* **Project Title**: Unraveling Customer Churn for a Fictious Telecom Company
* **Team Members**: Gayatri Shankar Shinde
* **Introduction:** In the dynamic landscape of telecommunications, minimizing customer churn is crucial for sustaining business growth and profitability. To address this imperative, I propose the development of a sophisticated churn prediction system, harnessing the power of machine learning techniques. By analyzing extensive historical customer data, our objective is to predict potential churners, empowering proactive retention strategies to be implemented effectively.

**Project Objective:** This project seeks to develop a robust churn prediction system tailored specifically to the telecommunications industry. Beyond its practical implications, we aim to delve into the inner workings of logistic regression, unraveling its fundamental mechanisms, and extend our understanding to the intricacies of multivariate linear regression.

**Dataset: Description and Summary:** The customer churn data sourced by the IBM Developer Platform. It includes a target label indicating whether or not the customer left within the last month, and other dependent features that cover demographics, services that each customer has signed up for, and customer account information.

|  |  |
| --- | --- |
| Number of rows: 7043, Number of features: 21, Missing values: 0 | |
| **There are 17 categorical features** | **Unique** |
| CustomerID: Customer ID unique for each customer | 7043 |
| gender: Whether the customer is a male or a female | 2 |
| SeniorCitizen: Whether the customer is a senior citizen or not (1, 0) | 2 |
| Partner: Whether the customer has a partner or not (Yes, No) | 2 |
| Dependent: Whether the customer has dependents or not (Yes, No) | 2 |
| PhoneService: Whether the customer has a phone service or not (Yes, No) | 73 |
| MultipeLines: Whether the customer has multiple lines or not (Yes, No, No phone service) | 2 |
| InternetService: Customer’s internet service provider (DSL, Fiber optic, No) | 3 |
| OnlineSecurity: Whether the customer has online security or not (Yes, No, No internet service) | 3 |
| OnlineBackup: Whether the customer has an online backup or not (Yes, No, No internet service) | 3 |
| DeviceProtection: Whether the customer has device protection or not (Yes, No, No internet service) | 3 |
| TechSupport: Whether the customer has tech support or not (Yes, No, No internet service) | 3 |
| StreamingTV: Whether the customer has streaming TV or not (Yes, No, No internet service) | 3 |
| StreamingMovies: Whether the customer has streaming movies or not (Yes, No, No internet service) | 3 |
| Contract: The contract term of the customer (Month-to-month, One year, Two years) | 3 |
| PaperlessBilling: The contract term of the customer (Month-to-month, One year, Two years) | 3 |
| PaymentMethod: The customer’s payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic)) | 2 |
| **Next, there are 3 numerical features** | **Unique** |
| Tenure: Number of months the customer has stayed with the company | 4 |
| MonthlyCharges: The amount charged to the customer monthly | 1585 |
| TotalCharges: The total amount charged to the customer | 6531 |
| **Finally, there’s a prediction feature** | **Unique** |
| Churn: Whether the customer churned or not (Yes or No) | 2 |
| Number of records labeled as 1: 1869 Number of records labeled as 0: 5174 | |

**Machine Learning Tools**: For this project, we plan to utilize various machine learning tools and techniques. We will explore algorithms such as logistic regression, decision trees classifier, random forests classifier. Additionally, we may employ techniques like feature engineering, hyperparameter tuning, and ensemble methods to enhance model performance. I am using the following Libraries: Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn.

**How the Data Set Helps:** The dataset serves as the foundation for our churn prediction system. By examining features such as customer demographics, usage patterns, and subscription details, we can identify potential predictors of churn. These predictors will be used to train machine learning models, which will then be deployed to predict future churn behavior. The churn prediction system will accurately identify customers at high risk of churning, enabling targeted retention efforts. Insights gained from the predictive models will inform the development of personalized retention strategies tailored to specific customer segments. By reducing churn rates and increasing customer retention, the telecommunications company will experience improved customer satisfaction, revenue stability, and long-term growth prospects.