Essentials of Data Science Laboratory - 2304102L - 2304102L

Search course ctrl + k

- 1. Practical 1
- 2. Practical 2
- 3. Practical 3
- 4. Practical 4
- 5. Practical 5

0 units left

Completed

Pickup from where you left off!

Scatter Plot for Age vs. Fare by Survived

Practical 5 • Lab Assignment

Resume

Practical 1

Unit • 100% completed

Practical 2

Unit • 100% completed

Practical 3

Unit • 100% completed

1.1.1. Calculate Momentum

12:05

Write a program that accepts the mass of an object (in kilograms) and its velocity (in meters per second), then calculates and displays the momentum of the object. The momentum p is calculated using the formula:

$$p = m \times v$$

where:
m is the mass of the object (in kilograms).
v is the velocity of the object (in meters per second).

Input Format:
A single floating-point number representing the mass of the object in kilograms.
A single floating-point number representing the velocity of the object in meters per second.

Output Format:
The output will display calculated momentum with appropriate units (kgm/s) (rounded up to 2 decimal places).

Sample Test Cases +

Explorer

calculate...

Submit

```
1  m=float(input())
2  v=float(input())
3  p=m*v
4  print('%0.2f'%p,end="")
5  print("kgm/s")
```

Debugger

Terminal Test cases

1.1.2. Conditional Calculation Based on the Number of Digits

16:12

Write a Python program that accepts an integer n as input. Depending on the number of digits in n.

Constraints:
$1 \le n \le 999$

Input Format:
The input consists of a single integer n.

Output Format:
If n is a single-digit number, print its square.
If n is a two-digit number, print its square root (rounded to two decimal places).
If n is a three-digit number, print its cube root (rounded to two decimal places).
Else print "Invalid".

Sample Test Cases +

Explorer

condition...

Submit

```
1  n=int(input())
2  if(0<n<10):
3      print(n*n)
4  elif(10<=n<100):
5      p=n**0.5
6      print( "%0.2f"%p)
7  elif(100<=n<=999):
8      r=n**(1/3)
9      print("%.2f"%r)
10 else:
11     print("Invalid")
```

Debugger

Terminal Test cases

1.1.3. Age and Salary Calculation

38:16

Write a Python program that reads the birth date and salary of employees.

Input Format:
The input consists of:
A string representing the birth date of the employee in the format $DD-MM-YYYY$.
A floating-point number representing the salary of the employee in rupees.

Output Format:
The output should include:
The age of the employee.
The salary of the employee in dollars.

Note:
1INR=0.012USD

Sample Test Cases +

Explorer

birthDate...

Submit

```python
def calculate_age(birthdate):
    date_object = datetime.strptime(birthdate, "%d-%m-%Y")
    today = datetime.today()
    if ((today.month, today.day ) < (date_object.month, date_object.day)):
        age = today.year-date_object.year-((today.month, today.day ) < (date_object.month, date_object.day))
        return age
    elif((today.month, today.day ) > (date_object.month,date_object.day)):
        age = today.year-date_object.year- ((today.month, today.day) > (date_object.month, date_object.day))
        return age


def convert_salary_to_dollars(salary_in_rupees):
    salary=salary_in_rupees*0.012
    return salary

birthdate = input()
salary_in_rupees = float(input())
age = calculate_age(birthdate)
salary_in_dollars = convert_salary_to_dollars(salary_in_rupees)
print(f"Age: {age}")
print(f"Salary in dollars: {salary_in_dollars: 2f}")
```

Debugger

Terminal Test cases

1.1.4. Reverse a Number

15:54

You are given an integer number. Your task is to reverse the digits of the number and print the reversed number.

Input Format
The input is an integer.

Output Format
Print a single integer which is the reversed number.

Sample Test Cases +

Explorer

reverseN...

Submit

Debugger

```
1  num=int(input())
2  reverse=0
3  while num != 0:
4      digit = num % 10
5      reverse = reverse*10 + digit
6      num = num//10
7
8  print(reverse)
```

Terminal Test cases

1.1.5. Multiplication Table

10:48

Write a Python program that takes an integer as input and prints the multiplication table for that integer from 1 to 10.

Input Format:
The first line of input contains an integer that represents the number for which the multiplication table is to be printed.

Output Format:
Print the multiplication table for the given number .

Sample Test Cases +

Explorer

multiplica...

Submit

```
1  n=int(input())
2  i=1
3  while i<=10:
4      print(n,"x",i,"=",n*i)
5      i=i+1
```

Debugger

Terminal Test cases

1.2.1. Pass or Fail

15:36

Write a Python program that accepts the number of courses and the marks of a student in those courses.
The grade is determined based on the aggregate percentage:

- If the aggregate percentage is greater than 75, the grade is Distinction.
- If the aggregate percentage is greater than or equal to 60 but less than 75, the grade is First Division.
- If the aggregate percentage is greater than or equal to 50 but less than 60, the grade is Second Division.
- If the aggregate percentage is greater than or equal to 40 but less than 50, the grade is Third Division.

Input Format:
The first input will be an integer n, the number of courses.
The second input will be n integers representing the marks of the student in each of the n courses, separated by a space.

Output Format:
If the student passes all courses:

- Print the aggregate percentage (rounded to two decimal places).
- Print the grade based on the aggregate percentage.

If the student fails any course (marks < 40 in any course), print:

Sample Test Cases +

Explorer passorFa... Submit Debugger

```
1  n=int(input())
2  marks= list(map(int, input().split()))
3  if any(mark<40 for mark in marks):
4      print("Fail")
5  else:
6      totalmarks = sum(marks)
7      Aggregate_percentage=(totalmarks/(n*100))*100
8      print(f"Aggregate Percentage: {Aggregate_percentage:.2f}")
9      if(Aggregate_percentage > 75):
10         print("Grade: Distinction")
11     elif(60<=Aggregate_percentage < 75):
12         print("Grade: First Division")
13     elif(50<=Aggregate_percentage < 60):
14         print("Grade: Second Division")
15     elif(40<=Aggregate_percentage > 50):
16         print("Grade: Third Division")
17     else:
18         print("Fail")
19
20
```

Terminal Test cases

1.2.2. Fibonacci series using Recursive Function

05:28

Write a Python program to find the Fibonacci series of a given number of terms using recursive function calls.

Expected Output-1:
Enter terms for Fibonacci series: 5
0 1 1 2 3

Expected Output-2:
Enter terms for Fibonacci series: 9
0 1 1 2 3 5 8 13 21

Instructions:

- Your input and output must follow the input and output layout mentioned in the visible sample test case.
- Hidden test cases will only pass when users' input and output match the expected input and output.

Sample Test Cases +

Explorer

fib.py

Submit

Debugger

```
1  def fib(n):
2      if n <= 0:
3          return 0
4      elif n == 1:
5          return 1
6      else:
7          return fib(n-1) + fib(n-2)
8
9
10
11
12
13
14
15 n=int(input("Enter terms for Fibonacci series: "))
16 for i in range (n):
17     print(fib(i),end=" ")
```

Terminal Test cases

1.2.3. Pattern - 1

19:59

Write a Python program to print a pattern of asterisks in the form of a right-angled triangle.

Input Format:
The input is an integer, representing the number of rows in the pattern.

Output Format
The output should display the pattern of asterisks (*), with each row containing an increasing number of asterisks.

Note:
Refer to the displayed test cases for the sample pattern.

Sample Test Cases +

Explorer

rightangl...

Submit

Debugger

```
1  n=int(input())
2  for i in range(1,n+1):
3      for j in range(0,i):
4          print("*",end=" ")
5      print()
```

Terminal Test cases

1.2.4. Pattern - 2

01:20

Write a Python program to print a right-angled triangle pattern of numbers.

Input Format:
The input is an integer, representing the number of rows in the pattern.

Output Format:
The output should display the pattern of numbers, with each row containing increasing numbers starting from 1 up to the row number.

Note:
Refer to the displayed test cases for the sample pattern.

Sample Test Cases +

Explorer

numberP...

Submit

```
1  n=int(input())
2  for i in range(1,n+1):
3      for j in range(1,i+1):
4          print(j,end=" ")
5      print()
```

Terminal Test cases

Debugger

2.1.1. List operations

15:36

Write a Python program that implements a menu-driven interface for managing a list of integers. The program should have the following menu options:

1. Add
2. Remove
3. Display
4. Quit

The program should repeatedly prompt the user to enter a choice from the menu. Depending on the choice selected, the program should perform the following actions:

- **Add**: Prompts the user to enter an integer and add it to the integer list. If the input is not a valid integer, display "Invalid input".
- **Remove:** Prompts the user to enter an integer to remove from the list. If the integer is found in the list, remove it; otherwise, display "Element not found". If the list is empty, display "List is empty".
- **Display**: Displays the current list of integers. If the list is empty, display "List is empty".
- **Quit**: Exits the program.
- The program should handle invalid menu choices by displaying "Invalid choice".Ensure that the program continues to prompt the user until they choose to quit (option 4).

