EDA PROJECT COMMENTARY

Christian Park

Introduction

- For this project, I chose to analyze a data set that covered apps in the Apple Store. In order to provide succinct and focused analysis the variables I used during this project include
- "id" : App ID
- "track_name": App Name
- "size_bytes": Size (in Bytes)
- "price": Price amount
- "rating_count_tot": User Rating counts (for all version)
- "rating_count_ver": User Rating counts (for current version)
- "prime_genre": Primary Genre

Exploratory Guiding Questions

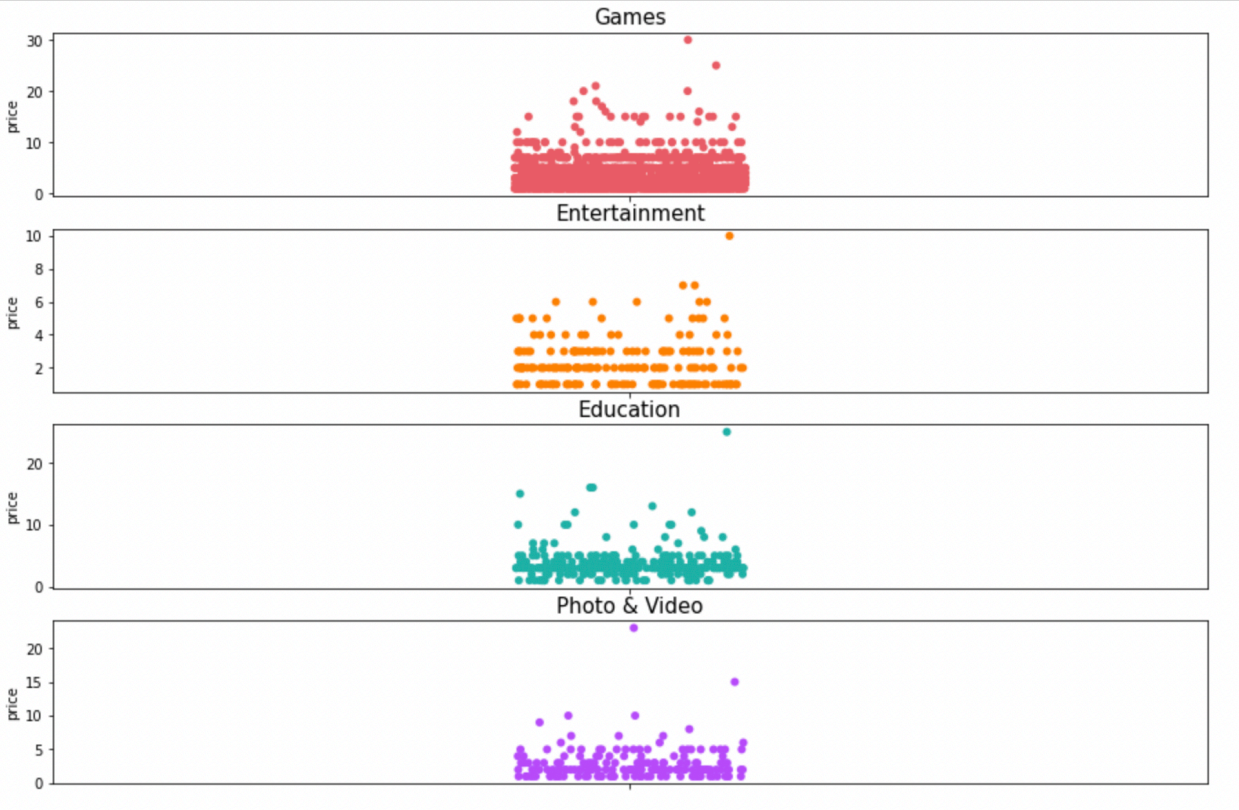
- 1. How is app price affected by the category it is in?
- 2. How does the distribution of free vs. paid apps vary in each category?
- 3. Is there a relationship between app price and the size of the app?

Understanding Distribution of Prices

	track_name	price	prime_genre	user_rating
115	Proloquo2Go - Symbol-based AAC	249.99	Education	4.0
162	NAVIGON Europe	74.99	Navigation	3.5
1136	Articulation Station Pro	59.99	Education	4.5
1479	LAMP Words For Life	299.99	Education	4.0
2181	Articulation Test Center Pro	59.99	Education	4.5
2568	KNFB Reader	99.99	Productivity	4.5
3238	FineScanner Pro - PDF Document Scanner App + OCR	59.99	Business	4.0

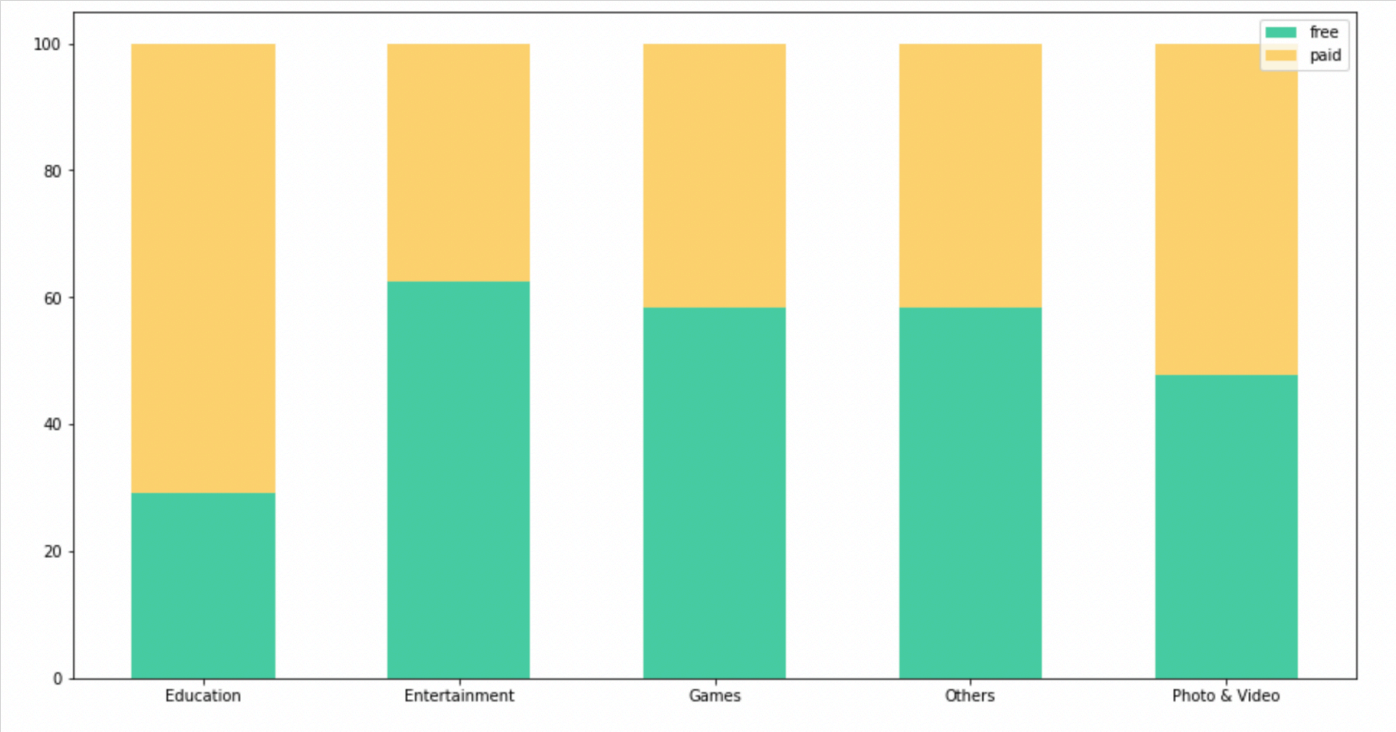
I performed analysis to search for outliers within the data and found that there were only 7 apps that were abnormally expensive compared to the rest of the apps within the data set. Therefore, in continuing with my data wrangling I removed these apps as they accounted for less than 1 percent of the data set. As a result of my data management, the max price of any app within the new data was $49.99 and the cheapest price was $.99.

How is price affected by category?

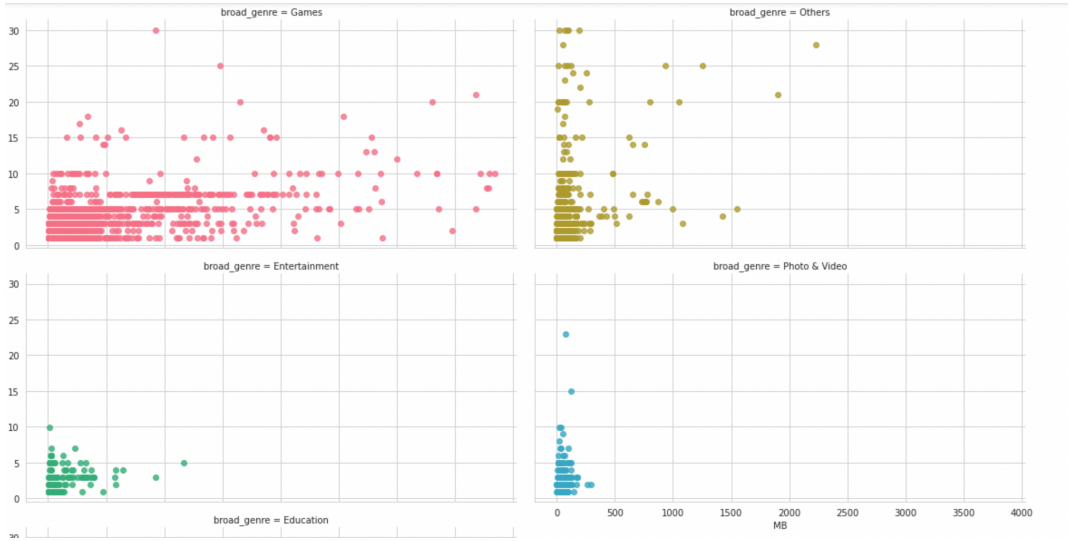


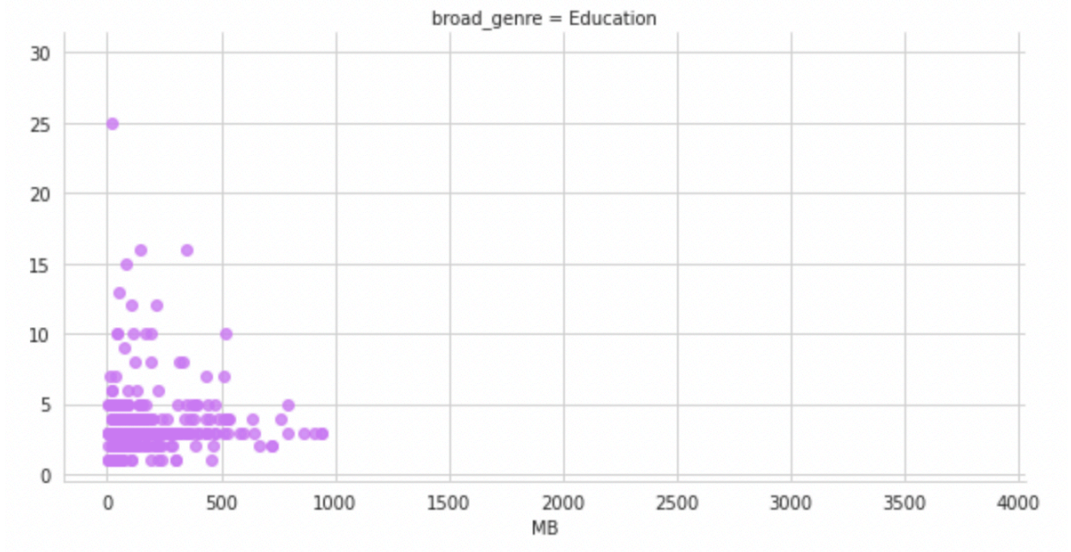
I was interested in understanding the relationship between price and category. These plots show that for Games, paid apps are priced very highly and leads all the way up to $25. Furthermore, it shows that Entertainment apps have a lower price range than the rest of the top categories. In order to do this analysis, I consolidated the different categories into the main genres.

How does distribution of free vs. paid apps vary by category?



Next, I wanted to explore the different variations of bundle mixes regarding free and paid apps by each category. By plotting the distributions, I found that the Education category has a high percentage of paid apps, while conversely Entertainment has a majority percentage of free apps. Similarly, Games and Others had a relatively close mix to Entertainment while Photos and Videos had a more equal split.





Does price increase with app size?

I tested the idea that as app size increase the price would increase as well. This concept followed the logic that you pay more if you get more. However, after analyzing the data set and plotting the distribution of price with relation to the size of the app, I found no evidence that as the size of the app in MB increase, there was no pattern of the price increasing with the size of the app.