APP Rating Prediction

Q1) Load the data file using pandas.

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

df=pd.read_csv("C:/Users/shiva/Desktop/Shivani/SimpliLearn/Python/
googleplay/googleplaystore.csv")

df # Data Frame of App Rating

Content Rating

```
App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART_AND_DESIGN
                                       Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                     Sketch - Draw & Paint
ART AND DESIGN
                    Pixel Draw - Number Art Coloring Book
ART AND DESIGN
. . .
. . .
10836
                                           Sya9a Maroc - FR
FAMILY
10837
                         Fr. Mike Schmitz Audio Teachings
FAMILY
                                    Parkinson Exercices FR
10838
MEDICAL
                             The SCP Foundation DB fr nn5n
10839
BOOKS AND REFERENCE
           iHoroscope - 2018 Daily Horoscope & Astrology
10840
LIFESTYLE
       Rating Reviews
                                       Size
                                                 Installs
                                                            Type Price
0
          4.1
                                         19M
                                                  10,000+
                                                            Free
                   159
                                                                      0
          3.9
1
                                                 500,000+
                                                            Free
                   967
                                         14M
                                                                      0
2
          4.7
                 87510
                                       8.7M
                                               5,000,000+
                                                            Free
                                                                      0
3
          4.5
                                         25M
                                              50,000,000+
                215644
                                                            Free
                                                                      0
4
          4.3
                   967
                                        2.8M
                                                 100,000+
                                                            Free
                                                                      0
                                                             . . .
          4.5
10836
                    38
                                         53M
                                                   5,000+
                                                            Free
                                                                      0
10837
          5.0
                     4
                                        3.6M
                                                      100 +
                                                            Free
                                                                      0
                     3
10838
                                       9.5M
                                                   1,000+
                                                                      0
          NaN
                                                            Free
10839
          4.5
                   114
                        Varies with device
                                                   1,000+
                                                            Free
                                                                      0
10840
          4.5
                                              10,000,000+
                398307
                                         19M
                                                            Free
                                                                      0
```

Last Updated \

Genres

```
0
            Everyone
                                     Art & Design
                                                     January 7, 2018
1
            Everyone Art & Design; Pretend Play
                                                    January 15, 2018
2
            Everyone
                                     Art & Design
                                                      August 1, 2018
3
                                                        June 8, 2018
                 Teen
                                     Art & Design
4
            Everyone
                         Art & Design;Creativity
                                                       June 20, 2018
                  . . .
. . .
                                        Education
                                                       July 25, 2017
10836
            Evervone
10837
            Everyone
                                        Education
                                                        July 6, 2018
10838
            Everyone
                                          Medical
                                                    January 20, 2017
10839
          Mature 17+
                                Books & Reference
                                                    January 19, 2015
10840
                                        Lifestyle
                                                       July 25, 2018
            Everyone
              Current Ver
                                    Android Ver
0
                     1.0.0
                                   4.0.3 and up
                                   4.0.3 and up
1
                     2.0.0
2
                     1.2.4
                                   4.0.3 and up
3
       Varies with device
                                     4.2 and up
4
                       1.1
                                     4.4 and up
                       . . .
. . .
10836
                      1.48
                                     4.1 and up
                                     4.1 and up
10837
                         1
                         1
10838
                                     2.2 and up
10839
       Varies with device Varies with device
10840
       Varies with device Varies with device
[10841 rows x 13 columns]
Q2) Check for null values in the data. Get the number of null values for each column.
df.describe() # Gives the statistical summary of the data frame
            Rating
       9367.000000
count
          4.193338
mean
std
          0.537431
          1.000000
min
25%
          4.000000
50%
          4.300000
75%
          4.500000
         19.000000
max
df.info() # Provides the concise summary of the data frame
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
```

Non-Null Count Dtype

object

object

float64

10841 non-null

10841 non-null

9367 non-null

Data columns (total 13 columns):

#

- - -

0

1

2

Column

_ _ _ _ _ _

Rating

Category

App

3	Reviews	10841 non-nul	l object
4	Size	10841 non-nul	l object
5	Installs	10841 non-nul	l object
6	Туре	10840 non-nul	l object
7	Price	10841 non-nul	l object
8	Content Rating	10840 non-nul	l object
9	Genres	10841 non-nul	l object
10	Last Updated	10841 non-nul	l object
11	Current Ver	10833 non-nul	l object
12	Android Ver	10838 non-nul	l object
4+,,,,,	oc. floo+64/1\	ab = a a + /12\	

dtypes: float64(1), object(12)
memory usage: 1.1+ MB

df.isnull() # Isnull function gives the boolean value of each column

u	5		director.	92100 0110	20000	racae e.	cacii c	o camir
`	Арр	Category	Rating	Reviews	Size	Installs	Туре	Price
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
1083	6 False	False	False	False	False	False	False	False
1083	7 False	False	False	False	False	False	False	False
10838	8 False	False	True	False	False	False	False	False
10839	9 False	False	False	False	False	False	False	False
10840	0 False	False	False	False	False	False	False	False
	Conten	t Rating	Genres	Last Upda	ted Cu	rrent Ver	Androi	d Ver
0		False	False	Fa	lse	False		False
1		False	False	Fa	lse	False		False
2		False	False	Fa	lse	False		False
3		False	False	Fa	lse	False		False

4	False	False	False	False	False
10836	False	False	False	False	False
10837	False	False	False	False	False
10838	False	False	False	False	False
10839	False	False	False	False	False
10840	False	False	False	False	False

[10841 rows x 13 columns]

df.isnull().sum() # Gives the sum of null values column wise

Here Rating column has 1474 Null Values, Type column has 1 null value, Content Rating has 1 null value, # Current Version has 8 null values and Android Version has 3 Null Values

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	

df.isnull().sum().sum() # Gives the total count of NULL Values

#Q3) Drop records with nulls in any of the columns.

df # App Rating Data Frame

```
App
Category \
          Photo Editor & Candy Camera & Grid & ScrapBook
ART AND DESIGN
                                       Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                     Sketch - Draw & Paint
ART AND DESIGN
                    Pixel Draw - Number Art Coloring Book
ART_AND_DESIGN
. . .
                                                        . . .
10836
                                          Sya9a Maroc - FR
FAMILY
10837
                         Fr. Mike Schmitz Audio Teachings
FAMILY
                                    Parkinson Exercices FR
10838
MEDICAL
                            The SCP Foundation DB fr nn5n
10839
BOOKS AND REFERENCE
10840
           iHoroscope - 2018 Daily Horoscope & Astrology
LIFESTYLE
       Rating Reviews
                                       Size
                                                 Installs
                                                           Type Price
0
          4.1
                   159
                                        19M
                                                  10,000+
                                                           Free
                                                                     0
          3.9
1
                                        14M
                                                 500,000+
                                                                     0
                   967
                                                           Free
2
          4.7
                                       8.7M
                                               5,000,000+
                87510
                                                           Free
                                                                     0
3
          4.5
               215644
                                        25M
                                             50,000,000+
                                                                     0
                                                           Free
4
          4.3
                   967
                                       2.8M
                                                 100,000+
                                                           Free
                                                                     0
          4.5
                    38
                                        53M
                                                   5,000+
                                                                     0
10836
                                                           Free
10837
          5.0
                     4
                                       3.6M
                                                     100+
                                                           Free
                                                                     0
                     3
                                                                     0
10838
          NaN
                                       9.5M
                                                   1,000+
                                                           Free
          4.5
                                                                     0
10839
                   114
                        Varies with device
                                                   1,000+
                                                           Free
10840
          4.5
               398307
                                             10,000,000+
                                                                     0
                                        19M
                                                           Free
      Content Rating
                                                        Last Updated \
                                           Genres
                                     Art & Design
0
                                                     January 7, 2018
            Everyone
1
            Everyone Art & Design; Pretend Play
                                                    January 15, 2018
2
                                     Art & Design
                                                      August 1, 2018
            Everyone
                                                        June 8, 2018
3
                 Teen
                                     Art & Design
4
            Everyone
                         Art & Design;Creativity
                                                       June 20, 2018
. . .
                                                       July 25, 2017
10836
            Everyone
                                        Education
                                                        July 6, 2018
10837
            Everyone
                                        Education
                                                    January 20, 2017
10838
            Everyone
                                          Medical
                               Books & Reference
                                                    January 19, 2015
10839
          Mature 17+
                                                       July 25, 2018
10840
            Everyone
                                        Lifestyle
```

```
Current Ver
                                  Android Ver
0
                    1.0.0
                                 4.0.3 and up
1
                    2.0.0
                                 4.0.3 and up
                                 4.0.3 and up
2
                    1.2.4
3
       Varies with device
                                   4.2 and up
4
                                   4.4 and up
                      1.1
10836
                                   4.1 and up
                     1.48
10837
                        1
                                   4.1 and up
10838
                        1
                                   2.2 and up
10839
      Varies with device Varies with device
10840
      Varies with device Varies with device
[10841 rows x 13 columns]
df.dropna(subset=["Rating"],axis=0,inplace=True) # Dropping the Null
Values of Rating Column
df.dropna(subset=["Type"],axis=0,inplace=True) # Dropping the Null
Values of Type Column
df.dropna(subset=["Content Rating"],axis=0,inplace=True) # Dropping
the Null Values of Content Rating
df.dropna(subset=["Current Ver"],axis=0,inplace=True) # Dropping the
Null Values of Current Version
df.dropna(subset=["Android Ver"],axis=0,inplace=True) # Dropping the
Null values of Android Version
#Q4) Variables seem to have incorrect type and inconsistent
formatting. You need to fix them:
# 4) Size column has sizes in Kb as well as Mb. To analyze, you'll
need to convert these to numeric.
#1) Extract the numeric value from the column
df=pd.read csv("C:/Users/shiva/Desktop/Shivani/SimpliLearn/Python/
googleplay/googleplaystore.csv")
df
                                                     App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART AND DESIGN
                                     Coloring book moana
ART AND DESIGN
```

```
U Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                     Sketch - Draw & Paint
ART AND DESIGN
                    Pixel Draw - Number Art Coloring Book
ART AND_DESIGN
                                                         . . .
. . .
10836
                                          Sya9a Maroc - FR
FAMILY
                         Fr. Mike Schmitz Audio Teachings
10837
FAMILY
                                    Parkinson Exercices FR
10838
MEDICAL
10839
                            The SCP Foundation DB fr nn5n
BOOKS AND REFERENCE
           iHoroscope - 2018 Daily Horoscope & Astrology
LIFESTYLE
       Rating Reviews
                                                 Installs
                                                            Type Price
                                       Size
                                                                        \
          4.1
                   159
                                        19M
                                                  10,000+
                                                            Free
                                                                     0
          3.9
                   967
1
                                        14M
                                                 500,000+
                                                            Free
                                                                     0
2
          4.7
                                       8.7M
                                               5,000,000+
                 87510
                                                            Free
                                                                     0
          4.5
                                              50,000,000+
3
               215644
                                        25M
                                                                     0
                                                            Free
4
          4.3
                   967
                                       2.8M
                                                 100,000+
                                                                     0
                                                            Free
10836
          4.5
                    38
                                        53M
                                                   5,000+
                                                            Free
                                                                     0
          5.0
                     4
10837
                                       3.6M
                                                     100 +
                                                            Free
                                                                     0
                     3
10838
          NaN
                                       9.5M
                                                   1,000+
                                                            Free
                                                                     0
          4.5
                                                                     0
10839
                        Varies with device
                                                   1,000+
                                                            Free
                   114
10840
          4.5
               398307
                                        19M
                                              10,000,000+
                                                                     0
                                                            Free
      Content Rating
                                                        Last Updated
                                            Genres
                                     Art & Design
                                                     January 7, 2018
0
            Everyone
                                                    January 15, 2018
1
            Everyone Art & Design; Pretend Play
2
            Everyone
                                     Art & Design
                                                      August 1, 2018
3
                 Teen
                                     Art & Design
                                                        June 8, 2018
                                                       June 20, 2018
4
            Everyone
                       Art & Design; Creativity
                                                       July 25, 2017
            Everyone
10836
                                        Education
10837
            Everyone
                                        Education
                                                        July 6, 2018
                                                    January 20, 2017
10838
            Everyone
                                          Medical
                                Books & Reference
                                                    January 19, 2015
10839
          Mature 17+
10840
            Everyone
                                        Lifestyle
                                                       July 25, 2018
               Current Ver
                                    Android Ver
0
                     1.0.0
                                   4.0.3 and up
1
                     2.0.0
                                   4.0.3 and up
2
                     1.2.4
                                   4.0.3 and up
3
       Varies with device
                                     4.2 and up
```

```
4
                       1.1
                                    4.4 and up
. . .
                       . . .
                                    4.1 and up
10836
                      1.48
10837
                         1
                                    4.1 and up
                         1
10838
                                    2.2 and up
10839
       Varies with device Varies with device
       Varies with device Varies with device
10840
[10841 rows x 13 columns]
df["Size"] # Fetching the values from the size column
0
                         19M
1
                         14M
2
                        8.7M
3
                         25M
4
                        2.8M
                 . . .
10836
                         53M
                        3.6M
10837
10838
                        9.5M
10839
         Varies with device
10840
                         19M
Name: Size, Length: 10841, dtype: object
# Q4) 2--Multiply the value by 1,000, if size is mentioned in Mb
df
                                                       App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART_AND_DESIGN
                                      Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                    Sketch - Draw & Paint
ART AND DESIGN
                    Pixel Draw - Number Art Coloring Book
ART AND DESIGN
. . .
. . .
10836
                                          Sya9a Maroc - FR
FAMILY
                         Fr. Mike Schmitz Audio Teachings
10837
FAMILY
                                   Parkinson Exercices FR
10838
MEDICAL
                            The SCP Foundation DB fr nn5n
10839
```

