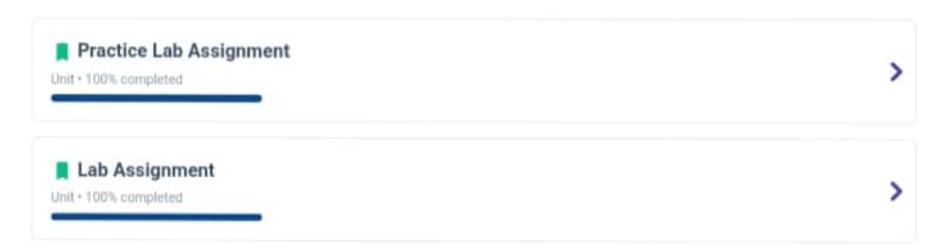


# **Practical 5**

# About this unit

Practical 5





A ==

Sample Test Cases

c Prev Reset Submit Next >



Write a Python program to analyze and visualize data from the Titanic dataset based on the following instructions:

(IIII) A L Z 2 -

#### Dataset Information:

5.2.1. Titanic Dataset

The dataset is stored in a CSV file named titanic.csv and has been loaded using the pandas library. It contains the following columns:

- Pclass: Passenger class (1 = First, 2 = Second, 3 = Third).
- Gender: Gender of the passenger (male/female).
- Age: Age of the passenger.
- Survived: Survival status (0 = Did not survive, 1 = Survived).
- Fare: Ticket fare paid by the passenger.

#### Visualization:

To represent these trends, you will create 5 visualizations using Matplotlib. The visualizations should be arranged in a 3x2 grid (3 rows and 2 columns).

#### Visualization Details:

Write the code to create a series of visualizations as follows:

## Bar Plot (Pclass Distribution):

- Create a bar plot to show the distribution of passengers across the different passenger classes (Pclass).
- Use the color skyblue for the bars.
- Title the plot as "Passenger Class Distribution".
- Label the x-axis as "Pclass" and the y-axis as "Count".

#### Pie Chart (Gender Distribution):

- · Create a pie chart to display the distribution of male and female passengers.
- Use lightblue for males and lightcoral for females.
- · Include percentages on the slices (use autopct='%1.1f%%').
- Title the plot as "Gender Distribution".

# Histogram (Age Distribution):

- · Create a histogram to visualize the distribution of passengers' ages.
- Use lightgreen for the bars with black edges (edgecolor = 'black').
- . Set the number of bins to 8 for the histogram.
- · Title the plot as "Age Distribution".
- · Label the x-axis as "Age" and the y-axis as "Frequency".

## Bar Plot (Survival Count):

- Create a bar plot to show the count of passengers who survived and those who did not, based on the Survived
- · Use the colors lightblue for survivors (1) and lightcoral for non-survivors (0).
- . Title the plot as "Survival Count".
- . Label the x-axis as "Survived (0 = No, 1 = Yes)" and the yaxis as "Count".

## Scatter Plot (Fare vs Age):

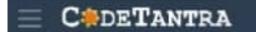
- . Create a scatter plot to visualize the relationship between the Fare and Age of passengers.
- · Use orange for the data points.
- . Title the plot as "Fare vs Age".
- Label the x-axis as "Age" and the y-axis as "Fare".

Note: Refer to the displayed plot in the sample test cases for better understanding.