Sample Test Cases +

Explorer

listOps.py

Submit

Debugger

```
a=[ ]
while True:
    print("1. Add")
    print("2. Remove")
    print("3. Display")
    print("4. Quit")
    n=int(input("Enter choice: "))
    if n==1:
        add=int(input("Integer: "))
        a.append(add)
        print(f"List after adding: {a}")
    elif n==2:
        if len(a)==0:
            print("List is empty")
        elif len(a)!=0:
            remove=int(input("Integer: "))
            if remove not in a :
                print("Element not found")
            else:
                a.remove(remove)
                print(f"List after removing: {a}")
    elif n==3:
        if len(a)==0:
            print("List is empty")
        else:
```

Terminal Test cases

< Prev Reset Submit Next >

2.1.2. Dictionary Operations

30:07

Write a Python program to perform the following dictionary operations:

- Create an empty dictionary and display it.
- Ask the user how many items to add, then input key-value pairs.
- Show the dictionary after adding items.
- Ask the user to update a key's value. Print "Value updated" if the key exists, otherwise print "Key not found".
- Retrieve and print a value using a key. If not found, print "Key not found".
- Use get() to retrieve a value. If the key doesn't exist, print "Key not found".
- Delete a key-value pair. If the key exists, delete and print "Deleted". If not, print "Key not found".
- Display the updated dictionary.

Note: Refer to visible test cases.

Sample Test Cases +

Explorer

dictOpera...

Submit

Debugger

```
1  dict = {}
2  print("Empty Dictionary:",dict)
3
4  n = int(input("Number of items: "))
5  for _ in range (n):
6      key = input("key: ")
7      value = input("value: ")
8      dict[key] = value
9  print("Dictionary:",dict)
10
11 update_key = input("Enter the key to update: ")
12 if update_key in dict:
13     new_value = input("Enter the new value: ")
14     dict[update_key] = new_value
15     print("Value updated")
16 else:
17     print("Key not found")
18
19 retrieve_key = input("Enter the key to retrieve: ")
20 if retrieve_key in dict:
21     print(f"Key: {retrieve_key}, Value: {dict[retrieve_key]}")
22 else:
23     print("Key not found")
24
25 get_key = input("Enter the key to get using the get() method: ")
```

Terminal Test cases

< Prev Reset Submit Next >

2.1.2. Dictionary Operations

30:07

Write a Python program to perform the following dictionary operations:

- Create an empty dictionary and display it.
- Ask the user how many items to add, then input key-value pairs.
- Show the dictionary after adding items.
- Ask the user to update a key's value. Print "Value updated" if the key exists, otherwise print "Key not found".
- Retrieve and print a value using a key. If not found, print "Key not found".
- Use get() to retrieve a value. If the key doesn't exist, print "Key not found".
- Delete a key-value pair. If the key exists, delete and print "Deleted". If not, print "Key not found".
- Display the updated dictionary.

Note: Refer to visible test cases.

Sample Test Cases +

Explorer

dictOpera...

Submit

Debugger

```
    print("Key not found")

retrieve_key = input("Enter the key to retrieve: ")
if retrieve_key in dict:
    print(f"Key: {retrieve_key}, Value: {dict[retrieve_key]}")
else:
    print("Key not found")

get_key = input("Enter the key to get using the get() method: ")
value = dict.get(get_key,"Key not found")
if value != "Key not found":
    print(f"Key: {get_key}, Value: {value}")
else:
    print(value)

deleted_key = input("Enter the key to delete: ")
if deleted_key in dict:
    del dict[deleted_key]
    print("Deleted")
else:
    print("Key not found")

print("Updated Dictionary:",dict)
```

Terminal Test cases

2.2.1. Linear search Technique

02:42

Write a program to check whether the given element is present or not in the array of elements using linear search.

Input format:

- The first line of input contains the array of integers which are separated by space
- The last line of input contains the key element to be searched

Output format:

- If the element is found, print the index.
- If the element is not found, print **Not found**.

Sample Test Case:
Input:
1 2 3 4 3 5 6
3
Output:
2

Sample Test Cases +

Explorer

CTP1709...

Submit

Debugger

```
1  arr = list(map(int,input().split(" ")))
2
3  key = int(input())
4
5  for i in range(len(arr)):
6      if arr[i] == key:
7          print(i)
8          break
9
10 if arr[i] != key:
11     print("Not found")
```

Terminal Test cases

2.2.2. Captain of the Team

01:03

You are provided with the heights of 11 cricket players (in centimeters). Your task is to identify the tallest player, who will be selected as the captain of the team.

Input Format:
The first line of input will contain 11 integers, each representing the height of a player (in centimeters), each separated by a space.

Output Format
The output should be the height (in centimeters) of the tallest player.

Sample Test Cases +

Explorer

captainof...

Submit

```
1  heights = list(map(int,input().split(" ")))
2
3  captain = max(heights)
4
5  print(captain)
```

Debugger

Terminal Test cases

< Prev Reset Submit Next >

3.1.1. Numpy array operations

18:31

Write a python program to demonstrate the usage of ndim, shape and size for a Numpy Array. The program should create a NumPy array using the entered elements and display it. Assume all input elements are valid numeric values.

Input Format:

- User inputs the number of rows and columns with space separated values.
- User inputs elements of the array row-wise followed line by line, separated by spaces.

Output Format:

- The created NumPy array based on the input dimensions and elements.
- Dimensions (ndim): Number of dimensions of the array.
- Shape: Tuple representing the shape of the array (number of rows, number of columns).
- Size: Total number of elements in the array.

Note: Use reshape() function to reshape the input array with the specified number of rows and columns.

Sample Test Cases +

Explorer

numpyarr...

Submit

```
1  import numpy as np
2  rows,cols= list(map(int,input().split()))
3  matrix= []
4  for i in range(rows):
5      row = list(map(int,input().split()))
6      matrix.append(row)
7  matrix= np.array(matrix).reshape(rows,cols)
8
9  print(matrix)
10 print(matrix.ndim)
11 print(matrix.shape)
12 print(matrix.size)
13
```

Terminal Test cases

Debugger

3.2.1. Numpy: Matrix Operations

04:12

The given code takes two 3×3 matrices, `matrix_a`, and `matrix_b`, as input from the user and converts them into NumPy arrays.

Task:

You are required to compute and display the results of the following matrix operations:

1. **Addition** (`matrix_a + matrix_b`)
2. **Subtraction** (`matrix_a - matrix_b`)
3. **Element-wise Multiplication** (`matrix_a * matrix_b`)
4. **Matrix Multiplication** (`matrix_a . matrix_b`)
5. **Transpose of Matrix A**

Input Format:

- The user will input 3 rows for `matrix_a`, each containing 3 integers separated by spaces.
- Similarly, the user will input 3 rows for `matrix_b`, each containing 3 integers separated by spaces.

Output Format:

The program should display the results of the operations in the following order:

1. The result of Addition.
2. The result of Subtraction.

Sample Test Cases +

Explorer

matrixOp...

Submit

Debugger

```
1  import numpy as np
2
3  # Input matrices
4  print("Enter Matrix A:")
5  matrix_a = np.array([list(map(int, input().split())) for i in range(3)])
6
7  print("Enter Matrix B:")
8  matrix_b = np.array([list(map(int, input().split())) for i in range(3)])
9
10
11 # Addition
12 print("Addition (A + B):")
13 print(matrix_a + matrix_b)
14 # Subtraction
15 print("Subtraction (A - B):")
16 print(matrix_a - matrix_b)
17 # Multiplication (element-wise)
18 print("Element-wise Multiplication (A * B):")
19 print(matrix_a * matrix_b)
20 # Matrix multiplication (dot product)
21 print("A dot B:")
22 print(np.dot(matrix_a,matrix_b))
23 # Transpose
24 print("Transpose of A:")
25 print(matrix_a.T)
```

Terminal Test cases

< Prev Reset Submit Next >

3.2.2. Numpy: Horizontal and Vertical Stacking of Arrays

03:05

You are given two arrays `arr1` and `arr2`. You need to perform horizontal and vertical stacking operations on them using NumPy.

- **Horizontal Stacking**: Stack the two matrices horizontally (side by side).
- **Vertical Stacking**: Stack the two matrices vertically (one below the other).

Input Format:

- The program should first prompt the user to input two 3x3 arrays.
- Each array consists of 3 rows, and each row contains 3 space-separated integers.
- The user will input the two arrays row by row.

Output Format:

- The program should display the result of the Horizontal Stack (side-by-side stacking) of the two arrays.
- The program should then display the result of the Vertical Stack (one below the other) of the two arrays.

Sample Test Cases +

Explorer

stacking.py

Submit

```python
import numpy as np

# Input matrices
print("Enter Array1:")
arr1 = np.array([list(map(int, input().split())) for i in range(3)])

print("Enter Array2:")
arr2 = np.array([list(map(int, input().split())) for i in range(3)])

# Perform horizontal stacking (hstack)
horizontal_stack = np.hstack((arr1,arr2))


# Perform vertical stacking (vstack)
vertical_stack = np.vstack((arr1,arr2))
print("Horizontal Stack:")
print(horizontal_stack)
print("Vertical Stack:")
print(vertical_stack)
```

Debugger

Terminal Test cases

< Prev Reset Submit Next >

3.2.3. Numpy: Custom Sequence Generation

00:53

Write a Python program that takes the following inputs from the user:

- Start value: The starting point of the sequence.
- Stop value: The sequence should end before this value.
- Step value: The increment between each number in the sequence.

The program should then generate a sequence using `numpy` based on these inputs and print the generated sequence.

Input Format:

- The user will input three integer values: start, stop, and step, each on a new line.

Output Format:

- The program should print the generated sequence based on the input values.

Sample Test Cases +

Explorer

customS...

