**Exploring Titanic Dataset with Pandas**

**1. Import the Required Libraries**

You should import Pandas (import pandas as pd) and Matplotlib (import matplotlib.pyplot as plt) for data manipulation and visualization.

import pandas as pd

import matplotlib.pyplot as plt

import pandas as pd

Explain: Importing pandas is able to calculate big datas.It’s Python excel.

**2. Load the Dataset**

Load the Titanic dataset from a CSV file named 'titanic.csv' into a Pandas DataFrame.

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

print(ti)

PassengerId Survived Pclass Name Sex ... Parch Ticket Fare Cabin Embarked

0 1 0 3 Braund, Mr. Owen Harris male ... 0 A/5 21171 7.2500 NaN S

1 2 1 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female ... 0 PC 17599 71.2833 C85 C

2 3 1 3 Heikkinen, Miss. Laina female ... 0 STON/O2. 3101282 7.9250 NaN S

3 4 1 1 Futrelle, Mrs. Jacques Heath (Lily May Peel) female ... 0 113803 53.1000 C123 S

4 5 0 3 Allen, Mr. William Henry male ... 0 373450 8.0500 NaN S

.. ... ... ... ... ... ... ... ... ... ... ...

886 887 0 2 Montvila, Rev. Juozas male ... 0 211536 13.0000 NaN S

887 888 1 1 Graham, Miss. Margaret Edith female ... 0 112053 30.0000 B42 S

888 889 0 3 Johnston, Miss. Catherine Helen "Carrie" female ... 2 W./C. 6607 23.4500 NaN S

889 890 1 1 Behr, Mr. Karl Howell male ... 0 111369 30.0000 C148 C

890 891 0 3 Dooley, Mr. Patrick male ... 0 370376 7.7500 NaN Q

Explain:Before knowing data type, we need to mean and load titianic

**3. Data Exploration**

Display the first few rows of the dataset to understand its structure.

Print information about the columns and their data types.

Show summary statistics for numerical columns (e.g., mean, min, max, etc.).

# Data Exploration

print(ti.info())

print(ti.shape)

print(ti['Age'].mean())

print(ti['Age'].median())

print(ti['Age'].min())

print(ti['Age'].max())

print(sum(ti['Age']))

print(sum(ti['Age'])/len(ti['Age']))#Age average

print(ti['Fare'].mean())

print(ti['Fare'].median())

print(ti['Fare'].min())

print(ti['Fare'].max())

print(sum(ti['Fare'])/len(ti['Fare']))#Fare average

Output:

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 PassengerId 891 non-null int64

1 Survived 891 non-null int64

2 Pclass 891 non-null int64

3 Name 891 non-null object

4 Sex 891 non-null object

5 Age 714 non-null float64

6 SibSp 891 non-null int64

7 Parch 891 non-null int64

8 Ticket 891 non-null object

9 Fare 891 non-null float64

10 Cabin 204 non-null object

11 Embarked 889 non-null object

Output(mean、median…..):

(891, 12)

29.69911764705882

28.0

0.42

80.0

nan

nan

32.204207968574636

14.4542

0.0

512.3292

32.204207968574636

Explain:Viewing datas type and calculating mean、median….

**4. Data Cleaning**

Identify and handle missing values in the dataset (e.g., fill missing age values with median).

Perform any necessary data transformations or cleaning steps.

# Data Cleaning

print(ti.isnull().sum())

ti=ti.dropna(subset=["Embarked"])

ti['Cabin']=ti['Cabin'].fillna("NoCabin")

print(ti.shape)

print(ti.isnull().sum())

ti = ti.fillna(ti['Age'].median())

print(ti)

print(ti.isnull().sum())

Output

PassengerId 0

Survived 0

Pclass 0

Name 0

Sex 0

Age 0

SibSp 0

Parch 0

Ticket 0

Fare 0

Cabin 0

Embarked 0

Explain:looking for and fill missing values(field:Embarked、Cabin and Age)

**5. Data Analysis**

Calculate and display the count of passengers by gender.

Compute the average age of passengers.

Determine the survival rate by passenger class.

# Data Analysis

print(ti['Sex'].value\_counts())

print(sum(ti['Age']))

print(sum(ti['Age'])/len(ti['Age']))#average

print(ti['Survived'].value\_counts(normalize=True))

print(ti['Pclass'].value\_counts())

Sr=ti.groupby('Pclass')['Survived'].mean()

Explain: Recalculating data and arrange survived rate by customer level.

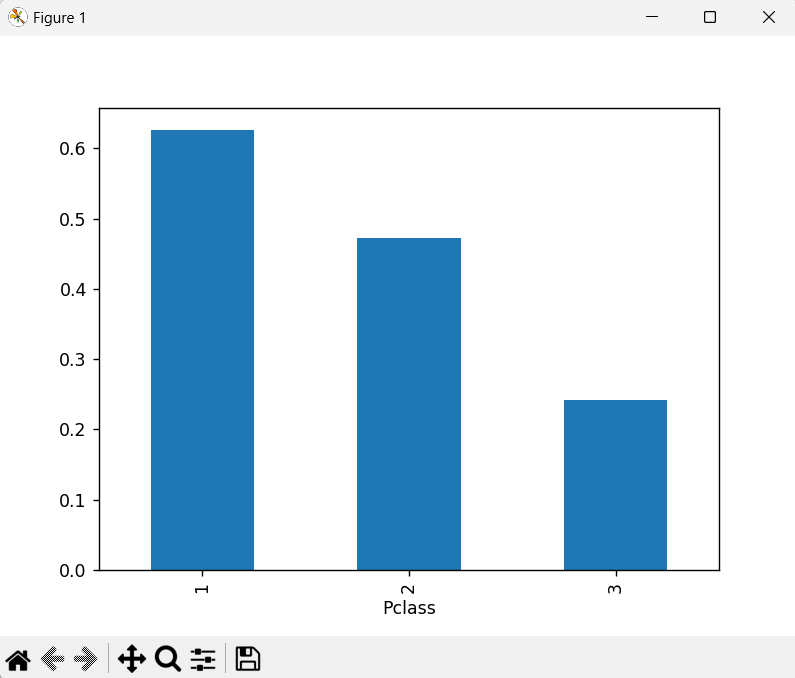
**6. Data Visualization**

Create a bar chart to visualize the survival rate by passenger class.

Add appropriate labels and a title to the chart.

Sr.plot.bar()

plt.show()



According to bar chart ,Pclass 1 survival rate is highest than others.

**7. Conclusion**

Summarize your findings or insights gained from analyzing the Titanic dataset.

Reflect on the importance of data cleaning and exploration in data analysis.

Ans:Solvling missing values ​​and arranging datas make data analysis be more complete.

thus ,arranging datas is very important.

My URL: <https://jason261109.github.io/Titanic/>

<https://github.com/Jason261109/Titanic.git>