Sample Test Cases

```
titanicData...
Explorer
         import pandas as pd
         import matplotlib.pyplot as plt
   2
   3
   4
         # Load the Titanic dataset from the
         CSV file
         df = pd.read_csv('titanic.csv')
   5
   6
   7
         # Set up the figure for 5 subplots
   8
         fig, axes = plt.subplots(3, 2,
         figsize=(12, 12))
   9
         # write the code...
  10
         # Plot 1: Count of passengers by
  11
         class
  12
         axes[0,
         0].bar(df['Pclass'].value_counts().in
         dex, df['Pclass'].value_counts(),
         color='skyblue')
         axes[0, 0].set_title("Passenger
  13
         Class Distribution")
  14
         axes[0, 0].set_xlabel("Pclass")
  15
         axes[0, 0].set_ylabel("Count")
  16
         # plot 2: Gender distribution
  17
         axes[0,
  18
         1].pie(df['Gender'].value_counts(),
         labels=df['Gender'].value_counts().in
         dex, autopct='%1.1f%%', colors=
         ['lightblue', 'lightcoral'])
         axes[0, 1].set_title("Gender
  19
         Distribution")
  20
  21
         # plot 3: Age distribution
  22
         axes[1, 0].hist(df['Age'].dropna().
         bins=8, color='lightgreen',
         edgecolor='black')
  23
         axes[1, 0].set_title("Age
         Distribution")
         axes[1, 0].set_xlabel("Age")
  24
  25
         axes[1, 0].set_ylabel("Frequency")
  26
  27
         # plot 4: Survival count
  28
         axes[1,
         1].bar(df['Survived'].value_counts().
         df['Survived'].value_counts(), color=
         ['lightblue', 'lightcoral'])
  29
         axes[1, 1].set_title("Survival
         Count")
         axes[1, 1].set_xlabel("Survived (0 =
  30
         No, 1 = Yes)")
  31
         axes[1, 1].set_ylabel("Count")
  32
  33
         # plot 5: Fare vs Age
  34
         axes[2, 0].scatter(df['Age'],
         df['Fare'], color='orange',
         edgecolors='black')
  35
         axes[2, 0].set_title("Fare vs Age")
         axes[2, 0].set_xlabel("Age")
  36
  37
         axes[2, 0].set_ylabel("Fare")
  38
  39
         plt.tight_layout()
         plt.show()
  40
    X.
```

C Prev Reset Submit Next >



5.2.2. Histogram of passenger information ... (IIIII) 🧀 📞 🗷 🔗

Write a Python code to plot a histogram for the distribution of the 'Age' column from the Titanic dataset. The histogram should display the frequency of different age ranges with the following specifications:

- 1. Use 30 bins for the histogram.
- Set the edge color of the bars to black (k).
- 3. Label the x-axis as 'Age' and the y-axis as 'Frequency'.
- 4. Add the title "Age Distribution" to the histogram.

The Titanic dataset contains columns as shown below,

a S u r v v e r d	P c l a s s	N a m e	Sex	A g e	Sbsp	P a r c h	T c k e t	Fare	C a b i n	E m b a r k e d
-------------------	-------------	------------------	-----	-------	------	-----------------------	-----------------------	------	-----------------------	-----------------

#### Sample Data:

PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ti 1,0,3, "Braund, Mr. Owen Harris", male, 22,1,0,A/5 21171,7 2,1,1, "Cumings, Mrs. John Bradley (Florence Briggs Thay 3,1,3,"Heikkinen, Miss. Laina", female, 26,0,0,STON/02. 3 4,1,1, "Futrelle, Mrs. Jacques Heath (Lily May Peel)", fe 5,0,3, "Allen, Mr. William Henry", male, 35,0,0,373450,8.0 6,0,3, "Moran, Mr. James", male, ,0,0,330877,8.4583, ,Q 7,0,1,"McCarthy, Mr. Timothy 3",male,54,0,0,17463,51.86 8.0.3, "Palsson, Master. Gosta Leonard", male.2,3,1,34990 9,1,3, "Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg 10,1,2,"Nasser, Mrs. Nicholas (Adele Achem)", female, 14,

Note: Refer to the visible test case for better reference.

Explorer Histogram... import pandas as pd import matplotlib.pyplot as plt 2 3 # Load the Titanic dataset 4 data = pd.read\_csv('Titanic-5 Dataset.csv') 6 7 # Data Cleaning 8 data['Age'].fillna(data['Age'].median (), inplace=True) 9 data['Embarked'].fillna(data['Embarke d'].mode()[0], inplace=True) data.drop('Cabin', axis=1, 10 inplace=True) 11 12 # Convert categorical features to numeric 13 data['Sex'] = data['Sex'].map({'male': 0, 'female': 1}) data = pd.get\_dummies(data, columns= 14 ['Embarked'], drop\_first=True) 15 # Write your code here for Histogram 16 plt.hist(data['Age'], bins=30, 17 edgecolor='k') 18 plt.xlabel('Age') 19 plt.ylabel('Frequency') plt.title('Age Distribution') 20 21 plt.show()

A ===



5.2.3. Bar plot of survival rate of passengers (1989) A 📞 🗷 🔗



BarPiotOf...

Write a Python code to plot a bar chart that shows the count of passengers who survived and did not survive in the Titanic dataset. The chart should display the following specifications:

- 1. Use the 'Survived' column to show the count of survivors (0 = Did not survive, 1 = Survived).
- 2. Set the chart type to 'bar'.
- Add the title "Survival Count" to the chart.
- Label the x-axis as 'Survived' and the y-axis as 'Count'.

The Titanic dataset contains columns as shown below,

S S P T T E M B S P T T A B B B B B B B B B B B B B B B B B
---

#### Sample Data:

PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ti 1,0,3, "Braund, Mr. Owen Harris", male, 22,1,0,A/5 21171,7 2,1,1, "Cumings, Mrs. John Bradley (Florence Briggs Thay 3,1,3,"Heikkinen, Miss. Laina", female, 26,0,0,STON/02. 3 4,1,1,"Futrelle, Mrs. Jacques Heath (Lily May Peel)",fe 5,0,3, "Allen, Mr. William Henry", male, 35,0,0,373450,8.0 6,0,3, "Moran, Mr. James", male, ,0,0,330877,8.4583, ,Q 7,0,1,"McCarthy, Mr. Timothy J",male,54,0,0,17463,51.86 8.0,3,"Palsson, Master. Gosta Leonard", male,2,3,1,34990 9,1,3, "Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg 10,1,2,"Nasser, Mrs. Nicholas (Adele Achem)", female, 14,

Note: Refer to the visible test case for better reference.

Explorer import pandas as pd 2 import matplotlib.pyplot as plt 3 4 # Load the Titanic dataset data = pd.read csv('Titanic-5 Dataset.csv') 6 7 # Data Cleaning 8 data['Age'].fillna(data['Age'].median (), inplace=True) 9 data['Embarked'].fillna(data['Embarke d'].mode()[0], inplace=True) data.drop('Cabin', axis=1, 10 inplace=True) 11 12 # Convert categorical features to numeric 13 data['Sex'] = data['Sex'].map({'male': 0, 'female': 1}) 14 data = pd.get\_dummies(data, columns= ['Embarked'], drop\_first=True) 15 # Write your code here for Bar Plot 16 for Survival Rate 17 18 survival counts = data['Survived'].value\_counts() survival counts.plot(kind='bar') 19 plt.title('Survival Count') 20 plt.xlabel('Survived') 21 plt.ylabel('Count') 22 23 plt.show() 24

**X**.

===



5.2.4. Bar Plot for Survival by Gender



Write a Python code to plot a stacked bar chart that shows the count of passengers who survived and did not survive, grouped by gender, in the Titanic dataset. The chart should display the following specifications:

- 1. Group the data by the 'Sex' column, then use the value\_counts() function to count the occurrences of survivors (0 = Did not survive, 1 = Survived) for each gender.
- Use a stacked bar chart to display the survival counts.
- Add the title "Survival by Gender" to the chart.
- 4. Label the x-axis as 'Gender' and the y-axis as 'Count'.
- The legend should indicate 'Not Survived' and 'Survived'.

The Titanic dataset contains columns as shown below,

a s s s s s s s s s s s s s s s s s s s	p c	N a m e	S e x	A g e	8 - b 8 p	P a r c h	T i c k e t	F a r e	C a b i n	E m b a r k e d
---	-----	------------------	-------	-------	-----------	-----------	-------------	------------------	-----------	--------------------------------

#### Sample Data:

Passengerld, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ti 1,0,3, "Braund, Mr. Owen Harris", male, 22,1,0,A/5 21171,7 2,1,1, "Cumings, Mrs. John Bradley (Florence Briggs Thay 3,1,3,"Heikkinen, Miss. Laina", female, 26,0,0,5TON/02. 3 4,1,1, "Futrelle, Mrs. Jacques Heath (Lily May Peel)", fee 5,0,3, "Allen, Mr. William Henry", male, 35,0,0,373450,8.0 6,0,3, "Moran, Mr. James", male, .0,0,330877,8.4583, .Q 7,0,1, "McCarthy, Mr. Timothy J", male, 54,0,0,17463,51.86 8,0,3,"Palsson, Master, Gosta Leonard", male,2,3,1,34990 9,1,3, "Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg 10,1,2,"Nasser, Mrs. Nicholas (Adele Achem)", female, 14,

Note: Refer to the visible test case for better reference.

