

4.1.1. Pandas - series creation and manipulation

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

```
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
8 series= pd.Series(numbers)
9 # Grouping by even and odd numbers and calculating the mean
10 grouped = series.groupby(series%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
14 grouped.index]
15 print("Mean of even and odd numbers:")
16 print(grouped)
```

4.1.2. Dictionary to dataframe2804

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:

Sample Test Cases

datafram...

```
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)
```

Terminal

Test cases

4.1.3. Student Information

05:38

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

studentin... studentdat...

Submit

1 import pandas as pd

2

3 # Read the text file into a DataFrame

4 file = input()

5 data = pd.read_csv(file, sep="\s+", header=None, names=["Name",

6 "Age", "Grade"])

7 print("First five rows:")

8 print(data.head(5))

9

10 # write your code here..

11 age=round(data["Age"].mean(),2)

12 print("Average age:",age)

13 print("Students with a grade up to B")

14 df=pd.DataFrame(data)

15 a=df[df["Grade"]<="B"]

16 print(a)

Terminal Test cases

4.2.1. Month with the Highest Total Sales

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

Sample Test Cases

monthFor... sales_dat... Submit

```
1 import pandas as pd
2
3 # Prompt the user for the file name
4 file_name = input()
5
6 # Load the data
7 df = pd.read_csv(file_name)
8 df["sales"] = df["Quantity"].multiply(df["Price"])
9 df["month"] = pd.to_datetime(df["Date"]).dt.strftime("%Y-%m")
10
11 # Find the month with the highest total sales
12 best_month = df.groupby("month")["sales"].sum().idxmax()
13 highest_sales = df["sales"].sum()
14
15 print(f"Best month: {best_month}")
16 print(f"Total sales: ${highest_sales:.2f}")
17
```

Terminal Test cases

Prev Reset Submit Next

CODETANTRA

Home

202401070216@mitaoe.ac.inSupportLogout

4.2.2. Best Selling Product22.01

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

Sample Test Cases

monthFor...sales_dat...

```
1 import pandas as pd
2
3 # Prompt the user for the file name
4 file_name = input()
5
6 # Load the data
7 df = pd.read_csv(file_name)
8
9
10 # Find the product with the highest total quantity sold
11 product_sales = df.groupby("Product")["Quantity"].sum()
12 best_product=product_sales.idxmax()
13 highest_quantity=product_sales.max()
14
15 # Display the result
16 print(f"Best selling product: {best_product}")
17 print(f"Total quantity sold: {highest_quantity}")
18
```

TerminalTest cases

< Prev

Reset

Submit

Next >

4.2.3. City that Sold the Most Products

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

Sample Test Cases

```
1 import pandas as pd
2
3 # Prompt the user for the file name
4 file_name = input()
5
6 # Load the data
7 df = pd.read_csv(file_name)
8
9 # write the code..
10 city_sales=df.groupby("City")["Quantity"].sum()
11 best_city = city_sales.idxmax()
12
13 # Display the result
14 print(f"City sold the most products: {best_city}")
15
```


4.2.4. Most Frequently Sold Product Pairs10:02

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	4	30	Los Angeles

Sample Test Cases+

frequentl...sales_dat...

Submit

1import pandas as pd2from itertools import combinations3from collections import Counter45# Prompt user to input the file name6file_name = input()78# Read data from the specified CSV file9df = pd.read_csv(file_name)1011# write the code12date_products = {}13for date, group in df.groupby("Date"):14 products = group["Product"].unique()15 if len(products) > 1:16 date_products[date] = products17 pair_count = Counter()18 for products in date_products.values():19 pairs = combinations(sorted(products), 2)20 pair_count.update(pairs)2122if pair_count:

TerminalTest cases

PrevResetSubmitNext

4.2.5. Titanic Dataset Analysis and Data Cleaning

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,

Sample Test Cases

+

Explorer titanicDat... Submit

7 # 1. Display the first 5 rows of the dataset
8 print(data.head())
9
10 # 2. Display the last 5 rows of the dataset
11 print(data.tail())
12
13 # 3. Get the shape of the dataset
14 print(data.shape)
15
16 # 4. Get a summary of the dataset (info)
17 print(data.info())
18
19 # 5. Get basic statistics of the dataset
20 print(data.describe())
21
22 # 6. Check for missing values
23 print(data.isnull().sum())
24
25 # 7. Fill missing values in the 'Age' column with the median age
26 median_age=data["Age"].median()
27 data["Age"].fillna(median_age, inplace=True)
28

Debugger

Terminal Test cases

< Prev Reset Submit Next >

4.2.5. Titanic Dataset Analysis and Data Cleaning05:53

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,

Sample Test Cases

titanicDat...

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

Terminal Test cases

< Prev

Reset

Submit

Next >

4.2.6. Titanic Dataset Analysis and Data Cleaning - 2 08:44

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,

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

Sample Test Cases

titanicDat...