Submit

```
1  import numpy as np
2
3  # Take user input for the start, stop, and step of the sequence
4  start = int(input())
5  stop = int(input())
6  step = int(input())
7
8  # Generate the sequence using np.arange()
9  sequence = np.arange(start, stop, step)
10 # Print the generated sequence
11 print(sequence)
```

Terminal Test cases

Debugger

3.2.4. Numpy: Arithmetic and Statistical Operations, Mathematical Operations, Bitw... 02:45

You are given two arrays `A` and `B`. Your task is to complete the function `array_operations`, which will convert these lists into NumPy arrays and perform the following operations:

1. **Arithmetic Operations**:
   - Compute the element-wise sum, difference, and product of the two arrays.
2. **Statistical Operations**:
   - Calculate the mean, median, and standard deviation of array `A`.
3. **Bitwise Operations**:
   - Perform bitwise `AND`, bitwise `OR`, and bitwise `XOR` on the arrays (ex: A_i `OR` B_i).

Input Format:
- The first line contains space-separated integers representing the elements of array `A`.
- The second line contains space-separated integers representing the elements of array `B`.

Output Format:
- For each operation (arithmetic, statistical, and bitwise), print the results in the specified format as shown in sample test cases.

Sample Test Cases +

Explorer

different...

Submit

```python
import numpy as np

def array_operations(A, B):

    # Convert A and B to NumPy arrays
    A = np.array(A)
    B = np.array(B)
    # Arithmetic Operations
    sum_result = A + B
    diff_result = A - B
    prod_result = A * B

    # Statistical Operations
    mean_A = np.mean(A)
    median_A = np.median(A)
    std_dev_A = np.std(A)

    # Bitwise Operations
    and_result = A & B
    or_result = A | B
    xor_result = A ^ B

    # Output results with one space between each element
    print("Element-wise Sum:", ' '.join(map(str, sum_result)))
    print("Element-wise Difference:", ' '.join(map(str, diff_result)))
```

Terminal Test cases

Debugger

3.2.5. Numpy: Copying and Viewing Arrays

02:33

The given code takes a list of integers as input and converts it into a NumPy array. Your task is to complete the code by:

- Creating a view of the `original_array` and assigning it to `view_array`.
- Creating a copy of the `original_array` and assigning it to `copy_array`.

After completing these steps, observe how modifying the view affects the `original_array`, while modifying the copy does not.

Input Format:

- A single line of space-separated integers.

Output Format:

- After modifying the view:

```
Original array after modifying view: <original_array>
View array: <view_array>
```

- After modifying the copy:

```
Original array after modifying copy: <original_array>
Copy array: <copy_array>
```

Sample Test Cases +

Explorer

copyAnd...

Submit

Debugger

```python
import numpy as np

inputlist = list(map(int,input().split(" ")))

# Original array
original_array = np.array(inputlist)

# Create a view
view_array =original_array.view()

# Create a copy
copy_array =original_array.copy()

# Modify the view
view_array[0] = 99
print("Original array after modifying view:", original_array)
print("View array:", view_array)

# Modify the copy
copy_array[1] = 88
print("Original array after modifying copy:", original_array)
print("Copy array:", copy_array)

```

Terminal Test cases

3.2.6. Numpy: Searching, Sorting, Counting, Broadcasting

04:00

The given code in the editor takes a single array, `array1`, as space-separated integers as input from the user.
Additionally, it takes the following inputs:

- `search_value`: The value to search for in the array.
- `count_value`: The value to count its occurrences in the array.
- `broadcast_value`: The value to add for broadcasting across the array.

You need to complete the code to perform the following operations:

1. **Searching**: Find the indices where `search_value` appears in `array1` and print these indices.
2. **Counting**: Count how many times `count_value` appears in `array1` and print the count.
3. **Broadcasting**: Add `broadcast_value` to each element of `array1` using broadcasting, and print the resulting array.
4. **Sorting**: Sort `array1` in ascending order and print the sorted array.

Input Format:

1. A single line containing space-separated integers representing `array1`.
2. An integer `search_value` represents the value to search for in the array.
3. An integer `count_value` represents the value to count in the array.
4. An integer `broadcast_value` represents the value to add to each element of the array.

Sample Test Cases +

Explorer

arrayOpe...

Submit

```
1  import numpy as np
2
3  # Input array from the user
4  array1 = np.array(list(map(int, input().split())))
5
6  # Searching
7  search_value = int(input("Value to search: "))
8  count_value = int(input("Value to count: "))
9  broadcast_value = int(input("Value to add: "))
10
11 # Find indices where value matches in array1
12 a=np.where(array1==search_value)[0]
13 print(a)
14 # Count occurrences in array1
15 b=np.count_nonzero(array1==count_value)
16 print(b)
17 # Broadcasting addition
18 c=array1+broadcast_value
19 print(c)
20 # Sort the first array
21 d=np.sort(array1)
22 print(d)
```

Terminal Test cases

Debugger

< Prev Reset Submit Next >

3.2.7. Student Data Analysis and Operations

59:45

Write a Python program that takes the file name of a CSV file containing student details, including roll numbers and their marks in three subjects as input, reads the data, and performs the following operations:

- **Print all student details**: Display the complete details of all students, including roll numbers and marks for all subjects.
- **Find total students**: Determine the total number of students in the dataset.
- **Print all student roll numbers**: Extract and print the roll numbers of all students.
- **Print Subject 1 marks**: Extract and print the marks of all students in Subject 1.
- **Find minimum marks in Subject 2**: Identify the lowest marks in Subject 2.
- **Find maximum marks in Subject 3**: Identify the highest marks in Subject 3.
- **Print all subject marks**: Display the marks of all students for each subject.
- **Find total marks of students**: Compute the total marks for each student across all subjects.
- **Find the average marks of each student**: Compute the average marks for each student.
- **Find average marks of each subject**: Compute the average marks for all students in each subject.
- **Find average marks of Subject 1 and Subject 2**: Compute the average marks for Subject 1 and Subject 2.
- **Find average marks of Subject 1 and Subject 3**: Compute the average marks for Subject 1 and Subject 3.
- **Find the roll number of the student with maximum marks in Subject 3**: Identify the student with the highest marks in Subject 3 and print their roll number.

Sample Test Cases +

Explorer

Operatio...

Submit

Debugger

```python
import numpy as np

a = np.loadtxt("Sample.csv", delimiter=',', skiprows=1)


# 1. Print all student details
print("All student Details:\n",a)

# 2. print total students
r,c=a.shape
print("Total Students:",r)

# 3. Print all student Roll numbers
print("All Student Roll Nos",a[:,0])

# 4. Print subject 1 marks
print("Subject 1 Marks", a[:,1])

# 5. print minimum marks of Subject 2
print("Min marks in Subject 2", np.min(a[:,2]))

# 6. print maximum marks of Subject 3
print("Max marks in Subject 3",np.max(a[:,3]))

# 7. Print All subject marks
```

Terminal Test cases

< Prev Reset Submit Next >

3.2.7. Student Data Analysis and Operations

59:45

Write a Python program that takes the file name of a CSV file containing student details, including roll numbers and their marks in three subjects as input, reads the data, and performs the following operations:

- **Print all student details**: Display the complete details of all students, including roll numbers and marks for all subjects.
- **Find total students**: Determine the total number of students in the dataset.
- **Print all student roll numbers**: Extract and print the roll numbers of all students.
- **Print Subject 1 marks**: Extract and print the marks of all students in Subject 1.
- **Find minimum marks in Subject 2**: Identify the lowest marks in Subject 2.
- **Find maximum marks in Subject 3**: Identify the highest marks in Subject 3.
- **Print all subject marks**: Display the marks of all students for each subject.
- **Find total marks of students**: Compute the total marks for each student across all subjects.
- **Find the average marks of each student**: Compute the average marks for each student.
- **Find average marks of each subject**: Compute the average marks for all students in each subject.
- **Find average marks of Subject 1 and Subject 2**: Compute the average marks for Subject 1 and Subject 2.
- **Find average marks of Subject 1 and Subject 3**: Compute the average marks for Subject 1 and Subject 3.
- **Find the roll number of the student with maximum marks in Subject 3**: Identify the student with the highest marks in Subject 3 and print their roll number.

Sample Test Cases +

Explorer

Operatio...

