

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from textblob import TextBlob

import warnings
warnings.filterwarnings('ignore')
plt.style.use('seaborn')
```

C:\Users\patil\anaconda3\lib\site-packages\pandas\core\computation\expressions.py:21: UserWarning: Pandas requires version '2.8.4' or newer of 'numexpr' (version '2.8.3' currently installed).
 from pandas.core.computation.check import NUMEXPR_INSTALLED
C:\Users\patil\anaconda3\lib\site-packages\pandas\core\arrays\masked.py:61: UserWarning: Pandas requires version '1.3.6' or newer of 'bottleneck' (version '1.3.5' currently installed).
 from pandas.core import (

```
In [2]: apps = pd.read_csv(r'D:\Oasis\Task 8\archive (11)\datasets\apps.csv')
reviews = pd.read_csv(r'D:\Oasis\Task 8\archive (11)\datasets\user_reviews.csv')

apps.head()
reviews.head()
```

```
Out[2]:
```

	App	Translated_Review	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity
0	10 Best Foods for You	I like eat delicious food. That's I'm cooking ...	Positive	1.00	0.533333
1	10 Best Foods for You	This help eating healthy exercise regular basis	Positive	0.25	0.288462
2	10 Best Foods for You	NaN	NaN	NaN	NaN
3	10 Best Foods for You	Works great especially going grocery store	Positive	0.40	0.875000
4	10 Best Foods for You	Best idea us	Positive	1.00	0.300000

```
In [3]: apps.isnull().sum()
reviews.isnull().sum()
```

```
Out[3]: App                                0
Translated_Review                        26868
Sentiment                               26863
Sentiment_Polarity                       26863
Sentiment_Subjectivity                   26863
dtype: int64
```

```
In [4]: apps['Installs'] = apps['Installs'].str.replace('+','').str.replace(',','')
apps['Installs'] = pd.to_numeric(apps['Installs'], errors='coerce')
```

```
In [5]: apps['Price'] = apps['Price'].str.replace('$','')
apps['Price'] = pd.to_numeric(apps['Price'], errors='coerce')
```

```
In [6]: def convert_size(size):
    if isinstance(size, str): # only process strings
        size = size.strip()
        if 'M' in size:
            return float(size.replace('M', '')) * 1_000_000
        elif 'k' in size:
            return float(size.replace('k', '')) * 1_000
        else:
            return np.nan
    return np.nan

apps['Size'] = apps['Size'].replace('Varies with device', np.nan)

