**Introduction:**

This project is based on analyzing the categories on the Google Play Store and identifying if there is a correlation between popularity and category. We will be exploring this data and using multiple different data points in collaboration with categories to make inferences.

**Data:**

The data is the cleaned ‘googleplaystore.csv’ that was wrangled in user-story 1. This data provides the category for each app, installs, ratings, reviews, etc. For this project we will be utilizing the Categories & Ratings columns.

**Method:**

First, we analyzed the number of records in each category, this data is not concatenated and creates a unique key value pair in each row.

A screenshot of a computer

Description automatically generated

Second, we analyzed based on the ratings per category, I found that the dot plot best captured this distribution and showed the density differences on the lower and upper 25% quartiles. This helps visualize which categories perform best and worst based on their ratings.

[Please see next page for dot plot]

, A black and white page with text

Description automatically generated

Lastly, we looked at the distribution of total amount of installs by category, this was done by summing the number of installs by each category. This is depicted by a bar chart which shows an obvious disparity showing that Communication & Game perform well with the total amount of installations.

A graph of a number of installs

Description automatically generated A white text on a white background

Description automatically generated

**Analysis:**

Upon my analysis I found that although ‘Family’ was the most popular category by app developers, education ranked high in ratings and was generally found to have an exceptional distribution showing a dense population over 4.0.

However, Communication and Gaming were most popular in terms of total downloads dauting over thirty billion installations across each individual category. Education and family did not exceed well in this category showing over a 70% difference from Communication and Gaming.

**Conclusion:**

Based on my findings I can reasonably assume that achieving installs, ratings, or number of apps does not correlate to success in other categories. This can be seen with the high number of ‘Family” applications on the market but showing only 30% of the installations of the most popular categories ‘Communication’ and ‘Gaming’. This is also shown on the dot plot showing density above 4.0 for ‘Education’ and ‘Family’ showing a much more even distribution indicating presence of low values.