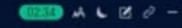
```
Explorer
   BarPiotOf...
         import pandas as pd
         import matplotlib.pyplot as plt
   2
   3
   4
         # Load the Titanic dataset
         data = pd.read csv('Titanic-
   5
         Dataset.csv')
   6
   7
         # Data Cleaning
         data['Age'].fillna(data['Age'].median
   8
         (), inplace=True)
   9
         data['Embarked'].fillna(data['Embarke
         d'].mode()[0], inplace=True)
         data.drop('Cabin', axis=1,
  10
         inplace=True)
  11
  12
         # Convert categorical features to
         numeric
  13
         data['Sex'] =
         data['Sex'].map({'male': 0,
         'female': 1})
         data = pd.get_dummies(data, columns=
  14
         ['Embarked'], drop_first=True)
  15
         # Write your code here for Bar Plot
  16
         for Survival by Gender
  17
         survival by gender =
  18
         data.groupby('Sex')
         ['Survived'].value_counts().unstack()
         .fillna(0)
  19
         survival_by_gender.columns = ['Not
         Survived', 'Survived']
         survival_by_gender.index = ['0', '1']
  20
         survival_by_gender.plot(kind='bar',
  21
         stacked=True)
         plt.title('Survival by Gender')
  22
         plt.xlabel('Gender')
  23
  24
         plt.ylabel('Count')
  25
         plt.legend(title=None)
  26
         plt.show()
```

**X**.

==



5.2.5. Bar Plot for Survival by Pclass



Write a Python code to plot a stacked bar chart that shows the count of passengers who survived and did not survive, grouped by passenger class (Pclass), in the Titanic dataset. The chart should display the following specifications:

- 1. Group the data by the Pclass column and count the number of survivors (0 = Did not survive, 1 = Survived) for each class using value\_counts().
- Use a stacked bar chart to display the survival counts.
- 3. Add the title "Survival by Pclass" to the chart.
- Label the x-axis as 'Pclass' and the y-axis as 'Count'.
- 5. The legend should indicate 'Not Survived' and 'Survived'.

The Titanic dataset contains columns as shown below.

a Survive d	p c l a s s	N a m e	S e x	A g e	S i b S p	Багср	T i c k e t	F a r e	C a b i n	E m b a r k e d
-------------	-------------	------------------	-------------	-------	-----------	-------	-------------	------------------	-----------------------	-----------------

#### Sample Data:

PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ti 1,0,3,"Braund, Mr. Owen Harris", male, 22,1,0,A/5 21171,7 2,1,1, "Cumings, Mrs. John Bradley (Florence Briggs Thay 3,1,3,"Heikkinen, Miss. Laina", female, 26,0,0,5TON/02. 3 4,1,1,"Futrelle, Mrs. Jacques Heath (Lily May Peel)",fe 5,0,3,"Allen, Mr. William Henry", male, 35,0,0,373450,8.0 6,0,3, "Moran, Mr. James", male,,0,0,330877,8,4583,,Q 7,0,1,"McCarthy, Mr. Timothy J",male,54,0,0,17463,51.86 8,0,3,"Palsson, Master. Gosta Leonard", male,2,3,1,34990 9,1,3, "Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg 10,1,2,"Nasser, Mrs. Nicholas (Adele Achem)", female, 14,

## Note:

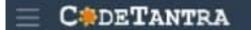
- Refer to the visible test case for better reference.
- . Ensure you use the groupby() function with value\_counts() to count the survivors and non-survivors for each Polass.
- Do not manually use size() or unstack() without value\_counts(). Use the value\_counts() method for counting survival status directly.

Explorer BarPiotOf... import pandas as pd import matplotlib.pyplot as plt 2 3 4 # Load the Titanic dataset 5 data = pd.read\_csv('Titanic-Dataset.csv') 6 7 # Data Cleaning data['Age'].fillna(data['Age'].median 8 (), inplace=True) data['Embarked'].fillna(data['Embarke 9 d'].mode()[0], inplace=True) data.drop('Cabin', axis=1, 10 inplace=True) 11 12 # Convert categorical features to numeric data['Sex'] = 13 data['Sex'].map({'male': 0, 'female': 1}) 14 data = pd.get\_dummies(data, columns= ['Embarked'], drop\_first=True) 15 # Write your code here for Bar Plot 16 for Survival by Pclass 17 survival by class = 18 data.groupby('Pclass') ['Survived'].value\_counts().unstack() .fillna(0) 19 survival\_by\_class.columns = ['Not Survived', 'Survived'] survival\_by\_class.plot(kind='bar', 20 stacked=True) 21 plt.title('Survival by Pclass') 22 plt.xlabel('Pclass') plt.ylabel('Count') 23 24 plt.legend(title=None) 25 plt.show() 26

Sample Test Cases

**X**.

==



5.2.6. Bar Plot for Survival by Embarked



BarPiotOf...

