

ASSIGNMENT

Problem Statement:

You are given a dataset containing information about the passengers of the Titanic. Your task is to perform exploratory data analysis (EDA) on the dataset using the following methods:

- Load the dataset(Titanic.csv) into a pandas dataframe and print the first 5 rows using the head() method.
- Use the info() method to get information about the dataset. In particular, pay attention to the data types of each column and whether there are any missing values.
- Use the describe() method to describe the dataset. Pay attention to the range of values for each numeric column and whether there are any outliers.
- Use the groupby() method to group the data based on the Pclass column and calculate the mean fare for each class.
- Use the value_counts() method to get the frequency count of the Embarked column.
- Create a new column in the dataframe called AgeRange that categorises passengers into age ranges: "Child" for ages 0-12, "Teen" for ages 13-19, "Adult" for ages 20-59, and "Senior" for ages 60 and above.
- Use the pivot_table() method to create a pivot table that shows the survival rate of passengers based on their sex, class, and age range.
- Create a bar chart that shows the total number of passengers in each age range.
- Create a scatter plot that shows the relationship between age and fare. Color the points based on whether the passenger survived or not.

Submission:

Create a Jupyter notebook and include the code for each task along with comments explaining what each line of code does. You should also include the necessary imports at the beginning of the notebook.

Use Following Data set-

https://drive.google.com/file/d/1adqfsBEIj2FPpORWoPrzIvmX9j7y_O4m/view?usp=drive_link