BOOKS_AND_REFERENCE 10840 iHoroscope - 2018 Daily Horoscope & Astrology LIFESTYLE

0 1 2 3 4	Rating Reviews 4.1 159 3.9 967 4.7 87510 4.5 215644 4.3 967		Size 19M 14M 8.7M 25M 2.8M	10 500 5,000 50,000	stalls 0,000+ 0,000+ 0,000+ 0,000+ 0,000+	Type Free Free Free Free	Price 0 0 0 0	\
10836 10837 10838 10839 10840	4.5 38 5.0 4 NaN 3 4.5 114 4.5 398307		53M 3.6M 9.5M device 19M	1	0,000+ 100+ 1,000+ 1,000+ 0,000+	Free Free Free Free Free	 0 0 0 0	
0 1 2 3 4	Content Rating Everyone Everyone Everyone Teen Everyone	Art & Design	Art & D ;Pretend Art & D Art & D	Play esign esign	Janua Januar Augu Ju	st Upo ry 7, y 15, st 1, ne 8, e 20,	2018 2018 2018 2018	\
10836 10837 10838 10839 10840	Everyone Everyone Everyone Mature 17+ Everyone	Book	Educ Me s & Refe	ation ation dical rence style	Ju Januar Januar	-	2018 2017 2015	
0 1 2 3 4	2	.0.0 4 .0.0 4 .2.4 4	Android 1.0.3 and 1.0.3 and 1.0.3 and 4.2 and 4.4 and	up up up up				
	Varies with de Varies with de			up up ice				
[10841	rows x 13 colu	mns]						
import	numny as no #	Converting t	he colum	ins to r	numeric			

import numpy as np # Converting the columns to numeric

df["Size"]=df["Size"].apply(lambda x:np.NaN if x=="Varies with device" else x)

```
df["Size"]=df["Size"].apply(lambda x:np.float64(x.replace('M',''))*1e6
if type(x)!=float and 'M' in x else x )
df["Size"]=df["Size"]*1000
df["Size"] # Multiplying the size column by 1000
         19000000000.0
0
1
         14000000000.0
2
          8700000000.0
3
         25000000000.0
4
          2800000000.0
10836
         53000000000.0
10837
          3600000000.0
10838
          9500000000.0
10839
                   NaN
         19000000000.0
10840
Name: Size, Length: 10841, dtype: object
#Q4)2- Reviews is a numeric field that is loaded as a string field.
Convert it to numeric (int/float).
df["Reviews"] # Data Frame of Reviews and the data type is object
0
            159
1
            967
2
          87510
3
         215644
            967
10836
             38
10837
              4
              3
10838
10839
            114
10840
         398307
Name: Reviews, Length: 10841, dtype: object
df["Reviews"]=pd.to numeric(df["Reviews"],errors="coerce")
df["Reviews"] # Converting the reviews to numeric
            159.0
1
            967.0
2
          87510.0
3
         215644.0
            967.0
10836
             38.0
10837
              4.0
10838
              3.0
```

10839 114.0 10840 398307.0

Name: Reviews, Length: 10841, dtype: float64

#04) 3-- Installs field is currently stored as string and has values like 1,000,000+.

#4) 3--1) Treat 1,000,000+ as 1,000,000

import pandas as pd

df=pd.read_csv("C:/Users/shiva/Desktop/Shivani/SimpliLearn/Python/
googleplay/googleplaystore.csv")

df # The original data frame of App Rating

App Category Photo Editor & Candy Camera & Grid & ScrapBook ART_AND_DESIGN Coloring book moana ART AND DESIGN U Launcher Lite - FREE Live Cool Themes, Hide ... ART_AND DESIGN Sketch - Draw & Paint ART AND DESIGN Pixel Draw - Number Art Coloring Book ART AND DESIGN Sya9a Maroc - FR 10836 **FAMILY** Fr. Mike Schmitz Audio Teachings 10837 **FAMILY** Parkinson Exercices FR 10838 MEDICAL 10839 The SCP Foundation DB fr nn5n **BOOKS AND REFERENCE** 10840 iHoroscope - 2018 Daily Horoscope & Astrology LIFESTYLE

	Rating	Reviews	Size	Installs	Type	Price	\
0	4.1	159	19M	10,000+	Free	0	
1	3.9	967	14M	500,000+	Free	0	
2	4.7	87510	8.7M	5,000,000+	Free	0	
3	4.5	215644	25M	50,000,000+	Free	0	
4	4.3	967	2.8M	100,000+	Free	0	
10836	4.5	38	53M	5,000+	Free	0	
10837	5.0	4	3.6M	100+	Free	0	
10838	NaN	3	9.5M	1,000+	Free	0	

```
114 Varies with device
                                                   1,000+
10839
          4.5
                                                           Free
                                                                     0
10840
          4.5 398307
                                        19M
                                             10,000,000+
                                                           Free
                                                                     0
      Content Rating
                                           Genres
                                                        Last Updated \
0
            Everyone
                                     Art & Design
                                                     January 7, 2018
1
                       Art & Design;Pretend Play
                                                    January 15, 2018
            Everyone
2
            Everyone
                                     Art & Design
                                                      August 1, 2018
                                                        June 8, 2018
3
                 Teen
                                     Art & Design
4
                                                       June 20, 2018
            Everyone
                         Art & Design; Creativity
. . .
                  . . .
                                                       July 25, 2017
10836
            Everyone
                                        Education
10837
            Everyone
                                        Education
                                                        July 6, 2018
                                                    January 20, 2017
10838
                                          Medical
            Everyone
10839
          Mature 17+
                                Books & Reference
                                                    January 19, 2015
                                                       July 25, 2018
10840
            Everyone
                                        Lifestyle
               Current Ver
                                    Android Ver
0
                                   4.0.3 and up
                     1.0.0
1
                     2.0.0
                                   4.0.3 and up
2
                     1.2.4
                                   4.0.3 and up
3
                                     4.2 and up
       Varies with device
4
                                     4.4 and up
                       1.1
                       . . .
10836
                      1.48
                                     4.1 and up
10837
                         1
                                     4.1 and up
10838
                         1
                                     2.2 and up
10839
       Varies with device
                            Varies with device
10840
       Varies with device Varies with device
[10841 \text{ rows } x 13 \text{ columns}]
#Q3) 2-- remove '+', ',' from the field, convert it to integer
df["Installs"] # Data Frame of Install Column
0
              10,000+
1
            500,000+
2
          5,000,000+
3
         50,000,000+
             100,000+
             . . .
               5,000+
10836
10837
                 100+
10838
               1,000+
10839
               1,000+
10840
         10,000,000+
Name: Installs, Length: 10841, dtype: object
df['Installs'] = df['Installs'].str.replace('+',' ',regex=True)
df['Installs'] = df['Installs'].str.replace(',','',regex=True)
df["Installs"]
```

```
0
            10000
1
           500000
2
          5000000
3
         50000000
4
           100000
10836
             5000
10837
               100
10838
              1000
10839
              1000
10840
         10000000
Name: Installs, Length: 10841, dtype: object
# Q4) 4--Price field is a string and has $ symbol. Remove '$' sign,
and convert it to numeric.
df["Price"]
0
         0
1
         0
2
         0
3
         0
         0
10836
         0
10837
         0
10838
         0
10839
         0
10840
         0
Name: Price, Length: 10841, dtype: object
df["Price"] = df["Price"].replace({'\$': ''}, regex=True)
Replacing the $ sign from Price column
df["Price"]
0
         0
1
         0
2
         0
3
         0
         0
10836
         0
10837
         0
10838
         0
10839
         0
10840
Name: Price, Length: 10841, dtype: object
df.dtypes # Here Price column is of object data type
```

```
object
App
Category
                   object
Rating
                  float64
Reviews
                   object
Size
                   object
Installs
                   object
Tvpe
                   object
Price
                   object
Content Rating
                   object
Genres
                   object
Last Updated
                   object
Current Ver
                   object
Android Ver
                   object
dtype: object
df["Price"]=pd.to numeric(df["Price"],errors="coerce")
df["Price"] # Converting the Price column to numeric
0
         0.0
1
         0.0
2
         0.0
3
         0.0
4
         0.0
        . . .
10836
         0.0
10837
         0.0
10838
         0.0
10839
         0.0
10840
         0.0
Name: Price, Length: 10841, dtype: float64
# Q5) Sanity checks:
#1) Average rating should be between 1 and 5 as only these values
#are allowed on the play store. Drop the rows that have a value
outside this range.
import pandas as pd
df=pd.read csv("C:/Users/shiva/Desktop/Shivani/SimpliLearn/Python/
googleplay/googleplaystore.csv")
df # The data frame of App Rating Prediction
                                                      App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART AND DESIGN
                                      Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
```

```
ART AND DESIGN
                                      Sketch - Draw & Paint
ART AND DESIGN
                    Pixel Draw - Number Art Coloring Book
ART AND DESIGN
. . .
. . .
                                           Sya9a Maroc - FR
10836
FAMILY
10837
                          Fr. Mike Schmitz Audio Teachings
FAMILY
                                     Parkinson Exercices FR
10838
MEDICAL
                             The SCP Foundation DB fr nn5n
10839
BOOKS AND REFERENCE
10840
           iHoroscope - 2018 Daily Horoscope & Astrology
LIFESTYLE
       Rating Reviews
                                        Size
                                                  Installs
                                                             Type Price
                                                                         \
0
           4.1
                   159
                                         19M
                                                   10,000+
                                                             Free
                                                                       0
           3.9
                   967
                                                                       0
1
                                         14M
                                                  500,000+
                                                             Free
2
          4.7
                 87510
                                        8.7M
                                               5,000,000+
                                                             Free
                                                                       0
3
          4.5
                                              50,000,000+
                                                                       0
                215644
                                         25M
                                                             Free
          4.3
4
                                        2.8M
                                                  100,000+
                                                             Free
                                                                       0
                   967
                                                              . . .
           . . .
                   . . .
                                                                     . . .
                    38
                                         53M
                                                    5,000+
                                                                      0
10836
          4.5
                                                             Free
10837
          5.0
                     4
                                        3.6M
                                                      100+
                                                             Free
                                                                       0
                     3
                                        9.5M
                                                                       0
10838
          NaN
                                                    1.000 +
                                                             Free
          4.5
10839
                   114
                         Varies with device
                                                    1,000+
                                                             Free
                                                                       0
                                                                       0
10840
          4.5
                398307
                                         19M
                                              10,000,000+
                                                             Free
      Content Rating
                                                         Last Updated
                                            Genres
                                      Art & Design
0
                                                      January 7, 2018
             Everyone
1
                       Art & Design; Pretend Play
                                                     January 15, 2018
             Everyone
2
                                      Art & Design
             Everyone
                                                       August 1, 2018
3
                 Teen
                                      Art & Design
                                                         June 8, 2018
4
                          Art & Design; Creativity
                                                        June 20, 2018
             Everyone
                                                        July 25, 2017
             Everyone
                                         Education
10836
                                                         July 6, 2018
             Everyone
                                         Education
10837
10838
             Everyone
                                           Medical
                                                     January 20, 2017
                                Books & Reference
                                                     January 19, 2015
10839
          Mature 17+
                                                        July 25, 2018
10840
             Everyone
                                         Lifestyle
               Current Ver
                                    Android Ver
0
                     1.0.0
                                   4.0.3 and up
1
                                   4.0.3 and up
                     2.0.0
2
                     1.2.4
                                   4.0.3 and up
3
       Varies with device
                                     4.2 and up
4
                                      4.4 and up
                        1.1
```

```
. . .
10836
                     1.48
                                   4.1 and up
                                   4.1 and up
10837
                        1
10838
                        1
                                   2.2 and up
      Varies with device Varies with device
10839
10840
      Varies with device Varies with device
[10841 rows x 13 columns]
df["Rating"] # Viewing the Rating Column
0
         4.1
1
         3.9
2
         4.7
3
         4.5
         4.3
        . . .
        4.5
10836
10837
         5.0
10838
        NaN
10839
         4.5
         4.5
10840
Name: Rating, Length: 10841, dtype: float64
df.dropna(subset=["Rating"],axis=0,inplace=True) # Dropping the Nan
Values of Rating Column
df["Rating"]
         4.1
1
         3.9
2
         4.7
3
         4.5
         4.3
10834
        4.0
10836
        4.5
10837
         5.0
10839
         4.5
         4.5
10840
Name: Rating, Length: 9367, dtype: float64
df[df.Rating>5] # Data Frame of Rating Column
                                           App Category Rating
Reviews \
10472 Life Made WI-Fi Touchscreen Photo Frame
                                                    1.9
                                                            19.0
3.0M
        Size Installs Type Price Content Rating
Genres \
```

. . .

```
10472 1,000+ Free 0 Everyone
                                                 NaN February 11,
2018
      Last Updated Current Ver Android Ver
10472
            1.0.19 4.0 and up
                                       NaN
df.drop([10472],inplace=True) # Dropping the row number
df["Rating"] # Dropping the row number 10472, because rating is 19
which is greater than 5
         4.1
0
1
         3.9
2
         4.7
3
         4.5
4
         4.3
10834
        4.0
10836
        4.5
10837
         5.0
10839
         4.5
10840
         4.5
Name: Rating, Length: 9366, dtype: float64
#Q5)2--Reviews should not be more than installs as only those who
installed can review the app.
#If there are any such records, drop them.
df # The data frame
                                                     App
Category \
          Photo Editor & Candy Camera & Grid & ScrapBook
ART AND DESIGN
                                     Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
ART_AND_DESIGN
                                   Sketch - Draw & Paint
ART AND DESIGN
                   Pixel Draw - Number Art Coloring Book
ART AND DESIGN
. . .
                                           FR Calculator
10834
FAMILY
                                        Sya9a Maroc - FR
10836
FAMILY
                        Fr. Mike Schmitz Audio Teachings
10837
FAMILY
                           The SCP Foundation DB fr nn5n
10839
```

0 1 2 3 4 10834 10836 10837 10839 10840	Rating Reviews 4.1 159 3.9 967 4.7 87510 4.5 215644 4.3 967 4.0 7 4.5 38 5.0 4 4.5 114 4.5 398307	2	19M 14M 5 8.7M 5,6 25M 50,6 8.8M 1 8.6M 53M 8.6M rice	Installs 10,000+ 600,000+ 000,000+ 000,000+ 500+ 5,000+ 1,000+ 000,000+	Type Free Free Free Free Free Free Free Fr	Price 0 0 0 0 0 0 0	\
0 1 2 3 4 10834 10836 10837 10839	Content Rating Everyone Everyone Teen Everyone Everyone Everyone Everyone Mature 17+	Art & Design;Pre Art Art Art & Design;C	& Design Creativity Creativity Education Education Reference	Janua Januar Augu Jun Jun Jul Jul Januar	st 1, ne 8, e 20, e 18, y 25, ly 6, y 19,	2018 2018 2018 2018 2018 2017 2017 2018 2015	\
10840 0 1 2 3 4 10834 10836 10837 10840	Varies with dev	.0.0 4.0.3 .0.0 4.0.3 .2.4 4.0.3 vice 4.2 1.1 4.40.0 4.1 1.48 4.1 vice Varies with		e Jul	y 25,	2018	
[9366	rows x 13 column	ns]					
df["Re	views"]						
0 1 2	159 967 87510						

```
3
         215644
            967
              7
10834
10836
             38
10837
10839
            114
10840
         398307
Name: Reviews, Length: 9366, dtype: object
Total count of reviews=df["Reviews"].count() # Count of Reviews
Column
Total count of reviews
9366
Total_count_of_installs=df["Installs"].count() # Count of Installs
Column
Total count of installs
9366
```

The Total Count of Reviews column is same as that of Installs column which is 9336. Hence,need to drop any values.

#Q4) 3--For free apps (type = "Free"), the price should not be >0. Drop any such rows.

freeaps=df[["Price","Type"]] # Clubing the Price and Type Column of
data frame

freeaps

```
Price Type
0
          0 Free
1
          0 Free
2
          0 Free
3
          0 Free
4
          0 Free
        . . .
         0 Free
10834
10836
         0 Free
         0 Free
10837
10839
         0 Free
10840
         0 Free
```

[9366 rows x 2 columns]

Q5) Performing univariate analysis:

#Boxplot for Price. Are there any outliers? Think about the price of usual apps on Play Store

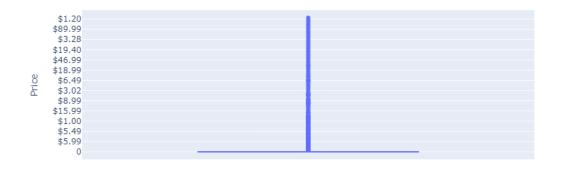
import pandas as pd

df=pd.read_csv("C:/Users/shiva/Desktop/Shivani/SimpliLearn/Python/
googleplay/googleplaystore.csv")

df # The data frame

```
App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART AND DESIGN
                                        Coloring book moana
ART AND DESIGN
       \overline{\sf U} Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                     Sketch - Draw & Paint
ART AND DESIGN
                    Pixel Draw - Number Art Coloring Book
ART AND DESIGN
. . .
                                                         . . .
. . .
10836
                                           Sya9a Maroc - FR
FAMILY
10837
                         Fr. Mike Schmitz Audio Teachings
FAMILY
                                    Parkinson Exercices FR
10838
MEDICAL
                             The SCP Foundation DB fr nn5n
10839
BOOKS AND REFERENCE
10840
           iHoroscope - 2018 Daily Horoscope & Astrology
LIFESTYLE
       Rating Reviews
                                        Size
                                                 Installs
                                                            Type Price
0
          4.1
                   159
                                         19M
                                                   10,000+
                                                            Free
                                                                      0
          3.9
                                                 500,000+
                                                            Free
                                                                      0
1
                   967
                                         14M
2
          4.7
                87510
                                        8.7M
                                               5,000,000+
                                                            Free
                                                                      0
3
          4.5
               215644
                                         25M
                                              50,000,000+
                                                            Free
                                                                      0
4
          4.3
                   967
                                        2.8M
                                                 100,000+
                                                            Free
                                                                      0
                                         . . .
                                                              . . .
10836
          4.5
                    38
                                         53M
                                                    5,000+
                                                            Free
                                                                      0
10837
          5.0
                     4
                                        3.6M
                                                      100+
                                                            Free
                                                                      0
10838
          NaN
                     3
                                        9.5M
                                                    1,000+
                                                            Free
                                                                      0
10839
          4.5
                   114
                        Varies with device
                                                    1,000+
                                                            Free
                                                                      0
10840
          4.5
                                         19M
                                              10,000,000+
               398307
                                                            Free
                                                                      0
      Content Rating
                                                         Last Updated \
                                            Genres
0
             Everyone
                                     Art & Design
                                                      January 7, 2018
1
             Everyone Art & Design; Pretend Play January 15, 2018
```

```
Art & Design
2
            Everyone
                                                      August 1, 2018
3
                 Teen
                                     Art & Design
                                                        June 8, 2018
4
            Everyone
                         Art & Design;Creativity
                                                       June 20, 2018
                  . . .
                                                       July 25, 2017
10836
            Everyone
                                        Education
10837
            Everyone
                                        Education
                                                        July 6, 2018
10838
                                          Medical
                                                    January 20, 2017
            Everyone
                                                    January 19, 2015
10839
          Mature 17+
                                Books & Reference
10840
            Everyone
                                        Lifestyle
                                                       July 25, 2018
               Current Ver
                                    Android Ver
0
                     1.0.0
                                   4.0.3 and up
1
                     2.0.0
                                   4.0.3 and up
2
                     1.2.4
                                   4.0.3 and up
3
       Varies with device
                                     4.2 and up
4
                                     4.4 and up
                       1.1
. . .
                       . . .
10836
                      1.48
                                     4.1 and up
10837
                         1
                                     4.1 and up
10838
                         1
                                     2.2 and up
10839
       Varies with device
                            Varies with device
10840
       Varies with device Varies with device
[10841 \text{ rows } x 13 \text{ columns}]
# Creating the Box Plot
import plotly.express as px
y=df["Price"] # Placing the Price column on y axis
fig=px.box(df,y="Price")
fig # The Box Plot, placing the price value on y axis
```



From the above mentioned figure, it is clear that thick line near 0 is the box part of our box plot. Above the box and upper fence are some points showing outliers. The Box point values can clearly be viewed by just hovering over the box plot

```
# Finding the outliers through statistical functions
df # The App Rating data frame
```

```
App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART_AND_DESIGN
                                       Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                     Sketch - Draw & Paint
ART AND DESIGN
                    Pixel Draw - Number Art Coloring Book
ART AND DESIGN
. . .
10836
                                           Sva9a Maroc - FR
FAMILY
                         Fr. Mike Schmitz Audio Teachings
10837
FAMILY
                                    Parkinson Exercices FR
10838
MEDICAL
10839
                             The SCP Foundation DB fr nn5n
BOOKS AND REFERENCE
           iHoroscope - 2018 Daily Horoscope & Astrology
10840
LIFESTYLE
       Rating Reviews
                                       Size
                                                 Installs
                                                            Type Price
          4.1
                                                            Free
0
                   159
                                        19M
                                                  10,000+
                                                                      0
1
          3.9
                   967
                                         14M
                                                 500,000+
                                                            Free
                                                                      0
2
          4.7
                 87510
                                       8.7M
                                               5,000,000+
                                                            Free
                                                                      0
3
          4.5
                215644
                                        25M
                                              50,000,000+
                                                            Free
                                                                      0
4
          4.3
                                       2.8M
                                                 100,000+
                   967
                                                            Free
                                                                      0
                                         . . .
          . . .
                                                             . . .
          4.5
                    38
                                         53M
                                                   5,000+
                                                                      0
10836
                                                            Free
          5.0
10837
                     4
                                       3.6M
                                                      100+
                                                            Free
                                                                      0
10838
          NaN
                     3
                                       9.5M
                                                   1,000+
                                                            Free
                                                                      0
10839
          4.5
                   114
                        Varies with device
                                                   1,000+
                                                            Free
                                                                      0
                                        19M
10840
          4.5
                398307
                                              10,000,000+
                                                            Free
                                                                      0
      Content Rating
                                                         Last Updated
                                            Genres
0
                                     Art & Design
                                                     January 7, 2018
            Everyone
1
            Everyone Art & Design; Pretend Play
                                                    January 15, 2018
2
                                                       August 1, 2018
            Everyone
                                     Art & Design
```

```
3
                                   Art & Design
                                                      June 8, 2018
                Teen
4
            Everyone
                        Art & Design; Creativity
                                                     June 20, 2018
10836
                                       Education
                                                     July 25, 2017
            Everyone
                                                      July 6, 2018
10837
            Everyone
                                       Education
                                                  January 20, 2017
10838
            Everyone
                                        Medical
                              Books & Reference
                                                  January 19, 2015
10839
          Mature 17+
10840
                                                     July 25, 2018
            Everyone
                                       Lifestyle
              Current Ver
                                  Android Ver
                    1.0.0
                                 4.0.3 and up
0
                                 4.0.3 and up
1
                    2.0.0
2
                    1.2.4
                                 4.0.3 and up
3
       Varies with device
                                   4.2 and up
4
                      1.1
                                   4.4 and up
                                   4.1 and up
10836
                     1.48
10837
                        1
                                   4.1 and up
10838
                                   2.2 and up
                        1
10839
      Varies with device Varies with device
10840 Varies with device Varies with device
[10841 rows x 13 columns]
def finding outliers(df):
    q1=df.quantile(0.25)
    q3=df.quantile(0.75)
    iqr=q3-q1
    outliers = df[((df<(q1-1.5*iqr)) | (df>(q3+1.5*iqr)))]
    return outliers
df["Price"]=pd.to numeric(df["Price"],errors="coerce") # Converting
the price to numeric
outliers = finding outliers(df["Price"])
print("number of outliers: "+ str(len(outliers)))
number of outliers: 0
# Visualization of Outliers through describe method
df.describe()["Price"]
count
         10040.0
mean
             0.0
```

```
      std
      0.0

      min
      0.0

      25%
      0.0

      50%
      0.0

      75%
      0.0

      max
      0.0
```

Name: Price, dtype: float64

Outliers can also be viewed by describe method on Price column. Here through describe method one can analyze that # mean value of Price is 0 and max value is also 0 which clearly indicates that mean value is not sensitive to max #as both are 0. Hence There Price column has no outlier

#Q5) 5- Boxplot for Reviews

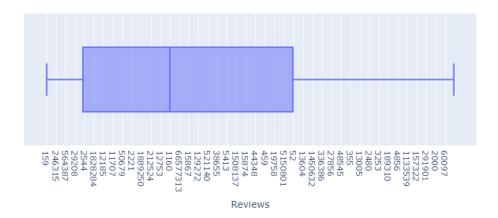
#Are there any apps with very high number of reviews? Do the values seem right?

import plotly.express as px

```
y=df["Reviews"]
```

fig=px.box(df,y)

fig # This figure shows the BoxPlot for Reviews



df.dtypes # Checking the data types of Data Frame

object
object
float64
object
object
object
object
float64

```
Content Rating
                   object
Genres
                   object
Last Updated
                   object
Current Ver
                   object
Android Ver
                   object
dtype: object
df["Reviews"].max()
'9992'
df[2544:]
                                                   App
Category \
2544
                                             Facebook
SOCIAL
2545
                                             Instagram
SOCIAL
2546
                                        Facebook Lite
SOCIAL 
2547
         Messages, Text and Video Chat for Messenger
SOCIAL
                                               Tumblr
2548
SOCIAL
. . .
. . .
10836
                                     Sya9a Maroc - FR
FAMILY
10837
                    Fr. Mike Schmitz Audio Teachings
FAMILY
                               Parkinson Exercices FR
10838
MEDICAL
                       The SCP Foundation DB fr nn5n
10839
BOOKS AND REFERENCE
10840 iHoroscope - 2018 Daily Horoscope & Astrology
LIFESTYLE
       Rating
                Reviews
                                        Size
                                                     Installs
Price
2544
          4.1
               78158306 Varies with device
                                              1,000,000,000+
0.0
               66577313 Varies with device
2545
          4.5
                                              1,000,000,000+
0.0
2546
          4.3
                8606259 Varies with device
                                                 500,000,000+
```

0.0

0.0 2548

0.0

2547

4.4

4.4

. . .

49173

. . .

2955326 Varies with device

4.0M

. . .

Type

Free

Free

Free

Free

Free

. . .

10,000,000+

100,000,000+

. . .

10836 0.0	4.5	38		53M	5,000	+ Free
10837	5.0	4	3	3.6M	100	+ Free
0.0 10838	NaN	3	Q	9.5M	1,000	+ Free
0.0 10839 0.0 10840 0.0	4.5	114	Varies with dev	vice	1,000	+ Free
	4.5	398307		19M 10,	000,000	+ Free
	Content	_	Genres	Last U	Ipdated	
2544	t Ver \	Teen	Social	August 3	, 2018	Varies with
device 2545		Teen	Social	July 31	., 2018	Varies with
device 2546		Teen	Social	August 1	., 2018	Varies with
device 2547	Ev	eryone	Social	June 4	, 2018	
1.24 2548 device	Matu	re 17+	Social	August 1	., 2018	Varies with
10836	Ev	eryone	Education	July 25	, 2017	
1.48 10837	Ev	eryone	Education	July 6	, 2018	
1 10838	Ev	eryone	Medical	January 20	, 2017	
1 10839	Matu	re 17+ B	ooks & Reference	January 19	, 2015	Varies with
device 10840 device	Ev	eryone	Lifestyle	July 25	, 2018	Varies with
2544 2545 2546 2547 2548 10836 10837	Varies Varies Varies	Android Vowith device with device 4.1 and continuous 4.1 and continuou	ce ce up ce up			
10838 10839 10840		2.2 and with devious with deviction in the contract of the con	ce			

[8297 rows x 13 columns]

df.iat[2544,0] # Use iat property for finding the name of App

'Facebook'

By using Iat property,I found that Facebook App is having highest number of reviews. The review count of Facebook App is 78158306.0 which is the highest

#Q5) c- Histogram for Rating

#How are the ratings distributed? Is it more toward higher ratings?

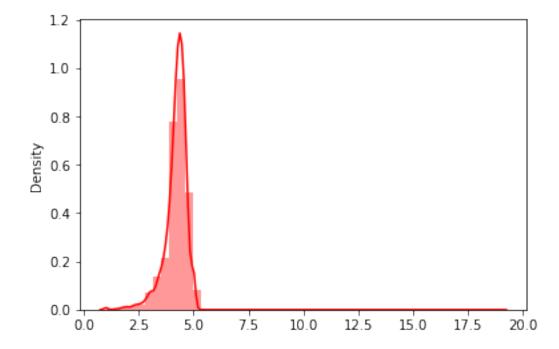
import seaborn as sns

sns.distplot(x=df["Rating"],hist=True,color="r",label="Actual Value")

C:\Users\shiva\anaconda3\lib\site-packages\seaborn\
distributions.py:2619: FutureWarning:

`distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

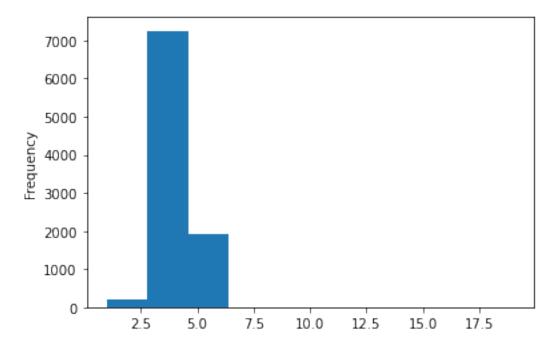
<AxesSubplot:ylabel='Density'>



import matplotlib.pyplot as plt

```
df["Rating"].plot(kind="hist")
```

<AxesSubplot:ylabel='Frequency'>



Histogram is the graph of the Frequency distributions of values. It can be created by matplotlib or seaborn library. From the above mentioned, histogram it is clear the values are not uniformly distributed towards ratings. It is not even maximum towards highest ratings. It is maximum for ratings which are in the range 0f 3 to 3.9(approx)

```
#Q5) c--Histogram for Size
```

```
import pandas as pd
import numpy as np
import seaborn as sns

df=pd.read_csv("C:/Users/shiva/Desktop/Shivani/SimpliLearn/Python/
googleplay/googleplaystore.csv")
```

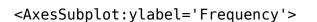
df # The original data frame

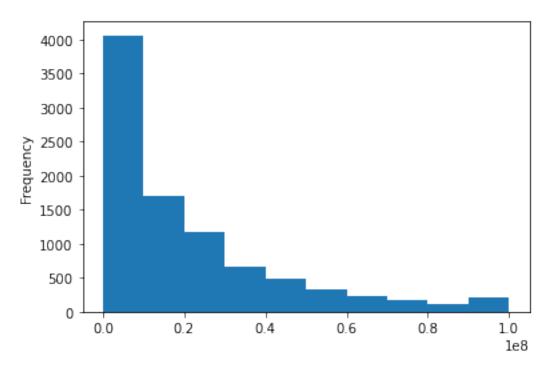
```
App
Category \
0 Photo Editor & Candy Camera & Grid & ScrapBook
ART_AND_DESIGN
Coloring book moana
ART_AND_DESIGN
U Launcher Lite - FREE Live Cool Themes, Hide ...
ART_AND_DESIGN
Sketch - Draw & Paint
```

```
ART AND DESIGN
                    Pixel Draw - Number Art Coloring Book
ART AND DESIGN
. . .
10836
                                           Sya9a Maroc - FR
FAMILY
                          Fr. Mike Schmitz Audio Teachings
10837
FAMILY
10838
                                     Parkinson Exercices FR
MEDICAL
10839
                             The SCP Foundation DB fr nn5n
BOOKS AND REFERENCE
            iHoroscope - 2018 Daily Horoscope & Astrology
10840
LIFESTYLE
       Rating Reviews
                                        Size
                                                  Installs
                                                             Type Price
           4.1
                                         19M
                                                   10,000+
                                                             Free
0
                   159
                                                                       0
           3.9
                                                  500,000+
                                                                       0
1
                   967
                                         14M
                                                             Free
2
          4.7
                 87510
                                        8.7M
                                                5,000,000+
                                                             Free
                                                                       0
3
                                               50,000,000+
          4.5
                215644
                                         25M
                                                             Free
                                                                       0
4
          4.3
                   967
                                        2.8M
                                                  100,000+
                                                             Free
                                                                       0
                                         . . .
                                                              . . .
           . . .
                                                        . . .
          4.5
                    38
                                         53M
                                                    5,000+
                                                             Free
                                                                       0
10836
          5.0
                     4
                                        3.6M
                                                             Free
                                                                       0
10837
                                                      100 +
                     3
                                                    1,000+
                                                                       0
10838
          NaN
                                        9.5M
                                                             Free
10839
          4.5
                   114
                         Varies with device
                                                    1,000+
                                                             Free
                                                                       0
10840
          4.5
                                                                       0
                398307
                                         19M
                                               10,000,000+
                                                             Free
      Content Rating
                                            Genres
                                                         Last Updated
0
             Everyone
                                      Art & Design
                                                      January 7, 2018
1
             Everyone
                      Art & Design; Pretend Play
                                                     January 15, 2018
2
                                      Art & Design
                                                       August 1, 2018
             Everyone
3
                                                          June 8, 2018
                 Teen
                                      Art & Design
                                                        June 20, 2018
4
             Everyone
                         Art & Design; Creativity
10836
             Everyone
                                         Education
                                                        July 25, 2017
                                                          July 6, 2018
10837
             Everyone
                                         Education
10838
             Everyone
                                           Medical
                                                     January 20, 2017
                                                     January 19, 2015
10839
          Mature 17+
                                Books & Reference
10840
             Everyone
                                         Lifestyle
                                                        July 25, 2018
               Current Ver
                                     Android Ver
0
                     1.0.0
                                    4.0.3 and up
1
                     2.0.0
                                    4.0.3 and up
2
                     1.2.4
                                    4.0.3 and up
3
       Varies with device
                                      4.2 and up
                        1.1
                                      4.4 and up
                        . . .
. . .
10836
                       1.48
                                      4.1 and up
```

```
10837
                         1
                                    4.1 and up
10838
                         1
                                    2.2 and up
10839
       Varies with device Varies with device
10840
      Varies with device Varies with device
[10841 rows x 13 columns]
df["Size"]
0
                         19M
1
                         14M
2
                        8.7M
3
                         25M
4
                        2.8M
                 . . .
10836
                         53M
10837
                        3.6M
10838
                        9.5M
10839
         Varies with device
10840
                         19M
Name: Size, Length: 10841, dtype: object
df["Size"]=df["Size"].apply(lambda x:np.NaN if x=="Varies with device"
else x)
df["Size"] # Modification of size column by removing the string k and
M from Size
0
         19000000.0
         14000000.0
1
2
          8700000.0
3
         25000000.0
4
          2800000.0
10836
         53000000.0
10837
          3600000.0
10838
          9500000.0
10839
                NaN
         19000000.0
10840
Name: Size, Length: 10841, dtype: object
df.dtypes
                    object
App
                    object
Category
Rating
                   float64
Reviews
                    object
Size
                    object
Installs
                    object
Type
                    object
Price
                    object
Content Rating
                    object
```

```
Genres
                    object
Last Updated
                    object
Current Ver
                    object
Android Ver
                    object
dtype: object
df["Size"]=pd.to_numeric(df["Size"],errors="coerce") # Converting the
size to numeric
df["Size"]
0
         19000000.0
1
         14000000.0
2
          8700000.0
3
         25000000.0
4
          2800000.0
10836
         53000000.0
10837
          3600000.0
10838
          9500000.0
10839
                NaN
10840
         19000000.0
Name: Size, Length: 10841, dtype: float64
```

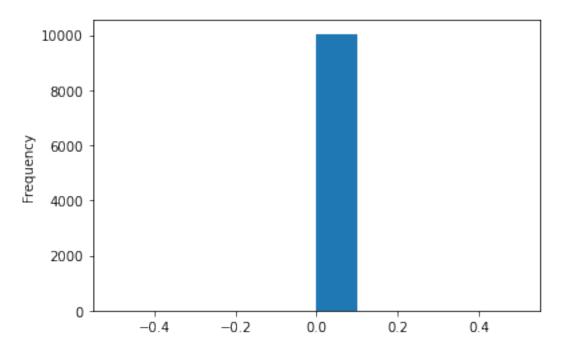




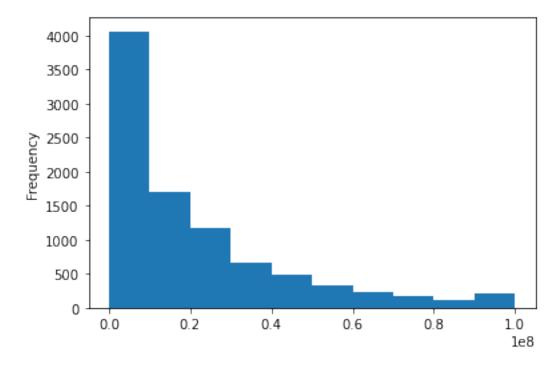
df["Size"].plot(kind="hist") # Histogram of Size Column

Q5) -- Note down your observations for the plots made above. Which of these seem to have outliers?

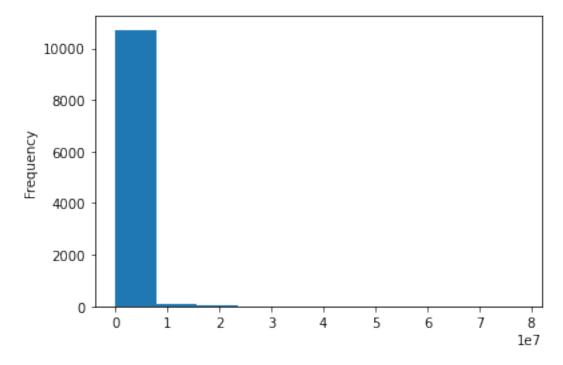
```
df.dtypes
App
                    object
Category
                    object
Rating
                   float64
Reviews
                    object
Size
                   float64
Installs
                    object
Type
                    object
Price
                    object
Content Rating
                    object
Genres
                    object
Last Updated
                    object
Current Ver
                    object
Android Ver
                    object
dtype: object
df["Price"]=pd.to_numeric(df["Price"],errors="coerce")
df["Reviews"]=pd.to numeric(df["Reviews"],errors="coerce")
df.dtypes # Converting the Price and Reviews to Numeric
App
                    object
Category
                    object
                   float64
Rating
                   float64
Reviews
                   float64
Size
Installs
                    object
Type
                    object
                   float64
Price
Content Rating
                    object
Genres
                    object
Last Updated
                    object
Current Ver
                    object
Android Ver
                    object
dtype: object
The below mentioned are the histogram of Price, Size, Ratings and Reviews
df["Price"].plot(kind="hist") # Histogram of Price Column
<AxesSubplot:ylabel='Frequency'>
```



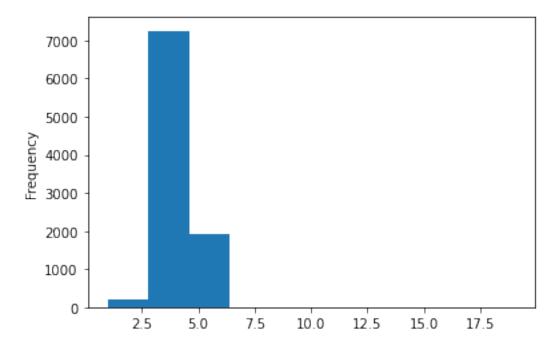
df["Size"].plot(kind="hist") # Histogram of size column
<AxesSubplot:ylabel='Frequency'>



df["Reviews"].plot(kind="hist") # Histogram of Reviews column
<AxesSubplot:ylabel='Frequency'>



df["Rating"].plot(kind="hist") # Histogram of Ratings column
<AxesSubplot:ylabel='Frequency'>



For Outlier Analysis, use the describe method on the above mentioned columns df["Price"].describe()

count 10040.0 mean 0.0

```
std     0.0
min     0.0
25%     0.0
50%     0.0
75%     0.0
max     0.0
Name: Price, dtype: float64
```

I applied the describe method on Price column. Here mean is 0 and max is also 0. The mean is not sensitive to max. Hence, this column doesn't have outlier

```
df["Reviews"].describe()
count
         1.084000e+04
mean
         4.441529e+05
         2.927761e+06
std
         0.000000e+00
min
25%
         3.800000e+01
50%
         2.094000e+03
         5.477550e+04
75%
         7.815831e+07
max
Name: Reviews, dtype: float64
```

Outliers of the Review Column can be analyzed through Interquartile Range(IQR) ie by IQR Method.

```
def find_outliers_IQR(df):
    q1=df.quantile(0.25)
    q3=df.quantile(0.75)
    IQR=q3-q1
    outliers = df[((df<(q1-1.5*IQR)) | (df>(q3+1.5*IQR)))]
    return outliers
outliers = find_outliers_IQR(df["Reviews"])
print("number of outliers: "+ str(len(outliers)))
print("max outlier value: "+ str(outliers.max()))
print("min outlier value: "+ str(outliers.min()))
number of outliers: 1924
max outlier value: 78158306.0
min outlier value: 137144.0
```

There are total 1924 outliers in Reviews Column

df["Rating"].describe() # Applying the describe menthod on Rating
Column

```
9367,000000
count
            4.193338
mean
std
            0.537431
            1.000000
min
25%
            4.000000
50%
            4.300000
75%
            4.500000
           19.000000
max
Name: Rating, dtype: float64
def find outliers IQR(df):
    q1=df.quantile(0.25)
    q3=df.quantile(0.75)
    IQR=q3-q1
    outliers = df[((df<(q1-1.5*IQR)) | (df>(q3+1.5*IQR)))]
    return outliers
outliers = find outliers IQR(df["Rating"])
print("number of outliers: "+ str(len(outliers)))
print("max outlier value: "+ str(outliers.max()))
print("min outlier value: "+ str(outliers.min()))
number of outliers: 504
max outlier value: 19.0
min outlier value: 1.0
```

The Rating Column also has the Outliers. From describe method, I analyze that mean value is sensitive to max The maximum value of Rating is 19.0 and the mean value is 4.1 which clearly indicates that mean value is very small to max,hence this column has outliers. The 504 is total number of outliers of Rating Column. The minimum value is 1.0 and maximum value is 19.0

```
10836
                         53M
10837
                        3.6M
10838
                        9.5M
10839
         Varies with device
10840
                         19M
Name: Size, Length: 10841, dtype: object
import numpy as np
df["Size"]=df["Size"].apply(lambda x:np.NaN if x=="Varies with device"
else x)
df["Size"] # Data Frame of Size Column
         19000000.0
1
         14000000.0
2
          8700000.0
3
         25000000.0
          2800000.0
         53000000.0
10836
10837
          3600000.0
10838
          9500000.0
10839
                NaN
         19000000.0
10840
Name: Size, Length: 10841, dtype: object
df["Size"]=pd.to numeric(df["Size"],errors="coerce")
df["Size"] # Converting Size column to numeric
0
         19000000.0
1
         14000000.0
2
          8700000.0
3
         25000000.0
          2800000.0
10836
         53000000.0
10837
          3600000.0
10838
          9500000.0
10839
                NaN
10840
         19000000.0
Name: Size, Length: 10841, dtype: float64
df["Size"].plot(kind="hist") # Histogram of Size Column
<AxesSubplot:ylabel='Frequency'>
```

```
3500 -

3000 -

2500 -

1500 -

1000 -

500 -

0.0 0.2 0.4 0.6 0.8 1.0 le8
```

```
df["Size"].describe()
         8.829000e+03
count
         2.227054e+07
mean
         2.262869e+07
std
min
         1.000000e+06
25%
         5.400000e+06
50%
         1.400000e+07
75%
         3.100000e+07
         1.000000e+08
max
Name: Size, dtype: float64
def finding outliers(df):
    q1=df.quantile(0.25)
    q3=df.quantile(0.75)
    iqr=q3-q1
    outliers = df[((df<(q1-1.5*iqr)) | (df>(q3+1.5*iqr)))]
    return outliers
outliers = finding outliers(df["Size"])
print("Number of outliers: "+ str(len(outliers)))
print("max outlier value: "+ str(outliers.max()))
print("min outlier value: "+ str(outliers.min()))
```

Number of outliers: 523

max outlier value: 100000000.0 min outlier value: 70000000.0

The size column has 523 outliers. The minimum outlier value is 70000000.0 and maximum outlier value is 100000000.0

```
# Q6)- Outlier treatment:
```

Q6)1-Price: From the box plot, it seems like there are some apps with very high price.

#A price of \$200 for an application on the Play Store is very high and suspicious!

#6)1-Check out the records with very high price

#6) 1-Is 200 indeed a high price?

import pandas as pd
import numpy as np

df=pd.read_csv("C:/Users/shiva/Desktop/Shivani/SimpliLearn/Python/
googleplay/googleplaystore.csv")

df

App

Category \
0 Photo Editor & Candy Camera & Grid & ScrapBook

ART_AND_DESIGN
Coloring book moana

ART AND DESIGN

2 $\overline{\mathsf{U}}$ Launcher Lite — FREE Live Cool Themes, Hide ... ART AND DESIGN

AILI_AIND_DESTON

3 Sketch - Draw & Paint

ART_AND_DESIGN

4 Pixel Draw - Number Art Coloring Book

ART_AND_DESIGN

. . .

10836 Sya9a Maroc - FR

FAMILY

10837 Fr. Mike Schmitz Audio Teachings

FAMILY

10838 Parkinson Exercices FR

MEDICAL

10839 The SCP Foundation DB fr nn5n

BOOKS AND REFERENCE

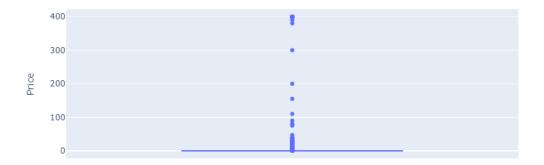
10840 iHoroscope - 2018 Daily Horoscope & Astrology

LIFESTYLE

	Rating	Reviews	Size	Installs	Type	Price	\
0	4.1	159	19M	10,000+	Free	0	
1	3.9	967	14M	500,000+	Free	0	

```
Free
2
          4.7
                87510
                                        8.7M
                                               5,000,000+
                                                                      0
3
                                              50,000,000+
          4.5
                215644
                                        25M
                                                            Free
                                                                      0
4
          4.3
                   967
                                        2.8M
                                                  100,000+
                                                            Free
                                                                      0
                                         . . .
          . . .
                   . . .
                                                              . . .
          4.5
                                                    5,000+
                                                                      0
10836
                    38
                                         53M
                                                            Free
10837
          5.0
                     4
                                        3.6M
                                                      100+
                                                            Free
                                                                      0
                     3
                                                                      0
10838
          NaN
                                        9.5M
                                                    1.000 +
                                                            Free
10839
          4.5
                                                    1,000+
                                                                      0
                   114
                        Varies with device
                                                            Free
10840
          4.5
                398307
                                         19M
                                              10,000,000+
                                                            Free
                                                                      0
      Content Rating
                                            Genres
                                                         Last Updated \
0
             Everyone
                                     Art & Design
                                                      January 7, 2018
1
             Everyone Art & Design; Pretend Play
                                                     January 15, 2018
2
             Everyone
                                     Art & Design
                                                       August 1, 2018
                                                         June 8, 2018
3
                 Teen
                                      Art & Design
4
                        Art & Design; Creativity
                                                        June 20, 2018
             Everyone
. . .
10836
                                         Education
                                                        July 25, 2017
             Everyone
10837
             Everyone
                                         Education
                                                         July 6, 2018
10838
             Everyone
                                           Medical
                                                     January 20, 2017
                                                     January 19, 2015
10839
          Mature 17+
                                Books & Reference
                                                        July 25, 2018
10840
             Everyone
                                         Lifestyle
               Current Ver
                                    Android Ver
0
                     1.0.0
                                   4.0.3 and up
1
                     2.0.0
                                   4.0.3 and up
2
                                   4.0.3 and up
                     1.2.4
3
       Varies with device
                                     4.2 and up
4
                       1.1
                                      4.4 and up
                        . . .
10836
                       1.48
                                      4.1 and up
10837
                          1
                                      4.1 and up
10838
                          1
                                      2.2 and up
       Varies with device
                             Varies with device
10839
10840
       Varies with device
                            Varies with device
[10841 rows x 13 columns]
df['Price'] = df['Price'].str.replace('$','',regex=True)
df["Price"]
0
         0
1
         0
2
         0
3
         0
4
         0
10836
         0
10837
         0
```

```
10838
         0
10839
         0
10840
Name: Price, Length: 10841, dtype: object
df["Price"]=pd.to numeric(df["Price"],errors="coerce") #Changing the
Price Column to Numeric
df["Price"]
0
         0.0
1
         0.0
2
         0.0
3
         0.0
         0.0
10836
         0.0
10837
         0.0
10838
         0.0
10839
         0.0
10840
         0.0
Name: Price, Length: 10841, dtype: float64
df["Price"].max() # Maximum Price
400.0
df["Price"].min() # Minimum Price
0.0
# Creating a Box Plot of Price Column
import numpy as np
import plotly.express as px
y=df["Price"]
#create a box plot
fig = px.box(df, y="Price")
fig.show()
```



From the above mentioned box plot, it is clear that 400 is the maximum value of Price and 0 is the minimum value of Price. 200 is not the highest price. The values which are in the range of 300 to 400 comes under higher price range

```
#Q6)1-2 Drop these as most seem to be junk apps
```

```
df["Price"] # Data Frame of Price Column
0
         0.0
1
         0.0
2
         0.0
3
         0.0
         0.0
10836
         0.0
10837
         0.0
10838
         0.0
10839
         0.0
10840
         0.0
Name: Price, Length: 10841, dtype: float64
df[9719:9720]
```

