Google Play Store
App Install Analysis

Team Members:
Dimitrije Adzic
Taylor Thompson
Travis Grogan
Yan Sun
Yang Song
Yael Yana Reinus

Problem Statement

How can app developers focus their times working on apps in the most fruitful categories?

Data

We used the Google Play Store Apps dataset which is web-scraped data on each application in the store

Each row in the dataset represents an app and has values for category, rating, size, price, and more

https://www.kaggle.com/lava18/google-play-store-apps?select=googleplaystore.csv

DATA CLEANING

1. Dropped the "na" values
2. Replaced "$" and ' '
3. Changed price on some apps from string to float items

About Data

Index	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Bible du Semeur-BDS (French)	BOOKS_AND_REFERENCE	4.5	313	6.9M	50000	Free	0	Everyone	Books & …	August 4, …	0.7	1.0 and up
1	Tamilnadu Electricity Info	TOOLS	3.6	216	811k	10000	Free	0	Everyone	Tools	November 9…	2.0	1.0 and up
2	DC Metro	MAPS_AND_NAVIGATION	3.4	59	1.3M	10000	Free	0	Everyone	Maps & N…	January 26…	1.0	1.5 and up
3	BNCR Token Celular	FINANCE	3.7	913	70k	100000	Free	0	Everyone	Finance	July 20, 2…	1.0	1.5 and up
4	DS Tower Defence	GAME	3.2	768	1.4M	100000	Free	0	Everyone	Arcade	June 5, 20…	1.0	1.5 and up
5	4 in a row	GAME	4.3	47698	4.1M	5e+06	Free	0	Everyone	Board	April 12, …	1.0.21	1.5 and up
6	Age of Conquest: N. America	FAMILY	4.5	659	3.3M	10000	Paid	$3.…	Everyone	Strategy	August 4, …	1.0.28	1.5 and up
7	Active Soccer 2 DX	SPORTS	3.6	28	48M	500	Paid	$2.…	Everyone	Sports	March 28, …	1.0.3	1.5 and up
8	.R	TOOLS	4.5	259	203k	10000	Free	0	Everyone	Tools	September …	1.1.06	1.5 and up
9	AH Alarm Panel	TOOLS	3.9	7	81k	100	Paid	$4.…	Everyone	Tools	May 12, 20…	1.2.0	1.5 and up
10	CS16Client	GAME	4.3	8668	9.1M	500000	Free	0	Teen	Action	April 1, 2…	1.33	1.5 and up

Analysis & Libraries

Analysis:

