**INFO 5810 Assignment -2**

The dataset used in this assignment is TitanicPassengers.xlsx.

There are 1309 rows and 15 columns. The name and datatype of each column are:

Passenger Class (integer),

Survived (integer),

Name (nominal),

Sex (nominal),

Age (real),

Siblings and Spouses (integer),

Parents and Children (integer),

Ticket# (Nominal),

Fare (real),

Cabin (nominal),

Port (nominal),

Lifeboat (nominal), and

Body (integer).

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Data imported successfully.

1.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Missing values in Age: 263 values.

A screenshot of a computer

Description automatically generated

Missing values in other rows:

Fare (1), Cabin (1014), Port (2), Lifeboat (823), Body (1188), Home/Destination (564).

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Remove Duplicates: There are no duplicate values in the dataset.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

2.

A screenshot of a computer

Description automatically generated

Count of survived is 500 and not survived is 809 and the proportion of survived is:

500 / (500+809) = 0.381 = 38.1%

People are in range from 1 month to 80 years and maximum in the range of 20-30 years and count is 361.

A screenshot of a graph

Description automatically generated

The distribution of Age Attribute:

A screenshot of a computer

Description automatically generated

The distribution of Survived Attribute:

A blue graph on a white background

Description automatically generated

3.

A screenshot of a computer

Description automatically generated

The number of passengers in first class is 323. Total passengers count 1309.

The proportion of passengers in first class is 323/1309 = 0.246 is 24.6%

A screenshot of a computer

Description automatically generated

The most common Home/Destination of passengers is New York, NY. Count is 64.

4.

I have taken **Survived** attribute and changed it into **category** and selected inputs as **sex, age and Passenger Class** to build a **logistic regression** using turbo prep tab.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The AUC Score is 82.4%.

A screenshot of a computer

Description automatically generated

The chance of Female baby i.e., less than 10 years old and survived on travelling in First class is 84%.

A screenshot of a computer

Description automatically generated

The chance of Female baby i.e., less than 10 years old and survived on travelling in Third class is 65%.

A screenshot of a computer

Description automatically generated

The chance of Male i.e., 40 years old and survived on travelling in First class is 46%.

A screenshot of a computer

Description automatically generated

The chance of Male i.e., 40 years old and survived on travelling in Third class is 26%.

A screenshot of a computer

Description automatically generated

According to previous data, passengers in first class, regardless of gender, had a considerably greater chance of survival than those in other classes. Furthermore, women have a much higher survival rate than men, regardless of class. In my perspective, the Titanic disaster spared a significant number of ladies. When all male passengers in all classes are taken into account, a 40-year-old male passengers in third class has an exceedingly low chance of surviving.

5.

A screenshot of a computer

Description automatically generated

From above image we can say that female has the high survival rate and more over compared to other classes First class has the high survival rate. Passengers in third class of age above 60 has the less survival rate.

6.

A screenshot of a computer

Description automatically generated

Auc value generated by Decision Tree is 80.1%.

A screenshot of a computer

Description automatically generated

Sex attribute as the highest weight as shown in below screenshot.

A screenshot of a computer

Description automatically generated