, or lack of commitment compared to automatic payments.
- **Visualization Details**:
  - The plot was customized with a figure size of 5x4 inches for clarity, rotated x-axis labels (45 degrees) for readability, and included bar labels to display exact counts for both churned and non-churned customers.
  - The title "Churn by Payments Methods" clearly conveyed the focus, and the hue parameter differentiated churn status, making the visualization intuitive.

Conclusions

The EDA reveals that payment method is a critical factor influencing customer churn, with electronic check users exhibiting a 44.99% churn rate, far exceeding the rates for mailed check (18.70%), bank transfer (16.77%), and credit card (15.30%). This discrepancy suggests that electronic checks may be associated with operational or experiential issues that drive customers away. Conversely, automatic payment methods correlate with higher retention, likely due to their convenience and contractual commitment. The dataset, now cleaned and with transformed features, is well-prepared for advanced analyses, such as predictive modeling to identify at-risk customers or segmentation to tailor retention strategies.

Recommendations

1. **Investigate Electronic Check Issues**:
    - Conduct qualitative research (e.g., customer surveys) to identify specific pain points with electronic checks, such as processing errors, delays, or cumbersome interfaces.
    - Analyze transaction failure rates or support tickets related to electronic checks to quantify operational issues.
2. **Promote Automatic Payment Methods**:
    - Offer incentives (e.g., discounts, loyalty rewards) to encourage customers to switch to bank transfers or credit cards, which have churn rates below 17%.
    - Simplify the enrollment process for automatic payments to increase adoption.
3. **Enhance Visualization and Analysis**:
    - Create additional visualizations to explore churn by other attributes, such as contract type (month-to-month vs. one/two-year), tenure, or internet service type (DSL vs. Fiber optic), to uncover further drivers of churn.
    - Use box plots or histograms to examine the distribution of tenure and MonthlyCharges by churn status, potentially revealing thresholds where churn risk increases.
4. **Advance to Predictive Modeling**:
    - Develop machine learning models (e.g., logistic regression, random forests) to predict churn probability, leveraging the cleaned dataset and insights from this EDA.
    - Focus on feature importance to quantify the impact of payment method alongside other variables like contract type or tenure.
5. **Targeted Retention Strategies**:
    - Design retention campaigns for electronic check users, offering personalized support or incentives to remain with the company.
    - Segment customers by tenure and payment method to prioritize high-risk groups (e.g., short-tenure customers using electronic checks).

Next Steps

- Expand the EDA to include correlations between numerical variables (e.g., tenure vs. TotalCharges) and churn, using heatmaps or scatter plots.
- Perform statistical tests (e.g., chi-square for categorical variables like PaymentMethod, t-tests for numerical variables like MonthlyCharges) to validate the significance of observed patterns.

- Prepare the dataset for modeling by encoding categorical variables (e.g., one-hot encoding for PaymentMethod, Contract) and scaling numerical features.

This comprehensive EDA provides a robust foundation for understanding customer churn and actionable insights to reduce churn rates, particularly by addressing issues with electronic check payments and promoting automatic payment methods.