Flight Price:

Q1. Load the flight price dataset and examine its dimensions. How many rows and columns does the dataset have?

In [46]:

```
import pandas as pd
df=pd.read_excel('flight_price.xlsx')
df.head()
```

Out[46]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additio
0	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m	non-stop	
1	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m	2 stops	
2	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h	2 stops	
3	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m	1 stop	
4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	

In [2]:

```
df.shape
```

Out[2]:

```
(10683, 11)
```

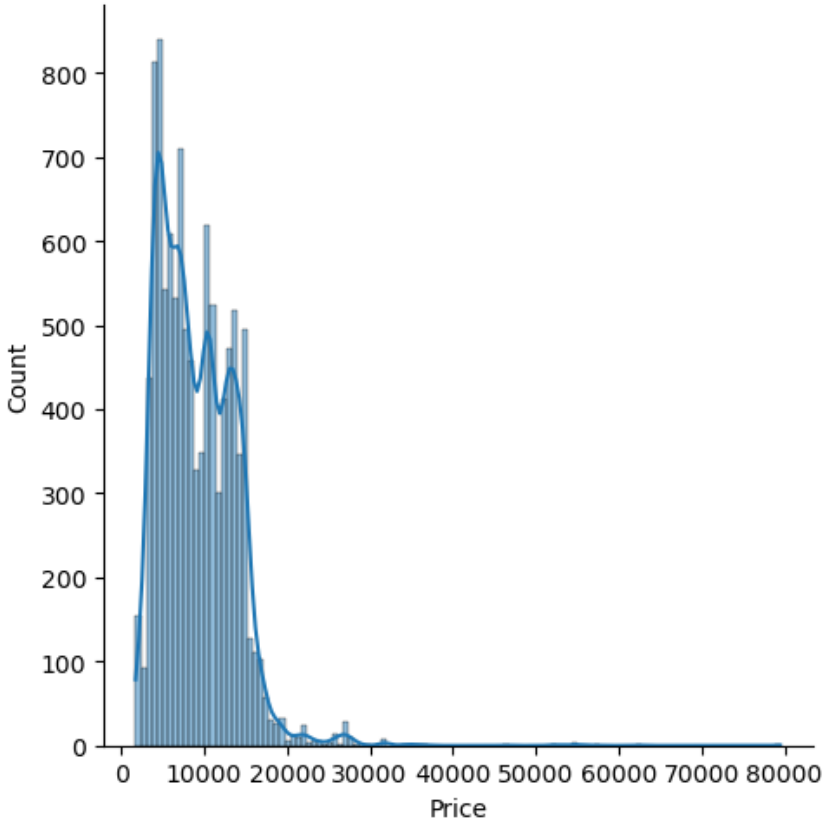
Q2. What is the distribution of flight prices in the dataset? Create a histogram to visualize the distribution.

In [3]:

```
import seaborn as sns
sns.displot(df['Price'],kde = True)
```

Out[3]:

```
<seaborn.axisgrid.FacetGrid at 0x28e80d6e160>
```



The distribution of Price is Right Skewed

Q3. What is the range of prices in the dataset? What is the minimum and maximum price?

In [4]:

```
Max =df.Price.max()
print('max price is ',Max)
Min = df.Price.min()
print('min price is ',Min)
```

```
max price is  79512
min price is  1759
```

Q4. How does the price of flights vary by airline? Create a boxplot to compare the prices of different airlines.

