TASK 02

Perform data cleaning and exploratory data analysis (EDA) on a dataset of your choice, such as the Titanic dataset from Kaggle. Explore the relationships between variables and identify patterns and trends in the data.

Titanic Dataset

jupyter SCT_DS_2 Last Checkpoint: 1 hour ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

TASK 2 - BY Gaurav Singh Yadav

What was the task : - Perform data cleaning and exploratory data analysis (EDA) on a dataset of your choice, such as the Titanic dataset from Kaggle. Explore the relationships between variables and identify patterns and trends in the data. Titanic Dataset

Step 1: download requried libaries.

[ ]:
```
%pip install pandas
%pip install numpy
%pip install matplotlib
%pip install seaborn
%pip install plotly
```

Step 2: Data Loading and Initial Exploration

[21]:
```
# Import essential libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

Q.3 what is the distributation of passenger classes and Embarked?

[8]:
```
passenger_class = train['Pclass'].value_counts()
print("Passenger class distribution:")
print(passenger_class)
```

```
Passenger class distribution:
Pclass
3    491
1    216
2    184
Name: count, dtype: int64
```

[9]:
```
embarked_counts = train['Embarked'].value_counts()
print("passengers embarked from each port:")
print(embarked_counts)
```

```
passengers embarked from each port:
Embarked
S    646
C    168
Q     77
Name: count, dtype: int64
```

Q.4 what is the number of survived passengers?

[10]:
```
survived_counts = train['Survived'].sum()
print("Number of passengers who survived:", survived_counts)
```

```
Number of passengers who survived: 342
```

[12]:
```
survival_rate = train['Survived'].mean()*100
print("survival rate: {:.2f}%".format(survival_rate))
```

```
survival rate: 38.38%
```

jupyter SCT_DS_2 Last Checkpoint: 2 minutes ago

File Edit View Run Kernel Settings Help

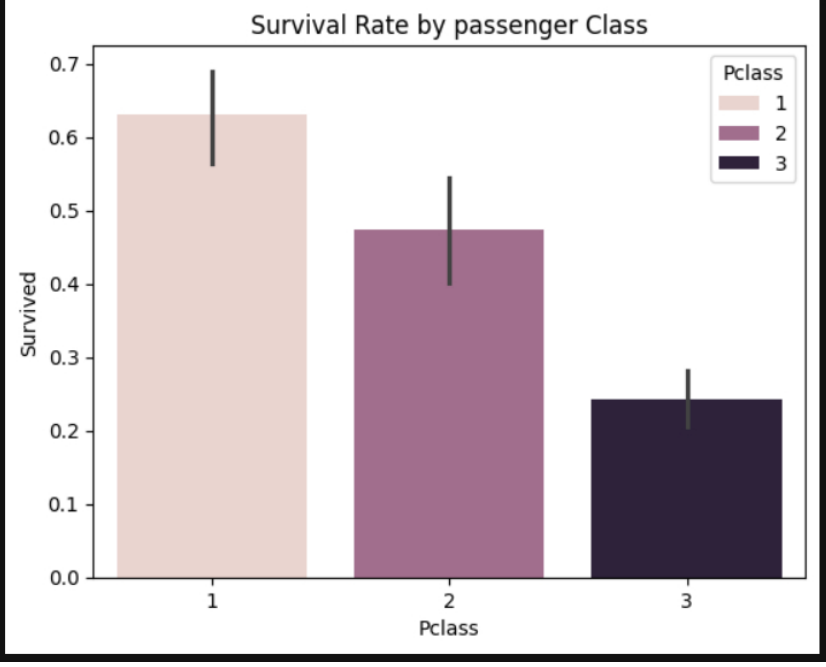
Trusted

Code

JupyterLab Python 3 (ipykernel)

Q.2 Plotting survival rate by passenger class

[29]:
```
sns.barplot(x='Pclass', y='Survived', hue='Pclass', data=train)
plt.title("Survival Rate by passenger Class")
plt.show()
```



jupyter SCT_DS_2 Last Checkpoint: 2 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

Q.3 plotting survival rate by embarked port

[31]:
```
sns.barplot(x='Embarked', y='Survived', hue='Sex', data=train)
plt.title("Survival Rate by Embarked Port")
plt.show()
```