

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
```

```
In [12]: xls = pd.ExcelFile('C:/Users/Vinayak BB/Desktop/Datasets/Virtual_Gaming.xlsx')
df1 = pd.read_excel(xls, 'User Demographics')
df2 = pd.read_excel(xls, 'Daily User-wise Revenue data')
```

```
In [13]: df1,df2
```

```
Out[13]: (   User Id State (entered by user) Gender
0    A9247B21      Uttar Pradesh      M
1    F3C79376      Chhattisgarh      F
2    1CC6DCF8      Jharkhand      M
3    69A3C687      Uttar Pradesh      M
4    FF5B5B7F      Karnataka      M
..      ...      ...      ...
995  6132EE3F      West Bengal      M
996  6F7CB4FF      Madhya Pradesh      M
997  B509489A      Utharakhand      M
998  C9A9F3D3      West Bengal      M
999  4E1E927A      Uttar Pradesh      F

[1000 rows x 3 columns],
   Date    User id  Revenue collected
0  2021-04-01  FF5B5B7F              53
1  2021-04-01  18427542              68
2  2021-04-01  C0307F5B              85
3  2021-04-01  C56C2457              90
4  2021-04-01  E7C82918              75
...      ...      ...      ...
14985 2021-04-30  F268AC1E              48
14986 2021-04-30  2A8C083D              56
14987 2021-04-30  1DECEC9C              68
14988 2021-04-30  6F7CB4FF              39
14989 2021-04-30  C9A9F3D3              5

[14990 rows x 3 columns])
```

```
In [17]: df2.rename(columns={"User id":"User Id"},inplace=True)
df1.rename(columns={"State (entered by user)":"State"},inplace=True)
```

```
In [18]: df1,df2
```

```
Out[18]: (
   User Id      State Gender
0  A9247B21  Uttar Pradesh    M
1  F3C79376  Chhattisgarh    F
2  1CC6DCF8    Jharkhand    M
3  69A3C687  Uttar Pradesh    M
4  FF5B5B7F    Karnataka    M
..      ...      ...      ...
995 6132EE3F  West Bengal    M
996 6F7CB4FF  Madhya Pradesh    M
997 B509489A  Utharakhand    M
998 C9A9F3D3  West Bengal    M
999 4E1E927A  Uttar Pradesh    F

[1000 rows x 3 columns],
   Date      User Id  Revenue collected
0  2021-04-01  FF5B5B7F                53
1  2021-04-01  18427542                68
2  2021-04-01  C0307F5B                85
3  2021-04-01  C56C2457                90
4  2021-04-01  E7C82918                75
...      ...      ...      ...
14985 2021-04-30  F268AC1E                48
14986 2021-04-30  2A8C083D                56
14987 2021-04-30  1DECEC9C                68
14988 2021-04-30  6F7CB4FF                39
14989 2021-04-30  C9A9F3D3                5

[14990 rows x 3 columns])
```

```
In [33]: df3 =pd.merge(df1,df2,on="User Id")
df4=df3.groupby(["User Id"]).first()
```

```
In [51]: df5=df3.groupby(["State"]).sum()  
df5
```

Out[51]:

Revenue collected	
State	
AP	32068
Andhra Pradesh	7062
Assam	17208
Bihar	39235
Chathisgarh	11750
Chattisgarh	38025
Chhattisgarh	25462
Delhi	22436
Gujarat	11686
HP	10228
Haryana	34359
Himachal Pradesh	5822
J&K	1576
Jammu & Kashmir	27165
Jammu and Kashmir	3814
Jharkhand	21795
KAR	3904
Kar	30237
Karnataka	58407
MP	3927
Madhya Pradesh	40016
Maharashtra	6257
New Delhi	15481

Revenue collected	
State	
Odisha	1422
Orissa	20600
Punjab	8208
Rajasthan	34506
TN	26050
Tamil Nadu	39647
Telangana	33099
Telengana	18624
UP	1556
Utharakhand	44657
Uttar Pradesh	16080
Uttarakhand	15168
WB	35778
West Bengal	34462

```
In [54]: Q1=df5.sort_values(["Revenue collected"],ascending=False)[0:5]
Q1#question1-Top5 states in terms of revenue
```

Out[54]:

Revenue collected	
State	
Karnataka	58407
Utharakhand	44657
Madhya Pradesh	40016
Tamil Nadu	39647
Bihar	39235

```
In [71]: display(Q1)
```

Revenue collected	
State	
Karnataka	58407
Utharakhand	44657
Madhya Pradesh	40016
Tamil Nadu	39647
Bihar	39235

```
In [119]: df5=df3.groupby(["Date"]).sum()[0:5]
df5.sort_values("Revenue collected",ascending=False)#revenue collected by date top first five
```

Out[119]:

Revenue collected	
Date	
2021-04-03	48946
2021-04-04	48806
2021-04-05	18643
2021-04-01	18381
2021-04-02	18264

```
In [120]: df3["Date"].value_counts().mean()#Q2 Answer Avg number of User Active Per Day
```

Out[120]: 499.6666666666667

```
In [189]: A=df1['User Id'].value_counts().sum()
A#number of user
```

Out[189]: 1000

```
In [199]: df3.groupby(['Date']).mean()#Revenue Collected Per Day
```

Out[199]:

Revenue collected	
Date	
2021-04-01	52.971182
2021-04-02	53.560117
2021-04-03	51.467928
2021-04-04	51.537487
2021-04-05	54.037681
2021-04-06	54.247093
2021-04-07	53.936416
2021-04-08	53.742775
2021-04-09	55.396450
2021-04-10	52.331197
2021-04-11	52.317647
2021-04-12	54.356725
2021-04-13	52.976744
2021-04-14	55.278107
2021-04-15	53.704142
2021-04-16	54.675595
2021-04-17	52.659893
2021-04-18	52.599574
2021-04-19	53.970326
2021-04-20	53.885057
2021-04-21	53.492711
2021-04-22	53.331378
2021-04-23	54.397059
2021-04-24	52.664509

Revenue collected

Date	
2021-04-25	53.175081
2021-04-26	55.363363
2021-04-27	54.373529
2021-04-28	53.542522
2021-04-29	53.885196
2021-04-30	55.257576

```
In [204]: df3.groupby(['User Id', 'Date']).mean()
#Q3-Avg Revenue Collected Per User Per Day
```

Out[204]:

Revenue collected		
User Id	Date	
00395C20	2021-04-01	50
	2021-04-02	33
	2021-04-03	37
	2021-04-04	34
	2021-04-05	54
...
FFAD51FD	2021-04-11	49
	2021-04-17	40
	2021-04-18	37
	2021-04-24	41
	2021-04-25	41

14990 rows × 1 columns


```
In [205]: df3.groupby(['Date', 'User Id']).mean()  
#Avg Revenue Collected Per Day Per User
```

Out[205]:

Revenue collected		
Date	User Id	
2021-04-01	00395C20	50
	00583276	8
	008B558C	36
	01DA0F63	43
	024FF02E	58
...
2021-04-30	FCA529B3	100
	FD322A9E	117
	FDC87F68	32
	FE4BE4E9	35
	FF5B5B7F	57

14990 rows × 1 columns

In []: