



# Google Play Store Apps

Dylan Creeden  
Edward Lee  
Freddy Lopez  
Islom Mashanlo  
Joanne Trac

Group 3 | CIS 9650 | Spring 2022 | Professor Tseng

# About the Dataset

2.3M+ records of Android Apps that were available on the Google Play Store

Author: Gautham Prakash

Published: June 2021

Source: Kaggle.com

24 attributes:

- App Name
- Category
- Rating
- Price
- Release Date
- Installs, etc.

# Data Cleaning

Cleansed the dataset based on Price column using z-scores

1. Calculated standard deviation and mean on Price column for all Paid apps
2. If z-score for app was greater than 4 standard deviations, we removed said app due to its price being an outlier. 296 rows were removed
3. Dataset was already formatted properly (i.e. dates, uppercase)
4. Dropped unnecessary columns ('app\_id', 'minimum\_android', 'developer\_id', 'developer\_website', 'developer\_email', 'privacy\_policy', 'ad\_supported', 'in\_app\_purchases', 'editors\_choice')
5. Created two new dataframes, Paid apps and Free apps

# Paid Apps vs. Free Apps

- Rating Comparison
- If Rating is positive, Paid apps have a higher average rating compare to Free
- Paid apps have a higher rating in 29 categories
- Free apps have a higher rating in 19 categories
- Largest difference in favor of Paid apps:
  - Category: Libraries & Demo (0.98)
- Largest difference in favor of Free apps:
  - Category: Dating (-1.01)

Categories 1 -24 Rating Diff.

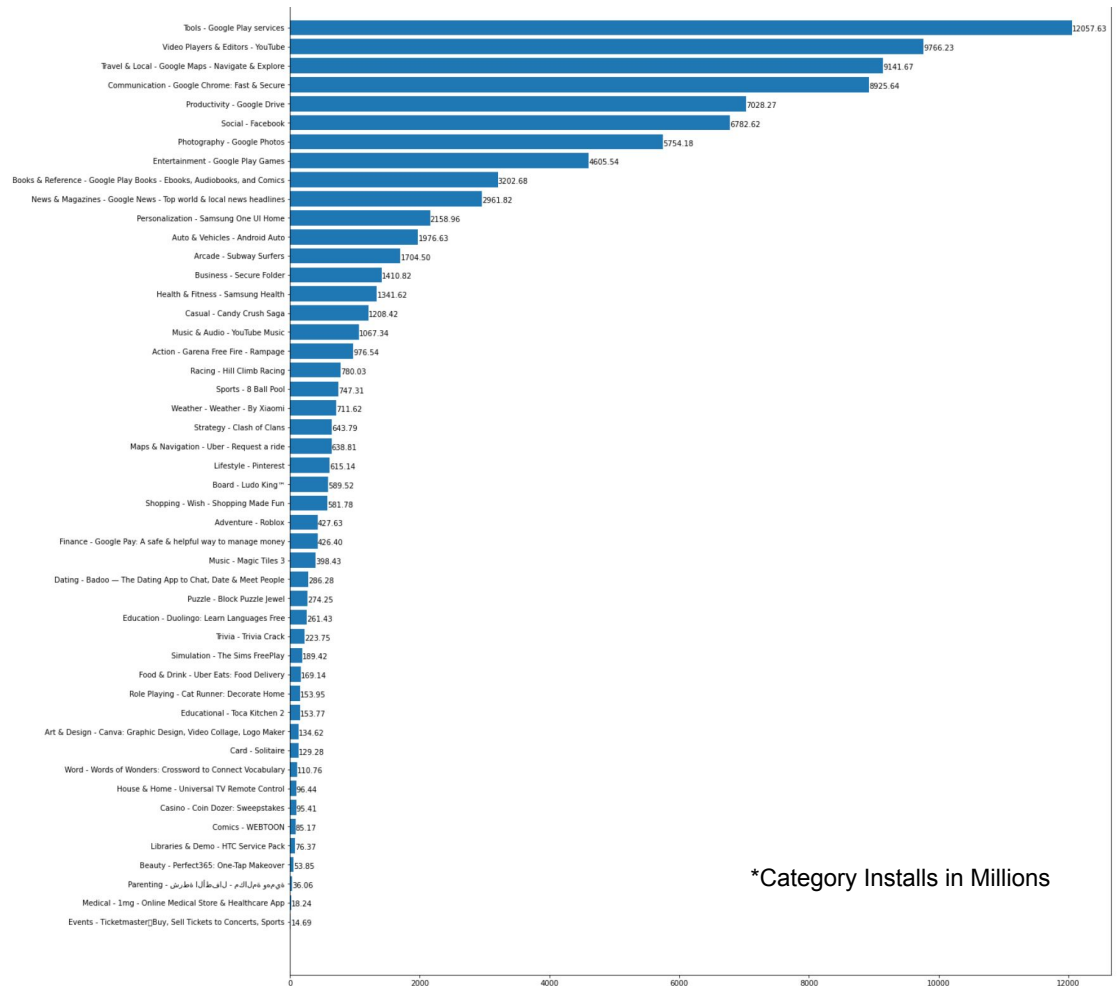
Personalization	0.229025
Weather	0.198635
Beauty	0.116033
Lifestyle	0.056867
Casual	0.032995
Sports	-0.002803
Finance	-0.100991
Trivia	-0.116868
Medical	-0.154765
Entertainment	-0.181288
Video Players & Editors	-0.199423
Word	-0.202903
News & Magazines	-0.284533
Social	-0.368908
Parenting	-0.460780
Music	-0.467896
Events	-0.492782
Education	-0.568865
Casino	-0.585557
Educational	-0.613515
Maps & Navigation	-0.681899
Travel & Local	-0.774589
Books & Reference	-0.829757
Dating	-1.014935

Categories 25 - 48 Rating Diff.

Libraries & Demo	0.980463
Strategy	0.826994
Adventure	0.768957
Comics	0.731763
Health & Fitness	0.675203
Productivity	0.657781
Business	0.600905
Photography	0.564362
Board	0.542351
Role Playing	0.509229
Action	0.434429
Auto & Vehicles	0.409278
Communication	0.401907
Puzzle	0.390329
Music & Audio	0.386320
House & Home	0.385108
Shopping	0.350888
Food & Drink	0.344964
Arcade	0.321371
Tools	0.257784
Art & Design	0.251903
Racing	0.243681
Card	0.231080
Simulation	0.230646

# Installs by category

- Highest Install Category:
  - Tools - Google Play Services (12+ Billion)
- Lowest Install Category:
  - Events - Ticketmaster, Sports, Concerts (14+ Million)



\*Category Installs in Millions

# Average App Price By Category

- Max. average price: \$14.41
  - Medical apps
  - Business apps
  - Dating apps
- Min. average price: \$1.58
  - Personalization apps
  - Trivia apps
  - Arcade

