In [22]:
```
import pandas as pd
```

In [23]:
```
data=pd.read_csv(r'E:\Data Analyst Project\googleplaystore.csv')
```

This Google Play Store Apps can be found on: https://www.kaggle.com/datasets/lava18/google-play-store-apps

In [24]:
```
# I would like to appreciate to Data Thinkers on Youtube and the google play store apps dataset owner
```

Data Analysis of Google Play Store Apps

1. Show top 5 Rows of the Dataset

In [25]:
```
data.head(5)
```

Out[25]:

	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

2. Show the last 10 Rows of the Dataset

In [26]:
```
data.tail(10)
```

Out[26]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
10831	payermonstationnement.fr	MAPS_AND_NAVIGATION	NaN	38	9.8M	5,000+	Free	0	Everyone	Maps & Navigation	June 13, 2018	2.0.148.0	4.0 and up
10832	FR Tides	WEATHER	3.8	1195	582k	100,000+	Free	0	Everyone	Weather	February 16, 2014	6.0	2.1 and up
10833	Chemin (fr)	BOOKS_AND_REFERENCE	4.8	44	619k	1,000+	Free	0	Everyone	Books & Reference	March 23, 2014	0.8	2.2 and up
10834	FR Calculator	FAMILY	4.0	7	2.6M	500+	Free	0	Everyone	Education	June 18, 2017	1.0.0	4.1 and up
10835	FR Forms	BUSINESS	NaN	0	9.6M	10+	Free	0	Everyone	Business	September 29, 2016	1.1.5	4.0 and up
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

3. Show the summary of the Dataset

In [27]:
```
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
```

4. Show the categorical & numerical statistic

In [28]:
```
data.describe(include='all')
```

Out[28]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
count	10841	10841	9367.000000	10841	10841	10841	10840	10841	10840	10841	10841	10833	10838
unique	9660	34	NaN	6002	462	22	3	93	6	120	1378	2832	33
top	ROBLOX	FAMILY	NaN	0	Varies with device	1,000,000+	Free	0	Everyone	Tools	August 3, 2018	Varies with device	4.1 and up
freq	9	1972	NaN	596	1695	1579	10039	10040	8714	842	326	1459	2451
mean	NaN	NaN	4.193338	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	0.537431	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
min	NaN	NaN	1.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25%	NaN	NaN	4.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
50%	NaN	NaN	4.300000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	4.500000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
max	NaN	NaN	19.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5. Find the most downloaded App

In [29]:
```
data.columns
```

Out[29]:
```
Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
       'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
       'Android Ver'],
      dtype='object')
```

In [30]:
```
data['Installs'].dtype
```

Out[30]:
```
dtype('O')
```

In [31]:
```
data['Installs_1']=data['Installs'].str.replace(',','')
```

In [32]:
```
data['Installs_1']=data['Installs_1'].str.replace('+','')
```

In [33]:
```
data.drop(data.index[10472], inplace=True)
#This 10472 is an error row
```

In [34]:
```
data['Installs_1']=data['Installs_1'].astype('int')
```

In [35]:
```
index=data['Installs_1'].sort_values(ascending=False).head(5).index
data.iloc[index]
```

Out[35]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	Installs_1
3896	Subway Surfers	GAME	4.5	27711703	76M	1,000,000,000+	Free	0	Everyone 10+	Arcade	July 12, 2018	1.90.0	4.1 and up	1000000000
3943	Facebook	SOCIAL	4.1	78128208	Varies with device	1,000,000,000+	Free	0	Teen	Social	August 3, 2018	Varies with device	Varies with device	1000000000
335	Messenger – Text and Video Chat for Free	COMMUNICATION	4.0	56642847	Varies with device	1,000,000,000+	Free	0	Everyone	Communication	August 1, 2018	Varies with device	Varies with device	1000000000
3523	Google Drive	PRODUCTIVITY	4.4	2731211	Varies with device	1,000,000,000+	Free	0	Everyone	Productivity	August 6, 2018	Varies with device	Varies with device	1000000000
3565	Google Drive	PRODUCTIVITY	4.4	2731211	Varies with device	1,000,000,000+	Free	0	Everyone	Productivity	August 6, 2018	Varies with device	Varies with device	1000000000

