Google play store Category Analysis

AJITH C

2025-07-21

Problem

The majority of apps on the Google Play Store are Free, while Paid apps struggle with lower downloads. Developers often release Paid apps without understanding category-wise demand and user behavior. This imbalance leads to low visibility and financial losses for developers of Paid apps.

Solution

- Analyze category-wise distribution of Free and Paid apps.
- Study average installs and ratings for both Free and Paid apps across categories.
- Identify categories where Paid apps still perform well despite being fewer in number.
- Provide recommendations on which categories are suitable for Paid app launches.

ASK - Business Challenge

Problem

Imbalance between Free and Paid Apps

Questions

- Why is there a significant imbalance between Free and Paid apps in the Google Play Store?
- Why do developers often prefer to release Free apps rather than Paid apps?
- Why do Paid apps tend to have lower install counts compared to Free apps?
- Why do some Paid apps perform better in certain categories despite lower overall installs?
- Why is it important for developers to analyze category trends before deciding on app pricing models?

PREPARE - Data collection and Sources

Dataset Source:

Kaggle — Google Play Store Apps Dataset (https://www.kaggle.com/datasets/lava18/google-play-store-apps)

Key Attributes Used:

`App` , `Category` , `Rating` , `Reviews` , `Installs` , `Type` , `Price` , `Content Rating`

PROCESS - Data cleaning and preparation

- Removed 483 duplicates.
- Removed Removed 1 row with invalid `Rating` 19.
- Handled missing values in `Current ver` column.
- Converted `Installs` to numerical format (removed commas and plus signs).
- Filtered apps with realistic install counts and valid categories.

| | Case study steps | ✕ | | Untitled Project - Posit Cloud | ✕ | + |

kP1daU1Dek7U4T0zl1EeVBPT67ysbo/edit?gid=664040273#gid=664040273

s   Help

efaul... ▼   — 10 +   B *I* ⇔ A   ◇.   ⊞ ⁵⁵ ▼   ☰ ▼ ↓ ▼ ↦ ▼ A ▼

	E	F	G	H	I	J	L
ze	Installs	Type	Price	Content Rating	Genres	
)M	10,000+	Free	0	Everyone	Art & Design	J
IM	500,000+	Free	0	Everyone	Art & Design;Pre	J
7M	5,000,000+	Free	0	Everyone	Art & Design	
5M	50,000,000+	Free	0	Teen	Art & Design	
8M	100,000+	Free	0	Everyone	Art & Design;Cre	
6M	50,000+	Free	0	Everyone	Art & Design	N
)M						
)M						
3M						S
1M						€
3M						C
2M						
)M						
IM						
7M						
7M						
5M	500,000+	Free	0	Everyone	Art & Design	
7M	10,000+	Free	0	Everyone	Art & Design	N
)M	5,000,000+	Free	0	Everyone	Art & Design	
IM	10,000,000+	Free	0	Everyone	Art & Design	
IM	100,000+	Free	0	Everyone	Art & Design	
2M	100,000+	Free	0	Everyone	Art & Design	N
2M	500,000+	Free	0	Everyone 10+	Art & Design	
0M	100,000+	Free	0	Everyone	Art & Design;Act	
3M	50,000+	Free	0	Everyone	Art & Design	
0M	10,000+	Free	0	Everyone	Art & Design	
5M	500,000+	Free	0	Everyone	Art & Design;Cre	C
1M	100,000+	Free	0	Everyone	Art & Design	N

Remove duplicates   ✕

483 duplicate rows found and removed.

10359 unique rows remain.

OK

Q Search

Duplicates

Google Data Analytics Capstone ✕ Google_Playstore_AJITH_C

docs.google.com/spreadsheets/d/1hXf7la9qfQY

Google_Playstore_AJITH_C ☆ 🗔 ☁

File   Edit   View   Insert   Format   Data   Tools   Extensic

🔍 ↺ ↻ 🖨 🖌 100% ▾ | $ % .0↙ .0↗0 123 |

C9992   ▾ | ƒx 19

	A	B	C	D
1	**App** ⇌	**Category** ⇌	**Rating** 🔽	**Reviews** ⇌
9992	Life Made WI-Fi	1.9	19	3.0M
10360				
10361				
10362				
10363				
10364				
10365				
10366				
10367				
10368				
10369				
10370				
10371				
10372				
10373				
10374				
10375				
10376				
10377				
10378				
10379				
10380				
10381				
10382				
10383				
10384				
10385				
10386				

