Eetu Haikola 22/11/2023
*Tarvittaessa dokumentin kääntäminen suomenkiellelle myös onnistuu.*

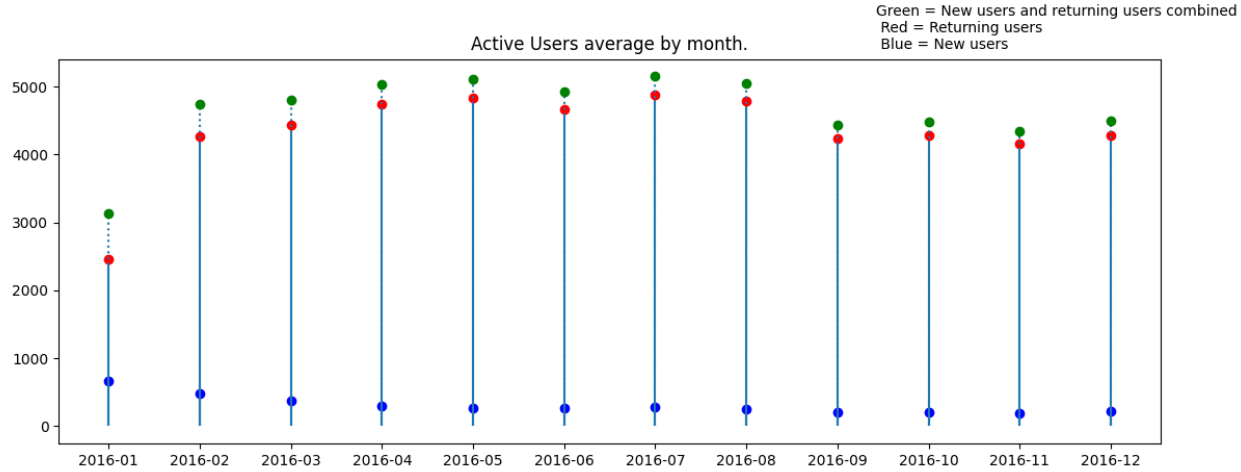
All of the code is written using Python 3.11.5.
Libraries: OS, NumPy, Matplotlib.pyplot, sqlite3, pandas, calendar

Python code and other files are available from link: https://github.com/asuzi/Data-localization

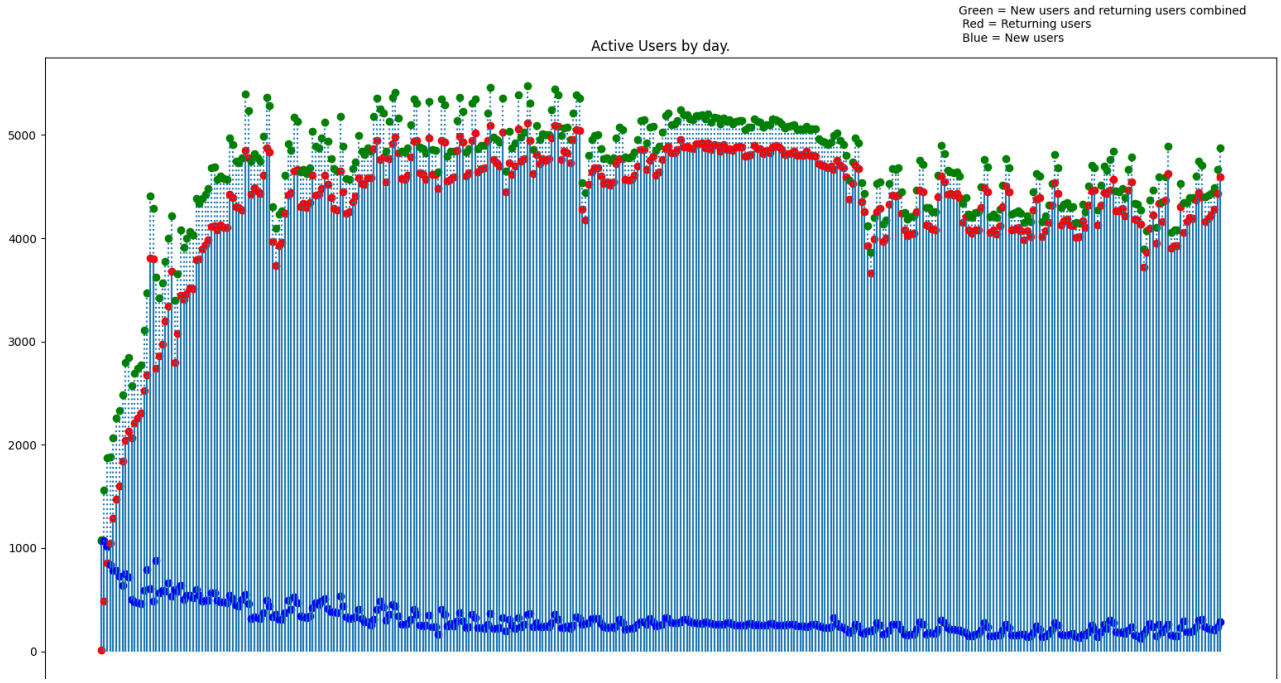
Task 1: Give us short description of datasets.

