Google Play

GOOGLE PLAYSTORE APPS RATING PREDICTION

PRESENTED BY :

SUBRATA DEY
DATA ANALYST INTERN

TOOLS USED : PYTHON, SQL, EXCEL, POWER BI

UNIFIED MENTOR PRIVATE LIMITED

DATE :

Executive Summary

- Analyzed approx. 9,000 Google Play Store apps to identify factors influencing app ratings.
- Average app rating across the platform is 4.19, indicating generally positive user sentiment.
- Games and Communication apps dominate installs, accounting for the highest user reach.
- Apps with higher installs and reviews tend to have stronger ratings, highlighting engagement as a key driver.
- Dashboard enables stakeholders to identify high-performing categories, genres, and platforms.

Business Impact

- Supports data-driven decisions for app development, pricing, and market targeting.

Problem Statement & Objectives

Problem Statement

With millions of apps on the Play Store, developers struggle to understand what drives app ratings and user adoption.

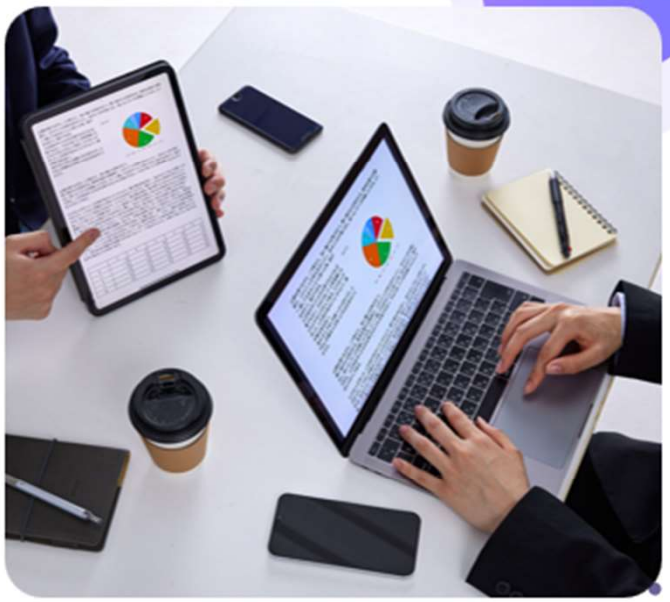
Objectives

- Analyze app performance across categories, installs, and ratings.
- Identify patterns affecting app ratings.
- Support rating prediction using analytical insights.

Why it matters ?

Higher ratings lead to:

- Better visibility.
- Increased installs.
- Higher revenue potential.