6. Find the highest average rating App based on the Content Rating

In [36]:
```
data['Content Rating'].nunique()
```

Out[36]:
```
6
```

In [37]:
```
data.groupby('Content Rating')['Rating'].mean().sort_values(ascending=False)
```

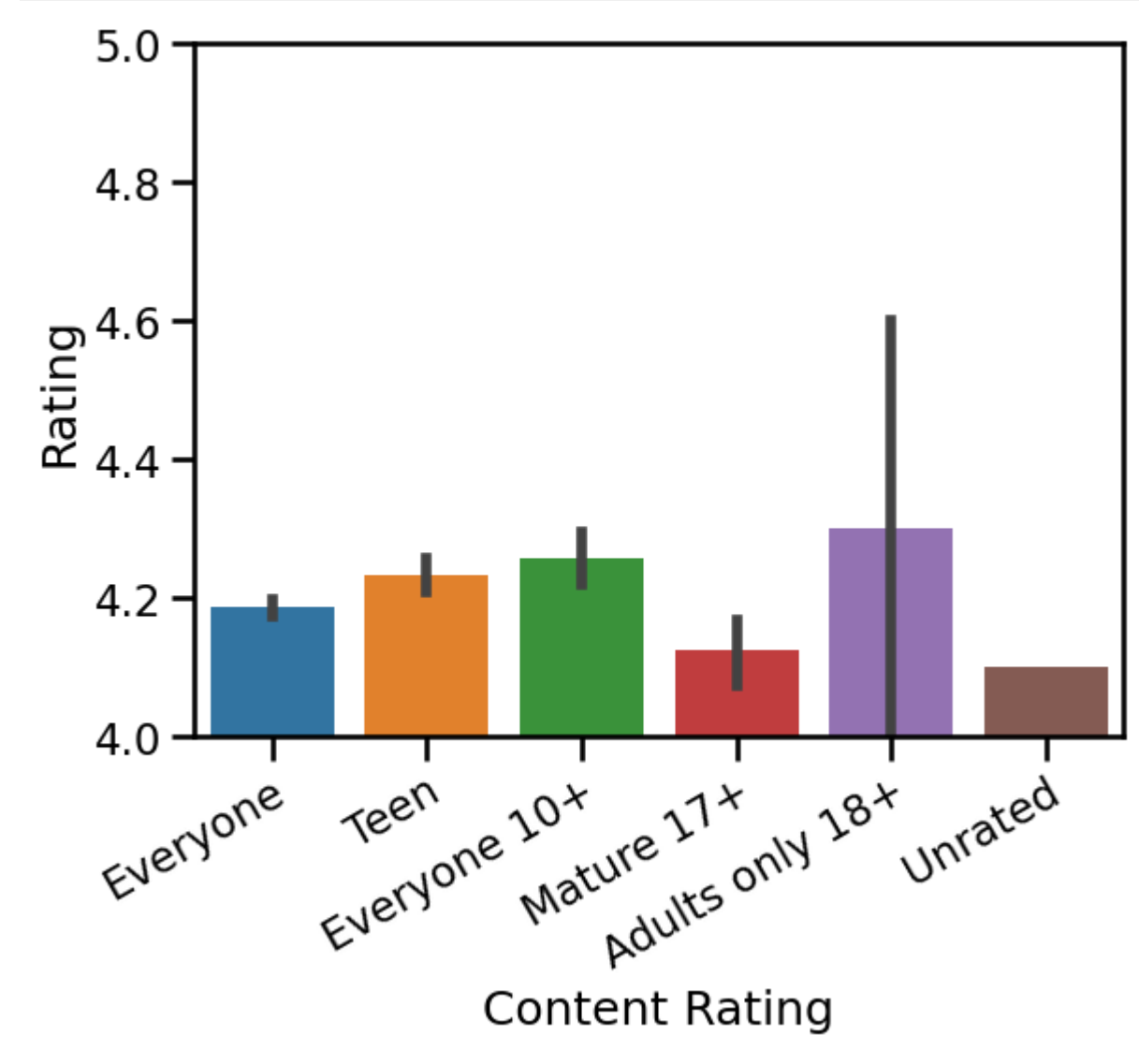
Out[37]:
```
Content Rating
Adults only 18+    4.300000
Everyone 10+       4.257179
Teen               4.233487
Everyone           4.186375
Mature 17+         4.123427
Unrated            4.100000
Name: Rating, dtype: float64
```

