Analysis of Apple App Store

Objective:

- To analyze and predict the rating of an application in Apple App store.
- To find mobile app profiles that are profitable for the App store.
- To analyze data to help developers understand the apps that are likely to attract more users.
- To enable team of developers to make data-driven decisions with respect to the kind of apps they build.

Data Dictionary:

Variables	Datatype	Description
id	int	App ID
track_name	object	App Name
size_bytes	int	Size (in bytes)
currency	object	Currency type
price	float	Price Amount
rating_count_tot	int	User Rating counts (for all versions)
rating_count_ver	int	User Rating counts (for current version)
user_rating	float	Average User Rating value (for all versions)
user_rating_ver	float	Average User Rating value (for current version)
ver	object	Latest Version code
cont_rating	object	Content rating
prime_genre	object	Primary Genre
sup_devices.num	int	Number of supporting devices
ipadSc_urls.num	int	Number of screenshots showed for displays
lang.num	int	Number of supported languages
vpp_lic	int	Vpp Device Based Licensing Enabled

Software Used:
Python, Tableau

Source:
Kaggle

Data set size:
11000+ rows and 17 columns

Data Cleaning (*Python: Jupyter Notebook):*

- Checking for Null values.
- Removed rows with null values and renamed null values in measure columns with a 0 for accurate aggregations.
- Removing the plus (+) sign from Content Rating column.

- Checking for duplicate values.

```
# Checking for duplicate Values
store_data.duplicated().sum()
```

```
0
```

- Head of the Dataset:

```
store_data.head()
```

	id	track_name	size_bytes	currency	price	rating_count_tot	rating_count_ver	user_rating	user_rating_ver	ver	cont_rating	prime_genre
0	281656475	PAC-MAN Premium	100788224	USD	3.99	21292	26	4.0	4.5	6.3.5	4+	Games
1	281796108	Evernote - stay organized	158578688	USD	0.00	161065	26	4.0	3.5	8.2.2	4+	Productivity
2	281940292	WeatherBug - Local Weather, Radar, Maps, Alerts	100524032	USD	0.00	188583	2822	3.5	4.5	5.0.0	4+	Weather
3	282614216	eBay: Best App to Buy, Sell, Save! Online Shop...	128512000	USD	0.00	262241	649	4.0	4.5	5.10.0	12+	Shopping
4	282935706	Bible	92774400	USD	0.00	985920	5320	4.5	5.0	7.5.1	4+	Reference

- Tail of the Dataset:

```
store_data.tail()
```

	App_id	App_Name	Size(in_bytes)	Currency type	Price_Amount	User Rating counts(all versions)	User Rating counts(current_version)	Avg_User_Rating_Value(all_versions)	Avg_User
11080	1187617475	Kubik	126644224	USD	0.00	142	75	4.5	
11081	1187682390	VR Roller-Coaster	120760320	USD	0.00	30	30	4.5	
11086	1187779532	Bret Michaels Emojis + Lyric Keyboard	111322112	USD	1.99	15	0	4.5	
11088	1187838770	VR Roller Coaster World - Virtual Reality	97235968	USD	0.00	85	32	4.5	
11096	1188375727	Escape the Sweet Shop Series	90898432	USD	0.00	3	3	5.0	

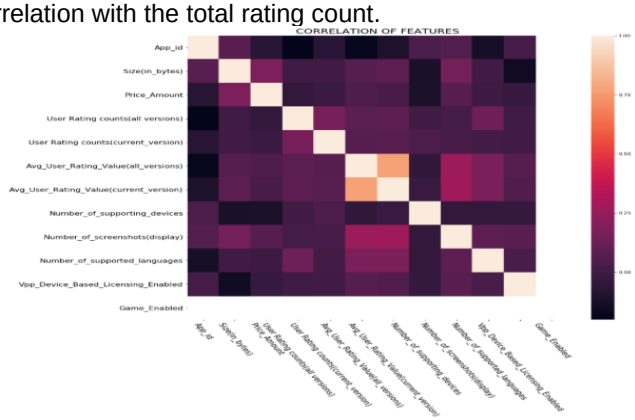
- Shape Before Cleaning: (11100,17)
- Shape After Cleaning: (7197,17)

- Encoding the Type column by converting free value to 0s and paid value to 1s based on price column.

