Aaron_JOJU_201_LAB11

Perform the Exploratory Data Analysis on your domain-based dataset and demonstrate the retrieved insights using “Matplotlib” modules.
Visualize hidden insights using appropriate plots (graphs) [Usage of line plot and scatter plot are mandatory]

```
import pandas as pd
import matplotlib.pyplot as plt
```

```
df = pd.read_csv('googleplaystore.csv')
```

```
print(df.head())
```

```
                                                 App        Category  Rating  \
0     Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1
1                                Coloring book moana  ART_AND_DESIGN     3.9
2  U Launcher Lite – FREE Live Cool Themes, Hide ...  ART_AND_DESIGN     4.7
3                              Sketch - Draw & Paint  ART_AND_DESIGN     4.5
4              Pixel Draw - Number Art Coloring Book  ART_AND_DESIGN     4.3

  Reviews  Size     Installs  Type Price Content Rating  \
0     159   19M      10,000+  Free     0       Everyone
1     967   14M     500,000+  Free     0       Everyone
2   87510  8.7M   5,000,000+  Free     0       Everyone
3  215644   25M  50,000,000+  Free     0           Teen
4     967  2.8M     100,000+  Free     0       Everyone

                      Genres      Last Updated         Current Ver  \
0               Art & Design   January 7, 2018               1.0.0
1  Art & Design;Pretend Play  January 15, 2018               2.0.0
2               Art & Design    August 1, 2018               1.2.4
3               Art & Design      June 8, 2018  Varies with device
4    Art & Design;Creativity     June 20, 2018                 1.1

    Android Ver
0  4.0.3 and up
1  4.0.3 and up
2  4.0.3 and up
3    4.2 and up
4    4.4 and up
```

```
print(df.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
None
```

```
print(df.describe())
```

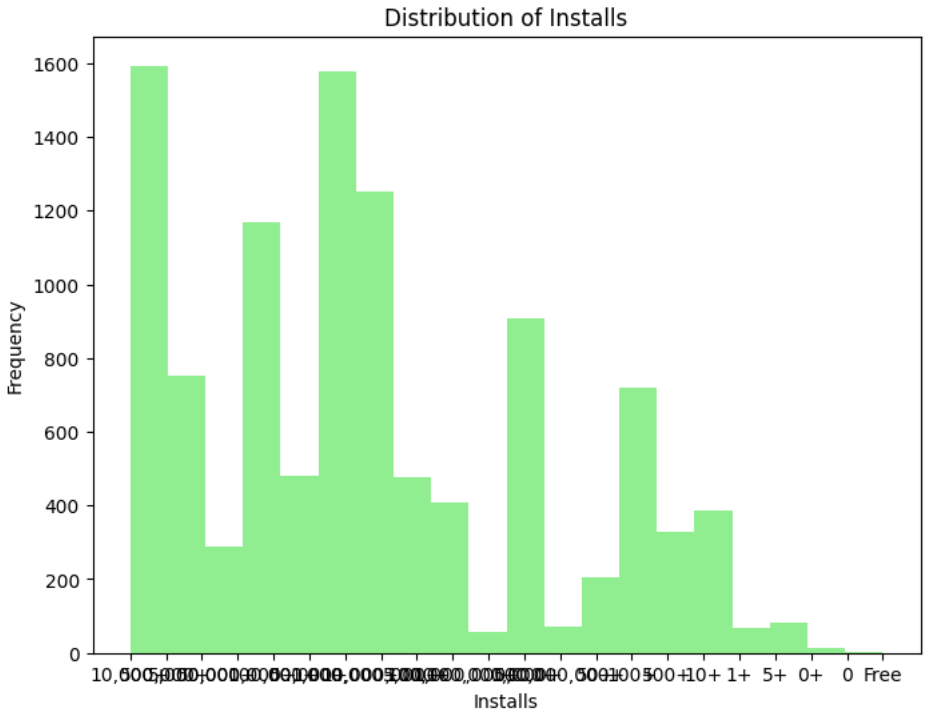
```
             Rating
count  9367.000000
mean      4.193338
std       0.537431
min       1.000000
25%       4.000000
50%       4.300000
75%       4.500000
max      19.000000
```

```
print(df.describe())
```

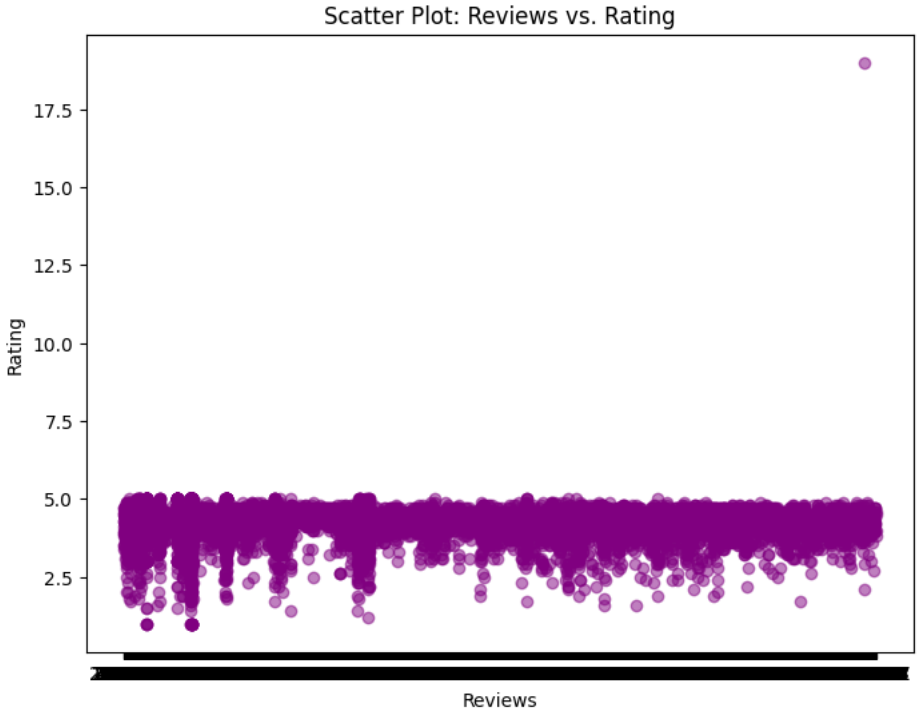
```
             Rating
count  9367.000000
mean      4.193338
std       0.537431
min       1.000000
25%       4.000000
```

```
50%        4.300000
75%        4.500000
max       19.000000
```

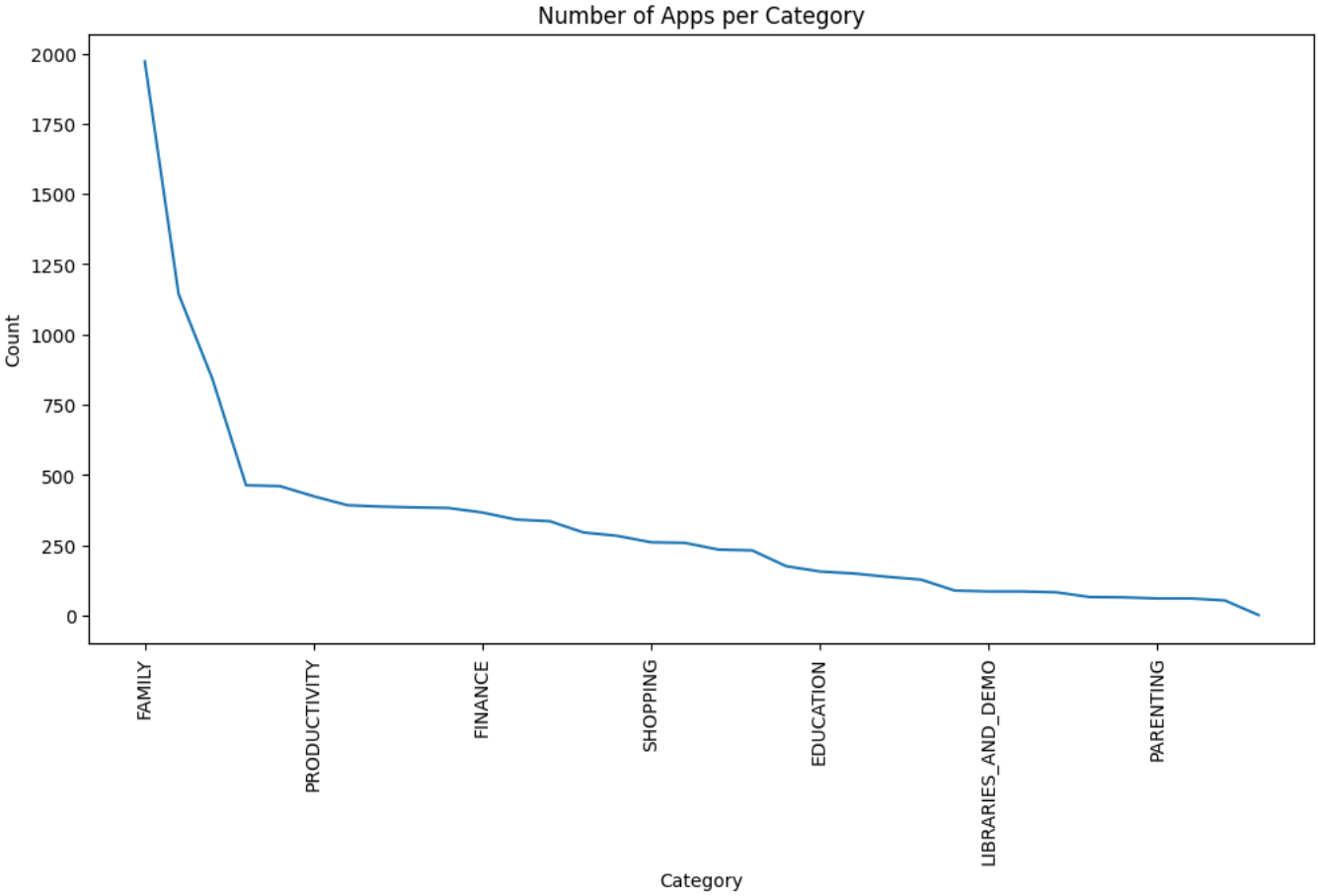
```
plt.figure(figsize=(8, 6))
plt.hist(df['Installs'], bins=20, color='lightgreen')
plt.xlabel('Installs')
plt.ylabel('Frequency')
plt.title('Distribution of Installs')
plt.show()
```




```
plt.figure(figsize=(8, 6))
plt.scatter(df['Reviews'], df['Rating'], alpha=0.5, color='purple')
plt.xlabel('Reviews')
plt.ylabel('Rating')
plt.title('Scatter Plot: Reviews vs. Rating')
plt.show()
```




```
plt.figure(figsize=(12, 6))
df['Category'].value_counts().plot(kind='line')
plt.xlabel('Category')
plt.ylabel('Count')
plt.title('Number of Apps per Category')
plt.xticks(rotation=90)
plt.show()
```


Number of Apps per Category
Count
2000
1750
1500
1250
1000
750
500
250
0
FAMILY
PRODUCTIVITY
FINANCE
SHOPPING
EDUCATION
LIBRARIES_AND_DEMO
PARENTING
Category