- Linear regression
- Summary Statistics
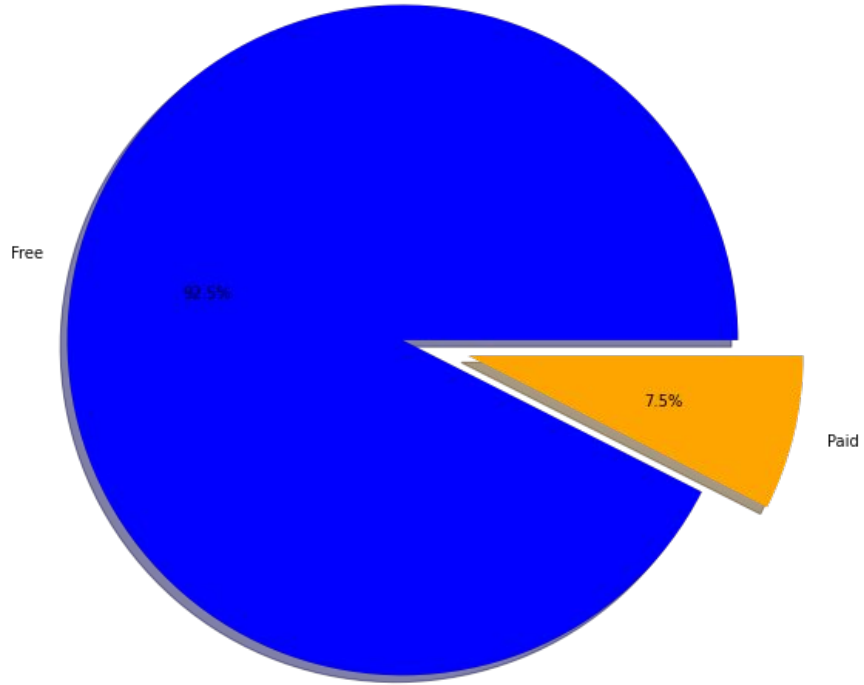