Analytics Home Assignment

Table of Contents

Assignment Instructions

This assignment uses Data Analysis to familiarize with the dataset provided in tasks_used_da, which contains the tasks performed by a given user at any given date for the timeframe provided from January 2017 to June 1, 2017). The data will be analyzed in a series of questions with reproducible queries and visualizations.

NOTE: Any sort of source data wont be included in this Git repository and resultant archive as requested.

Tools

For this assignment, we had the opportunity to choose the tools that would be necessary to complete the analysis. The following list will have the main reason of why the tools chosen were the most appropriate for the work requested:

- **SQL**: While one of the guidelines was specific about the time to be spent to 3 hours or less, decided to focus on SQL since it is the main skill I'm comfortable to answer the questions inside the timeframe given.

- **TablePlus**: There's many options to function as a versatile UI to connect to any database type. But with TablePlus connections can be easily sorted and label but most importantly when working with big queries this tool is able to make it easier by toggling lines and comments. Plus it includes a query history so that you can easily replay past queries or saved them for future reference.

- **Tableau Public**: We considered using Google Data Studio but encountered a technical limitation when the csv file is over the limit of 100 MB. With this limitation we considered the second optio which is Tableau, a very effective visualization tool that allows the creation of quick complex graphs handling more data and calculations on datasets.

Data Analysis

- Which are the top 10 accounts ids by number of users?
- Create a summary table at the account level that signals when an account is new (boolean). An account is new for the first day we see it run a task(s). The table should look similar to the one below in structure.
- Create a summary table at the account level. Add a column with the percentage difference in the number of tasks to the previous day.
- Add another column with the moving average of the tasks run in the last 7 days for each account.

- A lost account is an account with no tasks run on a given month. How many accounts did we lose (had no executed tasks) in February 2017?

NOTE: The queries for the data analysis can be found in the Queries folder in the file analysis_queries.sql

Visualization

(Optional) Create a visualization that represents the growth of new accounts in a way you would communicate to a peer or business stakeholder.

- We can start analyzing the data by looking how many accounts Planoly had active during Q1 and Q2 of 2017. Looking at the trend of Q1 we can say that if we had the rest of the data source for 2017 (we only have information until June 1, 2017), the number of accounts on Q2 would most likely stay short of surpassing the total of Q1. This could be caused by various situations or reasons, for example: deactivated accounts, social media platforms (Instagram/Pinterest) downtime or technical issues.

Number of Accounts (Quarter Review)

	Accounts
Q1	5,428,989
Q2	4,074,926
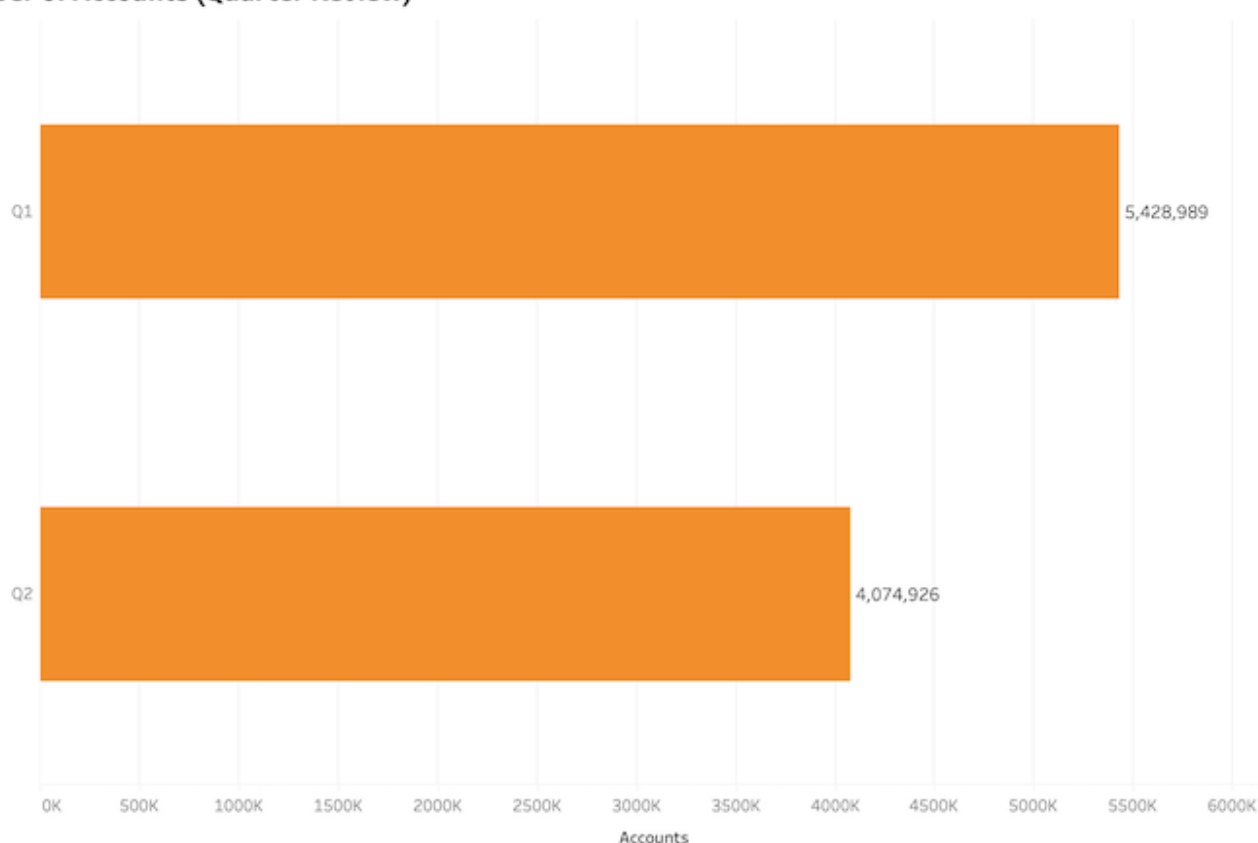
- Now we want to see how Planoly is doing on a monthly basis gaining new accounts. To consider a new account, we have to keep in mind that it would be the first date it had activity on the platform. The first thing to notice is January was the month with the most new accounts and a consistency of gaining an average of 3,603 through February and May. To understand how January had so many new accounts, we would have to get more data and analyze to identify what patterns were most used by the users. Consider:
  - What was the most used service from our platform?
  - Where there any special events during this month that would be important for our users? Besides well known days for sales or announcements.
  - How can we help them to explore other services offered in our platform?

Monthly Growth for New Accounts
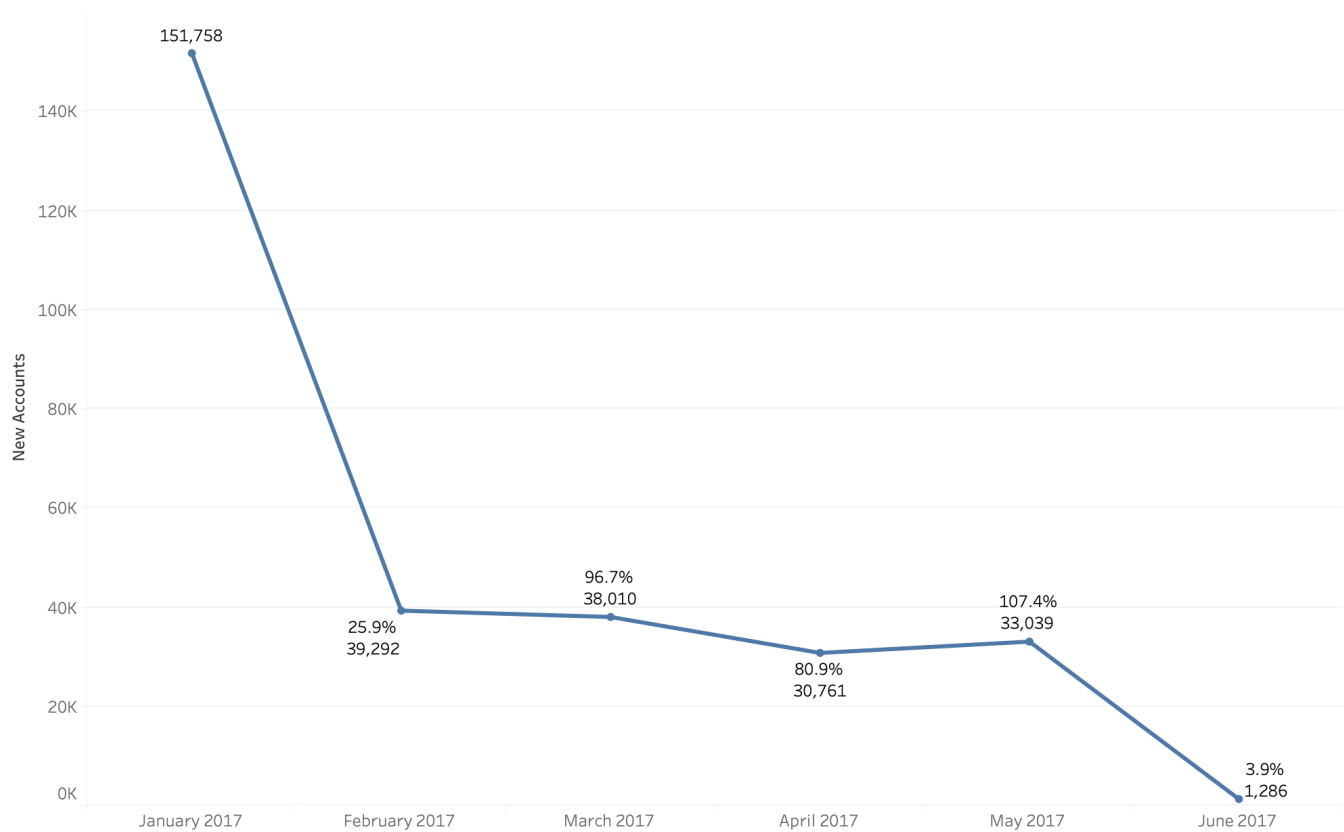
- For an extra analysis, we decided to do a visualization of the amount of tasks done every month. Were we can pretty quickly identify that the months of March and May had the most activity. These could be for several reasons considering the most of the users are businesses. For the month of March we have the beginning of Spring which is usually a perfect time for engagement and publishing new content for sales or new collections. Then for the month of May we have 'Cinco de Mayo' and Mother's Day which are also perfect days for businesses to grow engagement and announce sales or any type of special events.

Monthly Tasks