Anusha Cherukuri

Anusha Cherukuri - 10505580  
Anudeep Raju Basavaraju - 10507166  
Lakshmi Pranathi Tadavarthy - 10505692

Data visualization Assignment-2

Interactive Visual Analysis

# **Assignment 2: Interactive Visual Analysis**

## Step 1: Data Selection

With million of apps around nowadays, the following data set has become very key to getting top trending apps in iOS app store. This data set contains more than 7000 Apple iOS mobile application details. The data was extracted from the ***iTunes Search API*** at the Apple Inc website. R and Linux web scraping tools were used for this study.

* **Data collection date (from API):** July 2017
* **Dimension of the data set:** 7197 rows and 16 columns
* The dataset is selected from the Kaggle repository and can be found in the below link:
* <https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps>
* The dataset contains two csv files named:
* appleStore.csv
* appleStore\_description.csv

Here We make use of “**appleStore.csv”** only

## Step 2: Interactive Visual Analysis

Interactive Visual Analysis is to perform an exploratory analysis on the selected dataset using a visualization tool such as Python, matplotlib, Rshiny, D3, Seaborn and Bokeh etc. I have used the visualization tool named python, seaborn, matplotlib to Visualize my selected dataset.

We should consider two different phases of exploration.

In the **first phase**, we should seek to gain an overview of the shape & structure of your dataset such as: What variables does the dataset contain?

**What variables does the selected dataset contain?**

**appleStore.csv**

1. "id" : App ID
2. "track\_name": App Name
3. "size\_bytes": Size (in Bytes)
4. "currency": Currency Type
5. "price": Price amount
6. "rating\_count\_tot": User Rating counts (for all version)
7. "rating\_count\_ver": User Rating counts (for current version)
8. "user\_rating" : Average User Rating value (for all version)
9. "user\_rating\_ver": Average User Rating value (for current version)
10. "ver" : Latest version code
11. "cont\_rating": Content Rating
12. "prime\_genre": Primary Genre
13. "sup\_devices.num": Number of supporting devices
14. "ipadSc\_urls.num": Number of screenshots showed for display
15. "lang.num": Number of supported languages
16. "vpp\_lic": Vpp Device Based Licensing Enabled

* In the **second phase**, we should investigate your initial questions, as well as any new questions that arise during your exploration.
* The questions raised by exploring dataset are mentioned under each visualization

