

Free-to-play Mobile Game Data Analysis

Samuel Vitek

CONTENTS

1 INTRODUCTION - QUESTIONS.....	1
2 SHORT DESCRIPTION OF DATASETS	2
2.1 Relationships.....	2
2.2 Table “account”.....	2
2.3 Table “iap_purchase”	3
2.4 Table “account_date_session”.....	3
3 DAILY ACTIVE USERS	4
3.1 First quarter usage increase	5
3.2 Time between July and August.....	6
3.3 Anomaly between 3 rd and 9 th October.....	6
3.4 Halloween.....	8
4 SALES.....	9
4.1 Total users by country	9
4.2 Total revenue by country.....	10
4.3 Average revenue per user per country	10
5 CONCLUSION	13
REFERENCES.....	10

1 INTRODUCTION - QUESTIONS

This report answers questions about data from a hypothetical free-to-play mobile game.

Report is structured based on the following given questions [1]:

1. Give us short description of datasets.
2. Analyse the daily active users:
 - Compare DAU changes over time.
 - Can you identify any trends in data?
 - Can you find any ups or drops that are out of the normal behaviour?
 - What do you think why do they happen?
3. Analyse sales:
 - Analyse the geographic split of the revenue and the users.
 - Calculate average revenue per user per market.
 - What are your observations of the results?

Necessary data handling was done using Python libraries Pandas for data handling and Plotly for [inter-active data visualisations](#) and some of the data validation was done by SQL in PyCharm environment. Unfortunately, it is not possible to integrate the interactive visualizations into the pdf document and there was not enough time to create and deploy a dashboard with all the visual.

After downloading the data, it was necessary to analyse a sample, look at data quality of the datasets, clean the data (handling null values in this case), selecting an approach, creating visualisation, validating its correctness, and finally making observations.

All visualisations and code files are accessible in a public GitHub repository. [2]

2 SHORT DESCRIPTION OF DATASETS

This chapter describes general overview of given datasets, their relationships, and content of each table.

Three datasets were given by the task. Datasets “account”, “account_date_session” and “iap_purchase” where the dataset “account” contains all created user profiles, “iap_purchase” contains all in-app purchases made by the users, and “account_date_session” contains the number of sessions made by users. All data in those datasets are describing usage of the mobile game in the year 2016.

2.1 Relationships

As visible in the figure below these three tables are all related. Their relations are one to many from account to the other two based on accounts primary key column ‘account_id’ which is representing one user account.



FIGURE 1: Visualisation of relationships between the tables

2.2 Table “account”

This table’s purpose as mentioned before is to hold information about each created account (in this case in year 2016) and all necessary information about it. Table has 112 792 rows (i.e., 112 792 distinct accounts created in total in 2016). The following list goes through columns and their meanings:

- **account_id** – unique key distinguishing each account
- **created_time** – date of accounts creation in format ‘yyyy-MM-dd HH:mm:ss,SSS’
- **created_device** – device on which the account was created

- created_platform – device on which the account was created (iOS or Android)
- country_code – universal two letter code of country (e.g., ‘FI’ is code for Finland)
- created_app_store_id – key distinguishing from which app store the app was downloaded

There were null values found in column country_code. It was necessary to delete those rows, since later when geographical visualisations were needed, it was necessary to create a new column with full names of countries based on column country_code and then to use it to create the plot itself. That would not be possible with dataset containing null values. Other possible solution would be to fill the values based on other columns and previous data using machine learning and some type of classification.

