**Table of Contents.**

[1 Introduction 1](#_Toc414111172)

[2 Challenge. 1](#_Toc414111173)

[3 Your role. 1](#_Toc414111174)

[4 Details of the data 1](#_Toc414111175)

[5 Assignment 3](#_Toc414111176)

# Introduction

The Client is a Telecom company. They have an issue with customer churn. Currently, they want to improve their customer retention efforts. For this, they have engaged on a churn prediction exercise.

# Challenge.

The customer has stated that the objective of the churn prediction exercise is to predict churn. That is –

“The model should be able to identify (based on past data), who is going to churn in the next time period”.

# Your role.

You are the lead data miner for the project. In this phase, you have to confirm that you understand the business requirements (business understanding) as well as the data (data understanding) for the project.

# Details of the data

The customer has shared the data with your team.

The data is in the form of a text file “telco.txt”. The description of the various fields are given below.



The customer has stated that the target variable is “churn”.

**Data Description**

The data is described below

| **SN** | **Name** | **Description** |
| --- | --- | --- |
| 1 | Region | Geographic indicator |
| 2 | Tenure | Months with service |
| 3 | Age | Age in years |
| 4 | Marital | Marital status |
| 5 | Address | Years at current address |
| 6 | Income | Household income in thousands |
| 7 | Ed | Level of education |
| 8 | Employ | Years with current employer |
| 9 | Retire | Retired |
| 10 | Gender | Gender |
| 11 | Reside | Number of people in household |
| 12 | Tollfree | Toll free service |
| 13 | Equip | Equipment rental |
| 14 | Callcard | Calling card service |
| 15 | Wireless | Wireless service |
| 16 | Longmon | Long distance last month |
| 17 | Tollmon | Toll free last month |
| 18 | Equipmon | Equipment last month |
| 19 | Cardmon | Calling card last month |
| 20 | Wiremon | Wireless last month |
| 21 | Longten | Long distance over tenure |
| 22 | Tollten | Toll free over tenure |
| 23 | Equipten | Equipment over tenure |
| 24 | Cardten | Calling card over tenure |
| 25 | Wireten | Wireless over tenure |
| 26 | Multline | Multiple lines |
| 27 | Voice | Voice mail |
| 28 | Pager | Paging service |
| 29 | Internet | Internet |
| 30 | Called | Caller ID |
| 31 | Callwait | Call waiting |
| 32 | Forward | Call forwarding |
| 33 | Confer | 3-way calling |
| 34 | Ebill | Electronic billing |
| 35 | Loglong | Log-long distance |
| 36 | Logtoll | Log-toll free |
| 37 | Logequi | Log-equipment |
| 38 | Logcard | Log-calling card |
| 39 | Logwire | Log-wireless |
| 40 | Lninc | Log-income |
| 41 | Custcat | Customer category |
| 42 | Churn | Churn within last month |

# Assignment

**Exercise 1.**

On the basis of the problem statement and the data provided, carry out the following steps in CRISP DM for the client –

1. Business Understanding
2. Data Understanding

The response for this exercise will be a report in the form of a word document, with separate sections for the deliverables below (refer to the lecture on CRISP-DM Methodology).

|  |  |
| --- | --- |
| **Stage** | **Deliverables** |
| **Business Understanding** | Data Mining Goals |
|  | Data Mining Success Criteria |
| **Data Understanding** | Data Exploration Report |
|  | Data Quality Report |

**Exercise 2.**

Perform initial analysis to identify possible variables that could be impacting churn. Provide a report of the same.