The dataset contains three tables one for when and where the account was created, when and how much money spent to in-app purchases, when and how long and many sessions has been created by the user.

Task 2: Analyze the daily active users.




*MAU, Monthly active users. Calculated by getting the average activity per month. (Returning user + New user / days_in_month)*




*DAU, Daily active users. (Returning user + New user)*

1) Compare DAU changes over time.

For clarity I will be looking at the DAU changes over time from the Monthly perspective.

Looking at the average daily active users shows an upward curve from 2/2016 to 8/2016 with only exception being 6/2016 with slight drop in the active users. And from 9/206 to 12/2016 the active users drop by a little and stays a little bit lower than the previous months.

2) Can you identify any trends in data?
From 05/2016 and forward the average count for new users stays stable. Which in my opinion is very good trend to have as it tells your product is steadily growing and reaching out to new users.

The spike in new users in 01/2016 – 04/2016 may be due to successful launch or the database being a cut out of an already existing application meaning already existing users would be looked as “new" users.

3) Can you find any ups or drops that are out of the normal behavior?

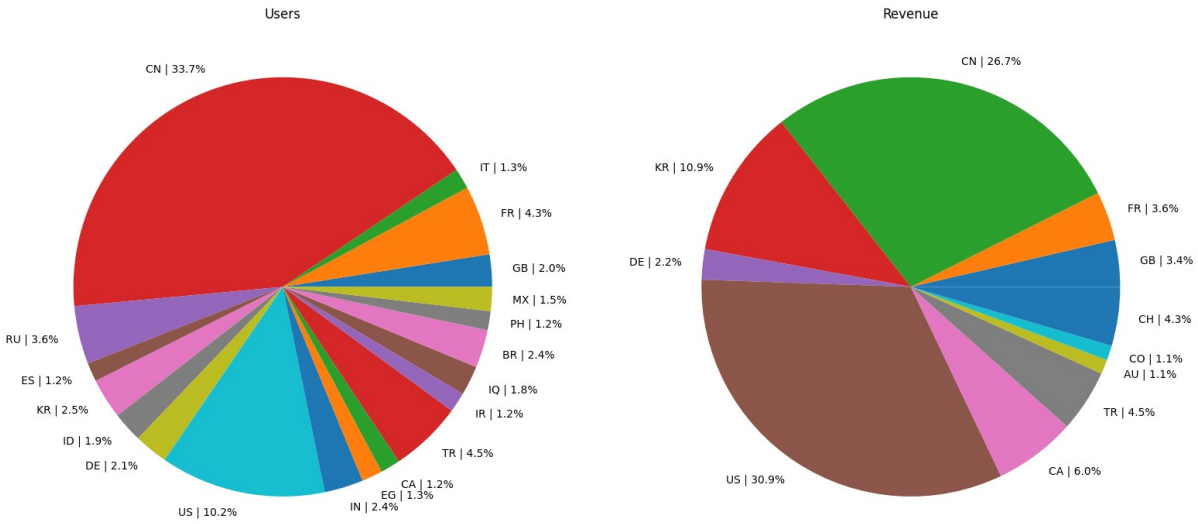
- 06/2016 has a slight drop in average active users comparing to 5/2016 and 7/2016.
- Months from 9/2016 to 12/2016 are averaging way less active users than the previous months.
- There is a high surge in new users at the start months of the year 2016 (1/2016 – 3/2016)

4) What do you think why do they happen?
The drop in 06/2016 may have been caused by many reasons making the application not feel like as relevant or rewarding as it maybe previously has. Maybe there has been issues with the application.

