

# Google Play Store Data Analysis Report

---

## Google Play Store Data Analysis Report

### Problem Statement

The Google Play Store hosts over **2.56 million apps**, impacting users' lives in areas like education, health, finance, and entertainment. Understanding trends in this app ecosystem can help developers and marketers make informed decisions.

---

### Objective

- Identify the **most popular app category**
  - Find the app with the **highest installs**
  - Detect the **app with the largest size**
  - Understand **user rating patterns**
  - Perform **feature classification and correlation** to reveal trends
- 

### Dataset Information

- **Total Rows:** 10,841

- **Total Columns:** 13 (after cleanup)

- Dataset includes attributes like:

- App Name
- Category
- Rating
- Reviews
- Size
- Installs
- Type (Free/Paid)
- Price
- Content Rating
- Genres
- Last Updated

---

## Steps Followed

### 1 Data Cleaning

- Converted Size, Installs, Price into numerical format
- Handled null values and type mismatches
- Extracted Day, Month, and Year from Last Updated

### 2 Feature Engineering

- Separated **numerical features**:  
Rating, Reviews, Size, Installs, Price, Day, Month, Year
  - And **categorical features**:  
App, Category, Type, Content Rating, Genres, Current Ver, Android Ver
- 

## Exploratory Data Analysis (EDA)

### Univariate Analysis

- KDE plots for numerical features showed:
  - Ratings mostly concentrated between **3.5 to 4.5**
  - Reviews and installs are **heavily right-skewed**
  - Most apps are **free** and rated **for Everyone**

### Categorical Feature Analysis

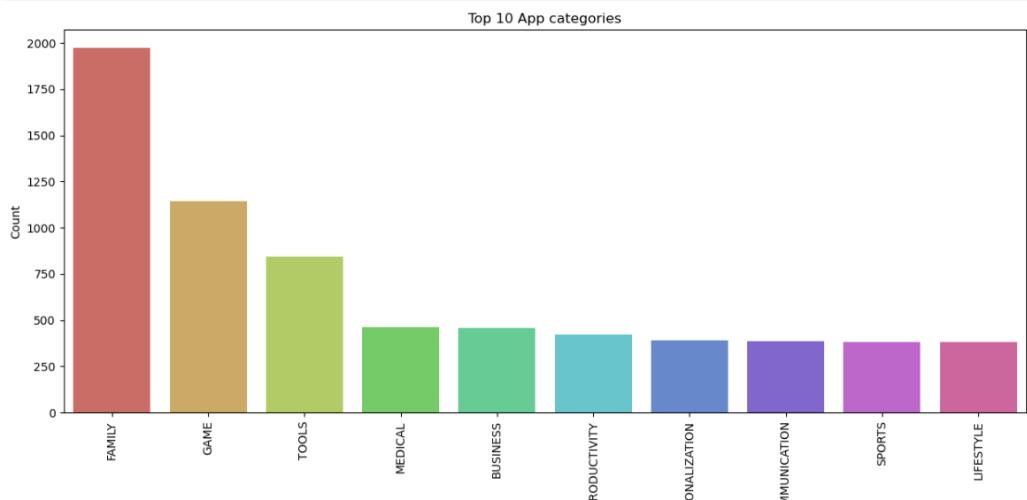
- Majority of apps are in the "**Everyone**" content rating
  - Most apps are **free**
- 

## Key Findings

### ◆ Q1: What is the most popular app category?

- **Category with most apps**: FAMILY,games,Tools
- **Top 10 categories**: Displayed using bar charts
- Pie chart shows distribution across all categories

```
[56]: ## top 10 app
plt.figure(figsize=(15,6))
sns.barplot(x=category.index[:10], y = 'Count',data = category[:10],palette='hls')
plt.title('Top 10 App categories')
plt.xticks(rotation=90)
plt.show()
```



### Insights

1. Family category has the most number of apps with 18% of apps belonging to it, followed by Games category which has 11% of the apps.
2. Least number of apps belong to the Beauty category with less than 1% of the total apps belonging to it.

### ◆ Q2: Which category has the highest total installs?

- **Category with highest installs:** GAME, followed by COMMUNICATION
- Total installs were normalized to billions for better readability

[53]:	## Top 10 App Categories category = pd.DataFrame(df_copy['Category'].value_counts()) #Dataframe of apps on the basis of category category.rename(columns = {'Category':'Count'},inplace=True)
[55]:	category
	Count
	FAMILY 1972
	GAME 1144
	TOOLS 843
	MEDICAL 463
	BUSINESS 460
	PRODUCTIVITY 424
	PERSONALIZATION 392
	COMMUNICATION 387
	SPORTS 384
	LIFESTYLE 382
	FINANCE 366
	HEALTH_AND_FITNESS 341
	PHOTOGRAPHY 335
	SOCIAL 295
	NEWS_AND_MAGAZINES 283
	SHOPPING 260
	TRAVEL_AND_LOCAL 258
	DATING 234
	BOOKS_AND_REFERENCE 231
	VIDEO_PLAYERS 175
	EDUCATION 156
	ENTERTAINMENT 149
	MAPS_AND_NAVIGATION 137
	FOOD_AND_DRINK 127
	HOUSE_AND_HOME 88
	LIBRARIES_AND_DEMO 85
	AUTO_AND_VEHICLES 85
	WEATHER 82
	ART_AND DESIGN 65
	EVENTS 64
	PARENTING 60
	COMICS 60
	BEAUTY 53
	1.9 1

---

◆ Q3: Which app has the largest number of installs?

