

Analyzing Play Store app data to provide developers with insights to conquer the Android Market.

CHIKLI Hugo

HELBURG Aline

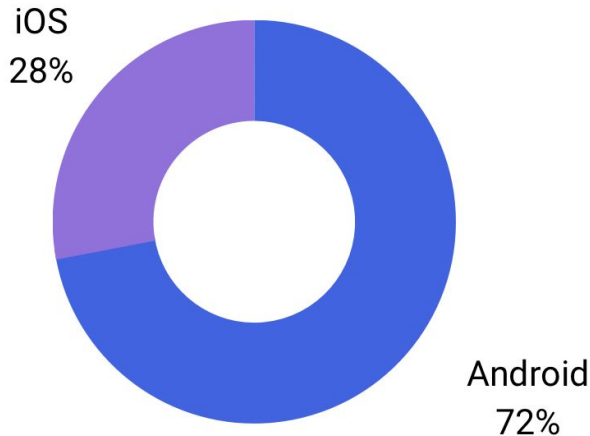
MICHEL Pauline

RIGUIDEL Emma



10,000+ app data to analyze Android market and users' segmentation.

Our dataset consists of a list of **more than 10,000 apps available on Google Play Store**. Android is the leader when comparing the mobile operating systems in terms of global usage strength, as it holds 72% share of the global mobile operating system market whereas iOS holds 28%.



Therefore, having information about applications available on Google Play Store, such as the *number of downloads*, *number of reviews*, *categories*, *ratings*, *prices*, and other metrics can be highly useful for several reasons.

It can **assist developers** in **understanding their competition**, **identifying gaps in the market**, and **optimizing their app's performance**. It can also provide valuable insights to researchers and analysts, who can use this data to draw conclusions about trends in the app market, user preferences, and the overall health of the mobile app industry.

The dataset is available on Kaggle and via this [link](#).

An interactive tool for developers to get tailored insights on the Android Market.

This tool is **designed for mobile app developers**. It provides an analysis of the app market to help them make the most of market opportunities and gaps.

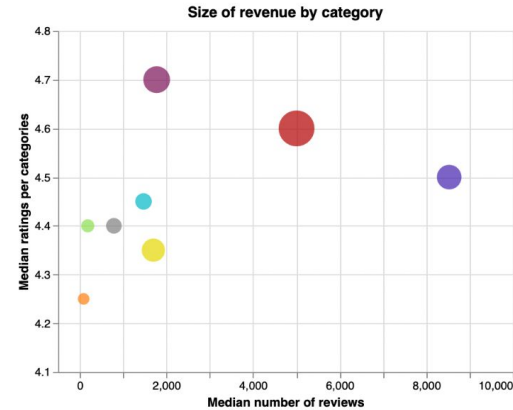
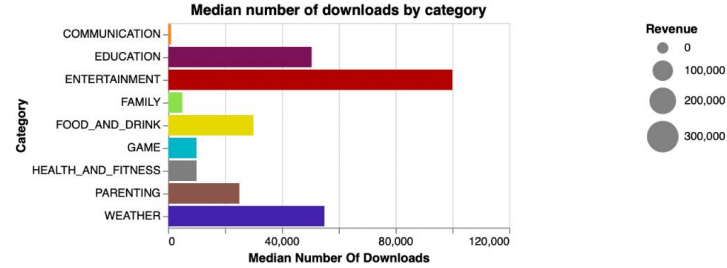
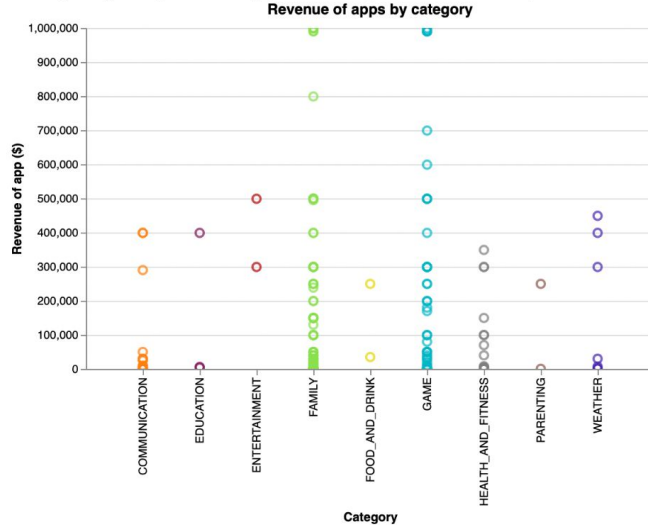
The interactive tool **visualises the interplay between different features of the dataset**, such as, the category of an application, the ratings of the application, revenues generated and many others. By identifying the usage patterns of different categories of users, for different categories of app, we aim to provide key information for developers to make the most of the Android app market.