Write a Python code to plot a stacked bar chart showing the survival count for passengers based on their embarkation location in the Titanic dataset.

The chart should display the following specifications:

- Use the Embarked column to determine the embarkation. location. After converting this column into dummy variables (using pd.get\_dummies()), plot the survival count based on the Embarked\_Q column (representing passengers who embarked from Queenstown) in relation to survival.
- Set the chart type to 'bar' and make it stacked.
- Add the title "Survival by Embarked" to the chart.
- 4. Label the x-axis as 'Embarked' and the y-axis as 'Count'.
- 5. Include a legend to distinguish between survivors and non-survivors (label the legend as 'Survived' and 'Not Survived').

The Titanic dataset contains columns as shown below.

S u P c l a s e d	N S e m x	A D B S P	p T i c k c e t	F a b i n	E m b a r k e d
-------------------	-----------	-----------	-----------------	-----------	-----------------------

## Sample Data:

PassengerId, Survived, Pclass, Name, Sex, Age, SihSp, Parch, Ti 1,0,3, "Braund, Mr. Owen Harris", male, 22,1,0,A/5 21171,7 2,1,1, "Cumings, Mrs. John Bradley (Florence Briggs Thay 3,1,3,"Heikkinen, Miss. Laina", female, 26,0,0,5TON/O2. 3 4,1,1,"Futrelle, Mrs. Jacques Heath (Lily May Peel)", fel 5,0,3,"Allen, Mr. William Henry", male, 35,0,0,373450,8.0 6,0,3, "Moran, Mr. James", male, ,0,0,330877,8.4583, ,Q 7,0,1,"McCarthy, Mr. Timothy J", male, 54,0,0,17463,51.86 8,0,3, "Palsson, Master. Gosta Leonard", male,2,3,1,34990 9,1,3, "Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg 10,1,2, "Nasser, Mrs. Nicholas (Adele Achem)", female, 14,

Note: Refer to the visible test case for better reference.

```
Explorer
         import pandas as pd
         import matplotlib.pyplot as plt
   2
   3
   4
         # Load the Titanic dataset
         data = pd.read_csv('Titanic-
   5
         Dataset.csv')
   6
   7
         # Data Cleaning
   8
         data['Age'].fillna(data['Age'].median
         (), inplace=True)
         data['Embarked'].fillna(data['Embarke
   9
         d'].mode()[0], inplace=True)
         data.drop('Cabin', axis=1,
  10
         inplace=True)
  11
  12
         # Convert categorical features to
         numeric
  13
         data['Sex'] =
         data['Sex'].map({'male': 0,
         'female': 1})
  14
         data = pd.get_dummies(data, columns=
         ['Embarked'], drop_first=True)
  15
         # Write your code here for Bar Plot
  16
         for Survival by Embarked
  17
  18
         grouped = data.groupby('Embarked Q')
         ['Survived'].value_counts().unstack()
         .fillna(0)
  19
         grouped.columns = ['Not Survived',
         'Survived'1
         grouped.plot(kind='bar',
  20
         stacked=True)
         plt.title('Survival by Embarked')
  21
  22
         plt.xlabel('Embarked')
         plt.ylabel('Count')
  23
         plt.legend(title=None)
  24
  25
         plt.show()
  26
```



5.2.7. Box plot for Age Distribution



Write a Python code to plot a boxplot that shows the distribution of the 'Age' column from the Titanic dataset across different passenger classes. The boxplot should display the following specifications:

- 1. Use the Pclass column to group the data for the boxplot.
- 2. Set the title of the plot to "Age by Pclass".
- 3. Remove the default subtitle with plt.suptitle(").
- Label the x-axis as 'Pclass' and the y-axis as 'Age'.

The Titanic dataset contains columns as shown below,

a s u r v i v e d d	Pclass	N a m e	S e x	A g e	S - b S p	Parch	T c k e t	F a r e	C a b i	E m b a r k e d
---------------------	--------	------------------	-------	-------	-----------	-------	-----------------------	------------------	------------------	-----------------------

#### Sample Data:

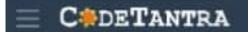
PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ti 1,0,3, "Braund, Mr. Owen Harris", male, 22,1,0,A/5 21171,7 2,1,1, "Cumings, Mrs. John Bradley (Florence Briggs Thay 3,1,3, "Heikkinen, Miss. Laina", female, 26,0,0,STON/O2. 3 4,1,1, "Futrelle, Mrs. Jacques Heath (Lily May Peel)", fe 5,0,3, "Allen, Mr. William Henry", male, 35,0,0,373450.8.0 6,0,3, "Moran, Mr. James", male,0,0,330877,8.4583.,Q 7,0,1, "McCarthy, Mr. Timothy J", male,54,0,0,17463,51.86 8,0,3, "Palsson, Master. Gosta Leonard", male,2,3,1,34990 9,1,3, "Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg 10,1,2, "Nasser, Mrs. Nicholas (Adele Achem)", female,14,

Note: Refer to the visible test case for better reference.

Explorer BoxPlotFo... import pandas as pd import matplotlib.pyplot as plt 2 3 4 # Load the Titanic dataset data = pd.read csv('Titanic-5 Dataset.csv') 6 7 # Data Cleaning 8 data['Age'].fillna(data['Age'].median (), inplace=True) 9 data['Embarked'].fillna(data['Embarke d'].mode()[0], inplace=True) data.drop('Cabin', axis=1, 10 inplace=True) 11 12 # Convert categorical features to numeric 13 data['Sex'] = data['Sex'].map({'male': 0, 'female': 1}) 14 data = pd.get\_dummies(data, columns= ['Embarked'], drop\_first=True) 15 # Write your code here for Box Plot 16 for Age by Pclass 17 plt.figure(figsize=(8, 6)) 18 19 data.boxplot(column='Age', by='Pclass') plt.suptitle('') 20 21 plt.title('Age by Pclass') plt.xlabel('Pclass') 22 23 plt.ylabel('Age') plt.show() 24 25

**X**.

===



5.2.8. Box Plot for Age by Survived



Write a Python code to plot a boxplot that shows the distribution of the 'Age' column from the Titanic dataset based on whether passengers survived or not. The boxplot should display the following specifications:

- 1. Use the Survived column to group the data for the boxplot (0 = Did not survive, 1 = Survived).
- 2. Set the title of the plot to "Age by Survival".
- 3. Remove the default subtitle with plt.suptitle(").
- Label the x-axis as 'Survived' and the y-axis as 'Age'.

The Titanic dataset contains columns as shown below.

as Suppose of Span Cab a
--

#### Sample Data:

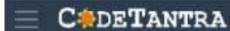
PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ti 1,0,3, "Braund, Mr. Owen Harris", male, 22,1,0,A/5 21171,7 2,1,1, "Cumings, Mrs. John Bradley (Florence Briggs Thay 3,1,3, "Heikkinen, Miss. Laina", female, 26,0,0,5TON/02. 3 4.1.1, "Futrelle, Mrs. Jacques Heath (Lily May Peel)", fee 5.0,3, "Allen, Mr. William Henry", male, 35,0,0,373450,8.0 6,0,3,"Moran, Mr. James", male,,0,0,330877,8.4583,,Q 7,0,1, "McCarthy, Mr. Timothy J", male, 54,0,0,17463,51.86 8,0,3,"Palsson, Master, Gosta Leonard", male,2,3,1,34990 9,1,3, "Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg 10,1,2,"Nasser, Mrs. Nicholas (Adele Achem)", female, 14,

Note: Refer to the visible test case for better reference.

Explorer BoxPlotFo... import pandas as pd import matplotlib.pyplot as plt 2 3 # Load the Titanic dataset 4 data = pd.read csv('Titanic-5 Dataset.csv') 6 7 # Data Cleaning data['Age'].fillna(data['Age'].median 8 (), inplace=True) 9 data['Embarked'].fillna(data['Embarke d'].mode()[0], inplace=True) data.drop('Cabin', axis=1, 10 inplace=True) 11 12 # Convert categorical features to numeric 13 data['Sex'] = data['Sex'].map({'male': 0, 'female': 1}) 14 data = pd.get\_dummies(data, columns= ['Embarked'], drop\_first=True) 15 # Write your code here for Box Plot 16 for Age by Survived 17 18 plt.figure(figsize=(8, 6)) data.boxplot(column='Age', 19 by='Survived') plt.suptitle('') 20 21 plt.title('Age by Survival') plt.xlabel('Survived') 22 23 plt.ylabel('Age') plt.show() 24 25

**X**.

===



5.2.9. Box Plot for Fare by Pclass



Write a Python code to plot a boxplot that shows the distribution of the 'Fare' column from the Titanic dataset based on the passenger class (Pclass). The boxplot should display the following specifications:

- Use the Polass column to group the data for the boxplot.
- Set the title of the plot to "Fare by Polass".
- 3. Remove the default subtitle with plt.suptitle(\*).
- Label the x-axis as 'Pclass' and the y-axis as 'Fare'.

