

Google Playstore Apps Market Analysis

Jei Park

Identifying the worldwide trends through the analysis of Google Playstore apps

Introduction

We are currently living in a world where almost everyone owns a smartphone. People share their lives with friends and family through social media apps such as Instagram and Snapchat, post an important event on Facebook, send quick emails with Gmail, take online classes and learn new skills with Coursera and Duolingo. Since the COVID-19 Pandemic, the online education market has been booming as more people are finding ways to learn skills and materials at home.

In this project, we would like to mainly focus on education apps and social media apps. We will use four data sets, Google Playstore apps, Coursera courses, Facebook daily users, and Facebook stock collected from Kaggle, to answer the questions below.

Research Questions

First, what does the overall education apps market look like and what can we observe from the top popular education apps? Is there a relationship between rating and categories of apps, and can we observe some relationship for the education apps? We will specifically take a look into the top popular educational apps to see what makes them popular and if there is a specific trend in the education industry.

Second, can we observe the growing interest in STEM-based courses? We would like to take a closer look at the skills category from the Coursera data set. We'll examine the most offered skills in Coursera. Why are those courses offered the most and what makes those courses popular these days?

Finally, what does the overall social media apps market look like? Is there a relationship between the number of downloads and the rating of an app, and if there is an outlier, what has caused the popular app to receive such a low rating? Can we see this kind of downward trend in their stock as well?

Section 0: Dataset Descriptions

Google Playstore apps dataset Description

The dataset contains information collected in June 2021 about various Android apps in Google Play Store. The dataset has 24 columns and 2312944 rows. It includes many interesting variables, such as category, rating, release date, number of installations of apps, as well as whether advertisements are supported inside apps. Using this dataset, we hope to explore more about app user preferences and gain insights from popular apps.

Specifically, we are interested in the variables **Category**, **Rating**, **Installs**, **Price**, **Size**, and **Released**. **Category** is a variable of type character that describes the category of an app, such as “Education” or “Music”. **Rating** is a variable of type integer that represents the average rating of an app. **Installs** counts the number of installations of an app. **Size** represents the size of an app.

Specifically, we are interested in the variables **Category**, **Rating**, **Installs**, **Price**, **Size**, and **Released**. **Category** is a variable of type character that describes the category of an app, such as “Education” or “Music”. **Rating** is a variable of type integer that represents the average rating of an app. **Installs** counts the number of installations of an app. **Size** represents the size of an app.

More importantly, we are interested in the variable **Released**, the app released date of an app. We would like to know which type of app has been released the most, and further investigate if there are specific trends or events that have encouraged more developers to develop that type of app.

App Name - name of the app

App Id - package name

Category - app category

Rating - average rating

Rating Count - number of rating

Installs - approximate install count **Minimum Installs** - approximate minimum app install count

Maximum Installs - approximate maximum app install count

Free - whether app is Free or Paid

Price - app price

Currency - app currency **Size** - size of the application package

Minimum Android - Minimum android version supported

Developer Id - Developer Id in Google Playstore

Developer Website - website of the developer

Developer Email - email-id of developer

Released - app launch date on google Playstore

Privacy Policy - privacy policy from the developer

Last Updated - last app update date

Content Rating - maturity level of app

Ad Supported - ad support in app

In app purchases - in-app purchases in app

Editor Choice - whether rated as editors choice

Coursera Dataset Description

The Coursera dataset is collected from Coursera from September 2021. The number of users of Coursera has increased since the COVID-19 pandemic since more people are taking classes online. We are interested in the category Skills, to identify what skills are offered the most from the Coursera classes. We are hoping to see if there are specific types of skills that are offered the most and further investigate why those skills are popular.

Course Name - the Name of the Course

University - the University or Industry Partner that offers the Course

Difficulty level - the Difficulty Level of the Course. Has 3 values: Beginner, Intermediate, Advanced. Also has Missing Values

Course Rating - rating on a 5-point scale with minimum step value 0.1. Missing Values represented by Not Calibrated

Course URL - the Course URL

Course Description - the Description of the Course. Has Missing Values

Skills - the Skill Tags associated with the Course extracted through NLP processes

Facebook Stock Dataset Description

The Facebook stock price dataset contains the daily Facebook Stock price from 2012 to 2021. It has a total of 8 variables: **Date**, **Open**, **High**, **Low**, **Close**, **Volume**, **Dividends**, **Stock Splits**. In this report, we used only two of these variables **Date** and **Close**, the date on which the stock price was recorded and the stock's closing price. Through visualizing and analyzing these variables, we are able to find a drop in the stock prices from September to October 2021 and relate to it the drop in the number of active users.