Submit

Debugger

```
25  # 7. Print All subject marks
26  print("All subject marks:", a[:,1:])
27
28  # 8. print Total marks of students
29  print("Total Marks",np.sum(a[:,1:],axis=1))
30
31  # 9. print average marks of each student
32  avg=np.mean(a[:,1:],axis=1)
33  print(np.round(avg,1))
34  # 10. print average marks of each subject
35  print("Average Marks of each subject",np.mean(a[:,1:],axis=0))
36
37  # 11. print average marks of S1 and S2
38  print("Average Marks of S1 and S2",np.mean(a[:,1:3],axis=0))
39
40  # 12. print average marks of S1 and S3
41  print("Average Marks of S1 and S3",np.mean(a[:,[1,3]],axis=0))
42
43  # 13. print Roll number who got maximum marks in Subject 3
44  i=np.argmax(a[:,3])
45  print("Roll no who got maximum marks in Subject 3",a[i,0])
46
47  # 14. print Roll number who got minimum marks in Subject 2
48  mn=np.argmin(a[:,2])
49  print("Roll no who got minimum marks in Subject 2",a[mn,0])
```

Terminal Test cases

3.2.7. Student Data Analysis and Operations

59:45

Write a Python program that takes the file name of a CSV file containing student details, including roll numbers and their marks in three subjects as input, reads the data, and performs the following operations:

- **Print all student details**: Display the complete details of all students, including roll numbers and marks for all subjects.
- **Find total students**: Determine the total number of students in the dataset.
- **Print all student roll numbers**: Extract and print the roll numbers of all students.
- **Print Subject 1 marks**: Extract and print the marks of all students in Subject 1.
- **Find minimum marks in Subject 2**: Identify the lowest marks in Subject 2.
- **Find maximum marks in Subject 3**: Identify the highest marks in Subject 3.
- **Print all subject marks**: Display the marks of all students for each subject.
- **Find total marks of students**: Compute the total marks for each student across all subjects.
- **Find the average marks of each student**: Compute the average marks for each student.
- **Find average marks of each subject**: Compute the average marks for all students in each subject.
- **Find average marks of Subject 1 and Subject 2**: Compute the average marks for Subject 1 and Subject 2.
- **Find average marks of Subject 1 and Subject 3**: Compute the average marks for Subject 1 and Subject 3.
- **Find the roll number of the student with maximum marks in Subject 3**: Identify the student with the highest marks in Subject 3 and print their roll number.

Sample Test Cases +

Explorer

Operatio...

Submit

Debugger

```
# 15. print Roll number who got 24 marks in Subject 2
whr=np.where(a[:,2]==24)
print("Roll no who got 24 marks in Subject 2",a[whr,0])

# 16. print count of students who got marks in Subject 1 < 40
ct=np.count_nonzero(a[:,1]<40)
print("Count of students who got marks in Subject 1 < 40",ct)

# 17. print count of students who got marks in Subject 2 > 90
ct1=np.count_nonzero(a[:,2]>90)
print("Count of students who got marks in Subject 2 > 90:",ct1)

# 18. print count of students in each subject who got marks >= 90

print("Count of students in each subject who got marks >= 90:" ,np.
count_nonzero(a[:,1:]>=90,axis=0))

# 19. print count of subjects in which each student got marks >= 90
print("Roll no:",a[:,0])
print("Count of subjects in which student got marks >=
90:",np.count_nonzero(a[:,1:]>=90,axis=1))

# 20. Print S1 marks in ascending order
srt=np.sort(a[:,1])
print(srt)
```

Terminal Test cases

< Prev Reset Submit Next >

3.2.7. Student Data Analysis and Operations

59:45

Write a Python program that takes the file name of a CSV file containing student details, including roll numbers and their marks in three subjects as input, reads the data, and performs the following operations:

- **Print all student details**: Display the complete details of all students, including roll numbers and marks for all subjects.
- **Find total students**: Determine the total number of students in the dataset.
- **Print all student roll numbers**: Extract and print the roll numbers of all students.
- **Print Subject 1 marks**: Extract and print the marks of all students in Subject 1.
- **Find minimum marks in Subject 2**: Identify the lowest marks in Subject 2.
- **Find maximum marks in Subject 3**: Identify the highest marks in Subject 3.
- **Print all subject marks**: Display the marks of all students for each subject.
- **Find total marks of students**: Compute the total marks for each student across all subjects.
- **Find the average marks of each student**: Compute the average marks for each student.
- **Find average marks of each subject**: Compute the average marks for all students in each subject.
- **Find average marks of Subject 1 and Subject 2**: Compute the average marks for Subject 1 and Subject 2.
- **Find average marks of Subject 1 and Subject 3**: Compute the average marks for Subject 1 and Subject 3.
- **Find the roll number of the student with maximum marks in Subject 3**: Identify the student with the highest marks in Subject 3 and print their roll number.

Sample Test Cases +

Explorer

Operatio...

Submit

Debugger

```
61  print("Count of students who got marks in Subject 2 > 90:",ct1)
62
63  # 18. print count of students in each subject who got marks >= 90
64
65  print("Count of students in each subject who got marks >= 90:" ,np.
    count_nonzero(a[:,1:]>=90,axis=0))
66
67  # 19. print count of subjects in which each student got marks >= 90
68  print("Roll no:",a[:,0])
69  print("Count of subjects in which student got marks >=
    90:",np.count_nonzero(a[:,1:]>=90,axis=1))
70
71  # 20. Print S1 marks in ascending order
72  srt=np.sort(a[:,1])
73  print(srt)
74  # 21. Print S1 marks >= 50 and <= 90
75  print(a[(a[:,1]>=50)& (a[:,1]<90)])
76  print(a)
77  ip=np.where(a[:,1]==79)
78  print(ip)
79
80
81
82
83
```

Terminal Test cases

< Prev Reset Submit Next >

4.1.1. Pandas - series creation and manipulation

47:04

Write a Python program that takes a list of numbers from the user, creates a Pandas series from it, and then calculates the mean of even and odd numbers separately using the **groupby** and **mean()** operations.

Input Format:

- The user should enter a list of numbers separated by space when prompted.

Output Format:

- The program should display the mean of even and odd numbers separately.
- Each mean value should be displayed with a label indicating whether it corresponds to even or odd numbers.

Sample Test Cases +

Explorer

seriesMa...

Submit

```
1  import pandas as pd
2
3  # Take inputs from the user to create a list of numbers
4  numbers = list(map(int, input().split()))
5
6  # Create a Pandas series from the list of numbers
7  data_list=numbers
8  series_from_list=pd.Series(data_list)
9  # Grouping by even and odd numbers and calculating the mean
10 grouped =series_from_list.groupby(series_from_list % 2 == 0).mean()
11
12 # Display the mean of even and odd numbers with labels
13 grouped.index = ['Even' if is_even else 'Odd' for is_even in
   grouped.index]
14 print("Mean of even and odd numbers:")
15 print(grouped)
16
```

Debugger

Terminal Test cases

4.1.2. Dictionary to dataframe

48:30

A dictionary of lists has been provided to you in the editor. Create a DataFrame from the dictionary of lists and perform the listed operations, then display the DataFrame before and after each manipulation.

Create the DataFrame:

- Convert the dictionary to a Pandas DataFrame.

Add a new row:

- Take inputs from the user for the new row data (name, age).
- Add the new row to the DataFrame.
- Display the DataFrame after adding the new row.

Modify a row:

- Modify a specific row by changing the age. Take the row index and new age value from the user.
- Display the DataFrame after modifying the row.

Delete a row:

- Take the row index to be deleted from the user.
- Remove the specified row.
- Display the DataFrame after deleting the row.

Add a new column:

Sample Test Cases +

Explorer

datafram...

Submit

Debugger

```
1  import pandas as pd
2
3  # Provided dictionary of lists
4  data = {
5      'Name': ['Alice', 'Bob', 'Charlie'],
6      'Age': [25, 30, 35],
7  }
8
9  # Convert the dictionary to a DataFrame
10 df = pd.DataFrame(data)
11
12 # Display the original DataFrame
13 print("Original DataFrame:")
14 print(df)
15
16 # Adding a new row
17 new_name=input("New name: ")
18 new_age=int(input("New age: "))
19 new_row={'Name': new_name,'Age':new_age}
20 df=pd.concat([df,pd.DataFrame([new_row])],ignore_index=True)
21 # Display the DataFrame after adding a new row
22 print("After adding a row:\n",df)
23
24 # Modifying a row
25 modify_index=int(input("Index of row to modify: "))
```

Terminal Test cases

4.1.2. Dictionary to dataframe

48:30

A dictionary of lists has been provided to you in the editor. Create a DataFrame from the dictionary of lists and perform the listed operations, then display the DataFrame before and after each manipulation.

Create the DataFrame:

- Convert the dictionary to a Pandas DataFrame.

Add a new row:

- Take inputs from the user for the new row data (name, age).
- Add the new row to the DataFrame.
- Display the DataFrame after adding the new row.

Modify a row:

- Modify a specific row by changing the age. Take the row index and new age value from the user.
- Display the DataFrame after modifying the row.

Delete a row:

- Take the row index to be deleted from the user.
- Remove the specified row.
- Display the DataFrame after deleting the row.

Add a new column:

Sample Test Cases +

Explorer

dataframe...

Submit

Debugger

```
32  # Deleting a row
33  delete_index=int(input("Index of row to delete: "))
34  df=df.drop(delete_index).reset_index(drop=True)
35  # Display the DataFrame after deleting a row
36  print("After deleting a row:")
37  print(df)
38
39  # Adding a new column
40  gender_input=input("Enter genders separated by space: ")
41  genders=gender_input.split()
42  df["Gender"]=genders
43  # Display the DataFrame after adding a new column
44  print("After adding a new column:")
45  print(df)
46
47  # Modifying a column
48  df["Name"]=df["Name"].str.upper()
49  # Display the DataFrame after modifying a column
50  print("After modifying a column:")
51  print(df)
52
53  # Deleting a column
54  df=df.drop(columns=['Age'])
55  # Display the DataFrame after deleting a column
56  print("After deleting a column:")
57  print(df)
```

Terminal Test cases

< Prev Reset Submit Next >

4.1.3. Student Information

05:05

Write a program to read a text file containing student information (name, age, and grade) using Pandas. Perform the following tasks:

- Display the first five rows of the data frame.
- Calculate the average age of the students(limit the average age up to 2 decimal places).
- Filter out the students who have a grade above a certain threshold(consider the threshold grade is **'B'**).

