Yin Htoo Htoo Naing

Student ID: 100366895

CPSC-4800-W01

Instructor: Nasim

11 Mar 2022

Table of Contents

[Introduction 2](#_Toc97738395)

[Determine if the survival rate is associated to the class of passenger 3](#_Toc97738396)

[Determine if the survival rate is associated to the gender 3](#_Toc97738397)

[Determine the survival rate is associated to the age 4](#_Toc97738398)

[Correlation 5](#_Toc97738399)

# 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

Fig: Stacked bar chart for Pclass and Survived

According to the bar chart here, the most number of people who are survived from the Titanic, are from class 1, followed by class 2 and 3.

# Determine if the survival rate is associated to the gender

Chart, pie chart

Description automatically generated

Fig: Pie chart for Sex(Gender) of Survival

According to the pie chart comparing the gender, more female passenger survived than the male passenger from Titanic. Based on the calculation in python, 74% of the female passenger survived while only 18% of the male passenger survived in Titanic.

# Determine the survival rate is associated to the age

Chart, bar chart

Description automatically generated

Fig: Stacked bar chart for Age and Survived

According to the stacked bar chart showing the number people who survived or unsurvived in each age group, most people who survived are from the age group 20-29 from Titanic. The most passengers from the Titanic are from the age group 20-29 as well.

Chart, bar chart

Description automatically generated

Fig: bar chart showing the proportion of survivors by age group/ Sex (Gender)

The average age of the children(below 18 years old) who are survived: 9 years old in female passengers and 8 years old in male passengers.

Table

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 a weak linear relationship.

Based on the correlation coefficient -0.543351 between Sex and Survived, we can interpret that there is a moderate linear relationship.

Based on the correlation coefficient -0.077221 between Age and Survived, we can interpret that there is a weak linear relationship.