2.3 Table “iap_purchase”

Table “iap_purchase” holds all necessary details about in app purchases made by different users in the year 2016. Size of this table is 9 909 rows in other words there were 9 909 purchases made in 2016. The following list goes through columns and their meanings:

- account_id – foreign key making a possibility of distinguishing which account made the concrete purchase
- created_time - date of purchases in format ‘yyyy-MM-dd HH:mm:ss,SSS’
- package_id_hash – unique key with purpose of distinguishing what was purchased in each purchase (i.e., which package)
- iap_price_usd_cents – price paid for the purchase in USD cents (1USD = 100 cents)
- app_store_id - key distinguishing from which app store the purchase was made from

2.4 Table “account_date_session”

This table has all information about sessions made by users. It holds records of how many times each user opened the game and in which day the user did so. Size of this table is 1 698 974 rows. The following list goes through columns and their meanings:

- account_id – foreign key making a possibility of distinguishing which account opened the game
- date – date of purchases in format ‘yyyy-MM-dd’
- session_count – how many times the user opened the game that day
- session_duration_sec – sum of how long were the sessions

3 DAILY ACTIVE USERS

Daily active users (DAU) is a metric of how many users were actively using the app during each day in this case in 2016. This metric brings very significant amounts insights.



FIGURE 2: Daily Active Users chart with local regression trendline

The chart consists of two lines. The blue line represents the exact number of users, who used the app in the specific day whereas the red line smoothens the into some overall trend. Since the data itself is not descriptive enough for some points it is only possible to make an assumption not a conclusion.

3.1 Overall trends

To the overall overview of the plot there are visible regular ups and downs quite often (excluding time between the beginning of July and the end of August). The ups in the blue line are representing weekends where users usually tend to have more time to play. The drops in between weekends are the workdays when users have less time. This trend is quite commonly occurring in entertainment app usages.

Regarding previous statements, it is also visible that during the weekdays DAU tends to go often once a week a little higher than the other weekdays. There is not enough data and business insights to be able

to make a conclusion, but it might be caused by some weekly rewards bringing the users back to the game during the week.

There are four recurring main dips in the usage. These might have been caused by a regular update of the app (might be necessary regularly for example because of starting a new season in the game) where users would not be able to use the app until they update it which could also possibly explain that the first two weeks after the dip have lower usage rates.

3.2 First quarter usage increase

Uptrend in the first quarter of the year is most probably caused by release of the mobile game. Since the marketing strategies usually begin before the game release the rapid increase in users at the beginning would make sense.