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

Sample Test Cases +

Explorer

studentin... studentdat...

Submit

```python
import pandas as pd

# Read the text file into a DataFrame
file = input()
data = pd.read_csv(file, sep="\s+", header=None, names=["Name", "Age", "Grade"])
print("First five rows:")
print(data.head(5))
# write your code here..
age=round(data['Age'].mean(),2)
print("Average age:",age)
print("Students with a grade up to B")
df=pd.DataFrame(data)
a=df[df['Grade']<='B']
print(a)
```

Debugger

Terminal Test cases

4.2.1. Month with the Highest Total Sales

04:13

Write a Python program that takes the file name of a CSV file as input, reads the data, and performs the following operations:

- The CSV file contains the columns: `Date`, `Product`, `Quantity`, `Price`, and `City`.
- Group the data by Month and calculate the total sales for each month.
- Find the month with the highest total sales and display it.
- Also, display the total sales for the best month.

Sample Data:

```
Date,Product,Quantity,Price,City
2025-01-01,Product A,5,20,New York
2025-01-01,Product B,3,15,Los Angeles
2025-01-02,Product A,7,20,New York
2025-01-02,Product C,4,30,Chicago
2025-01-03,Product B,2,15,Chicago
2025-01-03,Product A,8,20,Los Angeles
2025-01-04,Product C,6,30,New York
2025-01-04,Product B,5,15,Los Angeles
2025-01-05,Product A,3,20,Chicago
2025-01-05,Product C,10,30,Los Angeles
```

Note:

Sample Test Cases +

Explorer

monthFor... sales_dat...

Submit

```python
import pandas as pd

# Prompt the user for the file name
file_name = input()

# Load the data
df = pd.read_csv(file_name)
df['sales']=df['Quantity'].multiply(df['Price'])
df['month']=pd.to_datetime(df['Date']).dt.strftime("%Y-%m")
# Find the month with the highest total sales
best_month =df.groupby('month')['sales'].sum().idxmax()
highest_sales = df['sales'].sum()

print(f"Best month: {best_month}")
print(f"Total sales: ${highest_sales:.2f}")

```

Debugger

Terminal Test cases

< Prev Reset Submit Next >

4.2.2. Best Selling Product

25:34

Write a Python program that takes the file name of a CSV file as input, reads the data, and performs the following operations:

- The CSV file contains the columns: `Date`, `Product`, `Quantity`, `Price`, and `City`.
- Find the product that sold the most in terms of quantity sold.
- Display the product that sold the most and the total quantity sold for that product.

Sample Data:

```
Date,Product,Quantity,Price,City
2025-01-01,Product A,5,20,New York
2025-01-01,Product B,3,15,Los Angeles
2025-01-02,Product A,7,20,New York
2025-01-02,Product C,4,30,Chicago
2025-01-03,Product B,2,15,Chicago
2025-01-03,Product A,8,20,Los Angeles
2025-01-04,Product C,6,30,New York
2025-01-04,Product B,5,15,Los Angeles
2025-01-05,Product A,3,20,Chicago
2025-01-05,Product C,10,30,Los Angeles
```

Note:
The data cannot be displayed in the file. You can refer to the sample data provided for insights.

Sample Test Cases +

Explorer

monthFor... sales_dat...

Submit

```python
import pandas as pd

# Prompt the user for the file name
file_name = input()

# Load the data
df = pd.read_csv(file_name)


# Find the product with the highest total quantity sold
product_sales = df.groupby("Product")["Quantity"].sum()
best_product = product_sales.idxmax()
highest_quantity = product_sales.max()

# Display the result
print(f"Best selling product: {best_product}")
print(f"Total quantity sold: {highest_quantity}")

```

Terminal Test cases

Debugger

4.2.3. City that Sold the Most Products

02:00

Write a Python program that takes the file name of a CSV file as input, reads the data, and performs the following operations:

- The CSV file contains the columns: `Date`, `Product`, `Quantity`, `Price`, and `City`.
- Group the data by `City` and calculate the total quantity of products sold for each city.
- Find the city that sold the most products (based on the total quantity sold).

Sample Data:

```
Date,Product,Quantity,Price,City
2025-01-01,Product A,5,20,New York
2025-01-01,Product B,3,15,Los Angeles
2025-01-02,Product A,7,20,New York
2025-01-02,Product C,4,30,Chicago
2025-01-03,Product B,2,15,Chicago
2025-01-03,Product A,8,20,Los Angeles
2025-01-04,Product C,6,30,New York
2025-01-04,Product B,5,15,Los Angeles
2025-01-05,Product A,3,20,Chicago
2025-01-05,Product C,10,30,Los Angeles
```

Note:
The data cannot be displayed in the file. You can refer to the sample data provided for insights.

Sample Test Cases +

Explorer

monthFor... sales_dat...

Submit

```python
import pandas as pd

# Prompt the user for the file name
file_name = input()

# Load the data
df = pd.read_csv(file_name)

# write the code..
city_sales = df.groupby("City")["Quantity"].sum()


best_city = city_sales.idxmax()
# Display the result
print(f"City sold the most products: {best_city}")

```

Debugger

Terminal Test cases

< Prev Reset Submit Next >

4.2.4. Most Frequently Sold Product Pairs

24:43

Write a Python program that takes the file name of a CSV file as input, reads the data, and performs the following operations:

- The CSV file contains the following columns: Date, Product, Quantity, Price, and City.
- For each date, find all pairs of products that were sold together (i.e., two products sold on the same date).
- Output the product pair/s that was sold most frequently.

Sample Data:

```
Date,Product,Quantity,Price,City
2025-01-01,Product A,5,20,New York
2025-01-01,Product B,3,15,Los Angeles
2025-01-02,Product A,7,20,New York
2025-01-02,Product C,4,30,Chicago
2025-01-03,Product B,2,15,Chicago
2025-01-03,Product A,8,20,Los Angeles
2025-01-04,Product C,6,30,New York
2025-01-04,Product B,5,15,Los Angeles
2025-01-05,Product A,3,20,Chicago
2025-01-05,Product C,10,30,Los Angeles
```

Explanation:
Transactions:

Sample Test Cases +

Explorer

frequentl... sales_dat...

Submit

```python
import pandas as pd
from itertools import combinations
from collections import Counter

# Prompt user to input the file name
file_name = input()

# Read data from the specified CSV file
df = pd.read_csv(file_name)

# write the code

date_products = {}

for date, group in df.groupby('Date'):
    products = group['Product'].unique()
    if len(products) > 1:
        date_products[date] = products

pair_counter = Counter()

for products in date_products.values():
    pairs = combinations(sorted(products), 2)
    pair_counter.update(pairs)

```

Terminal Test cases

Debugger

Prev Reset Submit Next

4.2.4. Most Frequently Sold Product Pairs

24:43

Write a Python program that takes the file name of a CSV file as input, reads the data, and performs the following operations:

- The CSV file contains the following columns: Date, Product, Quantity, Price, and City.
- For each date, find all pairs of products that were sold together (i.e., two products sold on the same date).
- Output the product pair/s that was sold most frequently.

Sample Data:

```
Date,Product,Quantity,Price,City
2025-01-01,Product A,5,20,New York
2025-01-01,Product B,3,15,Los Angeles
2025-01-02,Product A,7,20,New York
2025-01-02,Product C,4,30,Chicago
2025-01-03,Product B,2,15,Chicago
2025-01-03,Product A,8,20,Los Angeles
2025-01-04,Product C,6,30,New York
2025-01-04,Product B,5,15,Los Angeles
2025-01-05,Product A,3,20,Chicago
2025-01-05,Product C,10,30,Los Angeles
```

Explanation:
Transactions:

Sample Test Cases +

Explorer

frequentI... sales_dat... Submit

Debugger

```
10
11  # write the code
12
13  date_products = {}
14
15  for date, group in df.groupby('Date'):
16      products = group['Product'].unique()
17      if len(products) > 1:
18          date_products[date] = products
19
20  pair_counter = Counter()
21
22  for products in date_products.values():
23      pairs = combinations(sorted(products), 2)
24      pair_counter.update(pairs)
25
26  if pair_counter:
27      max_count = max(pair_counter.values())
28
29      for pair, count in pair_counter.items():
30          if count == max_count:
31              print(f"{pair[0]} and {pair[1]}: {count} times")
32  else:
33      print("No product pairs found.")
34  # Output the most frequent product pairs
```

Terminal Test cases

4.2.5. Titanic Dataset Analysis and Data Cleaning

17:59

You are provided with the Titanic dataset containing information about passengers on the Titanic. Your task is to write Python code to answer the following questions based on the dataset. For each question, perform necessary data cleaning, transformations, and calculations as required.

1. Display the first 5 rows of the dataset.
2. Display the last 5 rows of the dataset.
3. Get the shape of the dataset (number of rows and columns).
4. Get a summary of the dataset (using .info()).
5. Get basic statistics (mean, standard deviation, etc.) of the dataset using .describe().
6. Check for missing values and display the count of missing values for each column.
7. Fill missing values in the 'Age' column with the median age.
8. Fill missing values in the 'Embarked' column with the most frequent value (mode).
9. Drop the 'Cabin' column due to many missing values.
10. Create a new column, 'FamilySize' by adding the 'SibSp' and 'Parch' columns.

The Titanic dataset contains columns as shown below,

Passengerid	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Test Cases +

Explorer

titanicDat...

