**Exploring Titanic Dataset with Pandas**

**1. Import the Required Libraries**

Data Analysis and visualization for “Titanic” file.

import pandas as pd

import matplotlib.pyplot as plt

Using pandas and matplotlib help us to analyze and visualize the data through their professional segment.

**2. Load the Dataset**

Read the “Titanic” file through pandas.

df = pd.read\_csv('c:/Users/Max/Downloads/titanic.csv')

This command makes the python to read the file for us.

**3. Data Exploration**

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

print(df.head())

This Function helps you to test if your object has right type of data in it.

Print information about the columns and their data types.

print(df.info())

This prints information about the DataFrame which including the index dtype and columns, non-null values and memory usage.

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

print(df.describe())

Generate the descriptive statistics which include the summarize of the central tendency, dispersion and shape of a dataset's distribution, excluding NaN values.

**4. Data Cleaning**

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

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

*Fillna: Using a specific value to fill out all the NA/NaN value.*

This helps to fill the missing value with the median of the age values and allows to modify the DataFrame directly on the original variable without needing to assign the result back to the DataFrame.

Perform any necessary data transformations or cleaning steps.

df.drop(['PassengerId', 'Name', 'Ticket', 'Cabin'], axis=1, inplace=True)

Inspect the whole dataset, drop the unnecessary or take too many percentages of the data type, this helps to decrease the challenge of handling full data, and focus on the specific result that we need.

**5. Data Analysis**

Calculate and display the count of passengers by gender.

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

Followed this code, we could easily get the result of the gender of the passengers from “Titanic” file.

Compute the average age of passengers.

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

Then we need to find the average age of the passengers, through this code we could get the result.

Determine the survival rate by passenger class.

class\_survival = df.groupby('Pclass')['Survived'].mean()

print("Survival Rate by Passenger Class:")

print(class\_survival)

This method adapts the a pandas series.

Divide the DataFrame into groups based on the unique values in the "Pclass" column, which indicates whether a passenger was in first, second, or third class.

For each of these groups, isolate the "Survived" column, which contains 1s and 0s to signify whether each passenger in that class survived or not.

Determine the average value of the "Survived" column within each group. Since "Survived" only contains 1s and 0s, this average will equal the fraction of passengers who survived in each class.

**6. Data Visualization**

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

Add appropriate labels and a title to the chart.

class\_survival.plot(kind='bar')

plt.xlabel('Passenger Class')

plt.ylabel('Survival Rate')

plt.title('Survival Rate by Passenger Class')

plt.show()

Codes shows the result within chart, and to make us more easily understandable the meanings of each content, we added some labels to more easily understand.

* *Because the Series uses "Pclass" as its index, these class labels can be directly used for the plot's x-axis values without any additional processing.*
* *The survival rates contained in the Series will be directly used as the plot's y-axis values, allowing the data to be visualized easily.*

**7. Conclusion**

**一張含有 文字, 螢幕擷取畫面, 陳列, 數字 的圖片

自動產生的描述**

Fig. 1 Survival Rate by Passenger Class

According to the figure 1, we could see that the survival rate is highly related to the passenger class. With higher level of class, the survival rate has positive effect. In contrast, with lower level of class, the survival rate goes lower.

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

The analysis emphasizes the crucial role of comprehensive data preparation, which involves data cleansing and exploratory analysis. Cleaning the data and making it easily understandable provides a robust basis for conducting the analysis and obtaining dependable insights. In summary, the project underscores the vital importance of data cleansing by demonstrating how appropriate data management and analytics methods can convert raw data into significant information and insights.