apps['Size'] = apps['Size'].apply(convert_size)
```

```
In [7]: apps['Reviews'] = pd.to_numeric(apps['Reviews'], errors='coerce')
apps['Rating'] = pd.to_numeric(apps['Rating'], errors='coerce')
```

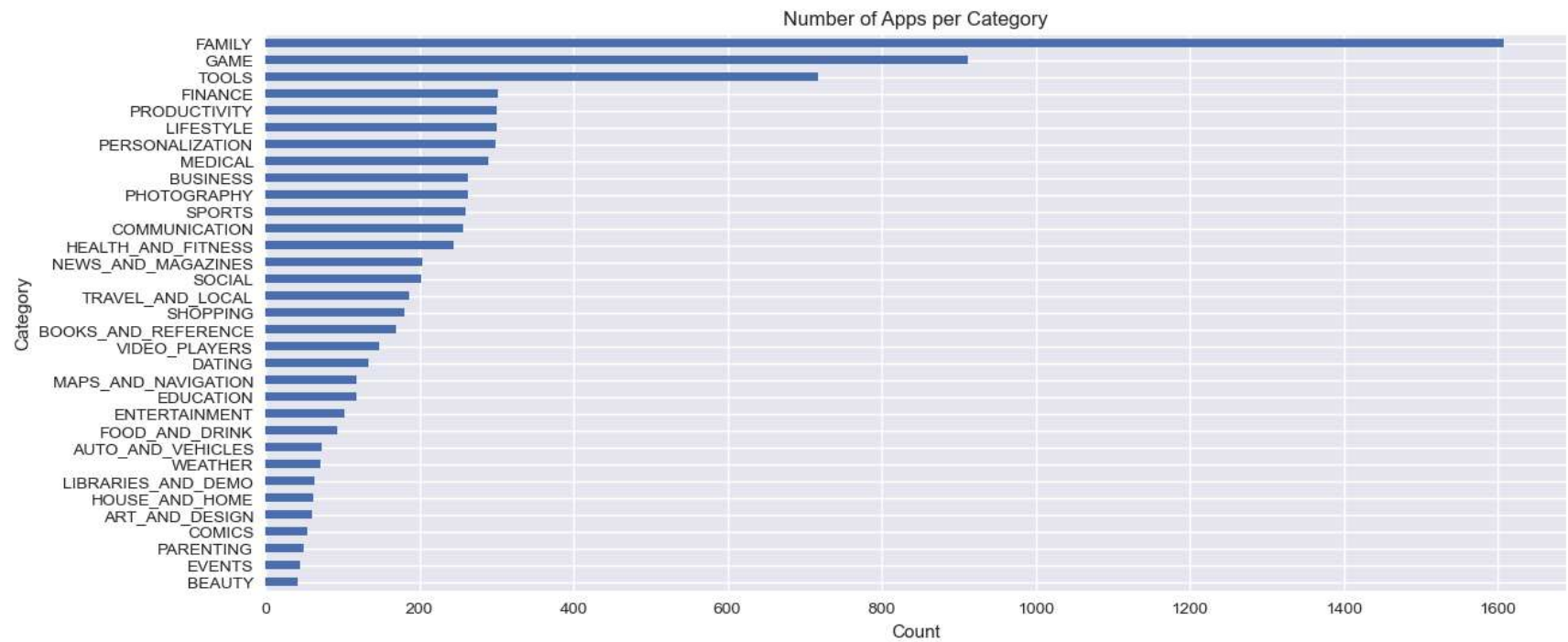
```
In [8]: apps_clean = apps.dropna(subset=['Rating'])
apps_clean.head()
```

Out[8]:

Unnamed: 0	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Andriod
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	NaN	10000	Free	0.0	Everyone	Art & Design	January 7, 2018	1.0.0	4.1
1	Coloring book moana	ART_AND_DESIGN	3.9	967	NaN	500000	Free	0.0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.1
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	NaN	5000000	Free	0.0	Everyone	Art & Design	August 1, 2018	1.2.4	4.1
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	NaN	50000000	Free	0.0	Teen	Art & Design	June 8, 2018	Varies with device	4.2
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	NaN	100000	Free	0.0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4

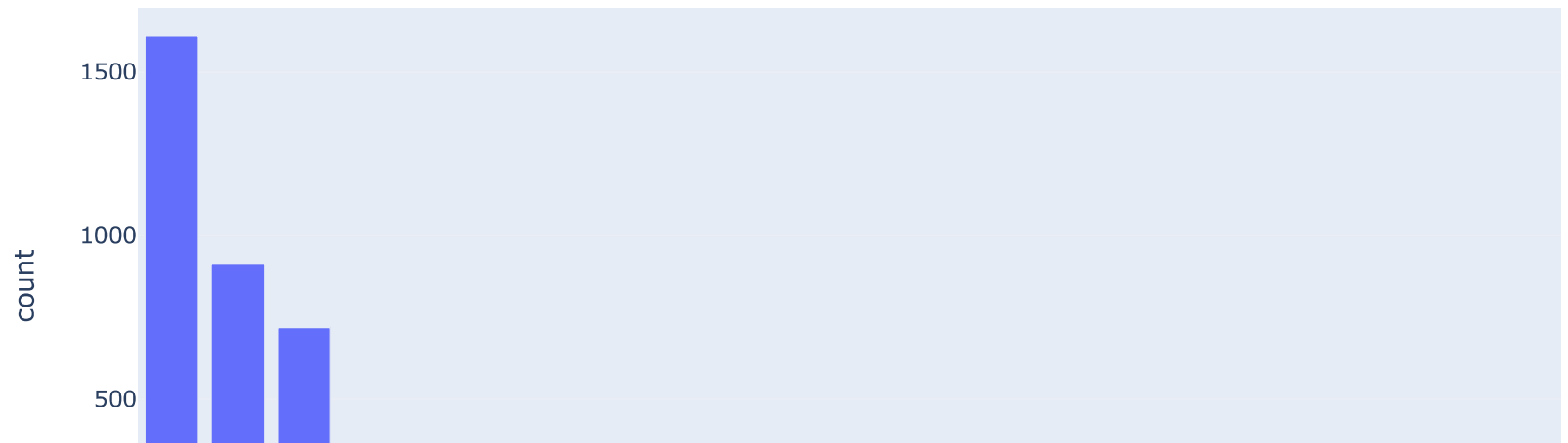
In [9]:

```
plt.figure(figsize=(14,6))
apps_clean['Category'].value_counts().sort_values().plot(kind='barh')
plt.title("Number of Apps per Category")
plt.xlabel("Count")
plt.ylabel("Category")
plt.show()
```

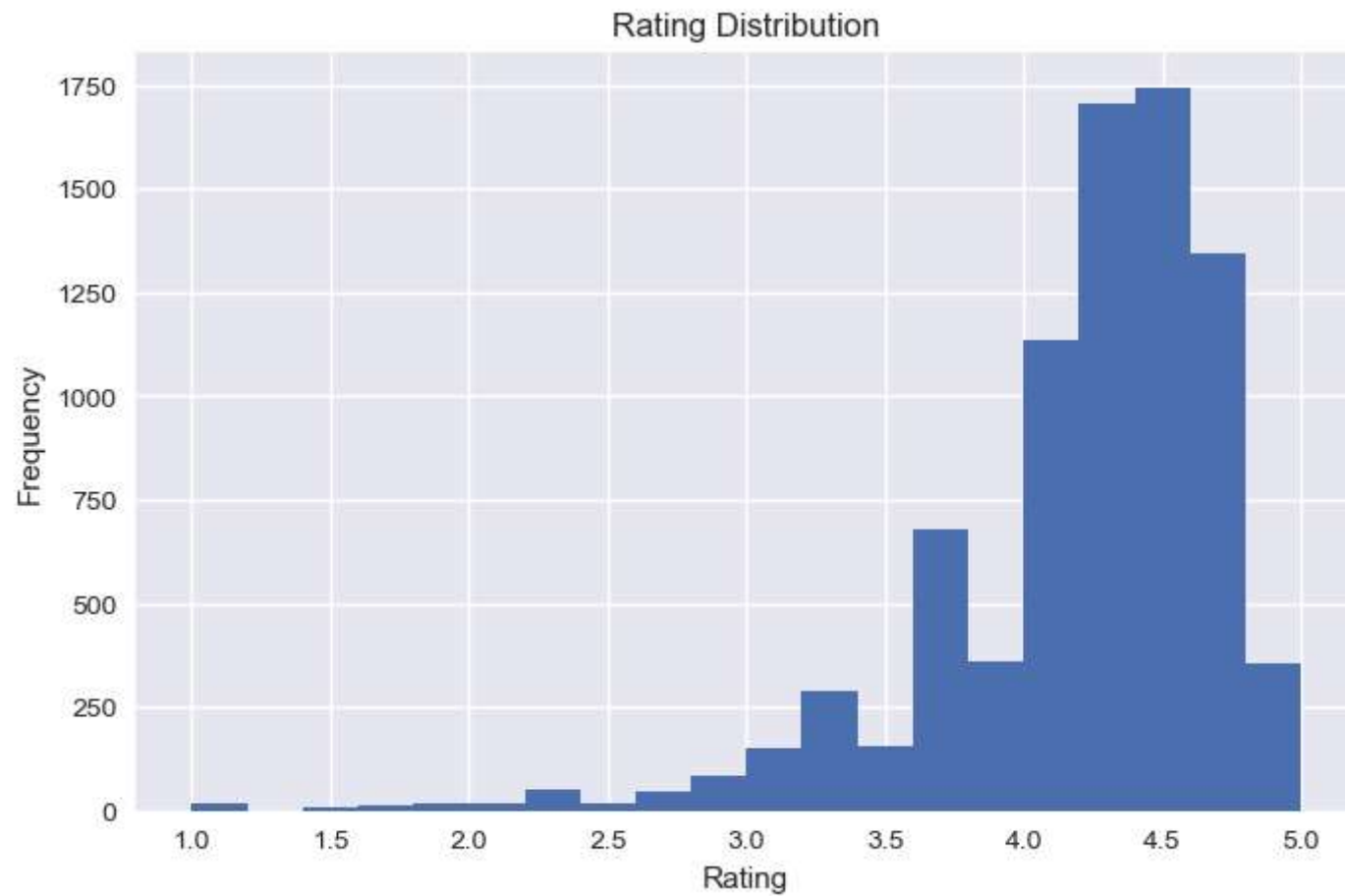