The Ep Cook Book App which is at row number 9719 in the data frame has price of 200\$. The rating is Nan(missing) value and the price is 200 which can be considerd high and suspicious. Although, it is not the highest one. The highest price is 400 \$. This row can be dropped by drop method

```
9720
                Dr.Slender Ep 1 Guide (Eng)
                                                  GAME
                                                           4.0
                                                                    454
27M
                Type Price Content Rating
      Installs
                                                 Genres
                                                             Last
Updated \
9718
      500,000+
                         0.0
                                 Mature 17+ Adventure
                Free
                                                          January 2,
2018
9720
       10,000+
                Free
                         0.0
                                       Teen
                                                 Arcade December 2,
2014
     Current Ver Android Ver
9718
           1.3.7 4.3 and up
9720
                  2.3 and up
               1
Here, I use the drop method for removing the record at 9719 row number. The Ep Cook
Book App is removed now and is not visible in the updated data frame
#06) 2-Reviews: Very few apps have very high number of reviews. These
are all star apps that
#don't help with the analysis and,
#in fact, will skew it. Drop records having more than 2 million
reviews.
import pandas as pd
df=pd.read csv("C:/Users/shiva/Desktop/Shivani/SimpliLearn/Python/
googleplay/googleplaystore.csv")
df # The original data frame
                                                       App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART AND DESIGN
                                      Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                    Sketch - Draw & Paint
ART AND_DESIGN
                   Pixel Draw - Number Art Coloring Book
ART AND DESIGN
. . .
10836
                                          Sva9a Maroc - FR
FAMILY
                         Fr. Mike Schmitz Audio Teachings
10837
FAMILY
                                   Parkinson Exercices FR
10838
MEDICAL
10839
                            The SCP Foundation DB fr nn5n
```

0 1 2 3 4 10836 10837 10838 10839 10840	3.9 9 4.7 875 4.5 2156 4.3 9 4.5 5.0 NaN	59 67 610 644 667 38 4 3	8 2 3 9 with dev	25M 50 .8M 53M .6M .5M ice	Installs 10,000+ 500,000+ ,000,000+ 100,000+ 5,000+ 100+ 1,000+ 1,000+	Type Free Free Free Free Free Free Free Fr	Price 0 0 0 0 0 0 0	\
0 1 2 3 4 10836 10837 10838	Content Ratin Everyon Everyon Tee Everyon Everyon Everyon Everyon	de Art & D de Art & D de Art & de Art & de de	esign;Pre Art Art Design;C	& Desi & Desi	gn Januar ay Januar gn Augu gn Jur ty Jur on Jul on Ju	ast Upo ary 7, ry 15, ust 1, une 8, ne 20, uly 6,	2018 2018 2018 2018 2018	\
10839 10840	Mature 17 Everyon	' +	Books &		ce Januar	y 20, ry 19, .y 25,	2015	
0 1 2 3 4 	Curre Varies with	ent Ver 1.0.0 2.0.0 1.2.4 device 1.1 	4.0.3 4.0.3 4.0.3 4.2 4.4	oid Ver and up and up and up and up and up and up				
10837 10838 10839 10840	Varies with Varies with	1 1 device Va	4.1	and up and up device				
[10841	rows x 13 co	lumns]						
df["Re	views"] # Dat	a Frame of	Reviews	Column				
0 1 2	159 967 87510							

```
3
         215644
            967
10836
             38
10837
              4
10838
              3
10839
            114
10840
         398307
Name: Reviews, Length: 10841, dtype: object
df["App"]
0
            Photo Editor & Candy Camera & Grid & ScrapBook
1
                                        Coloring book moana
2
         U Launcher Lite - FREE Live Cool Themes, Hide ...
                                      Sketch - Draw & Paint
3
4
                      Pixel Draw - Number Art Coloring Book
10836
                                           Sya9a Maroc - FR
10837
                           Fr. Mike Schmitz Audio Teachings
                                     Parkinson Exercices FR
10838
10839
                              The SCP Foundation DB fr nn5n
             iHoroscope - 2018 Daily Horoscope & Astrology
10840
Name: App, Length: 10841, dtype: object
df["Reviews"]=pd.to numeric(df["Reviews"],errors="coerce") #
Converting Reviews to Numeric
df["Reviews"].max()
78158306.0
Here, The maximum value of Reviews column is 78158306
df["Reviews"].replace(to replace=["3.0M"],value="3000000") #
Replacing the 3.0 M to 3000000
            159.0
0
1
            967.0
2
          87510.0
3
         215644.0
            967.0
10836
             38.0
              4.0
10837
10838
              3.0
10839
            114.0
10840
         398307.0
Name: Reviews, Length: 10841, dtype: float64
df["Reviews"].max()
```

78158306.0

df[df.Reviews>2000000] # Finding the Records of Reviews Column which are greater than 2000000

Datina	,	,	Арр	Category
Rating 139	\	Wattpad 🕮 Free Bo	oks B00KS	S_AND_REFERENCE
4.6 335	Messenger — Text a	nd Video Chat for F	ree	COMMUNICATION
4.0 336		WhatsApp Messen	ger	COMMUNICATION
4.4 338	Google	Chrome: Fast & Sec	ure	COMMUNICATION
4.3 340		Gma	ail	COMMUNICATION
4.3				
9166	Modern	Combat 5: eSports	FPS	GAME
4.3 9841		Google Ea	rth TR/	AVEL_AND_LOCAL
4.3 10186		Farm Heroes Sa	aga	FAMILY
4.4 10190		Fallout Shel	ter	FAMILY
4.6 10327		Garena Free F	ire	GAME
4.5				
139	Reviews 2914724.0 Varies	Size with device 100	Installs ,000,000+	Type Price \ Free 0
335 336			,000,000+ ,000,000+	Free 0 Free 0
338 340	9642995.0 Varies	with device 1,000	,000,000+ ,000,000+	Free 0 Free 0
9166 9841		with device 100	,000,000+ ,000,000+	Free 0 Free 0
10186 10190	7615646.0 2721923.0		,000,000+ ,000,000+	Free 0 Free 0
10327	5534114.0		,000,000+	Free 0
	Content Rating	Genres La	ast Update	d
Current 139		ks & Reference Aug	ust 1, 2018	8 Varies with
device 335	Everyone	Communication Aug	ust 1, 2018	8 Varies with
device 336	Everyone	Communication Augu	ust 3, 2018	8 Varies with

```
device
            Everyone
                          Communication
                                          August 1, 2018 Varies with
338
device
340
            Everyone
                          Communication
                                          August 2, 2018 Varies with
device
. . .
                                     . . .
. . .
9166
          Mature 17+
                                  Action
                                           July 24, 2018
3.2.1c
                         Travel & Local
                                          June 18, 2018
9841
            Everyone
9.2.17.13
                                  Casual August 7, 2018
10186
            Everyone
5.2.6
10190
                             Simulation
                                           June 11, 2018
                Teen
1.13.12
10327
                Teen
                                  Action August 3, 2018
1.21.0
              Android Ver
       Varies with device
139
335
       Varies with device
336
       Varies with device
338
       Varies with device
340
       Varies with device
9166
               4.0 and up
               4.1 and up
9841
10186
               2.3 and up
               4.1 and up
10190
10327
             4.0.3 and up
```

There are total 453 rows in the given data frame which have reviews greater than 2.0 M.Hence,I have to drop these rows as these are not required in our data

df.drop(df[df["Reviews"] >= 2000000].index, inplace = True) # Drop
the rows in the data frame where reviews>2000000

df["Reviews"]

[453 rows x 13 columns]

```
0
             159.0
1
             967.0
2
           87510.0
3
          215644.0
             967.0
              38.0
10836
10837
               4.0
10838
               3.0
             114.0
10839
```

10840 398307.0 Name: Reviews, Length: 10388, dtype: float64

df

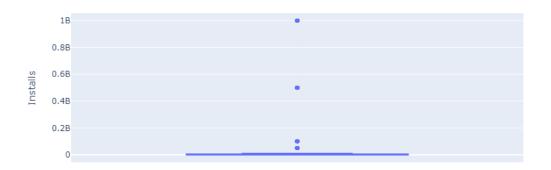
Calaaaa				Арр			
Categor	Phot		Candy Camera & Grid	& ScrapBook			
ART_AND							
2		her Lite –	- FREE Live Cool Them	es, Hide			
3	_DESIGN		Sketch - I	Oraw & Paint			
4	_DESIGN	Pixel	. Draw - Number Art Co	oloring Book			
ART_AND	_DESIGN						
 10836			Sya9a	a Maroc - FR			
FAMILY 10837			Fr. Mike Schmitz Aud	io Teachings			
FAMILY 10838		Parkinson Exercices FR					
MEDICAL 10839	-						
The SCP Foundation DB fr nn5n B00KS_AND_REFERENCE 10840 iHoroscope - 2018 Daily Horoscope & Astrology LIFESTYLE							
5 .	•	Reviews	Size	Installs	Туре		
Price 0	4.1	159.0	19M	10,000+	Free	0	
1	3.9	967.0	14M	500,000+	Free	0	
2	4.7	87510.0	8.7M	5,000,000+	Free	0	
3	4.5	215644.0	25M	50,000,000+	Free	0	
4	4.3	967.0	2.8M	100,000+	Free	0	
10836	4.5	38.0	53M	5,000+	Free	0	
10837	5.0	4.0	3.6M	100+	Free	0	
10838	NaN	3.0	9.5M	1,000+	Free	0	

```
10839
          4.5
                  114.0 Varies with device
                                                   1,000+
                                                           Free
                                                                    0
10840
          4.5 398307.0
                                         19M 10,000,000+
                                                           Free
                                                                    0
      Content Rating
                                          Genres
                                                      Last Updated
0
            Everyone
                                   Art & Design
                                                   January 7, 2018
1
            Everyone Art & Design; Pretend Play
                                                  January 15, 2018
2
                                                    August 1, 2018
            Evervone
                                   Art & Design
3
                                   Art & Design
                                                      June 8, 2018
                Teen
4
            Everyone
                     Art & Design; Creativity
                                                     June 20, 2018
10836
            Everyone
                                       Education
                                                     July 25, 2017
                                                      July 6, 2018
10837
            Everyone
                                       Education
                                                  January 20, 2017
            Everyone
10838
                                         Medical
                                                  January 19, 2015
10839
          Mature 17+
                              Books & Reference
                                                     July 25, 2018
10840
            Everyone
                                      Lifestyle
              Current Ver
                                  Android Ver
0
                    1.0.0
                                 4.0.3 and up
1
                    2.0.0
                                 4.0.3 and up
2
                    1.2.4
                                 4.0.3 and up
3
       Varies with device
                                   4.2 and up
4
                      1.1
                                   4.4 and up
. . .
                      . . .
                                   4.1 and up
10836
                     1.48
10837
                        1
                                   4.1 and up
10838
                        1
                                   2.2 and up
10839
       Varies with device Varies with device
10840
      Varies with device Varies with device
[10388 rows x 13 columns]
#06)3-Installs: There seems to be some outliers in this field too.
#Apps having very high number of installs should be dropped from the
analysis.
df # The data frame
                                                      App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART AND DESIGN
                                     Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                   Sketch - Draw & Paint
ART AND DESIGN
                   Pixel Draw - Number Art Coloring Book
```

ART_AN	D_DESIGN			
10836			Sya9a Maroc - FR	
FAMILY 10837			Fr. Mike Schmitz Audio Teachings	
FAMILY 10838			Parkinson Exercices FR	
MEDICA 10839		DENCE	The SCP Foundation DB fr nn5n	
10840 LIFEST			2018 Daily Horoscope & Astrology	
Price	Rating	Reviews	Size Installs Type	
0	4.1	159.0	19M 10,000+ Free	0
1	3.9	967.0	14M 500,000+ Free	0
2	4.7	87510.0	8.7M 5,000,000+ Free	0
3	4.5	215644.0	25M 50,000,000+ Free	0
4	4.3	967.0	2.8M 100,000+ Free	0
10836	4.5	38.0	53M 5,000+ Free	0
10837	5.0	4.0	3.6M 100+ Free	0
10838	NaN	3.0	9.5M 1,000+ Free	0
10839	4.5	114.0	Varies with device 1,000+ Free	0
10840	4.5	398307.0	19M 10,000,000+ Free	0
0 1 2 3 4	Ev Ev	eryone eryone Ai eryone Teen	Genres Last Updated Art & Design January 7, 2018 rt & Design;Pretend Play January 15, 2018 Art & Design August 1, 2018 Art & Design June 8, 2018 Art & Design;Creativity June 20, 2018	\
10836 10837 10838	Ev Ev	eryone eryone eryone	Education July 25, 2017 Education July 6, 2018 Medical January 20, 2017	

```
10839
          Mature 17+
                               Books & Reference
                                                   January 19, 2015
10840
            Everyone
                                        Lifestyle
                                                       July 25, 2018
              Current Ver
                                   Android Ver
0
                     1.0.0
                                  4.0.3 and up
1
                     2.0.0
                                  4.0.3 and up
2
                     1.2.4
                                  4.0.3 and up
3
       Varies with device
                                     4.2 and up
4
                                     4.4 and up
. . .
                       . . .
10836
                      1.48
                                     4.1 and up
10837
                         1
                                     4.1 and up
10838
                         1
                                     2.2 and up
10839
       Varies with device
                           Varies with device
10840
       Varies with device Varies with device
[10388 rows x 13 columns]
df["Installs"]
0
             10,000+
1
            500,000+
2
          5,000,000+
3
         50,000,000+
4
            100,000+
10836
              5,000+
10837
                100+
10838
              1,000+
10839
              1,000+
         10,000,000+
10840
Name: Installs, Length: 10841, dtype: object
# Finding the Outlier through BoxPlot Method of Installs Column
df["Installs"]
0
             10,000+
1
            500,000+
2
          5,000,000+
3
         50,000,000+
4
            100,000+
10836
              5,000+
10837
                100+
10838
              1,000+
10839
              1,000+
10840
         10,000,000+
Name: Installs, Length: 10841, dtype: object
df['Installs'] = df['Installs'].str.replace('+','',regex=True)
```

```
df['Installs'] = df['Installs'].str.replace(',',','',regex=True)
df["Installs"]
0
            10000
1
           500000
2
          5000000
3
         50000000
           100000
10836
             5000
10837
              100
10838
             1000
10839
             1000
10840
         10000000
Name: Installs, Length: 10841, dtype: object
df["Installs"]=pd.to_numeric(df["Installs"],errors="coerce")
df["Installs"]
0
            10000.0
1
           500000.0
2
          5000000.0
3
         50000000.0
           100000.0
10836
             5000.0
10837
              100.0
10838
             1000.0
10839
             1000.0
10840
         10000000.0
Name: Installs, Length: 10841, dtype: float64
df["Installs"].max()
1000000000.0
import numpy as np
import plotly.express as px
y=df["Installs"]
#create a box plot
fig = px.box(df, y="Installs")
fig.show()
```



The above mentioned is the box plot of Installs column

```
df["Installs"].min()
0.0
df["Installs"].max()
1000000000.0
df["Installs"].median()
100000.0
df[df.Installs>100000.0]
                                                      App
Category \
                                      Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
ART_AND_DESIGN
                                    Sketch - Draw & Paint
ART_AND_DESIGN
                                         Infinite Painter
ART AND DESIGN
                                     Garden Coloring Book
ART_AND_DESIGN
                             Fatal Raid - No.1 Mobile FPS
10803
GAME
                   Castle Clash: RPG War and Strategy FR
10809
FAMILY
10815
                               Golden Dictionary (FR-AR)
BOOKS AND REFERENCE
10826
               Frim: get new friends on local chat rooms
```

1 2 3 7 8	Rating Reviews 3.9 967 4.7 87510 4.5 215644 4.1 36815 4.4 13791	Size 14M 8.7M 25M 29M 33M	Installs 500000.0 5000000.0 50000000.0 1000000.0	Type Price \ Free 0 Free 0 Free 0 Free 0 Free 0	
10803 10809 10815 10826 10840	4.3 56496 4.7 376223 4.2 5775 4.0 88486 Va 4.5 398307	 81M 24M 4.9M ries with device 19M	1000000.0 1000000.0 500000.0 5000000.0 10000000.0	Free 0 Free 0 Free 0 Free 0 Free 0	
1 2 3 7 8	Content Rating Everyone Art Everyone Teen Everyone Everyone	& Design;Pretend Art & D Art & D Art & D Art & D	Design A Design Design	Last Updated uary 15, 2018 ugust 1, 2018 June 8, 2018 June 14, 2018 mber 20, 2017	\
10803 10809 10815 10826 10840	Teen Everyone Everyone Mature 17+ Everyone	Str Books & Refe	rategy erence Social M	ugust 7, 2018 July 18, 2018 July 19, 2018 arch 23, 2018 July 25, 2018	
1 2 3 7 8	Current Ver 2.0.6 1.2.4 Varies with device 6.1.61.1 2.9.2	4.0.3 and 4.0.3 and 4.2 and	d up d up d up d up		
10803 10809 10815 10826 10840	1.5.447 1.4.2 7.0.4.6 Varies with device Varies with device	4.2 and Varies with dev	d up d up vice		

[4950 rows x 13 columns]

In the Installs column of the data frame, the minimum value is 0 and the maximum value is 100000000.0. The median value of the data frame is 100000.0. Here, I have to drop those columns which are greater than the median value ie 100000.0

is >100000.0. # The below mentioned data frame has installs column which is >100000.0(higher number of installs) App Category \ Coloring book moana ART AND DESIGN U Launcher Lite - FREE Live Cool Themes, Hide ... ART AND DESIGN Sketch - Draw & Paint ART AND DESIGN Infinite Painter ART AND DESIGN Garden Coloring Book ART AND DESIGN . . . 10803 Fatal Raid - No.1 Mobile FPS GAME Castle Clash: RPG War and Strategy FR 10809 FAMILY 10815 Golden Dictionary (FR-AR) BOOKS AND REFERENCE Frim: get new friends on local chat rooms 10826 SOCIAL 10840 iHoroscope - 2018 Daily Horoscope & Astrology LIFESTYLE Rating Reviews Size Installs Type Price \ 3.9 500000.0 Free 1 967 14M 0 2 4.7 87510 8.7M 5000000.0 Free 0 3 4.5 215644 25M 50000000.0 Free 7 4.1 36815 29M 1000000.0 Free 0 8 4.4 13791 33M 1000000.0 Free 0 4.3 56496 81M 1000000.0 0 10803 Free 4.7 376223 24M 10809 1000000.0 Free 0 4.2 10815 5775 4.9M 500000.0 Free 0 10826 4.0 88486 Varies with device 5000000.0 Free 0 10840 4.5 398307 19M 10000000.0 Free 0 Content Rating Last Updated \ Genres 1 Everyone Art & Design; Pretend Play January 15, 2018 2 Everyone Art & Design August 1, 2018 3 Art & Design June 8, 2018 Teen 7 June 14, 2018 Everyone Art & Design

8

Everyone

Art & Design September 20, 2017

df[df.Installs>100000.0] # The installs column of the data frame which

```
10803
                                                       August 7, 2018
                Teen
                                          Action
                                                        July 18, 2018
10809
            Everyone
                                        Strategy
            Everyone
                               Books & Reference
10815
                                                       July 19, 2018
                                                      March 23, 2018
10826
          Mature 17+
                                          Social
                                                        July 25, 2018
10840
            Everyone
                                       Lifestyle
              Current Ver
                                   Android Ver
1
                    2.0.0
                                  4.0.3 and up
2
                    1.2.4
                                  4.0.3 and up
3
                                    4.2 and up
       Varies with device
7
                 6.1.61.1
                                    4.2 and up
8
                    2.9.2
                                    3.0 and up
                  1.5.447
10803
                                    4.0 and up
10809
                    1.4.2
                                    4.1 and up
                  7.0.4.6
10815
                                    4.2 and up
       Varies with device Varies with device
10826
       Varies with device Varies with device
10840
[4950 rows x 13 columns]
df.drop(df[df['Installs']>100000.0].index, inplace = True) # Dropping
the column which has higher number of installs
df["Installs"]
0
          10000.0
4
         100000.0
5
          50000.0
6
          50000.0
9
          10000.0
10835
             10.0
10836
           5000.0
10837
            100.0
10838
           1000.0
10839
           1000.0
Name: Installs, Length: 5891, dtype: float64
df # The revised data frame after deleting higher number of installs
                                                   App
Category \
       Photo Editor & Candy Camera & Grid & ScrapBook
ART AND DESIGN
                Pixel Draw - Number Art Coloring Book
ART AND DESIGN
                            Paper flowers instructions
ART_AND_DESIGN
              Smoke Effect Photo Maker - Smoke Editor
ART AND DESIGN
```

```
Kids Paint Free - Drawing Fun
9
ART AND DESIGN
. . .
10835
                                                FR Forms
BUSINESS
                                       Sya9a Maroc - FR
10836
FAMILY
10837
                      Fr. Mike Schmitz Audio Teachings
FAMILY
10838
                                 Parkinson Exercices FR
MEDICAL
                         The SCP Foundation DB fr nn5n
10839
BOOKS AND REFERENCE
                                                         Type Price
       Rating Reviews
                                       Size
                                              Installs
0
          4.1
                   159
                                        19M
                                               10000.0
                                                         Free
                                                                  0
4
          4.3
                   967
                                       2.8M
                                              100000.0
                                                         Free
                                                                  0
5
          4.4
                                       5.6M
                   167
                                               50000.0
                                                         Free
                                                                  0
6
          3.8
                   178
                                        19M
                                               50000.0
                                                                  0
                                                         Free
9
                                                                  0
          4.7
                   121
                                       3.1M
                                               10000.0
                                                         Free
                                        . . .
                                                   . . .
                                                         . . .
          . . .
                   . . .
. . .
                                                                 . . .
10835
          NaN
                    0
                                       9.6M
                                                  10.0
                                                         Free
                                                                  0
                    38
          4.5
                                                5000.0
                                                                  0
10836
                                         53M
                                                         Free
          5.0
                    4
                                        3.6M
                                                                  0
10837
                                                100.0
                                                         Free
                    3
                                       9.5M
                                                1000.0
10838
          NaN
                                                         Free
                                                                  0
10839
          4.5
                   114 Varies with device
                                                1000.0
                                                         Free
                                                                  0
      Content Rating
                                                         Last Updated
                                          Genres
0
            Everyone
                                   Art & Design
                                                      January 7, 2018
                                                        June 20, 2018
4
            Everyone Art & Design; Creativity
5
            Everyone
                                   Art & Design
                                                      March 26, 2017
                                                      April 26, 2018
                                   Art & Design
6
            Everyone
9
            Everyone Art & Design; Creativity
                                                         July 3, 2018
. . .
10835
            Everyone
                                       Business
                                                  September 29, 2016
                                                        July 25, 2017
10836
            Everyone
                                      Education
                                                         July 6, 2018
10837
            Everyone
                                      Education
10838
            Everyone
                                         Medical
                                                     January 20, 2017
          Mature 17+
                            Books & Reference
                                                    January 19, 2015
10839
               Current Ver
                                    Android Ver
0
                     1.0.0
                                   4.0.3 and up
4
                                     4.4 and up
                       1.1
5
                         1
                                     2.3 and up
6
                       1.1
                                   4.0.3 and up
9
                                   4.0.3 and up
                       2.8
                        . . .
10835
                     1.1.5
                                     4.0 and up
                      1.48
                                     4.1 and up
10836
```

```
10837
                         1
                                     4.1 and up
10838
                         1
                                     2.2 and up
10839 Varies with device Varies with device
[5891 rows x 13 columns]
#Q6)3-1-Find out the different percentiles - 10, 25, 50, 70, 90, 95,
df # The revised data frame
                                                     App
Category \
       Photo Editor & Candy Camera & Grid & ScrapBook
ART AND DESIGN
                Pixel Draw - Number Art Coloring Book
ART AND DESIGN
                            Paper flowers instructions
ART_AND_DESIGN
              Smoke Effect Photo Maker - Smoke Editor
ART AND DESIGN
                         Kids Paint Free - Drawing Fun
ART AND DESIGN
. . .
10835
                                               FR Forms
BUSINESS
10836
                                       Sya9a Maroc - FR
FAMILY
10837
                      Fr. Mike Schmitz Audio Teachings
FAMILY
10838
                                 Parkinson Exercices FR
MEDICAL
10839
                         The SCP Foundation DB fr nn5n
BOOKS AND REFERENCE
       Rating Reviews
                                       Size
                                             Installs
                                                        Type Price
          4.1
                                        19M
                                              10000.0
                                                        Free
0
                   159
                                                                 0
4
          4.3
                   967
                                       2.8M
                                             100000.0
                                                                 0
                                                        Free
5
          4.4
                                       5.6M
                                              50000.0
                   167
                                                        Free
                                                                 0
                                       19M
                                              50000.0
6
          3.8
                   178
                                                        Free
                                                                 0
9
          4.7
                                       3.1M
                                              10000.0
                   121
                                                        Free
                                                                 0
                   . . .
                                        . . .
10835
                    0
                                                  10.0
                                                                 0
          NaN
                                       9.6M
                                                        Free
                    38
          4.5
                                        53M
                                               5000.0
10836
                                                        Free
                                                                 0
10837
          5.0
                     4
                                       3.6M
                                                100.0
                                                        Free
                                                                 0
                     3
10838
          NaN
                                       9.5M
                                               1000.0
                                                                 0
                                                        Free
                   114 Varies with device
10839
                                               1000.0
          4.5
                                                        Free
                                                                 0
```