Submit

```python
import pandas as pd
import numpy as np

# Load the Titanic dataset
data = pd.read_csv('Titanic-Dataset.csv')

print(data.head())

# 2. Display the last 5 rows of the dataset
print(data.tail())

# 3. Get the shape of the dataset
print(data.shape)

# 4. Get a summary of the dataset (info)
print(data.info())

# 5. Get basic statistics of the dataset
print(data.describe())

# 6. Check for missing values
print(data.isnull().sum())

# 7. Fill missing values in the 'Age' column with the median age
median_age = data['Age'].median()
```

Terminal Test cases

Debugger

4.2.5. Titanic Dataset Analysis and Data Cleaning

17:59

You are provided with the Titanic dataset containing information about passengers on the Titanic. Your task is to write Python code to answer the following questions based on the dataset. For each question, perform necessary data cleaning, transformations, and calculations as required.

1. Display the first 5 rows of the dataset.
2. Display the last 5 rows of the dataset.
3. Get the shape of the dataset (number of rows and columns).
4. Get a summary of the dataset (using .info()).
5. Get basic statistics (mean, standard deviation, etc.) of the dataset using .describe().
6. Check for missing values and display the count of missing values for each column.
7. Fill missing values in the 'Age' column with the median age.
8. Fill missing values in the 'Embarked' column with the most frequent value (mode).
9. Drop the 'Cabin' column due to many missing values.
10. Create a new column, 'FamilySize' by adding the 'SibSp' and 'Parch' columns.

The Titanic dataset contains columns as shown below,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Test Cases +

Explorer

titanicDat...

Submit

Debugger

```
15  # 4. Get a summary of the dataset (info)
16  print(data.info())
17
18  # 5. Get basic statistics of the dataset
19  print(data.describe())
20
21  # 6. Check for missing values
22  print(data.isnull().sum())
23
24  # 7. Fill missing values in the 'Age' column with the median age
25  median_age = data['Age'].median()
26  data['Age'].fillna(median_age, inplace=True)
27
28  # 8. Fill missing values in the 'Embarked' column with the mode
29  mode_embarked = data['Embarked'].mode()[0]
30  data['Embarked'].fillna(mode_embarked, inplace=True)
31
32  # 9. Drop the 'Cabin' column due to many missing values
33  data.drop('Cabin', axis=1, inplace=True)
34
35  # 10. Create a new column 'FamilySize' by adding 'SibSp' and 'Parch'
36  data['FamilySize'] = data['SibSp'] + data['Parch']
37
38
39
```

Terminal Test cases

4.2.6. Titanic Dataset Analysis and Data Cleaning - 2

18:54

You are provided with the Titanic dataset containing information about passengers on the Titanic. Your task is to write Python code to answer the following questions based on the dataset.

1. Create a new column 'IsAlone' which is 1 if the passenger is alone (FamilySize = 0), otherwise 0.
2. Convert the 'Sex' column to numeric values (male: 0, female: 1).
3. One-hot encode the 'Embarked' column, dropping the first category.
4. Get the mean age of passengers.
5. Get the median fare of passengers.
6. Get the number of passengers by class.
7. Get the number of passengers by gender.
8. Get the number of passengers by survival status.
9. Calculate the survival rate of passengers.
10. Calculate the survival rate by gender.

The Titanic dataset contains columns as shown below,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Test Cases +

Explorer titanicDat... Submit Debugger

```python
import pandas as pd
import numpy as np

# Load the Titanic dataset
data = pd.read_csv('Titanic-Dataset.csv')
data['FamilySize'] = data['SibSp'] + data['Parch']

data['IsAlone'] = np.where(data['FamilySize'] == 0, 1, 0)

# 2. Convert 'Sex' to numeric (male: 0, female: 1)
data['Sex'] = data['Sex'].map({'male': 0, 'female': 1})

# 3. One-hot encode the 'Embarked' column
data = pd.get_dummies(data, columns=['Embarked'], drop_first=True)

# 4. Get the mean age of passengers
mean_age = data['Age'].mean()
print(mean_age)

# 5. Get the median fare of passengers
median_fare = data['Fare'].median()
print(median_fare)

# 6. Get the number of passengers by class
passengers_by_class = data['Pclass'].value_counts()
```

Terminal Test cases

4.2.6. Titanic Dataset Analysis and Data Cleaning - 2

18:54

You are provided with the Titanic dataset containing information about passengers on the Titanic. Your task is to write Python code to answer the following questions based on the dataset.

1. Create a new column 'IsAlone' which is 1 if the passenger is alone (FamilySize = 0), otherwise 0.
2. Convert the 'Sex' column to numeric values (male: 0, female: 1).
3. One-hot encode the 'Embarked' column, dropping the first category.
4. Get the mean age of passengers.
5. Get the median fare of passengers.
6. Get the number of passengers by class.
7. Get the number of passengers by gender.
8. Get the number of passengers by survival status.
9. Calculate the survival rate of passengers.
10. Calculate the survival rate by gender.

The Titanic dataset contains columns as shown below,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Test Cases +

Explorer

titanicDat...

Submit

Debugger

```
19
20  # 5. Get the median fare of passengers
21  median_fare = data['Fare'].median()
22  print(median_fare)
23
24  # 6. Get the number of passengers by class
25  passengers_by_class = data['Pclass'].value_counts()
26  print(passengers_by_class)
27
28  # 7. Get the number of passengers by gender
29  passengers_by_gender = data['Sex'].value_counts().sort_index()
30  print(passengers_by_gender)
31
32  # 8. Get the number of passengers by survival status
33  passengers_by_survival = data['Survived'].value_counts().sort_index()
34  print(passengers_by_survival)
35
36  # 9. Calculate the survival rate
37  survival_rate = data['Survived'].mean()
38  print(survival_rate)
39
40  # 10. Calculate the survival rate by gender
41  survival_rate_by_gender = data.groupby('Sex')['Survived'].mean()
42  print(survival_rate_by_gender)
43
```

Terminal Test cases

< Prev Reset Submit Next >

4.2.7. Titanic Dataset Analysis and Data Cleaning - 3

08:18

You are provided with the Titanic dataset containing information about passengers on the Titanic. Your task is to write Python code to answer the following questions based on the dataset.

1. Calculate the survival rate by class.
2. Calculate the survival rate by embarkation location (Embarked_S).
3. Calculate the survival rate by family size (FamilySize).
4. Calculate the survival rate by being alone (IsAlone).
5. Get the average fare by passenger class (Pclass).
6. Get the average age by passenger class (Pclass).
7. Get the average age by survival status (Survived).
8. Get the average fare by survival status (Survived).
9. Get the number of survivors by class (Pclass).
10. Get the number of non-survivors by class (Pclass).

The Titanic dataset contains columns as shown below,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Test Cases +

Explorer titanicDat... Submit Debugger

```
13  print(data.groupby('Embarked_S')['Survived'].mean())
14
15  # 3. Calculate the survival rate by family size
16  print(data.groupby('FamilySize')['Survived'].mean())
17
18  # 4. Calculate the survival rate by being alone
19  print(data.groupby('IsAlone')['Survived'].mean())
20
21  # 5. Get the average fare by class
22  print(data.groupby('Pclass')['Fare'].mean())
23
24  # 6. Get the average age by class
25  print(data.groupby('Pclass')['Age'].mean())
26
27  # 7. Get the average age by survival status
28  print(data.groupby('Survived')['Age'].mean())
29
30  # 8. Get the average fare by survival status
31  print(data.groupby('Survived')['Fare'].mean())
32
33  # 9. Get the number of survivors by class
34  print(data[data['Survived']==1]['Pclass'].value_counts())
35
36  # 10. Get the number of non-survivors by class
37  print(data[data['Survived'] == 0]['Pclass'].value_counts())
```

Terminal Test cases

< Prev Reset Submit Next >

4.2.8. Titanic Dataset Analysis and Data Cleaning - 4

30:28

You are provided with the Titanic dataset containing information about passengers on the Titanic. Your task is to write Python code to answer the following questions based on the dataset.

1. Get the number of survivors by gender (Sex).
2. Get the number of non-survivors by gender (Sex).
3. Get the number of survivors by embarkation location (Embarked_S).
4. Get the number of non-survivors by embarkation location (Embarked_S).
5. Calculate the percentage of children (Age < 18) who survived.
6. Calculate the percentage of adults (Age >= 18) who survived.
7. Get the median age of survivors.
8. Get the median age of non-survivors.
9. Get the median fare of survivors.
10. Get the median fare of non-survivors.

The Titanic dataset contains columns as shown below,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Test Cases +

Explorer

titanicDat...

Submit

Debugger

```python
import pandas as pd
import numpy as np

# Load the Titanic dataset
data = pd.read_csv('Titanic-Dataset.csv')
data = pd.get_dummies(data, columns=['Embarked'], drop_first=True)


survivors_by_gender = data[data['Survived'] == 1]['Sex'].value_counts()
print(survivors_by_gender)

# 2. Get the number of non-survivors by gender
non_survivors_by_gender = data[data['Survived'] == 0]
['Sex'].value_counts()
print(non_survivors_by_gender)

# 3. Get the number of survivors by embarked location
survivors_by_embarked_s = data[data['Survived'] == 1]
['Embarked_S'].value_counts()
print(survivors_by_embarked_s)

# 4. Get the number of non-survivors by embarked location
non_survivors_by_embarked_s = data[data['Survived'] == 0]
['Embarked_S'].value_counts()
print(non_survivors_by_embarked_s)
```

Terminal Test cases

4.2.8. Titanic Dataset Analysis and Data Cleaning - 4

30:28

You are provided with the Titanic dataset containing information about passengers on the Titanic. Your task is to write Python code to answer the following questions based on the dataset.

