#### **Project Title**

Customer Churn Analysis

#### **Project Summary**

To Identify patterns/reasons for customer churn rate and to develop strategies to retain customers. Customer PII is potentially at risk.

#### **Business Objectives**

Identify patterns and reasons for customer churn to develop strategies to retain customers.

#### **Data Sources**

*List all data sources that will be used in the project. Include details such as the type of data, format, and how it will be accessed or collected.*

| **Data Source Name** | **Type (e.g., CSV, Database, API)** | **Description** | **Location/Access Method** |
| --- | --- | --- | --- |
| Telco customer churn (11.1.3+) | CSV | 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. Multiple important demographics are included for each customer, as well as a Satisfaction Score, Churn Score, and Customer Lifetime Value (CLTV) index. | Kaggle: <https://www.kaggle.com/datasets/alfathterry/telco-customer-churn-11-1-3> |

#### **Data Description**

*Provide a detailed description of the data, including key attributes, data types, and any relevant data dictionaries or schemas.*

| **Attribute Name** | **Data Type** | **Description** | **Example Value** |
| --- | --- | --- | --- |
| CustomerID | String | A unique ID that identifies each customer | 3488-PGMQJ |
| Gender | Character | The customer’s gender: Male, Female | Male |
| Age | Integer | The customer’s current age, in years, at the time the fiscal quarter ended. | 78 |
| Under 30 | Boolean | Discloses whether or not customers are under 30. | Yes |
| Senior Citizen | Boolean | Indicates if the customer is 65 or older: Yes, No | Yes |
| Married | Boolean | Indicates if the customer is married: Yes, No | No |
| Dependents | Boolean | Indicates if the customer lives with any dependents: Yes, No. Dependents could be children, parents, grandparents, etc. | Yes |
| Number of Dependents | Integer | Indicates the number of dependents that live with the customer. | 3 |
| Country | Character | The country of the customer’s primary residence. | United States |
| State | Character | The state of the customer’s primary residence. | California |
| City | Character | The city of the customer’s primary residence. | Los Angeles |
| Zip Code | Integer | The zip code of the customer’s primary residence. | 90005 |
| Latitude | Float | The latitude of the customer’s primary residence. | 34.059281 |
| Longitude | Float | The longitude of the customer’s primary residence. | -118.30742 |
| Population | Integer | A current population estimate for the entire Zip Code area. | 43019 |
| Quarter | String | The fiscal quarter that the data has been derived from (e.g. Q3). | Q3 |
| Referred a Friend | Boolean | Indicates if the customer has ever referred a friend or family member to this company: Yes, No | No |
| Number of Referrals | Boolean | Indicates the number of referrals to date that the customer has made. | Yes |
| Tenure in Months | Integer | Indicates the total amount of months that the customer has been with the company by the end of the quarter specified above. | 18 |
| Offer | Character | 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. | Offer E |
| Phone Service | Boolean | Indicates if the customer subscribes to home phone service with the company: Yes, No | No |
| Avg Monthly Long Distance Charges | Float | Indicates the customer’s average long distance charges, calculated to the end of the quarter specified above. | 19.76 |
| Multiple Lines | Boolean | Indicates if the customer subscribes to multiple telephone lines with the company: Yes, No | Yes |
| Internet Service | Boolean | Indicates if the customer subscribes to Internet service with the company: No, DSL, Fiber Optic, Cable. | Yes |
| Internet Type | Character | Indicate the type of internet the customer has. | DSL |
| Avg Monthly GB Download | Integer | Indicates the customer’s average download volume in gigabytes, calculated to the end of the quarter specified above. | 17 |
| Online Security | Boolean | Indicates if the customer subscribes to an additional online security service provided by the company: Yes, No | Yes |
| Online Backup | Boolean | Indicates if the customer subscribes to an additional online backup service provided by the company: Yes, No | No |
| Device Protection Plan | Boolean | Indicates if the customer subscribes to an additional device protection plan for their Internet equipment provided by the company: Yes, No | Yes |
| Premium Tech Support | Boolean | Indicates if the customer subscribes to an additional technical support plan from the company with reduced wait times: Yes, No | No |
| Streaming TV | Boolean | 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. | Yes |
| Streaming Movies | Boolean | 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. | No |
| Streaming Music | Boolean | 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. | Yes |
| Unlimited Data | Boolean | Indicates if the customer has paid an additional monthly fee to have unlimited data downloads/uploads: Yes, No | Yes |
| Contract | String | Indicates the customer’s current contract type: Month-to-Month, One Year, Two Year. | Month-to-Month |
| Paperless Billing | Boolean | Indicates if the customer has chosen paperless billing: Yes, No | Yes |
| Payment Method | Character | Indicates how the customer pays their bill: Bank Withdrawal, Credit Card, Mailed Check | Bank Withdrawal |
| Monthly Charge | Float | Indicates the customer’s current total monthly charge for all their services from the company. | 80.65 |
| Total Charges | Float | Indicates the customer’s total charges, calculated to the end of the quarter specified above. | 39.65 |
| Total Refunds | Float | Indicates the customer’s total refunds, calculated to the end of the quarter specified above. | 45.61 |
| Total Extra Data Charges | Integer | Indicates the customer’s total charges for extra data downloads above those specified in their plan, by the end of the quarter specified above. | 20 |
| Total Long Distance Charges | Float | Indicates the customer’s total charges for long distance above those specified in their plan, by the end of the quarter specified above. | 390.8 |
| Total Revenue | Float | Indicates the total amount of revenue generated from the customer. | 2995.07 |
| Satisfaction Score | Integer | A customer’s overall satisfaction rating of the company from 1 (Very Unsatisfied) to 5 (Very Satisfied). | 3 |
| Customer Status | Character | Indicates the status of the customer at the end of the quarter: Churned, Stayed, or Joined | Churned |
| Churn Label | Boolean | Yes = the customer left the company this quarter. No = the customer remained with the company. Directly related to Churn Value. | No |
| Churn Score | Integer | 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. | 91 |
| CLTV | Integer | 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. | 5302 |
| Churn Category | Character | 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. | Dissatisfaction |
| Churn Reason | Character | A customer’s specific reason for leaving the company. Directly related to Churn Category. | Limited range of services |