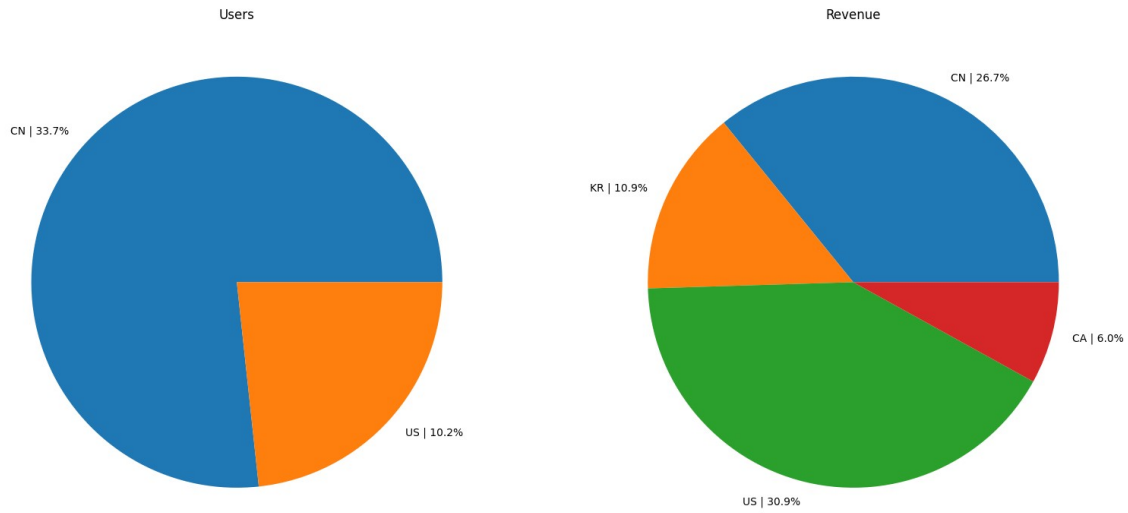
How ever the drop in 06/2016 comes just two months before the drop in 09/2016. Maybe the reason is same for the drop as it was for 06/2016 or maybe the drops in users are connected somehow.

The high surge in new users at the starting months of the 2016 may be caused by the database being a slice of already existing dataset making the already returning users seem like new users.

Task 3: Analyze sales.




*This pie plot shows all of the users and the revenue per country only if they contribute 1% or more.*




*This plot shows all of the users and the revenue per country only if they contribute 5% or more.*

1) Analyze the geographic split of the revenue and the users.
The geographic split of the revenue and the users can be seen from the pie plots above. How ever both of these plots are filtered for readability.

2) Calculate average revenue per user per market.
This can be found in the task_3_2.txt file, which includes all countries and all data. (Data precision can be changed from the code).

3) What are your observations of the results?
Top 3 -
Countries with most users

```
- CN |User: 33.73% |Revenue: 26.73% |Avg. revenue 0.30 USD per user.
- US |User: 10.24% |Revenue: 30.92% |Avg. revenue 1.14 USD per user.
- TR |User: 4.49%  |Revenue: 4.55%  |Avg. revenue 0.38 USD per user.
```

Countries with most revenue

```
- US |User: 10.24% |Revenue: 30.92% |Avg. revenue 1.14 USD per user.
- CN |User: 33.73% |Revenue: 26.73% |Avg. revenue 0.30 USD per user.
- KR |User: 2.50%  |Revenue: 10.90% |Avg. revenue 1.65 USD per user.
```

Countries with highest average revenue per user

```
- CH |User: 0.18%  |Revenue: 4.30%  |Avg. revenue 9.23 USD per user.
- CA |User: 1.22%  |Revenue: 5.99%  |Avg. revenue 1.85 USD per user.
- KR |User: 2.50%  |Revenue: 10.90% |Avg. revenue 1.65 USD per user.
```

Countries with low user % but high revenue may mean that there are probably some high spenders. For example these users could enjoy more pack-deals and separate purchases.

Countries with high user % and high revenue may mean that there may be more users willing to spend money in application, but not too much. For example these users might enjoy subscription type of products.

Countries with high or medium user % and low revenue may be countries where their local currency is not worth as much as a dollar making their revenue seem low even if they are “high

spenders”. For countries like this it would be beneficial to offer cheaper prices to match the local market.