# **INFORMATICS PRACTICES**

PROJECT: CLASS XII



PROJECTNAME: GOOGLE PLAY STORE APP ANALYSIS
CONCEPT USED: PANDAS AND MATPLOTLIB
SUBMITTED TO: MRS. PURTI MALHOTRA

Kritika Tripathi

12<sup>TH</sup> GRADE SCIENCE

Delhi Public School

Maruti Kunj

## **INDEX**

S. No.	DESCRIPTION	PAGE NO.
1	Certificate	2
2	Acknowledgment & References	3
3	Introduction	4
4	Technology Used	5
4	Source Code	6
5	Output Screen	11
6	Hardware & Software requirements	27

### **CERTIFICATE**

This is to certify that the <u>Google Play Store Apps Analysis</u> Informatics Practices project is developed by <u>Kritika Tripathi</u> under my guidance in the session 2020-2021. The work done by her is original.

**Teacher Name: Mrs. Purti Malhotra** 

**Informatics Practices** 

Date:

### **ACKNOWLEDGMENT**

I would like to express my sincere gratitude to my computer teacher Mrs. Purti Malhotra for her vital support, guidance, and encouragement without which this project would not come forth from my side, who helped me complete the project by giving ideas, thoughts and made this project easy and accurate. I wish to thank my parents for their undivided support and interest who inspired me and encouraged me to go my own way, without which I would be unable to complete my project.

### **REFERENCES**

- 1. https://www.kaggle.com/lava18/google-play-store-apps
- 2. https://www.geeksforgeeks.org
- 3. http://stackoverflow.com
- 4. https://www.androidcentral.com/google-play-store/home
- 5. Class notes

### INTRODUCTION

#### **Google Play Store Apps Analysis**

Google Play, which was originally born and referred to by Google as the Android Market, is Google's official store and portal for Android apps, games, and other content. There are millions of Apps available on this Android Market having a variety of categories, types, genres, and content ratings. All apps for Android phones are housed in the Google Play Store, and each app has its own web page where information about the app is available. Though the list of these applications is larger than imagined, this project is a short analysis of more than ten thousand apps in the play store.

The aim of the project is to inform the users about the variety of apps available on Play Store by various kinds of Data Manipulations along with their Visualisations.

To process this large amount is not something a human analyst can perform, so the need for data analysis models arise. This program introduces a special assistant, Anna, who will help the user throughout the analysis.

### **TECHNOLOGY USED**

### ❖ Python 3.7.0

#### > PANDAS:

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

To import: import pandas as pd

#### > MATPLOTLIB:

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy.

To import: import matplotlib.pyplot as plt

#### > NUMPY:

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

To import: import numpy as np

### **SOURCE CODE**

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import sys
df= pd.read_csv("googleplaystore.csv")
df['Reviews'] = pd.to_numeric(df['Reviews'])
df['Installs'] = pd.to_numeric(df['Installs'])
df['Rating'] = pd.to_numeric(df['Rating'])
#Data analysis:
def csv():
  print(df)
def count():
#Number of apps in total
  counts = df["App"].count()
  print("There are ", str(counts)," in the analysis.")
def appinfo():
  #Printing Information of any user specified app
  row = input("Please enter the name of any App from the given to analyse it: ")
  print(df[df["App"]==row])
def col():
  field= str(input("Enter the name of the field you want me to display:"))
  field1=df[["App", field]].copy()
  print(field1)
  g=input("Do you want me to sort the values in this field for you? (Yes/No):\n")
  if q== "Yes":
    field2= field1.sort values(by=field)
    print(field2)
  else:
    print("Okay! Thank you so much!")
def drop():
  print(df.columns)
  delete= eval(input("Enter the field(s) you want me to delete (in square brackets):\n"))
  df.drop(delete, axis=1)
  print(df)
  print(delete," Successfully deleted!")
def max():
  print("These are the columns\n", df.columns)
  field=eval(input("Enter the column names as list in square bracket"))
  print('Print the maximum values of the ',field,' columns')
  print(df[field].max())
```