**Initial Analysis Questions**

## How do you visualize price distribution of paid apps?

## How does the price distribution get affected by category?

## What about paid apps Vs Free apps?

## Are paid apps good enough?

## As the size of the app increases do they get pricier?

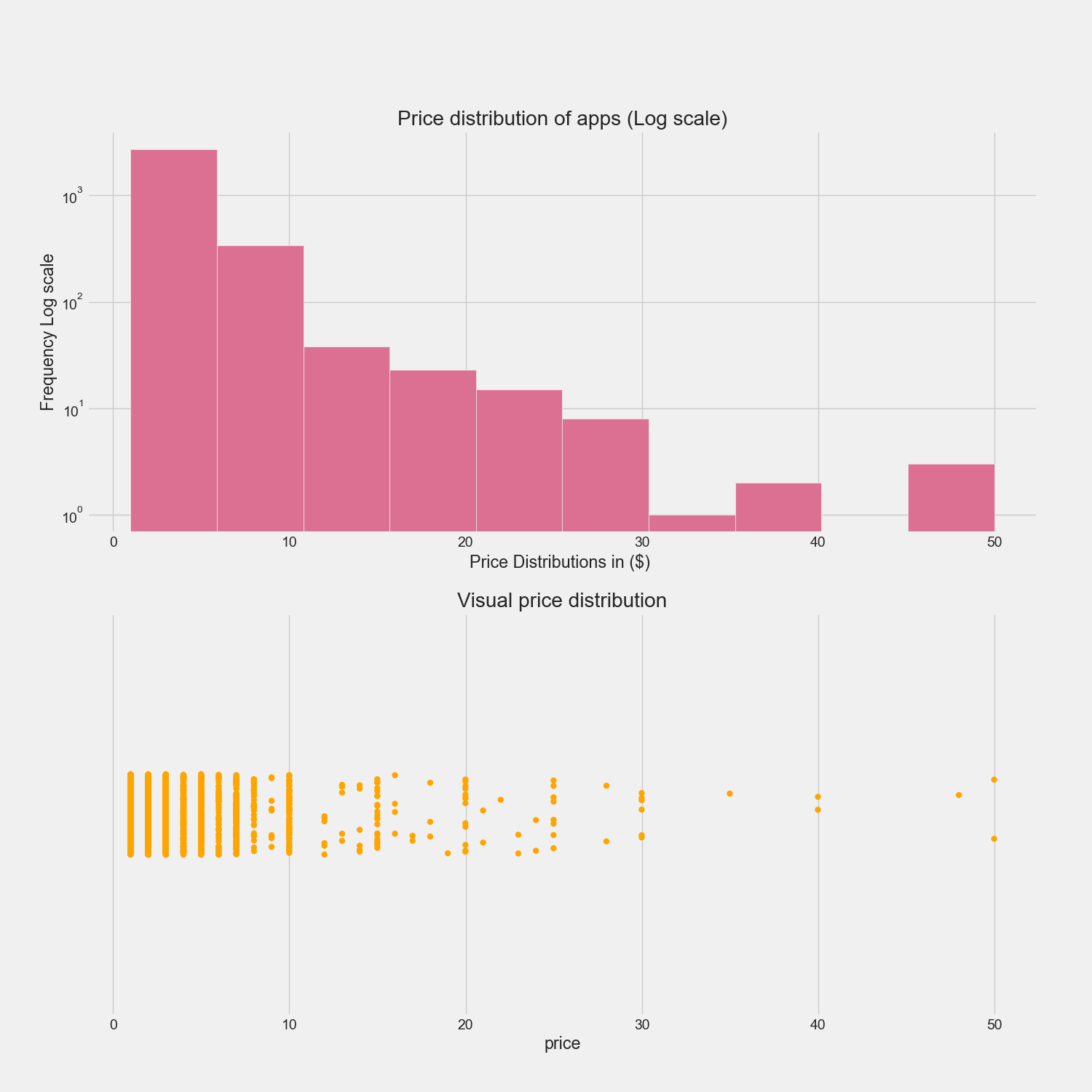
## How are the apps distributed category wise? Can we split by paid category?

**VISUALIZATIONS**

Visualization 1:

## **Question1:**

## **How do you visualize price distribution of paid apps?**

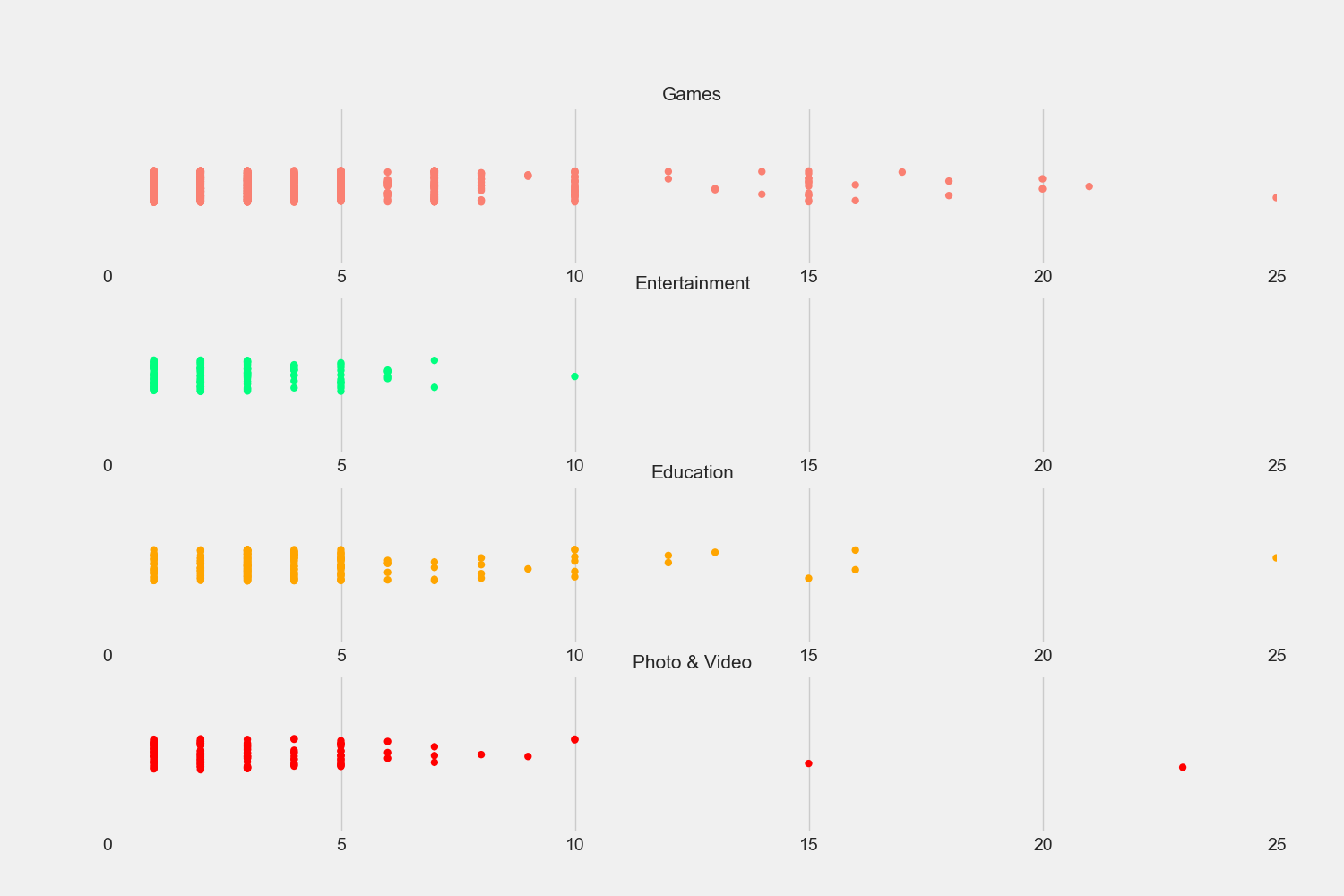


* From the above visualization, it is clearly observed that the price of the apps varies according to the number of downloads.
* As the number of downloads increases, the demand for the app increases so that the price of the app will be increased.
* Similarly, if a particular app has a smaller number of downloads, the price of the app may get decreased and, in few cases, the app may be available for free.

Visualization 2:

**Question 2:**

**How does the price distribution get affected by category?**



* From the above visualization the price is distributed according to the categories.
* We can observe that price distribution for games is more when compared to the entertainment, education, photo &video.
* As most of the population in the world shows interest towards gaming apps, so the demand for the gaming apps is high and the price is high as well.