- Renamed columns as per business requirement and in a more readable format.

```
App_id
App_Name
Size(in_bytes)
Currency type
Price_Amount
User Rating counts(all versions)
User Rating counts(current_version)
Avg_User_Rating_Value(all_versions)
Avg_User_Rating_Value(current_version)
Latest_Version_code
Content_rating
Primary_ Genre
Number_of_supporting_devices
Number_of_screenshots(display)
Number_of_supported_languages
Vpp_Device_Based_Licensing_Enabled
Game_Enabled
```

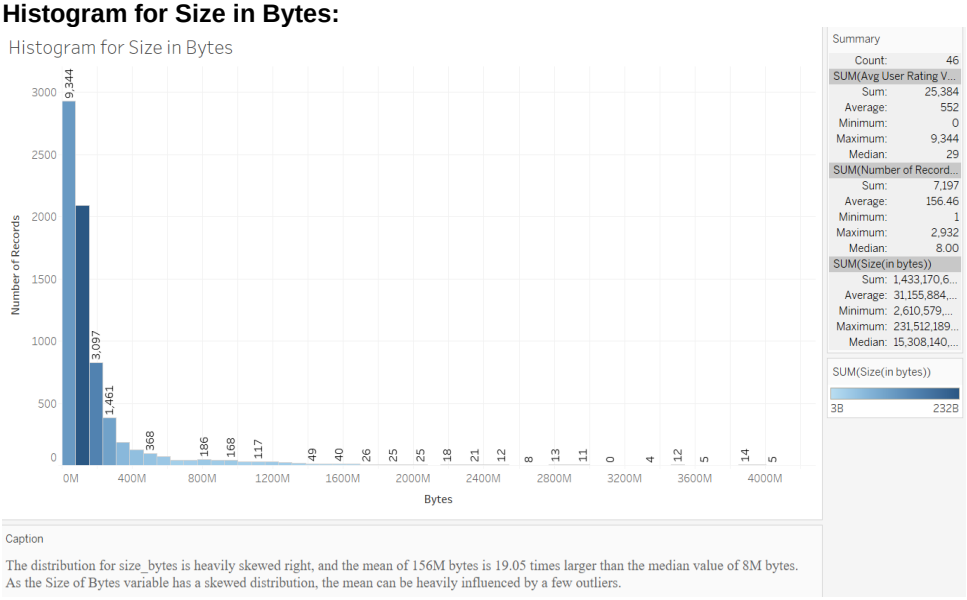
- Number of languages app support) shows the highest correlation with the total rating count.



.

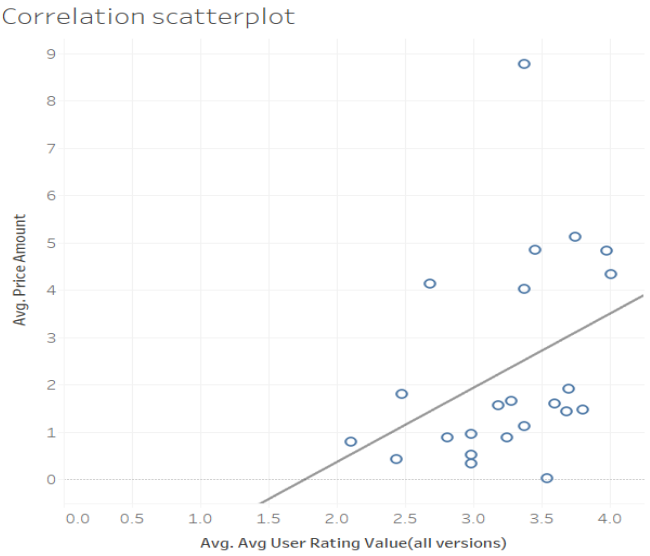
Exploratory Data Analysis/ Data Visualizations (Tableau):