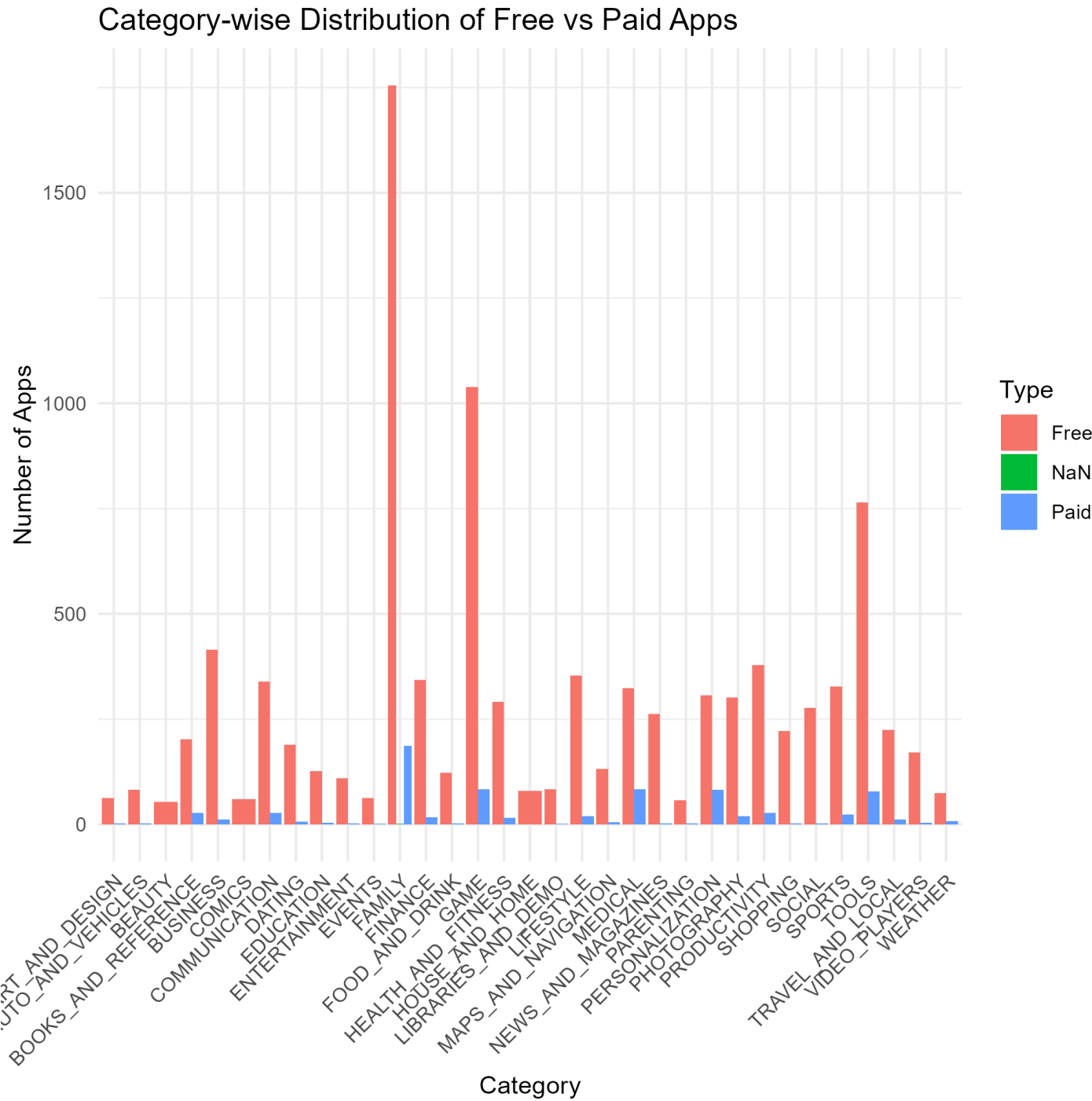➕ ☰   **googleplaystore** ▾

Ratings

ANALYZE - Data Exploration and Findings
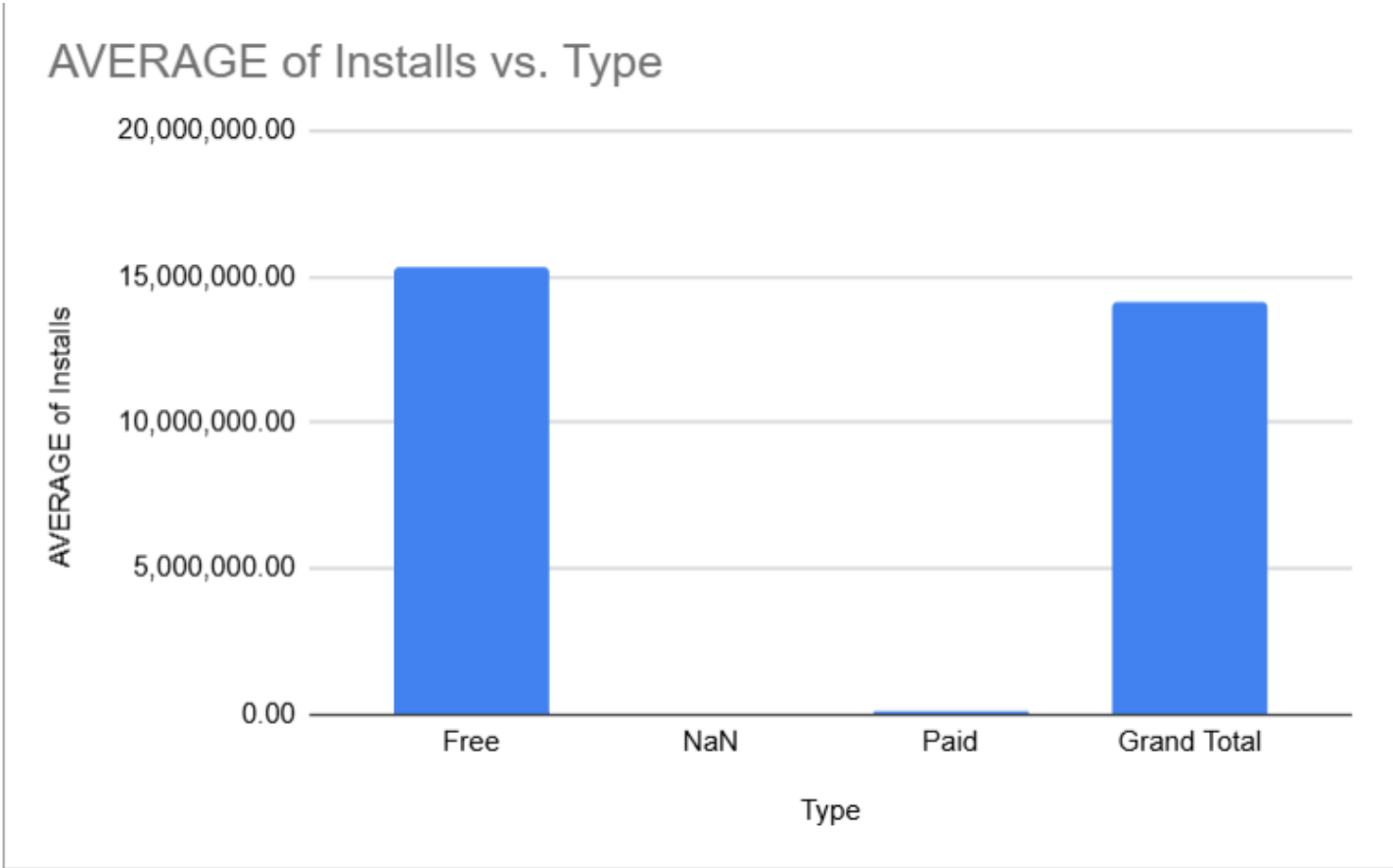
Category-wise Distribution of Free vs Paid Apps