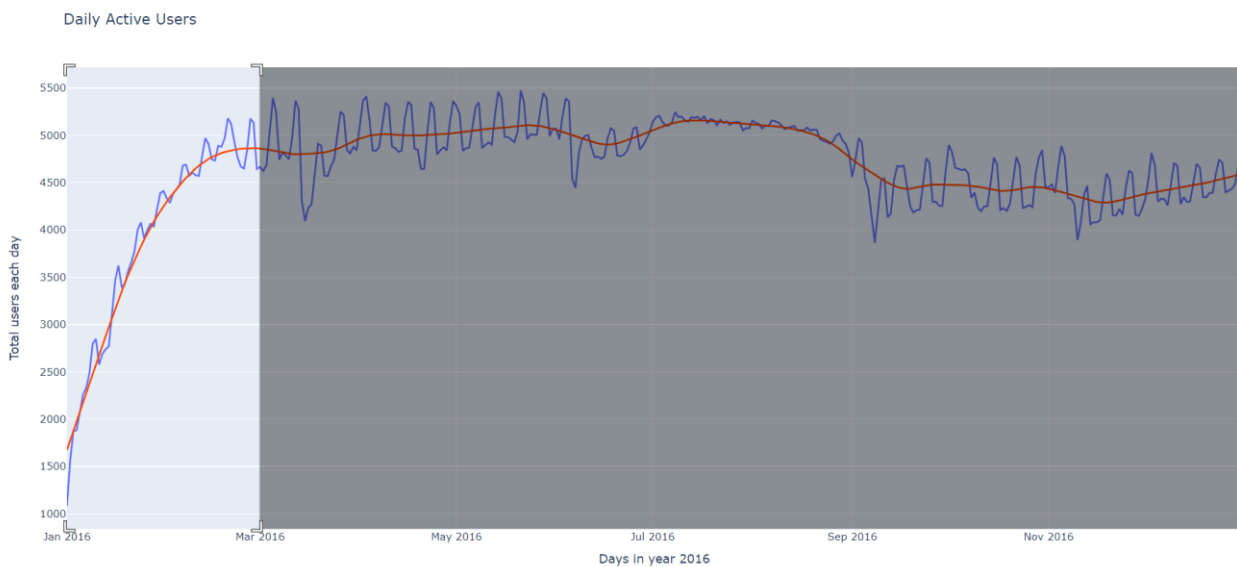


FIGURE 3: First quarter uptrend

3.3 Time between July and August

After second dip (probably caused by update, see chapter 3.1) the DAU goes slowly back to normal but visibly with decreasing deviations between weekends and weekdays until it reaches July, where the deviations almost match the red smoothed trendline.

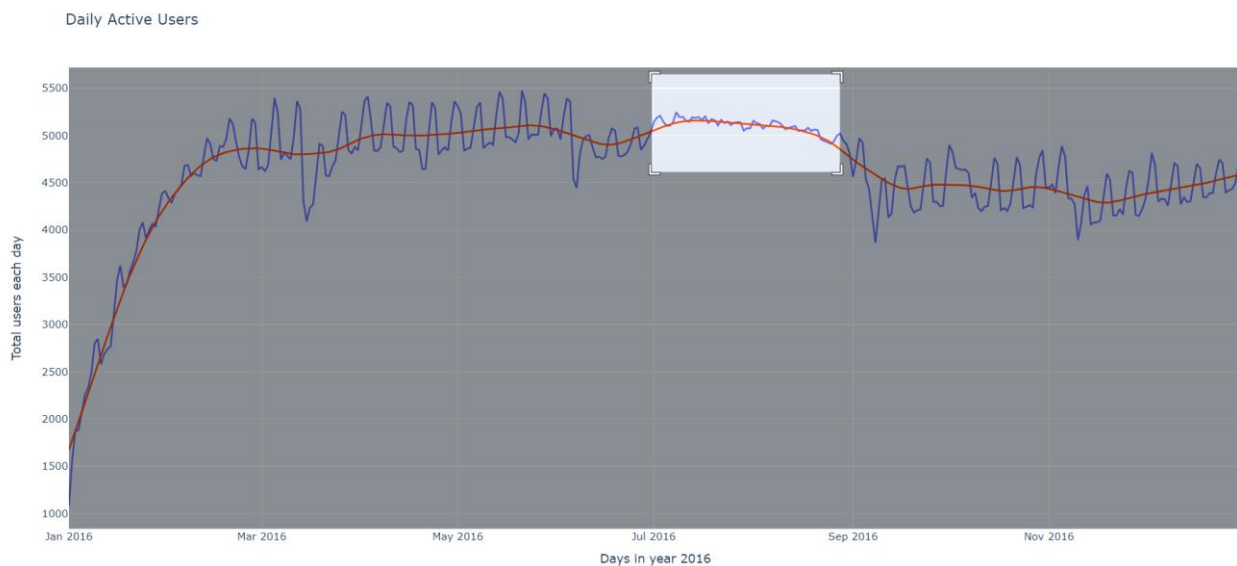


FIGURE 4: Summertime DAU

The most probable target audience playing mobile games are young people, mostly still attending school. This would explain the fact that during July and August, when schools have holiday, the daily app usage deviations noticeably settle down.

3.4 Anomaly between 3rd and 9th October

There is observable behaviour which does not seem normal between 3rd and 7th October where even during the weekdays there is unusually high number of users accessing the app and then during the 8th and 9th October unexpected dip.

