

TELECOM CHURN INSIGHTS

ANALYTICS, PREDICTION & VIZUALIZATION



DM5 | SARANYALAKSHMI SUBBURAM | JULY 2023

INTRODUCTION

This project deals with various data analytics and business strategy insights that have been implemented using tools such as SQL, PYTHON (Machine Learning), and Power BI. The scope of this project is to deal with the telecom industry churn analysis, deriving insights, recommending actionable solutions and presenting the insights through user interactive visualization.

DOMAIN BACKDROP

The telecom domain is a sector that encompasses various communication technologies, services, and infrastructure. It includes telecommunications companies that provide voice, data, and internet services to customers. In this highly competitive industry, customer churn (the rate at which customers switch to competitors or discontinue services) holds great importance. Churn impacts telecom companies in multiple ways, including financial implications, loss of market share, and reputation. It is crucial for telecom companies to understand the factors that drive churn, such as contract type, service quality, pricing, and customer satisfaction. By effectively predicting and managing churn, companies can implement targeted retention strategies, improve customer loyalty, and maintain a competitive edge in the market.

METHODOLOGY

SQL Random Data Generation

In this section, the process of generating random data using SQL for telecom churn prediction. The importance of data generation for training and testing machine learning models and provide an overview of the techniques used in generating realistic datasets. By simulating customer behavior and churn events, we ensure the availability of diverse and representative data for accurate predictions.

Creating Power BI Visuals

This section focuses on creating Power BI visuals to gain valuable insights from the generated data as a single-person project. It is demonstrated how Power BI can be used to visualize churn patterns, customer demographics, and other relevant metrics. Through interactive dashboards and reports, we enable ourselves to analyse key performance indicators and identify actionable churn-related trends for informed decision-making.

Data Preparation, Cleaning, and Performing EDA: Univariate, Bivariate, and Multivariate Analysis

Before conducting any analysis or building predictive models, it is essential to prepare and clean the data. This column explores the crucial steps involved in data preparation, cleaning, and performing Exploratory Data Analysis (EDA) using univariate, bivariate, and multivariate analysis techniques.

Data Preparation: This phase involves tasks such as handling missing values, removing duplicates, standardizing data formats, and encoding categorical variables. Data preparation ensures that the dataset is ready for analysis and model building.

Data Cleaning: Data cleaning focuses on identifying and handling outliers, correcting inconsistencies, dealing with noisy data, and addressing data quality issues. By eliminating or mitigating these problems, we ensure the accuracy and reliability of the analysis results.

Univariate Analysis: Univariate analysis involves analyzing individual variables to understand their distributions, central tendencies, and spread. Techniques such as histograms, box plots, and summary statistics are used to gain insights into the characteristics and patterns of each variable.

Bivariate Analysis: Bivariate analysis examines the relationship between two variables. It helps identify correlations, dependencies, and associations between variables. Scatter plots, correlation matrices, and hypothesis testing techniques aid in understanding the strength and nature of relationships.

Multivariate Analysis: Multivariate analysis explores the relationships between multiple variables simultaneously. Techniques like cluster analysis, factor analysis, and dimensionality reduction methods like Principal Component Analysis (PCA) are applied to uncover complex patterns and dependencies within the dataset.

Data preparation, cleaning, and EDA play vital roles in extracting meaningful insights from data. By performing univariate, bivariate, and multivariate analysis, we gain a comprehensive understanding of the dataset's characteristics and relationships between variables. These steps lay the foundation for further analysis and model development, enabling informed decision-making and data-driven strategies.

From the above analysis, three **CRUCIAL INSIGHTS** have been derived.

High CHURN Factors: **HIGH** Churn seen in case of Month-to-month contracts, lower call duration, No Tech support, No movie & tv subscriptions, Fiber Optics Internet, and gender_female

Low CHURN Factors: **LOW** Churn is seen in case of Higher monthly charges, with tech_support, the customers who are in yearly contract, With Movie & Tv subscription, With Internet service_DSL, and gender_male.

Factors with NO IMPACT on CHURN: Factors like Availability of Phone Service, online backup and # of multiple lines have almost **NO impact** on Churn.

