



Google Play

Data Cleaning, Analysis and Visualization



Google Play

Objective

- Clean and preprocess the Google Play Store dataset by handling missing, duplicate, and inconsistent data.
- Analyze app data to identify patterns, trends, and correlations.
- Visualize key insights using charts and graphs for better understanding and communication
- Support decision making through meaningful data-driven insights.

Importing Libraries and Data



Google Play

```
[1]: import pandas as pd
```

```
[2]: data = pd.read_csv("googleplaystore.csv")  
data
```

```
[2]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up

Head Data



Google Play

```
[3]: # Printing First Twenty Rows  
data.head(20)
```

[3]:	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up
5	Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	March 26, 2017	1.0	2.3 and up
6	Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	April 26, 2018	1.1	4.0.3 and up

Tail Data



Google Play

```
[4]: # Printing Last 5 Rows  
data.tail(5)
```

[4]:	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free	0	Everyone	Education	July 25, 2017	1.48	4.1 and up
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100+	Free	0	Everyone	Education	July 6, 2018	1.0	4.1 and up
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9.5M	1,000+	Free	0	Everyone	Medical	January 20, 2017	1.0	2.2 and up
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0	Mature 17+	Books & Reference	January 19, 2015	Varies with device	Varies with device
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	Everyone	Lifestyle	July 25, 2018	Varies with device	Varies with device



Info Data

```
[5]: # Getting Information about the Data
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype
---  -
0   App              10841 non-null  object
1   Category         10841 non-null  object
2   Rating           9367 non-null   float64
3   Reviews          10841 non-null  object
4   Size             10841 non-null  object
5   Installs         10841 non-null  object
6   Type             10840 non-null  object
7   Price            10841 non-null  object
8   Content Rating   10840 non-null  object
9   Genres           10841 non-null  object
10  Last Updated     10841 non-null  object
11  Current Ver      10833 non-null  object
12  Android Ver      10838 non-null  object
dtypes: float64(1), object(12)
memory usage: 1.1+ MB
```

Missing Value



Google Play

```
[6]: missing_value = data.isnull().sum()  
missing_value
```

```
[6]: App          0  
     Category     0  
     Rating      1474  
     Reviews      0  
     Size         0  
     Installs     0  
     Type         1  
     Price        0  
     Content Rating 1  
     Genres       0  
     Last Updated  0  
     Current Ver   8  
     Android Ver   3  
     dtype: int64
```



Mean

```
[7]: mean_value = data["Rating"].mean()  
     data["Rating"].fillna(mean_value, inplace = True)
```

```
[8]: data.describe()
```

```
[8]:
```

	Rating
count	10841.000000
mean	4.193338
std	0.499557
min	1.000000
25%	4.100000
50%	4.200000
75%	4.500000
max	19.000000

```
[9]: data.describe().T
```

```
[9]:
```

	count	mean	std	min	25%	50%	75%	max
Rating	10841.0	4.193338	0.499557	1.0	4.1	4.2	4.5	19.0

Data Describe



Google Play

```
[10]: data.describe(include='object')
```

	App	Category	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
count	10841	10841	10841	10841	10841	10840	10841	10840	10841	10841	10833	10838
unique	9660	34	6002	462	22	3	93	6	120	1378	2832	33
top	ROBLOX	FAMILY	0	Varies with device	1,000,000+	Free	0	Everyone	Tools	August 3, 2018	Varies with device	4.1 and up
freq	9	1972	596	1695	1579	10039	10040	8714	842	326	1459	2451

```
[11]: data.describe(include='object').T
```

	count	unique	top	freq
App	10841	9660	ROBLOX	9
Category	10841	34	FAMILY	1972
Reviews	10841	6002	0	596
Size	10841	462	Varies with device	1695
Installs	10841	22	1,000,000+	1579

Duplicate Data



Google Play

10836	Sya9a Maroc - FR	FAMILY	4.500000	38	53M	5,000+	Free	0	Everyone	Education	July 25, 2017	1.48	4.1 and up
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.000000	4	3.6M	100+	Free	0	Everyone	Education	July 6, 2018	1.0	4.1 and up
10838	Parkinson Exercices FR	MEDICAL	4.193338	3	9.5M	1,000+	Free	0	Everyone	Medical	January 20, 2017	1.0	2.2 and up
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.500000	114	Varies with device	1,000+	Free	0	Mature 17+	Books & Reference	January 19, 2015	Varies with device	Varies with device
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.500000	398307	19M	10,000,000+	Free	0	Everyone	Lifestyle	July 25, 2018	Varies with device	Varies with device

10358 rows x 13 columns

```
4]: data.duplicated().sum()
```

```
4]: 0
```

Category Data



Google Play

```
[15]: data["Category"].value_counts()
```

```
[15]: Category
FAMILY                1943
GAME                  1121
TOOLS                  843
BUSINESS               427
MEDICAL               408
PRODUCTIVITY          407
PERSONALIZATION       388
LIFESTYLE              373
COMMUNICATION          366
FINANCE                360
SPORTS                 351
PHOTOGRAPHY           322
HEALTH_AND_FITNESS    306
SOCIAL                 280
NEWS_AND_MAGAZINES    264
TRAVEL_AND_LOCAL      237
BOOKS_AND_REFERENCE   230
SHOPPING              224
DATING                 196
VIDEO_PLAYERS         175
MAPS_AND_NAVIGATION   137
EDUCATION              130
FOOD_AND_DRINK        124
ENTERTAINMENT          111
AUTO_AND_VEHICLES      85
LIBRARIES_AND_DEMO     85
```

Category Data



Google Play

```
[16]: data["Category"].unique()
```

```
[16]: array(['ART_AND_DESIGN', 'AUTO_AND_VEHICLES', 'BEAUTY',  
          'BOOKS_AND_REFERENCE', 'BUSINESS', 'COMICS', 'COMMUNICATION',  
          'DATING', 'EDUCATION', 'ENTERTAINMENT', 'EVENTS', 'FINANCE',  
          'FOOD_AND_DRINK', 'HEALTH_AND_FITNESS', 'HOUSE_AND_HOME',  
          'LIBRARIES_AND_DEMO', 'LIFESTYLE', 'GAME', 'FAMILY', 'MEDICAL',  
          'SOCIAL', 'SHOPPING', 'PHOTOGRAPHY', 'SPORTS', 'TRAVEL_AND_LOCAL',  
          'TOOLS', 'PERSONALIZATION', 'PRODUCTIVITY', 'PARENTING', 'WEATHER',  
          'VIDEO_PLAYERS', 'NEWS_AND_MAGAZINES', 'MAPS_AND_NAVIGATION',  
          '1.9'], dtype=object)
```



Google Play

Rating

```
[17]: import numpy as np
unique = 'ART_AND_DESIGN', 'AUTO_AND_VEHICLES', 'BEAUTY', 'BOOKS_AND_REFERENCE', 'BUSINESS', 'COMICS', 'COMMUNICATION', 'DATING', 'EDUCATION', 'ENTERTAINME
data["Category"] = data["Category"].apply(lambda x: x if x in unique else np.nan)
data["Category"].unique()
data.dropna(subset = ["Category"], inplace = True)
data.Category.isnull().sum()
```

```
[17]: 0
```

```
[18]: data.Rating.isnull().sum()
```

```
[18]: 0
```

```
[19]: data.Rating.value_counts()
```

```
[19]: Rating
4.193338    1465
4.400000    1032
4.300000    1016
4.500000     977
4.200000     888
4.600000     768
4.100000     657
4.000000     539
4.700000     484
3.900000     372
3.800000     293
5.000000     271
3.700000     231
4.800000     228
```



Google Play

Rating

```
1.800000    8
2.100000    8
1.600000    4
1.400000    3
1.500000    3
1.200000    1
Name: count, dtype: int64
```

```
[20]: data["Rating_category"] = data["Rating"].apply (lambda x: "High" if x>3.5 else "Low")
```

```
[21]: data["Rating_category"].value_counts()
```

```
[21]: Rating_category
High    9477
Low      880
Name: count, dtype: int64
```



PIP Install

```
[22]: # Pip Install
!pip install numpy
```

Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: numpy in c:\programdata\anaconda3\lib\site-packages (1.26.4)

```
[23]: # Packages
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
[24]: data1 = pd.read_csv("googleplaystore.csv")
data1
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with ...	4.2 and up



Data Type

```
data1.shape
```

```
(10841, 13)
```

```
data1.columns
```

```
Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',  
      'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',  
      'Android Ver'],  
      dtype='object')
```

```
data1.dtypes
```

App	object
Category	object
Rating	float64
Reviews	object
Size	object
Installs	object
Type	object
Price	object
Content Rating	object
Genres	object
Last Updated	object
Current Ver	object
Android Ver	object
dtype:	object

Distribution App Ratings



Google Play

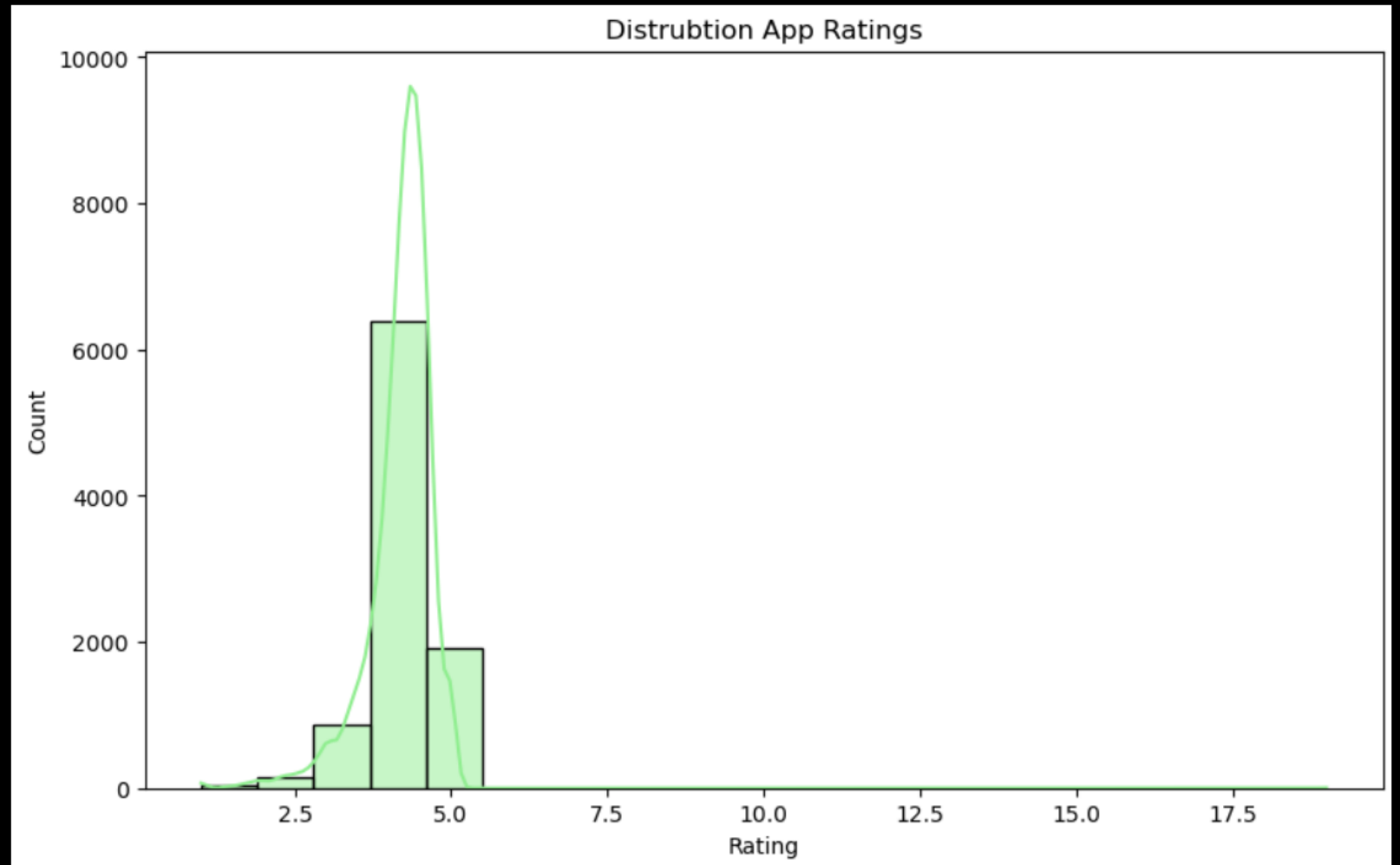
```
[32]: # Distribution App Ratings
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(10,6))
sns.histplot(data1["Rating"].dropna(),kde=True,bins=20,color="lightgreen")
plt.title("Distrubtion App Ratings")
plt.xlabel("Rating")
plt.ylabel("Frequency")
plt.show()
```

Histoplot Graph



Google Play

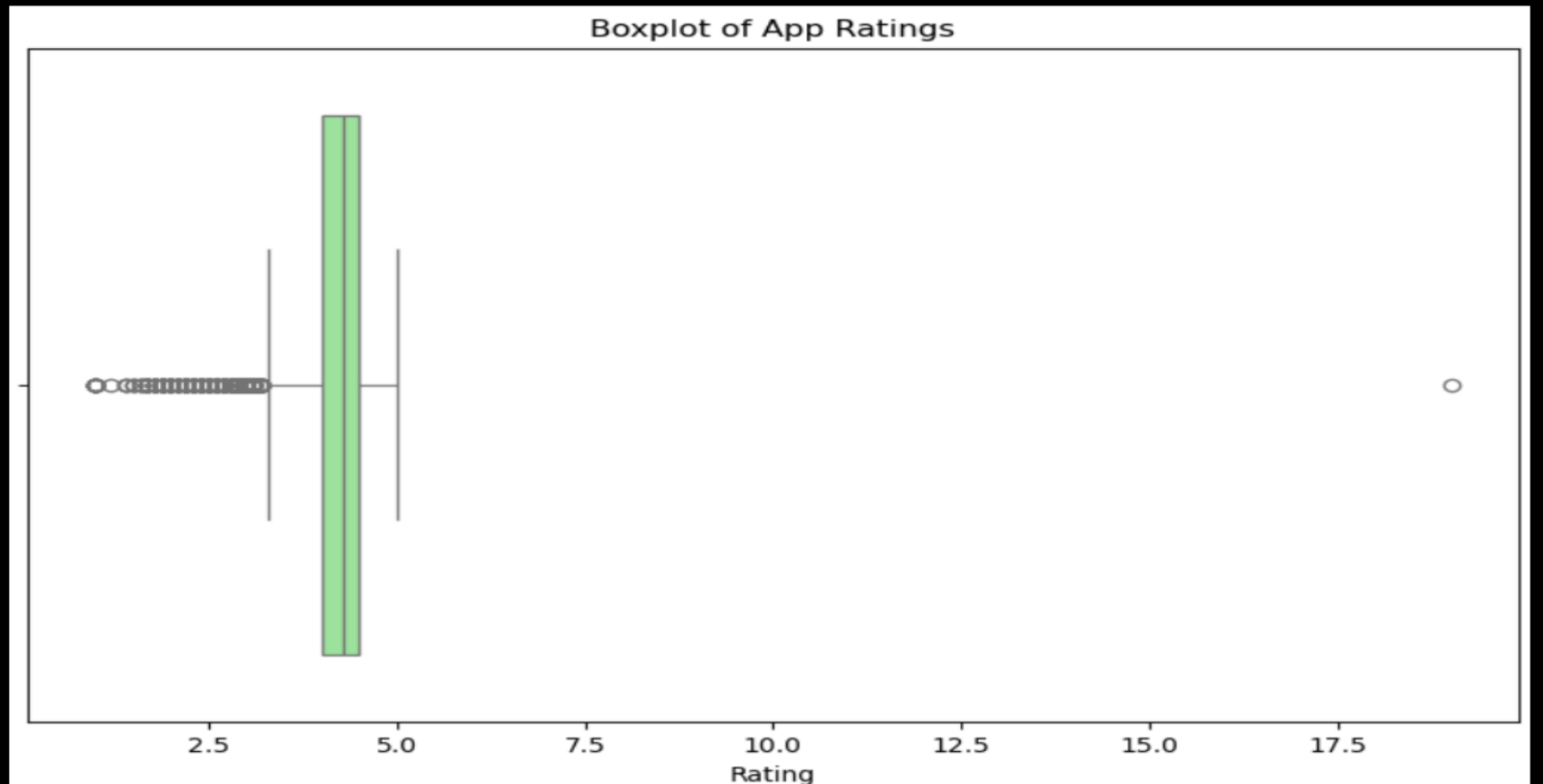


Boxplot Graph



Google Play

```
plt.figure(figsize=(10,6))
sns.boxplot(x = 'Rating',data=data1, color='lightgreen')
plt.title("Boxplot of App Ratings")
plt.xlabel("Rating")
plt.show()
```



Outlier and Boxplot



Google Play

```
: Q1=data1["Rating"].quantile(0.25)
: Q3=data1["Rating"].quantile(0.75)
: IQR=Q3-Q1
: lower_bound=Q1-1.5*IQR
: upper_bound=Q1+1.5*IQR

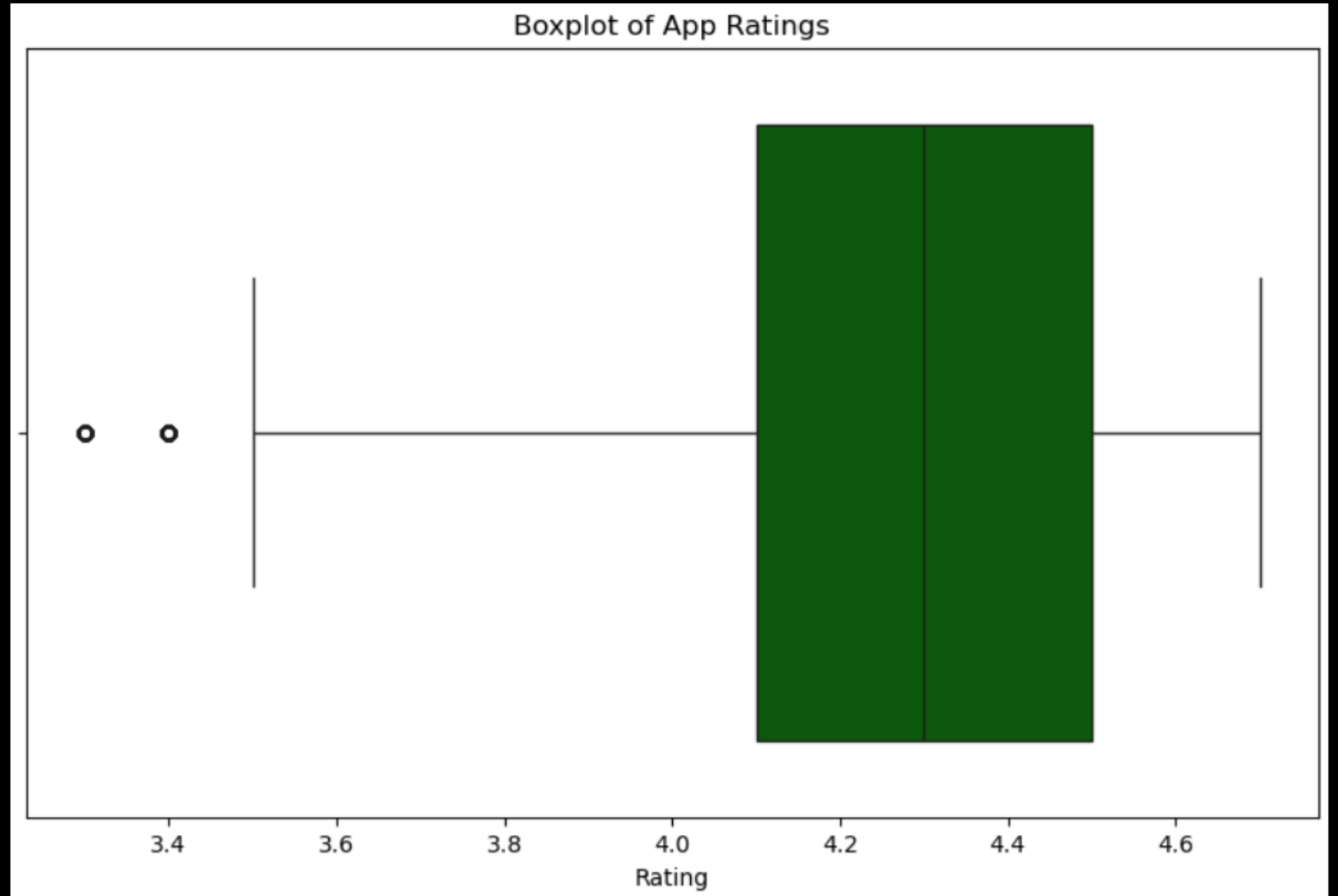
: filterdata1=data1[(data1["Rating"]>=lower_bound)&(data1["Rating"]<=upper_bound)]

: plt.figure(figsize=(10,6))
: sns.boxplot(x = 'Rating' ,data=filterdata1,color='Darkgreen')
: plt.title("Boxplot of App Ratings")
: plt.xlabel("Rating")
: plt.show()
```

Boxplot Outlier used for Size



Google Play





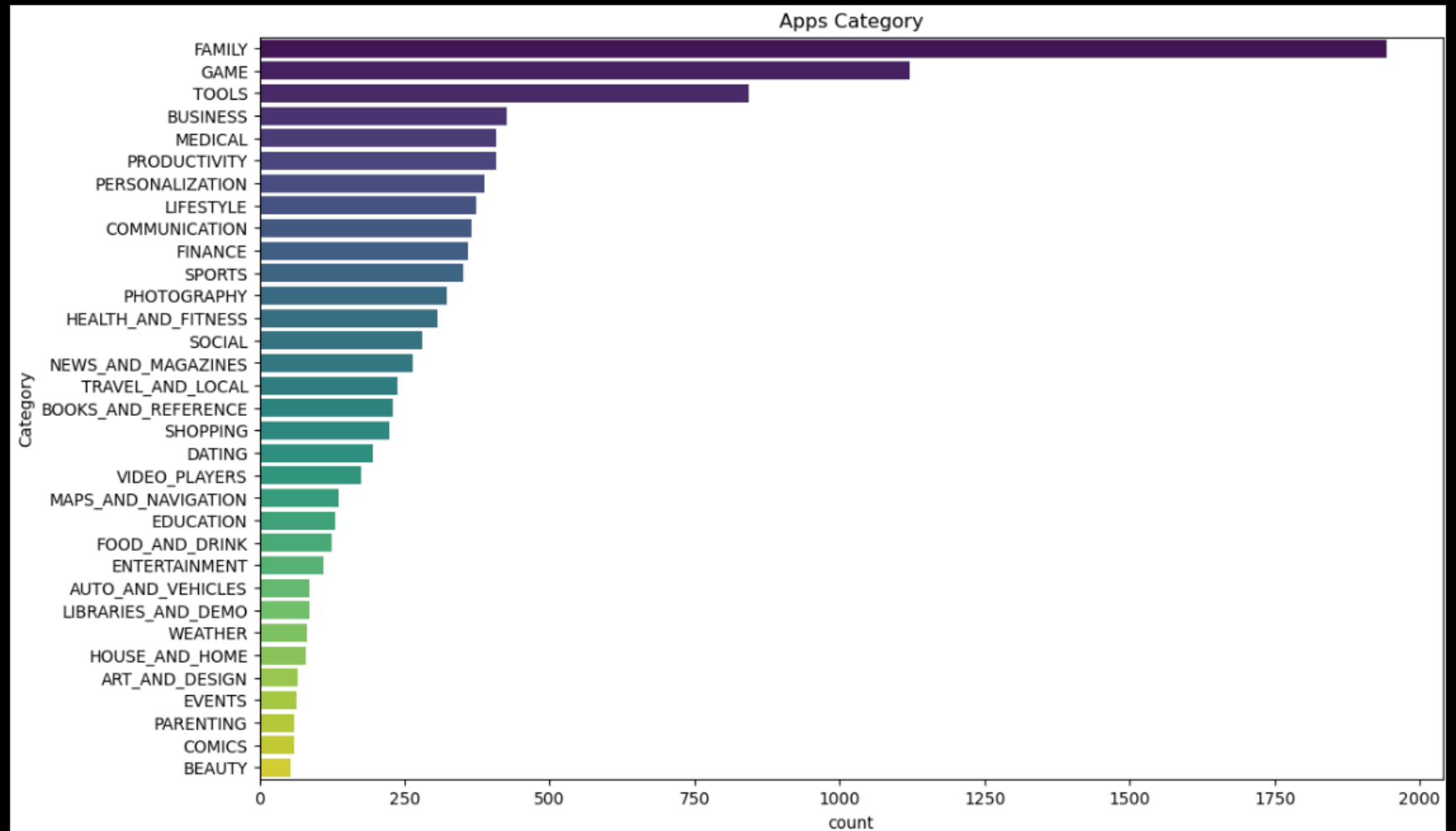
Google Play

Count plot

```
plt.figure(figsize=(13,8))
sns.countplot(y='Category', data=data,order=data['Category'].value_counts().index,palette='viridis')
plt.title('Apps Category')
plt.xlabel('Count')
plt.ylabel('Category')
plt.show()
```



Count plot





Google Play

Thank You