**Customer Churn Analysis Using Python**

**Introduction**

*i) Context*

Customer churn, or the rate at which customers discontinue a company's subscription services, is a key concern for companies in the subscription-based industry. Understanding why customers churn and identifying associated patterns are crucial for companies striving to enhance customer retention and profitability. This report presents a comprehensive **customer churn analysis** conducted using Python, leveraging data science tools and techniques to reveal insights into churn behaviour.

ii)*Objective*

The goal of this project is to interpret patterns that contribute to higher customer churn rates and come up with recommendations to help reduce this phenomenon from occurring.

**Data Description**

The dataset has 64,374 records of customer information containing the following columns:

| **Column** | **Description** |
| --- | --- |
| Customer ID | Unique identifier for each customer |
| Age | Age of the customer |
| Gender | Gender of the customer |
| Tenure | The duration of the customer’s subscription, relative to their contract length |
| Usage Frequency | Monthly usage frequency |
| Support Calls | Number of support calls made by the customer |
| Payment Delay | Number of delayed payments |
| Subscription Type | Subscription plan(Basic, Premium, Standard) |
| Contract Length | Contract type(Monthly, Quarterly, Yearly) |
| Total Spend | Total amount spent by the customer |
| Last Interaction | Months since last interaction |
| Churn | Churn status( 1 = Churned, 0 = Retained) |

By exploring these variables, we aim to uncover patterns and identify potential indicators of churn.

**Data Cleaning and Preprocessing**

We started by importing the relevant Python libraries in Jupyter Notebook to conduct our analysis(pandas,numpy, matplotlib and seaborn). Then we loaded the dataset, which is a zip file. Afterwards, we checked the overall characteristics of the dataset, such as the number of rows and columns and their respective names, the datatypes of each column, the first five records and the last five records of the dataset. We followed by inspecting for missing values and duplicates(none were found), and finally running the descriptive statistics of the dataset.

**Exploratory Data Analysis(EDA)**

i)*Univariate Analysis*

* Usage Frequency Distribution: The usage frequency of customers is normally distributed, as evidenced by approximately equal mean and median values shown by the box plot.
* Gender Distribution: The company has more female customers compared to male ones. Female customers comprise 53.4% of the total customers, and male customers consist of 46.4% of the total customers.
* Distribution of Tenure: The histogram representing the distribution of tenure is right-skewed, meaning that a lot of customers are new, or churn is higher among newer customers.
* Distribution of customers by age: The majority of customers are between 50 and 59 years old.
* Churn Distribution: The number of active customers(33881) is slightly more than that of churned customers(30493). This is a major cause for concern.
* Distribution of Subscription Types: Most customers subscribe to the Standard Package. There are notably slight differences in the packages customers subscribe to(Standard-21502, Basic-21451, Premium-21421).

ii)*Multivariate Analysis*

* Subscription Type vs Churn: We discovered that most of the customers subscribed to the Premium package, whereas the basic package was the least subscribed to.
* Contract Length vs Churn: The resulting analysis shows that customers who subscribed on a monthly basis churned the most(11421). Conversely, the majority of the retained customers subscribed on a quarterly basis(11657).
* Age Group vs Churn: The customer base mainly consists of individuals within the age range of 50 and 59 years. There are very few customers aged between 18 and 19 years, assuming that an individual must be of legal age to subscribe to the various packages on offer.
* Gender vs Churn: Female customers churned the most compared to male customers.
* Correlation Matrix: From the matrix, it is observed that payment delay has the highest correlation to churn(0.56). Last interaction has the least correlation to churn(-0.00)

**Feature Engineering**

We created a new feature, ‘Adjusted Tenure’, to standardize the duration of the customers’ subscriptions to months for easier analysis and derivation of insights. We also created new variables, namely: ‘churn\_data’ and ‘contract\_churn’ to conveniently create bar charts to analyze the relationships between ‘Subscription Types’ and ‘Churn’, and ‘Contract Length’ and ‘Churn’ respectively. Additionally, we binned the ages into groups to identify key insights in the different age groups.

**Key Insights**

* The majority of the customers subscribing to the company’s various packages are aged between 50 and 59 years.
* Female customers churned the most from the company’s services, with a higher percentage discontinuing their subscriptions compared to those that were retained.
* Customers who had a higher number of delayed payments had the highest probability of churning from the company’s services.
* Customers with a high number of support calls were more likely to churn.
* More customers subscribed on a quarterly basis than on a yearly basis.
* The percentage of customers who have churned(47.4%) is almost similar to the percentage of customers that have been retained(52.6%), which is a huge cause for concern.

**Recommendations**

* Tailor-make services across the various packages to suit the needs and preferences of female customers.
* Increase the benefits across all the subscription packages to lure new customers and raise their chances of retaining the current ones.
* Act swiftly on support calls made by customers to ensure that the same problems are not repeated in the future.
* Offer discounts on yearly and quarterly payment plans to entice customers to subscribe to those payment terms and thus increase their chances of continuing to subscribe to the services offered in the long run.
* Segment the customers based on age groups and thus tailor their services to suit the needs and preferences of individuals across all age groups.

*For further details and a comprehensive breakdown of the data, refer to the full analysis* [*here.*](http://localhost:8888/notebooks/customer_churn_analysis.ipynb)