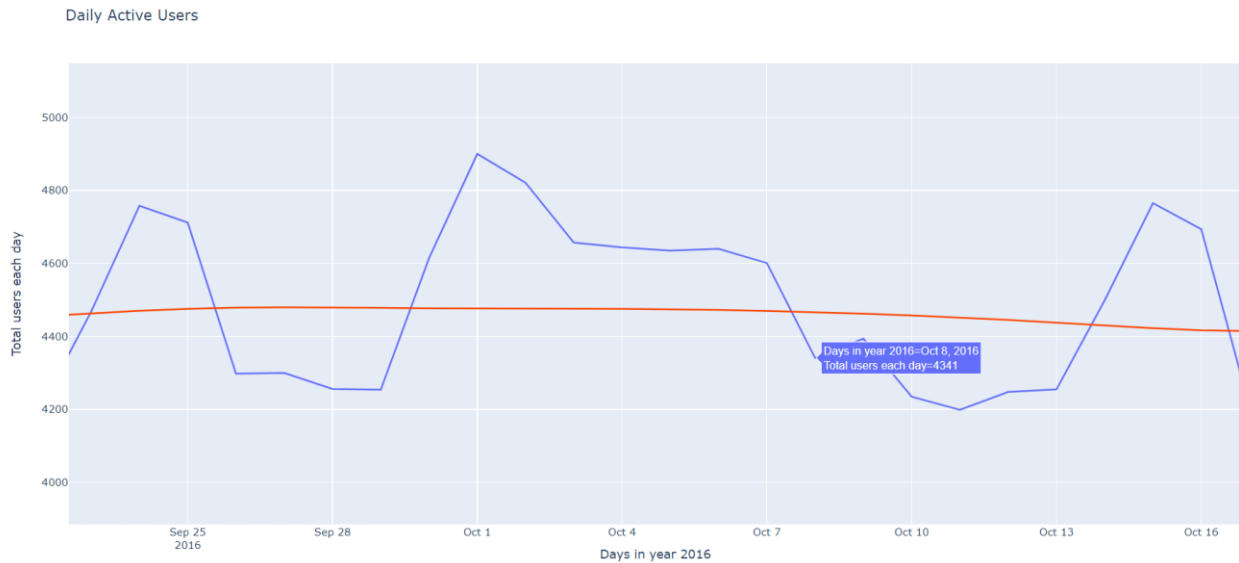


FIGURE 5: Time segment between 23rd September and 17th October

After examining the proportioning of countries in overall users (chapter 4.1) it was possible to find out that almost 44% of all users are either from USA or China. Looking at historical events happening during that week, helped finding the reason for this behaviour. China has during the week between 1st and 7th October national holidays because of celebration called “Golden Week”. [3]

1 Oct	Saturday	National Day	National holiday
2 Oct	Sunday	National Day Golden Week holiday	National holiday
3 Oct	Monday	National Day Golden Week holiday	National holiday
4 Oct	Tuesday	National Day Golden Week holiday	Common local holiday
5 Oct	Wednesday	National Day Golden Week holiday	Common local holiday
6 Oct	Thursday	National Day Golden Week holiday	Common local holiday
7 Oct	Friday	National Day Golden Week holiday	Common local holiday
8 Oct	Saturday	Special Working Day	Working day on weekend
9 Oct	Sunday	Special Working Day	Working day on weekend

TABLE 1: Table of national holidays in China in 2016 [4]

That means that during the week the largest audience by country has much more time to spend playing the game. It is recommended to take combination of these facts in consideration. Creating a special in game event with unique in app purchases might increase the revenue and active usage.

TABLE 1 shows that when the national holiday week in China ends, it is followed by two workdays, where the people in China must work, even though it is a weekend. That has the opposite effect, and the DAU will decrease because of it.

3.5 Halloween

Last visible anomaly is after 30th of October when the celebration of Halloween happens. During the following weekdays the daily usage is above the average weekday's daily usage.

