The problem statement for working on EDA on the Titanic dataset is to explore and analyze the dataset to gain insights and understanding of the factors that affected the survival of passengers on the Titanic.

Classification Problem: <https://www.kaggle.com/c/titanic>

Solution :

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

titanic\_df = pd.read\_csv('titanic.csv')

titanic\_df.head()

titanic\_df.describe()

titanic\_df.info()

sns.histplot(titanic\_df['Age'], kde=False)

sns.scatterplot(data=titanic\_df, x="Age", y="Fare", hue="Survived")

sns.boxplot(data=titanic\_df, x="Survived", y="Age")

titanic\_df.groupby(['Sex'])['Survived'].mean()

titanic\_df.groupby(['Pclass'])['Fare'].median()

Conclusions:

- Women were more likely to survive than men.

- Passengers in first class had a higher survival rate than those in second or third class.

- The average age of passengers was around 29 years old.