Titanic Dataset – Python, Pandas & Visualization Exercise

Dataset: titanic.csv

Step 1: Import libraries and load the dataset Import pandas and read the CSV into a DataFrame.

##

```
import pandas as pd

df = pd.read_csv("students_titanic.csv")

df.head()
```

##

Step 2: Explore the dataset Display the shape, column names, and data types.

##

```
print(df.shape)

print(df.columns)

print(df.info())
```

##

Step 3: Basic descriptive statistics Display statistical summaries of numeric variables.

##

```
df.describe()
```

##

Step 4: Select specific columns Display only the columns Name, Age, and Sex.

##

```
df[["Name", "Age", "Sex"]].head()
```

##

Step 5: Average age Calculate the average age of all passengers.

##

df["Age"].mean()

##

Step 6: Count passengers by gender

Count how many males and females are onboard.

##

df["Sex"].value_counts()

##

Step 7: Passengers older than 40

Select passengers where Age > 40.

##

df[df["Age"] > 40]

##

Step 8: Survival rate by gender

Compute the mean of Survived grouped by Sex.

##

df.groupby("Sex")["Survived"].mean()

##

Step 9: Average fare by class

Compute average fare for each passenger class.

##

df.groupby("Pclass")["Fare"].mean()

##

Step 10: Sort by Fare

Sort passengers by fare (highest first).

##

df.sort_values(by="Fare", ascending=False).head()

##

Step 11: Plot a histogram of passenger ages

Visualize how ages are distributed using a histogram.

##

import matplotlib.pyplot as plt

plt.hist(df["Age"], bins=10, edgecolor="black") # Histogram of Age column, 10 bins, #black borders for clarity

plt.hist(df["Age"], bins=10, edgecolor="black") #

plt.title("Age Distribution of Passengers")

plt.xlabel("Age")

plt.ylabel("Number of Passengers")

plt.show()

##

Step 12: Plot a circular (pie) diagram of gender distribution

Show the proportion of males and females using a pie chart.

##

df["Sex"].value_counts().plot.pie(

autopct="%1.1f%%",# If a slice represents **63.45%** it wwill show 63.5 on the plot

startangle=90, # the first slice starts **at the top** (12 o'clock position).

colors=["skyblue", "lightcoral"],

ylabel=""

)

plt.title("Gender Distribution")

```
 plt.show()
```

##

Step 13: Visualize the correlation between Age and Fare

Plot a scatter plot of Age vs Fare and calculate their correlation.

##

```
plt.scatter(df["Age"], df["Fare"], alpha=0.7)
plt.title("Correlation between Age and Fare")
plt.xlabel("Age")
plt.ylabel("Fare")
plt.show()
corr = df["Age"].corr(df["Fare"])           # Compute Pearson correlation between Age and Fare
print("Correlation between Age and Fare:", round(corr, 2))  # Print rounded correlation
```

##