#### **Data Cleaning and Preprocessing**

* Missing data value will be removed from the dataset - to protect the purity of our data.
* Outliers will be displayed via ‘Box and Whisker Plots’.
* Data Normalisation via Min-Max Scaling

#### **Exploratory Data Analysis (EDA)**

*Detail the exploratory techniques and visualisations that will be used to understand the data better and uncover initial insights.*

1. Descriptive Statistics

| **Column(s)** | **Statistics** |
| --- | --- |
| * Age (stayed & churned) * Number of Dependents (stayed & churned) * Tenure in Months (stayed & churned) * Satisfaction Score (stayed & churned) * Churn Score (stayed & churned) * CLTV (stayed & churned) * Avg Monthly Long Distance Charges (stayed & churned) * Avg Monthly GB Download (stayed & churned) * Monthly Charge (stayed & churned) * Total Charges (stayed & churned) * Total Refunds (stayed & churned) * Total Extra Data Charges (stayed & churned) * Total Long Distance Charges (stayed & churned) * Total Revenue (stayed & churned) | min, max, mean, median, mode, std. Dev. |
| * Population (stayed & churned) | mean, std. Dev. |

1. Visualisations

| **Column(s)** | **Visualisations** |
| --- | --- |
| * Churn Category | Treemap |
| * Customer Status * Payment Method * Contract * Internet Type | Pie chart |
| * Total revenue avg. * Total Refunds avg. | Bar chart (stayed vs churned) |
| * Internet Service * Phone Service * Referred a Friend * Senior Citizen * Married * Dependents * Gender | Stacked bar chart (stayed vs churned) |
| * Tenure in Months * Number of Referrals * Number of Dependents * Age | Stacked histogram (stayed vs churned) |
| * City | Map chart (stayed vs churned) |

#### 

#### **Tools and Technologies**

| **Tool/Technology** | **Purpose** |
| --- | --- |
| Kaggle | Extract data needed to perform analysis. |
| Google Sheets | Clean the dataset. |
| Python - Jupyter Notebook & Pandas | Analyse the dataset. |
| Tableau Desktop | Visualise data to stakeholders. |
| Google Slides | Share insights and provide recommendations. |

#### **Project Timeline**

| **Task/Milestone** | **Start Date** | **End Date** | **Notes** |
| --- | --- | --- | --- |
| Project Planning | 18-July-2024 | 20-July-2024 | N/A |
| Data Collection | 20-July-2024 | 20-July-2024 | N/A |
| Data Cleaning & Preprocessing | 20-July-2024 | 20-July-2024 | Some columns will still have a blank dataset- depending on the nature of their importance. |
| Exploratory Data Analysis | 20-July-2024 | 20-July-2024 | After the EDA was performed eye-opening insights were discovered. |
| Report & Presentation | 26-July-2024 | 27-July-2024 |  |

#### **Deliverables**

* Cleaned and processed dataset - to remove missing data values and anomalies that could negatively influence the analysis
* Analytical models and results - to help analysts interpret the data in a meaningful way.
* Visualisations and dashboards - to enable stakeholder grasp the data.
* Presentation - providing insights to churn rate and recommendation of how we can reduce it.

#### **Risks and Mitigations**

| **Risk** | **Likelihood (Low/Medium/High)** | **Impact (Low/Medium/High)** | **Mitigation Strategy** |
| --- | --- | --- | --- |
| PII | Low | Medium | Anonymization of customer’s PII. |

#### **Stakeholders**

| **Stakeholder Name** | **Role/Responsibility** |
| --- | --- |
| Me | Data analysis |

#### **Project Dependencies**

N/A

#### **Notes and Assumptions**

* I hypothesise the churn rate is heavily influenced by the pricing of the services provided - by the company.
* ‘Quarter’ column removed - description of data made column redundant.