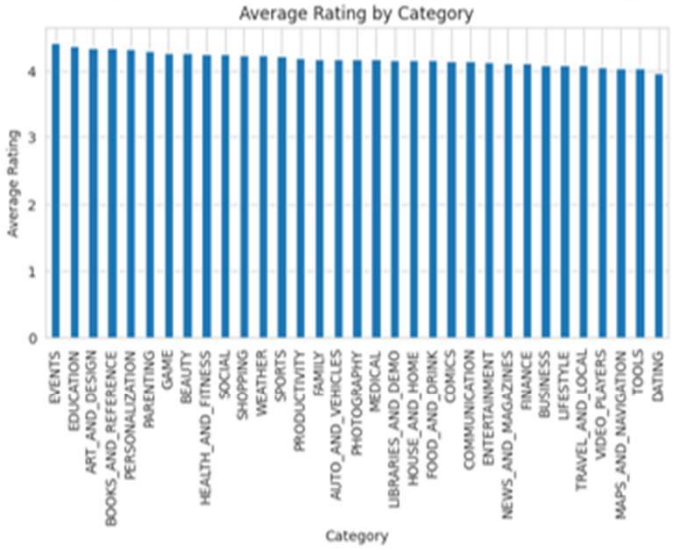
Data Preparation and Cleaning

Data Loading & Cleaning

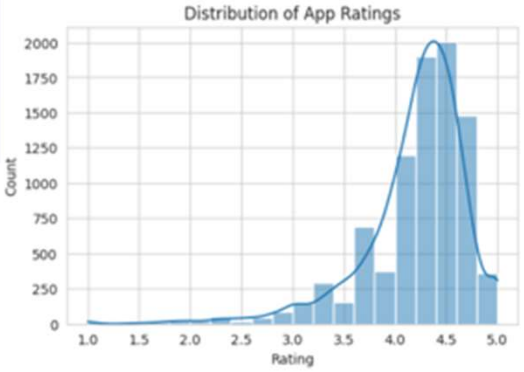
```
# Load the dataset:
df = pd.read_csv('/content/googleplaystore.csv')
# Looking for any Missing Values:
df.isnull().sum()
# Look for % of missing value / Null Value:
df.isnull().sum()/df.shape[0]*100
# Drop rows with missing Rating (target variable)
df2 = df.dropna(subset=['Rating'])  # Remove rows where Rating is null
df2.isnull().sum()
# Fill missing values for categorical columns
# Fill Type with most frequent value (ie Mode)
df2['Content Rating'].fillna(df2['Content Rating'].mode()[0],
inplace=True)
df2['Android Ver'].fillna(df2['Android Ver'].mode()[0], inplace=True)
# Clean and convert corrupted Installs column
# Remove '+' and ',' characters
df2['Installs'] = df2['Installs'].astype(str)  # makes the code
re-runnable without errors
df2['Installs'] = df2['Installs'].str.replace(',', '', regex=False)
df2['Installs'] = df2['Installs'].str.replace('+', '', regex=False)
# Convert Installs to numeric (invalid values become NaN)
df2['Installs'] = pd.to_numeric(df2['Installs'], errors='coerce')
# Drop rows where Installs could not be converted
df2 = df2.dropna(subset='Installs')
# Convert to integer
df2['Installs'] = df2['Installs'].astype(int)
```

Price and Rating Relationship

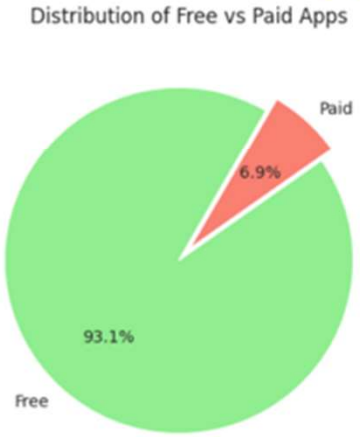
Category vs Average Rating



App Rating distribution



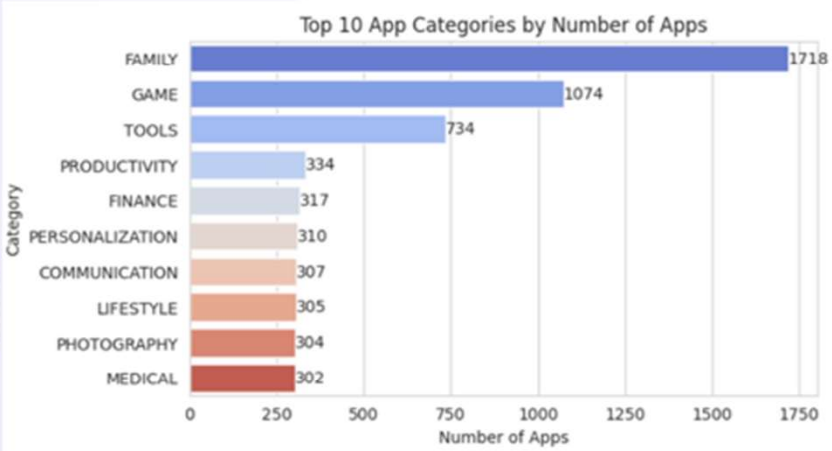
Free vs Paid apps



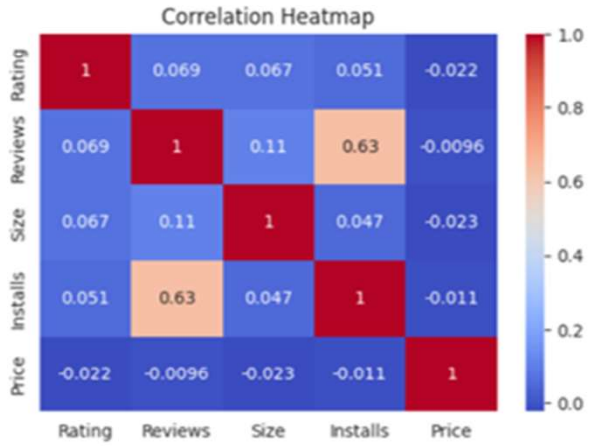
- Most paid apps are low-priced, expensive apps are rare, and ratings show no price correlation.
- Most app categories exceed 4.0 ratings, with niche categories slightly outperforming others consistently.
- Most apps have high ratings, skewed toward 4.0–5.0, reflecting strong overall user satisfaction.
- Free apps dominate the Play Store, while paid apps represent only a small minority.