Category-wise Distribution of Free vs Paid Apps

Graph

- **Observation**: Free apps dominate most categories.
- **Exception**: Some categories have a noticeable share of Paid apps.
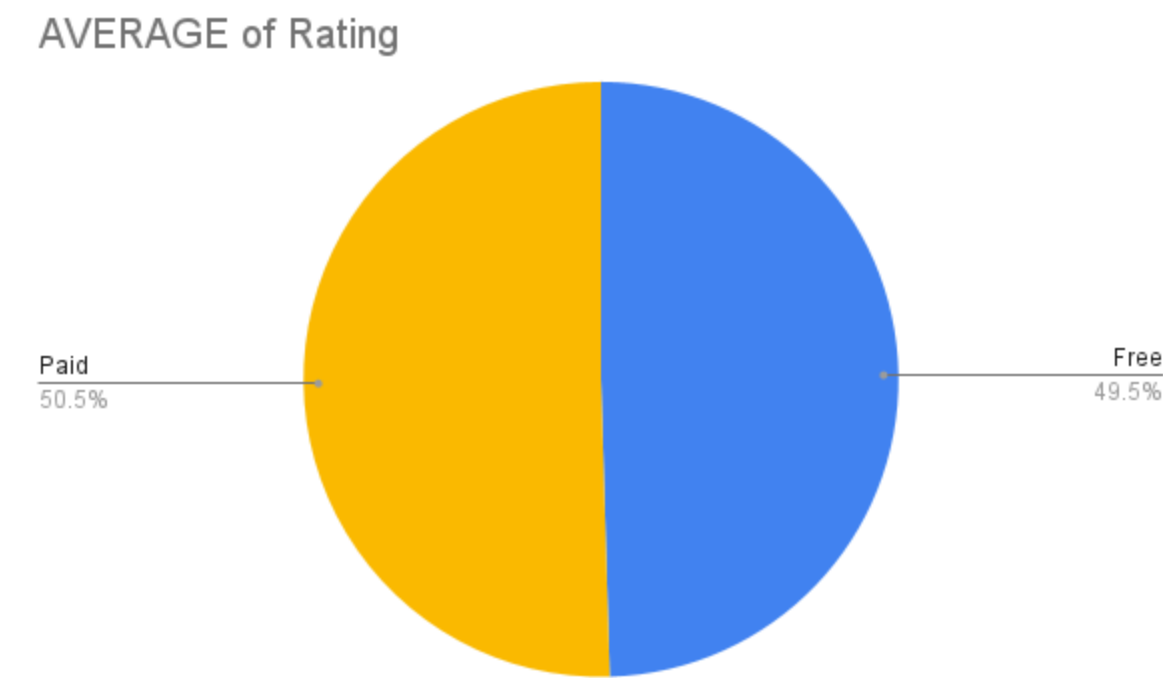
Average Installs: Free vs Paid Apps

Average
Observation

- Free apps have significantly higher average installs overall.
- Paid apps see better installs in niche categories

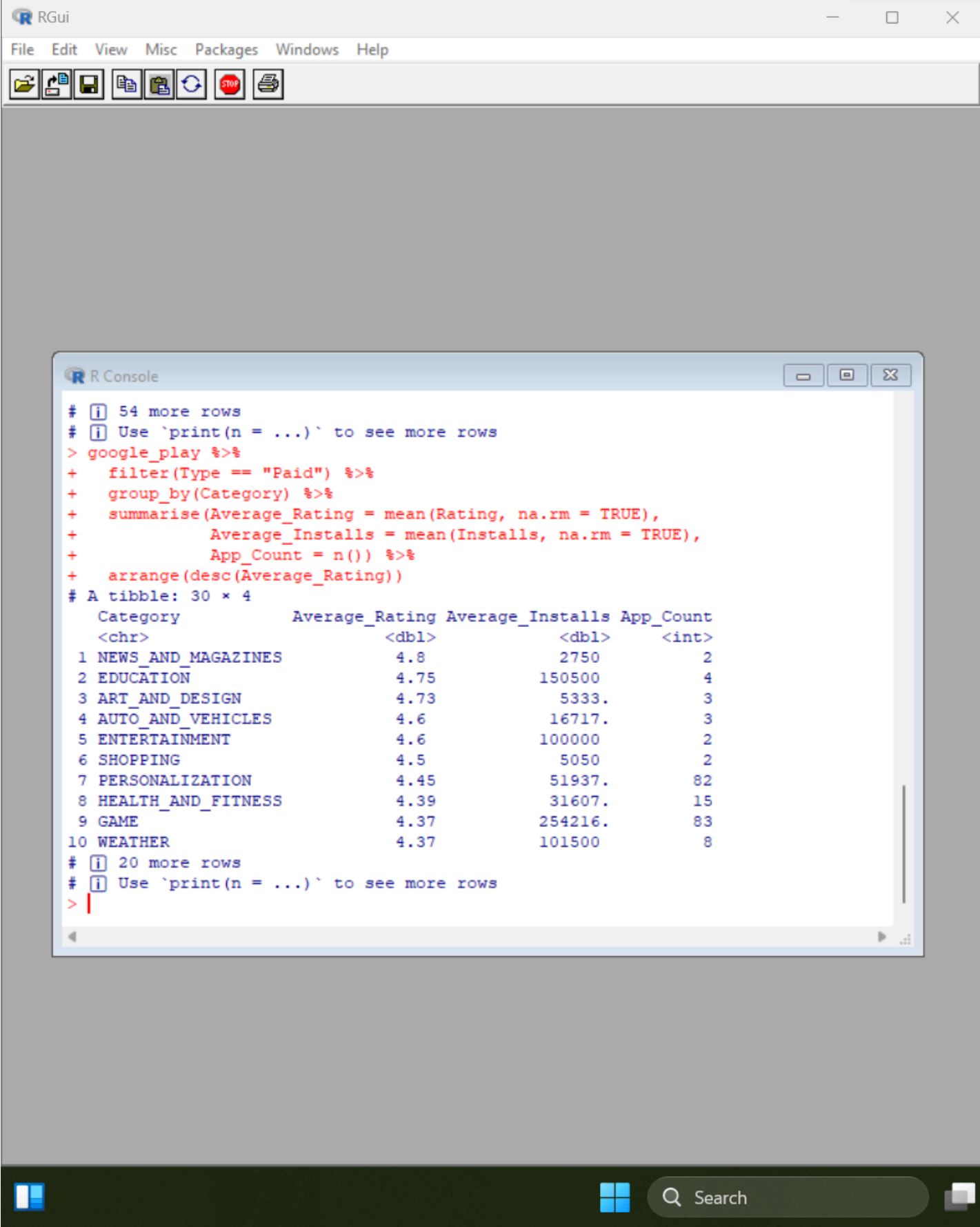## Average Ratings: Free vs Paid Apps

Average
Observation

- Paid apps tend to have slightly higher average ratings in selected categories.

SHARE — Insights & Recommendations

- Categories like **NEWS_AND_MAGAZINES**, **EDUCATION**, and **ART_AND_DESIGN** have the highest average ratings among Paid apps, suggesting they are well-received by users despite having fewer apps.
- This indicates that Paid apps can succeed in **niche categories** where specialized content or quality features are valued.
- On the other hand, categories like **GAME** and **PERSONALIZATION** have a large number of Paid apps and high average installs, but slightly lower ratings — indicating high competition.

- Developers should prioritize **Free or Premium models** for competitive categories like **Games**, while targeting **Paid models** for niche categories like **Education** and **News & Magazines**.
- Analyzing category-specific trends helps developers make informed pricing decisions and optimize their app launch strategies.


```
#  ℹ 54 more rows
#  ℹ Use `print(n = ...)` to see more rows
> google_play %>%
+   filter(Type == "Paid") %>%
+   group_by(Category) %>%
+   summarise(Average_Rating = mean(Rating, na.rm = TRUE),
+             Average_Installs = mean(Installs, na.rm = TRUE),
+             App_Count = n()) %>%
+   arrange(desc(Average_Rating))
# A tibble: 30 × 4
   Category          Average_Rating Average_Installs App_Count
   <chr>                      <dbl>            <dbl>     <int>
 1 NEWS_AND_MAGAZINES          4.8              2750         2
 2 EDUCATION                   4.75           150500         4
 3 ART_AND_DESIGN              4.73             5333.        3
 4 AUTO_AND_VEHICLES           4.6             16717.        3
 5 ENTERTAINMENT               4.6            100000         2
 6 SHOPPING                    4.5              5050         2
 7 PERSONALIZATION             4.45            51937.       82
 8 HEALTH_AND_FITNESS          4.39            31607.       15
 9 GAME                        4.37           254216.       83
10 WEATHER                     4.37           101500         8
#  ℹ 20 more rows
#  ℹ Use `print(n = ...)` to see more rows
>
```

Average

ACT — Business Action Plan

For High-Competition Categories like games,entertainment:

- Focus on Free apps or Free + premium models to attract users in crowded categories.
- Use in-app purchases or ads for how you make money instead of upfront payment.
- Differentiate your app through unique features or better user experience.

For Niche Categories like education,News:

- Consider launching Paid apps since users in these categories show higher acceptance of premium apps.
- Focus on delivering high-quality, specialized content or features that justify the price.
- Highlight ratings and user satisfaction in marketing strategies.

Overall Strategic Actions:

- Conduct category analysis before deciding app pricing.
- Monitor category performance periodically to adapt business strategy.
- Use insights from data to make data-driven decisions for future app launches.

Tools Used

- R
- Spreadsheets
- Tableau
- SQL

Connect with me

Linkedin (www.linkedin.com/in/ajith-c-073a45332)

GitHub (https://github.com/Ajith-coder777)