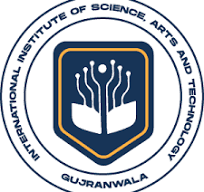
**Assignment no 4**

****

**Name: Khadija Anwar**

**Roll no:0671**

**Semester: 4th**

**Department: IT**

**Submitted to: Sir Zubair**

**Course Title: Artificial intelligence**

**Department of Information Technology**

**International Institute of Science, Arts and Technology**

**(IISAT), Gujranwala**

**Assignment 4**

**Data Analysis and Visualization Using Pandas**

Objective:

The objective of this assignment is for students to use a data set to perform data analysis and visualization using Pandas, and explain the data set in detail.

**Instructions:**

**Data Set Selection:**

The Titanic dataset from Kaggle will be used for this assignment.

**Data Loading:**

Load the selected data set into a Pandas Data Frame.

**Data Exploration:**

Load the selected data set into a Pandas Data Frame.

**Data Cleaning:**

Clean the data by handling missing values, duplicates, and performing any necessary data transformations.

**Data Visualization:**

Use Pandas built-in visualization capabilities along with Matplotlib and Seaborn to create various graphs and charts.

**Analysis and Insights:**

After each visualization, provide an analysis and the insights you derived from it.

**Survival Count:**

The count plot of survival shows that more passengers did not survive compared to those who did.

**Distribution of Age:**

The age distribution plot shows that most passengers were between 20 and 40 years old, with fewer older passengers. This information helps us understand the age demographics of the passengers.

**Count of Passengers by Embarked:**

Most passengers embarked from port 'S' (Southampton), followed by 'C' (Cherbourg) and 'Q' (Queenstown). This gives us an idea of the distribution of passengers based on their boarding locations.

**Survival Rate by Pclass:**

Passengers in the 1st class had a higher survival rate compared to those in the 2nd and 3rd classes. This suggests that socio-economic status played a significant role in survival, with higher-class passengers having better access to lifeboats.

**Survival Rate by Sex:**

Females had a significantly higher survival rate compared to males, indicating that women were given priority during the evacuation.

**Survival Rate by Embarked:**

Passengers who embarked from 'C' (Cherbourg) had a higher survival rate compared to those from 'S' (Southampton) and 'Q' (Queenstown). This could be due to various factors, including socio-economic status and lifeboat access.

**Survival Rate by Age:**

Younger passengers, especially children, had a slightly higher survival rate. This suggests that younger individuals were also given priority during the evacuation.

**Survival Rate by Fare:**

Passengers who paid higher fares had a higher survival rate, indicating that those in higher economic classes had better chances of survival.

By following these steps, you will have performed a comprehensive data analysis and visualization on the Titanic dataset using the training data from Kaggle. This will help you understand the structure, characteristics, and insights derived from the data.