Yin Htoo Htoo Naing

Student ID: 100366895

CPSC-4800-W01

Instructor: Nasim

11 Mar 2022

Introduction

The purpose of this report to perform exploratory data analysis and answer the following three hypotheses based on the Titanic dataset.

The following is how I observe the variables and data types of this dataset and the first five observations.

Table

Description automatically generated

A picture containing graphical user interface

Description automatically generated

Determine if the survival rate is associated to the class of passenger

Chart, bar chart

Description automatically generated

Determine if the survival rate is associated to the gender

Chart, pie chart

Description automatically generated

Determine the survival rate is associated to the age

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Correlation

Here is the Correlation matrix between numerical variables.

Table

Description automatically generated

Based on the correlation coefficient -0.338481 between Pclass and Survived, we can interpret that there is weak linear relationship. From the pearson correlation test, p-value is

(-0.3384810359610166, 2.537047387977477e-25)

Based on the correlation coefficient -0.069809 between Age and Survived, we can interpret that there is weak linear relationship.