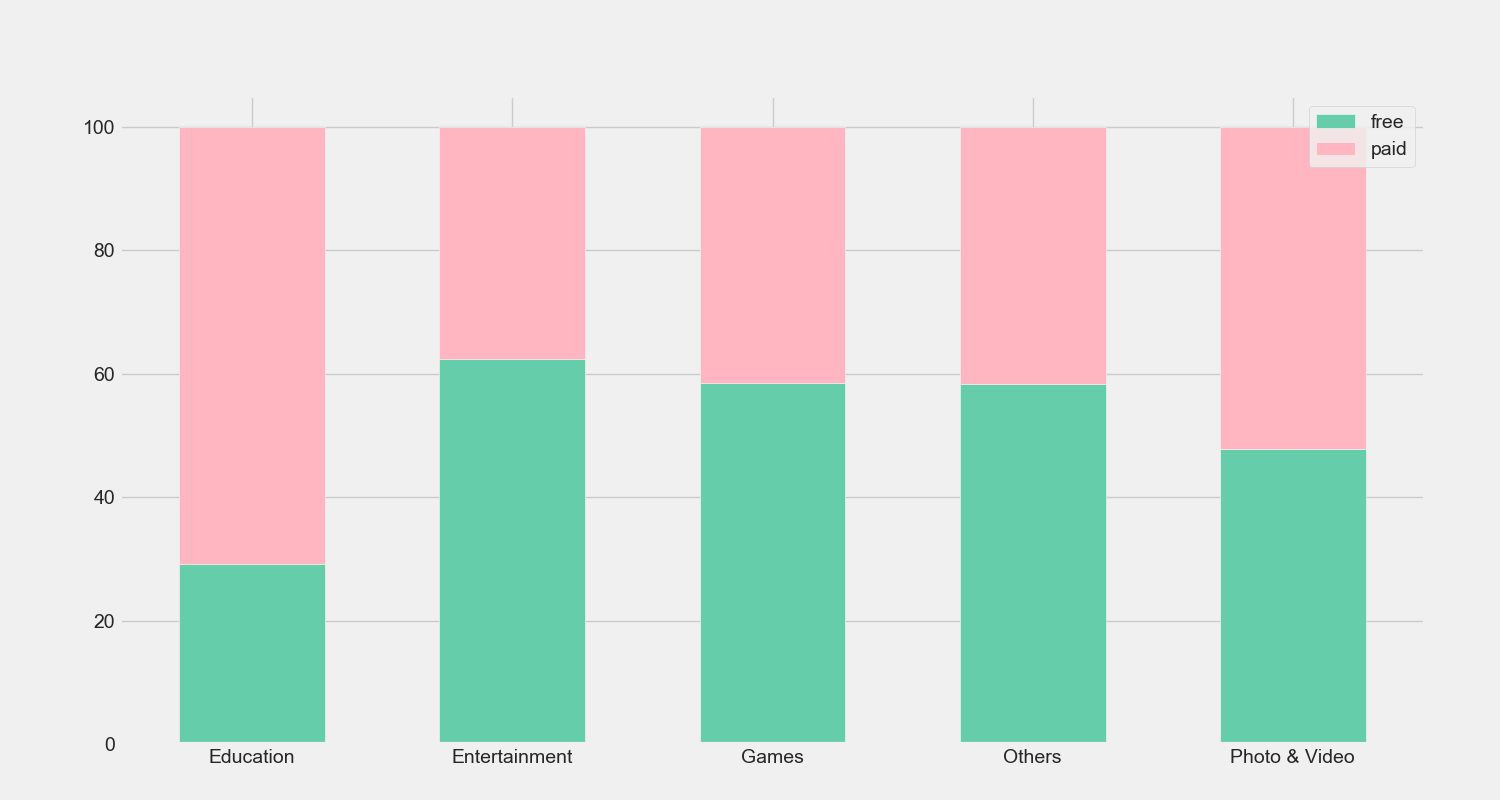
**INSIGHTS:**

* Paid gaming apps are highly priced and distribution extends till 25$.
* Paid entertainment apps have lower price range

Visualization 3:

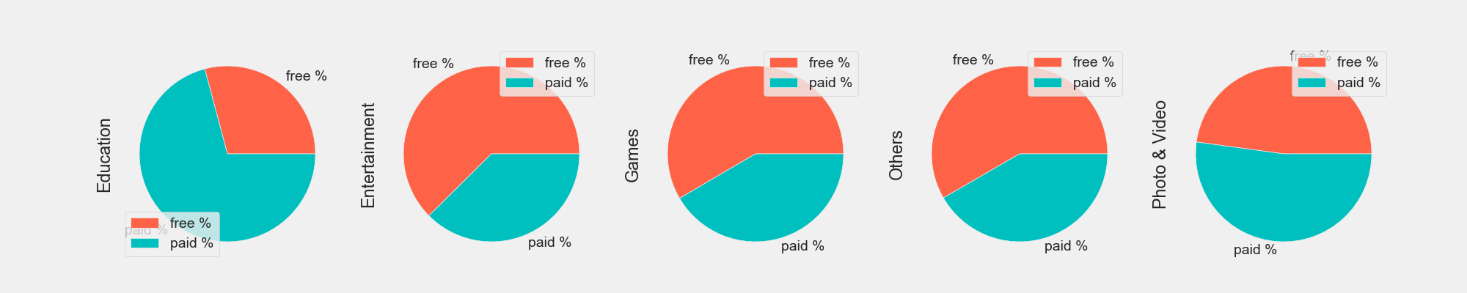
**Question 3:**

**What about paid apps Vs Free apps?**



* From the above visualization we can observe that there are many paid apps for education compared to entertainment, games, photo & video and others.
* Whereas for entertainment, free apps are more when compared to paid apps.
* Coming to games and others availability of free and paid are very similar.
* For photo & video there are almost equal amount of free and paid apps.

Visualization 4:



This visualization is same as like above visualization.

But here it is represented in pie charts manner.

Visualization 5:

## **Question 5:**

## **Are paid apps good enough?**



* From the above visualization it is observed that the ratings of user for free apps in entertainment is high when compared to other categories.
* Coming to games category the user ratings for free and paid apps are nearly similar.
* For others category the free apps have highest user rating compared to paid apps.
* In education category user ratings for paid apps are more compared to free apps.

