PROJECT RESULTS REPORT

TELCO CUSTOMER CHURN ANALYSIS

PRESENTED BY: CHAN

horizontal line

# Placeholder image

# Introduction

“Predict behavior to retain customers. You can analyse all relevant customer data and develop focused customer retention programs.” - IBM Sample Datasets.

This project addresses the critical issue of customer churn within the telecommunication sector. As highlighted in previous studies, understanding challenges faced in retaining customers is paramount for maintaining a competitive advantage. This investigation aligns with prior studies that underscore the substantial problem of customer attrition in telecommunications. While strategic marketing campaigns are crucial for acquiring new subscribers, retaining existing customers is demonstrably more cost-effective, albeit often more challenging. Consequently, the ability to accurately predict a customer's likelihood of terminating their relationship with a telecommunications provider represents a significant strategic asset.

Drawing upon established frameworks for churn analysis, such as the clustering and classification techniques recommended by Batch et al. (2021) and the ensemble and clustering-based model combinations proposed by Fathian et al. (2016), this project will further examine the relationship between various customer attributes and their propensity to churn. Employing both unsupervised and supervised machine learning methodologies, this analysis will re-investigate these critical connections within the context of a specific telecommunications provider's data. The findings will be comprehensively presented, supported by visualisations to elucidate key patterns and insights. Ultimately, the primary objective of this project is to generate actionable intelligence that can inform strategic marketing investments and enable the development of targeted customer retention programs.

## Context and Data

The Telco customer churn data contains information about a fictional telco company that provided home phone and Internet services to 7043 customers in California in Q3. It indicates which customers have left, stayed, or signed up for their service. Each row represents a customer, each column contains customer’s attributes described on the column Metadata. The “Churn” column is our target. Below are details of the dataset.

**Demographics:**

* CustomerID: A unique ID that identifies each customer.
* Count: A value used in reporting/dashboarding to sum up the number of customers in a filtered set.
* Gender: The customer’s gender: Male, Female
* Age: The customer’s current age, in years, at the time the fiscal quarter ended.
* Senior Citizen: Indicates if the customer is 65 or older: Yes, No
* Married: Indicates if the customer is married: Yes, No
* Dependents: Indicates if the customer lives with any dependents: Yes, No. *Dependents could be children, parents, grandparents, etc.*
* Number of Dependents: Indicates the number of dependents that live with the customer.

**Location:**

* CustomerID: A unique ID that identifies each customer.
* Count: A value used in reporting/dashboarding to sum up the number of customers in a filtered set.
* Country: The country of the customer’s primary residence.
* State: The state of the customer’s primary residence.
* City: The city of the customer’s primary residence.
* Zip Code: The zip code of the customer’s primary residence.
* Lat Long: The combined latitude and longitude of the customer’s primary residence.
* Latitude: The latitude of the customer’s primary residence.
* Longitude: The longitude of the customer’s primary residence.

**Population:**

* ID: A unique ID that identifies each row.
* Zip Code: The zip code of the customer’s primary residence.
* Population: A current population estimate for the entire Zip Code area.

**Services:**

* CustomerID: A unique ID that identifies each customer.
* Count: A value used in reporting/dashboarding to sum up the number of customers in a filtered set.
* Quarter: The fiscal quarter that the data has been derived from (e.g. Q3).
* Referred a Friend: Indicates if the customer has ever referred a friend or family member to this company: Yes, No
* Number of Referrals: Indicates the number of referrals to date that the customer has made.
* Tenure in Months: Indicates the total amount of months that the customer has been with the company by the end of the quarter specified above.
* Offer: Identifies the last marketing offer that the customer accepted, if applicable. Values include None, Offer A, Offer B, Offer C, Offer D, and Offer E.
* Phone Service: Indicates if the customer subscribes to home phone service with the company: Yes, No
* Avg Monthly Long Distance Charges: Indicates the customer’s average long distance charges, calculated to the end of the quarter specified above.
* Multiple Lines: Indicates if the customer subscribes to multiple telephone lines with the company: Yes, No
* Internet Service: Indicates if the customer subscribes to Internet service with the company: No, DSL, Fiber Optic, Cable.
* Avg Monthly GB Download: Indicates the customer’s average download volume in gigabytes, calculated to the end of the quarter specified above.
* Online Security: Indicates if the customer subscribes to an additional online security service provided by the company: Yes, No
* Online Backup: Indicates if the customer subscribes to an additional online backup service provided by the company: Yes, No
* Device Protection Plan: Indicates if the customer subscribes to an additional device protection plan for their Internet equipment provided by the company: Yes, No
* Premium Tech Support: Indicates if the customer subscribes to an additional technical support plan from the company with reduced wait times: Yes, No
* Streaming TV: Indicates if the customer uses their Internet service to stream television programing from a third party provider: Yes, No. The company does not charge an additional fee for this service.
* Streaming Movies: Indicates if the customer uses their Internet service to stream movies from a third party provider: Yes, No. The company does not charge an additional fee for this service.
* Streaming Music: Indicates if the customer uses their Internet service to stream music from a third party provider: Yes, No. The company does not charge an additional fee for this service.
* Unlimited Data: Indicates if the customer has paid an additional monthly fee to have unlimited data downloads/uploads: Yes, No
* Contract: Indicates the customer’s current contract type: Month-to-Month, One Year, Two Year.
* Paperless Billing: Indicates if the customer has chosen paperless billing: Yes, No
* Payment Method: Indicates how the customer pays their bill: Bank Withdrawal, Credit Card, Mailed Check
* Monthly Charge: Indicates the customer’s current total monthly charge for all their services from the company.
* Total Charges: Indicates the customer’s total charges, calculated to the end of the quarter specified above.
* Total Refunds: Indicates the customer’s total refunds, calculated to the end of the quarter specified above.
* Total Extra Data Charges: Indicates the customer’s total charges for extra data downloads above those specified in their plan, by the end of the quarter specified above.
* Total Long Distance Charges: Indicates the customer’s total charges for long distance above those specified in their plan, by the end of the quarter specified above.

