CPSC 4800 - Assignment 3

## Part – 2

Niha Siddikha Sachin  
100374481

**Purpose:**

For this assignment, we will be working on the Titanic Dataset from Kaggle which is a very famous dataset to perform some basic exploratory data analysis.

**Data Cleaning:**First we import the dataset into a panda dataframe called ‘titanic’.

Text, application

Description automatically generated

It can be inferred that the dataset has 891 rows with 12 columns and the columns ‘Age’ and ‘Cabin’ have the most null values. In order to clean the data, ‘Cabin’ column was dropped and dropna() was performed on the Titanic dataset.

**Analysis:**‘Survived =1’ represents the passengers who survived and ‘Survived = 0’ represents passengers who did not survive.

Chart, bar chart

Description automatically generated

From the graph, we can see that there were a greater number of passengers who did not survive (more than 500) compared to those who did (approx. 350).

Chart, histogram

Description automatically generated

**Sex:** From the passengers who survived, more were female passengers and from the passengers who did not survive, more were male passengers. Therefore, survival rate is associated to the gender.

Chart, histogram

Description automatically generated

**Pclass:** The highest number of passengers who did not survive belonged to the third class and passengers belonging to first class are more likely to survive. Therefore, survival rate is associated to the class of passenger.

Let us analyse further whether the sex and passenger class combined impact the survival rate of passengers.

Chart

Description automatically generated

With the above graph, we have a clearer understanding of whether passengers survived based on their sex and passenger class as the highest survivors were females from the upper class (1 and 2) and the highest numbers of passengers who did not survive were males belonging to the lower class.

Chart, histogram

Description automatically generated

**Age:** We can infer that majority number of passengers travelling onboard the Titanic belong to the age group between 20-40 and there are very few passengers belonging to the age group 70-80.

Chart, box and whisker chart

Description automatically generated

The box plot shows that there were higher number of older passengers in the first class as compared to the second and third class with a few exceptions or outliers.