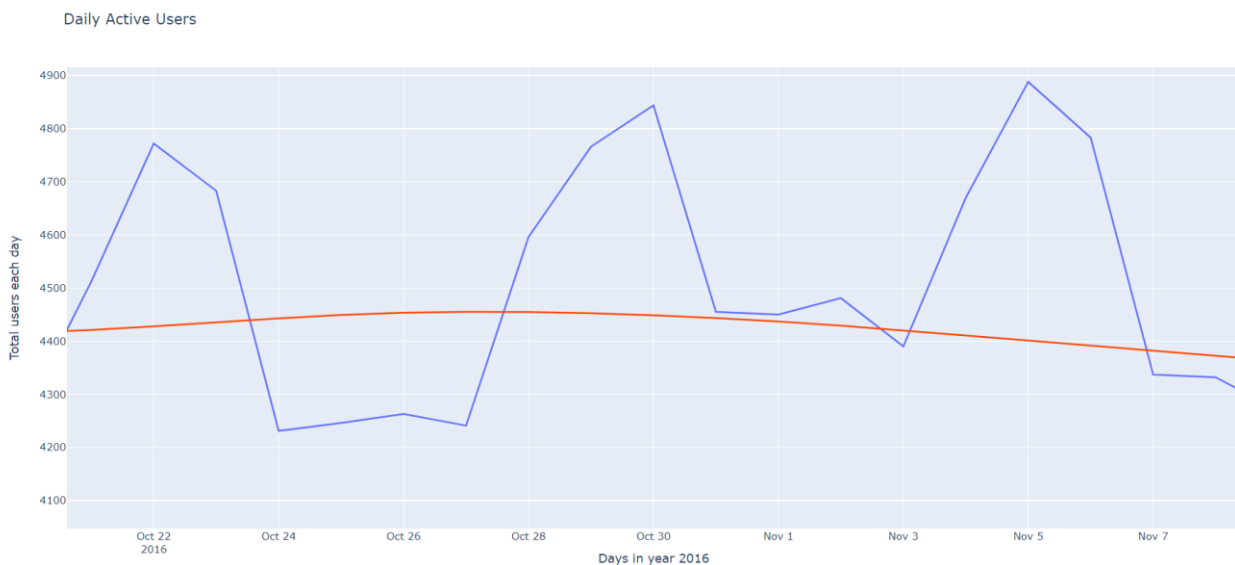


FIGURE 6: Halloween time

There most probably are Halloween events in the game during this week which might bring more users to the game, but this only an assumption which would need to be verified.

4 SALES

Based on requirements of the task, several visualisations were created in order to make observations and conclusions about total users by country, revenue by country and average revenue by user by country.

4.1 Total users by country

Measuring total users by country enabled in depth analysis of data anomaly in DAU (see chapter 3.3). This measure consists of sum of accounts created in each country.