```
print('Print the minimum values of the ', field, ' columns')
  print(df[field].min())
def catmax():
  #Maximum Rating per Category
  print("The maximum ratings for each Category is:")
  print(pd.pivot_table(df, index=["Category"], values=["Rating"], aggfunc="max"))
  g=input("Do you want me to name the Apps with top Ratings? (Yes/No):\n")
  if \alpha== 'Yes':
    print("Here are the apps with maximum rating per Category:")
    df1=df.drop(['Reviews', 'Size', 'Price', 'Genres', "Current Ver", 'Installs', 'Type',
           'Content Rating','Android Ver'], axis=1)
    df2=df1.groupby(["Category"])
    df3=df2.apply(lambda x: x.sort_values(["Rating"], ascending=False)).reset_index(drop=True)
    df4=df3.groupby('Category').head(1)
    print(df4)
    cmap = plt.cm.tab10
    colors = cmap(np.arange(len(df)) % cmap.N)
    df4.plot.barh(x='Category', y='Rating', color=colors)
    plt.show()
  elif q=="No":
    print("Okay! Thank you so much!")
  else:
    print("Sorry! I'm unable to understand what you are trying to say.")
def field():
  #sorting according to a particular row
  field=input("Write any of the given fields to print the number of Apps according to it:\na. Category\n"
"b. Installs\nc. Content Rating\nd. Type\ne. Android Ver\nf. Current Ver\ng. Genres\n"
         "h. Current Ver\n")
  print("The Field has the following distribution of Apps:")
  unique= df[field].unique()
  print(unique)
  print("There are ", len(unique)," ", field, "in total.")
  q=input("Do you want me to show you the app information for every distribution")
  df.set index(field)
  if q=="Yes":
    a=str(input("For which Distribution?\n"))
    b= df[df[field] == a]
    print(b)
    num = pd.pivot_table(df, index=[field], values=["App"], aggfunc="count")
    num.plot.pie(y="App", figsize=(10, 5), labeldistance=None)
    plt.legend(prop={'size': 6}, ncol=2, bbox_to_anchor=(0.85, 1.025), loc="upper left")
```

```
plt.show()
  else:
    print("Okay! Thank you so much!")
def maxmin():
  #Minimum/Maximum Rating/Reviews/ Installs
  field=str(input("Select a field from the given list:\n"
     "a. Reviews\n"
     "b. Rating\n"
     "c. Installs"))
  df1=df[["App","Reviews","Rating","Installs"]].copy()
  n=int(input("Enter the number of apps you want to be displayed:"))
  df1.sort values(by=field, ascending=False)
  print("The app with maximum ", field," is:")
  print(df1.head(n))
  print("The app with minimum ", field, " is:")
  print(df1.tail(n))
def agg():
  field=input("1. Category 2. Genres 3. Type 4. Content Rating\n"
        "Enter the name of the field from the above list:\n")
  perform=input("1. Installs 2. Rating 3. Reviews 4. Price\n"
         "Enter the name of the field from the above list whose mean you want to find:\n")
  df.groupby(field)
  print(pd.pivot table(df, index=[field], values=[perform], aggfunc="mean"))
def reviews():
  reviews = df[['Category', 'Reviews']].copy()
  reviews['Reviews'] = pd.to numeric(reviews['Reviews'])
  num = pd.pivot_table(df, index=["Category"], values=["Reviews"], aggfunc="sum")
  print(num)
  num.boxplot(color='blue', vert=0, notch=True)
  plt.show()
def price():
  price=df[["App","Price"]].copy()
  price["Price"]=pd.to_numeric(price["Price"])
  price.plot(color="blue")
  plt.title("Apps Pricing")
  plt.xlabel("No. Of Apps")
  plt.ylabel("Price")
  plt.show()
def andver():
  print("Play store has these many Android Version ranges:")
  print(df["Android Ver"].unique())
  version= str(input("Type your Android Version Range as per the above ranges:"))
  app=str(input("Enter the name of the app you want to use:"))
  b = df[df["App"] == app]
  if version in b.values:
    print("Yes! Your Android Version is up to date for ", app, '!')
  else:
    print("Sorry! The desired Android Version does not match.")
def hist():
  df.hist()
  plt.show()
```

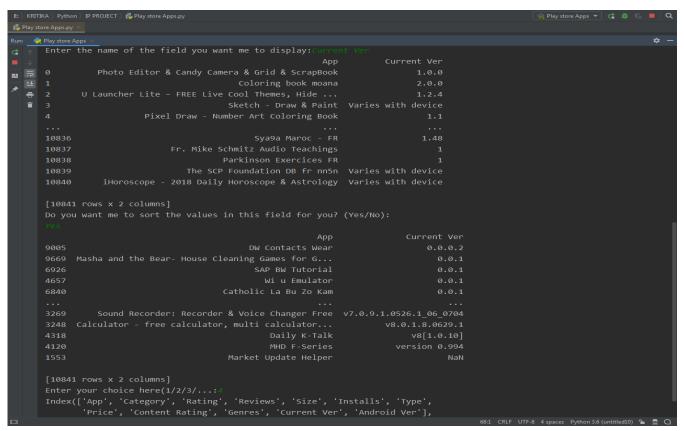
```
#menu
print("
              ", "Google Play Store Apps","
print("
********************************
name = input("Please enter you name:")
print("Hello ", name, "! I'm your Analysis Assistant, Anna, here to help you with the analysis!")
print("Please choose and option from the given menu:\n"
   "1. Display the number of Apps in the Data Analysis\n"
   "2. Find any App's information from the dataset\n"
   "3. Display a particular field\n4. Drop field(s) from data set\n"
   "4. Show maximum/ minimum for a particular field\n"
   "5. Show maximum Ratings- Category wise\n"
   "6. Display the number of apps according to a particular field with details\n"
   "7. Display the App with Maximum and minimum Rating/ Reviews/ Installs\n"
   "8. Tell the mean number of Installs/ Ratings/ Reviews/ Price according to "
   "Category/ Genre/ Type/ Content Rating\n"
   "9. Calculate the sum of reviews for each Category\n"
   "10. Show the price distribution of all apps in dataset\n"
   "11. Check if I can run my favourite app on my Android phone"
   "12. Make a histogram for the dataset\n"
   "13. Display th whole CSV File\n"
   "14. Exit")
while True:
  ch=int(input("Enter your choice here(1/2/3/...:"))
  if ch==1:
    count()
  if ch==2:
    appinfo()
  if ch==3:
    col()
  if ch==4:
    drop()
  if ch==5:
    max()
  if ch==6:
    catmax()
  if ch==7:
    field()
  if ch==8:
    maxmin()
  if ch==9:
    agg()
  if ch==10:
    reviews()
  if ch==11:
    price()
  if ch==12:
    andver()
  if ch==13:
    hist()
  if ch==14:
    csv()
  if ch==15:
    print("Thank you so much! Have a great day!")
```

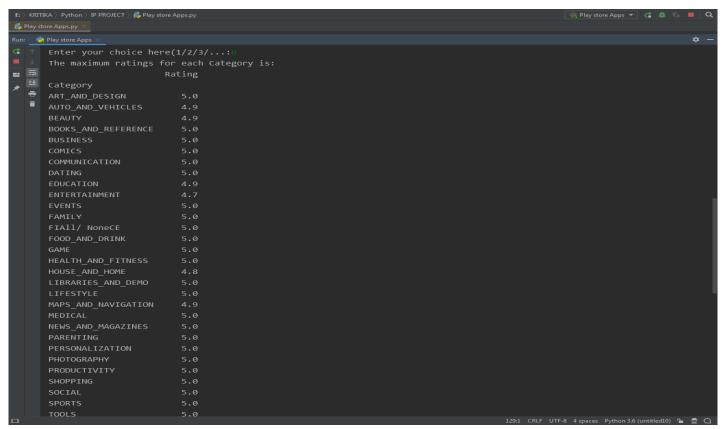
#### **OUTPUT SCREEN**

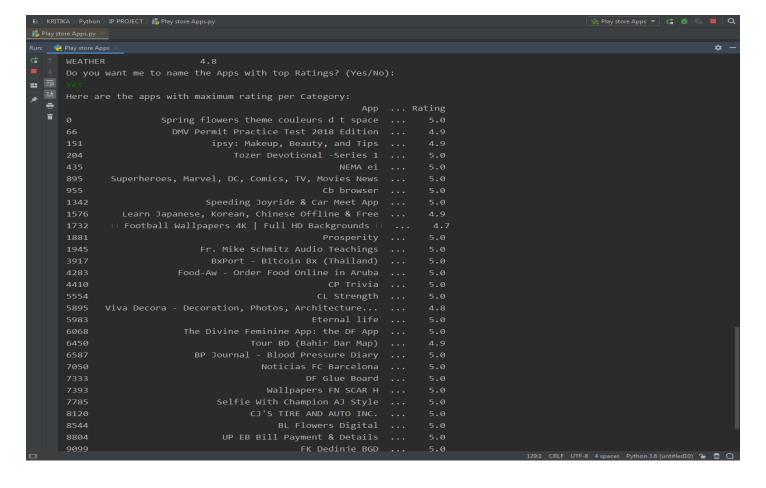
### **Functions performed:**

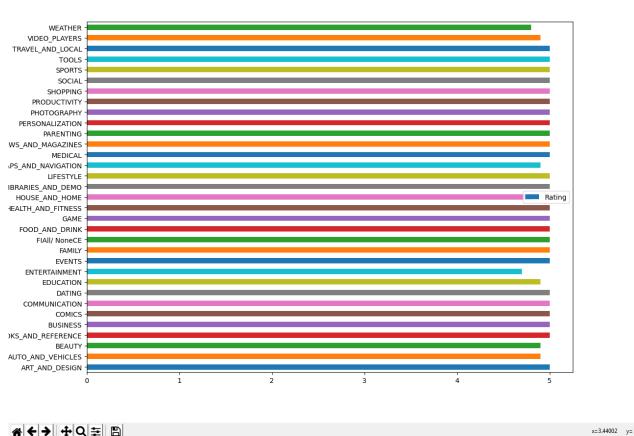
- 1. Number of Apps in the Data Analysis
- 2. Find any App's information from the dataset
  - 3. Drop field(s) from the dataset
    - 4. Display a particular field
  - 5. Maximum/ Minimum of a particular field
- 6. Maximum Ratings- Category wise (Bar Chart)
- 7. Number of apps according to a particular field (variety of Pie charts)
- 8. App with Maximum and minimum Rating/ Reviews/ Installs
  - 9. Mean of particular fields
  - 10. The sum of reviews for each Category (Box plot)
  - 11. Price distribution of all apps in dataset\n (Line Plot)
- 12. Check if user can run a particular app in his/her Android Phone
  - 13. Histogram of the whole dataset 14.Exit

```
| PROJECT | Proj
```



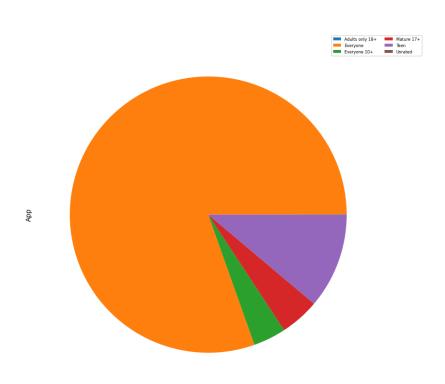




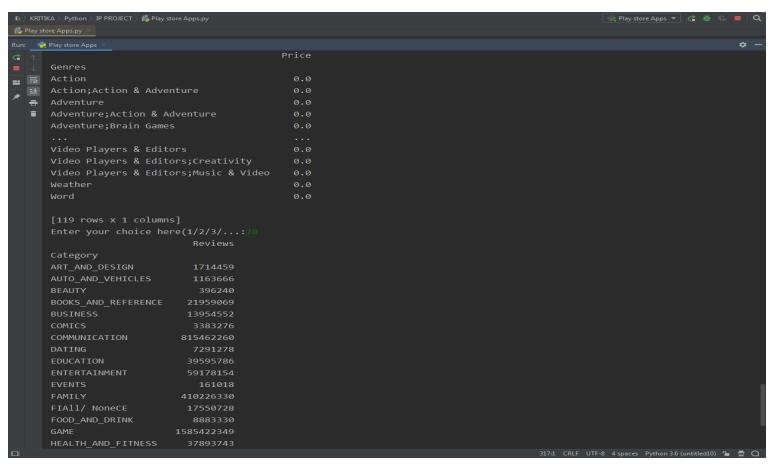


```
E MOTION System DEPROPORT & Polystone Appears

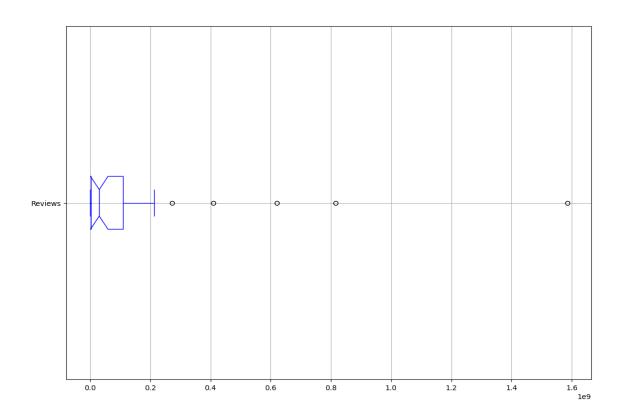
| Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears | Stylene Appears |
```



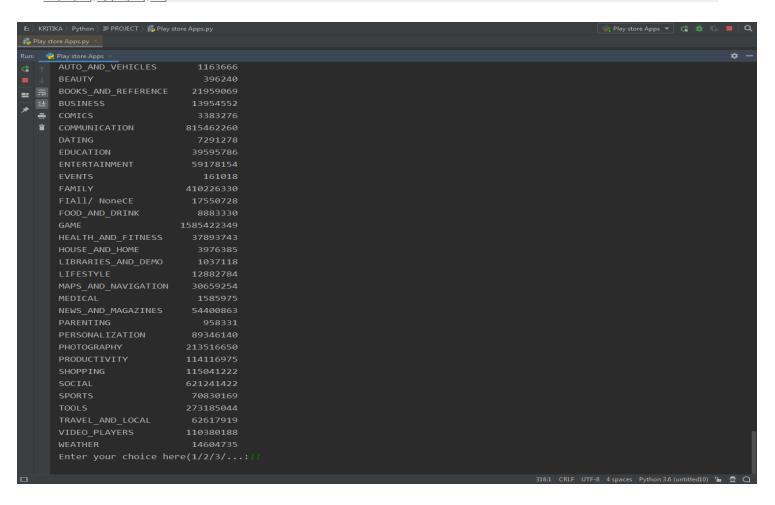
```
🗎 Play store Apps 🔻 🕻 🇯 🕟 📕
    Play store Apps
      Enter your choice here(1/2/3/...:8
  ⇒ a. Reviews
==
      b. Rating
  ÷
      Enter the number of apps you want to be displayed:
                                                                159
967
                                                                                   10000
                                        Coloring book moana
                                                                                   500000
                                                                           4.5 500000000
                     Pixel Draw - Number Art Coloring Book
      The app with minimum Reviews is:
                                                                                    5000
                           Fr. Mike Schmitz Audio Teachings
                                     Parkinson Exercices FR
                                                                                     1000
                              The SCP Foundation DB fr nn5n
                                                                                    1000
      1. Installs 2. Rating 3. Reviews 4. Price
      Enter the name of the field from the above list whose mean you want to find:
                                                 0.0
                                                 0.0
      Adventure
                                                 0.0
                                                 0.0
                                                                                            281:34 CRLF UTF-8 4 spaces Python 3.6 (untitled10) %
```



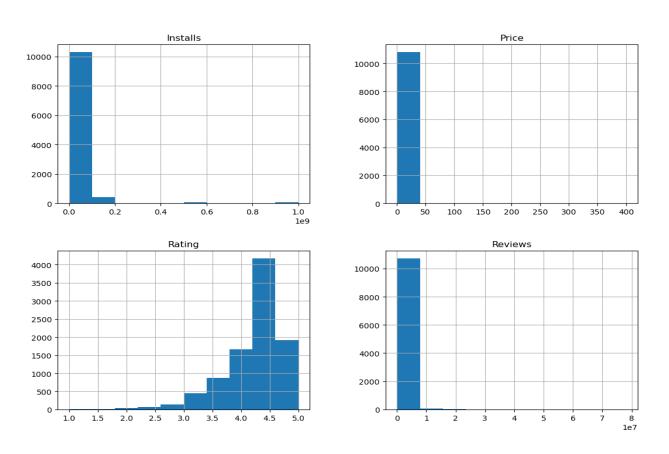
ifigure 1



#### # **←** → **+** Q = B



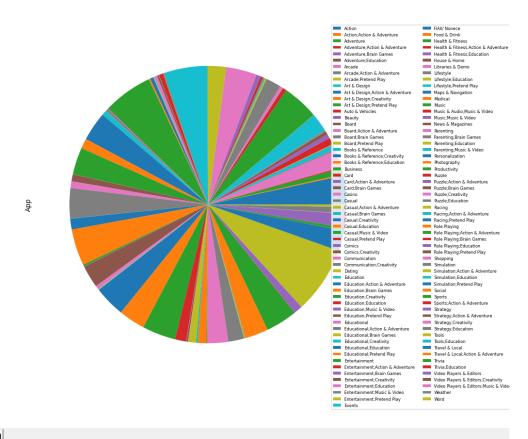
🗓 Figure 1



#### # **←** → **+** Q = B

```
Enter your choice here(1/2/3/...:14
                                                     App
                                                                Android Ver
0
         Photo Editor & Candy Camera & Grid & ScrapBook
                                                         ... 4.0.3 and up
                                     Coloring book moana
                                                         ... 4.0.3 and up
2
      U Launcher Lite - FREE Live Cool Themes, Hide ...
                                                          ... 4.0.3 and up
3
                                   Sketch - Draw & Paint
                                                                4.2 and up
                   Pixel Draw - Number Art Coloring Book
                                                               4.4 and up
10836
                                        Sya9a Maroc - FR
                                                              4.1 and up
10837
                       Fr. Mike Schmitz Audio Teachings
                                                                4.1 and up
10838
                                  Parkinson Exercices FR
                                                                2.2 and up
                           The SCP Foundation DB fr nn5n ...
10839
                                                                 All/ None
10840
           iHoroscope - 2018 Daily Horoscope & Astrology ...
                                                                 All/ None
[10841 rows x 12 columns]
Enter your choice here (1/2/3/...:15)
Thank you so much! Have a great day!
Process finished with exit code 0
```

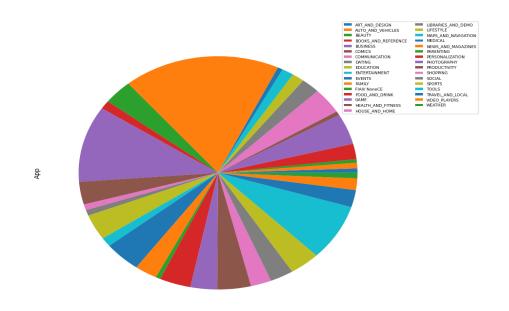




**☆ ← → + Q = B** 

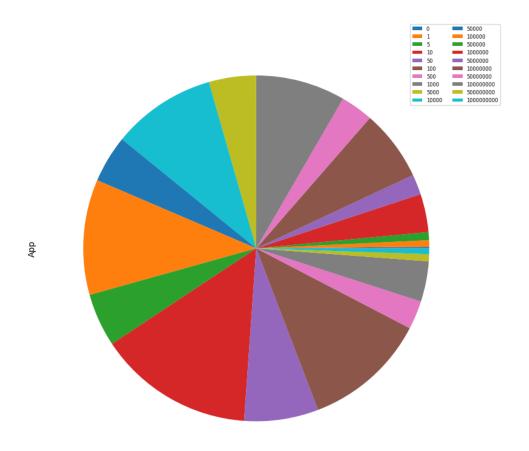
(Content Rating Distribution)

€ Figure 1



#### **☆←→** +Q = B

#### (Category distribution)



(Installs Distribution)

## **HARDWARE & SOFTWARE REQUIREMENTS**

## **Hardware Requirement**

Pentium 3/4/Core 2 Duo/Dual core/i3/i5/i7 With at least 256 MB RAM 2 MB free space on Hard Disk Color Monitor/LCD

## **Operating System & Compiler**

MS Windows Python with related libraries used of Data Analysis

- Pandas
- Matplotlib