In [38]:
```
import seaborn as sns
```

In [39]:
```
import matplotlib.pyplot as plt
```

In [40]:
```
import warnings
warnings.filterwarnings('ignore')
```

In [41]:
```
fig, ax = plt.subplots()
x=data['Content Rating']
ax.set_xticklabels(x, rotation=30, ha="right", rotation_mode="anchor")
sns.barplot(x=data['Content Rating'],y=data['Rating'])
plt.ylim(4, 5)
sns.set_context("talk")
plt.show()
```



7. Find the highest average rating App based on the Category

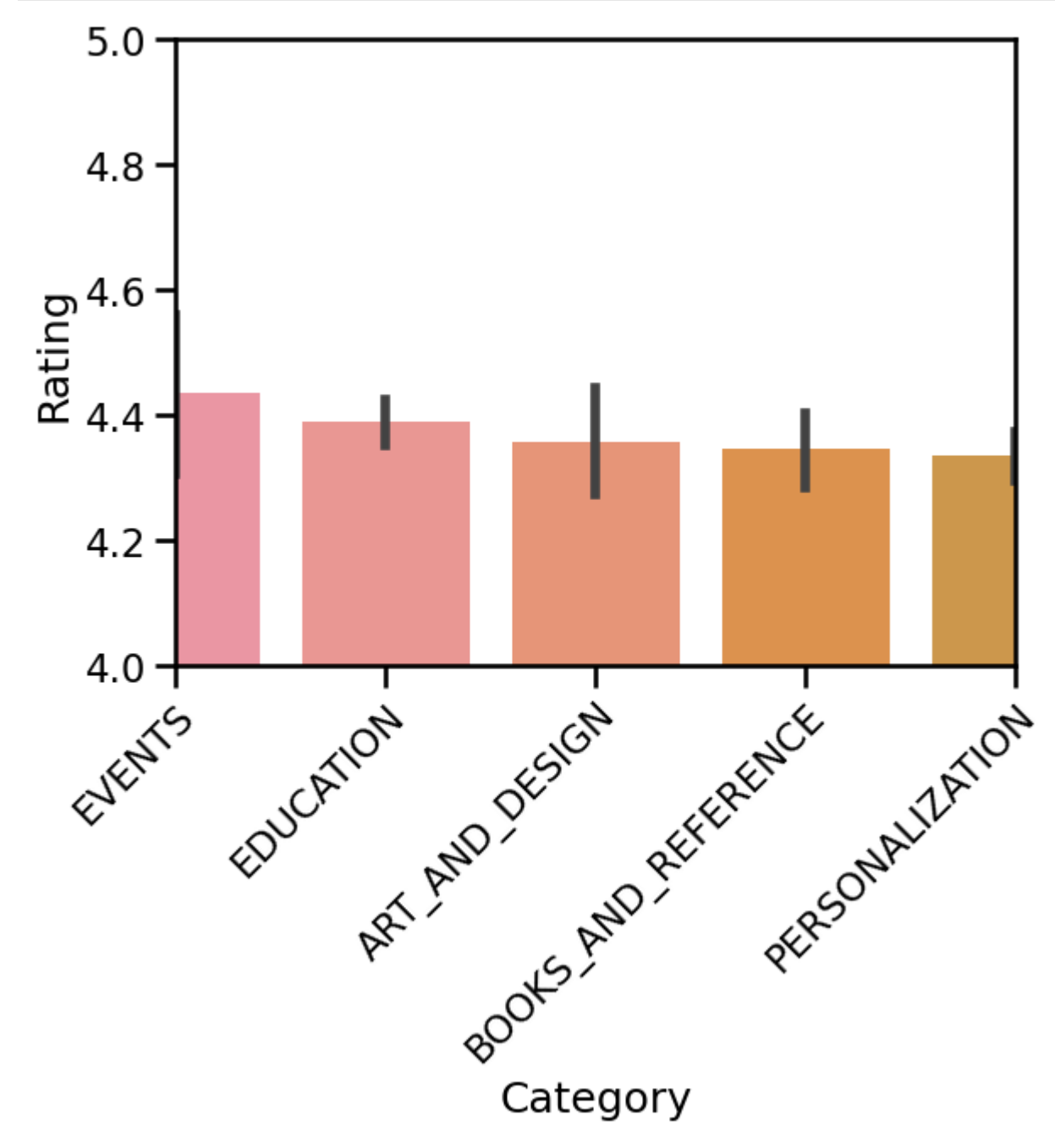
In [42]:
```
data.groupby('Category')['Rating'].mean().sort_values(ascending=False).head(5)
```