```
In [10]: fig = px.bar(  
    apps_clean['Category'].value_counts().reset_index(),  
    x='Category', y='count',  
    title="Interactive: App Count by Category"  
)  
fig.show()
```

Interactive: App Count by Category

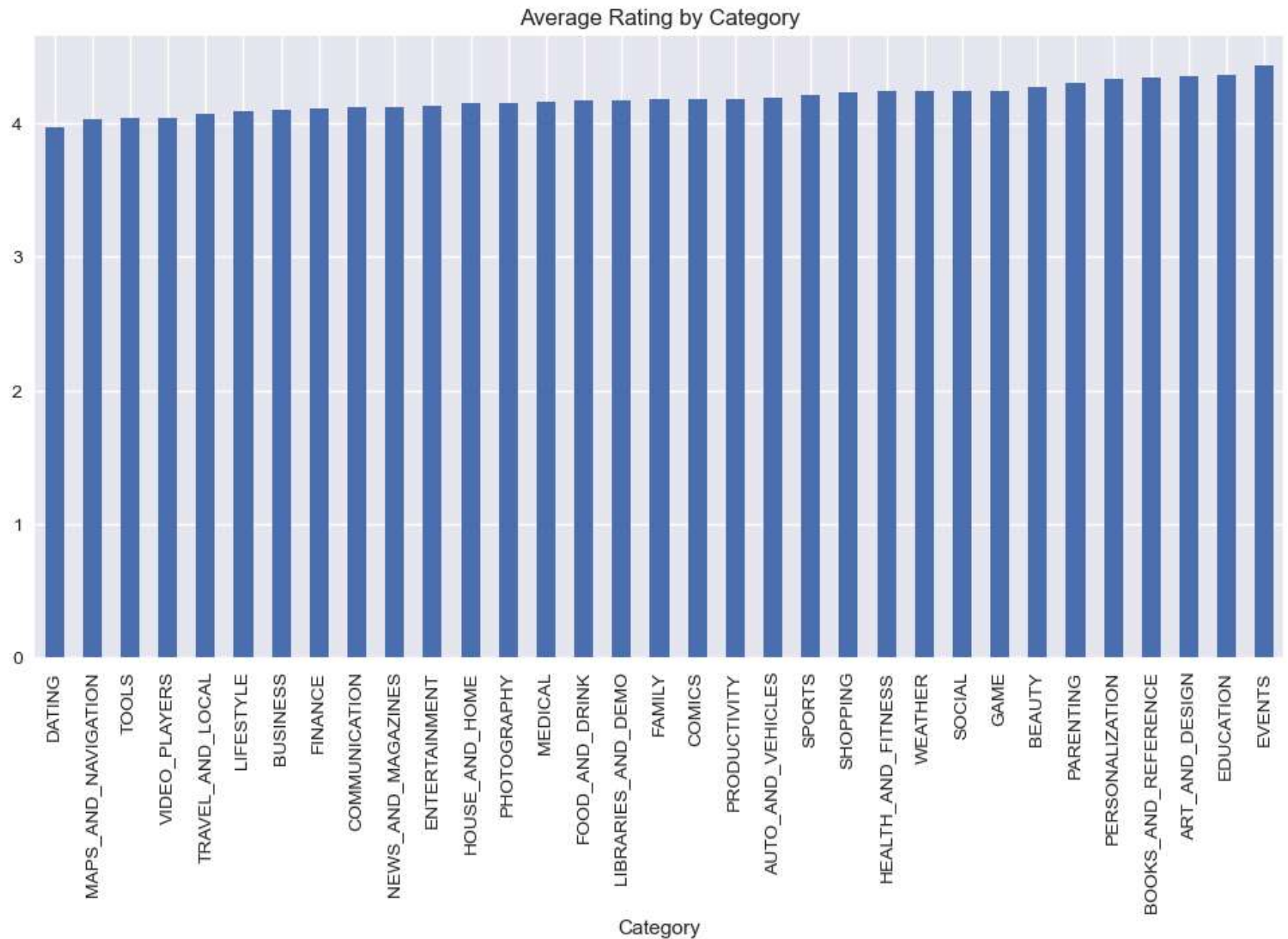


