Pandas-Live-Session

Use the "Run" button to execute the code.

```
!pip install jovian --upgrade --quiet
```

```python
import jovian
```

```python
# Execute this to save new versions of the notebook
jovian.commit(project="pandas-live-session")
```

Pandas

- Series
- DataFrame

Series

```python
import pandas as pd

sr = pd.Series(["Ram","Shyam","Jadu","Hari"])
sr
```

```
0      Ram
1    Shyam
2     Jadu
3     Hari
dtype: object
```

```python
sr = sr = pd.Series(["Ram","Shyam","Jadu","Hari"], index = ("A","B","C","D"))
sr
```

```
A      Ram
B    Shyam
C     Jadu
D     Hari
dtype: object
```

Data Frame

```python
employeeData = {

    "EmpId" : [10, 11 , 12 , 13],
    "Employee": ["Ram","Shyam","Jadu","Hari"],
    "City": ["Harayna","Banglore","Jaipur", "Bhubaneswar"]
}
```

```python
df = pd.DataFrame(employeeData)
df
```

	EmpId	Employee	City
0	10	Ram	Harayna
1	11	Shyam	Banglore
2	12	Jadu	Jaipur
3	13	Hari	Bhubaneswar

```python
dct = {"Employeename": "Kajal", "Age": 20, "Salary": 2533, "City": "Banglore"}
pd.DataFrame(dct, index = ["A"])
```

	Employeename	Age	Salary	City
A	Kajal	20	2533	Banglore

```python
dct = {"Employeename": ["Kajal"], "Age": [20], "Salary": [2533], "City": ["Banglore"]}

pd.DataFrame(dct)
```

	Employeename	Age	Salary	City
0	Kajal	20	2533	Banglore

```python
dct = {"Employeename": ("Kajal"), "Age": (20), "Salary": (2533), "City": ("Banglore")}

pd.DataFrame(dct, index = [1])
```

	Employeename	Age	Salary	City
1	Kajal	20	2533	Banglore

```python
pip install openpyxl
```

```
Collecting openpyxl
  Downloading openpyxl-3.0.10-py2.py3-none-any.whl (242 kB)
     |████████████████████████████| 242 kB 6.9 MB/s eta 0:00:01
Collecting et-xmlfile
  Downloading et_xmlfile-1.1.0-py3-none-any.whl (4.7 kB)
Installing collected packages: et-xmlfile, openpyxl
Successfully installed et-xmlfile-1.1.0 openpyxl-3.0.10
Note: you may need to restart the kernel to use updated packages.
```

Data Frame Operations

```python
path = r"sales.xlsx"
df = pd.read_excel(path)
df
```

	Market	Region	No_of_Orders	Profit	Sales
0	Africa	Western Africa	251	-12901.51	78476.06
1	Africa	Southern Africa	85	11768.58	51319.50
2	Africa	North Africa	182	21643.08	86698.89
3	Africa	Eastern Africa	110	8013.04	44182.60
4	Africa	Central Africa	103	15606.30	61689.99
5	Asia Pacific	Western Asia	382	-16766.90	124312.24
6	Asia Pacific	Southern Asia	469	67998.76	351806.60
7	Asia Pacific	Southeastern Asia	533	20948.84	329751.38
8	Asia Pacific	Oceania	646	54734.02	408002.98
9	Asia Pacific	Eastern Asia	414	72805.10	315390.77
10	Asia Pacific	Central Asia	37	-2649.76	8190.74
11	Europe	Western Europe	964	82091.27	656637.14
12	Europe	Southern Europe	338	18911.49	215703.93
13	Europe	Northern Europe	367	43237.44	252969.09
14	Europe	Eastern Europe	241	25050.69	108258.93
15	LATAM	South America	496	12377.59	210710.49
16	LATAM	Central America	930	74679.54	461670.28
17	LATAM	Caribbean	288	13529.59	116333.05
18	USCA	Western US	490	44303.65	251991.83
19	USCA	Southern US	255	19991.83	148771.91
20	USCA	Eastern US	443	47462.04	264973.98
21	USCA	Central US	356	33697.43	170416.31
22	USCA	Canada	49	7246.62	26298.81

```
df.head()
```

	Market	Region	No_of_Orders	Profit	Sales
0	Africa	Western Africa	251	-12901.51	78476.06
1	Africa	Southern Africa	85	11768.58	51319.50
2	Africa	North Africa	182	21643.08	86698.89
3	Africa	Eastern Africa	110	8013.04	44182.60
4	Africa	Central Africa	103	15606.30	61689.99

```
df.tail()
```

	Market	Region	No_of_Orders	Profit	Sales
18	USCA	Western US	490	44303.65	251991.83
19	USCA	Southern US	255	19991.83	148771.91
20	USCA	Eastern US	443	47462.04	264973.98
21	USCA	Central US	356	33697.43	170416.31

	Market	Region	No_of_Orders	Profit	Sales
22	USCA	Canada	49	7246.62	26298.81

```
df.head(3)
```

	Market	Region	No_of_Orders	Profit	Sales
0	Africa	Western Africa	251	-12901.51	78476.06
1	Africa	Southern Africa	85	11768.58	51319.50
2	Africa	North Africa	182	21643.08	86698.89

```
### How to get all the column names from excel

df.keys()
```

Index(['Market', 'Region', 'No_of_Orders', 'Profit', 'Sales'], dtype='object')

```
print(df.columns.to_list())
```

['Market', 'Region', 'No_of_Orders', 'Profit', 'Sales']

```
df1 = df.copy()
df.keys()
```

Index(['Market', 'Region', 'No_of_Orders', 'Profit', 'Sales'], dtype='object')

```
df.rename(columns={"Profit":"Profit_int"}, inplace = True)
df.head()
```

	Market	Region	No_of_Orders	Profit_int	Sales
0	Africa	Western Africa	251	-12901.51	78476.06
1	Africa	Southern Africa	85	11768.58	51319.50
2	Africa	North Africa	182	21643.08	86698.89
3	Africa	Eastern Africa	110	8013.04	44182.60
4	Africa	Central Africa	103	15606.30	61689.99

```
df.drop("Profit_int", axis =1).head(2)
```

	Market	Region	No_of_Orders	Sales
0	Africa	Western Africa	251	78476.06
1	Africa	Southern Africa	85	51319.50

```
# Axis = 0 ------rows
# axis = 1 ------Columns
```

```
df.drop(["Market","Sales"], axis = 1).head(10)
```

```
df
```

	Region	No_of_Orders	Profit_int
0	Western Africa	251	-12901.51
1	Southern Africa	85	11768.58
2	North Africa	182	21643.08
3	Eastern Africa	110	8013.04
4	Central Africa	103	15606.30
5	Western Asia	382	-16766.90
6	Southern Asia	469	67998.76
7	Southeastern Asia	533	20948.84
8	Oceania	646	54734.02
9	Eastern Asia	414	72805.10
10	Central Asia	37	-2649.76
11	Western Europe	964	82091.27
12	Southern Europe	338	18911.49
13	Northern Europe	367	43237.44
14	Eastern Europe	241	25050.69
15	South America	496	12377.59
16	Central America	930	74679.54
17	Caribbean	288	13529.59
18	Western US	490	44303.65
19	Southern US	255	19991.83
20	Eastern US	443	47462.04
21	Central US	356	33697.43
22	Canada	49	7246.62

How to select dataframe from pandas DataFrame

iloc and loc

```
path = r"sales.xlsx"
df = pd.read_excel(path)
df
```

	Market	Region	No_of_Orders	Profit	Sales
0	Africa	Western Africa	251	-12901.51	78476.06
1	Africa	Southern Africa	85	11768.58	51319.50
2	Africa	North Africa	182	21643.08	86698.89
3	Africa	Eastern Africa	110	8013.04	44182.60
4	Africa	Central Africa	103	15606.30	61689.99

	Market	Region	No_of_Orders	Profit	Sales
5	Asia Pacific	Western Asia	382	-16766.90	124312.24
6	Asia Pacific	Southern Asia	469	67998.76	351806.60
7	Asia Pacific	Southeastern Asia	533	20948.84	329751.38
8	Asia Pacific	Oceania	646	54734.02	408002.98
9	Asia Pacific	Eastern Asia	414	72805.10	315390.77
10	Asia Pacific	Central Asia	37	-2649.76	8190.74
11	Europe	Western Europe	964	82091.27	656637.14
12	Europe	Southern Europe	338	18911.49	215703.93
13	Europe	Northern Europe	367	43237.44	252969.09
14	Europe	Eastern Europe	241	25050.69	108258.93
15	LATAM	South America	496	12377.59	210710.49
16	LATAM	Central America	930	74679.54	461670.28
17	LATAM	Caribbean	288	13529.59	116333.05
18	USCA	Western US	490	44303.65	251991.83
19	USCA	Southern US	255	19991.83	148771.91
20	USCA	Eastern US	443	47462.04	264973.98
21	USCA	Central US	356	33697.43	170416.31
22	USCA	Canada	49	7246.62	26298.81

```
df.iloc[[0,10,20,22]]
```

	Market	Region	No_of_Orders	Profit	Sales
0	Africa	Western Africa	251	-12901.51	78476.06
10	Asia Pacific	Central Asia	37	-2649.76	8190.74
20	USCA	Eastern US	443	47462.04	264973.98
22	USCA	Canada	49	7246.62	26298.81

```
df.iloc[[1,3,5]]
```

	Market	Region	No_of_Orders	Profit	Sales
1	Africa	Southern Africa	85	11768.58	51319.50
3	Africa	Eastern Africa	110	8013.04	44182.60
5	Asia Pacific	Western Asia	382	-16766.90	124312.24

```
df.iloc[[1,3,5], 2:]
```

	No_of_Orders	Profit	Sales
1	85	11768.58	51319.50
3	110	8013.04	44182.60
5	382	-16766.90	124312.24

```
df.iloc[[1,3,5], ::-1]
```

	Sales	Profit	No_of_Orders	Region	Market
1	51319.50	11768.58	85	Southern Africa	Africa
3	44182.60	8013.04	110	Eastern Africa	Africa
5	124312.24	-16766.90	382	Western Asia	Asia Pacific

```
df.iloc[10::-1, ::-1]
```

	Sales	Profit	No_of_Orders	Region	Market
10	8190.74	-2649.76	37	Central Asia	Asia Pacific
9	315390.77	72805.10	414	Eastern Asia	Asia Pacific
8	408002.98	54734.02	646	Oceania	Asia Pacific
7	329751.38	20948.84	533	Southeastern Asia	Asia Pacific
6	351806.60	67998.76	469	Southern Asia	Asia Pacific
5	124312.24	-16766.90	382	Western Asia	Asia Pacific
4	61689.99	15606.30	103	Central Africa	Africa
3	44182.60	8013.04	110	Eastern Africa	Africa
2	86698.89	21643.08	182	North Africa	Africa
1	51319.50	11768.58	85	Southern Africa	Africa
0	78476.06	-12901.51	251	Western Africa	Africa

Loc()

```
df.loc[[1,2,3,5], "Sales":"Market"]
```

1
2
3
5

```
df.loc[[1,3,5],df.columns.to_list()[::-1]]
```

	Sales	Profit	No_of_Orders	Region	Market
1	51319.50	11768.58	85	Southern Africa	Africa
3	44182.60	8013.04	110	Eastern Africa	Africa
5	124312.24	-16766.90	382	Western Asia	Asia Pacific

```
df[["Sales", "Region", "No_of_Orders", "Profit"]].head()
```

	Sales	Region	No_of_Orders	Profit
0	78476.06	Western Africa	251	-12901.51

	Sales	Region	No_of_Orders	Profit
1	51319.50	Southern Africa	85	11768.58
2	86698.89	North Africa	182	21643.08
3	44182.60	Eastern Africa	110	8013.04
4	61689.99	Central Africa	103	15606.30

```
df.head()
```

	Market	Region	No_of_Orders	Profit	Sales
0	Africa	Western Africa	251	-12901.51	78476.06
1	Africa	Southern Africa	85	11768.58	51319.50
2	Africa	North Africa	182	21643.08	86698.89
3	Africa	Eastern Africa	110	8013.04	44182.60
4	Africa	Central Africa	103	15606.30	61689.99

```
df.set_index("Sales", inplace= True)
df.head()
```

Sales	Region	No_of_Orders	Profit
78476.06	Western Africa	251	-12901.51
51319.50	Southern Africa	85	11768.58
86698.89	North Africa	182	21643.08
44182.60	Eastern Africa	110	8013.04
61689.99	Central Africa	103	15606.30

```
df.iloc[[10,15]]
```

Sales	Region	No_of_Orders	Profit
8190.74	Central Asia	37	-2649.76
210710.49	South America	496	12377.59

```
df.iloc[9]
```

```
Region          Eastern Asia
No_of_Orders             414
Profit               72805.1
Name: 315390.77, dtype: object
```

```
df.loc[:, "Region":"No_of_Orders"].head()
```

Sales	Region	No_of_Orders

Sales	Region	No_of_Orders
78476.06	Western Africa	251
51319.50	Southern Africa	85
86698.89	North Africa	182
44182.60	Eastern Africa	110
61689.99	Central Africa	103

Split and Combine the Data

```
path = r"sales.xlsx"
df = pd.read_excel(path)
df.head()
```

