**Google Play Store Report and Analysis: 105-DB-Story#2**

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, “User would like to know how many installs are there are per genre based on the ART\_AND\_DESIGN category. The client would like to see a scatter plot visual of some sort maybe with some interactive features if possible.”

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 a scatter plot of the total installs per each app (Genre) in the ART\_AND\_DESIGN category.

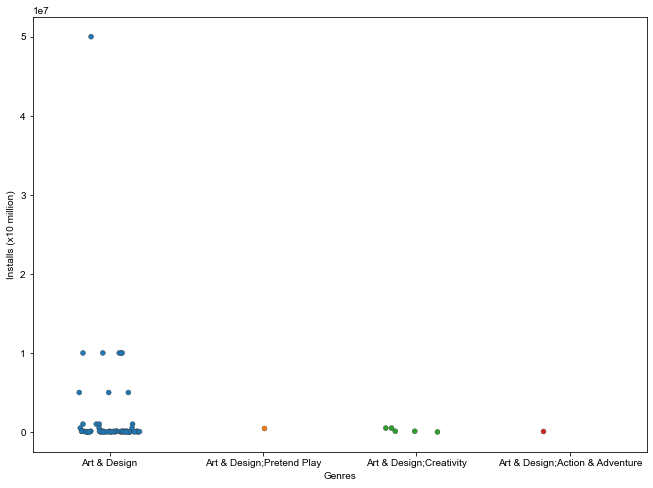
​Data:  
​

The task of presenting a scatterplot for the client’s use involves taking all rows from the category ART\_AND\_DESIGN and setting a dataframe with the columns Genres and Installs. In order to make a scatter plot, the Installs column showing numeric values as a string (i.e, 1,000,000+) must first be converted from the object data type 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. The final task for dataframe construction was to assign a variable to all the records in the category ART\_AND\_DESIGN using the loc[] method. The data for this story was now considered cleaned and ready for plotting.

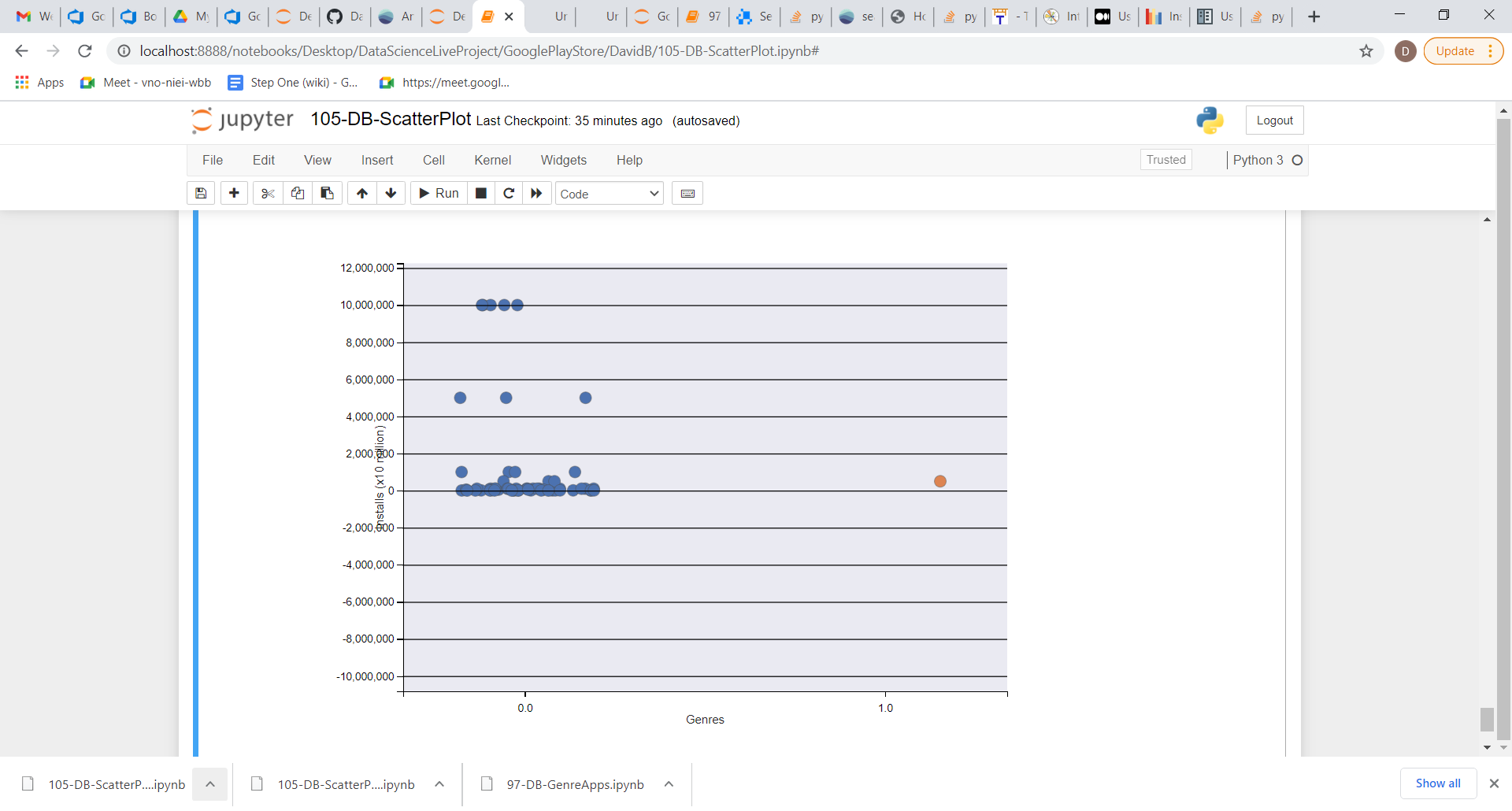
Results:

The stripplot method was used for data visualization of all the Installs contained in the four Genres of ART\_AND\_DESIGN. Due to the large-scale difference in the single highest value for one of the Installs vs. the majority of values for all other Installs, a scatterplot was coded with an interactive command such as the use of a magnifying glass for more detail in order to more accurately determine y-values.

Scatterplot #1 (No Interactive Commands):



Scatterplot #2 (With Interactive Commands):



Conclusions:

The above results indicate both non-interactive and interactive scatterplots of the number of Installs per app Genre from the category ART\_AND\_DESIGN from the Google Play Store. The visualizations may be used as an aid in choosing values for data cleaning during further statistical analysis.