- **Histogram for Size in Bytes:**



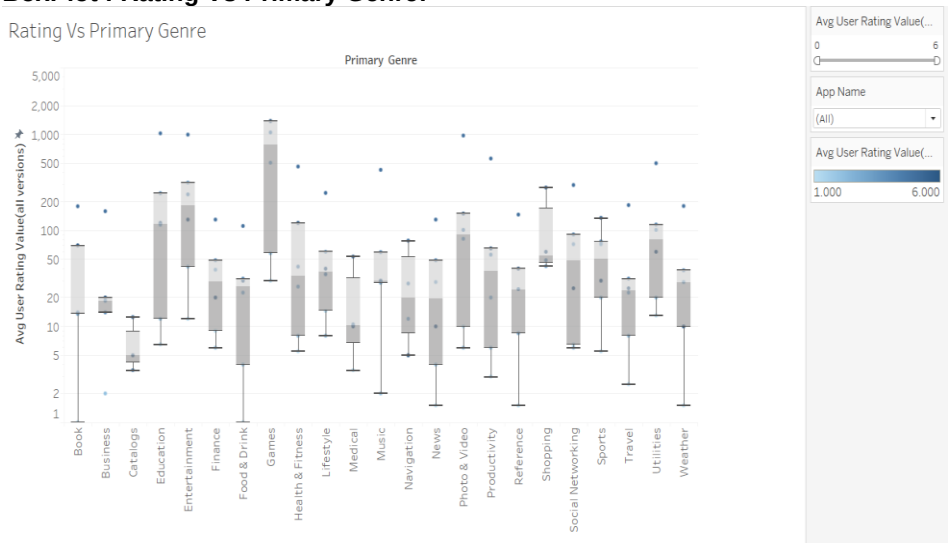
The distribution for size is heavily skewed right, and the mean of 156M bytes is 19.05 times larger than the median value of 8M bytes. As the Size of Bytes variable has a skewed distribution, the mean can be heavily influenced by a few outliers.

- **Correlation scatterplot:**



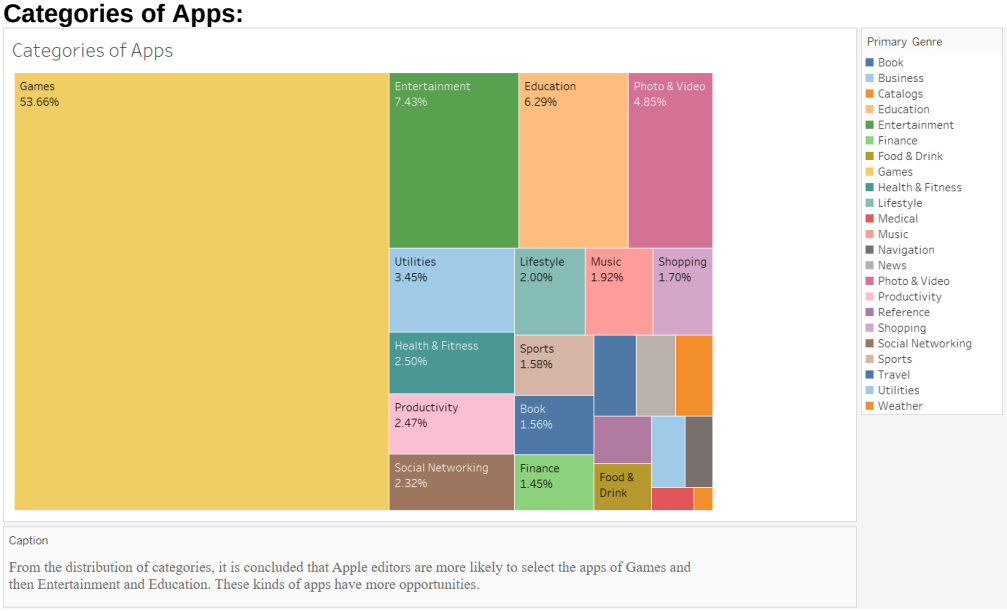
The chart shows a quite strong positive correlation between price and ratings, with more costly apps tending to have a higher rating. But correlation does not imply causation, putting a high price tag on your app will not necessarily improve your ratings.

- **BoxPlot : Rating Vs Primary Genre:**



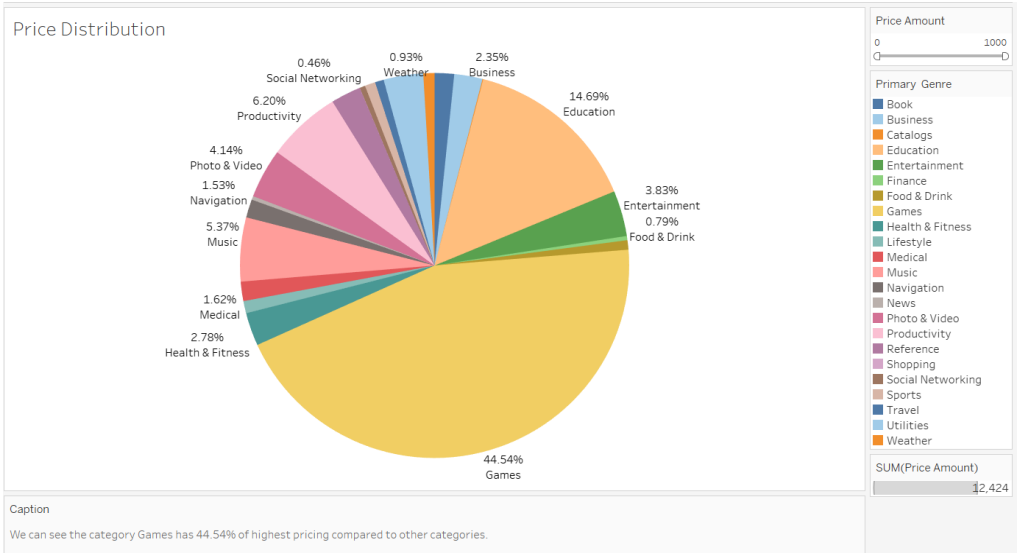
We can see the Games, Education, and Health & Fitness categories have high variances. Users leave the widest range of review scores for these apps. The catalogs category also has the lowest mean. However, if you visit the the catalogs section on the app store today, there are 0 apps listed. Maybe this category was discontinued? The Games genre's ratings are high and tightly distributed. People who are rating Games apps are consistently rating them well. For the highest mean rating, it's a three-way tie between Photos & Videos, Health & Fitness and Games all with a mean rating of 4.7. Apps in the Photos & Video category seem to perform very, so we have some stiff competition.