```
In [11]: plt.figure(figsize=(8,5))
plt.hist(apps_clean['Rating'].dropna(), bins=20)
plt.title("Rating Distribution")
plt.xlabel("Rating")
plt.ylabel("Frequency")
plt.show()
```

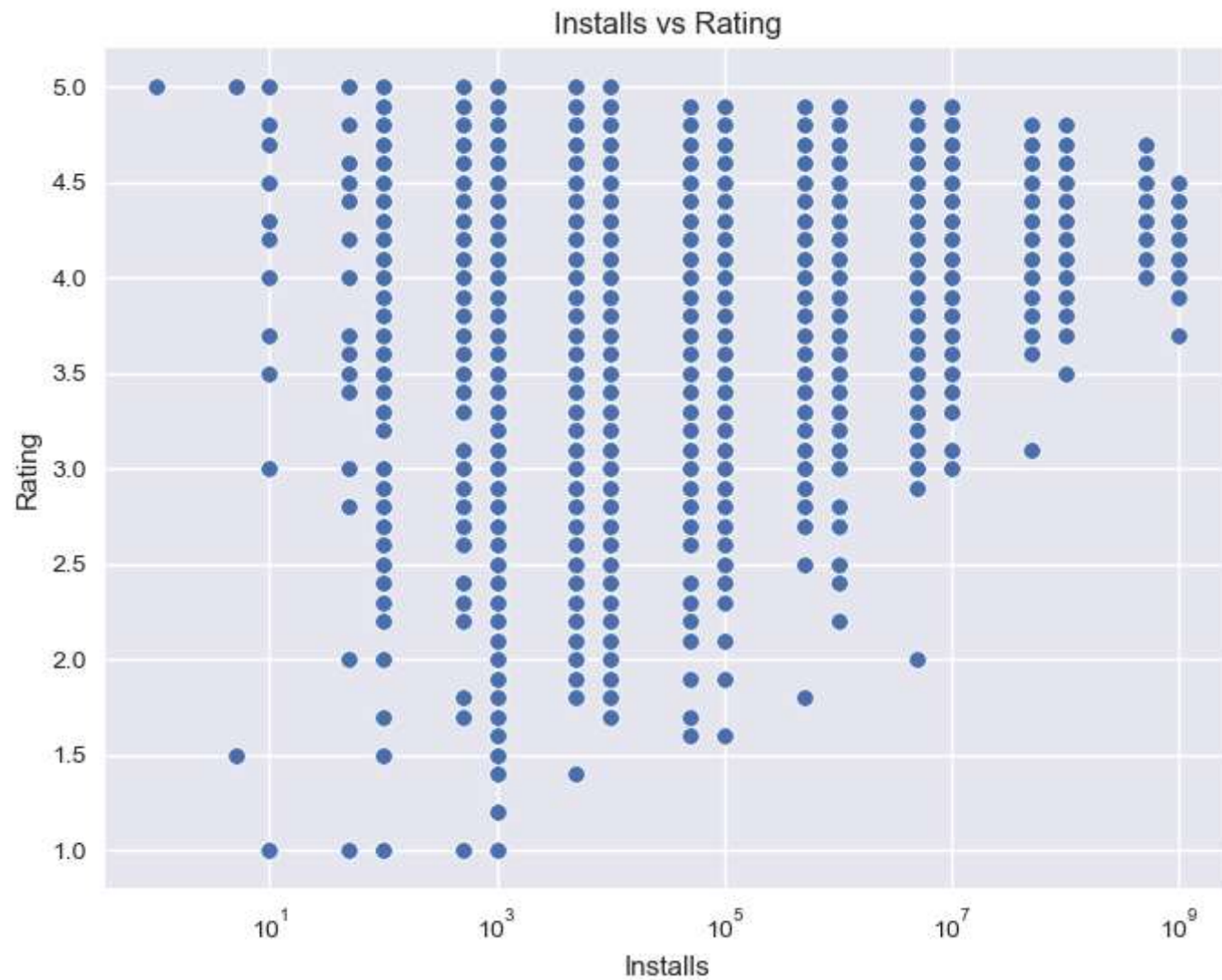


```
In [12]: rating_by_category = apps_clean.groupby('Category')['Rating'].mean().sort_values()

plt.figure(figsize=(12,6))
rating_by_category.plot(kind='bar')
plt.title("Average Rating by Category")
plt.show()
```

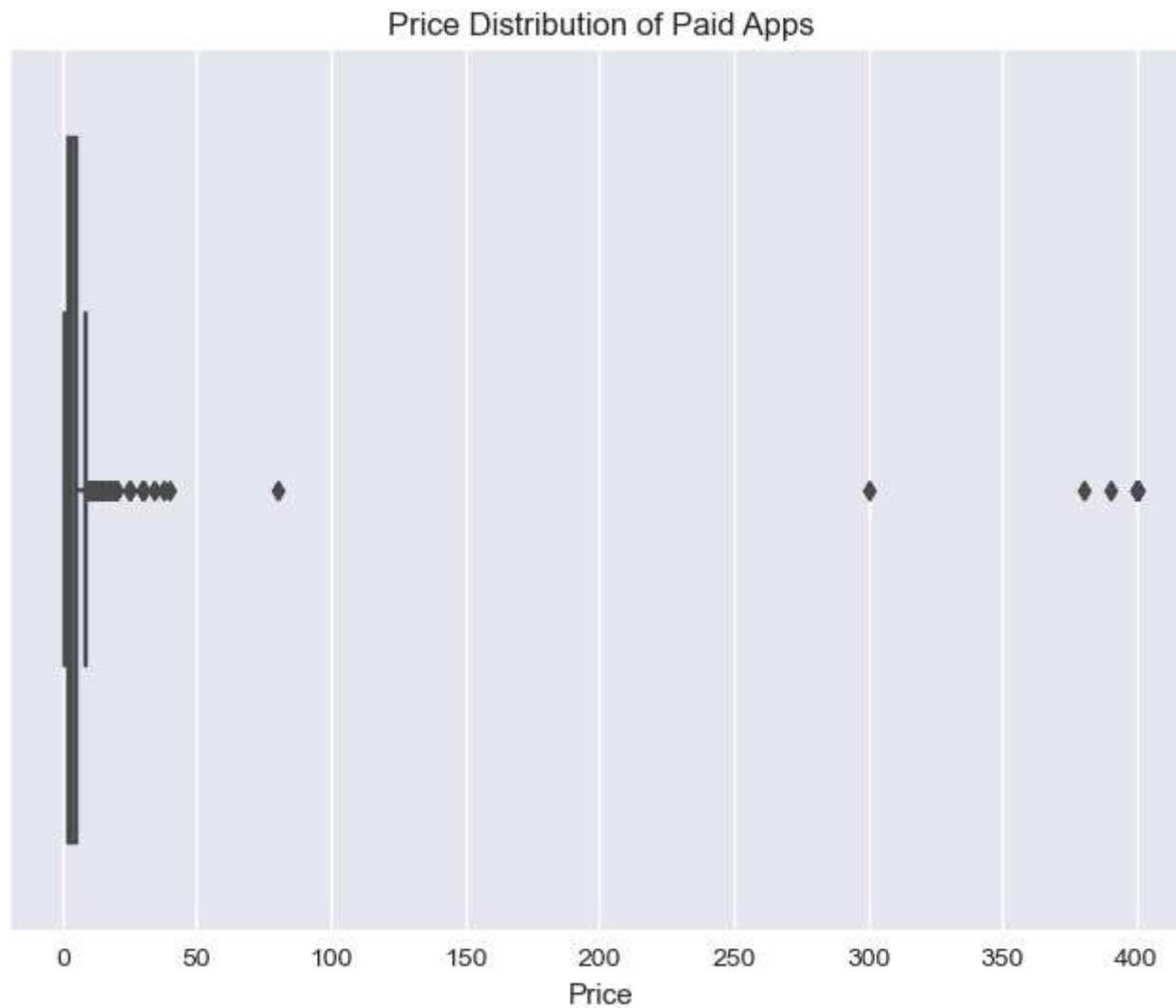


```
In [13]: plt.figure(figsize=(8,6))
sns.scatterplot(data=apps_clean, x='Installs', y='Rating')
plt.xscale('log')
plt.title("Installs vs Rating")
plt.show()
```



```
In [14]: paid_apps = apps_clean[apps_clean['Price']>0]
```

```
plt.figure(figsize=(8,6))
sns.boxplot(x=paid_apps['Price'])
plt.title("Price Distribution of Paid Apps")
plt.show()
```

```
In [15]: reviews = reviews.dropna(subset=['Translated_Review'])
```

```
In [16]: def get_sentiment(text):  
         return TextBlob(text).sentiment.polarity  
  
reviews['Sentiment_Polarity'] = reviews['Translated_Review'].apply(get_sentiment)  
  
reviews.head()
```

Out[16]:

	App	Translated_Review	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity
0	10 Best Foods for You	I like eat delicious food. That's I'm cooking ...	Positive	1.00	0.533333
1	10 Best Foods for You	This help eating healthy exercise regular basis	Positive	0.25	0.288462
3	10 Best Foods for You	Works great especially going grocery store	Positive	0.40	0.875000
4	10 Best Foods for You	Best idea us	Positive	1.00	0.300000
5	10 Best Foods for You	Best way	Positive	1.00	0.300000

In [17]:

```
sentiment_summary = reviews.groupby('App')['Sentiment_Polarity'].mean().reset_index()  
sentiment_summary.head()
```

Out[17]:

	App	Sentiment_Polarity
0	10 Best Foods for You	0.470733
1	104 找工作 - 找工作 找打工 找兼職 履歷健檢 履歷診療室	0.392405
2	11st	0.185943
3	1800 Contacts - Lens Store	0.318145
4	1LINE – One Line with One Touch	0.196290

In [18]:

```
apps_sentiment = apps_clean.merge(sentiment_summary, on='App', how='left')  
apps_sentiment.head()
```

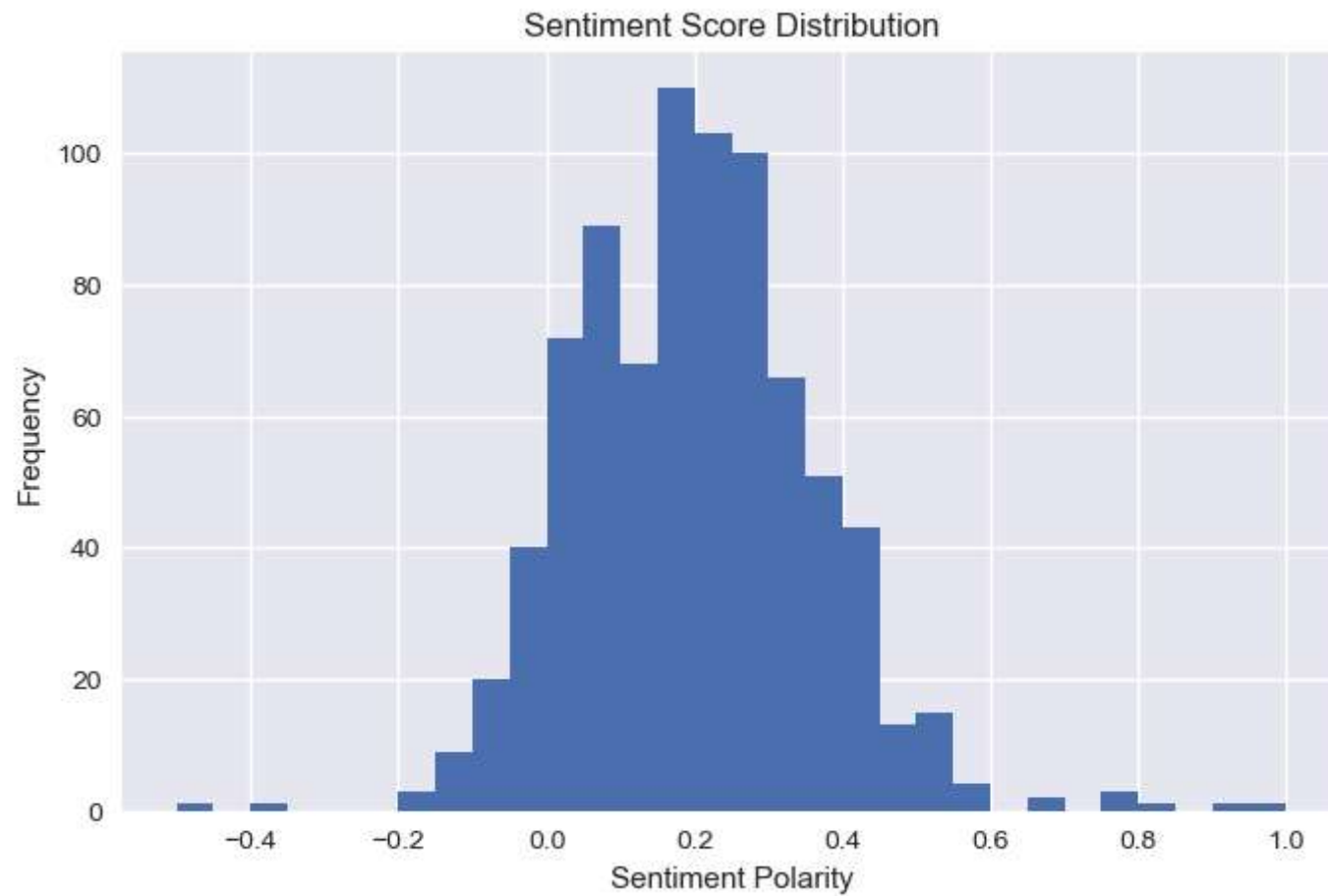
Out[18]:

Unnamed: 0	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Andriod
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	NaN	10000	Free	0.0	Everyone	Art & Design	January 7, 2018	1.0.0	4.1
1	Coloring book moana	ART_AND_DESIGN	3.9	967	NaN	500000	Free	0.0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.1
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	NaN	5000000	Free	0.0	Everyone	Art & Design	August 1, 2018	1.2.4	4.1
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	NaN	50000000	Free	0.0	Teen	Art & Design	June 8, 2018	Varies with device	4.2
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	NaN	100000	Free	0.0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4

```

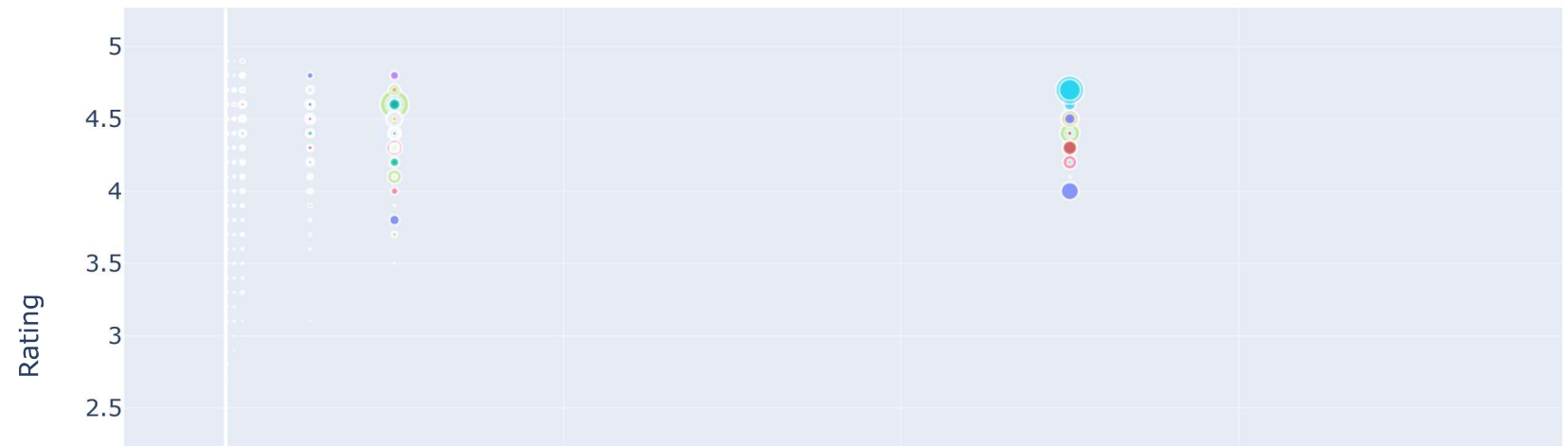
In [19]: plt.figure(figsize=(8,5))
plt.hist(apps_sentiment['Sentiment_Polarity'].dropna(), bins=30)
plt.title("Sentiment Score Distribution")
plt.xlabel("Sentiment Polarity")
plt.ylabel("Frequency")
plt.show()

```



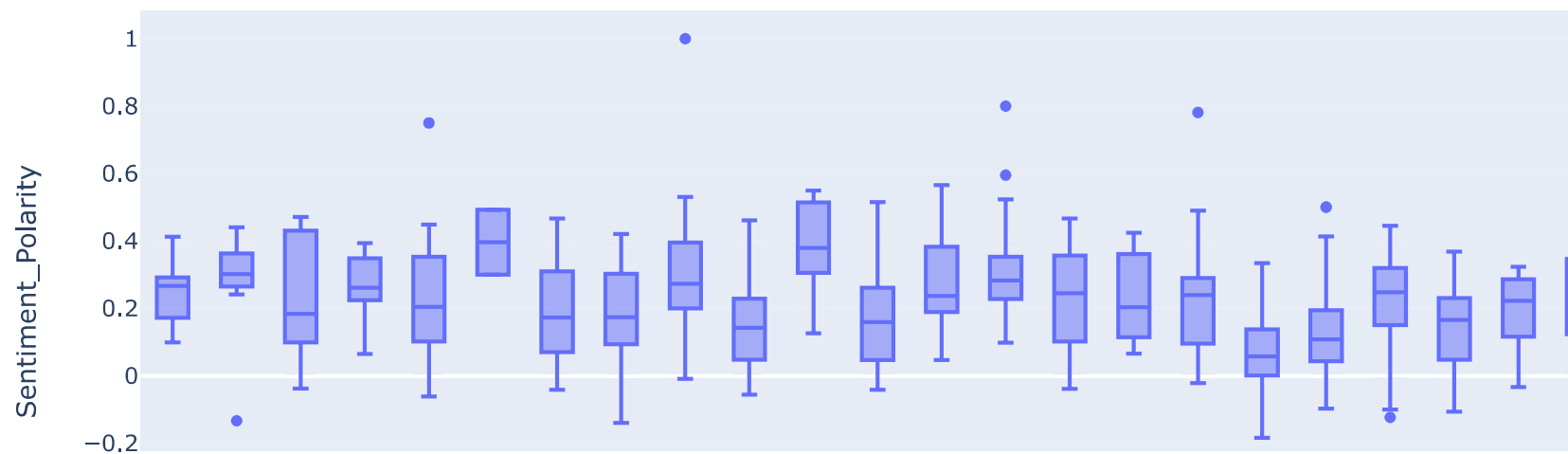
```
In [20]: fig = px.scatter(  
    apps_clean, x='Installs', y='Rating',  
    color='Category',  
    size='Reviews',  
    title="Interactive: Installs vs Rating"  
)  
fig.show()
```

Interactive: Installs vs Rating



```
In [21]: fig = px.box(  
    apps_sentiment, x='Category', y='Sentiment_Polarity',  
    title="Sentiment by Category (Interactive)"  
)  
fig.show()
```

Sentiment by Category (Interactive)



In []: