Inclass - Lab (Day 2)

Table of Content

Let's begin with some hands-on practice exercises

1. Data Structures with Examples:

1. Write a program to create a dataframe from list of lists

Use the below nested list:

```
data = [['Joy', 21], ['Mia', 15], ['Sam', 14]]
```

In [2]:
```
# type your code here
import pandas as pd
data = [['Joy', 21], ['Mia', 15], ['Sam', 14]]
p1 = pd.Series(data)
print(p1)
```

```
0    [Joy, 21]
1    [Mia, 15]
2    [Sam, 14]
dtype: object
```

2. Extract the second word from the given string

Use the below string:

```
sentence = 'Be confident and be yourself'
```

In [5]:
```
# type your code here
sentence = 'Be confident and be yourself'
res = sentence.split()
print(res[1])
```

```
confident
```

2. Create and Manipulate the DataFrames:

3. Split the 'Name' column into two separate columns as First_Name and Last_Name

Use the dataframe given below:

Name	Salary
Emma Larter	3200
Mia Junior	4500
Sophia Depp	3600
James Smith	5596

In [8]:
```
# type your code here
df = pd.DataFrame({'Name': ['Emma Larter','Mia Junior','Sophia Depp','James Smith
                   'Salary':[3200,4500,3600,5596]})
df[['First_Name','Last_Name']] = df.Name.str.split(expand=True)
print(df)
```

```
          Name  Salary First_Name Last_Name
0  Emma Larter    3200       Emma    Larter
1   Mia Junior    4500        Mia    Junior
2  Sophia Depp    3600     Sophia      Depp
3  James Smith    5596      James     Smith
```

4. Select the rows from below dataframe where marks are less than 60

Use the dataframe given below:

Name	Marks
John	90
Robert	50
Jonny	89
Mia	55

In [10]:
```
# type your code here
df4 = pd.DataFrame({'Name':['John','Robert','Johnny','Mia'],'Marks':[90,50,89,55]
df4[df4.Marks<60]
```

Out[10]:

	Name	Marks
1	Robert	50
3	Mia	55

5. Write a program to select the rows where the sales are higher than 60000 for Mumbai

Use the dataframe given below:

Store	Location	Sales
A	Mumbai	40000
B	Pune	45000
A	Hyderabad	50000
C	Mumbai	90000
D	Pune	89000
A	Delhi	87000
D	Hyderabad	85000
A	Pune	78000
C	Mumbai	89000
B	Pune	70000

In [4]:
```
# type your code here
df5 = pd.DataFrame({'Store':['A','B','A','C','D','A','D','A','C','B'],'Location':

                    'Sales':[40000,45000,50000,90000,89000,87000,85000,78000,89000
df5[(df5.Location=='Mumbai') & (df5.Sales>60000)]
```

Out[4]:

	Store	Location	Sales
3	C	Mumbai	90000
8	C	Mumbai	89000

6. Write a program to select stores located in Hyderabad

Use the dataframe given below:

Store	Location	Sales
A	Mumbai	40000

B	Pune	45000
A	Hyderabad	50000
C	Mumbai	90000
D	Pune	89000
A	Delhi	87000
D	Hyderabad	85000
A	Pune	78000
C	Mumbai	89000
B	Pune	70000

In [14]:
```
# type your code here
df5[df5.Location=='Hyderabad']
```

Out[14]:

	Store	Location	Sales
2	A	Hyderabad	50000
6	D	Hyderabad	85000

7. Write a program to replace the 'Store' column with the values A, B, C, and D with Store_A, Store_B, Store_C and Store_D respectively

Use the dataframe given below:

Store	Location	Sales
A	Mumbai	40000
B	Pune	45000
A	Hyderabad	50000
C	Mumbai	90000
D	Pune	89000
A	Delhi	87000
D	Hyderabad	85000
A	Pune	78000
C	Mumbai	89000
B	Pune	70000

In [20]:
```
# type your code here
df5['Store'] = df5['Store'].replace(['A','B','C','D'],['Store_A','Store_B','Store
df5
```

Out[20]:

	Store	Location	Sales
0	Store_A	Mumbai	40000
1	Store_B	Pune	45000
2	Store_A	Hyderabad	50000
3	Store_C	Mumbai	90000
4	Store_D	Pune	89000
5	Store_A	Delhi	87000
6	Store_D	Hyderabad	85000
7	Store_A	Pune	78000
8	Store_C	Mumbai	89000
9	Store_B	Pune	70000

8. Create a dataframe in pandas using the given list.

Use the below list

```
city = ['Delhi', 'Mumbai', 'Chennai', 'Kolkata']
```

In [21]:
```
# type your code here
city = ['Delhi', 'Mumbai', 'Chennai', 'Kolkata']
df8 = pd.Series(city)
df8
```