Churn Prediction with ML Models

Addressing Imbalanced Data in Telecom Churn Prediction using SMOTE-ENN

Imbalanced data is a common challenge in telecom churn prediction, where the number of churned customers is significantly lower than non-churners. To overcome this issue, the SMOTE-ENN (Synthetic Minority Over-sampling Technique with Edited Nearest Neighbours) approach is employed. SMOTE-ENN combines the SMOTE algorithm, which generates synthetic samples for the minority class, with the ENN technique, which removes noisy samples from both classes. By oversampling the minority class and under sampling the majority class, SMOTE-ENN effectively balances the dataset, leading to improved model performance and churn prediction accuracy. This technique enables more accurate identification of churn patterns and empowers telecom companies to take proactive measures for customer retention and business growth.

In the next section, we explore the application of machine learning models, including Principal Component Analysis (PCA), for telecom churn prediction. We discuss **four** models: Decision Tree Classification, Random Forest Classification, PCA, and Logistic Regression. Here are their model summaries:

1. Decision Tree Classification:

- Model Summary: Decision trees create a flowchart-like structure to make predictions based on features such as call duration, plan type, and customer demographics. It partitions the data based on attribute values and produces a tree-like model for classification.
- Accuracy: 74.71%

2. Random Forest Classification:

- Model Summary: Random Forest combines multiple decision trees to improve accuracy and reduce overfitting. It randomly selects features and creates an ensemble of decision trees to make predictions.
- Accuracy: 88.23%

3. PCA (Principal Component Analysis):

- Model Summary: PCA is a dimensionality reduction technique that transforms the original features into a lower-dimensional space. It identifies the most significant components of the data to retain relevant information.
- Accuracy: 72.94%

4. Logistic Regression:

- Model Summary: Logistic regression models the probability of a binary outcome (churn or non-churn) based on independent variables. It estimates the coefficients of the input features to make predictions.
- Accuracy: 91.76%

The difference in accuracy percentages can be attributed to various factors.

- Decision trees may have a lower accuracy due to their tendency to overfit the training data.
- Random Forest Classification, by combining multiple trees, can provide better accuracy and reduce overfitting.
- PCA, while useful for dimensionality reduction, may not capture the complex relationships in the data as effectively as other models. Logistic Regression, with its statistical approach, provide higher accuracy due to its ability to model the relationship between features and the target variable more precisely.

ACTIONABLE INSIGHTS

Based on the gained insights on high and low churn factors, the actionable recommendations are provided as follows:

I. High Churn Factors:

- Month-to-month contracts: Encourage customers to opt for longer contract durations by offering incentives like discounts or additional services.
- Lower call duration: Enhance the quality of customer service and address any issues promptly to increase customer engagement.
- No tech support: Invest in providing reliable and efficient technical support to ensure customer satisfaction and reduce churn.
- No movie and TV subscriptions: Introduce personalized offers or bundles that include attractive entertainment options to increase customer retention.
- Fiber Optics Internet: Focus on delivering high-speed and reliable internet services to maintain customer satisfaction.
- Female customers: Understand specific needs and preferences of female customers and tailor marketing strategies accordingly.

2. Low Churn Factors:

- Higher monthly charges: Communicate the value and benefits of premium services to justify the higher charges and retain customers.
- With tech support: Continue providing excellent technical support services to enhance customer satisfaction and reduce churn.
- Yearly contracts: Promote the advantages of long-term contracts, such as discounted rates or added benefits, to encourage customer commitment.
- With movie and TV subscriptions: Highlight the convenience and entertainment options available with movie and TV subscriptions to retain customers.
- With DSL internet service: Emphasize the stability and reliability of DSL internet service to address any concerns customers may have.
- Male customers: Tailor marketing efforts to cater to the specific needs and preferences of male customers.

3. Factors with No Impact on Churn:

- Availability of Phone Service, online backup, and # of multiple lines: While these factors may not directly influence churn, it is still important to ensure the availability and reliability of these services to maintain overall customer satisfaction.

By focusing on these actionable recommendations based on churn factors, we can effectively control churn rates and enhance customer retention in your telecom business. Regularly monitor customer feedback and preferences to stay updated on evolving customer needs and adjust your strategies accordingly.

Dashboard via PowerBI

