# Section A

## Part 1: Research Question

The dataset consists of AnyTelecom’s observations on customer demographics, their telecom services, and whether they have churned (dropped AnyTelecom’s services within the past month). Which customers are at high risk of churn? To answer this question, I will use Naïve Bayes classification.

## Part 2: Analytical Goal

Build a machine learning model that classifies a customer as either “high” or “low” risk of churn.

# Section B

## Part 1: Classification Method

Naïve Bayes classification uses the probability of observing predictor variables to estimate the probability of observing the outcome Y=i, given a set of predictor variables. The expected outcome is a binary response that assigns a record to the class with the highest probability for a given set of predictor values (Bruce, Bruce, & Gedeck, 2019).

## Part 2: Assumption

One assumption of Naïve Bayes classification is that the method only works with categorical variables. To use continous variables with Naïve Bayes, continous variables must be converted to categorical variables (Bruce, Bruce, & Gedeck, 2019).

## Part 3: Packages

The following packages will be used to support the analysis:

[TABLE]

# Section C

## Part 1: Preprocessing Goal

Recall in Section B, Part 2 that continous variables must be converted to categorical variables for use in a Naïve Bayes classifier. One preprocessing goal is to convert continous predictor variables into categorical variables using the technique of binning.

## Part 2: Variables

Data descriptions for all predictor variables used in the analysis, including whether a particular variable is considered continuous or categorical, can be viewed in Section F, Part 1.

## Part 3: Steps

The data was prepared according to the following steps:

1. Data quality issues were identified.
2. Data quality issues were remediated.
3. Continuous variables were binned and converted to categorical variables.
4. All predictor variables were encoded.
   1. Binary variables were encoded using ordinal encoding.
   2. All other variables were encoded using one-hot encoding.

To view the associated data preparation code, refer to Section F, Part 1.

## Part 4: Preprocessing Output

Refer to the attachment “churn prepped1.csv” to view the output of the data cleaning process.

# Section D

## Part 1: Train/Test Split Output

## Part 2: Analysis

## Part 3: Code Output

# Section E

## Part 1: Model Evaluation

## Part 2: Results

## Part 3: Limitation

## Part 4: Recommendation

# Section F

## Part 1: Data Preparation Code

## Part 2: Classification Model Code

# Section G

## Part 1: Demonstration

# Section H

## Part 1: Web Sources

## Part 2: References