Out[42]:
```
Category
EVENTS                 4.435556
EDUCATION              4.389032
ART_AND_DESIGN         4.358065
BOOKS_AND_REFERENCE    4.346067
PERSONALIZATION        4.335987
Name: Rating, dtype: float64
```

In [43]:
```
data.groupby('Category')['Rating'].mean().sort_values(ascending=False).head(5)
```

Out[43]:
```
Category
EVENTS                 4.435556
EDUCATION              4.389032
ART_AND_DESIGN         4.358065
BOOKS_AND_REFERENCE    4.346067
PERSONALIZATION        4.335987
Name: Rating, dtype: float64
```

In [44]:
```
fig, ax = plt.subplots()
ax.set_xticklabels(x, rotation=30, ha="right", rotation_mode="anchor")
x=data['Content Rating']
plot_order = data.groupby('Category')['Rating'].mean().sort_values(ascending=False).index.values
sns.barplot(x=data['Category'],y=data['Rating'],order=plot_order)
plt.ylim(4, 5)
plt.xlim(0, 4)
plt.xticks(rotation=45)
sns.set_context("talk")
plt.show()
```



8. Find total number of App that has >4.9 rating

In [45]:
```
len(data[data['Rating']>=4.9])
```

Out[45]:
```
361
```

9. Find the average value of Reviews

In [46]:
```
data.columns
```

Out[46]:
```
Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
       'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
       'Android Ver', 'Installs_1'],
      dtype='object')
```

In [47]:
```
data['Reviews'].dtype
```

Out[47]:
```
dtype('O')
```

In [48]:
```
data['Reviews']
```

Out[48]:
```
0           159
1           967
2         87510
3        215644
4           967
          ...
10836        38
10837         4
10838         3
10839       114
10840    398307
Name: Reviews, Length: 10840, dtype: object
```

In [49]:
```
data['Reviews'].astype('float')
```

Out[49]:
```
0           159.0
1           967.0
2         87510.0
3        215644.0
4           967.0
           ...
10836        38.0
10837         4.0
10838         3.0
10839       114.0
10840    398307.0
Name: Reviews, Length: 10840, dtype: float64
```

In [50]:
```
data['Reviews']=data['Reviews'].astype('float')
```

In [51]:
```
data['Reviews'].dtype
```

Out[51]:
```
dtype('float64')
```

In [52]:
```
data['Reviews'].mean()
```

Out[52]:
```
444152.89603321033
```

10. Find the app that has most Reviews

In [53]:
```
data['Reviews'].max()
```

Out[53]:
```
78158306.0
```

In [54]:
```
data[data['Reviews'].max()==data['Reviews']]
```

Out[54]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	Installs_1
2544	Facebook	SOCIAL	4.1	78158306.0	Varies with device	1,000,000,000+	Free	0	Teen	Social	August 3, 2018	Varies with device	Varies with device	1000000000

11. Show the top 10 apps based on Reviews

In [55]:
```
index=data['Reviews'].sort_values(ascending=False).head(10).index
```

In [56]:
```
data.iloc[index]['App']
```

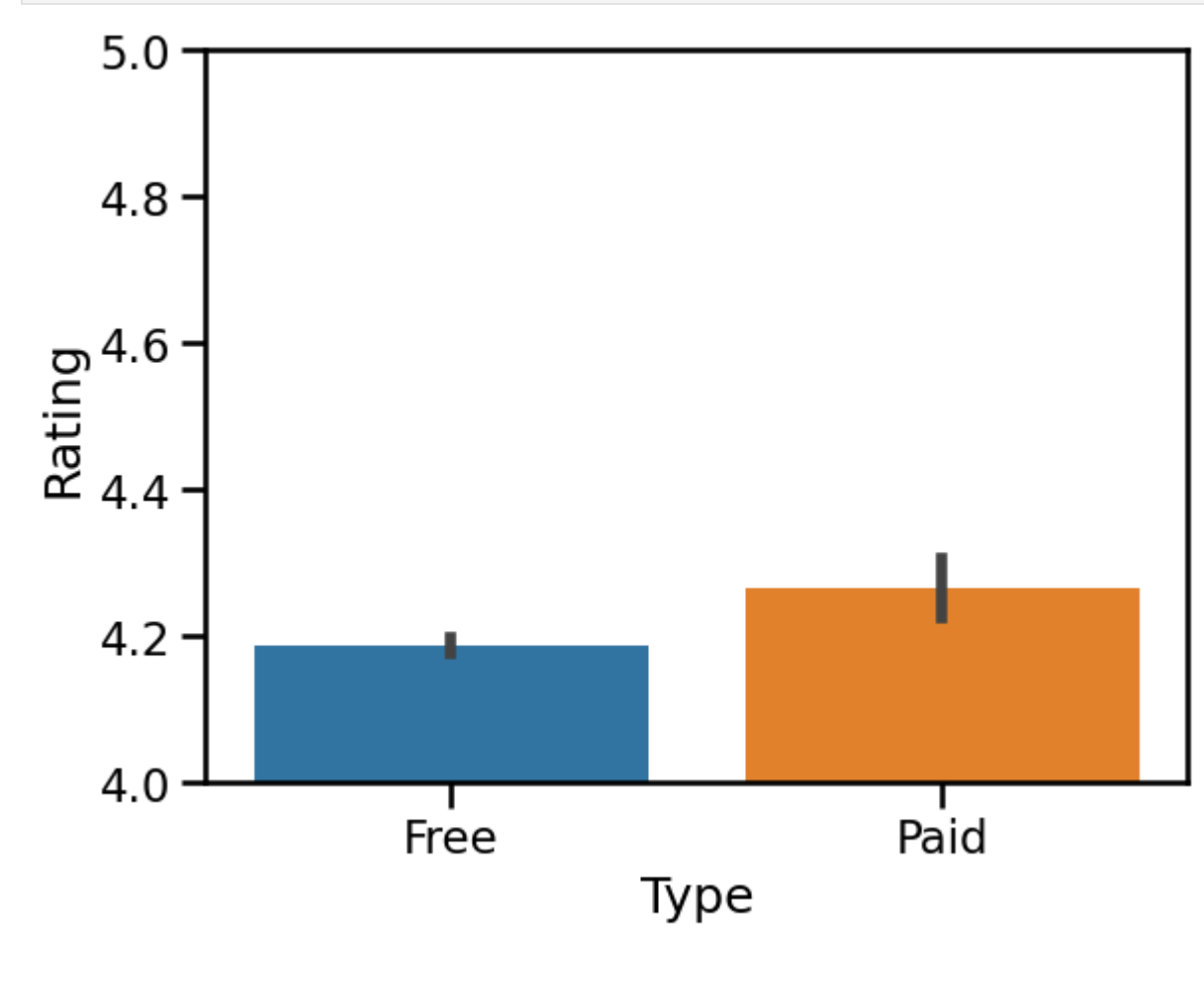
Out[56]:
```
2544                                    Facebook
3943                                    Facebook
381                           WhatsApp Messenger
336                           WhatsApp Messenger
3904                          WhatsApp Messenger
2604                                   Instagram
2611                                   Instagram
2545                                   Instagram
3909                                   Instagram
382     Messenger – Text and Video Chat for Free
Name: App, dtype: object
```

12. Find average rating of Free and Paid App

In [57]:
```
data.groupby('Type')['Rating'].mean()
```

Out[57]:
```
Type
Free    4.186203
Paid    4.266615
Name: Rating, dtype: float64
```

In [58]:
```
sns.barplot(x=data['Type'],y=data['Rating'])
sns.set_context("talk")
plt.ylim(4, 5)
plt.show()
```



Finish

In [ ]: