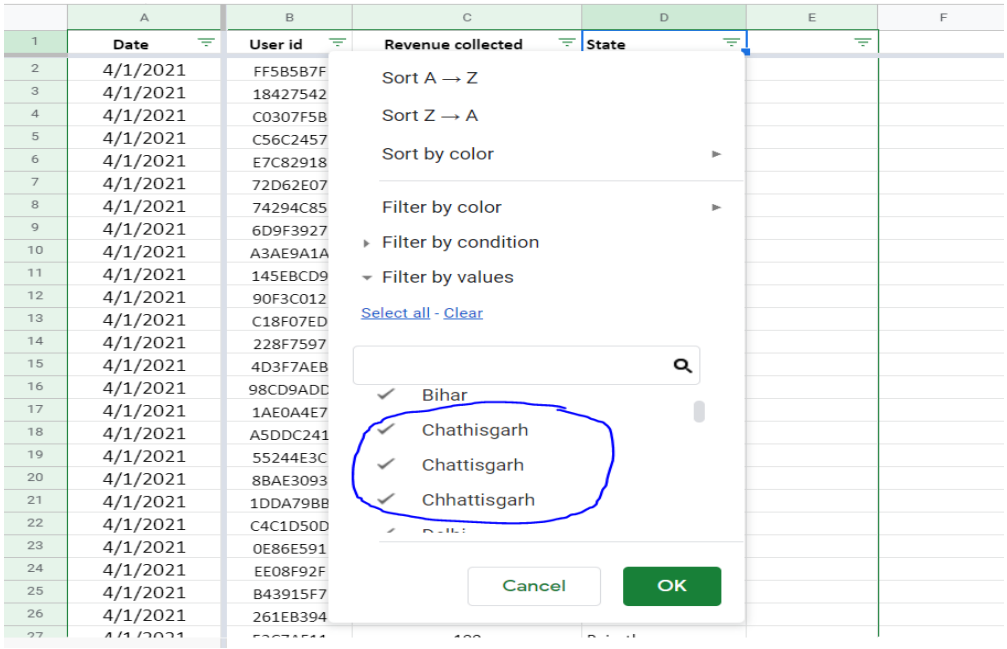
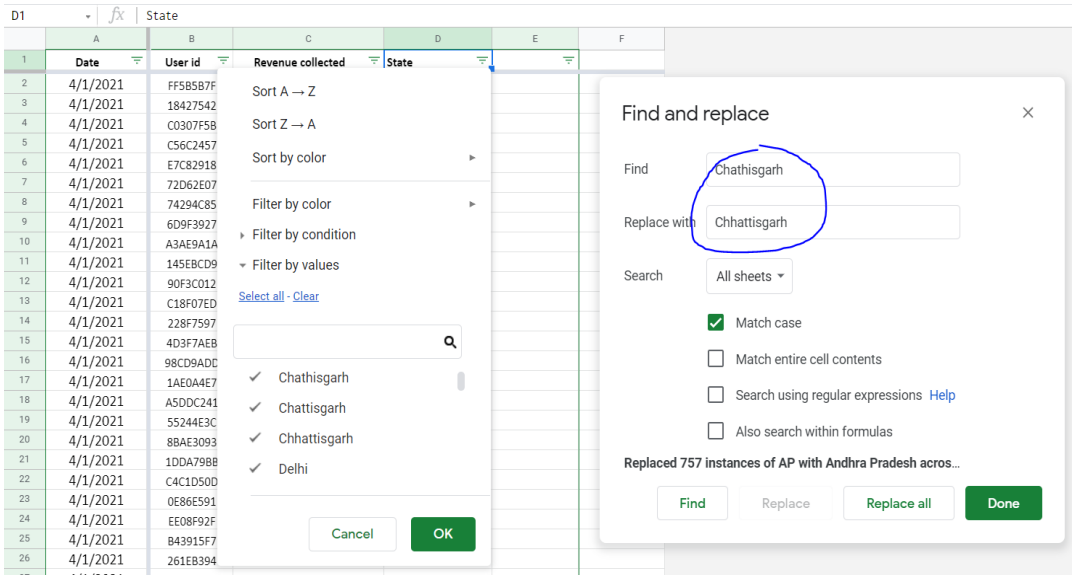
Documentation

Step 1: Cleaning the data

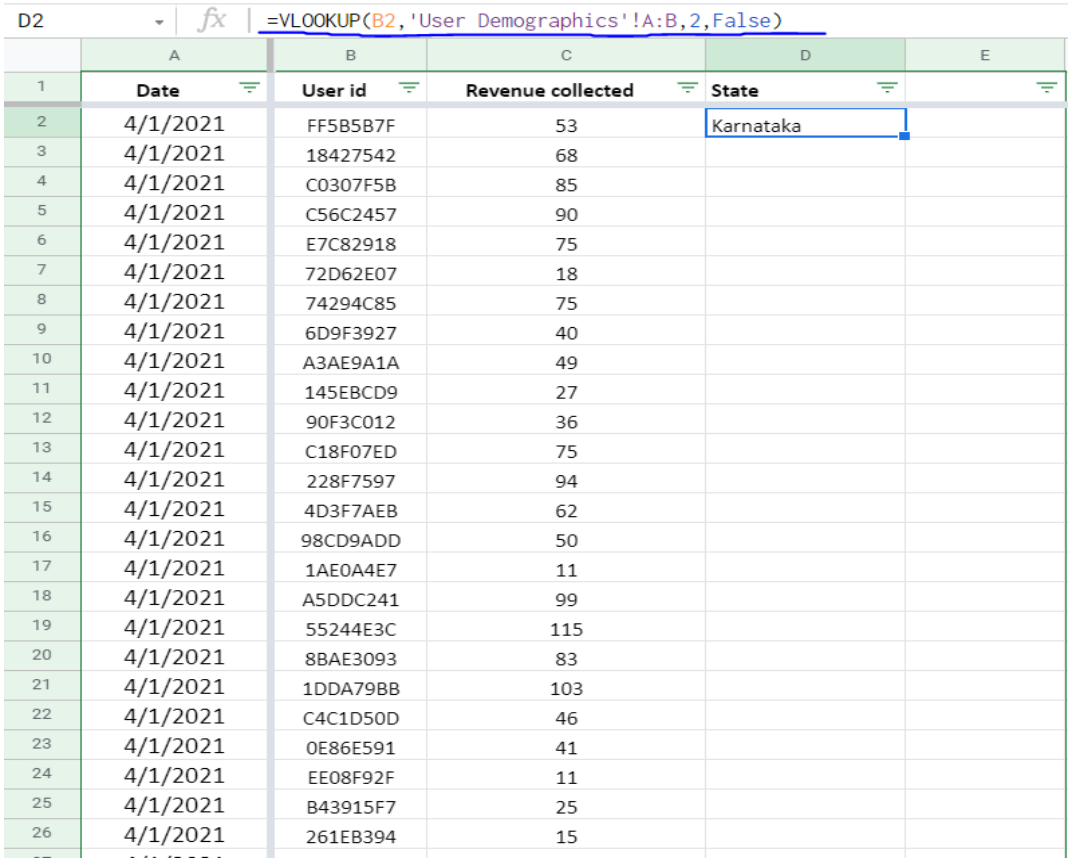
- Some states were named abbreviated in the state column at some places (E.g., Andhra Pradesh and AP). There are also spelling mistakes in the state column. By using the filter option, we can find out all those states.



- After finding those errors, we use the **find and replace** option to make the data consistent.



- We used the **VLOOKUP** function to aggregate and import states belonging to the user IDs in the revenue data. After using that, we used the filter function to verify the user IDs in both sheets. As it has not given NA in the states column, we can confirm that user ID data in both the sheets is consistent and the same.



	A	B	C	D	E
1	Date	User id	Revenue collected	State	
2	4/1/2021	FF5B5B7F	53	Karnataka	
3	4/1/2021	18427542	68		
4	4/1/2021	C0307F5B	85		
5	4/1/2021	C56C2457	90		
6	4/1/2021	E7C82918	75		
7	4/1/2021	72D62E07	18		
8	4/1/2021	74294C85	75		
9	4/1/2021	6D9F3927	40		
10	4/1/2021	A3AE9A1A	49		
11	4/1/2021	145EBCD9	27		
12	4/1/2021	90F3C012	36		
13	4/1/2021	C18F07ED	75		
14	4/1/2021	228F7597	94		
15	4/1/2021	4D3F7AEB	62		
16	4/1/2021	98CD9ADD	50		
17	4/1/2021	1AE0A4E7	11		
18	4/1/2021	A5DDC241	99		
19	4/1/2021	55244E3C	115		
20	4/1/2021	8BAE3093	83		
21	4/1/2021	1DDA79BB	103		
22	4/1/2021	C4C1D50D	46		
23	4/1/2021	0E86E591	41		
24	4/1/2021	EE08F92F	11		
25	4/1/2021	B43915F7	25		
26	4/1/2021	261EB394	15		

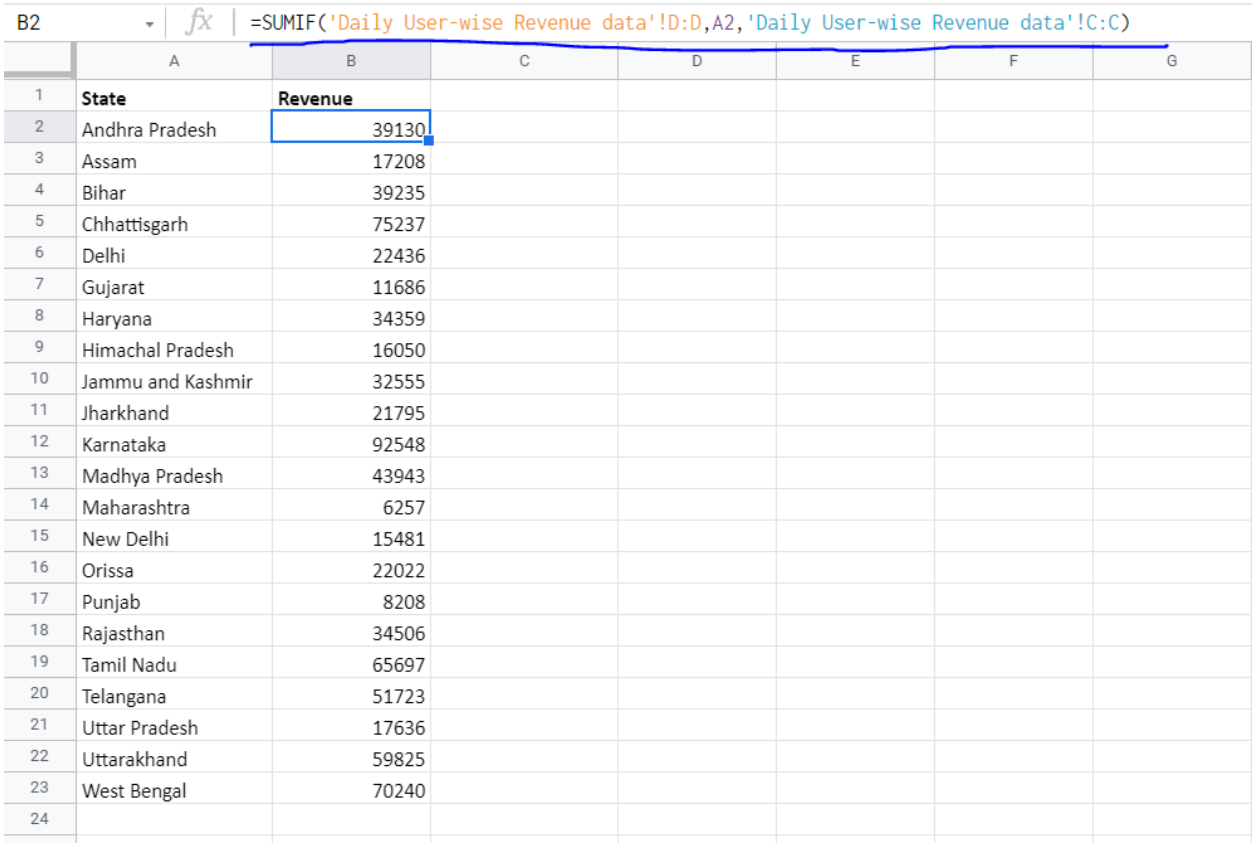
- We have also cleaned the revenue collected data by checking **MIN** and **MAX** of the revenue column is **1** and **170** and is consistent with the data.

Recommendation

- To avoid multiple entries for the names of states, we can use the **Data Validation** tool to use a consistent names dropdown menu since the number of states is limited.

Question 1: State-wise Revenue and Top 5 states

- We created a separate sheet, copy-pasted all the states, and removed duplicates. Then, we use the **SUMIF** function to add the revenue corresponding to each state.



B2 | =SUMIF('Daily User-wise Revenue data'!D:D,A2,'Daily User-wise Revenue data'!C:C)

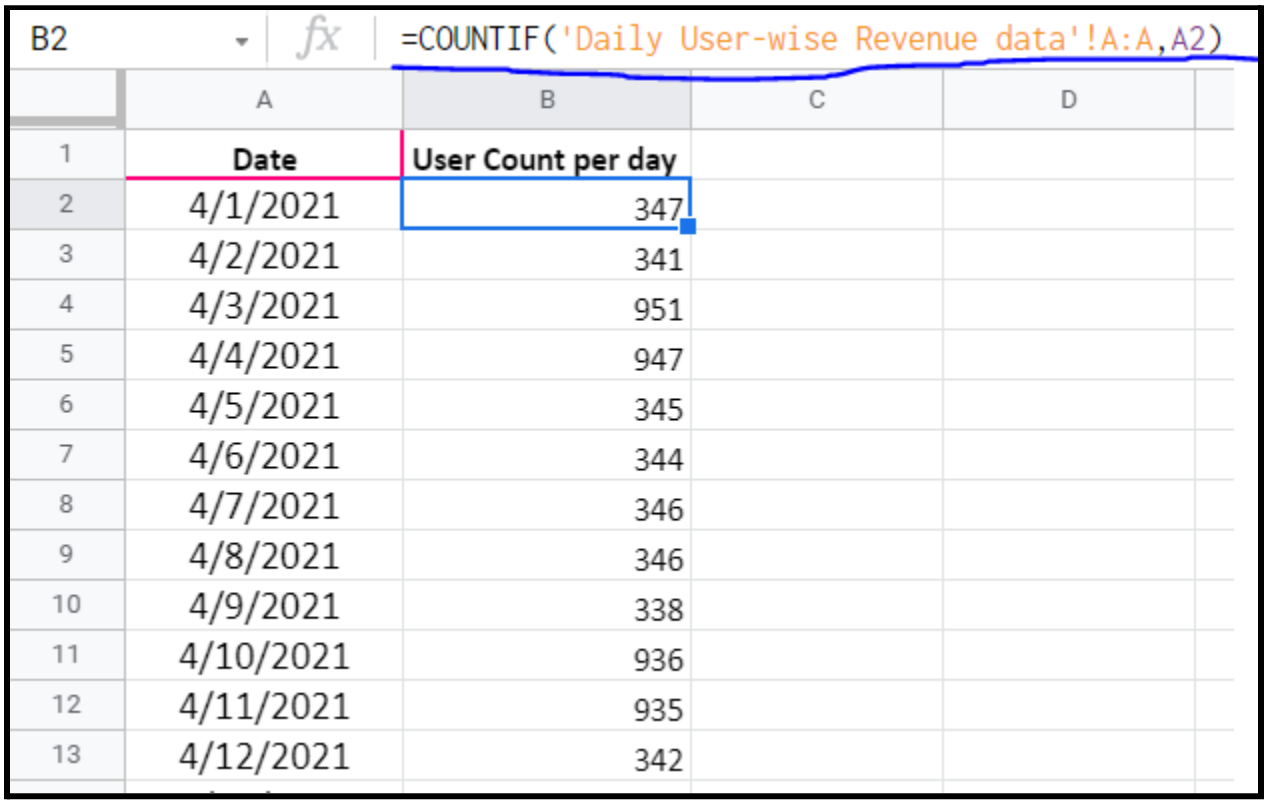
	A	B
1	**State**	**Revenue**
2	Andhra Pradesh	39130
3	Assam	17208
4	Bihar	39235
5	Chhattisgarh	75237
6	Delhi	22436
7	Gujarat	11686
8	Haryana	34359
9	Himachal Pradesh	16050
10	Jammu and Kashmir	32555
11	Jharkhand	21795
12	Karnataka	92548
13	Madhya Pradesh	43943
14	Maharashtra	6257
15	New Delhi	15481
16	Orissa	22022
17	Punjab	8208
18	Rajasthan	34506
19	Tamil Nadu	65697
20	Telangana	51723
21	Uttar Pradesh	17636
22	Uttarakhand	59825
23	West Bengal	70240
24		