Date - date **Open** - opening price

High - high price

Low - low price

Close - closing price

Volume - volume traded

Dividends - dividends

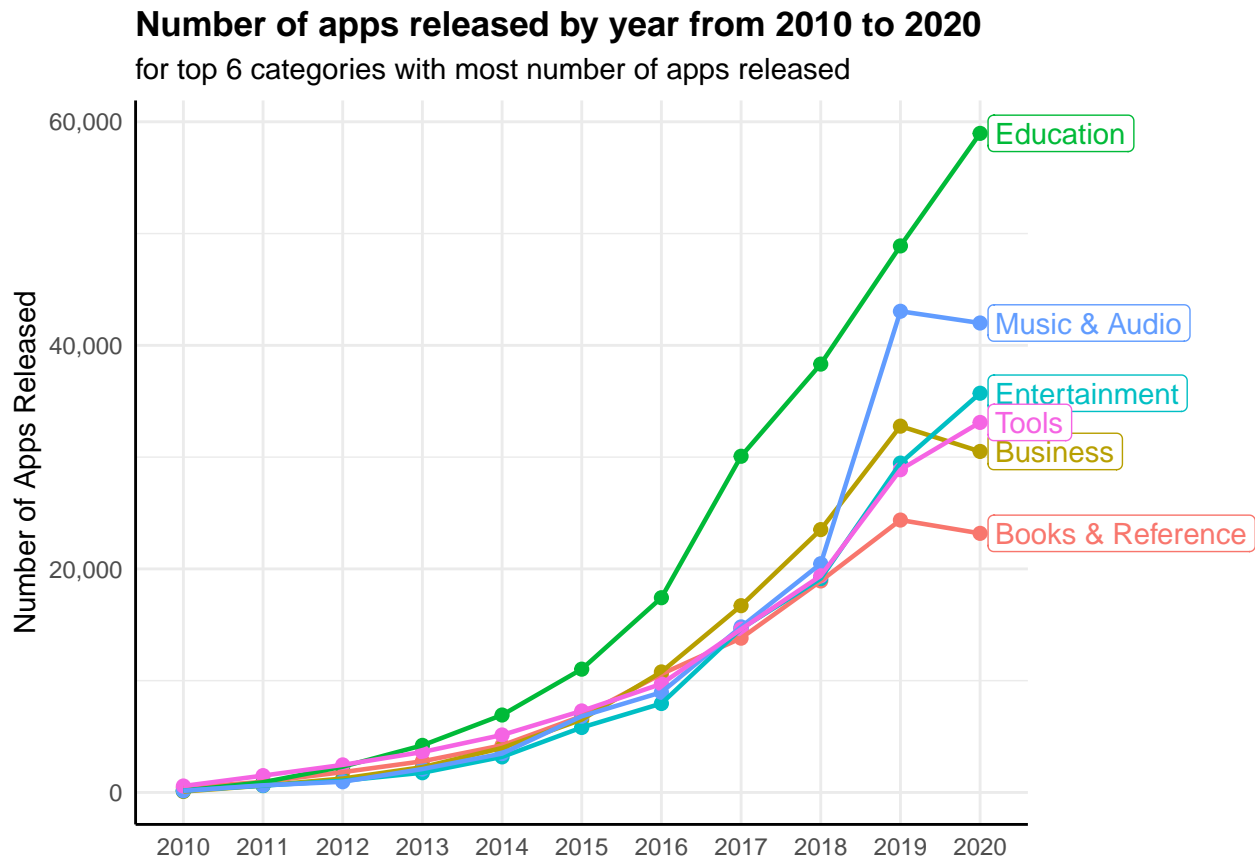
Stock Splits - stock splits

Facebook Daily Users Dataset Description

The Facebook daily users dataset records the number of daily active Facebook users worldwide from 2011 to 2022. It has 2 variables, **Time** and **Count**. **Time** represents quarter and year, and **Count** is the number of daily active users. Using this dataset, we hope to investigate the trends in daily active users and further connect these trends to other data on Facebook.