The Titanic dataset contains columns as shown below,

a S u r v e n g e r d d		N a m e	S e x	A g e	S i b S p	P a r c h	T i c k e t	F a r e	Cabin	Embarked
-------------------------	--	---------	-------------	-------	-----------	-----------	-------------	------------------	-------	----------

#### Sample Data:

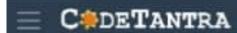
```
PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ti
1,0,3, "Braund, Mr. Onen Harris", male, 22,1,0,A/5 21171,7
2,1,1, "Cumings, Mrs. John Bradley (Florence Briggs Thay
3,1,3,"Heikkinen, Miss. Laina", female, 26,0,0,5TON/02. 3
4.1.1, "Futrelle, Mrs. Jacques Heath (Lily May Peel)", fee
5,0,3, "Allen, Mr. William Henry", male, 35,0,0,373450,8.0
6,0,3, "Moran, Mr. James", male,,0,0,330877,8.4583,,Q
7,0,1, "McCarthy, Mr. Timothy J", male, 54,0,0,17463,51.86
8,0,3,"Palsson, Master. Gosta Leonard", male,2,3,1,34990
9,1,3, "Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg
10,1,2,"Nasser, Mrs. Nicholas (Adele Achem)", female,14,
```

Note: Refer to the visible test case for better reference.

```
BoxPlotFo...
Explorer
         import pandas as pd
         import matplotlib.pyplot as plt
   2
   3
   4
         # Load the Titanic dataset
         data = pd.read_csv('Titanic-
   5
         Dataset.csv')
   6
   7
         # Data Cleaning
   8
         data['Age'].fillna(data['Age'].median
         (), inplace=True)
         data['Embarked'].fillna(data['Embarke
   9
         d'].mode()[0], inplace=True)
         data.drop('Cabin', axis=1,
  10
         inplace=True)
  11
         # Convert categorical features to
  12
         numeric
  13
         data['Sex'] =
         data['Sex'].map({'male': 0,
         'female': 1})
         data = pd.get_dummies(data, columns=
  14
         ['Embarked'], drop_first=True)
  15
  16
         # Write your code here for Box Plot
         for Fare by Pclass
  17
         plt.figure(figsize=(8, 6))
  18
         data.boxplot(column='Fare',
  19
         by='Pclass')
         plt.suptitle('')
  20
  21
         plt.title('Fare by Pclass')
         plt.xlabel('Pclass')
  22
  23
         plt.ylabel('Fare')
         plt.show()
  24
  25
```

**)**..

 $\blacksquare$ 



5.2.10. Scatter Plot for Age vs. Fare

Write a Python code to plot a scatter plot showing the relationship between the 'Age' and 'Fare' columns in the Titanic dataset. The scatter plot should display the following specifications:

(IIIII) A L Z Ø

- 1. Use the Age column for the x-axis and the Fare column for the y-axis.
- Set the title of the plot to "Age vs. Fare".
- 3. Label the x-axis as 'Age' and the y-axis as 'Fare'.

The Titanic dataset contains columns as shown below,

S U I V e d	P C N a m s e s	S e x	A g e	SIBSP	p a r c	T i c k e t	F a r e	C a b i n	Embarked
-------------	-----------------	-------	-------------	-------	---------	-------------	------------------	-----------	----------

#### Sample Data:

PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ti 1,0,3, "Braund, Mr. Owen Harris", male, 22,1,0,A/5 21171,7 2,1,1, "Cumings, Wrs. John Bradley (Florence Briggs Thay 3,1,3,"Heikkinen, Miss. Laina", female, 26,0,0,5TON/02. 3 4,1,1, "Futrelle, Mrs. Jacques Heath (Lily May Peel)", fe 5,0,3,"Allen, Mr. William Henry", male, 35,0,0,373450,8.0 6,0,3, "Moran, Mr. James", male, ,0,0,330877,8.4583, ,Q 7,0,1, "McCarthy, Mr. Timothy J", male, 54,0,0,17463,51.86 8,0,3, "Palsson, Master, Gosta Leonard", male,2,3,1,34990 9,1,3, "Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg 10,1,2,"Masser, Mrs. Nicholas (Adele Achem)", female, 14,

Note: Refer to the visible test case for better reference.

