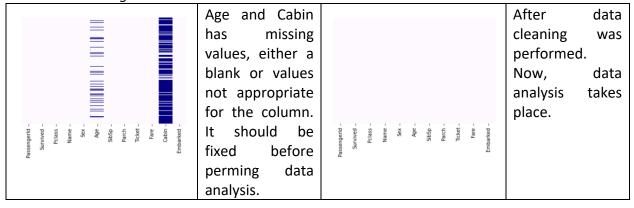Akmad, Mohammad Nur
CPSC 4800 – W01

Assignment #3 Part 2 – Titanic Data Set
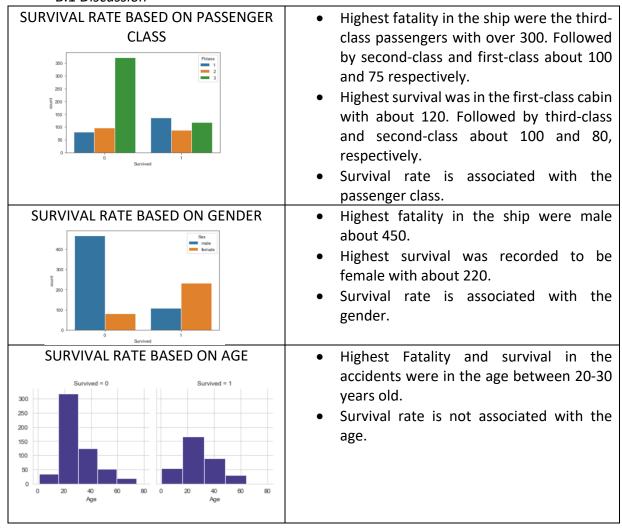
A. Overview

One of the unforgettable trategies in the world was the sinking of RMS Titanic in 1912 due to collision of its ship in one of the ice bergs. The information about the passengers is one of the famous data sets that can be downloaded in Kaggle. This information is examined to determine the factors that can affect the survival rate of the passengers. The objective of this study is to determine if: (1) survival rate is associated to the class of passenger; (2) survival rate is associated to the gender and (3) survival rate is associated to the age.

B. Findings

B.1 Missing Data

	Age and Cabin has missing values, either a blank or values not appropriate for the column. It should be fixed before perming data analysis.		After data cleaning was performed. Now, data analysis takes place.

B.1 Discussion

SURVIVAL RATE BASED ON PASSENGER CLASS	• Highest fatality in the ship were the third-class passengers with over 300. Followed by second-class and first-class about 100 and 75 respectively. • Highest survival was in the first-class cabin with about 120. Followed by third-class and second-class about 100 and 80, respectively. • Survival rate is associated with the passenger class.
SURVIVAL RATE BASED ON GENDER	• Highest fatality in the ship were male about 450. • Highest survival was recorded to be female with about 220. • Survival rate is associated with the gender.
SURVIVAL RATE BASED ON AGE	• Highest Fatality and survival in the accidents were in the age between 20-30 years old. • Survival rate is not associated with the age.

C. Results

Therefore, we can conclude that the class type and gender of the passenger is highly associated with the survival rate. On the other hand, age does not suggest any association with regards to the survival rate.