- **Categories of Apps:**



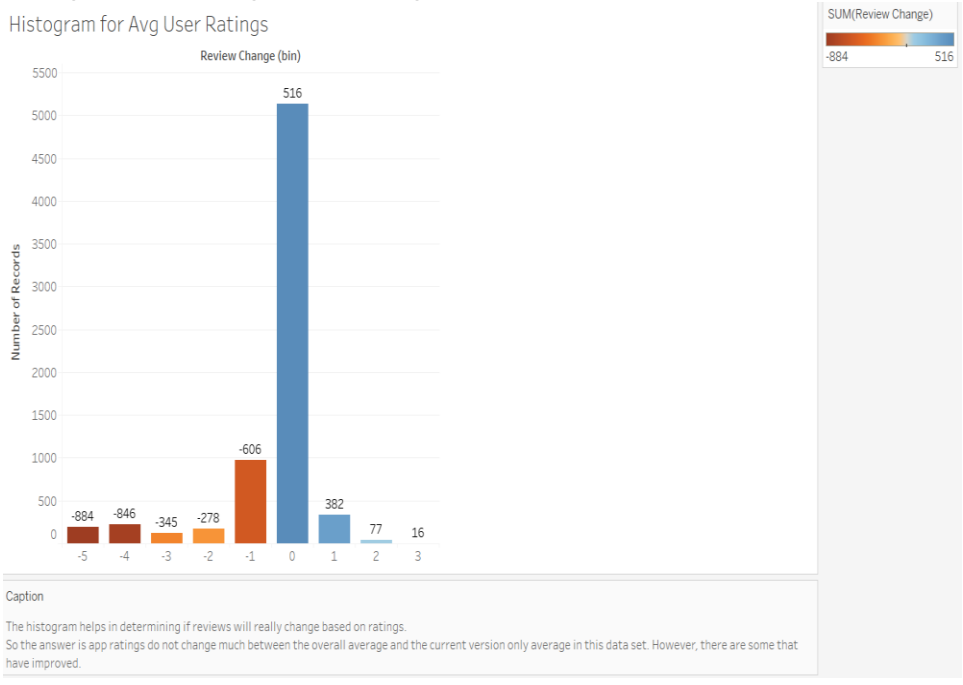
From the distribution of categories, it is concluded that Apple editors are more likely to select the apps of Games with 53.66% and then Entertainment (7.43%) and Education (6.29%). These apps have more opportunities and would improve the revenue of the apps. I have plotted a tree map of total number of apps against Primary Genre column.

- **Price Distribution:**



Here, we can see the category Games has 44.54% of higher pricing compared to other categories. We can infer from the pie chart, that a greater number of app purchases are under Games followed by Education with 14.69% of total revenue incurred. This would help the developer to make efficient decisions in enhancing Games and Education for higher profits.

- **Histogram for Average User Ratings:**

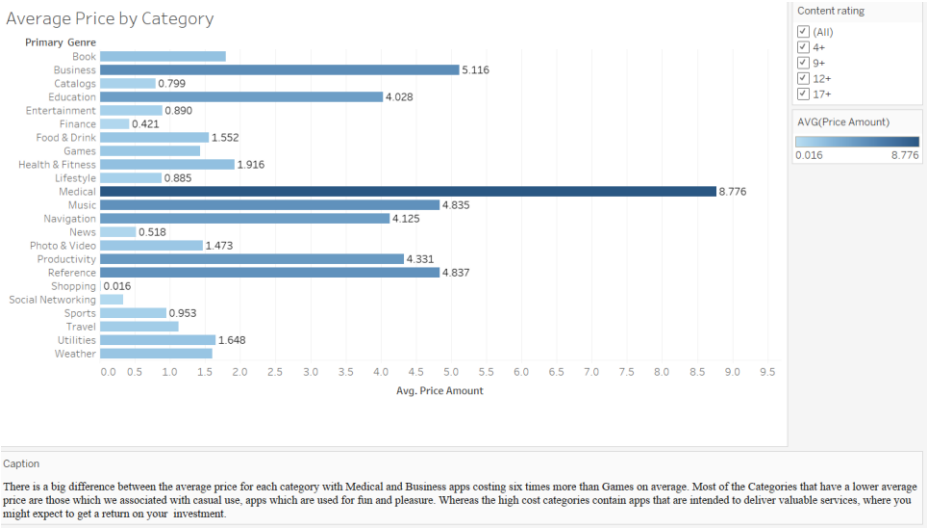


The histogram helps in determining if reviews will really change based on ratings. Here, app ratings do not change much between the overall average and the current version only average in this data set. However, there are some that have improved a lot.

Top 5 Apps with improved rating change

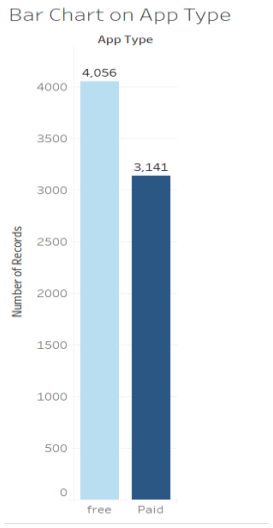
App Name	
VyStar Mobile Banki..	3.5000
Mannequin Challeng..	3.0000
Unblue	3.0000
VR-Crazy Car Traffic ..	3.0000
放置江湖: 大侠的成长..	3.0000

- **Average Price by Category:**

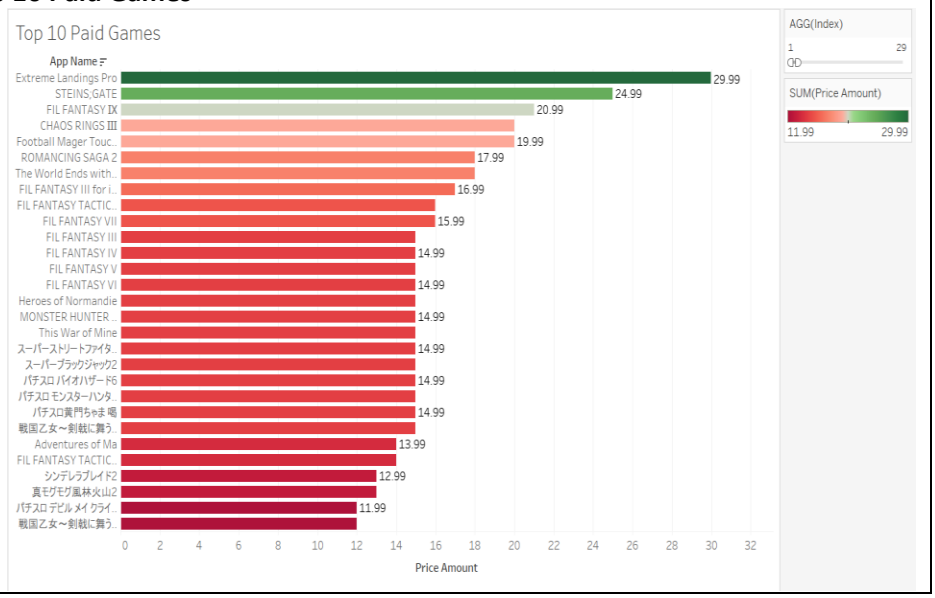


There is a big difference between the average price for each category with Medical and Business apps costing six times more than Games on average. Most of the Categories that have a lower average price are those which we associated with casual use, apps which are used for fun and pleasure. Whereas the high cost categories contain apps that are intended to deliver valuable services, where you might expect to get a return on investment.

- **Bar Chart on App Type**



- **Top 10 Paid Games**



Conclusion:

- Through our data analysis and data visualizations on the dataset, we could find the top categories of Apps to be Games, Education and Medical which would incur higher revenue growth in the market.
- We could visualize the top Apps in the Games category which generated highest profits and it gives the scope in establishing business strategies to improve efficiency and further profits in the industry.
- The EDA helped us in determining the apps which could be potentially discontinued, and which could add not much value to the customers, explained through the number of outliers in distribution graphs. Here, in this case, catalog category seems to be had discontinuity.
- Through the app statistics on different variables, we can see that app quality (as measured by user rating) has a bigger positive impact on total reviews than price does a negative impact.
- However, there are some categories which are less mature than we would expect, such as finance. Intuitively we all want mobile finance, so this may be a dynamic space to explore.

References:

- https://www.kaggle.com/hazemshokry/exploratory-data-analysis-7200-apps-in-ios-store/data
- https://blog.scottlogic.com/2014/03/20/app-store-analysis.html
- https://blog.scottlogic.com/2014/03/20/app-store-analysis.html