# **Project Proposal**

|  |  |
| --- | --- |
| Name: | K A L Bhathiya Bandara |
| Index No: | 239147H |

# Dataset

This dataset belongs to the Telecommunication domain and contains details about the customer and their churn status in California in Q2 2022. Information about demographics, location, tenure, subscription services, and status of 7043 customers are included in the dataset. The dataset contains 2 CSV files with the following attributes.

* telecom\_customer\_churn.csv

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Field** | **DataType** | **ID** | **Field** | **DataType** | **ID** | **Field** | **DataType** |
| 1 | CustomerID | String | 14 | Avg Monthly Long Distance Charges | Float | 27 | Contract | String |
| 2 | Gender | String | 15 | Multiple Lines | String | 28 | Paperless Billing | String |
| 3 | Age | Int | 16 | Internet Service | String | 29 | Payment Method | String |
| 4 | Married | String | 17 | Internet Type | String | 30 | Monthly Charge | Float |
| 5 | Number of Dependents | Int | 18 | Avg Monthly GB Download | Int | 31 | Total Charges | Float |
| 6 | City | String | 19 | Online Security | String | 32 | Total Refunds | Float |
| 7 | Zip Code | Int | 20 | Online Backup | String | 33 | Total Extra Data Charges | Int |
| 8 | Latitude | Float | 21 | Device Protection Plan | String | 34 | Total Long Distance Charges | Float |
| 9 | Longitude | Float | 22 | Premium Tech Support | String | 35 | Total Revenue | Float |
| 10 | Number of Referrals | Int | 23 | Streaming TV | String | 36 | Customer Status | String |
| 11 | Tenure in Months | Int | 24 | Streaming Movies | String | 37 | Churn Category | String |
| 12 | Offer | String | 25 | Streaming Music | String | 38 | Churn Reason | String |
| 13 | Phone Service | String | 26 | Unlimited Data | String |  |  |  |

* telecom\_zipcode\_population.csv

|  |  |  |
| --- | --- | --- |
| **ID** | **Field** | **DataType** |
| 1 | Zip Code | String |
| 2 | Population | Int |

Dataset link: <https://www.kaggle.com/datasets/shilongzhuang/telecom-customer-churn-by-maven-analytics>

GitHub repo: <https://github.com/BhathiyaBandra/MSC-Data_Science-Project>

# Beneficiary

The beneficiary of this analysis is the Marketing Department of the respective Telecommunication Service Provider. This analysis helps them to identify the Churn groups and design retention plans (special packages, promotion campaigns, loyalty points, etc.) to retain the customers.

# Research questions

This study tries to identify the factors behind customer churn and proposes. Following areas to be covered.

* + Identify high-value customers,
  + Identify Churn risks groups.

# Expected deliverables

Present findings in an interactive dashboard with charts/tables or a form of a report.

# Assumptions

Since the dataset does not have any Service Provider identification field in the dataset, I assume that all this data was collected from one operator in California.