Exploratory Data Analysis (EDA)

Top Categories by no of Apps

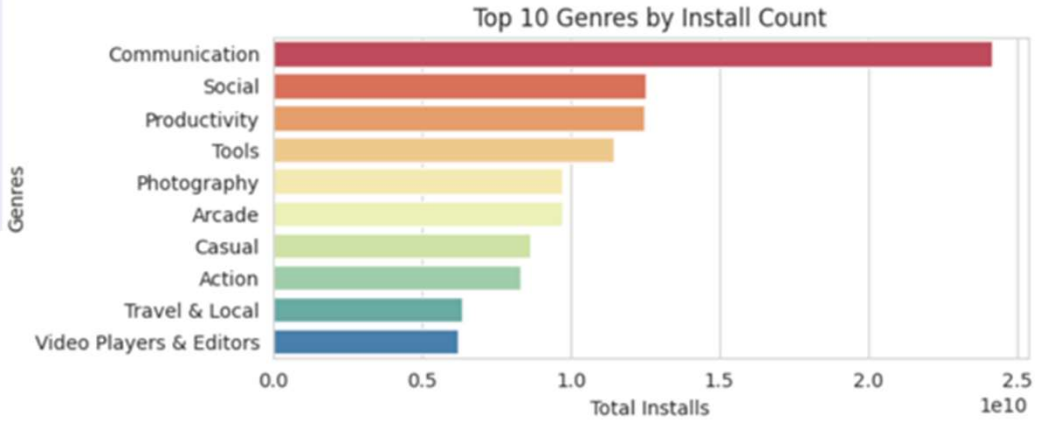


Correlation Analysis



Rating vs Reviews

Top Genres by Install Count



- Family and Games dominate app counts, while utility categories highlight diverse and practical user needs.
- Reviews and installs strongly correlate, while ratings weakly relate to size, price, and reviews.
- App ratings cluster high overall, while review counts show only weak influence on ratings.
- Communication and Social genres dominate installs, while utility and entertainment apps reflect diverse user needs.

SQL Analysis For Insights

1) What is the relationship between installs and app ratings?

```
select
    case                                -- Categorize a
        when Installs >= 1000000 then 'High Installs'
        else 'Low Installs'
    end as install_group,
    round(avg(Rating), 2) as avg_rating  -- Average ra
from playstore_app_data
where Rating is not null                -- Exclude miss
group by install_group;
```

install_group	avg_rating
Low Installs	4.11
High Installs	4.28

2) Top 10 app categories by number of apps

```
select Category,
   count(*) as app_count
from playstore_app_data
group by Category
order by app_count desc
limit 10;
```

Category	app_count
FAMILY	1718
GAME	1074
TOOLS	734
PRODUCTIVITY	334
FINANCE	317
PERSONALIZATION	310
COMMUNICATION	307
LIFESTYLE	305
PHOTOGRAPHY	304
MEDICAL	302

3) Which categories have the highest average app ratings?

```
select Category,
   round(avg(Rating), 2) as avg_rating
from playstore_app_data
where Rating is not null     -- Consider
group by Category
order by avg_rating desc;
```

Category	avg_rating
EVENTS	4.44
EDUCATION	4.38
ART_AND_DESIGN	4.36
BOOKS_AND_REFERENCE	4.35
PERSONALIZATION	4.33
PARENTING	4.3
BEAUTY	4.28
GAME	4.28
HEALTH_AND_FITNESS	4.26
SOCIAL	4.25
SHOPPING	4.25
WEATHER	4.24
SPORTS	4.23
PRODUCTIVITY	4.2
AUTO_AND_VEHICLES	4.19
FAMILY	4.19
LIBRARIES_AND_DEMO	4.18
MEDICAL	4.18
PHOTOGRAPHY	4.18
COMICS	4.16
FOOD_AND_DRINK	4.16

