Google Play Store Report and Analysis

1. 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 there are for each category.” 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 hypothesis I will be trying to test in this report will be that the entertainment/gaming categories will be the most installed.

1. Body

Data:

To start off this analysis I initially gathered all of the data into a Data Fram for inspection of rows which will be relevant to the user story and the hypothesis which will be tested. After the initial inspection I noticed numerous rows and columns would be unnecessary for this analysis and thus removed them. I decided I will be needing to keep 2 columns Category, and Installs.

The next step taken was to start cleaning the data and preparing it for later analysis. One of the first things I noticed was an issue in the Installs column. This is a numerical column and would be needed in operations and thus is changed to the integer data type as well as removing the symbols.

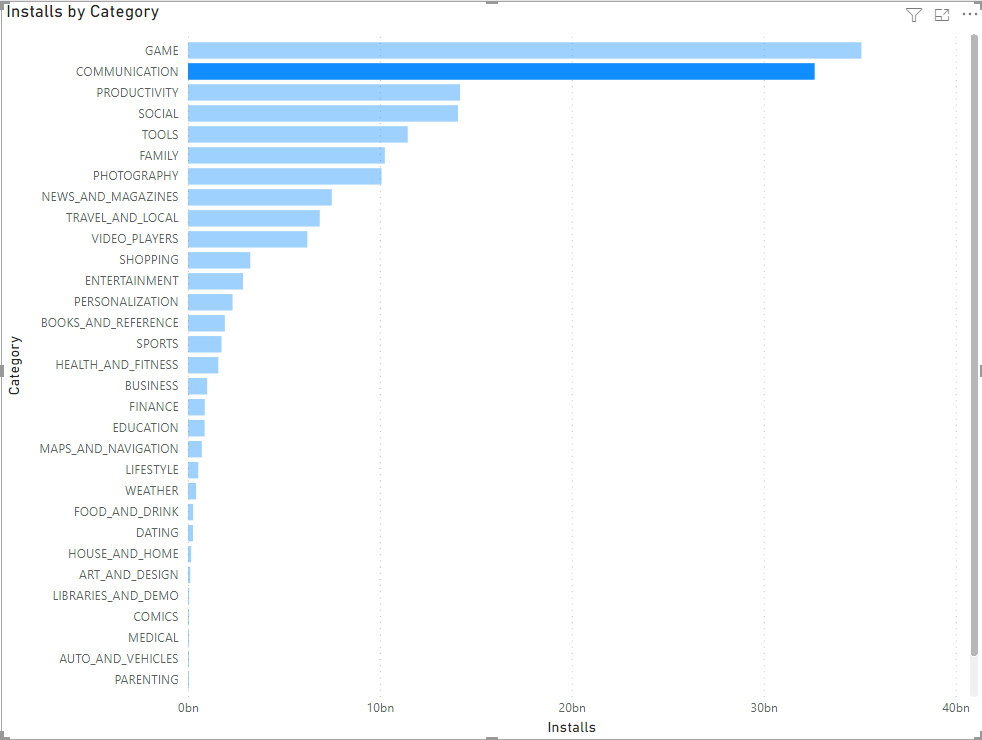
I now have the essential data needed to complete the request of User Story 125. I will discuss in the section below my method for analysis.

Method:

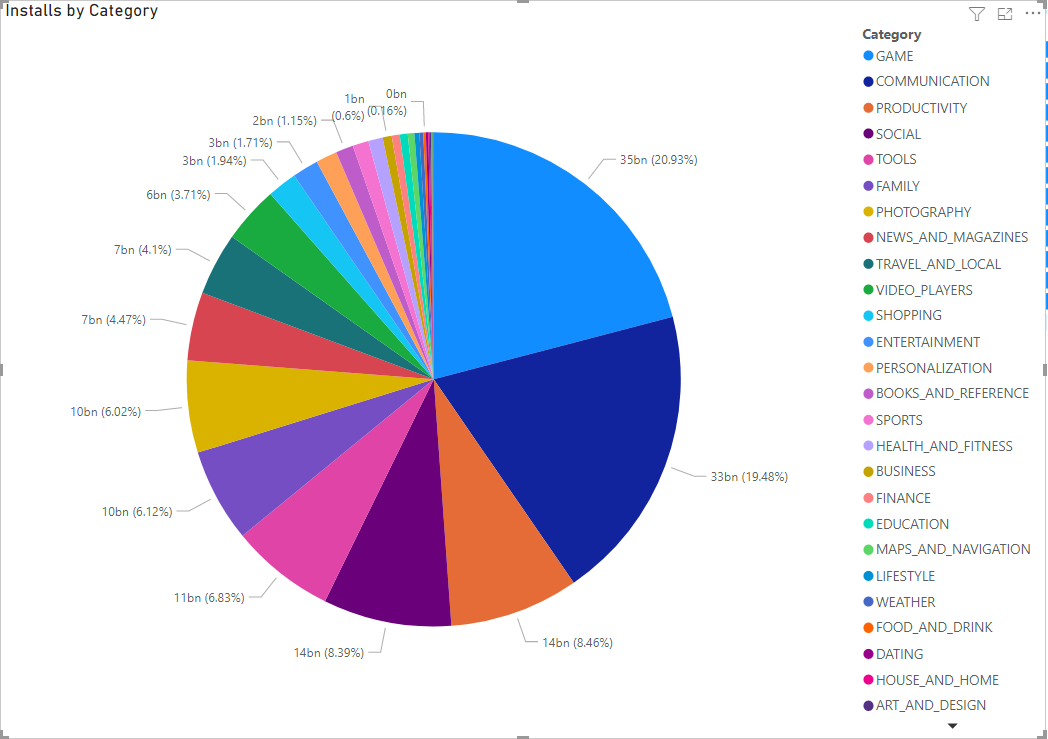
Once I had just the Installs and Category columns it was an easy step to getting the visual we wanted. With the software Power BI we can see a very nice, readable graph made by the Category and Installs columns.

Analysis:

Once the columns were trimmed to the 2 columns in question the visual of Installs per Category really shows the difference for each category.



For a better representation of how much of the overall percentage of installs the game and communication categories have over the others we can use a pie chart.



Results:

As you can see from the previous graphics, the Game and Communication categories have over double the next leading category in installs.

Conclusion:

To conclude this report for number of apps installed per category, the hypothesis “That the entertainment/gaming categories will be the most installed.” will be seen as mostly validated. Although, I will say that in my initial hypothesis I mentioned entertainment as a category and would like to correct that. I had assumed they would be the same category, that was a mistake.