**Status:**

* CustomerID: A unique ID that identifies each customer.
* Count: A value used in reporting/dashboarding to sum up the number of customers in a filtered set.
* Quarter: The fiscal quarter that the data has been derived from (e.g. Q3).
* Satisfaction Score: A customer’s overall satisfaction rating of the company from 1 (Very Unsatisfied) to 5 (Very Satisfied).
* Satisfaction Score Label: Indicates the text version of the score (1-5) as a text string.
* Customer Status: Indicates the status of the customer at the end of the quarter: Churned, Stayed, or Joined
* Churn Label: Yes = the customer left the company this quarter. No = the customer remained with the company. Directly related to Churn Value.
* Churn Value: 1 = the customer left the company this quarter. 0 = the customer remained with the company. Directly related to Churn Label.
* Churn Score: A value from 0-100 that is calculated using the predictive tool IBM SPSS Modeler. The model incorporates multiple factors known to cause churn. The higher the score, the more likely the customer will churn.
* Churn Score Category: A calculation that assigns a Churn Score to one of the following categories: 0-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, and 91-100
* CLTV: Customer Lifetime Value. A predicted CLTV is calculated using corporate formulas and existing data. The higher the value, the more valuable the customer. High value customers should be monitored for churn.
* CLTV Category: A calculation that assigns a CLTV value to one of the following categories: 2000-2500, 2501-3000, 3001-3500, 3501-4000, 4001-4500, 4501-5000, 5001-5500, 5501-6000, 6001-6500, and 6501-7000.
* Churn Category: A high-level category for the customer’s reason for churning: Attitude, Competitor, Dissatisfaction, Other, Price. When they leave the company, all customers are asked about their reasons for leaving. Directly related to Churn Reason.
* Churn Reason: A customer’s specific reason for leaving the company. Directly related to Churn Category.

Note: some attributes in the data will not be used due to the author's perspective and duplicated columns.

# Tasks and Results

Details of codes and results can be found in <https://github.com/tangchanthieng/Telo-Customer-Churn/blob/main/code/Telco_Churn_Prediction.ipynb>

## Task 1. Data Preprocessing

* Packages Import (Pandas, Numpy, Matplotlib.pyplot, Seaborn)
* Data Import
* Merging Datasets
* Data Cleaning (Duplicates, Missing values, Imputations, Distributions, Manipulation, and Transformation)
* Data Split for Unsupervised Learning

*Results: Merged data is exported as a CSV file for other purposes.*

## 

## Task 2. Unsupervised Learning

* K-Mean Clustering (PCA)
* K-Mean Clustering (Without PCA)

*Results: Comparisons of Clustering techniques with and without PCA.*

## Task 3. Supervised Learning

* Data Split (70:30)
* Baseline Model (KNN)
* Logistic Regression
* XGBoost (SHAP and LIME Interpretation)
* Modelling with PCA proxies (Services)

*Results: Improved accuracy of models after implementing the PCA proxy.*

## Task 4. Power BI Dashboard

* Summary Page
* Demographics Page
* Services Pages
* Status Page

*Results: Visualisations of figures and relations between attributes. Visually easier to draw insights.*

# 

# 

# 

# 

# Insights

## Retention Segmentation Strategy

Objective: Target at-risk personas with tailored communication and service offers.

Action Items:

* Age-focused personalization: Users aged 50+ are churning more. Introduce tech-literacy support, simplified UX, and proactive assistance.
* Single users and 30+ age bracket: Launch solo user plans with loyalty incentives (e.g., data rollovers, annual discounts).
* Monthly contract users: These are churn-prone. Push annual or 6-month lock-ins with clear value-added services (e.g., exclusive bundles, price freeze guarantees).

## Offer Optimization & Incentive Engineering

Objective: Build a compelling value proposition to offset price sensitivity and competitive poaching.

Action Items:

* Competitive benchmarking campaign: Explicitly highlight where your pricing or service beats competitors—call out gaps and plug them.
* Promotions targeting churn flags:
* Bundle internet with streaming services for stickiness.
* Offer online backup/security as part of the “premium customer care” tier.
* Offer free trials of protection plans (not selling well, but could reduce anxiety-driven churn).
* Referral rewards: Particularly in high-churn cities. Use network effects to create stickiness.

## 

## Customer Experience (CX) Revamp

Objective: Solve the root cause of dissatisfaction before it turns into churn.

Action Items:

* Fix internet and phone service quality: This is the red flag. Prioritize infrastructure upgrades or manage customer expectations better (via transparent service maps and real-time issue updates).
* City-based CX task forces: Only 5 cities have high satisfaction—replicate best practices from these locations.
* Post-signup onboarding: Especially for older or monthly-contract users. Offer step-by-step onboarding via video or agent call to set expectations.
* Paperless billing opt-out experience: Make it optional and explain benefits, especially to older users.

## Payment and Refund Policy Reform

Objective: Reduce friction in payments and stop revenue bleed from poorly structured refunds.

Action Items:

* Deprioritize bank withdrawal: Promote card/autopay methods with discounts.
* Redesign refund policy: 60% revenue loss via refunds is unsustainable. Consider refund caps, loyalty-linked refund privileges, or staggered refund models.
* Monthly GB usage insights: Offer flexible data plans and alerts if users are underusing, with a “smart downgrade” option.