App Store Landscape Report

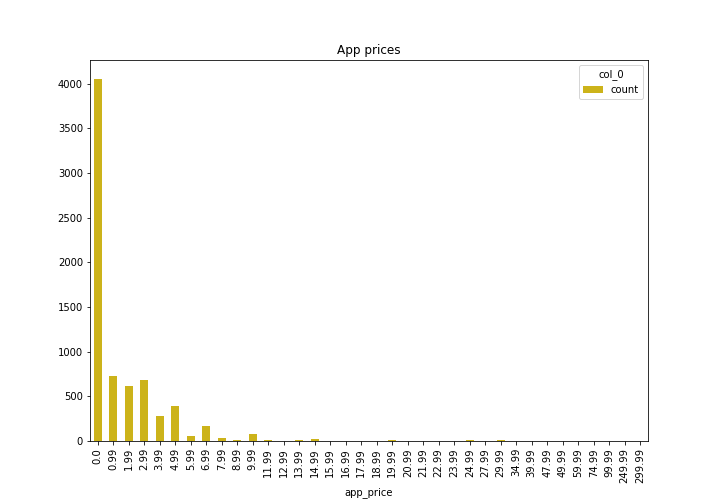
Summary:

Using data from a csv, this report was crafted to report on the current landscape of the app store. The goal is to analyze any trends and find insights for future Xbot projects.

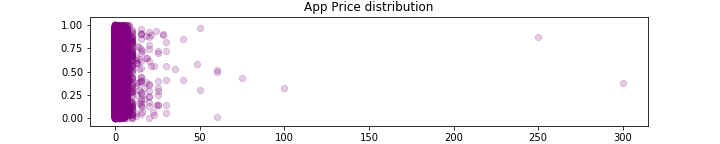
In the analysis of the app store data, python, pandas, matplotlib, and numpy were used in the jupyter notebook. The following link is for the jupyter notebook (ipynb) file used:

<http://localhost:8888/notebooks/Desktop/data_scientist.ipynb>

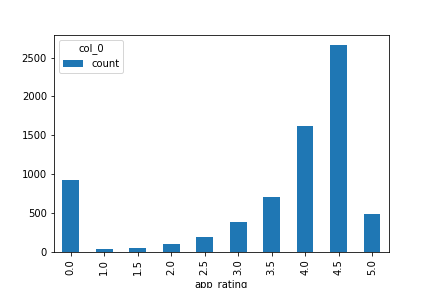
In my analysis, I discovered some common trends. This count of app prices shows the glaring trend that nearly all apps are free. With higher prices, there is less demand for apps because the market is so full of free alternatives. The figure below shows the lack of visibility of prices below around $11.99.



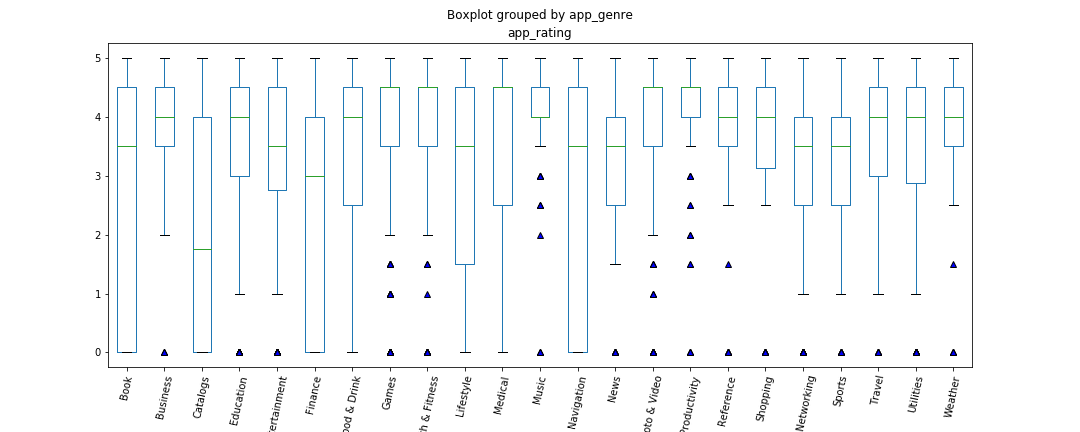
Here is the figure, shown as a scatterplot of prices:



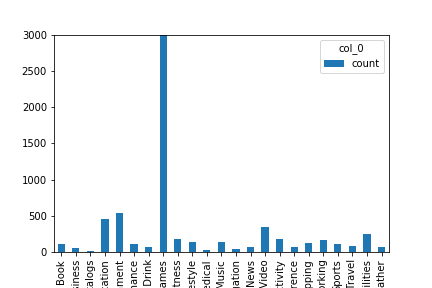
In addition to the price trend, the graph below shows two things. Firstly, many of the apps in the dataset have not been rated or have been rated as 0. Apart from this, the distribution shows a hard right-skew with the mean being much greater than the median. Most apps were rated between 4 and 4.5.



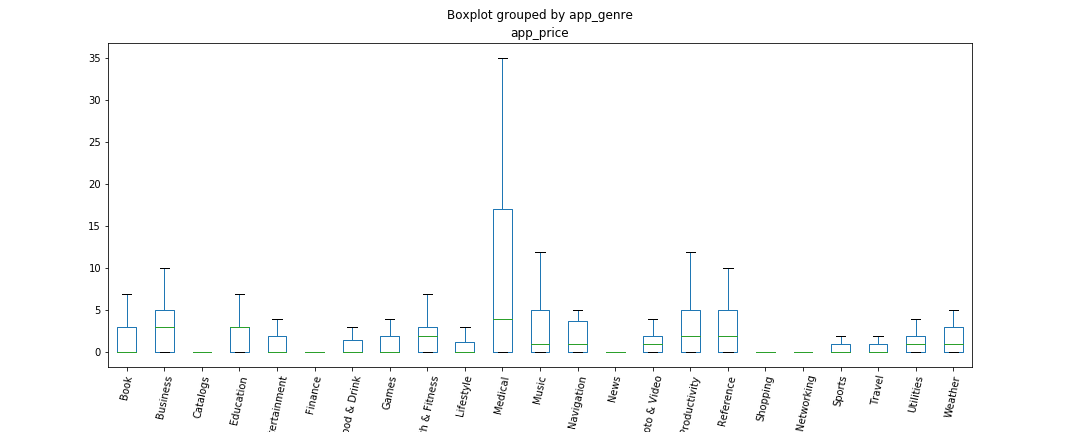
In an attempt to find trends between genres and ratings, this boxplot shows that games, health & fitness, productivity, photo & video, and music apps all have high average ratings and smaller standard deviations than the other genres.



In addition to the trend in price, plotting the number of apps in each category shows an extreme trend in game apps. A vast majority of all apps in the app store are game apps with entertainment and transportation apps coming in well behind in number as second and third.



This boxplot (without outliers) shows statistics on app price per genre. From this graph we determine that app categories of shopping, networking, finance, catalogs, and news are nearly always free while medical, music, business, and productivity apps have means between 1 and 5 dollars.



Notes on graph production using python & data visualization libraries:

In the production of these graphs, much of what I wanted to show required comparisons of categorical data counts and some other variable. To do this, I used matplotlib’s .boxplot() function and set columns to the values I wanted to compare by genre. From there, I rotated the x tick labels for readability and changed the marker shape and color.

Similar to .boxplot, the function .crosstab in pandas allowed me to count app prices and return their frequencies in a bar graph.

To orient myself with the data before trying some visualizations, I used the following lines of code to display tables.

app\_store\_df.head(20) #displays the first 20 rows of data

app\_store\_df.app\_genre.value\_counts() #shows a 2-d list of all names in app\_genre and their counts.

To save all of the figures from my jupyter notebook, I used plt.savefig(‘name.png’).