1import pandas as pd
2import numpy as np
3
4# Load the Titanic dataset
5data = pd.read_csv('Titanic-Dataset.csv')
6data['FamilySize'] = data['SibSp'] + data['Parch']
7
8# 1. Create a new column 'IsAlone' (1 if alone, 0 otherwise)
9data["IsAlone"] = np.where(data["FamilySize"] == 0, 1, 0)
10
11# 2. Convert 'Sex' to numeric (male: 0, female: 1)
12data["Sex"] = data["Sex"].map({"male": 0, "female": 1})
13
14# 3. One-hot encode the 'Embarked' column
15data = pd.get_dummies(data, columns=["Embarked"], drop_first=True)
16
17# 4. Get the mean age of passengers
18mean_age = data["Age"].mean()
19print(mean_age)
20
21
22# 5. Get the median fare of passengers

TerminalTest cases

< Prev Reset Submit Next >

08:44 A 🌙 ✎ 🔗 —

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.

[illegible]

+

Submit

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

Terminal Test cases

4.2.6. Titanic Dataset Analysis and Data Cleaning - 2

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,

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

Sample Test Cases

+

titanicDat... Submit

Explorer

31

7. Get the number of passengers by gender

32 passengers_by_gender = data["Sex"].value_counts().sort_index()

33 print(passengers_by_gender)

34

35

36

37

8. Get the number of passengers by survival status

38 passengers_by_survival =

39 data["Survived"].value_counts().sort_index()

40 print(passengers_by_survival)

41

42

9. Calculate the survival rate

43 survival_rate=data["Survived"].mean()

44 print(survival_rate)

45

46

10. Calculate the survival rate by gender

47 survival_rate_by_gender=data.groupby("Sex")["Survived"].mean()

48 print(survival_rate_by_gender)

49

50

51

Debugger

Terminal Test cases

< Prev Reset Submit Next >

05:13

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).

Sample Test Cases +

Terminal Test cases

4.2.7. Titanic Dataset Analysis and Data Cleaning - 3 05:13

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,

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

Sample Test Cases +

titanicDat... Submit

19 # 4. Calculate the survival rate by being alone
20 print(data.groupby("IsAlone")["Survived"].mean())
21
22 # 5. Get the average fare by class
23 print(data.groupby("Pclass")["Fare"].mean())
24
25
26 # 6. Get the average age by class
27 print(data.groupby("Pclass")["Age"].mean())
28
29 # 7. Get the average age by survival status
30 print(data.groupby("Survived")["Age"].mean())
31
32 # 8. Get the average fare by survival status
33 print(data.groupby("Survived")["Fare"].mean())
34
35 # 9. Get the number of survivors by class
36 print(data[data["Survived"] == 1]["Pclass"].value_counts())
37
38 # 10. Get the number of non-survivors by class
39 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 12:58

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,

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

Sample Test Cases +

titanicDat... Submit

Explorer

```
1 import pandas as pd
2 import numpy as np
3
4 # Load the Titanic dataset
5 data = pd.read_csv('Titanic-Dataset.csv')
6 data = pd.get_dummies(data, columns=['Embarked'], drop_first=True)
7
8
9 # 1. Get the number of survivors by gender
10 survivors_by_gender = data[data["Survived"] == 1]
11 ["Sex"].value_counts()
12 print(survivors_by_gender)
13
14 # 2. Get the number of non-survivors by gender
15 non_survivors_by_gender = data[data["Survived"] == 0]
16 ["Sex"].value_counts()
17 print(non_survivors_by_gender)
18
19 # 3. Get the number of survivors by embarked location
20 survivors_by_embarked_s = data[data["Survived"] == 1]
21 ["Embarked_S"].value_counts()
22 print(survivors_by_embarked_s)
```

Terminal Test cases

Debugger

4.2.8. Titanic Dataset Analysis and Data Cleaning - 4

12:55

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,

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

Sample Test Cases

+

titanicDat...

Submit

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

Terminal

Test cases

Debugger

```
# 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)

# 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)
```

12:58

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.

Sample Test Cases +

```
titanicDat... Submit
31 adults_survival_rate = adults[ 'Survived' ].median()
32 print(adults_survival_rate)
33
34 # 7. Get the median age of survivors
35 median_age_survivors = data[data["Survived"] == 1]["Age"].median()
36 print(median_age_survivors)
37
38 # 8. Get the median age of non-survivors
39 median_age_non_survivors = data[data["Survived"] == 0]
40 ["Age"].median()
41 print(median_age_non_survivors)
42
43 # 9. Get the median fare of survivors
44 median_fare_survivors = data[data["Survived"] == 1]
45 ["Fare"].median()
46 print(median_fare_survivors)
47
48 # 10. Get the median fare of non-survivors
49 median_fare_non_survivors = data[data["Survived"] == 0]
50 ["Fare"].median()
51 print(median_fare_non_survivors)
```