1. Get the number of survivors by gender (Sex).
2. Get the number of non-survivors by gender (Sex).
3. Get the number of survivors by embarkation location (Embarked_S).
4. Get the number of non-survivors by embarkation location (Embarked_S).
5. Calculate the percentage of children (Age < 18) who survived.
6. Calculate the percentage of adults (Age >= 18) who survived.
7. Get the median age of survivors.
8. Get the median age of non-survivors.
9. Get the median fare of survivors.
10. Get the median fare of non-survivors.

The Titanic dataset contains columns as shown below,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Test Cases +

Explorer

titanicDat...

Submit

Debugger

```
# 5. Calculate the percentage of children (Age < 18) who survived
children = data[data['Age'] < 18]
children_survival_rate = children['Survived'].mean()
print(children_survival_rate)

# 6. Calculate the percentage of adults (Age >= 18) who survived
adults = data[data['Age'] >= 18]
adults_survival_rate = adults['Survived'].mean()
print(adults_survival_rate)

# 7. Get the median age of survivors
median_age_survivors = data[data['Survived'] == 1]['Age'].median()
print(median_age_survivors)

# 8. Get the median age of non-survivors
median_age_non_survivors = data[data['Survived'] == 0]['Age'].median()
print(median_age_non_survivors)

# 9. Get the median fare of survivors
median_fare_survivors = data[data['Survived'] == 1]['Fare'].median()
print(median_fare_survivors)

# 10. Get the median fare of non-survivors
median_fare_non_survivors = data[data['Survived'] == 0]['Fare'].median()
print(median_fare_non_survivors)
```

Terminal Test cases

5.1.1. Stacked Plot

06:24

Create a stacked area plot to visualize the temperature variations for three different cities (City A, City B, and City C) across the months of the year. The temperature data is provided for each city in the editor.

Your task is to:

- Create a stacked area plot using the data.
- Label the x-axis as "Month", the y-axis as "Temperature", and provide the title "Temperature Variation" for the plot.
- Display the plot showing the temperature variation for each city throughout the months of the year.

Sample Test Cases +

Explorer

stackedpl...

Submit

```python
import matplotlib.pyplot as plt
import pandas as pd

# Data for Months and Temperature for three cities
data = {
    'Month': ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December'],
    'City_A_Temperature': [5, 7, 10, 13, 17, 20, 22, 21, 18, 12, 8, 6],
    'City_B_Temperature': [2, 3, 5, 6, 10, 14, 16, 17, 12, 9, 5, 3],
    'City_C_Temperature': [3, 4, 6, 8, 9, 12, 15, 14, 10, 7, 4, 2]
}

# Write your code...
df = pd.DataFrame(data)
plt.stackplot(df['Month'],df['City_A_Temperature'],df['City_B_Temperature'],df['City_C_Temperature'])
plt.title('Temperature Variation')
plt.xlabel('Month')
plt.ylabel('Temperature')
plt.show()

```

Debugger

Terminal Test cases

< Prev Reset Submit Next >

5.2.1. Titanic Dataset

35:47

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

Dataset Information:

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:

Sample Test Cases +

Explorer

titanicDat...

Submit

Debugger

```
1  import pandas as pd
2  import matplotlib.pyplot as plt
3
4  # Load the Titanic dataset from the CSV file
5  df = pd.read_csv('titanic.csv')
6
7  # Set up the figure for 5 subplots
8  fig, axes = plt.subplots(3, 2, figsize=(12, 12))
9
10
11 axes[0, 0].bar(df['Pclass'].value_counts().index,
   df['Pclass'].value_counts(), color='skyblue')
12 axes[0, 0].set_title("Passenger Class Distribution")
13 axes[0, 0].set_xlabel("Pclass")
14 axes[0, 0].set_ylabel("Count")
15
16 axes[0, 1].pie(df['Gender'].value_counts(),
   labels=df['Gender'].value_counts().index, autopct='%1.1f%%', colors=
   ['lightblue', 'lightcoral'])
17 axes[0, 1].set_title("Gender Distribution")
18
19 axes[1, 0].hist(df['Age'].dropna(), bins=8, color='lightgreen',
   edgecolor='black')
20 axes[1, 0].set_title("Age Distribution")
21 axes[1, 0].set_xlabel("Age")
```

Terminal Test cases

< Prev Reset Submit Next >

5.2.1. Titanic Dataset

35:47

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

Dataset Information:

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:

Sample Test Cases +

Explorer

titanicDat...

Submit

Debugger

```
16  axes[0, 1].pie(df['Gender'].value_counts(),
    labels=df['Gender'].value_counts().index, autopct='%1.1f%%', colors=
    ['lightblue', 'lightcoral'])
17  axes[0, 1].set_title("Gender Distribution")
18
19  axes[1, 0].hist(df['Age'].dropna(), bins=8, color='lightgreen',
    edgecolor='black')
20  axes[1, 0].set_title("Age Distribution")
21  axes[1, 0].set_xlabel("Age")
22  axes[1, 0].set_ylabel("Frequency")
23
24  axes[1, 1].bar(df['Survived'].value_counts().index,
    df['Survived'].value_counts(), color=['lightblue', 'lightcoral'])
25  axes[1, 1].set_title("Survival Count")
26  axes[1, 1].set_xlabel("Survived (0 = No, 1 = Yes)")
27  axes[1, 1].set_ylabel("Count")
28
29  axes[2, 0].scatter(df['Age'], df['Fare'], color='orange',
    edgecolors='black')
30  axes[2, 0].set_title("Fare vs Age")
31  axes[2, 0].set_xlabel("Age")
32  axes[2, 0].set_ylabel("Fare")
33
34  plt.tight_layout()
35  plt.show()
```

Terminal Test cases

5.2.2. Histogram of passenger information of Titanic

03:48

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.
2. 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,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Data:

```
PassengerId,Survived,Pclass,Name,Sex,Age,SibSp,Parch,Ticket,Fare,Cabin,Embarked
1,0,3,"Braund, Mr. Owen Harris",male,22,1,0,A/5 21171,7.25,,S
2,1,1,"Cumings, Mrs. John Bradley (Florence Briggs Thayer)",female,38,1,0,PC 17599,71.2833,C85,C
```

Sample Test Cases +

Explorer

Histogra...

Submit

Debugger

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the Titanic dataset
data = pd.read_csv('Titanic-Dataset.csv')

# Data Cleaning
data['Age'].fillna(data['Age'].median(), inplace=True)
data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)
data.drop('Cabin', axis=1, inplace=True)

# Convert categorical features to numeric
data['Sex'] = data['Sex'].map({'male': 0, 'female': 1})
data = pd.get_dummies(data, columns=['Embarked'], drop_first=True)

# Write your code here for Histogram
plt.hist(data['Age'], bins=30, edgecolor='k')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.title('Age Distribution')
plt.show()

```

Terminal Test cases

5.2.3. Bar plot of survival rate of passengers

02:48

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**'.
3. Add the title "**Survival Count**" to the chart.
4. Label the x-axis as '**Survived**' and the y-axis as '**Count**'.

The Titanic dataset contains columns as shown below,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Data:

```
PassengerId,Survived,Pclass,Name,Sex,Age,SibSp,Parch,Ticket,Fare,Cabin,Embarked
1,0,3,"Braund, Mr. Owen Harris",male,22,1,0,A/5 21171,7.25,,S
2,1,1,"Cumings, Mrs. John Bradley (Florence Briggs Thayer)",female,38,1,0,PC 17599,71.2833,C85,C
```

Sample Test Cases +

Explorer

BarPlotOf...

Submit

Debugger

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the Titanic dataset
data = pd.read_csv('Titanic-Dataset.csv')

# Data Cleaning
data['Age'].fillna(data['Age'].median(), inplace=True)
data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)
data.drop('Cabin', axis=1, inplace=True)

# Convert categorical features to numeric
data['Sex'] = data['Sex'].map({'male': 0, 'female': 1})
data = pd.get_dummies(data, columns=['Embarked'], drop_first=True)

# Write your code here for Bar Plot for Survival Rate
survival_counts = data['Survived'].value_counts()
survival_counts.plot(kind='bar')
plt.title('Survival Count')
plt.xlabel('Survived')
plt.ylabel('Count')
plt.show()

```

Terminal Test cases

5.2.4. Bar Plot for Survival by Gender

12:14

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.
2. Use a **stacked bar chart** to display the survival counts.
3. Add the title **"Survival by Gender"** to the chart.
4. Label the x-axis as **'Gender'** and the y-axis as **'Count'**.
5. The legend should indicate **'Not Survived'** and **'Survived'**.

The Titanic dataset contains columns as shown below,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Data:

Sample Test Cases +

Explorer

BarPlotOf...