- We use the **SORT RANGE** function to sort out the top 5 revenue-generating states.

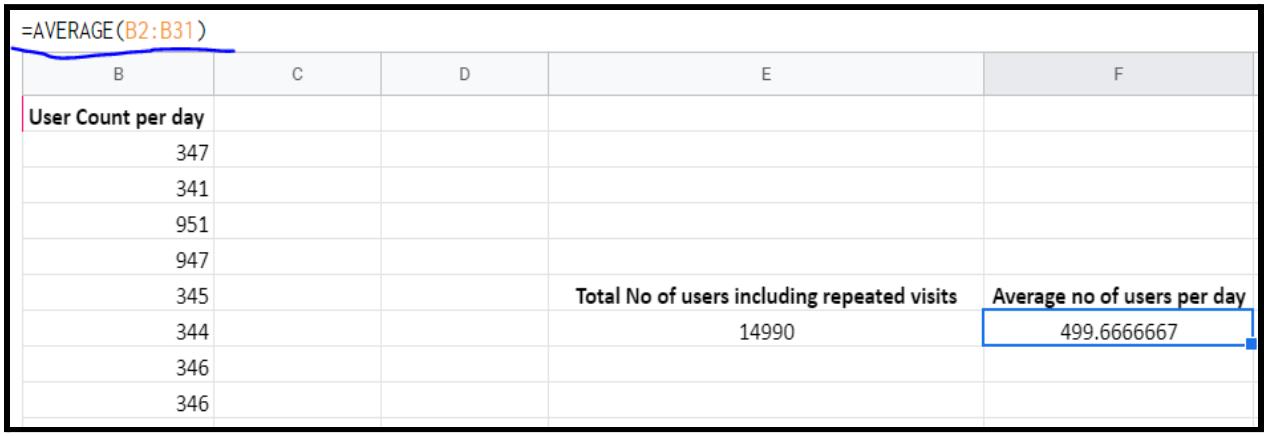
	A	B	C
1	**Ranking**	**State**	**Revenue**
2	1	Karnataka	92548
3	2	Chhattisgarh	75237
4	3	West Bengal	70240
5	4	Tamil Nadu	65697
6	5	Uttarakhand	59825

Question 2: Average number of users per day

- We can directly count the number of rows in the revenue sheet and divide it by 30 to get the average number of users per day, which is approximately **500**.
- But to do the procedure properly using maths, we have calculated users per day using the **COUNTIF** function and found the average using the **AVERAGE** function.



	A	B	C	D
1	Date	User Count per day		
2	4/1/2021	347		
3	4/2/2021	341		
4	4/3/2021	951		
5	4/4/2021	947		
6	4/5/2021	345		
7	4/6/2021	344		
8	4/7/2021	346		
9	4/8/2021	346		
10	4/9/2021	338		
11	4/10/2021	936		
12	4/11/2021	935		
13	4/12/2021	342		



