

```
import pandas as pd
import numpy as np
import re
```

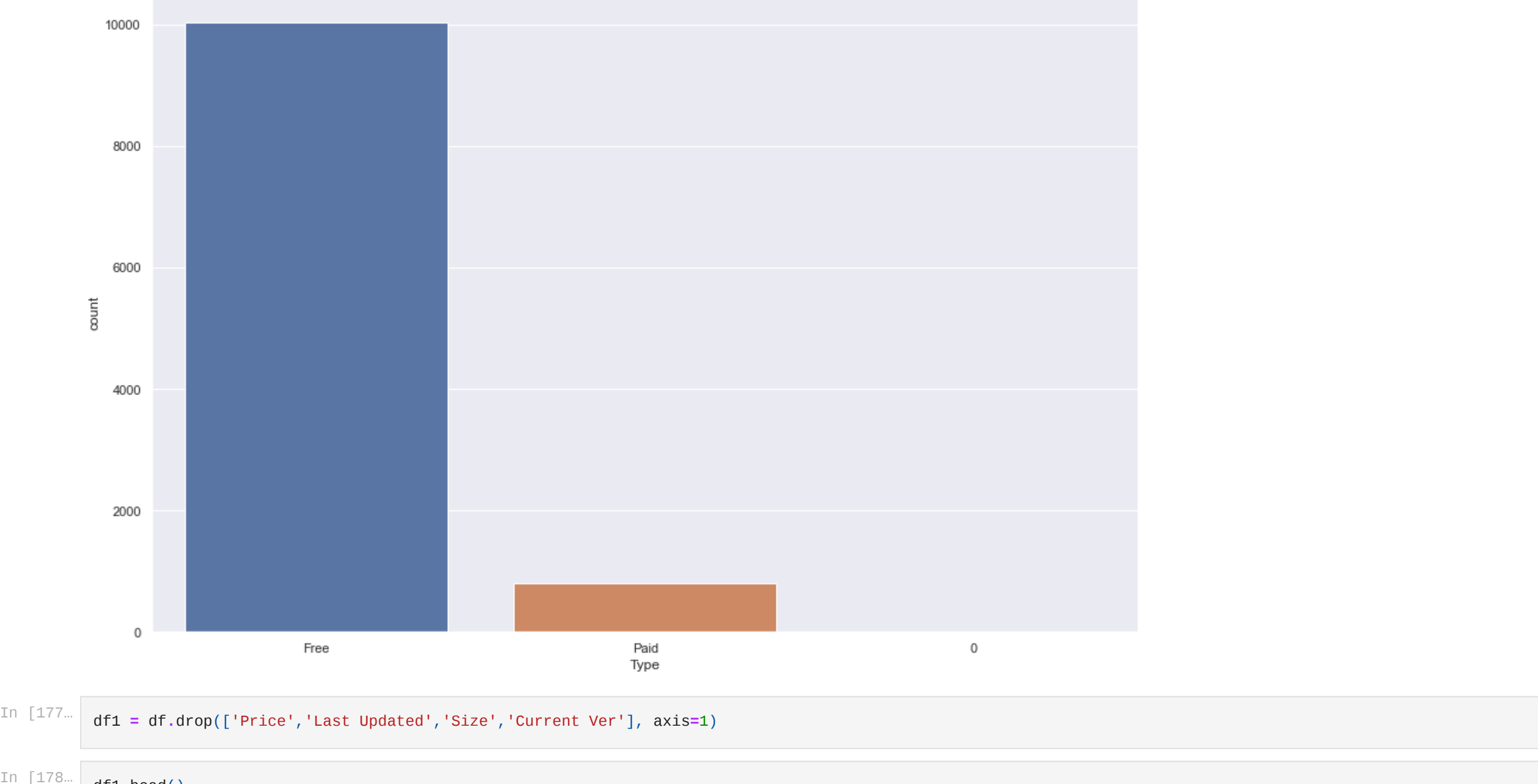
```
In [169]: import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
In [176]: df = pd.read_csv(r"C:\Users\BhwaJ Agrawal\Downloads\gstore.csv")
```

```
In [130]: df.head()
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	ContentRating	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	7-Jan-18	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	15-Jan-18	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	1-Aug-18	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	8-Jun-18	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	20-Jun-18	1.1	4.4 and up