Explorer AgeFareSc... import pandas as pd import matplotlib.pyplot as plt 2 3 # Load the Titanic dataset 4 data = pd.read\_csv('Titanic-5 Dataset.csv') 6 7 # Data Cleaning 8 data['Age'].fillna(data['Age'].median (), inplace=True) data['Embarked'].fillna(data['Embarke 9 d'].mode()[0], inplace=True) data.drop('Cabin', axis=1, 10 inplace=True) 11 12 # Convert categorical features to numeric 13 data['Sex'] = data['Sex'].map({'male': 0, 'female': 1}) data = pd.get\_dummies(data, columns= 14 ['Embarked'], drop\_first=True) 15 # Write your code here for Box Plot 16 for Fare by Pclass 17 plt.figure(figsize=(6.4,4.8)) 18 plt.scatter(data['Age'],data['Fare']) 19 plt.title('Age vs. Fare') 20 plt.xlabel('Age') 21 plt.ylabel('Fare') 22 23 24 plt.show() 25

=

# □ C⊕DETANTRA

5.2.11. Scatter Plot for Age vs. Fare by Surv... (1997) A C 2

Write a Python code to plot a scatter plot showing the relationship between the 'Age' and 'Fare' columns in the Titanic dataset, with points color-coded by survival status. The scatter plot should display the following specifications:

- Use the Age column for the x-axis and the Fare column for the y-axis.
- Color the points based on the Survived column: Red for passengers who did not survive (Survived = 0). Blue for passengers who survived (Survived = 1).
- 3. Set the title of the plot to "Age vs. Fare by Survival".
- 4. Label the x-axis as 'Age' and the y-axis as 'Fare'.

The Titanic dataset contains columns as shown below,

S u r v i v e d	P C I a s s	N a m e	53 e ×	A g e	8 - 68 - 6	barcp	T c k e t	F a r e	Cabin	E m b a r k e d
-----------------	-------------	---------	--------	-------------	------------	-------	-----------	------------------	-------	--------------------------

## Sample Data:

PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ti 1.0,3. "Braund, Mr. Owen Harris", male, 22,1,0,A/5 21171,7 2.1,1, "Cumings, Mrs. John Bradley (Florence Briggs Thay 3.1,3, "Heikkinen, Miss. Laina", female, 26,0,0,STON/O2. 3 4.1,1, "Futrelle, Mrs. Jacques Heath (Lily May Peel)", fer 5.0,3, "Allen, Mr. William Henry", male, 35,0,0,373450,8.0 6.0,3, "Moran, Mr. James", male,,0,0,330877,8.4583,,Q 7.0,1, "McCarthy, Mr. Timothy J", male,54,0,0,17463,51.86 8.0,3, "Palsson, Master, Gosta Leonard", male,2,3,1,34990 9.1,3, "Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg 10,1,2, "Nasser, Mrs. Nicholas (Adele Achem)", female,14,

Note: Refer to the visible test case for better reference.

Explorer AgeFareSc... import pandas as pd 2 import matplotlib.pyplot as plt 3 4 # Load the Titanic dataset data = pd.read csv('Titanic-5 Dataset.csv') 6 7 # Data Cleaning 8 data['Age'].fillna(data['Age'].median (), inplace=True) data['Embarked'].fillna(data['Embarke 9 d'].mode()[0], inplace=True) data.drop('Cabin', axis=1, 10 inplace=True) 11 12 # Convert categorical features to numeric 13 data['Sex'] = data['Sex'].map({'male': 0, 'female': 1}) 14 data = pd.get\_dummies(data, columns= ['Embarked'], drop\_first=True) 15 16 # Write your code here for Scatter Plot for Age vs. Fare by Survived 17 colors = data['Survived'].map({0: 'red', 1: 'blue'}) plt.scatter(data['Age'], 18 data['Fare'], c=colors) 19 # Set labels and title 20 plt.xlabel("Age") 21 plt.ylabel("Fare") 22 plt.title("Age vs. Fare by Survival") 23 24 # Show the plot 25 plt.show() 26 27 28 29 30

Reset

**3.** 

=