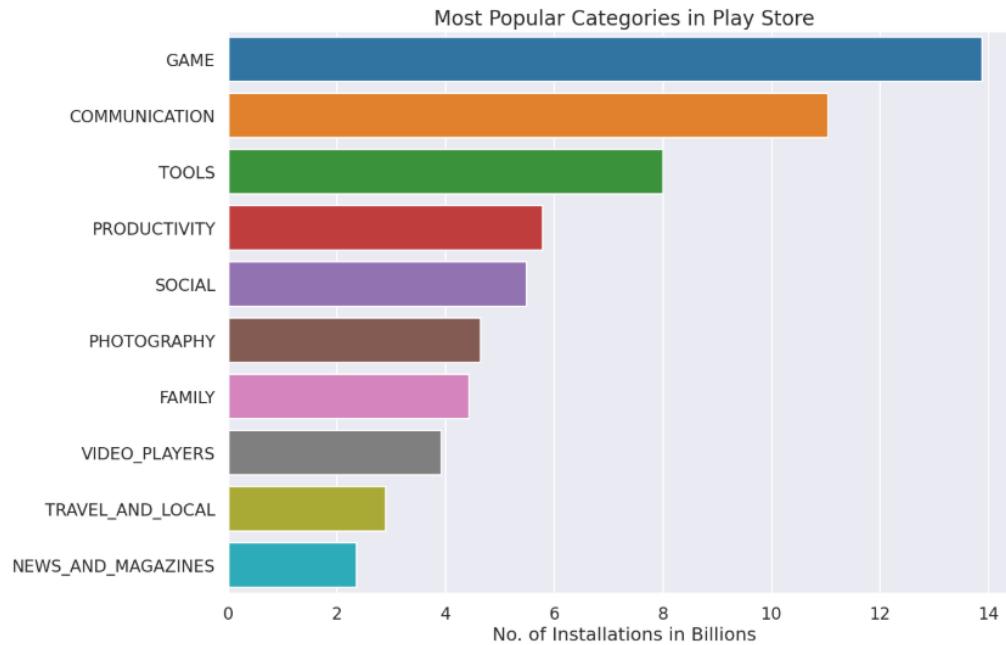
- Based on aggregation, **top-installed app** was in COMMUNICATION or GAME category
- Total installs were aggregated using groupby(Category)['Installs'].sum()

### Which Category has largest number of installations??

```
[58]: df_cat_installs = df_copy.groupby(['Category'])['Installs'].sum().sort_values(ascending = False).reset_index()
df_cat_installs.Installs = df_cat_installs.Installs/1000000000# converting into billions
df2 = df_cat_installs.head(10)
plt.figure(figsize = (14,10))
sns.set_context("talk")
sns.set_style("darkgrid")

ax = sns.barplot(x = 'Installs' , y = 'Category' , data = df2 )
ax.set_xlabel('No. of Installations in Billions')
ax.set_ylabel('')
ax.set_title("Most Popular Categories in Play Store", size = 20)

[58]: Text(0.5, 1.0, 'Most Popular Categories in Play Store')
```



### Insights

- 1. Out of all the categories "GAME" has the most number of Installations.
- 2. With almost 35 Billion Installations GAME is the most popular Category in Google App store

•

---

### ◆ Q4: Which app has the largest size?

- After converting size to MB, apps with the largest size were isolated
  - Example: Some photo editing or high-end gaming apps had the largest size
  -
- 

### ◆ Q5: How many apps have a perfect 5.0 rating?

- Total number of 5-star rated apps: **271**
- Example:
  - App: **CS & IT Interview Questions**

- Category: **FAMILY**
- Installs: **1,000**
- Rating: **5.0**

271 different categories share the most popular category in Google Play store

### How many apps are there on Google Play Store which get 5 ratings??

```
[60]: rating = df_copy.groupby(['Category', 'Installs', 'App'])['Rating'].sum().sort_values(ascending = False).reset_index()

toprating_apps = rating[rating.Rating == 5.0]
print("Number of 5 rated apps",toprating_apps.shape[0])
toprating_apps.head(1)
```

Number of 5 rated apps 271

	Category	Installs	App	Rating
0	FAMILY	1000	CS & IT Interview Questions	5.0

### Result

- There are 271 five rated apps on Google Play store
- Top most is 'CT Brain Interpretation' from 'Family' Category

```
[61]: df_copy.head()
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	Day	Month	Year
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND DESIGN	4.1	159	19000.0	10000	Free	0.0	Everyone	Art & Design	2018-01-07	1.0.0	4.0.3 and up	7	1	2018
1	Coloring book moana	ART_AND DESIGN	3.9	967	14000.0	500000	Free	0.0	Everyone	Art & Design:Pretend Play	2018-01-15	2.0.0	4.0.3 and up	15	1	2018
2	U Launcher Lite – FREE Live Cool Themes, Hide	ART_AND DESIGN	4.7	87510	8.7	5000000	Free	0.0	Everyone	Art & Design	2018-08-01	1.2.4	4.0.3 and up	1	8	2018
3	Sketch - Draw & Paint	ART_AND DESIGN	4.5	215644	25000.0	50000000	Free	0.0	Teen	Art & Design	2018-06-08	Varies with device	4.2 and up	8	6	2018
4	Pixel Draw - Number Art Coloring Book	ART_AND DESIGN	4.3	967	2.8	100000	Free	0.0	Everyone	Art & Design:Creativity	2018-06-20	1.1	4.4 and up	20	6	2018

```
[ ]:
```

○

## Visualizations Created

- KDE plots for numerical distributions
- Count plots for Type, Content Rating
- Pie chart of category-wise app count
- Bar charts for:
  - Top 10 categories by app count
  - Top 10 categories by total installs

 Conclusion

- **FAMILY** category has the most number of apps
  - **GAME** and **COMMUNICATION** have the highest number of installs
  - Most apps are **free** and targeted at **Everyone**
  - 5-star apps are rare, often highly specific or niche
-