```
In [119]: sns.countplot(x='Type',data=df)
```



```
In [177]: df1 = df.drop(['Price','Last Updated','Size','Current Ver'], axis=1)
```

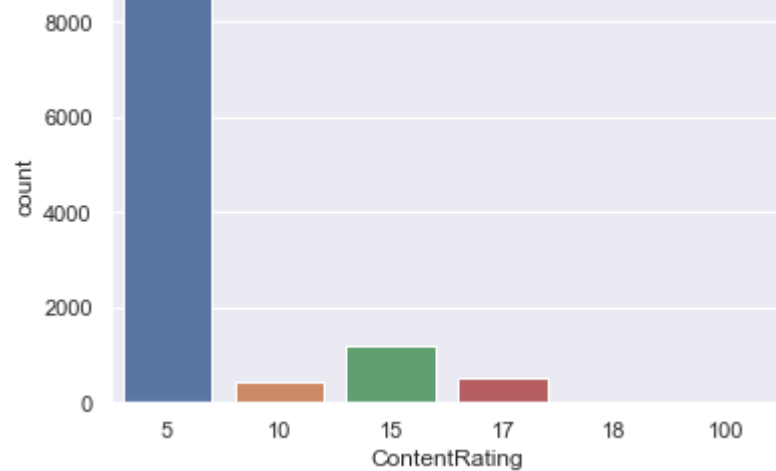
```
In [178]: df1.head()
```

	App	Category	Rating	Reviews	Installs	Type	ContentRating	Genres	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	10,000+	Free	Everyone	Art & Design	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	500,000+	Free	Everyone	Art & Design;Pretend Play	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	5,000,000+	Free	Everyone	Art & Design	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	50,000,000+	Free	Teen	Art & Design	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	100,000+	Free	Everyone	Art & Design;Creativity	4.4 and up

```
In [179]: df1.Type.replace({'Free', 'Paid'}, (0, 1), inplace=True)
```

```
In [180]: df1.ContentRating.replace({'Everyone', 'Everyone 10+', 'Teen', 'Mature 17+', 'Adults only 18+', 'Unrated'}, (5, 10,15,17,18,100), inplace=True)
```

```
In [182]: fig=sns.countplot(x='ContentRating',data=df1)
sns.set(rc = {'figure.figsize':(15,10)})
```



```
In [183]: df1.head()
```

	App	Category	Rating	Reviews	Installs	Type	ContentRating	Genres	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	10,000+	0.0	5	Art & Design	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	500,000+	0.0	5	Art & Design;Pretend Play	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	5,000,000+	0.0	5	Art & Design	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	50,000,000+	0.0	15	Art & Design	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	100,000+	0.0	5	Art & Design;Creativity	4.4 and up

```
In [184]: df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 9 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  int64
4    Installs    10841 non-null  object
5    Type        10840 non-null  float64
6    ContentRating 10841 non-null  int64
7    Genres      10841 non-null  object
8    Android Ver 10839 non-null  object
dtypes: float64(2), int64(2), object(5)
memory usage: 762.4+ KB