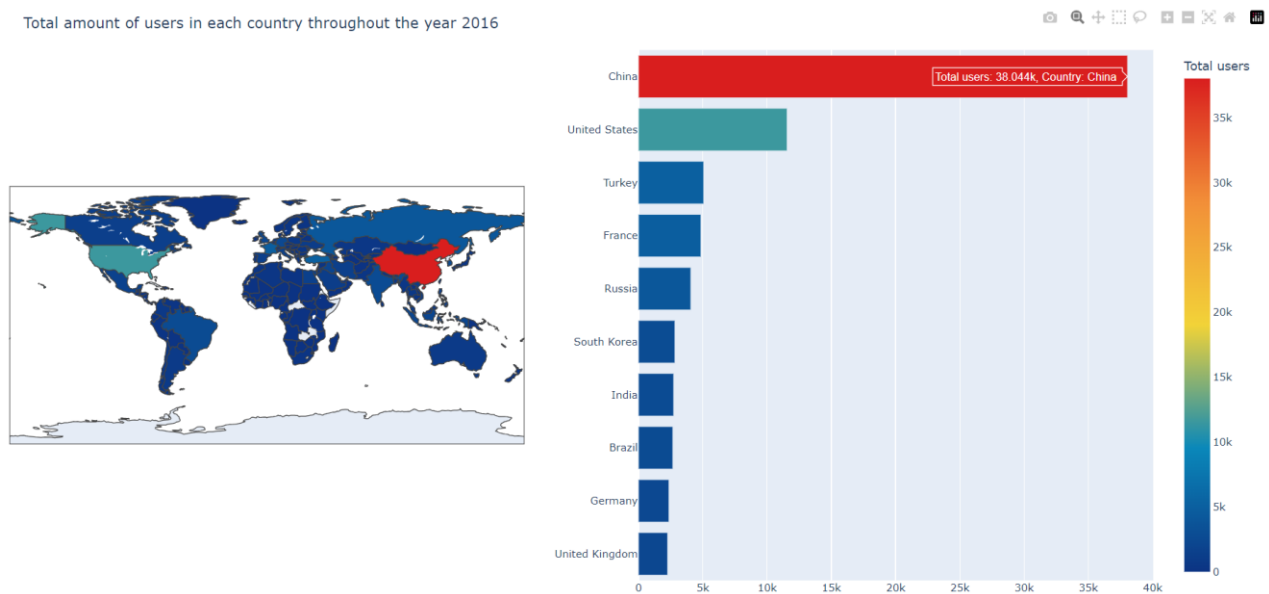


FIGURE 7: Geographic split of users

Based on data given by sub report above, China and United States of America are the countries with the largest audience. Outside of France, Europe is not as productive in numbers as it might have been expected. Compared to China, population of India was less ‘only’ by 100 million [5] yet number of users around half of the users of Turkey or France which are far not as large in population. It might be caused by lack of access to electricity, internet, and electronic devices, but it also could be unused potential.

4.2 Total revenue by country

Measuring total revenue by country shows which market is the most valuable in terms of revenue from in app purchases.

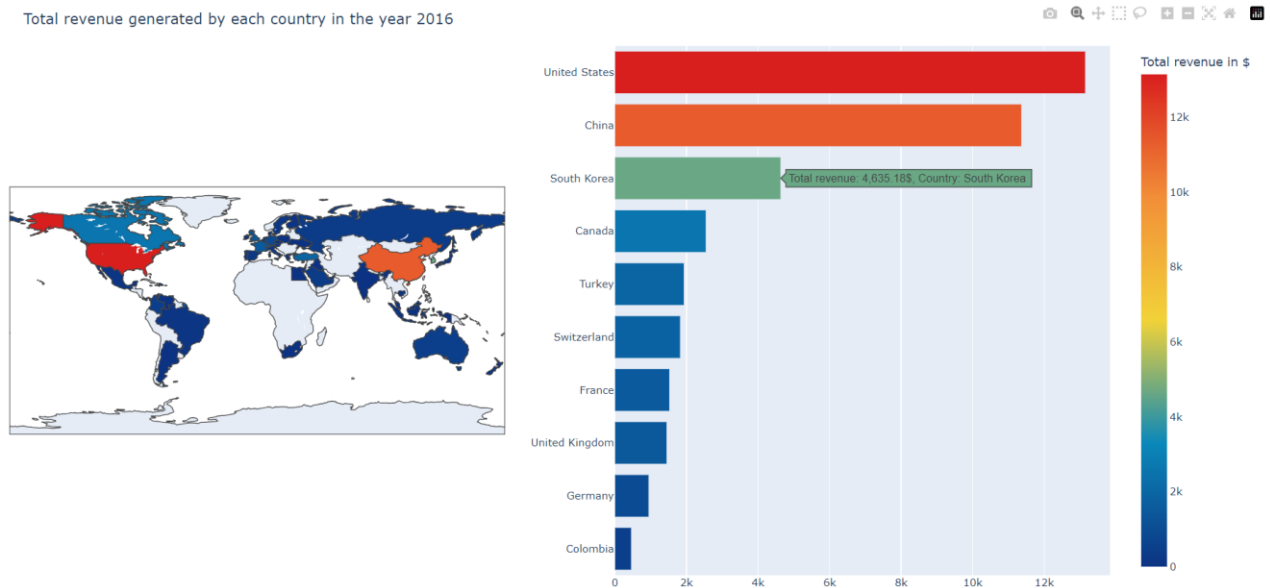


FIGURE 8: Total revenue by country

From figure above it is observable that even though China has the most users by far (see chapter 4.1), users from USA bring much more value in revenue. Surprisingly high revenue was brought by South Korea even though in FIGURE 7 we can see that South Korea has a little less than 3000 users.