4) How has the average app rating changed over time?

```
select year(`Last Updated`) as update_year,
   round(avg(Rating), 2) as avg_rating
from playstore_app_data
where Rating is not null
group by update_year
order by update_year;
```

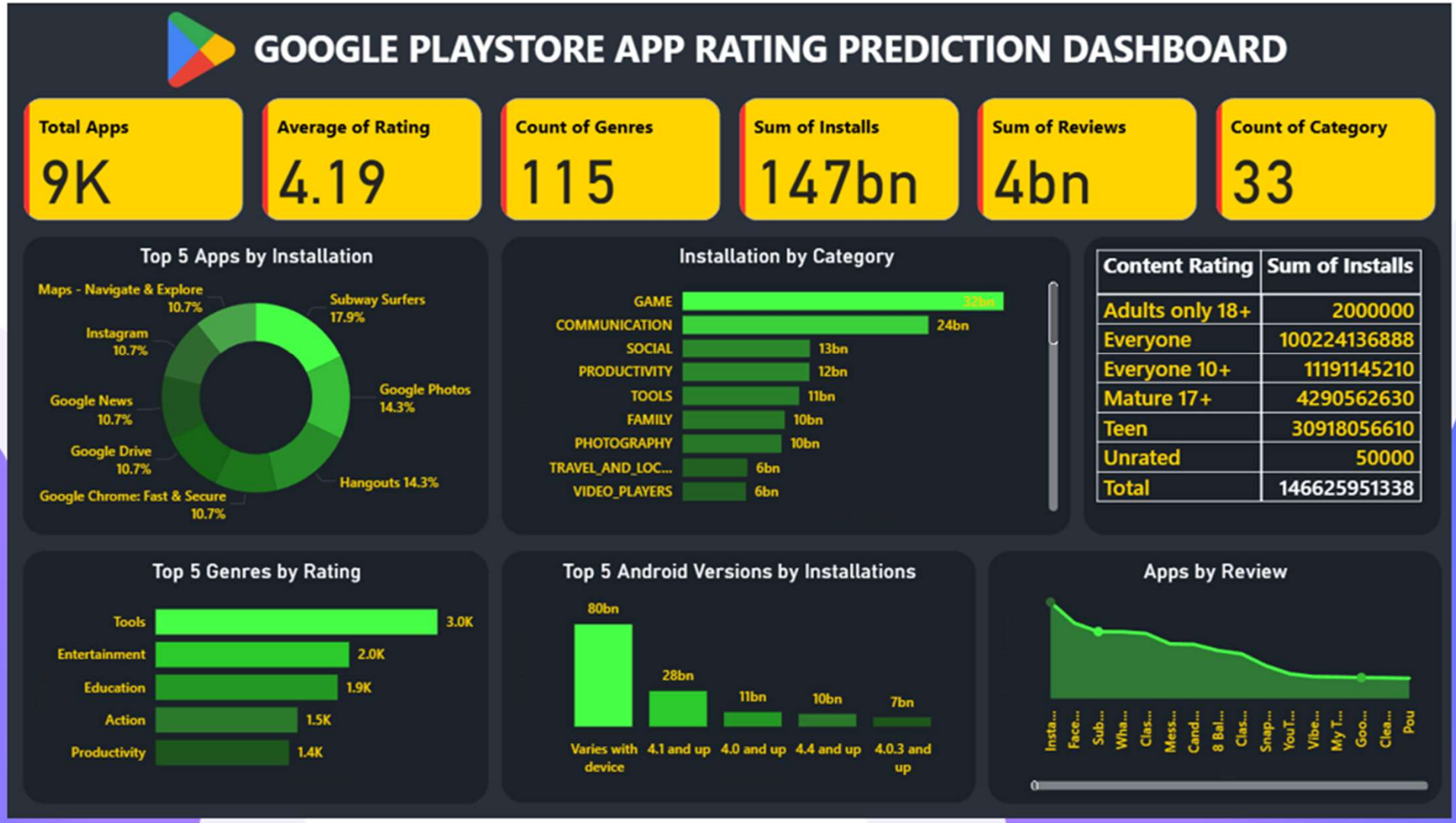
update_year	avg_rating
2010	4.2
2011	3.97
2012	3.79
2013	4.07
2014	4.04
2015	4.06
2016	4.04
2017	4.09
2018	4.24

5) Top-rated apps with at least 1 million installs

```
select App,
   Rating,
   Installs
from playstore_app_data
where Rating is not null    --
   and Installs >= 1000000
order by Rating desc
limit 10;
```

App	Rating	Installs
Period Tracker	4.9	1000000
ipsy: Makeup, Beauty, and Tips	4.9	1000000
Tickets + PDA 2018 Exam	4.9	1000000
PixPanda - Color by Number Pixel Art Coloring B...	4.9	1000000
Learn Japanese, Korean, Chinese Offline & Free	4.9	1000000
JW Library	4.9	10000000
StrongLifts 5x5 Workout Gym Log & Personal Tr...	4.9	1000000
Six Pack in 30 Days - Abs Workout	4.9	10000000
Lose Belly Fat in 30 Days - Flat Stomach	4.9	5000000
Six Pack in 30 Days - Abs Workout	4.9	10000000

Power BI Dashboard Insights



Recommendations

Recommendation 1 :

Target high-install categories (Games, Communication).

- ***Insight:*** These categories dominate user reach.
- ***Outcome:*** Faster adoption.

Recommendation 2 :

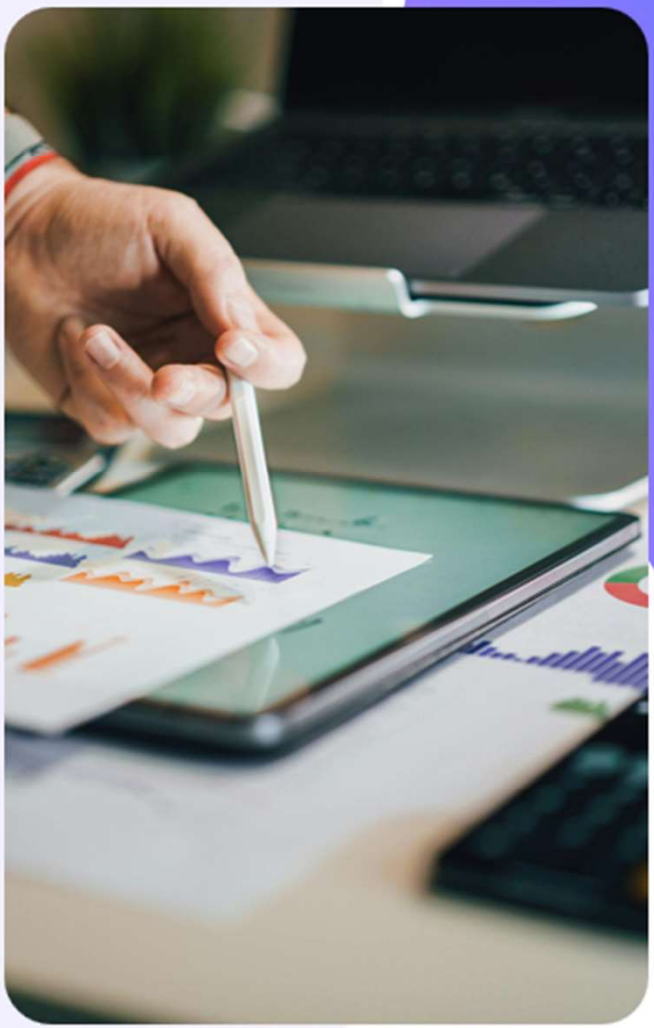
Improve engagement metrics.

- ***Insight:*** Reviews and installs strongly correlate with ratings.
- ***Outcome:*** Better app visibility.

Recommendation 3 :

Optimize Android version support.

- ***Insight:*** Wider compatibility leads to higher installs.
- ***Outcome:*** Expanded user base.



Conclusion

- Successfully analyzed Google Play Store app data.
- Identified key drivers of app ratings and installs.
- Delivered an interactive, insight-driven dashboard.
- Provided actionable recommendations for app strategy.

Thank you so much

By - Subrata Dey
Data Analyst Intern
Unified Mentor Private Limited
Date :