**DAV PROJECT**

**DATA ANALYSIS OF PLAY STORE DATASET**

Submitted to:

**Ms. HEMA BANATI**

**Ms. SHEETAL TANEJA**

Submitted in Partial Fulfillment of the Requirement for the Degree of

**Bachelor of Science in Computer Science**

Submitted by:

**Aditya (Roll No.20/49002)**



**Dyal Singh College (University of Delhi)**

**Department of Computer Science**

**About Dataset**

* **Source**

[**https://www.kaggle.com/datasets/mrushan3/playstore-analysis-data**](https://www.kaggle.com/datasets/mrushan3/playstore-analysis-data)

* **Description**

Google Play Store hosts about 3.48 million apps, all with different purposes and user bases.

With all the data we have, we will clean it, visualize it and extract meaningful insights from it which will help the user by recommending apps they use or according to their need and requirements on a daily basis according to their own preference by visualizing and analyzing their recently searched/explored apps.

### Content

The happiness scores and rankings use data from the Gallup World Poll. The attributes Names are: -

The dataset has 9 columns and 10841 rows.

1. App
2. category
3. Rating
4. Reviews
5. Size
6. Installs
7. Price
8. Content Rating
9. Genres

* **Queries**

# What category is preferred the most?

# What are the top 5 paid apps in terms of rating and installs?

1. Which 5 categories have engaged teens more?

# What is the Category of Apps having size more than 90MB

# Genres of expensive Apps (Apps of price greater than 200).

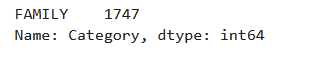
# Query 1: What category is preferred the most?

Code :

a = df['Category'].value\_counts()

most\_preffered\_category = a.iloc[:1]

most\_preffered\_category



* Apps stated under Family category is preferred the most.

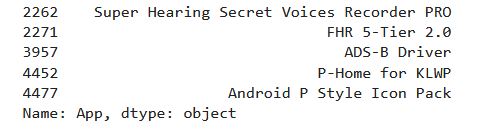
**Query 2: What are the top 5 paid apps in terms of rating and installs?**

Code:

df2 = df[df.Price != 0]

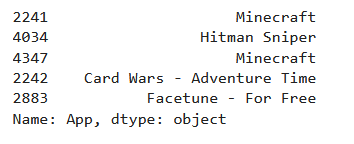
df2\_1 = df2.nlargest(5, ['Rating'])

print(df2\_1['App'])



df2\_2 = df2.nlargest(5, ['Installs'])

df2\_2['App']



🡪 In terms of rating and installation, the top 5 paid apps are as given above. For all the entries we selected the paid apps and then fetched the top 5 in these columns.

**Query 3: Which 5 categories have engaged teens more?**

Code :

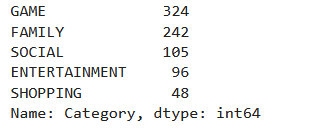
df3 = df[df['Content Rating'] == "Teen"]

df3

a = df3['Category'].value\_counts()

most\_teen = a.iloc[:5]

most\_teen



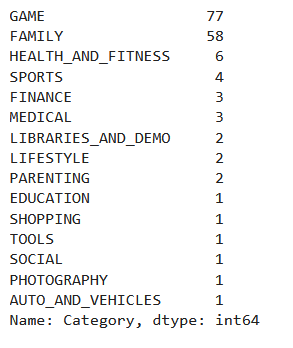
* Teens have been engaged in these 5 categories, (Games, Family, Social, Entertainment, Shopping)

# Query 4: What is the Category of Apps having size more than 90MB.

Code :

dfs = df[df["Size"] > 92160]

dfs['Category'].value\_counts()



* Categories of bigger apps, of size greater than 90Mb.

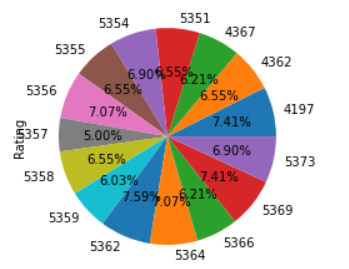
# Query 5: Ratings of expensive Apps (Apps of price greater than 200).

Code :

dfp = df.loc[df['Price'] > 200]

x = dfp["Rating"]

x.plot.pie(autopct='%1.2f%%')



* Pie Plot describing the ratings of expensive apps (apps greter than $200) with their percentage values.

Thanks and Regards

**Aditya (Roll No.20/49002)**