In [47]:

```
df['Airline'].unique()
```

Out[47]:

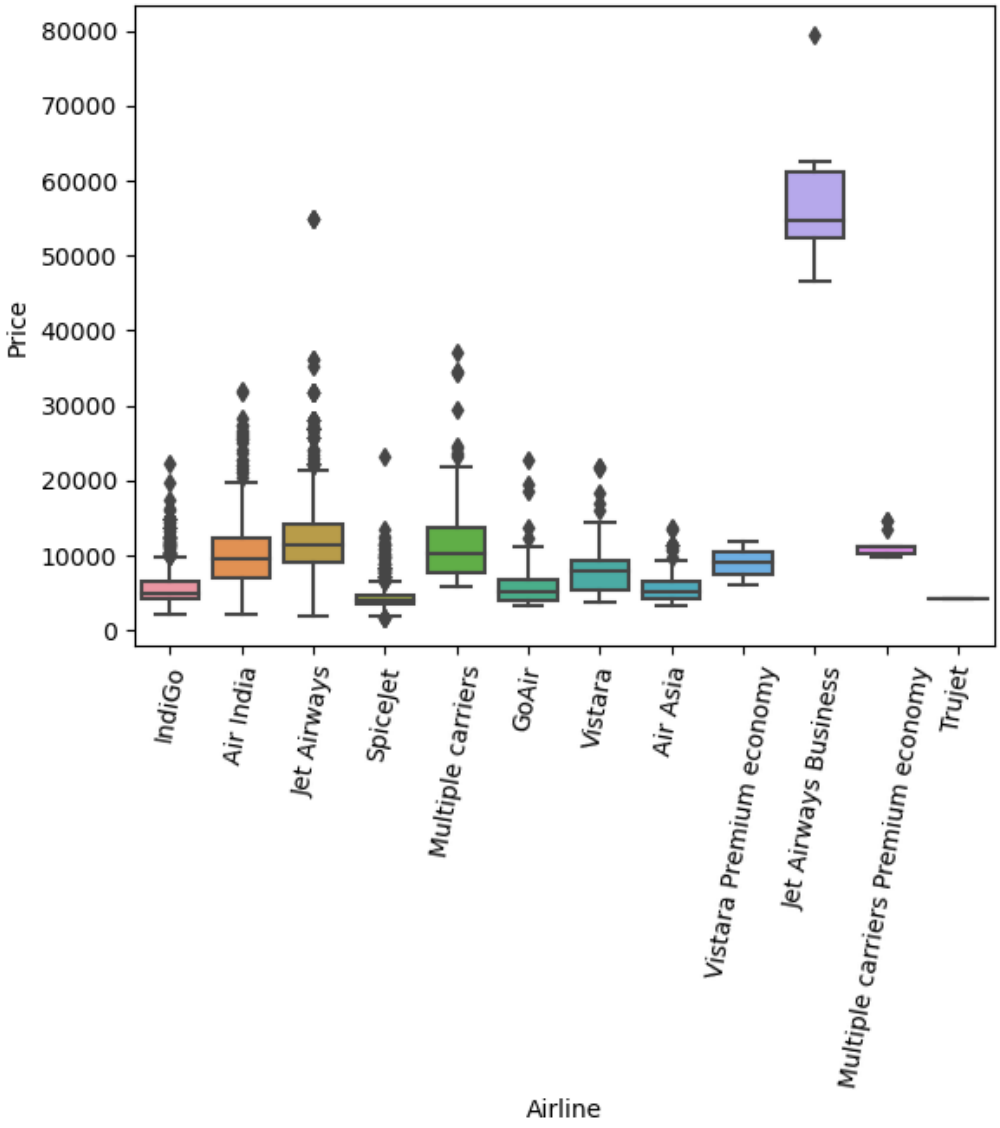
```
array(['IndiGo', 'Air India', 'Jet Airways', 'SpiceJet',
       'Multiple carriers', 'GoAir', 'Vistara', 'Air Asia',
       'Vistara Premium economy', 'Jet Airways Business',
       'Multiple carriers Premium economy', 'Trujet'], dtype=object)
```

In [48]:

```
from matplotlib import pyplot as plt
sns.boxplot(x='Airline', y='Price', data=df)
plt.xticks(rotation = 80)
```