EXTRA

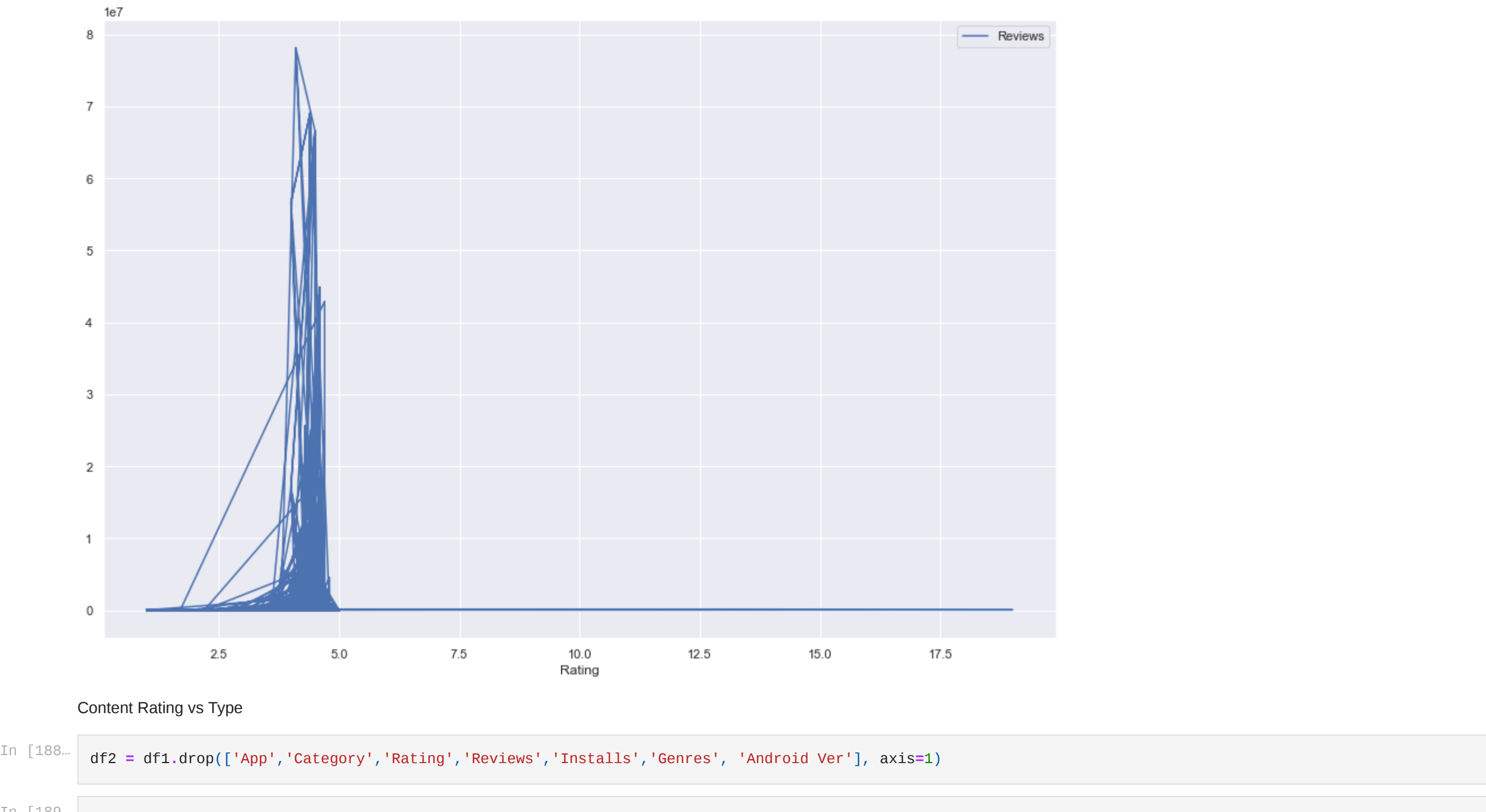
CONVERTING VARIOUS COLUMNS TO FLOAT
```

```
In [185]: df1["Reviews"] = pd.to_numeric(df1["Reviews"], downcast="float")
df1["ContentRating"] = pd.to_numeric(df1["ContentRating"], downcast="float")
df1["Type"] = pd.to_numeric(df1["Type"], downcast="float")
```

```
In [186]: df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 9 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  float32
4    Installs    10841 non-null  object
5    Type        10840 non-null  float32
6    ContentRating 10841 non-null  float32
7    Genres      10841 non-null  object
8    Android Ver 10839 non-null  object
dtypes: float32(3), float64(1), object(5)
memory usage: 635.3+ KB
```

```
In [187]: df1.plot(x='Rating', y='Reviews')
```



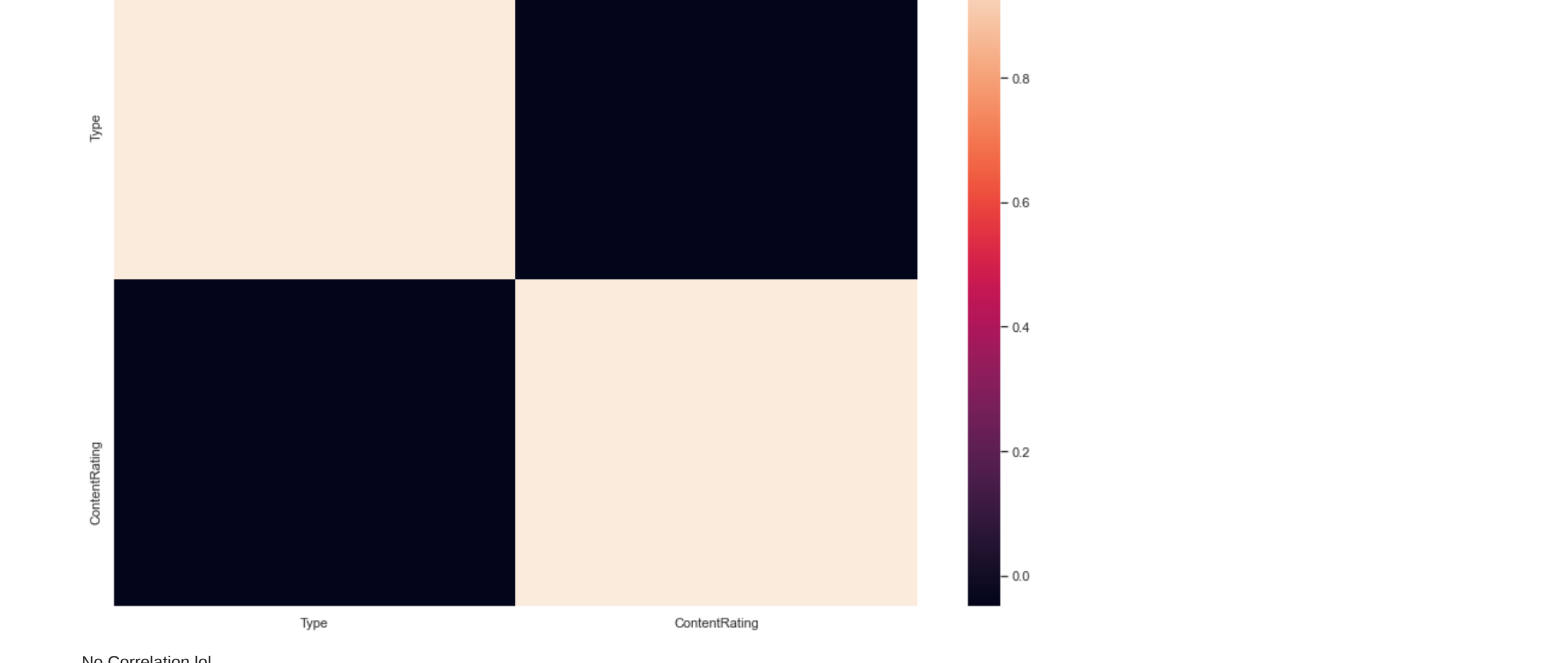
Content Rating vs Type

```
In [188]: df2 = df1.drop(['App','Category','Rating','Reviews','Installs','Genres', 'Android Ver'], axis=1)
```

```
In [189]: df2.head()
```

	Type	ContentRating
0	0.0	5.0
1	0.0	5.0
2	0.0	5.0
3	0.0	15.0
4	0.0	5.0

```
In [190]: sns.heatmap(df2.corr())
```



No Correlation lol

Removing + and , from Installs

```
In [191]: x = df1['Installs'].iloc[0]
x
```

```
Out[191]: '10,000+'
x
```

```
In [192]: x.split('+')
```

```
Out[192]: ['10,000', '']
```

```
In [193]: df1['Installs'] = df1['Installs'].apply(lambda Installs : Installs.split('+')[0])
```

```
In [194]: df1.head()
```

	App	Category	Rating	Reviews	Installs	Type	ContentRating	Genres	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159.0	10,000	0.0	5.0	Art & Design	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967.0	500,000	0.0	5.0	Art & Design;Pretend Play	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510.0	5,000,000	0.0	5.0	Art & Design	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644.0	50,000,000	0.0	15.0	Art & Design	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967.0	100,000	0.0	5.0	Art & Design;Creativity	4.4 and up

```
In [195]: y = df1['Installs'].iloc[0]
y
```

```
Out[195]: '10,000'
```

```
In [196]: y=re.sub(",","",y)
y
```

```
Out[196]: '10000'
```

```
In [197]: df1['Installs'] = df1['Installs'].apply(lambda Installs : re.sub(",","",Installs))
```

```
In [198]: df1.head()
```

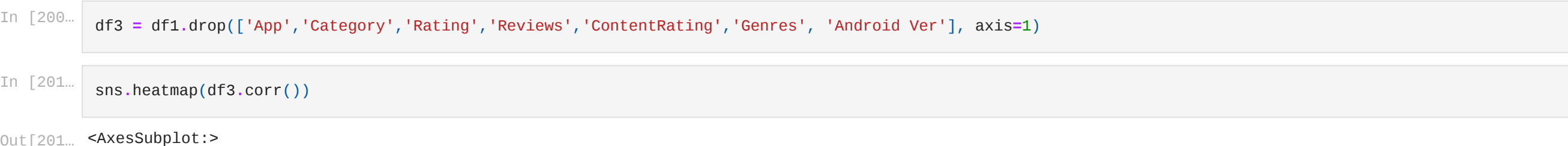
	App	Category	Rating	Reviews	Installs	Type	ContentRating	Genres	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159.0	10000	0.0	5.0	Art & Design	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967.0	500000	0.0	5.0	Art & Design;Pretend Play	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510.0	5000000	0.0	5.0	Art & Design	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644.0	50000000	0.0	15.0	Art & Design	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967.0	100000	0.0	5.0	Art & Design;Creativity	4.4 and up

```
In [199]: df1["Installs"] = pd.to_numeric(df1["Installs"], downcast="float")
```

Correlation between Installs and Type

```
In [200]: df3 = df1.drop(['App','Category','Rating','Reviews','ContentRating','Genres', 'Android Ver'], axis=1)
```

```
In [201]: sns.heatmap(df3.corr())
```



No correlation again

INSTALLS VS CATEGORY

```
In [212]: df4 = df1.drop(['App','Reviews','Installs','Category','Rating','Type', 'Android Ver'], axis=1)
```

```
In [213]: df4.head()
```

	ContentRating	Genres
0	5.0	Art & Design
1	5.0	Art & Design;Pretend Play
2	5.0	Art & Design
3	15.0	Art & Design
4	5.0	Art & Design;Creativity

```
In [214]: sns.heatmap(df4.corr())
```



```
In [ ]:
```