Out[21]:
```
0      Delhi
1     Mumbai
2    Chennai
3    Kolkata
dtype: object
```

9. Save a given dataframe to a csv file

Use the dataframe given below:

Store	Location	Sales
A	Mumbai	40000
B	Pune	45000
A	Hyderabad	50000
C	Mumbai	90000

D	Pune	89000
A	Delhi	87000
D	Hyderabad	85000
A	Pune	78000
C	Mumbai	89000
B	Pune	70000

In [24]:
```
# type your code here
df5['Store'] = df5['Store'].replace(['Store_A','Store_B','Store_C','Store_D'],['A
df5.to_csv('df5.csv')
```

10. Find the maximum sales for each store

Use the dataframe given below:

Store	Location	Sales
A	Mumbai	40000
B	Pune	45000
A	Hyderabad	50000
C	Mumbai	90000
D	Pune	89000
A	Delhi	87000
D	Hyderabad	85000
A	Pune	78000
C	Mumbai	89000
B	Pune	70000

In [53]:
```
# type your code here
df5
psales = df5.groupby('Store')
dfsales = psales['Sales'].max()
dfsales
```

Out[53]:
```
Store
A    87000
B    70000
C    90000
D    89000
Name: Sales, dtype: int64
```

11. Write a program to select all columns, except the 'Location' from the given dataframe

Use the dataframe given below:

Store	Location	Sales
A	Mumbai	40000
B	Pune	45000
A	Hyderabad	50000
C	Mumbai	90000
D	Pune	89000
A	Delhi	87000
D	Hyderabad	85000
A	Pune	78000
C	Mumbai	89000
B	Pune	70000

In [28]:
```
# type your code here
df5[['Store','Sales']]
```

Out[28]:

	Store	Sales
0	A	40000
1	B	45000
2	A	50000
3	C	90000
4	D	89000
5	A	87000
6	D	85000
7	A	78000
8	C	89000
9	B	70000

12. Write a program to insert a given column at the second position in a dataframe

Use the dataframe given below:

Name	Age
Emma Larter	34
Mia Junior	59
Sophia Depp	32
James Smith	40

Add the variable 'Salary' as second column

```
Salary = [35000, 24000, 55000, 40000]
```

In [32]:
```
# type your code here
df12 = pd.DataFrame({'Name':['Emma Robert','Mia Junior','Sophia Depp','James Smit
df12.insert(1,"Salary",[35000, 24000, 55000, 40000])
df12
```

Out[32]:

	Name	Salary	Age
0	Emma Robert	35000	34
1	Mia Junior	24000	59
2	Sophia Depp	55000	32
3	James Smith	40000	40

13. Write a program to get the third observation of the given dataframe

Use the dataframe given below:

Name	Age
Emma Larter	34
Mia Junior	59
Sophia Depp	32
James Smith	40

In [36]:
```
# type your code here
df13 = pd.DataFrame({'Name':['Emma Larter','Mia Junior','Sophia Depp','James Smit
df13.iloc[2]
```

Out[36]:
```
Name    Sophia Depp
Age              32
Name: 2, dtype: object
```

14. Write a program to create a dataframe from the provided lists

Use the following lists to create your arrays:

```
columns = ["Science", "Maths", "English"]

marks = [[55, 78, 88], [93, 63, 83],[94, 74, 64]]
```

Hint: Convert the list marks into an array

In [79]:
```
# type your code here
#f2 = pd.DataFrame(np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]]),
...             #       columns=['a', 'b', 'c'])
import numpy as np
columns = ["Science", "Maths", "English"]
marks = [[55, 78, 88], [93, 63, 83],[94, 74, 64]]

df14 = pd.DataFrame(np.array([[55, 78, 88], [93, 63, 83],[94, 74, 64]]),columns=[
df14
```

Out[79]:

	Science	Maths	English
0	55	78	88
1	93	63	83
2	94	74	64

15. Write a program to count total sales per store

Use the dataframe given below:

Store	Location	Sales
A	Mumbai	40000
B	Pune	45000
A	Hyderabad	50000
C	Mumbai	90000
D	Pune	89000
A	Delhi	87000
D	Hyderabad	85000
A	Pune	78000
C	Mumbai	89000
B	Pune	70000

In [5]:
```
# type your code here
group = df5.groupby(['Store','Location']).sum()
group
```

Out[5]:

		Sales
Store	**Location**	
A	**Delhi**	87000
	Hyderabad	50000
	Mumbai	40000
	Pune	78000
B	**Pune**	115000
C	**Mumbai**	179000
D	**Hyderabad**	85000
	Pune	89000