Out[48]:

```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11]),
 [Text(0, 0, 'IndiGo'),
  Text(1, 0, 'Air India'),
  Text(2, 0, 'Jet Airways'),
  Text(3, 0, 'SpiceJet'),
  Text(4, 0, 'Multiple carriers'),
  Text(5, 0, 'GoAir'),
  Text(6, 0, 'Vistara'),
  Text(7, 0, 'Air Asia'),
  Text(8, 0, 'Vistara Premium economy'),
  Text(9, 0, 'Jet Airways Business'),
  Text(10, 0, 'Multiple carriers Premium economy'),
  Text(11, 0, 'Trujet')])
```



Q5. Are there any outliers in the dataset? Identify any potential outliers using a boxplot and describe how they may impact your analysis.

In [7]:

```
plt.boxplot(x = df['Price'])
```

Out[7]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x28e81829520>,
  <matplotlib.lines.Line2D at 0x28e818296a0>],
 'caps': [<matplotlib.lines.Line2D at 0x28e81829910>,
  <matplotlib.lines.Line2D at 0x28e81829c40>],
 'boxes': [<matplotlib.lines.Line2D at 0x28e8180e4f0>],
 'medians': [<matplotlib.lines.Line2D at 0x28e81919730>],
 'fliers': [<matplotlib.lines.Line2D at 0x28e81937250>],
 'means': []}
```

Q6. You are working for a travel agency, and your boss has asked you to analyze the Flight Price dataset to identify the peak travel season. What features would

In [10]:

```
import warnings

warnings.filterwarnings("ignore")

df['Date_of_Journey']=pd.to_datetime(df['Date_of_Journey'])
df['month'] = df['Date_of_Journey'].dt.month
```

In [11]:

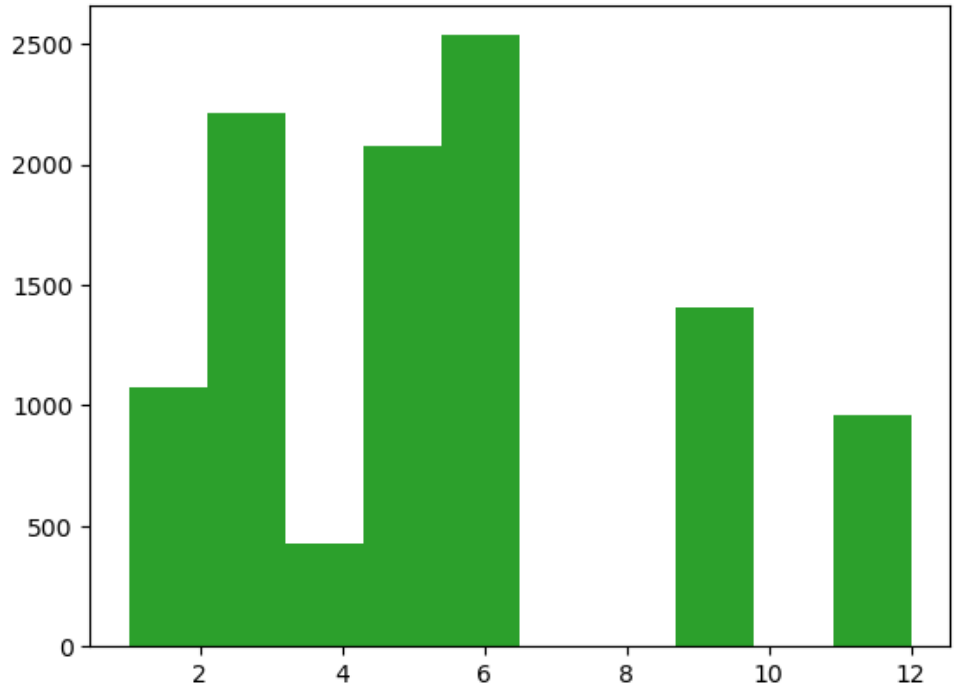
```
#add month colomn
df.head()
```

Out[11]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additio
0	IndiGo	2019-03-24	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m	non-stop	
1	Air India	2019-01-05	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m	2 stops	
2	Jet Airways	2019-09-06	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h	2 stops	
3	IndiGo	2019-12-05	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m	1 stop	
4	IndiGo	2019-01-03	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	

In [36]:

```
j = list(df['month'])
plt.hist(j)
plt.show()
```

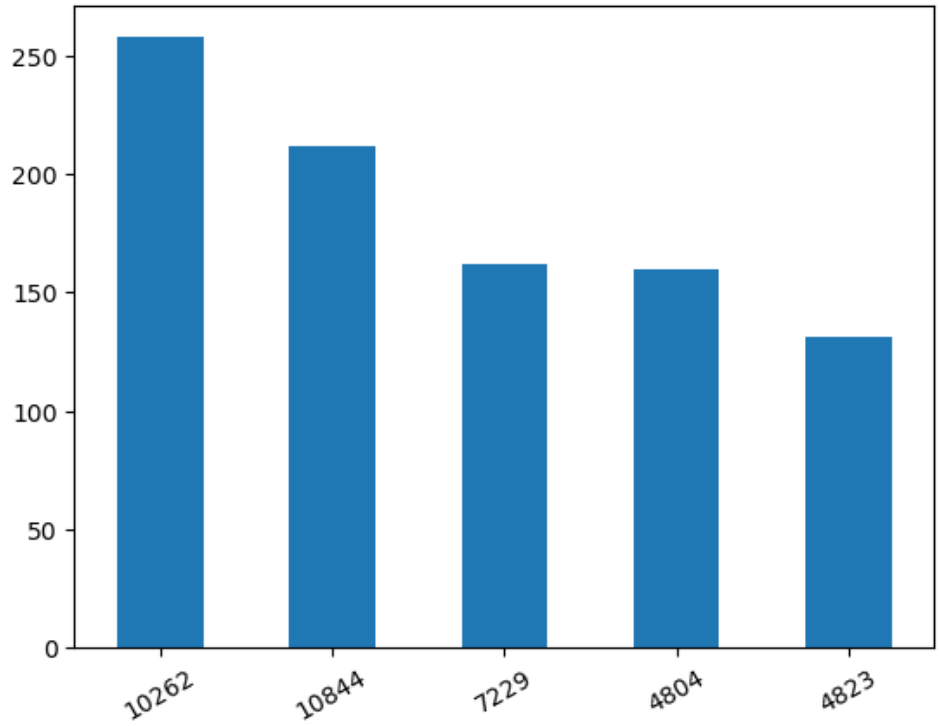


the peak travel season month is 4 ,5, 6 .

Q7. You are a data analyst for a flight booking website, and you have been asked to analyze the Flight Price dataset to identify any trends in flight prices. What features would you analyze to identify these trends, and what visualizations would you use to present your findings to your team?

In [37]:

```
df['Price'].value_counts().head().plot(kind = 'bar')
plt.xticks(rotation = 30)
plt.show()
```



I will use barplot to see the trend of ticket price. And from this I can analyza that Most people Booked a flight when ticket price is 10262 or less

Q8. You are a data scientist working for an airline company, and you have been asked to analyze the Flight Price dataset to identify the factors that affect flight prices. What features would you analyze to identify these factors, and how would you present your findings to the management team?

In [38]:

```
df.cov()
```

Out[38]:

	Price	month
Price	2.126463e+07	147.410006
month	1.474100e+02	8.925092

I will analyze either the the current season is on peak or not, and for that i will present barplot month in which maximum ticket are booked.

Flight Price dataset the month factors most affect flight prices.

Google Playstore:

Q9. Load the Google Playstore dataset and examine its dimensions. How many rows and columns does the dataset have?

In [52]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings

warnings.filterwarnings("ignore")

%matplotlib inline
df=pd.read_csv('https://raw.githubusercontent.com/krishnaik06/playstore-Dataset/main/googleplaystore.csv'
df.head()
```

Out[52]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Upd
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	Jan 7,
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	Jan 15,
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	Au 1,
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	Ju
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	Jun

In [53]:

```
df.shape
```

Out[53]:

```
(10841, 13)
```

In this dataset have 10841 row ,13 column

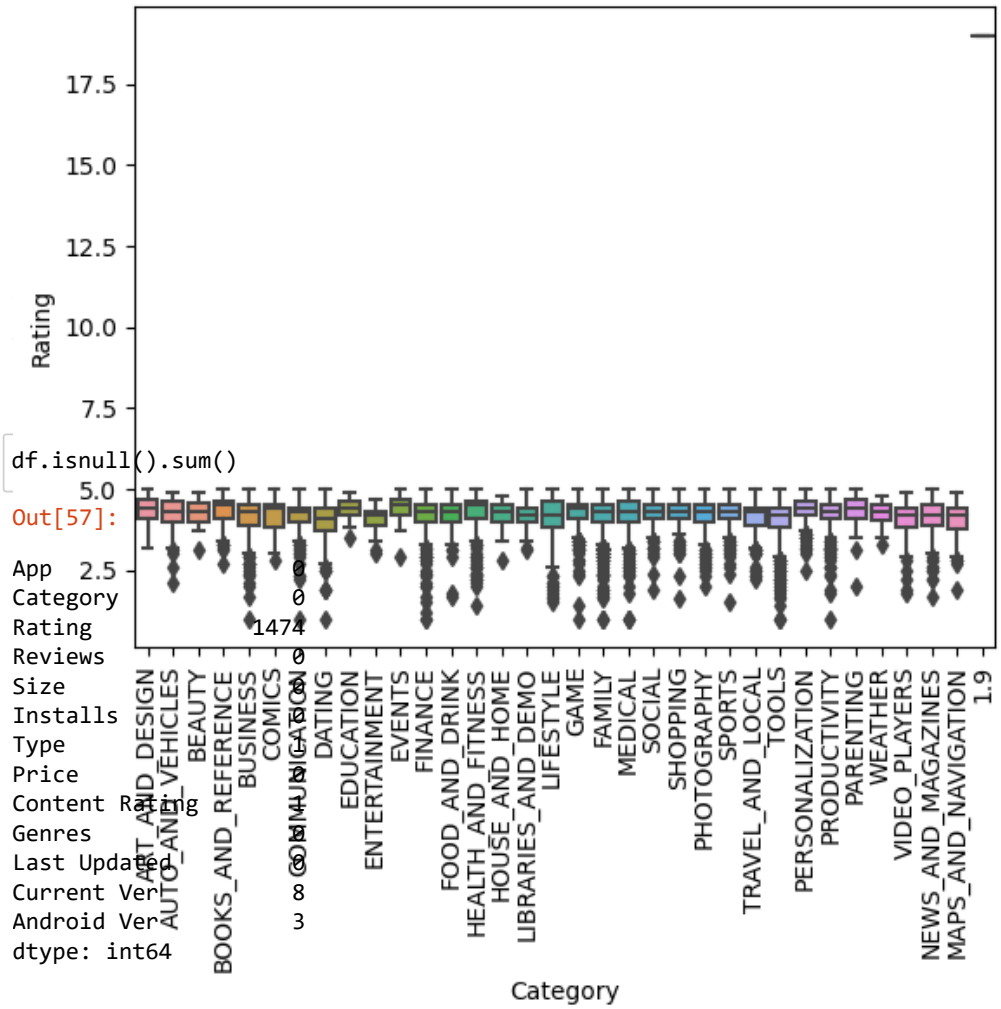
Q10. How does the rating of apps vary by category? Create a boxplot to compare the ratings of different app categories.

In [56]:

```
from matplotlib import pyplot as plt
sns.boxplot(x='Category', y='Rating', data=df)
plt.xticks(rotation = 90)
```

Out[56]:

```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
        17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33]),
 [Text(0, 0, 'ART_AND_DESIGN'),
  Text(1, 0, 'AUTO_AND_VEHICLES'),
  Text(2, 0, 'BEAUTY'),
  Text(3, 0, 'BOOKS_AND_REFERENCE'),
  Text(4, 0, 'BUSINESS'),
  Text(5, 0, 'COMICS'),
  Text(6, 0, 'COMMUNICATION'),
  Text(7, 0, 'DATING'),
  Text(8, 0, 'EDUCATION'),
  Text(9, 0, 'ENTERTAINMENT'),
  Text(10, 0, 'EVENTS'),
  Text(11, 0, 'FINANCE'),
  Text(12, 0, 'FOOD_AND_DRINK'),
  Text(13, 0, 'HEALTH_AND_FITNESS'),
  Text(14, 0, 'HOUSE_AND_HOME'),
  Text(15, 0, 'LIBRARIES_AND_DEMO'),
  Text(16, 0, 'LIFESTYLE'),
  Text(17, 0, 'GAME'),
  Text(18, 0, 'FAMILY'),
  Text(19, 0, 'MEDICAL'),
  Text(20, 0, 'SOCIAL'),
  Text(21, 0, 'SHOPPING'),
  Text(22, 0, 'PHOTOGRAPHY'),
  Text(23, 0, 'SPORTS'),
  Text(24, 0, 'TRAVEL_AND_LOCAL'),
  Text(25, 0, 'TOOLS'),
  Text(26, 0, 'PERSONALIZATION'),
  Text(27, 0, 'PRODUCTIVITY'),
  Text(28, 0, 'PARENTING'),
  Text(29, 0, 'WEATHER'),
  Text(30, 0, 'VIDEO_PLAYERS'),
  Text(31, 0, 'NEWS_AND_MAGAZINES'),
  Text(32, 0, 'MAPS_AND_NAVIGATION'),
  Text(33, 0, '1.9')])
```



:ribe how they may impact your

```
df.isnull().sum()
```

Out[57]:

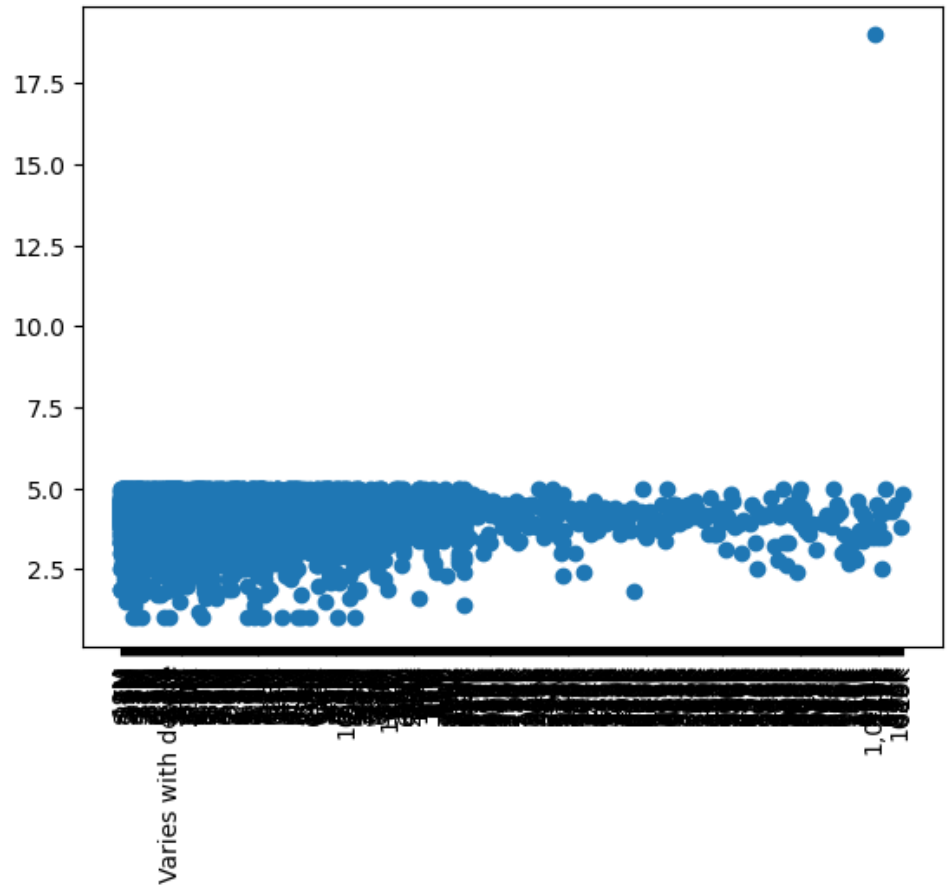
```
App                0
Category           0
Rating          1474
Reviews            0
Size               0
Installs           0
Type               1
Price              0
Content Rating     1
Genres             0
Last Updated       0
Current Ver        8
Android Ver        3
dtype: int64
```

there are rating ,type ,content rating ,Current Ver ,Android Ver feature missing values .

Q12. What is the relationship between the size of an app and its rating? Create a scatter plot to visualize the relationship.

In [64]:

```
k = list(df['Size'])
h = list(df['Rating'])
plt.scatter(x = k ,y = h)
plt.xticks(rotation = 90)
plt.show()
```



Q13. How does the type of app affect its price? Create a bar chart to compare average prices by app type.

In [ ]:

```
new_df= pd.DataFrame(df.groupby('Category')['Price'].mean().sort_values(ascending = False).reset_index())
sns.barplot(data = new_df,x = 'Category',y = 'Price')
plt.xticks(rotation = 90)
```

Q14. What are the top 10 most popular apps in the dataset? Create a frequency table to identify the apps with the highest number of installs.

In [ ]:

```
df = df.drop_duplicates(subset = ['App'],keep = 'first')

new = df[['App','Installs']].sort_values(by='Installs',ascending = False).head()
sns.barplot(data = new,x = 'App',y = 'Installs')
plt.xticks(rotation = 50)
```

Q15. A company wants to launch a new app on the Google Playstore and has asked you to analyze the Google Playstore dataset to identify the most popular app categories. How would you approach this task, and what features would you analyze to make recommendations to the company?

In [ ]:

```
new = pd.DataFrame(df.groupby('Category')['Installs'].sum().sort_values(ascending = False).reset_index())
sns.barplot(data = new,x = 'Category',y = 'Installs')
plt.xticks(rotation = 90)
```

Q16. A mobile app development company wants to analyze the Google Playstore dataset to identify the most successful app developers. What features would you analyze to make recommendations to the company, and what data visualizations would you use to present your findings?

In [74]:

```
df = df.drop_duplicates(subset = ['App'],keep = 'first')
new = df.sort_values(by = ['Rating','Installs'],ascending= False).head()
```

In [75]:

```
df.head()
```

Out[75]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Upd
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	Jan 7,
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	Jan 15,
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	Au 1,
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	Ju
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	Jun

Q17. A marketing research firm wants to analyze the Google Playstore dataset to identify the best time to launch a new app. What features would you analyze to make recommendations to the company, and what data visualizations would you use to present your findings?

To identify the best time to launch a new app in the Google Playstore, I would recommend analyzing the following features:

App release date: The release date of apps can provide insights into the best time to launch a new app based on historical trends.

Total number of downloads: The total number of downloads of apps can provide insights into the best time to launch a new app based on seasonal or other trends.

Average rating of apps: The average rating of apps can provide insights into the best time to launch a new app based on the competition in the market.

Category of apps: The category of apps can provide insights into the best time to launch a new app based on user behavior and preferences.

In [ ]:

```

```