Google Play Store Report and Analysis

Introduction:

This report will include data and information regarding the Google Play Store and will be used to answer a user story. The data utilized for answering the user story consists of 10,841 rows and 13 columns all related to Google Play Store statistics and metrics. These ranged from the category of the app, the overall rating, installs, current version, price, and other similar factors. Meaning there are 10,841 apps and relevant information contained within the data.

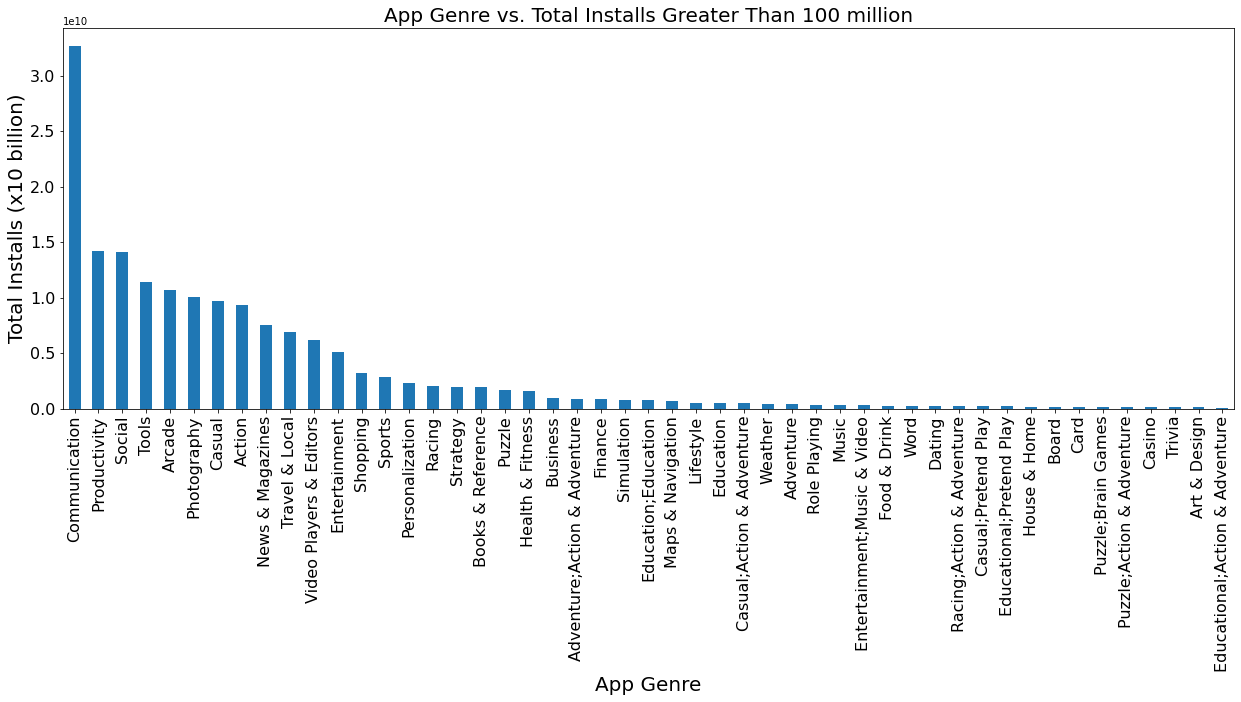
The question we are tasked with is, “Create a function that will sort out how many apps belong to each genre along with their mean installs.” In this report I will be discussing more in-depthly in the following sections about the data, methods used, analysis performed, and the overall results. The results I will present in this report will be the total and mean installs for each app.