16. Write a program to count the top five commonly used words in a list

Use the below list:

```
words_list = ['words', 'will', 'where', 'shall', 'we', 'will',
 'shall', 'we', 'words', 'where', 'shall',
                    'will', 'there', 'wow', 'should', 'shall', 'we', 'wh
ere', 'should', 'where', 'will',
                    'there', 'now', 'where', 'we', 'will', 'where', 'sho
uld', 'will', 'where']
```

In [48]:
```
# type your code here
words_list = ['words', 'will', 'where', 'shall', 'we', 'will', 'shall', 'we', 'wo
                    'will', 'there', 'wow', 'should', 'shall', 'we', 'where', 'sh
                    'there', 'now', 'where', 'we', 'will', 'where', 'should', 'wi
df16 = pd.unique(pd.Series(words_list))
df16[:5]
```

Out[48]:
```
array(['words', 'will', 'where', 'shall', 'we'], dtype=object)
```

3. Sort

17. Write a program to sort the dataframe by 'Sales' in ascending order

Use the dataframe given below:

Store	Location	Sales
A	Mumbai	40000
B	Pune	45000
A	Hyderabad	50000
C	Mumbai	90000
D	Pune	89000
A	Delhi	87000
D	Hyderabad	85000
A	Pune	78000
C	Mumbai	89000
B	Pune	70000

In [50]:
```
# type your code here
df5.sort_values(by='Sales')
```

Out[50]:

	Store	Location	Sales
0	A	Mumbai	40000
1	B	Pune	45000
2	A	Hyderabad	50000
9	B	Pune	70000
7	A	Pune	78000
6	D	Hyderabad	85000
5	A	Delhi	87000
4	D	Pune	89000
8	C	Mumbai	89000
3	C	Mumbai	90000

18. Write a program to sort the dataframe by 'Store' in ascending order and by 'Sales' in descending order simultaneously

Use the dataframe given below:

Store	Location	Sales
A	Mumbai	40000
B	Pune	45000
A	Hyderabad	50000
C	Mumbai	90000
D	Pune	89000

A	Delhi	87000
D	Hyderabad	85000
A	Pune	78000
C	Mumbai	89000
B	Pune	70000

In [64]:
```
# type your code here
df5.sort_values(by = ['Store','Sales'],ascending=[True,False])
```

Out[64]:

	Store	Location	Sales	Rank_sales
5	A	Delhi	87000	8.5
7	A	Pune	78000	8.5
2	A	Hyderabad	50000	8.5
0	A	Mumbai	40000	8.5
9	B	Pune	70000	5.5
1	B	Pune	45000	5.5
3	C	Mumbai	90000	3.5
8	C	Mumbai	89000	3.5
4	D	Pune	89000	1.5
6	D	Hyderabad	85000	1.5

4. Rank

19. Create a column 'Rank_sales' which contains the rank of each store based on sales of the product in ascending order

Use the dataframe given below:

Store	Location	Sales
A	Mumbai	40000
B	Pune	45000
A	Hyderabad	50000
C	Mumbai	90000
D	Pune	89000
A	Delhi	87000
D	Hyderabad	85000
A	Pune	78000
C	Mumbai	89000

B	Pune	70000

In [65]:
```
# type your code here
df5['Rank_sales'] = df5.Store.rank(axis=0,method='average',ascending=False)
df5
```

Out[65]:

	Store	Location	Sales	Rank_sales
0	A	Mumbai	40000	8.5
1	B	Pune	45000	5.5
2	A	Hyderabad	50000	8.5
3	C	Mumbai	90000	3.5
4	D	Pune	89000	1.5
5	A	Delhi	87000	8.5
6	D	Hyderabad	85000	1.5
7	A	Pune	78000	8.5
8	C	Mumbai	89000	3.5
9	B	Pune	70000	5.5

20. Rank the dataframe in descending order of sales by each store

Use the dataframe given below:

Store	Location	Sales
A	Mumbai	40000
B	Pune	45000
A	Hyderabad	50000
C	Mumbai	90000
D	Pune	89000
A	Delhi	87000
D	Hyderabad	85000
A	Pune	78000
C	Mumbai	89000
B	Pune	70000

In [73]:
```
# type your code here
df20 = df5.drop('Rank_sales',axis=1)
df20
df20['Group_rank'] = df20.groupby('Store')['Sales'].rank(ascending=False)
df20
```

Out[73]:

	Store	Location	Sales	Group_rank
0	A	Mumbai	40000	4.0
1	B	Pune	45000	2.0
2	A	Hyderabad	50000	3.0
3	C	Mumbai	90000	1.0
4	D	Pune	89000	1.0
5	A	Delhi	87000	1.0
6	D	Hyderabad	85000	2.0
7	A	Pune	78000	2.0
8	C	Mumbai	89000	2.0
9	B	Pune	70000	1.0