# BAN130

# NBB Section

# Project part 1-P1


Group Members:

1. Name - Udbhav Singh Chauhan

   Student ID - 146782206

   E-Mail ID - uschauhan@myseneca.ca

2. Name - Pruthviben Jamin Patel

   Email ID - pjpatel44@myseneca.ca

   Student ID – 139602205


1. What is the size of the dataset you have chosen for **your project**?

   The dataset that we are going to use is titanic dataset.
   There are 891 data values and 12 attributes/features of the dataset.

   **SAS Code:**

```
Proc print data=titanic;
Run;

Proc contents data=titanic;
Run;
```

2. What are the variables (description)used in the dataset?

   Variables:
   survival - Survival (0 = No; 1 = Yes)
   class - Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
   name - Name
   sex - Sex

age - Age
sibsp - Number of Siblings/Spouses Aboard
parch - Number of Parents/Children Aboard
ticket - Ticket Number
fare - Passenger Fare
cabin - Cabin
embarked - Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)
boat - Lifeboat (if survived)
body - Body number (if did not survive and body was recovered)

**SAS Code:**

```
Proc means data=titanic;
Run;

Proc univariate data=titanic;
Run;
```

3. In each dataset what do you think which variable will be used as a predicting variable for any business analytics project?

   In the dataset "Titanic" that we have chosen for our project, we have variable "survived" which will be our predicting variable as it will be deciding the outcome of our business analytic project

4. Write 4 types of questions that you can extract from the dataset that adds business value in the project.

   Using the following questions, we would be able to analyze the data and we would be able to predict what are the chances of survival of the passenger.

   What is the gender of passenger who survived?
   What is the age of the passenger who survived?
   What was the class of the passenger who survived?
   How many People had children who survived?

5. Define a proposal using the chosen dataset.

   The following steps we would perform:

   1. Dataset Description and Metadata
      a. First we will check what are different variables in the dataset.
      b. Which types of variables are there in the dataset.
      c. What is the size of the dataset.
      d. We will write Metadata for each attribute in order to understand the data and types of values.

   2. Data preprocessing and cleaning
      a. We will check each attribute for missing values.
      b. We will check for invalid values.
      c. We will try to fill-in the values.

   3. Analysis of attributes
      a. We will take each variable and perform statistical analysis.
      b. We will check min, max, mean, median, mode and frequency for the attributes.
   4. Visualization
      a. Here we will see how the attributes are corelated.
      b. Attributes have what types of values and how they have impact on the dataset using different types of visualizations.
      c. Which factors has more impact on the survival rate.
   5. Conclusion
      a. Here, we will conclude our project, that what are the chances of survival based on different attributes and how we can predict the survival of passengers.