Submit

```
import matplotlib.pyplot as plt

# Load the Titanic dataset
data = pd.read_csv('Titanic-Dataset.csv')

# Data Cleaning
data['Age'].fillna(data['Age'].median(), inplace=True)
data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)
data.drop('Cabin', axis=1, inplace=True)

# Convert categorical features to numeric
data['Sex'] = data['Sex'].map({'male': 0, 'female': 1})
data = pd.get_dummies(data, columns=['Embarked'], drop_first=True)

# Write your code here for Bar Plot for Survival by Gender
survival_by_gender = data.groupby('Sex')
['Survived'].value_counts().unstack().fillna(0)
survival_by_gender.columns = ['Not Survived', 'Survived']
survival_by_gender.index = ['0', '1']
survival_by_gender.plot(kind='bar', stacked=True)
plt.title('Survival by Gender')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.legend(title=None)
plt.show()
```

Terminal Test cases

Debugger

5.2.5. Bar Plot for Survival by Pclass

07:59

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()**.
2. Use a **stacked bar chart** to display the survival counts.
3. Add the title **"Survival by Pclass"** to the chart.
4. 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,

Passengerid	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Data:

Sample Test Cases +

Explorer

BarPlotOf...

Submit

```
1  import pandas as pd
2  import matplotlib.pyplot as plt
3
4  # Load the Titanic dataset
5  data = pd.read_csv('Titanic-Dataset.csv')
6
7  # Data Cleaning
8  data['Age'].fillna(data['Age'].median(), inplace=True)
9  data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)
10 data.drop('Cabin', axis=1, 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 Bar Plot for Survival by Pclass
17 survival_by_class = data.groupby('Pclass')
   ['Survived'].value_counts().unstack().fillna(0)
18 survival_by_class.columns = ['Not Survived', 'Survived']
19 survival_by_class.plot(kind='bar', stacked=True)
20 plt.title('Survival by Pclass')
21 plt.xlabel('Pclass')
22 plt.ylabel('Count')
23 plt.legend(title=None)
24 plt.show()
```

Terminal Test cases

Debugger

< Prev Reset Submit Next >

5.2.6. Bar Plot for Survival by Embarked

05:17

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:

1. 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.
2. Set the chart type to 'bar' and make it stacked.
3. 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,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Test Cases +

Explorer

BarPlotOf...

Submit

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the Titanic dataset
data = pd.read_csv('Titanic-Dataset.csv')

# Data Cleaning
data['Age'].fillna(data['Age'].median(), inplace=True)
data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)
data.drop('Cabin', axis=1, inplace=True)

# Convert categorical features to numeric
data['Sex'] = data['Sex'].map({'male': 0, 'female': 1})
data = pd.get_dummies(data, columns=['Embarked'], drop_first=True)

# Write your code here for Bar Plot for Survival by Embarked
grouped = data.groupby('Embarked_Q')
['Survived'].value_counts().unstack().fillna(0)
grouped.columns = ['Not Survived', 'Survived']
grouped.plot(kind='bar', stacked=True)
plt.title('Survival by Embarked')
plt.xlabel('Embarked')
plt.ylabel('Count')
plt.legend(title=None)
plt.show()
```

Terminal Test cases

Debugger

5.2.7. Box plot for Age Distribution

01:55

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('')**.
4. Label the x-axis as **'Pclass'** and the y-axis as **'Age'**.

The Titanic dataset contains columns as shown below,

Passengerid	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Data:

```
PassengerId,Survived,Pclass,Name,Sex,Age,SibSp,Parch,Ticket,Fare,Cabin,Embarked
1,0,3,"Braund, Mr. Owen Harris",male,22,1,0,A/5 21171,7.25,,S
2,1,1,"Cumings, Mrs. John Bradley (Florence Briggs Thayer)",female,38,1,0,PC 17599,71.2833,C85,C
```

Sample Test Cases +

Explorer

BoxPlotF...

Submit

Debugger

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the Titanic dataset
data = pd.read_csv('Titanic-Dataset.csv')

# Data Cleaning
data['Age'].fillna(data['Age'].median(), inplace=True)
data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)
data.drop('Cabin', axis=1, inplace=True)

# Convert categorical features to numeric
data['Sex'] = data['Sex'].map({'male': 0, 'female': 1})
data = pd.get_dummies(data, columns=['Embarked'], drop_first=True)

# Write your code here for Box Plot for Age by Pclass
plt.figure(figsize=(8, 6))
data.boxplot(column='Age', by='Pclass')
plt.suptitle('')
plt.title('Age by Pclass')
plt.xlabel('Pclass')
plt.ylabel('Age')
plt.show()

```

Terminal Test cases

5.2.8. Box Plot for Age by Survived

03:20

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('')**.
4. Label the x-axis as **'Survived'** and the y-axis as **'Age'**.

The Titanic dataset contains columns as shown below,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Data:

```
PassengerId,Survived,Pclass,Name,Sex,Age,SibSp,Parch,Ticket,Fare,Cabin,Embarked
1,0,3,"Braund, Mr. Owen Harris",male,22,1,0,A/5 21171,7.25,,S
```

Sample Test Cases +

Explorer

BoxPlotF...

Submit

Debugger

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the Titanic dataset
data = pd.read_csv('Titanic-Dataset.csv')

# Data Cleaning
data['Age'].fillna(data['Age'].median(), inplace=True)
data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)
data.drop('Cabin', axis=1, inplace=True)

# Convert categorical features to numeric
data['Sex'] = data['Sex'].map({'male': 0, 'female': 1})
data = pd.get_dummies(data, columns=['Embarked'], drop_first=True)

# Write your code here for Box Plot for Age by Survived
plt.figure(figsize=(8, 6))
data.boxplot(column='Age', by='Survived')
plt.suptitle('')
plt.title('Age by Survival')
plt.xlabel('Survived')
plt.ylabel('Age')
plt.show()
```

Terminal Test cases

5.2.9. Box Plot for Fare by Pclass

02:41

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:

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

The Titanic dataset contains columns as shown below,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Data:

```
PassengerId,Survived,Pclass,Name,Sex,Age,SibSp,Parch,Ticket,Fare,Cabin,Embarked
1,0,3,"Braund, Mr. Owen Harris",male,22,1,0,A/5 21171,7.25,,S
2,1,1,"Cumings, Mrs. John Bradley (Florence Briggs Thayer)",female,38,1,0,PC 17599,71.2833,C85,C
```

Sample Test Cases +

Explorer

BoxPlotF...

Submit

Debugger

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the Titanic dataset
data = pd.read_csv('Titanic-Dataset.csv')

# Data Cleaning
data['Age'].fillna(data['Age'].median(), inplace=True)
data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)
data.drop('Cabin', axis=1, inplace=True)

# Convert categorical features to numeric
data['Sex'] = data['Sex'].map({'male': 0, 'female': 1})
data = pd.get_dummies(data, columns=['Embarked'], drop_first=True)

# Write your code here for Box Plot for Fare by Pclass
plt.figure(figsize=(8, 6))
data.boxplot(column='Fare', by='Pclass')
plt.suptitle('')
plt.title('Fare by Pclass')
plt.xlabel('Pclass')
plt.ylabel('Fare')
plt.show()
```

Terminal Test cases

< Prev Reset Submit Next >

5.2.10. Scatter Plot for Age vs. Fare

02:10

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:

1. Use the **Age** column for the x-axis and the **Fare** column for the y-axis.
2. 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,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Data:

```
PassengerId,Survived,Pclass,Name,Sex,Age,SibSp,Parch,Ticket,Fare,Cabin,Embarked
1,0,3,"Braund, Mr. Owen Harris",male,22,1,0,A/5 21171,7.25,,S
2,1,1,"Cumings, Mrs. John Bradley (Florence Briggs Thayer)",female,38,1,0,PC 17599,71.2833,C85,C
3,1,3,"Heikkinen, Miss. Laina",female,26,0,0,STON/O2. 3101282,7.925,,S
```

Sample Test Cases +

Explorer

AgeFareS...

Submit

Debugger

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the Titanic dataset
data = pd.read_csv('Titanic-Dataset.csv')

# Data Cleaning
data['Age'].fillna(data['Age'].median(), inplace=True)
data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)
data.drop('Cabin', axis=1, inplace=True)

# Convert categorical features to numeric
data['Sex'] = data['Sex'].map({'male': 0, 'female': 1})
data = pd.get_dummies(data, columns=['Embarked'], drop_first=True)

# Write your code here for Box Plot for Fare by Pclass
plt.figure(figsize=(6.4,4.8))
plt.scatter(data['Age'],data['Fare'])
plt.title('Age vs. Fare')
plt.xlabel('Age')
plt.ylabel('Fare')
plt.show()

```

Terminal Test cases

< Prev Reset Submit Next >

5.2.11. Scatter Plot for Age vs. Fare by Survived

04:01

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:

1. Use the **Age** column for the x-axis and the **Fare** column for the y-axis.
2. 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,

Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

Sample Data:

Sample Test Cases +

Explorer

AgeFareS...

Submit

Debugger

```
1  import pandas as pd
2  import matplotlib.pyplot as plt
3
4  # Load the Titanic dataset
5  data = pd.read_csv('Titanic-Dataset.csv')
6
7  # Data Cleaning
8  data['Age'].fillna(data['Age'].median(), inplace=True)
9  data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)
10 data.drop('Cabin', axis=1, 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'})
18 plt.scatter(data['Age'],data['Fare'],c=colors)
19 plt.title('Age vs. Fare by Survival')
20 plt.xlabel('Age')
21 plt.ylabel('Fare')
22 plt.show()
23
24
```

Terminal Test cases

< Prev Reset Submit