To obtain **tailor-made findings for their specific needs**, developers have the possibility to **select** certain data, or to apply **filters** on it.



Three views for developers to get insights based on the app category.

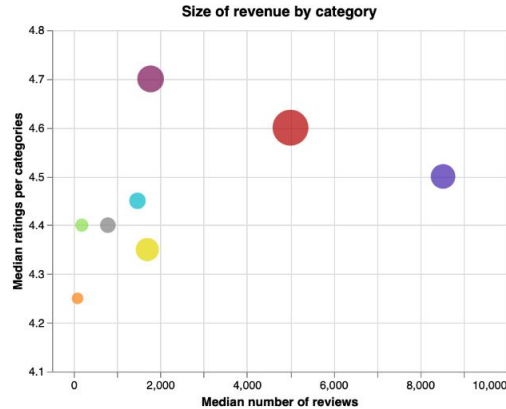
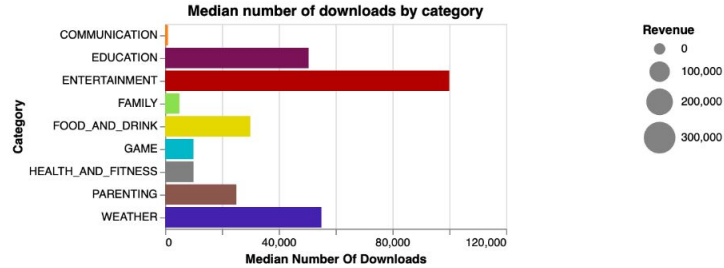
Analyzing Play Store app data to provide developers with insights to conquer the Android Market



Type_Type ☐ Free ☒ Paid

Age target

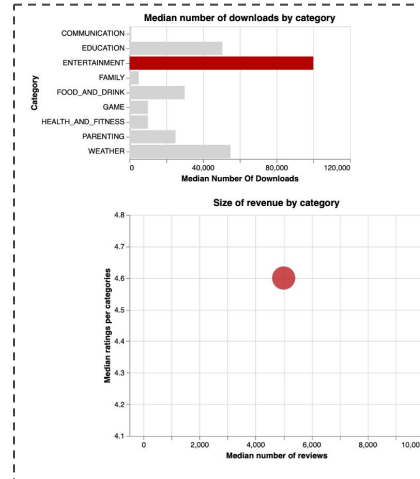
Two linked views to get insights about app's rating, revenue and size.



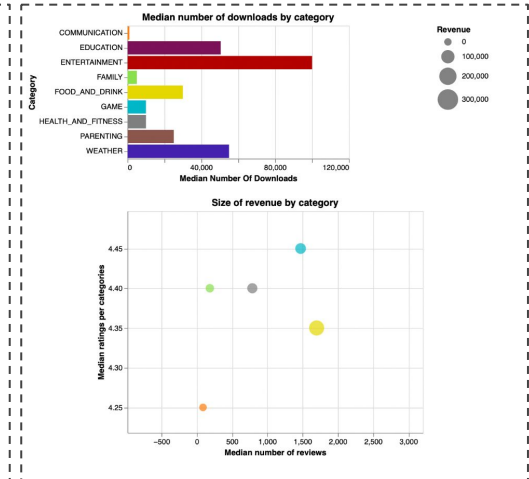
Type_Type ○ Free ● Paid

The first two views are **coordinated with linked highlighting**. A click interaction within one view triggers a change in the other view. The selection is based on the app category. Also, both views can be filtered based on if the app is paid or free.

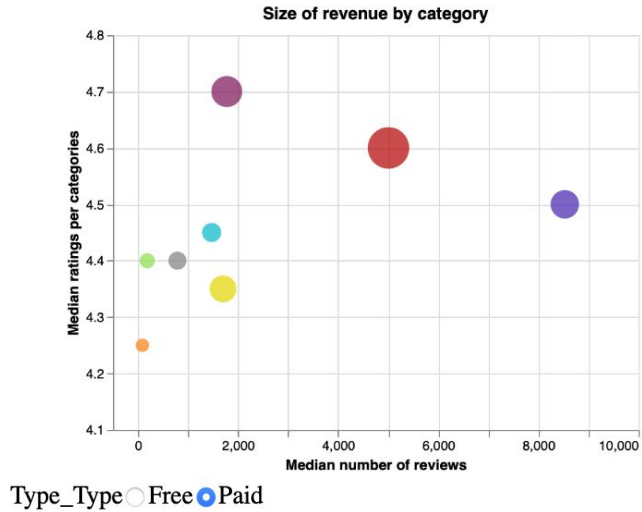
Example of category selection



Example of scale interactivity



The major view to help developers determine the most profitable app category.

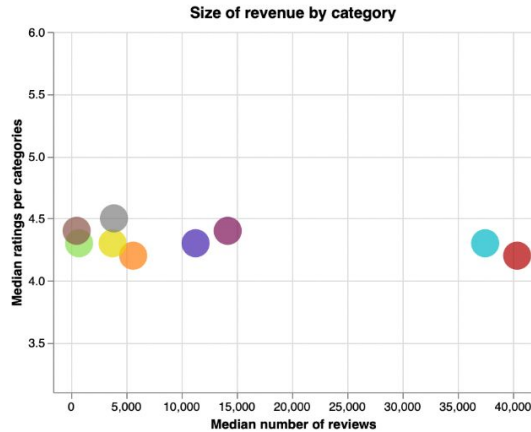
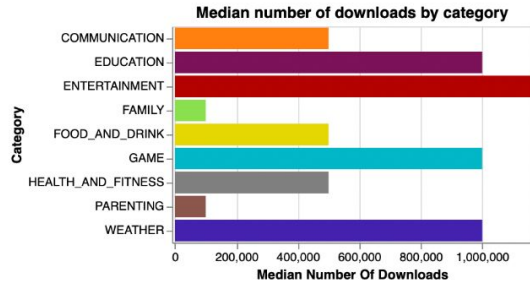


The first view, a bubble size chart, provides an analysis of the app market to help determine the most profitable app category, also considering the **expected technical level** of the app.

It displays three information: the **median revenue**, **median number of reviews** and **median rating** given to an app according to its category. The first information is given by the surface of the circle: the bigger the circle, the more profitable the category. The second and third ideas are given respectively by the x and y-coordinates of the point: the more the category is on the right, the more often the category has been rated; the further up the category, the higher the rating given to apps in that category.

One conclusion can be that a high revenue category with a certain number of low ratings, meaning **lower quality applications**, can be an opportunity for a developer to develop a fast (lower quality) application while getting as much revenue. For example, **Food and Drink apps** display a consequent revenue, similar to Education and Weather apps, while having on average a lower rating, and in volume. It means that the technical level expected for a new app in that category will not be so high, while developers will be able to earn significant revenue from it.

A second view giving an order of magnitude of the scope of app advertising.

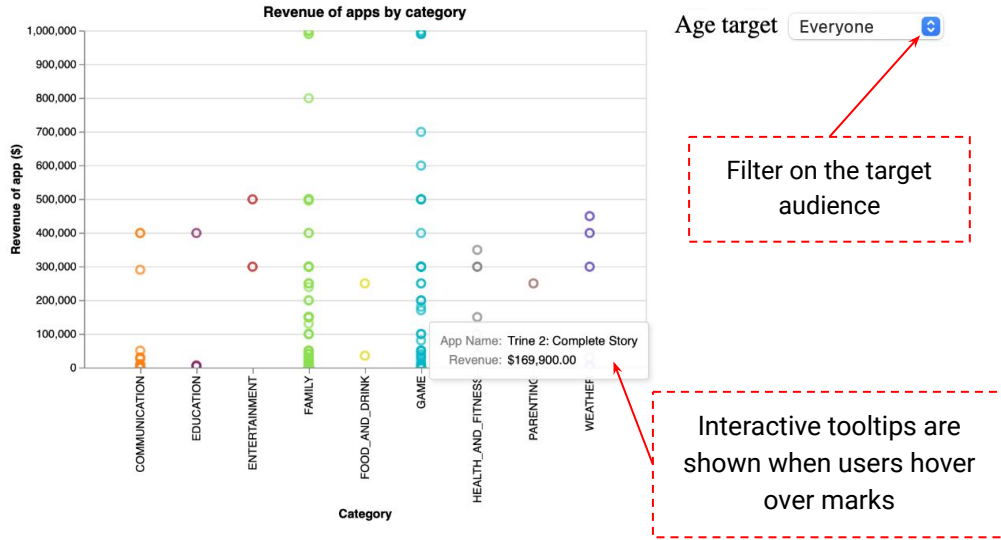


Type_Type ● Free ○ Paid

Alongside the scatter plot, the second linked view, a bar chart, shows developers the **total number of downloads** of an application according to its category. This is useful for developers who want to develop a free application, for which there is no income of their own. Knowing the number of downloads is useful since **more downloads means more advertising revenue**. Here we can guess that the **Entertainment category** would probably be the most profitable, since the total number of downloads is the highest, while the median rating is similar to other apps, for the highest total number of reviews.

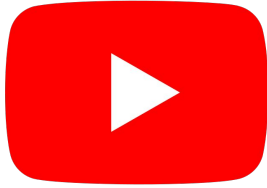


A third view to estimate the total revenue of application per category.

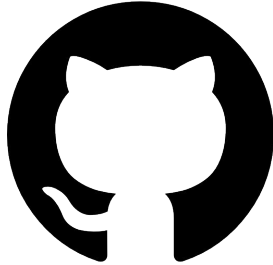


The third view shows that the most represented categories are mainly family and games. Also, most of the circles are in the lower part of the graph, meaning that **only a few apps manage to stand out from other applications** by generating very important revenues. Most of the latter are in the **family or games categories**, which could lead developers to see an opportunity. However, this effect is statistical: since there are more apps in these two categories, it is normal that there are more apps generating significant revenues.

In reality, getting into the development of an application of this type will not necessarily pay off, because the **market is already saturated** and we notice that the applications that generate a lot of money are very well known (*Call of Duty*, *Fruit Ninja*). Instead, another strategy could be to launch an app in a valuable category not saturated, such as **Entertainment**. A few apps exist, which means that the category is still not too competitive, and on average they generate more revenue than other categories (as seen in the previous slides).



[Data Visualization - La French](#)



[Github.com/emmarql/Data_Vizualization_Project](#)

CHIKLI Hugo

Code: Data preprocessing + Altair (1 view) + html

HELBURG Aline

Code: Data preprocessing +Altair (1 view + 1 filter) + script

MICHEL Pauline

Slides + script

RIGUIDEL Emma

Code: Data preprocessing + Altair (1 view + 1 filter) + script