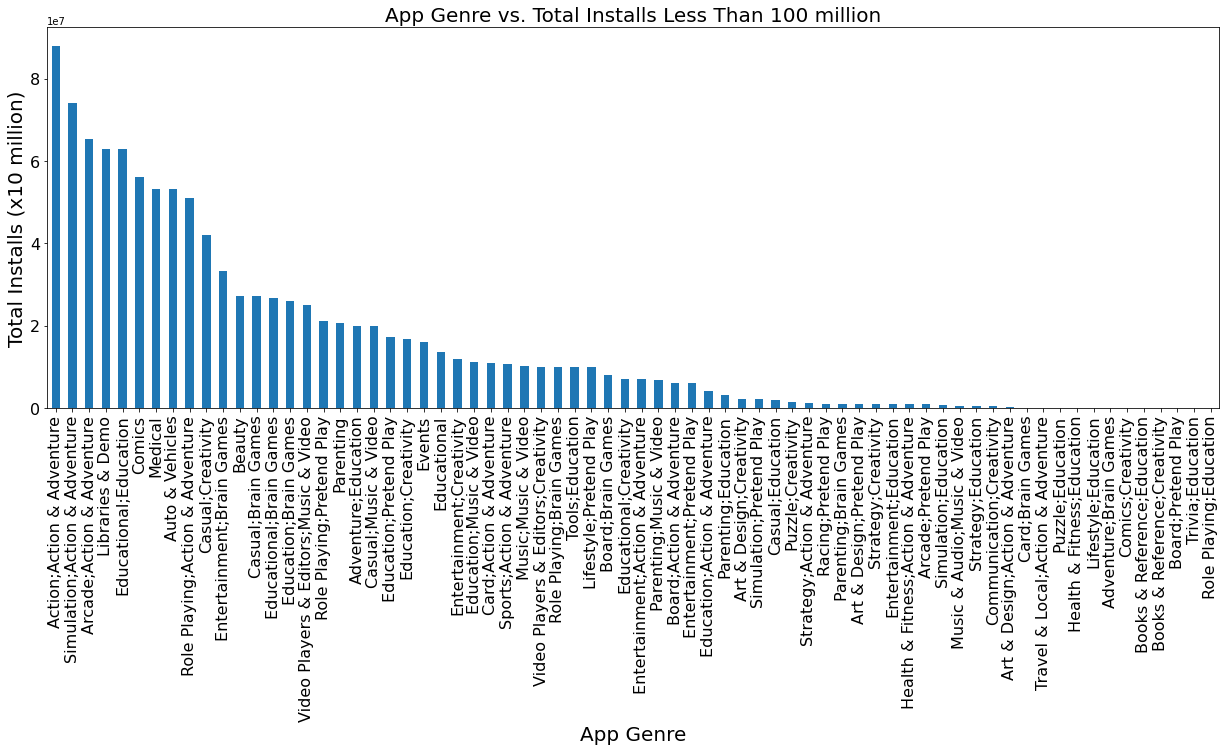
​Data:  
​

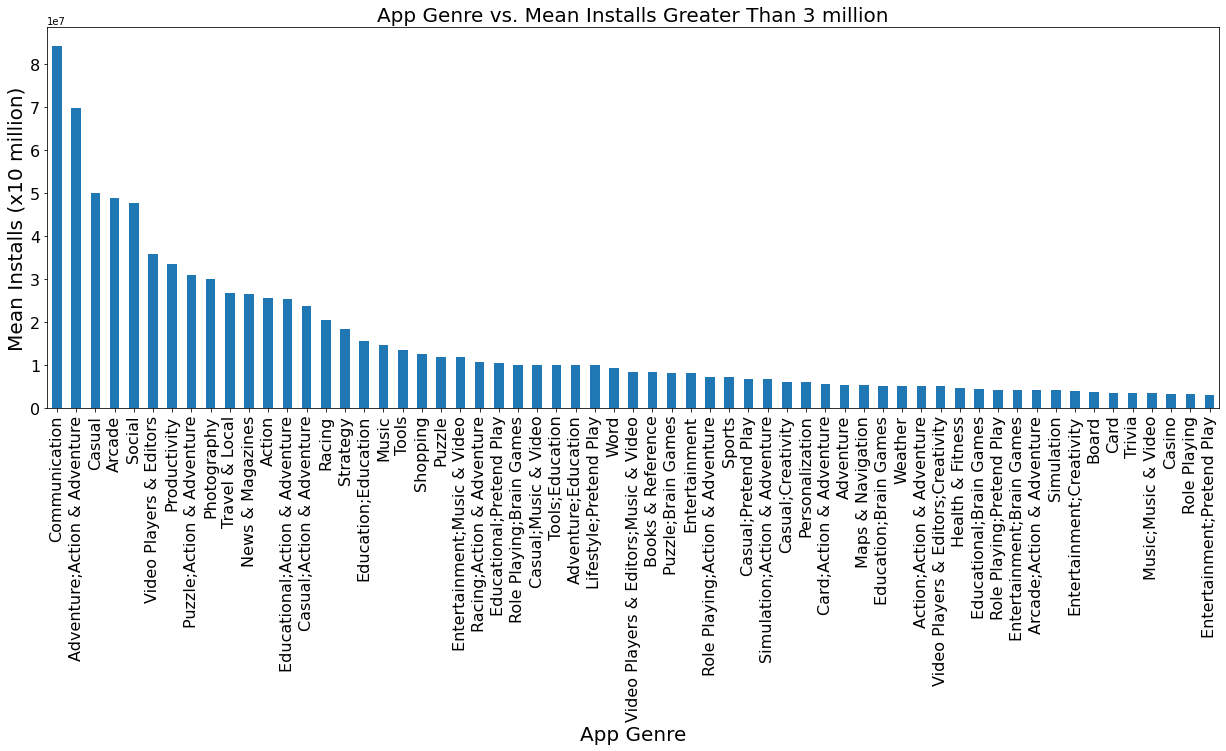
The task of deriving the total and mean installs will utilize two columns; Genres and Installs. The mathematics involved will utilize the sum and means functions. In order to add the values of the Installs column, the column must be converted from the object to the float data type. This requires the stripping of the plus sign from the end of each number in the cell and the removal of the commas using the replace() method for both. I then converted the object data type to the float data type using the astype() method. The Installs column was then checked for unique data types using the unique() method. This indicated all unique numbers except for the value Free, so I used the loc[] method to search the dataframe for this value. Since this was only one record consisting of a relatively low install value (1000), I used the drop() method to remove the record from the dataframe. All data for this story was now cleaned.

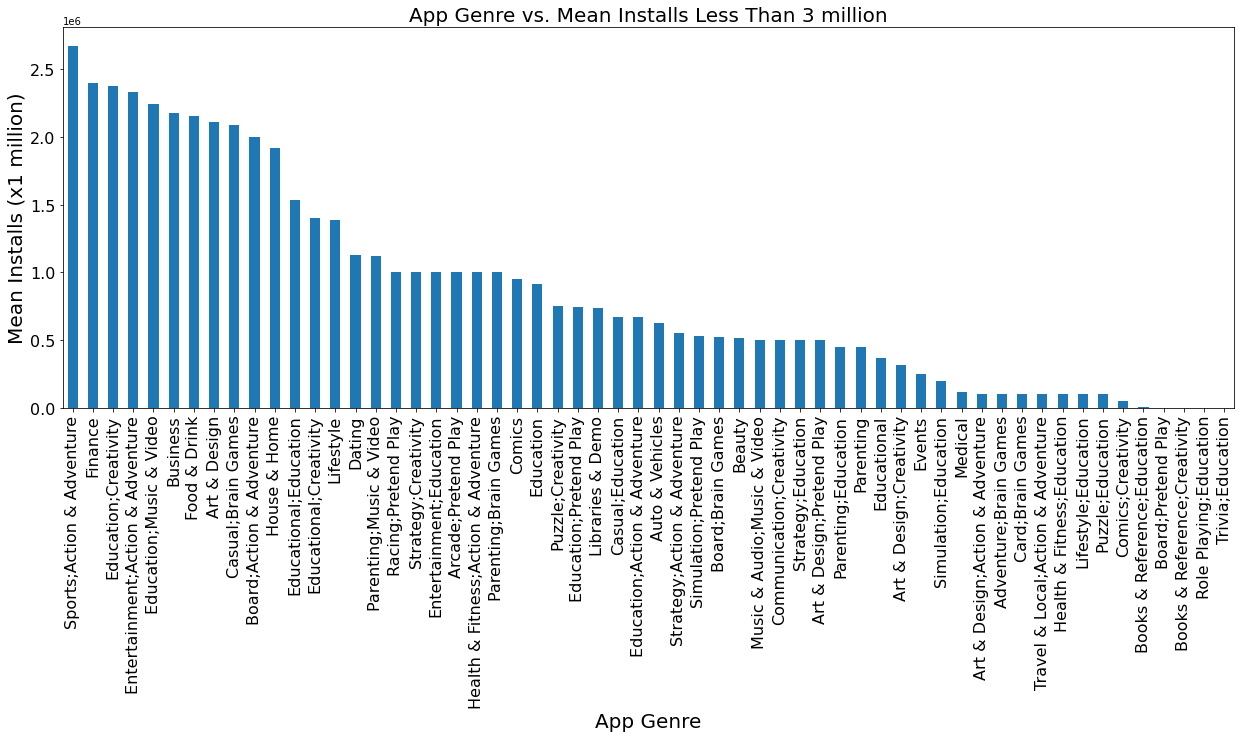
Results:

Visualizations of Total Installs and Mean Installs are presented below. Due to the large amount of Genres, both Total Installs and Mean Installs were split into plots separating each their respective high and low results.









Conclusions:

The above results indicate how many apps belong to each genre along with their mean installs from the Google Play Store. This includes all of the available data except for one ambiguous record which was determined to be erroneously entered. The visualizations may be used as an aid in choosing values for data cleaning during further statistical analysis.