4.3 Average revenue per user per country

In this measurement the value returned is the average revenue created by users of a specific country. It differs from average revenue per country in breaking down the average by distinct users.

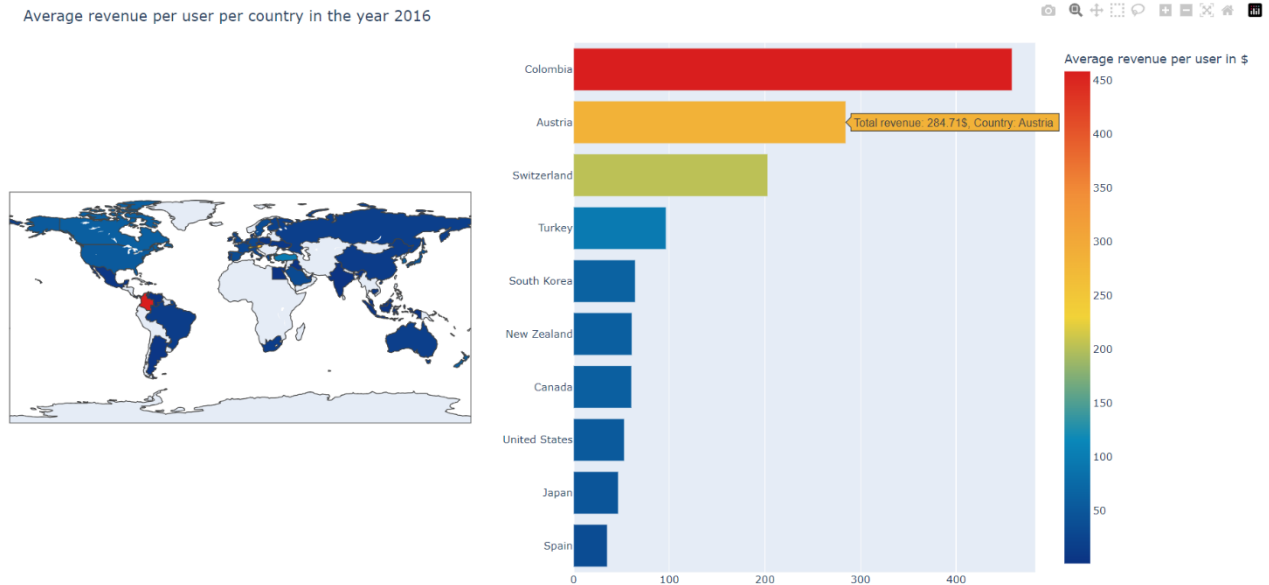


FIGURE 9: Average revenue per user per country

Surprisingly Colombia is having the biggest revenue per user which seems suspicious since it did not have more than 500\$ in total revenue according to FIGURE 8. That is already pointing at the problem, which is countries, where one user made all the revenue by himself, that causes Colombia to have the highest average revenue per user. The following figure shows the same data with filter set to countries where at least two different users made purchases.

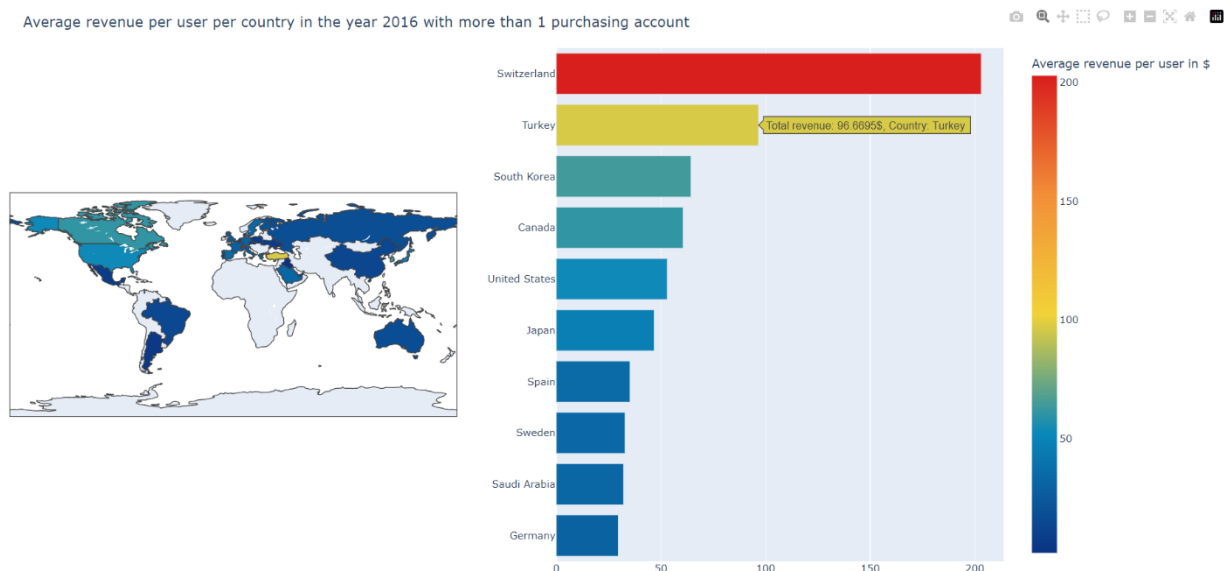


FIGURE 10: Average revenue per user per country with at least two different users made purchases.

From FIGURE 10 it is visible that Colombia and Austria are not in the bar chart on the right side due to filtering only countries with at least two different users made purchases. Based on this figure it is possible to conclude that Switzerland's users are the ones willing to pay the most in average for in app purchases. Another interesting observation is that users from Canada and Japan are also paying more in average.

5 CONCLUSION

At the beginning, the report is describing the relationship using `account_id` column between the three tables that were the datasets for this analysis. It describes the meaning of each table as well as meanings of each column and its data.

Secondly it went through Daily Active Users analysis, showing the chart over time analysing trends and unusual ups and downs throughout the year 2016. It was concluded that the first quarter the DAU went up due to launch of the app, later on during summer the usage got more stable because of students having holidays and one week in China is dedicated to holidays finishing with two days of work on the weekend which causes anomaly in October.

Lastly the report describes and visualises sales data in geographical manner. First it went through total amount of users which is the highest in China and second highest in USA. Second subset of plots was showing total revenue by country where USA is surpassing China and interestingly South Korea made the 3rd biggest revenue in the world. Last visualisations were average revenue per user per country which split into two visuals, one for all the countries with Colombia having the most and second with filter only for countries with more than one purchasing user where Switzerland proved to have the highest average per user.

Other interesting plots could be total users by platform in time, geographic split of number of paying users compared with free-to-play users, different dimensions of revenue in time.

REFERENCES

- [1] Supercell, Localization Data Intern Test. Available at: <https://www.dropbox.com/scl/fi/kcixwvtjg3ydtpylypdbe/Localization-Data-Intern-Test.pdf?dl=0&rlkey=uifet3rmmcefyfmzd5p9ndt9g>
- [2] GitHub repository, Samuel Vitek. Available at: https://github.com/SamuelVitek/mobile_game_analysis
- [3] Wikipedia contributors. Golden Week (China). Wikipedia, The Free Encyclopedia. October 4, 2023, 12:55 UTC. Available at: [https://en.wikipedia.org/w/index.php?title=Golden_Week_\(China\)&oldid=1178561819](https://en.wikipedia.org/w/index.php?title=Golden_Week_(China)&oldid=1178561819). Accessed November 30, 2023.
- [4] Time and Date AS, Holidays and Observances in China in 2016, Available at: <https://www.timeanddate.com/holidays/china/2016#tr49>
- [5] Statistics Times, India vs China by population, Available at: <https://population.un.org/wpp/Download/Standard/Population/>