	Market	Region	No_of_Orders	Profit	Sales
0	Africa	Western Africa	251	-12901.51	78476.06
1	Africa	Southern Africa	85	11768.58	51319.50
2	Africa	North Africa	182	21643.08	86698.89
3	Africa	Eastern Africa	110	8013.04	44182.60
4	Africa	Central Africa	103	15606.30	61689.99

```
df["Market"].value_counts()
```

```
Asia Pacific    6
Africa          5
USCA            5
Europe          4
LATAM           3
Name: Market, dtype: int64
```

```
df["Region"].value_counts()
```

```
Western Africa     1
Southern Europe    1
Central US         1
Eastern US         1
Southern US        1
Western US         1
Caribbean          1
Central America    1
South America      1
Eastern Europe     1
Northern Europe    1
Western Europe     1
Southern Africa    1
Central Asia       1
```

```
Eastern Asia          1
Oceania               1
Southeastern Asia     1
Southern Asia         1
Western Asia          1
Central Africa        1
Eastern Africa        1
North Africa          1
Canada                1
Name: Region, dtype: int64
```

```python
#Unique State Name

print(df["Region"].unique())
```

```
['Western Africa' 'Southern Africa' 'North Africa' 'Eastern Africa'
 'Central Africa' 'Western Asia' 'Southern Asia' 'Southeastern Asia'
 'Oceania' 'Eastern Asia' 'Central Asia' 'Western Europe'
 'Southern Europe' 'Northern Europe' 'Eastern Europe' 'South America'
 'Central America' 'Caribbean' 'Western US' 'Southern US' 'Eastern US'
 'Central US' 'Canada']
```

```python
print(df["Region"].nunique())  #No of Unique Regions
```

```
23
```

```python
data = df.copy()
```

```python
data.head(2)
```

	Market	Region	No_of_Orders	Profit	Sales
0	Africa	Western Africa	251	-12901.51	78476.06
1	Africa	Southern Africa	85	11768.58	51319.50

```python
df.head(2)
```

	Market	Region	No_of_Orders	Profit	Sales
0	Africa	Western Africa	251	-12901.51	78476.06
1	Africa	Southern Africa	85	11768.58	51319.50

```python
#Split my data based on Region of my data

df[df["Region"] == "Canada"]
```

	Market	Region	No_of_Orders	Profit	Sales
22	USCA	Canada	49	7246.62	26298.81

```
df["Region"] == "Canada"
```

```
0      False
1      False
2      False
3      False
4      False
5      False
6      False
7      False
8      False
9      False
10     False
11     False
12     False
13     False
14     False
15     False
16     False
17     False
18     False
19     False
20     False
21     False
22      True
Name: Region, dtype: bool
```

```
for state in df["Region"].unique():
    result = df[df["Region"] == state]
    result.to_csv(state + ".csv", index = False)
```

```
result
```

	Market	Region	No_of_Orders	Profit	Sales
22	USCA	Canada	49	7246.62	26298.81

```
df[(df["Region"] == "Western Africa ") & (df["No_of_Orders"]==50)]
```

Market	Region	No_of_Orders	Profit	Sales

```
df[(df["Region"] == "Canada") & (df["No_of_Orders"] == 49)]
```

	Market	Region	No_of_Orders	Profit	Sales
22	USCA	Canada	49	7246.62	26298.81

Pandas Pivot Table

```
path = r"sales.xlsx"
df = pd.read_excel(path)
df.head()
```

	Market	Region	No_of_Orders	Profit	Sales
0	Africa	Western Africa	251	-12901.51	78476.06
1	Africa	Southern Africa	85	11768.58	51319.50
2	Africa	North Africa	182	21643.08	86698.89
3	Africa	Eastern Africa	110	8013.04	44182.60
4	Africa	Central Africa	103	15606.30	61689.99

```
# There are 2 types of Pivto table in Pandas

pd.pivot?
```

```
pd.pivot_table?
```

```
# Based on the region, Can you display the region

pd.pivot_table(data = df, index = "Region", values =  "Sales")
```

	Sales
Region	
Canada	26298.81
Caribbean	116333.05
Central Africa	61689.99
Central America	461670.28
Central Asia	8190.74
Central US	170416.31
Eastern Africa	44182.60
Eastern Asia	315390.77
Eastern Europe	108258.93
Eastern US	264973.98
North Africa	86698.89
Northern Europe	252969.09
Oceania	408002.98
South America	210710.49
Southeastern Asia	329751.38
Southern Africa	51319.50
Southern Asia	351806.60
Southern Europe	215703.93
Southern US	148771.91

	Sales
Region	
Western Africa	78476.06
Western Asia	124312.24
Western Europe	656637.14
Western US	251991.83

```python
pd.pivot_table(data = df, index = ["Region","Market"], values = "Sales", aggfunc = sum
```

		Sales
Region	**Market**	
Canada	USCA	26298.81
Caribbean	LATAM	116333.05
Central Africa	Africa	61689.99
Central America	LATAM	461670.28
Central Asia	Asia Pacific	8190.74

```python
type(pd.pivot_table(data = df, index = ["Region","Market"], values = "Sales", aggfunc
```

pandas.core.frame.DataFrame