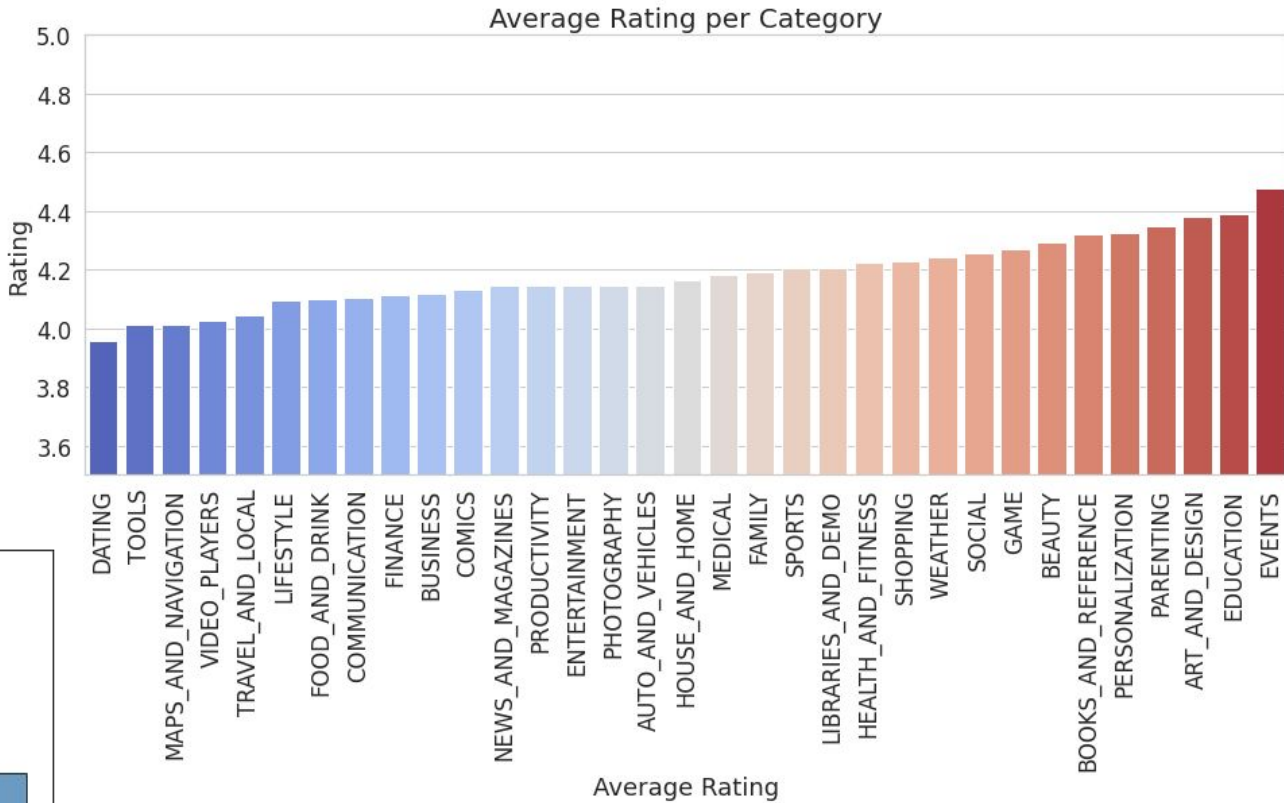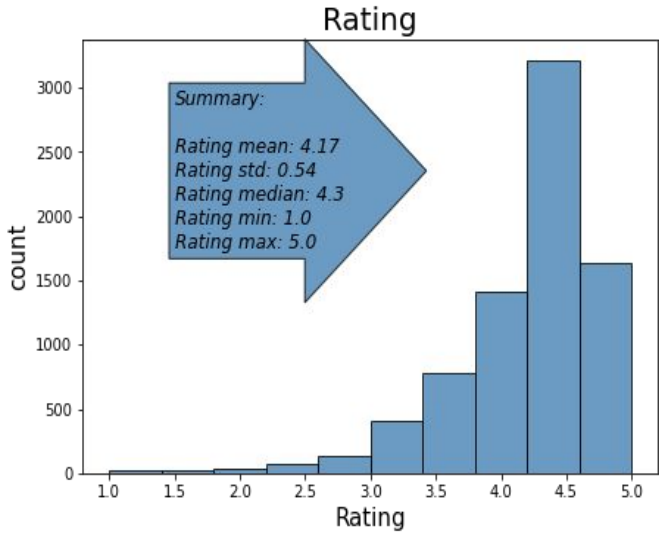
Libraries:

- Pandas
- Numpy
- Matplotlib
- Seaborn

FREE VS PAID

VAST MAJORITY OF APPS ARE FREE

Percent of Free Vs Paid Apps in store

Free 92.5%

Paid 7.5%

APPS WITH RATINGS

Average Rating per Category

Rating

Categories (left to right): DATING, TOOLS, MAPS_AND_NAVIGATION, VIDEO_PLAYERS, TRAVEL_AND_LOCAL, LIFESTYLE, FOOD_AND_DRINK, COMMUNICATION, FINANCE, BUSINESS, COMICS, NEWS_AND_MAGAZINES, PRODUCTIVITY, ENTERTAINMENT, PHOTOGRAPHY, AUTO_AND_VEHICLES, HOUSE_AND_HOME, MEDICAL, FAMILY, SPORTS, LIBRARIES_AND_DEMO, HEALTH_AND_FITNESS, SHOPPING, WEATHER, SOCIAL, GAME, BEAUTY, BOOKS_AND_REFERENCE, PERSONALIZATION, PARENTING, ART_AND_DESIGN, EDUCATION, EVENTS

Average Rating

Rating

Summary:

Rating mean: 4.17
Rating std: 0.54
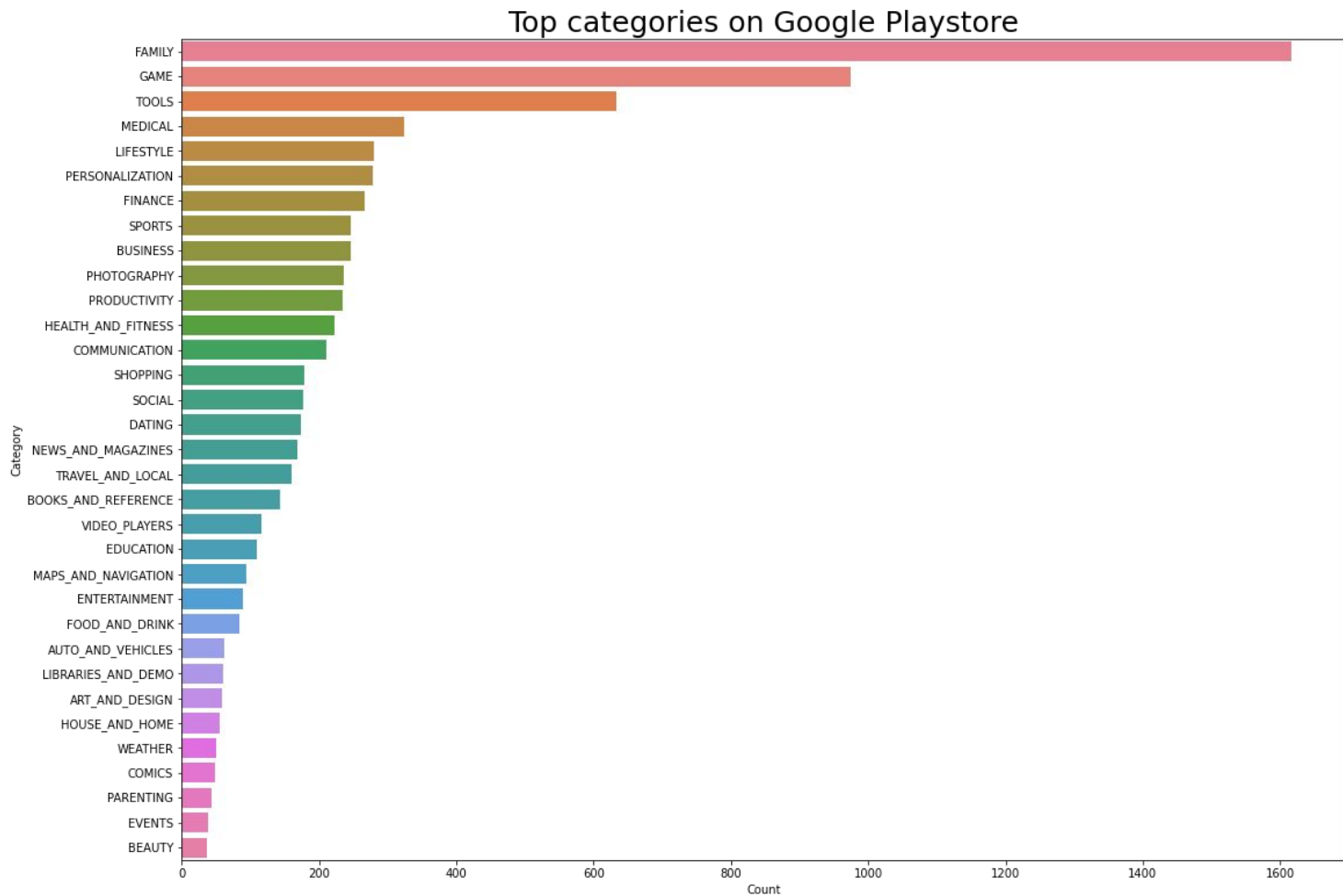Rating median: 4.3
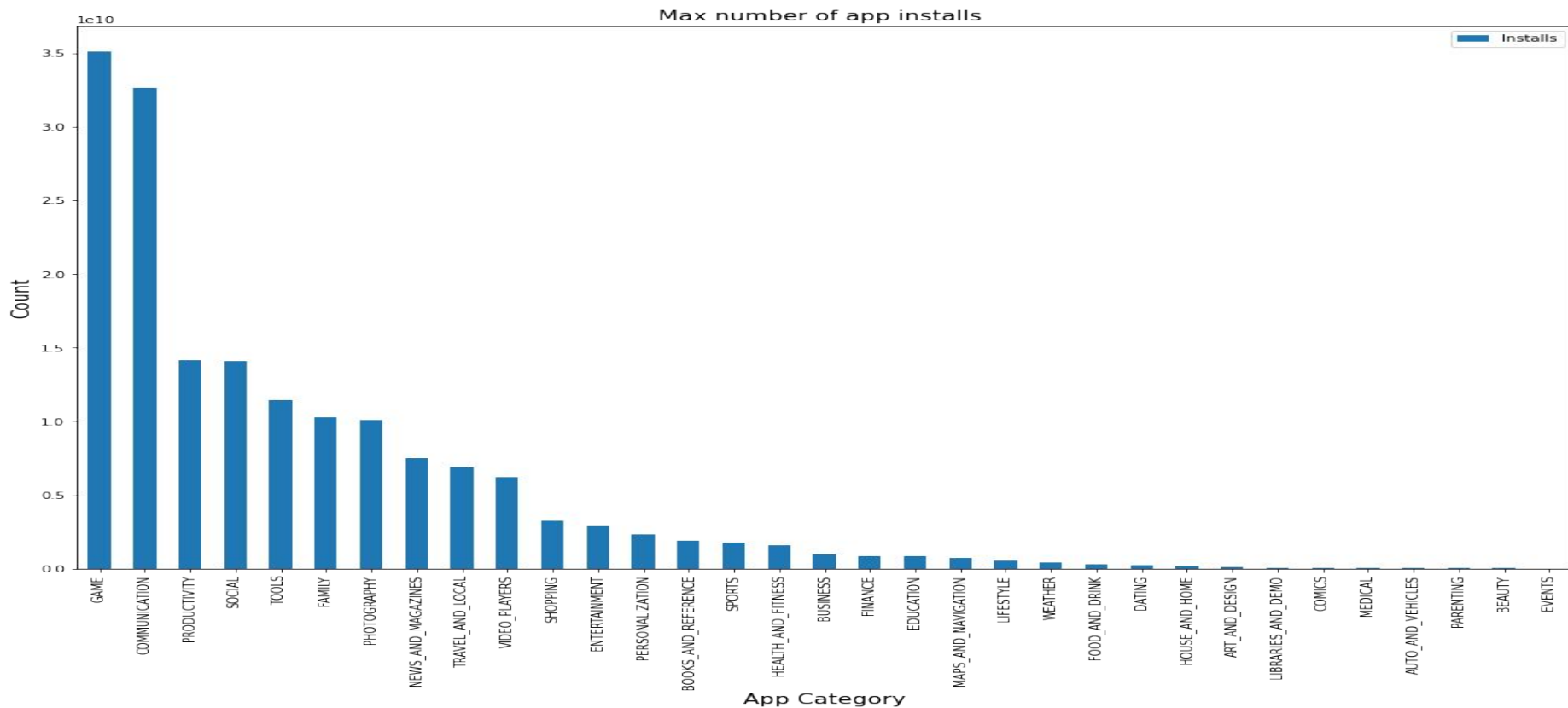Rating min: 1.0
Rating max: 5.0

How App Price Impacts Ratings

- Most expensive apps =/ the highest rating

- Free apps receive higher ratings and more of them
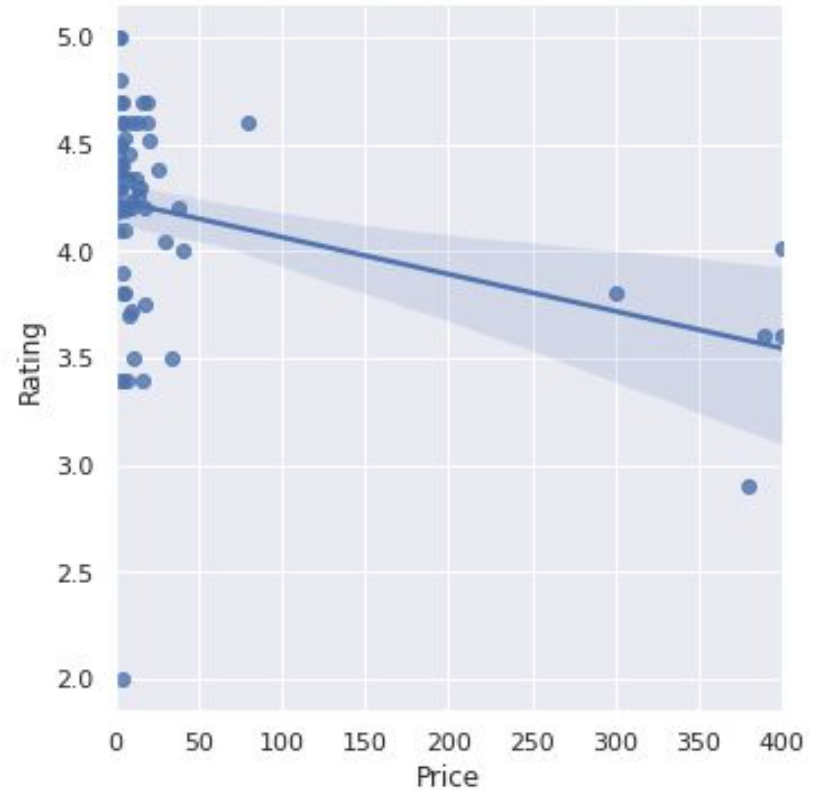
TOP 5 Categories:
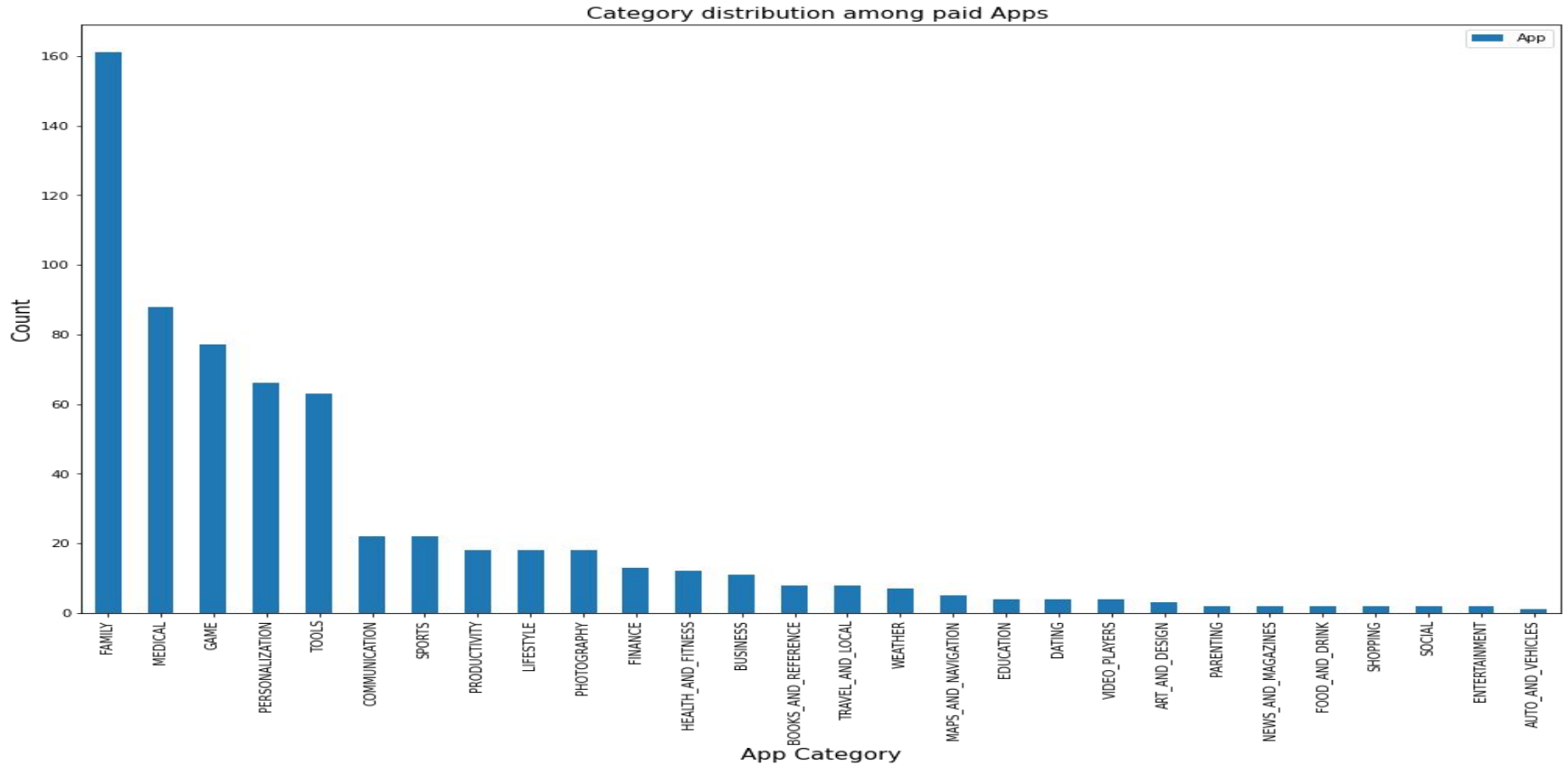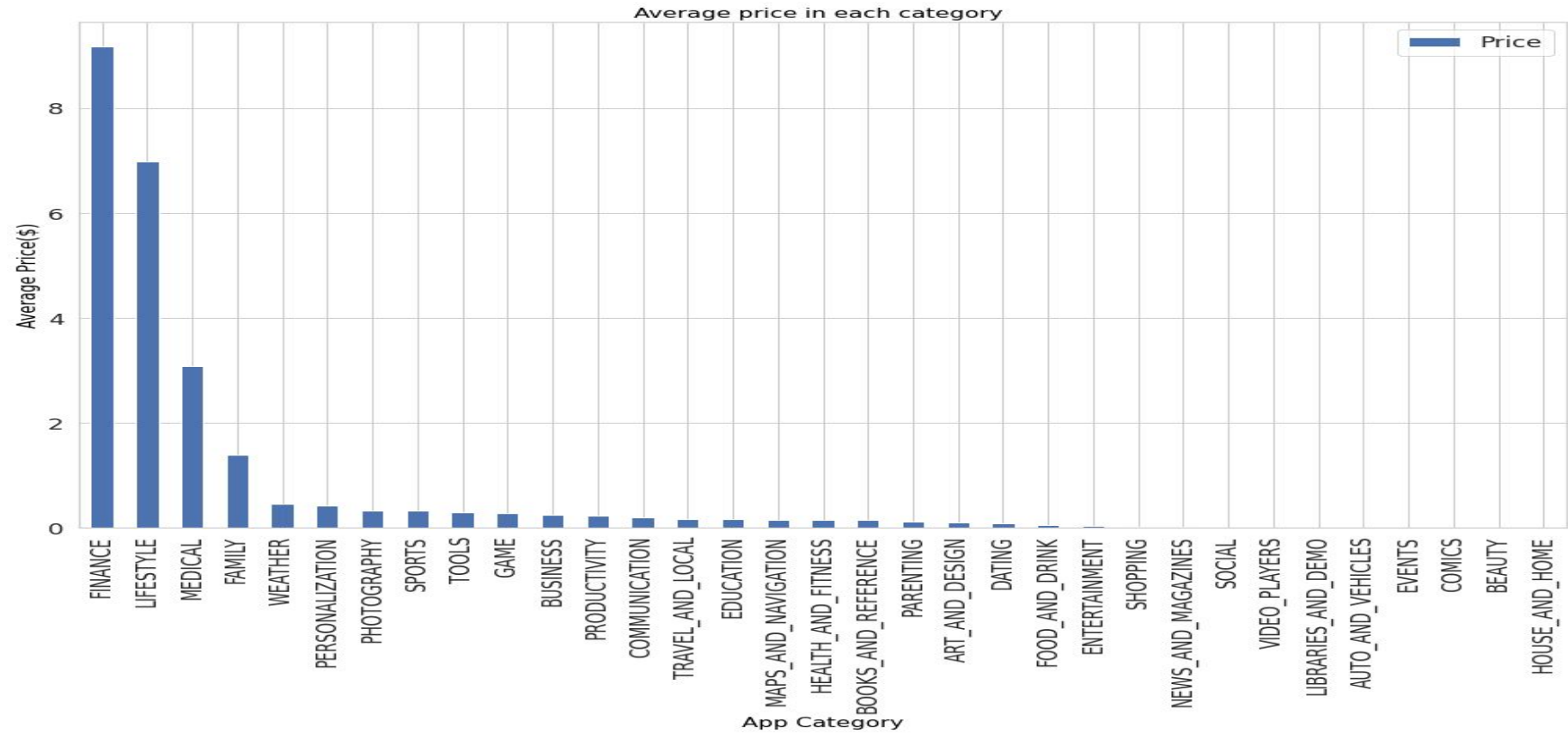Family
Game
Tools
Medical
Lifestyle

Top categories on Google Playstore

Installed Apps(Category distribution)

Max number of app installs

Paid Apps(Category distribution)

Category distribution among paid Apps

Average price in each category

Average price in each category

ADVICE TO DEVELOPERS

- Price greatly impacts app ratings

- Financial Apps Focus

- Potential to "Gamify" other categories

Challenges

- Doesn't Account for In-App Purchases/ "Freemium" business model
    - ○
- Outdated dataset