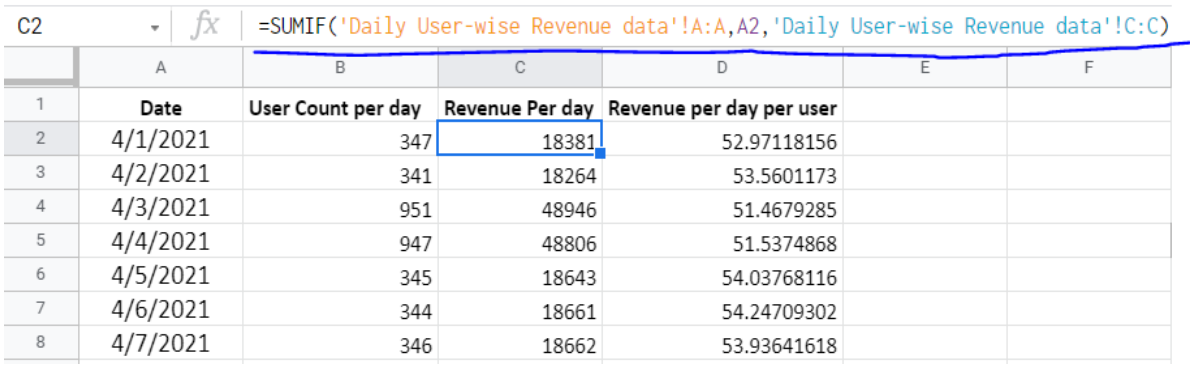
B	C	D	E	F
User Count per day				
347				
341				
951				
947				
345			Total No of users including repeated visits	Average no of users per day
344			14990	499.6666667
346				
346				

Question 3: Average revenue per user per day

- Here also we can find the average revenue per day per user directly with formula

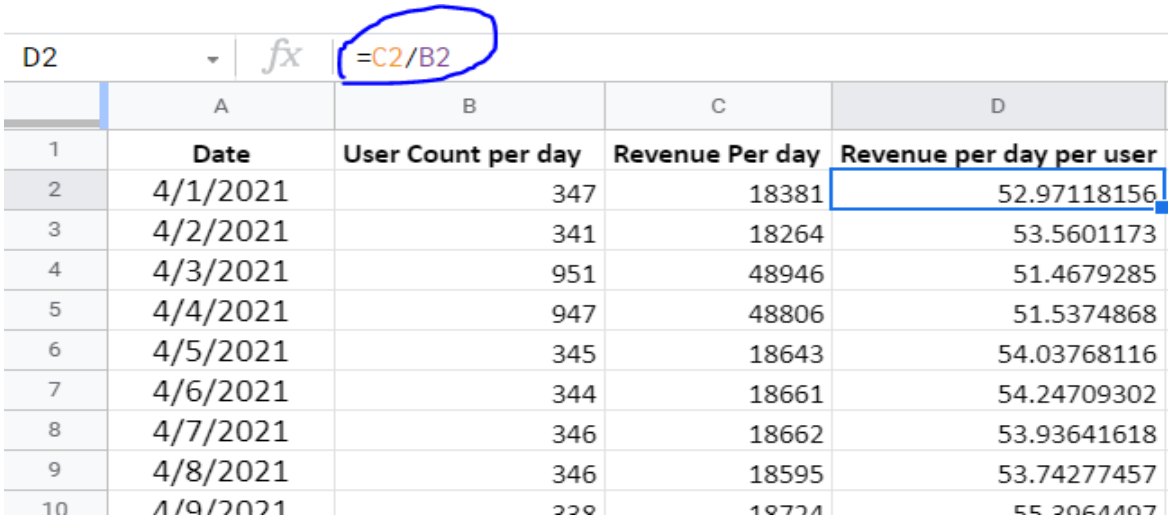
$$\frac{Total\ revenue}{Total\ number\ of\ users * No.of\ days}$$

- We get **53.6** as the answer.
- But to make the analysis proper, we use the **SUMIF** and the **AVERAGE** formulas in sheets to calculate the answer.



C2 =SUMIF('Daily User-wise Revenue data'!A:A,A2,'Daily User-wise Revenue data'!C:C)

	A	B	C	D	E	F
1	Date	User Count per day	Revenue Per day	Revenue per day per user		
2	4/1/2021	347	18381	52.97118156		
3	4/2/2021	341	18264	53.5601173		
4	4/3/2021	951	48946	51.4679285		
5	4/4/2021	947	48806	51.5374868		
6	4/5/2021	345	18643	54.03768116		
7	4/6/2021	344	18661	54.24709302		
8	4/7/2021	346	18662	53.93641618		



D2 =C2/B2

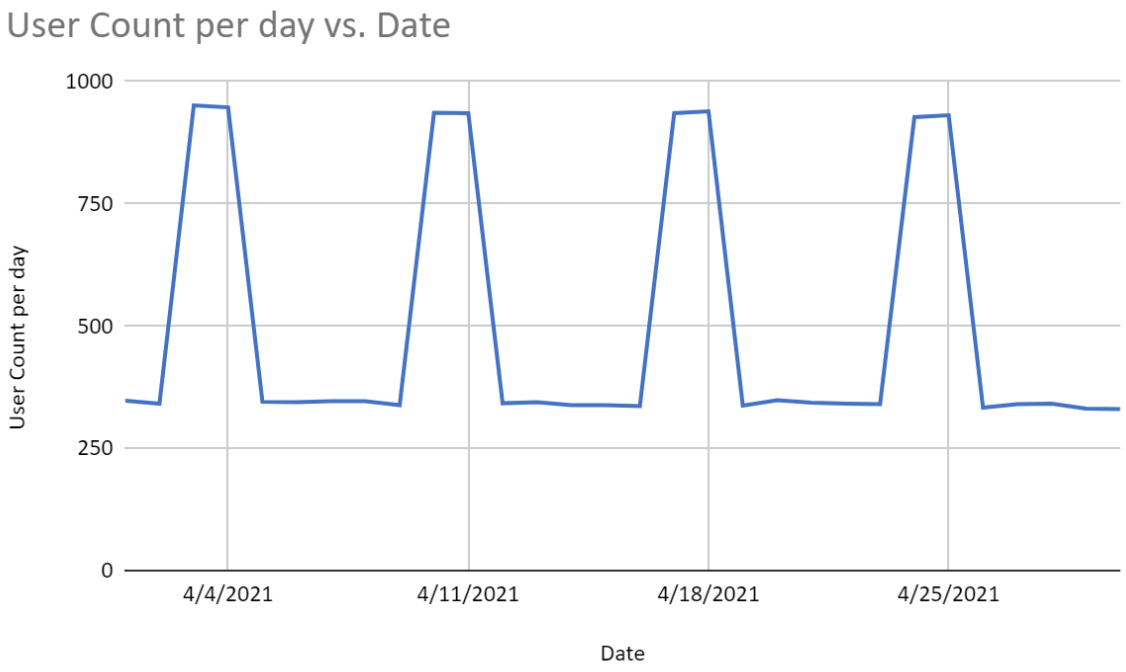
	A	B	C	D
1	Date	User Count per day	Revenue Per day	Revenue per day per user
2	4/1/2021	347	18381	52.97118156
3	4/2/2021	341	18264	53.5601173
4	4/3/2021	951	48946	51.4679285
5	4/4/2021	947	48806	51.5374868
6	4/5/2021	345	18643	54.03768116
7	4/6/2021	344	18661	54.24709302
8	4/7/2021	346	18662	53.93641618
9	4/8/2021	346	18595	53.74277457
10	4/9/2021	338	18724	55.3964497

=AVERAGE(D2:D31)

B	C	D	E	F	G
User Count per day	Revenue Per day	Revenue per day per user			
347	18381	52.97118156			
341	18264	53.5601173			
951	48946	51.4679285			
947	48806	51.5374868			Average revenue per day per user
345	18643	54.03768116			53.63783539
344	18661	54.24709302			
346	18662	53.93641618			
346	18595	53.74277457			

Bonus: Insights on the given data

- If we observe trends in the data, we can see that the number of users has spiked on weekends suggesting that the **users are engaging more on weekends.**



- If we analyze the data by segregating genders, we can see that the number of male users and revenue generated by them is twice the number of female users. But surprisingly, the **average revenue per user per day is more in female users**. We can point out that female users are slightly generating more revenue than male users per day irrespective of the number of users.

Gender	Users	Average users per day	Revenue	Average revenue per day per user
F	304	149.2	245867	1647.902145
M	696	350.4666667	551910	1574.786

Note:

We have not used the **paste values only** option and left the formulas for your reference.