**NOTE**:

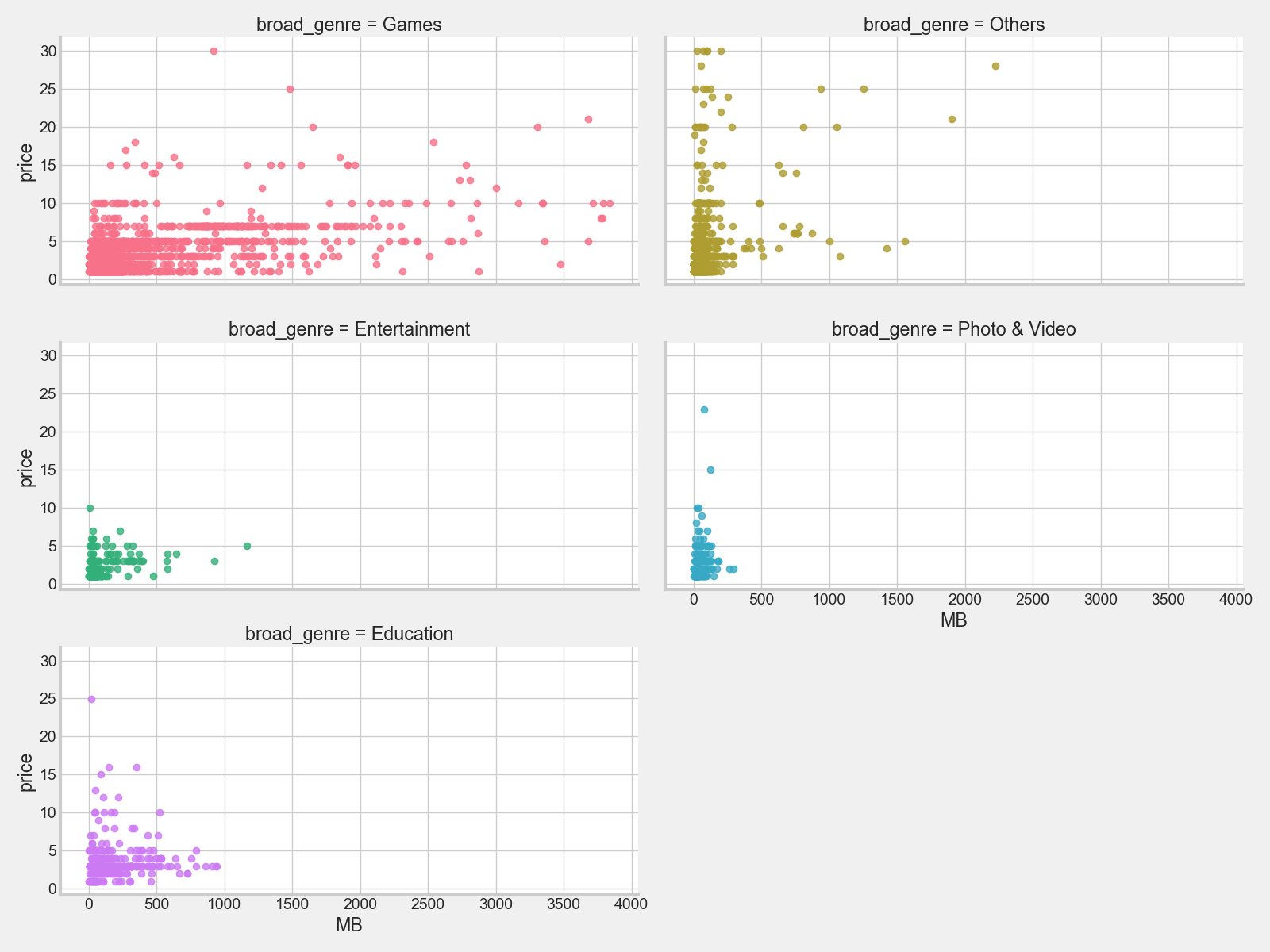
**Why can’t we use swarm plot?**

* Swarm is a non-overlap plot i.e., it will plot each point separately on the graph.
* It will have to plot each value separately on the graph.
* The graph will expand horizontally.

Visualization 6:

## **Question 6:**

## **As the size of the app increases do they get pricier?**

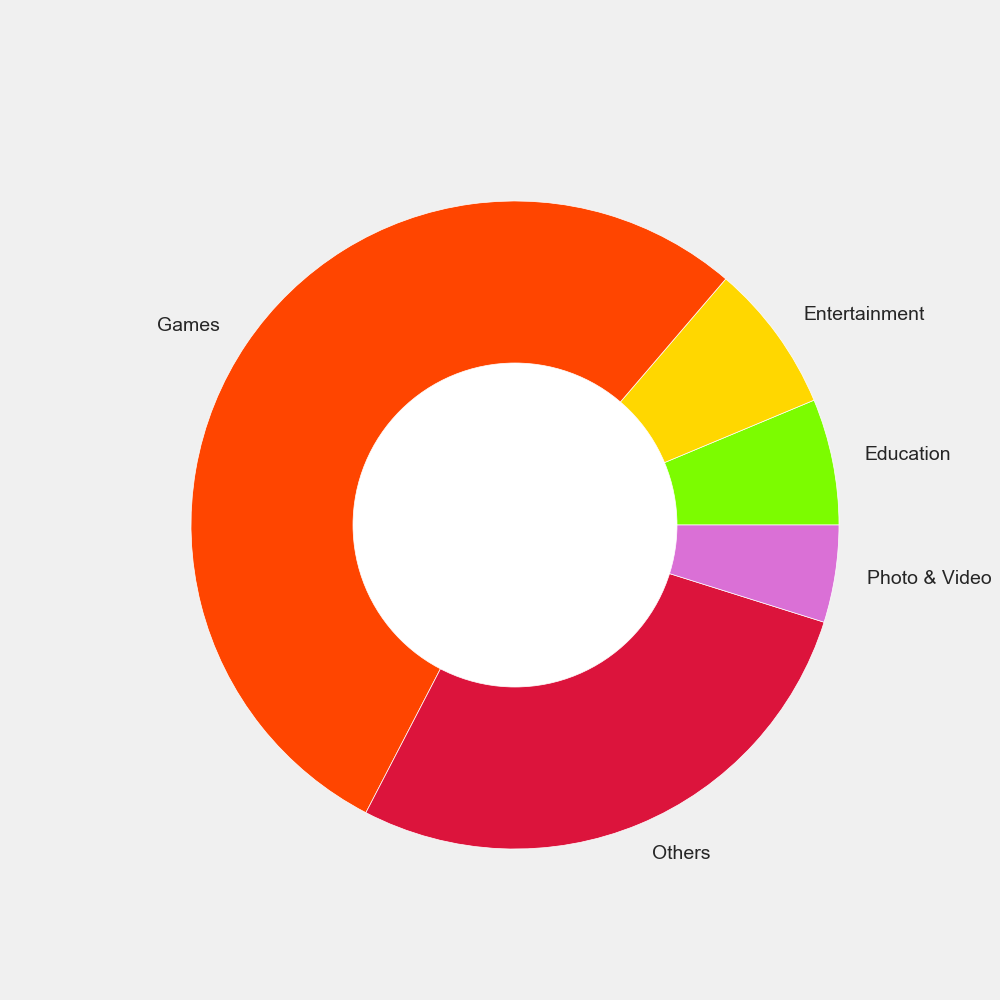


* From the above visualization we can observe that in the games category when the app size increases the price increases.
* Coming to others category it is less compared to games.
* In remaining 3 categories it is very less compared to other categories.

Visualization 7:

## **Question 7:**

## **How are the apps distributed category wise? Can we split by paid category?**

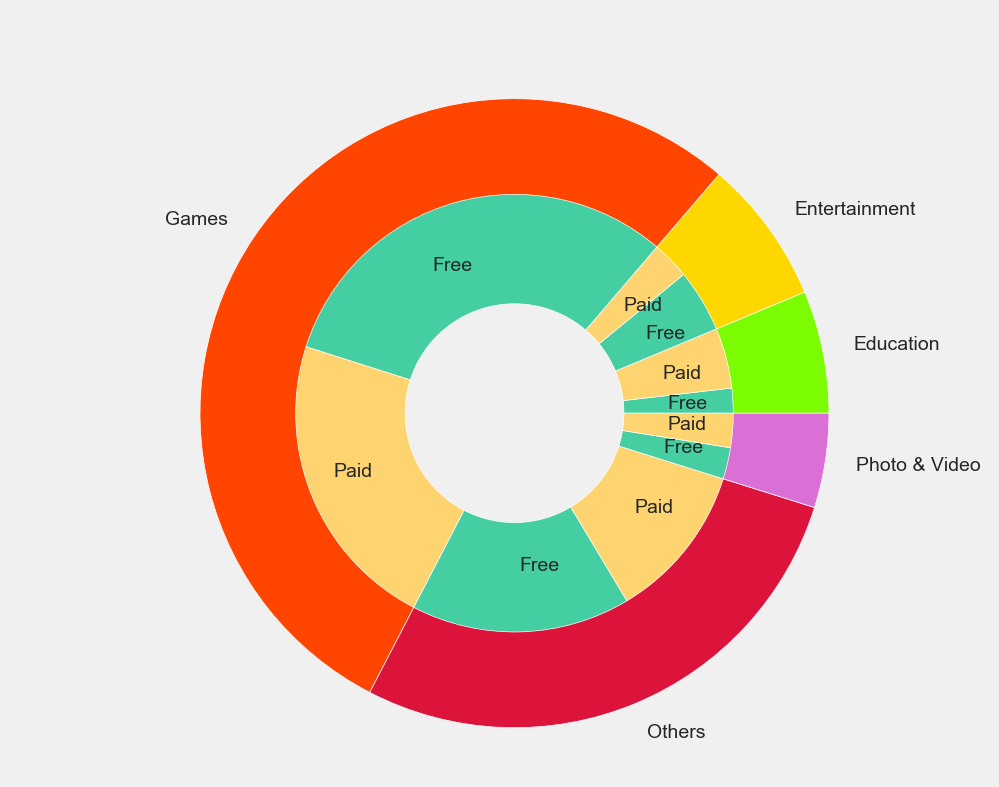


* From the above visualization we can observe that all the categories are divided based on paid wise.
* Games has the highest number of paid apps.
* Others occupy the second place in paid category compared to games.
* Remaining 3 categories has the equal number of paid apps.

Visualization 8:

**Question 8:**

The below Visualization shows the amount of free and paid apps in each category:



* This visualization is similar to the above visualization7.
* But in the visualization 7 it has just shown the divisions of category in a brief way.
* But in this current visualization it is explained and shown clearly that how many free apps and paid apps are there in each category.