Section 1: Education Apps

1.1 What are the categories with the most number of apps released?

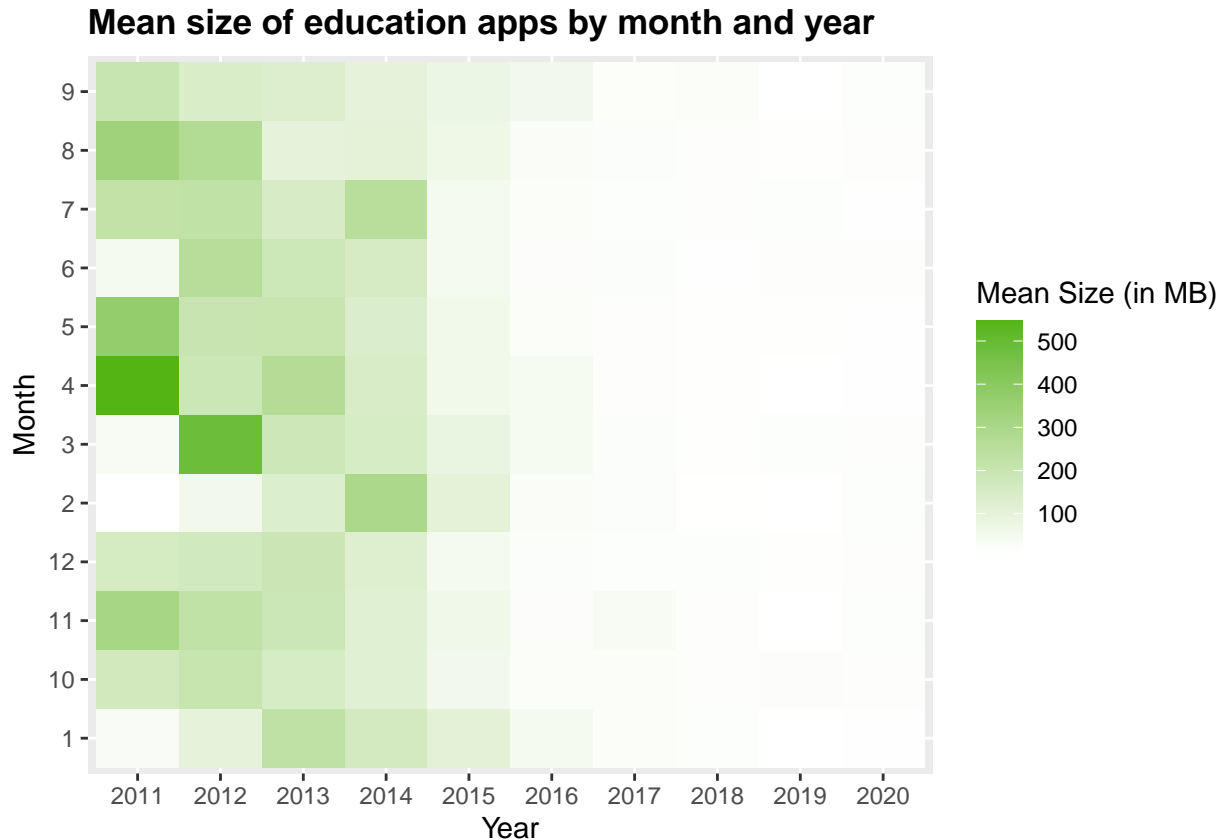


The line graph above shows the number of apps released each year from 2010 to 2020, for the top 6 categories with the most number of apps released. The graph indicates that the Education category has the most number of apps released, and the number has been steadily increasing over the years.

According to Technavio, a leading global technology research and advisory company, their research and analysis focused on the emerging market trends shows that from 2021 to 2026, due to the increasing awareness of education apps, growing internet, and increased use of smartphones and emphasis on technology by educational institutions, government initiatives will assist the growth in the education apps market. The potential growth of education apps from 2021 to 2026 is USD 70.55 billion.

The increased market demand for education apps is because of the growing interest in STEM-based apps. In the United States, there is an increasing demand for STEM jobs, so app developers are investing in creating apps that can help students learn coding, astronomy, space science, and more.

1.2 How has education apps changed in size as time passes?



The heat map shows the mean size of education apps by month and year from 2011 to 2020. From this heat map, we can observe that the mean size of education apps has declined over the years and fallen below 100 MB. This indicates a great improvement in education apps: app developers gradually have better optimization methods to reduce the file sizes. Moreover, this decrease in average size means that education apps now have a better chance to obtain and retain users, because apps with a small file size have more acceptance among the users. According to a blog from Farfetch, compare to large mobile apps, they take less time to download and generate lower user expectations that are easier to meet.

1.3 Chi-Square Independence Test: is there a relationship between Rating and Category

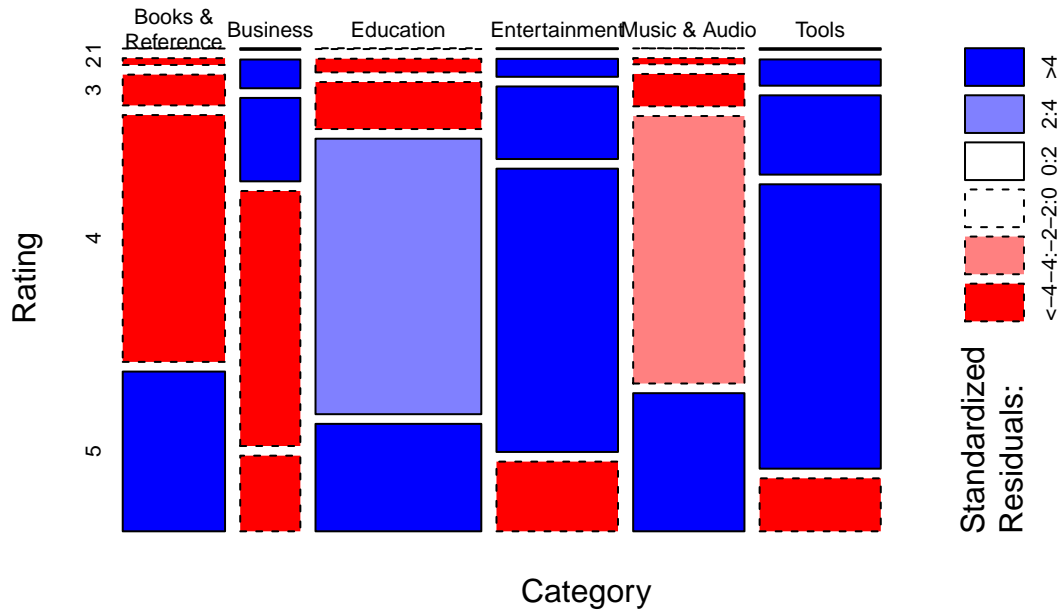
In this section, we will conduct a Chi-square test of independence to check if there is a relationship between rating and category of apps. Our null hypothesis is that rating and category are independent, and our alternative hypothesis is that the two variables are not independent.

```
##  
## Pearson's Chi-squared test  
##  
## data:  table2  
## X-squared = 23189, df = 20, p-value < 2.2e-16
```

From the Chi-square test shown above, we have an extremely low p-value, less than $2.2e-16$. This gives us

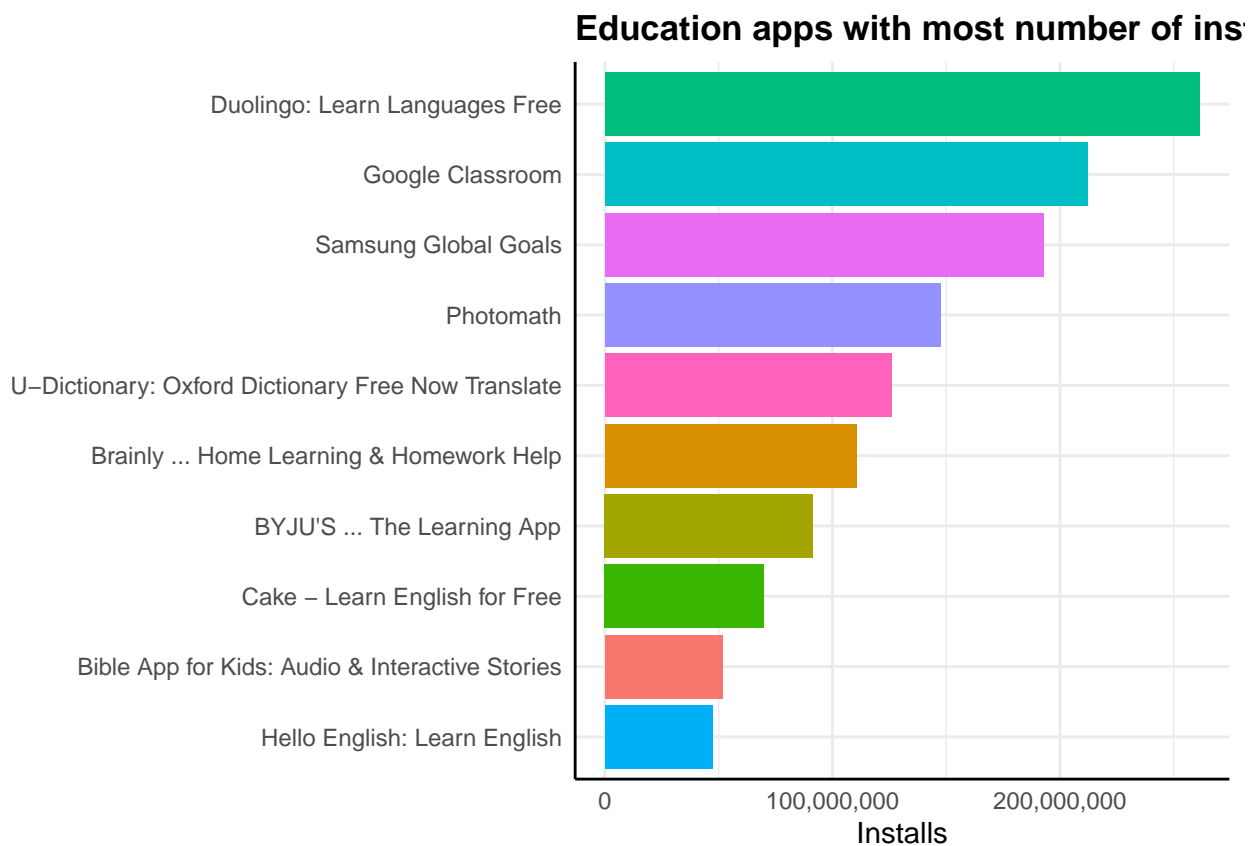
sufficient evidence to reject the null hypothesis and conclude that there is a relationship between rating and category.

Chi-square test of rating and category



Furthermore, we can have a deeper understanding of how rating and category are related by examining the mosaic plot above. This mosaic plot shows the conditional distributions of the ratings given a category, colored by standardized residuals, for the top 6 categories with the most number of apps released. A deeper blue color in a cell indicates a higher standardized residual between the corresponding row and column variables, while a deeper red color indicates a lower standard residual. From the plot, we notice a high standardized residual between 4-star ratings and the education category and an even higher one between 5-star ratings and the education category. This means that there is a strong positive correlation between the education category and 4 and 5-star ratings. Thus, we could conclude that among the top 6 app categories with the most number of apps released, education apps generally have higher ratings.

1.4 What are the most downloaded Education apps? Why are they popular?



This is a bar chart showing the name and number of installations for the top 10 education apps with the highest number of installations. From the chart, it is clear that Duolingo has the most number of installations among all education apps, followed by Google Classroom and Samsung Global Goals. We will now dive deeper into the reason behind the popularity of these apps to understand more about mobile app user preferences in the world.

Top 1 Education app: Duolingo

Duolingo is the world's most downloaded education app that offers more than 40 language lessons for free. According to the 2021 Duolingo Language Report, despite the COVID-19 Pandemic and reduced traveling, there was a record growth of new Duolingo Users who started studying a new language. From 2021, many new users joined Duolingo to learn Japanese and Korean, due to the increasing popularity of Asian cultures and K-culture. The growth of interest in the Korean language has surged in 2021 because of love for Korean entertainment and K-pop worldwide. Following the launch of one of the most popular Netflix series "Squid Game" in September 2021, the Duolingo report showed a surge in new learners in the United States studying Korean on Duolingo.

Top 2 Education app: Google Classroom

Google Classroom is the second most downloaded app in the education apps category. Google Classroom mobile app was introduced in January 2015 and became one of the most popular free hybrid learning

platforms. Over the past 2 years during the COVID-19 Pandemic, the education community quickly adopted the use of online classroom platforms, and today, more than 150 million students, teachers, and educators use Google Classroom for teaching and learning. 40 million new users joined the Google Classroom from 2020 to 2021, according to the Google Classroom Program's article. The increasing use of paperless assignments, quizzes, exams, and more will attract more Google Classroom users.

Top 3 Education app: Samsung Global Goals

Samsung Global Goals app is the top three most installed apps in the Google Playstore (This app is only available in the Google Play Store).

According to an article from the National Library of Medicine, due to the increase in demand for online shopping, the organic waste and inorganic waste have increased as well. The consumer behavior led many companies to produce more environmentally friendly and sustainable products.

In a recent survey by BCG, more than 70 % of the survey participants said that since the pandemic, people are more aware of environmental awareness, sustainable consumption, and social responsibility since the pandemic. Younger people (age 25-34) are more likely to believe their action of advocating sustainability will bring a positive environmental impact.

Section 2: Coursera

What are the most popular skills in Coursera?

In order to better understand the type of courses that are popular in online education, we narrow our focus to a specific app, Coursera. In this part of the report, we will explore the most common skills that are taught in Coursera. Below are two word clouds visualizing the most common words and bigrams in the Coursera skills description.



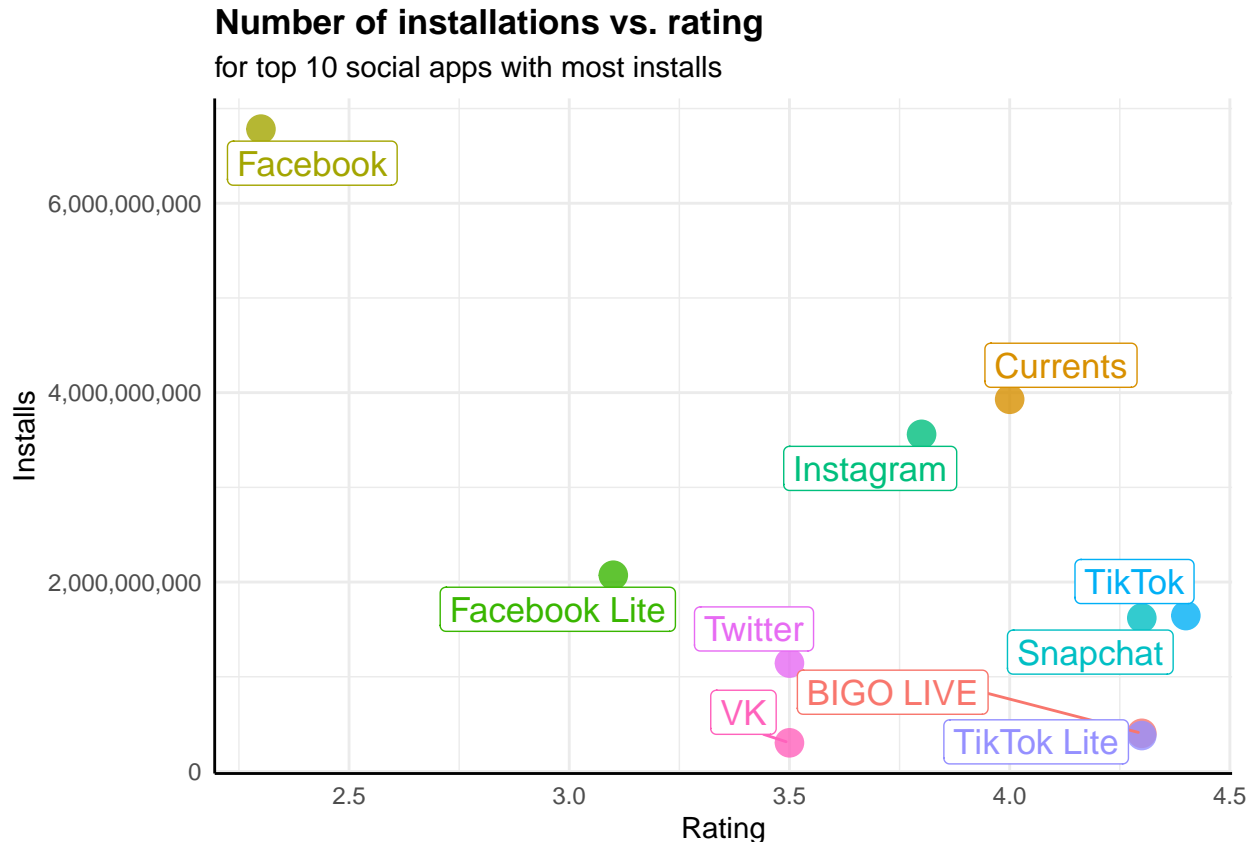


The first word cloud pictures the top 20 most common single words in the skills description of Coursera: science, data, management, business, analysis, computer, learning, development, programming, engineering, etc. Most of the words are directly related to STEM courses. The second word cloud illustrates the top 10 most common bigrams in the skills description of Coursera: computer science, data science, machine learning, etc. Most importantly, we could see that all of them are related to STEM courses. From these two graphs, we observe that the education app and website Coursera is offering more STEM classes. This aligns with the previous statements that the interest in STEM is increasing and that more app developers are invested in developing STEM education apps and courses.

Section 3: Social Media Apps

In this section, we shift our focus to social media mobile applications. First, we investigate what are the most popular social media apps.

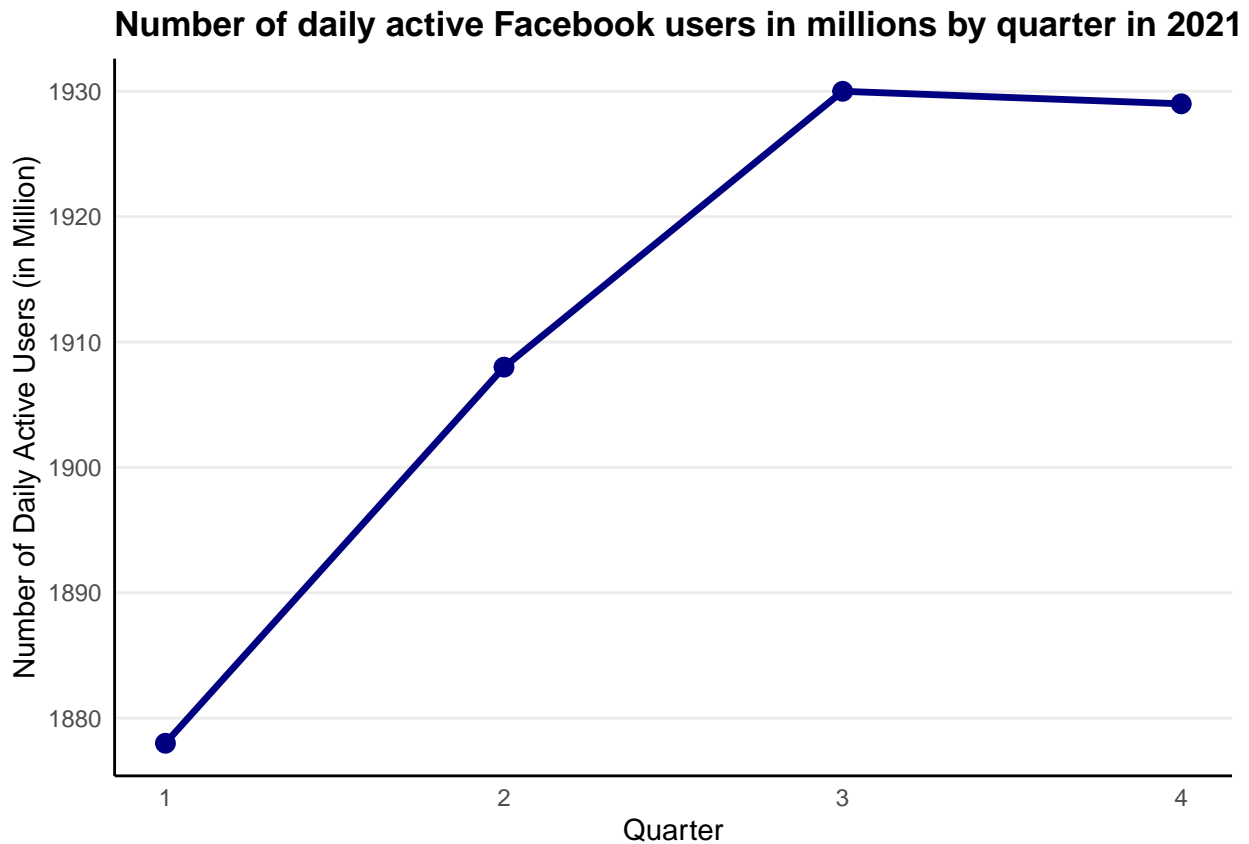
3.1 What are the most popular social media apps?



This scatter plot shows the average rating and number of installations for the top 10 social media apps with the highest number of installs. From the plot, we could observe that the point that represents Facebook is in the top left corner by itself, while all other apps seem to be more closely distributed along the bottom right corner. This indicates that even though Facebook has the highest number of installs, it has a particularly low rating among the most popular social media apps. To investigate the reason behind this peculiarity, we will focus our investigation on Facebook.

We found that the reason behind this low rating is a protest against Facebook's restriction and deletion of pro-Palestinian content and accounts. In May 2021, a month before the collection of this dataset, the Israeli-Palestinian conflict was escalating. Israeli warplanes have launched airstrikes on Gaza, killing more than 200 Palestinians and wounding thousands. According to Business Insider, during this period, many Facebook users found out that their pro-Palestine posts were removed, restricted, or flagged as harmful or dangerous by Facebook, thus feeling that they were being censored and silenced. As a result, they launched a campaign that calls people to give Facebook one-star ratings in the Google Play Store. This campaign was successful: the Facebook app experienced a massive drop in rating from over 4 to 2.4 in the Google Play store by May 22. The rating further dropped to 2.3 in June, which could be found in this dataset.

3.2 How has the number of daily active Facebook users changed in 2021?



The line graph above shows the Facebook daily active users (in millions) in 2021 by quarters. From the plot, we can see that the daily users were increasing during the first quarter and second quarter of 2021. However, from the third quarter of 2021 to the fourth quarter of 2021, the number of Facebook daily users suddenly dropped by almost 1 million. The number of daily active users is a key growth metric for social media apps: this drop in daily active users indicates a decline in the popularity and usage of Facebook, so what effect does it have on Facebook?

3.3 How has Facebook stock price changed in 2021?

We will now proceed to look at the stock price of 2021 to further understand if the drop in the number of daily users of Facebook has an influence on the stock price of Facebook.



Here we specifically identify the stock price of Facebook in 2021 by month. We are able to identify a similar pattern as the previous Facebook daily users line graph. From this graph, we can clearly see that from September to October 2021, the Facebook stock price dropped by almost 40 USD. According to an article from CNBC, the huge drop in the number of daily active users shown in the previous plot contributed to the drop in the stock prices.

Conclusion

Through our analysis of the popular education apps, Coursera's most offered skills, the relationship between the number of downloads of the social media apps and rating, Facebook stock price change over time, and Facebook daily users over time, we present the following conclusions:

- We see a growth in the number of education apps being released and greater interest in learning STEM-related education and skills.
- We observe that the mean size of education apps has declined over the years, which indicates a great improvement in education apps.
- We find a strong positive correlation between the education category and 4 and 5-star ratings. Among the top categories with the most number of apps released, education apps generally have higher ratings.
- People are becoming more interested in learning about different cultures and languages.
- Since the COVID-19 pandemic, more people are advocating for sustainability and have a greater interest in addressing environmental issues.
- The drop in Facebook rating, stock, and daily users shows that big social media companies should protect the user's freedom of thought on politics and social movements around the world.

Future Research

After the in-depth research on education apps and social media apps, we would like to ask the following questions in our future research: with the growing use of online education, how does this affect the learning experience for students? What are some ways to provide better online education learning experiences for students? What does the job market for STEM majors look like? Are there particular skills that companies are most looking for when hiring? How does the eco-friendly business market look overall? Would TikTok take over facebook's throne in social media apps? Would there be a positive and more diverse job growth for social app companies? Social media apps are now offering the option for shoppers to directly purchase products on social media, and provide more ways for companies to engage with customers. Most advertisements are made through social media nowadays and companies are noticing the importance of social media for promoting and branding their companies. In order to answer these questions, we have to do more research and data analysis on education apps, education and skills trends, and social media apps and consumer behavior. For long-term research on the effects of online education, we will need to perform an experiment on students who took online classes for most of their classes and compare them with students who took most in person. We would also need to do a survey on the consumer's interaction with advertisements and content to see what kind of content and type of advertisements effectively attracts users to purchase the products.