# Project Proposal
## Group 2

<u>NAMES:</u>  Aida Karimu, Jingxin Yao, Sunita Barik, Zihan Huang
<u>GROUP NUMBER:</u> 2
<u>CHOSEN CASE STUDY:</u> Predicting Customer Churn

<u>MOTIVATION FOR CHOOSING THE BUSINESS CASE/DATA:</u>
Telecommunication companies often use the analysis of customer churn or churn rates as one of their key business metrics. It costs far less to retain an existing customer than it does to acquire a new one. Companies in these industries often have customer service departments that try to win back defecting customers because recovered long-term customers are more valuable to the company than newly recruited ones.
The predictive analytics model is to predict the churn by assessing relevant business variables, such as contract type, payment method and monthly charge. The dataset has right balance of categorical and numerical variables and has missing data which will give us an opportunity to do preprocessing, missing data analysis, imputation, and prediction.

<u>DIMENSIONALITY AND VARIABLE IDENTIFICATION:</u>
**No. of Rows:** 7043 **No. of Column:** 21
**No. of Categorical:** 18 **No. of Numerical:** 3
**Missing Data Percentage:** 0.15%
**Target Variable:** Churn
**Predictor Variables:** Gender, SeniorCitizen, MonthlyCharges, TotalCharges, tenure, InternetService, MultipleLines, Partner, Dependents, Phone Service, Contract, PaymentMethod, StreamingTV, StreamingMovies, OnlineSecurity, OnlineBackup.

| # | Attribute Name | Data Type | Missing Data | Description |
|---|---|---|---|---|
| 1 | customerID | Nominal | No | unique customer identifier ( Does not provide any insight ) |
| 2 | gender | Nominal | No | gender of customer, |
| 3 | SeniorCitizen | Nominal | No | indicates if a customer is a senior citizen (1) or not (0) |
| 4 | Partner | Nominal | No | Indicates if the customer has a partner (Yes) or not (No) |
| 5 | Dependents | Nominal | No | Indicates if the customer has dependents (Yes) or not (No) |
| 6 | tenure | Numerical | No | the length of time that the customer has been a customer |
| 7 | PhoneService | Nominal | No | Indicates if the customer has phone service with the company (Yes) or not (No) |
| 8 | MultipleLines | Nominal | No | Whether the customer has multiple lines or not (Yes, No, No phone service) |
| 9 | InternetService | Nominal | No | Indicates if the customer has fiber optic, DSL or no internet service with the company |
| 10 | OnlineSecurity | Nominal | No | Whether the customer has online security or not (Yes, No, No internet service) |
| 11 | OnlineBackup | Nominal | No | Whether the customer has online backup or not (Yes, No, No internet service) |
| 12 | DeviceProtection | Nominal | No | Whether the customer has device protection or not (Yes, No, No internet service) |
| 13 | TechSupport | Nominal | No | Whether the customer has tech support or not (Yes, No, No internet service) |
| 14 | StreamingTV | Nominal | No | Whether the customer has streaming TV or not (Yes, No, No internet service) |
| 15 | StreamingMovies | Nominal | No | Whether the customer has streaming movies or not (Yes, No, No internet service) |
| 16 | Contract | Ordinal | No | The type of contract that the customer has with the company (Month-to-month, One year,Two year) |
| 17 | PaperlessBilling | Nominal | No | If the customer is enrolled in paperless billing (Yes) or not (No) |
| 18 | PaymentMethod | Nominal | No | The most recent payment method used by the customer to pay the company (Electronic check, Mailed check, Bank transfer (automatic), or Credit card (automatic) |
| 19 | MonthlyCharges | Numerical | No | The most recent amount that the customer is charged per month |
| 20 | TotalCharges | Numerical | Yes( 11 ) | The total amount that the customer has been charged |
| 21 | Churn | Nominal | No | Whether the customer has left the company (Yes) or not (No) ( Target Variable ) |

<u>INITIAL DATA EXPLORATION:</u>
**Missing value rows:**  "489" "754" "937" "1083" "1341" "3332" "3827" "4381" "5219" "6671" "6755"
**Variance:** tenure- 603.1681, MonthlyCharges- 905.4109

```
> summary(churn)
  customerID          gender          SeniorCitizen Partner    Dependents    tenure        PhoneService        MultipleLines
 Length:7043        Female:3488      0:5901         No :3641   No :4933   Min.   : 0.00   No : 682       No               :3390
 Class :character   Male  :3555      1:1142         Yes:3402   Yes:2110   1st Qu.: 9.00   Yes:6361       No phone service: 682
 Mode  :character                                              Median :29.00                  Yes              :2971
                                                               Mean   :32.37
                                                               3rd Qu.:55.00
                                                               Max.   :72.00

      InternetService        OnlineSecurity            OnlineBackup             DeviceProtection          TechSupport
 DSL        :2421   No                 :3498   No                 :3088   No                 :3095   No                 :3473
 Fiber optic:3096   No internet service:1526   No internet service:1526   No internet service:1526   No internet service:1526
 No         :1526   Yes                :2019   Yes                :2429   Yes                :2422   Yes                :2044



       StreamingTV             StreamingMovies          Contract     PaperlessBilling           PaymentMethod
 No                 :2810   No                 :2785   Month-to-month:3875   No :2872   Bank transfer (automatic):1544
 No internet service:1526   No internet service:1526   One year      :1473   Yes:4171   Credit card (automatic)  :1522
 Yes                :2707   Yes                :2732   Two year      :1695              Electronic check         :2365
                                                                                        Mailed check             :1612


 MonthlyCharges    TotalCharges      Churn
 Min.   : 18.25   Min.   :  18.8   No :5174
 1st Qu.: 35.50   1st Qu.: 401.4   Yes:1869
 Median : 70.35   Median :1397.5
 Mean   : 64.76   Mean   :2283.3
 3rd Qu.: 89.85   3rd Qu.:3794.7
 Max.   :118.75   Max.   :8684.8
                  NA's   :11
>
```