Genres

Last Updated \

Content Rating

```
0
            Everyone
                                  Art & Design
                                                    January 7, 2018
4
            Everyone Art & Design; Creativity
                                                      June 20, 2018
5
                                                     March 26, 2017
            Everyone
                                  Art & Design
6
            Evervone
                                  Art & Design
                                                     April 26, 2018
9
            Everyone Art & Design; Creativity
                                                       July 3, 2018
. . .
                                                 September 29, 2016
                                      Business
10835
            Evervone
                                                      July 25, 2017
10836
            Everyone
                                     Education
10837
            Everyone
                                     Education
                                                       July 6, 2018
10838
                                       Medical
                                                   January 20, 2017
            Everyone
10839
          Mature 17+
                             Books & Reference
                                                   January 19, 2015
              Current Ver
                                   Android Ver
0
                     1.0.0
                                  4.0.3 and up
4
                       1.1
                                    4.4 and up
5
                                    2.3 and up
                        1
6
                       1.1
                                  4.0.3 and up
9
                                  4.0.3 and up
                       2.8
. . .
                    1.1.5
                                    4.0 and up
10835
10836
                      1.48
                                    4.1 and up
10837
                         1
                                    4.1 and up
                         1
10838
                                    2.2 and up
10839
       Varies with device Varies with device
[5891 rows x 13 columns]
q1=df["Installs"].quantile(0.10)
                                  #10th Percentile
q2=df["Installs"].quantile(0.25)
                                  # 25th Percentile
q3=df["Installs"].quantile(0.50) # 50th Percentile
q4=df["Installs"].quantile(0.70) # 70th Percentile
q5=df["Installs"].quantile(0.90) # 90th Percentile
q6=df["Installs"].quantile(0.95) # 95th Percentile
q7=df["Installs"].quantile(0.99) # 99th Percentile
print(q1,q2,q3,q4,q5,q6,q7)
50.0 100.0 5000.0 10000.0 100000.0 100000.0 100000.0
#Q6)2-Decide a threshold as cutoff for outlier and drop records having
values more than that
df["Installs"].describe()
           5890.000000
count
          26304.496944
mean
std
          38941.624628
min
              0.00000
25%
            100.000000
50%
           5000.000000
```

75%

50000.000000

max 100000.000000

Name: Installs, dtype: float64

In this dataframe, the mean value of Installs column is 26304 and the maximum value of Install is 100000 The mean value is 26304. The mean value is sensitive to max which clearly indicates that Install column has outliers. Here, I am considering the mean 26304 as the threshold and will drop values which are greater than the threshold

df[df.Installs>26304]

```
App
Category \
                  Pixel Draw - Number Art Coloring Book
ART_AND_DESIGN
                              Paper flowers instructions
ART AND DESIGN
                Smoke Effect Photo Maker - Smoke Editor
ART AND DESIGN
                                   Mandala Coloring Book
13
ART AND DESIGN
       3D Color Pixel by Number - Sandbox Art Coloring
ART AND_DESIGN
. . .
. . .
10808
                                      lesparticuliers.fr
LIFESTYLE
10814
                                    FR: My Secret Pets!
FAMILY
10817
                                    HTC Sense Input - FR
T00LS
10830
                                       News Minecraft.fr
NEWS AND MAGAZINES
10832
                                                 FR Tides
WEATHER
       Rating Reviews
                        Size
                               Installs
                                          Type Price Content Rating
4
          4.3
                   967
                        2.8M
                               100000.0
                                          Free
                                                   0
                                                            Everyone
5
          4.4
                   167
                        5.6M
                                50000.0
                                          Free
                                                   0
                                                            Everyone
6
          3.8
                   178
                         19M
                                50000.0
                                         Free
                                                   0
                                                            Everyone
13
          4.6
                  4326
                         21M
                               100000.0
                                          Free
                                                   0
                                                            Everyone
14
          4.4
                  1518
                         37M
                               100000.0
                                          Free
                                                   0
                                                            Everyone
                                           . . .
                          . . .
                                          Free
10808
          NaN
                    96
                        1.0M
                                50000.0
                                                   0
                                                            Everyone
10814
          4.0
                   785
                         31M
                                50000.0
                                          Free
                                                   0
                                                                Teen
10817
          4.0
                   885
                        8.0M
                               100000.0
                                          Free
                                                   0
                                                            Everyone
10830
                        2.3M
                               100000.0
                                                   0
          3.8
                   881
                                          Free
                                                            Everyone
10832
          3.8
                  1195
                               100000.0
                                                   0
                        582k
                                         Free
                                                            Everyone
```

Genres Last Updated Current Ver

Android Ver

4 <i>A</i>	Art & Design;Creativity	June 20,	2018	1.1	4.4
and up 5	Art & Design	March 26,	2017	1	2.3
and up 6	Art & Design	April 26,	2018	1.1	4.0.3
and up 13	Art & Design	June 26,	2018	1.0.4	4.4
and up 14	Art & Design	August 3,	2018	1.2.3	2.3
and up					
• • •	• • •				
10808	Lifestyle	November 25,	2014	1.5	2.3
and up 10814	Entertainment	June 3,	2015	1.3.1	3.0
and up 10817	Tools	October 30,	2015	1.0.612928	5.0
and up 10830	News & Magazines	January 20,	2014	1.5	1.6
and up 10832	Weather	February 16,	2014	6	2.1
and up					

[1648 rows x 13 columns]

Hence, there are 1648 rows and 13 columns in the dataframe where the installation is greater than the 26304(the threshold value).

df.drop(df[df['Installs'] > 26304].index, inplace = True) # Drop the
rows which are greater than the threshold

df["Installs"] # Refined data frame of Installs columns

```
0
          10000.0
9
          10000.0
15
          5000.0
17
         10000.0
25
         10000.0
10835
            10.0
10836
          5000.0
10837
           100.0
10838
          1000.0
10839
          1000.0
```

Name: Installs, Length: 4243, dtype: float64

df # The final dataframe after threshold amount deletion

```
App Category \
0 Photo Editor & Candy Camera & Grid & ScrapBook
```

```
ART_AND_DESIGN
                         Kids Paint Free - Drawing Fun
ART_AND_DESIGN
                       Learn To Draw Kawaii Characters
ART AND DESIGN
17
                               350 Diy Room Decor Ideas
ART AND_DESIGN
                            Harley Quinn wallpapers HD
ART AND DESIGN
10835
                                                FR Forms
BUSINESS
                                       Sya9a Maroc - FR
10836
FAMILY
10837
                      Fr. Mike Schmitz Audio Teachings
FAMILY
10838
                                 Parkinson Exercices FR
MEDICAL
                         The SCP Foundation DB fr nn5n
10839
BOOKS AND REFERENCE
                                                         Type Price
       Rating Reviews
                                       Size
                                              Installs
0
          4.1
                                        19M
                                               10000.0
                                                         Free
                   159
9
          4.7
                   121
                                       3.1M
                                               10000.0
                                                         Free
                                                                  0
15
          3.2
                    55
                                       2.7M
                                                5000.0
                                                         Free
                                                                  0
                    27
17
          4.5
                                        17M
                                               10000.0
                                                         Free
                                                                  0
25
          4.8
                   192
                                       6.0M
                                               10000.0
                                                         Free
                                                                  0
. . .
          . . .
                   . . .
                                         . . .
                                                   . . .
                                                         . . .
                                                                 . . .
10835
          NaN
                    0
                                       9.6M
                                                                  0
                                                  10.0
                                                         Free
          4.5
                    38
10836
                                        53M
                                                5000.0
                                                         Free
                                                                  0
          5.0
                     4
                                       3.6M
                                                                  0
10837
                                                 100.0
                                                         Free
                     3
                                       9.5M
                                                1000.0
10838
          NaN
                                                         Free
                                                                  0
10839
          4.5
                   114
                        Varies with device
                                                1000.0
                                                         Free
      Content Rating
                                          Genres
                                                         Last Updated
0
            Everyone
                                   Art & Design
                                                     January 7, 2018
                                                         July 3, 2018
9
            Everyone Art & Design; Creativity
15
                                   Art & Design
                                                         June 6, 2018
            Everyone
                                                    November 7, 2017
17
                                   Art & Design
            Everyone
25
            Everyone
                                   Art & Design
                                                       April 25, 2018
. . .
                                                  September 29, 2016
10835
            Everyone
                                       Business
                                                        July 25, 2017
10836
            Everyone
                                      Education
                                                         July 6, 2018
                                      Education
10837
            Everyone
                                                    January 20, 2017
10838
            Everyone
                                        Medical
          Mature 17+
                              Books & Reference
                                                    January 19, 2015
10839
               Current Ver
                                    Android Ver
0
                     1.0.0
                                   4.0.3 and up
```

```
9
                                  4.0.3 and up
                      2.8
15
                      NaN
                                    4.2 and up
17
                        1
                                    2.3 and up
25
                      1.5
                                    3.0 and up
. . .
                       . . .
10835
                    1.1.5
                                    4.0 and up
                     1.48
10836
                                    4.1 and up
10837
                        1
                                    4.1 and up
10838
                        1
                                    2.2 and up
10839
      Varies with device Varies with device
[4243 rows x 13 columns]
#Q7)Bivariate analysis: Let's look at how the available predictors
relate to the variable of interest, i.e., our target variable rating.
Make scatter plots (for numeric features) and box plots (for character
features)
#to assess the relations between rating and the other features.
# 7)1-Make scatter plot/joinplot for Rating vs. Price
# 1-What pattern do you observe? Does rating increase with price?
df # The data frame
                                                   App
Category \
       Photo Editor & Candy Camera & Grid & ScrapBook
ART AND DESIGN
                        Kids Paint Free - Drawing Fun
ART_AND_DESIGN
                      Learn To Draw Kawaii Characters
ART AND DESIGN
                              350 Diy Room Decor Ideas
17
ART AND DESIGN
                           Harley Quinn wallpapers HD
ART_AND_DESIGN
. . .
                                                    . . .
10835
                                              FR Forms
BUSINESS
10836
                                      Sya9a Maroc - FR
FAMILY
10837
                     Fr. Mike Schmitz Audio Teachings
FAMILY
10838
                                Parkinson Exercices FR
MEDICAL
                        The SCP Foundation DB fr nn5n
10839
BOOKS_AND_REFERENCE
       Rating Reviews
                                      Size
                                            Installs
                                                      Type Price \
```

19M

10000.0

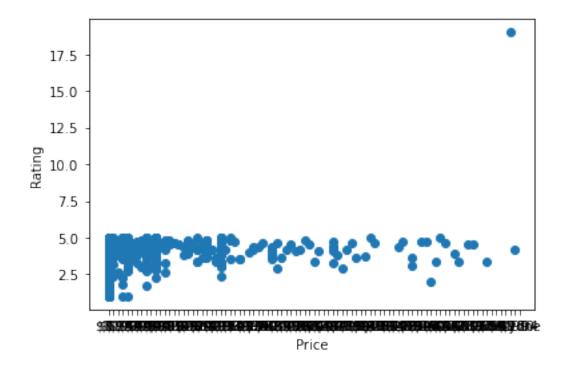
Free

0

4.1

159

```
9
          4.7
                   121
                                        3.1M
                                               10000.0
                                                         Free
                                                                   0
15
          3.2
                    55
                                        2.7M
                                                5000.0
                                                         Free
                                                                   0
17
          4.5
                    27
                                         17M
                                               10000.0
                                                         Free
                                                                   0
25
          4.8
                   192
                                        6.0M
                                               10000.0
                                                         Free
                                                                   0
. . .
           . . .
                   . . .
                                         . . .
                                                    . . .
                                                          . . .
10835
          NaN
                     0
                                        9.6M
                                                   10.0
                                                         Free
                                                                   0
          4.5
                    38
                                         53M
                                                5000.0
                                                                   0
10836
                                                         Free
10837
          5.0
                     4
                                        3.6M
                                                 100.0
                                                                   0
                                                         Free
10838
          NaN
                     3
                                        9.5M
                                                 1000.0
                                                         Free
                                                                   0
10839
          4.5
                   114
                        Varies with device
                                                1000.0
                                                                   0
                                                         Free
      Content Rating
                                          Genres
                                                         Last Updated
0
                                   Art & Design
                                                      January 7, 2018
             Everyone
9
             Everyone Art & Design; Creativity
                                                         July 3, 2018
                                                         June 6, 2018
15
             Everyone
                                   Art & Design
                                                     November 7, 2017
17
                                   Art & Design
             Everyone
25
                                                       April 25, 2018
             Everyone
                                   Art & Design
                                        Business
                                                   September 29, 2016
10835
             Everyone
10836
             Everyone
                                       Education
                                                        July 25, 2017
                                                         July 6, 2018
10837
             Everyone
                                       Education
                                                     January 20, 2017
10838
             Everyone
                                         Medical
          Mature 17+
                              Books & Reference
10839
                                                     January 19, 2015
               Current Ver
                                    Android Ver
0
                     1.0.0
                                   4.0.3 and up
9
                        2.8
                                   4.0.3 and up
15
                                      4.2 and up
                       NaN
17
                          1
                                      2.3 and up
25
                                      3.0 and up
                        1.5
                        . . .
. . .
10835
                     1.1.5
                                      4.0 and up
                      1.48
                                      4.1 and up
10836
                          1
10837
                                      4.1 and up
10838
                          1
                                      2.2 and up
10839
       Varies with device
                            Varies with device
[4243 rows x 13 columns]
import matplotlib.pyplot as plt
x=df["Price"]
y=df["Rating"]
plt.xlabel("Price")
plt.ylabel("Rating")
plt.scatter(x,y)
<matplotlib.collections.PathCollection at 0x18ee5abb820>
```



The above mentioned is the scatter plot of Rating vs Price. When the Price is increasing, the rating is increasing then at a certain point it gets decreased then increased and decreased. The Rating is not increasing in a linear fashion. It is scattered

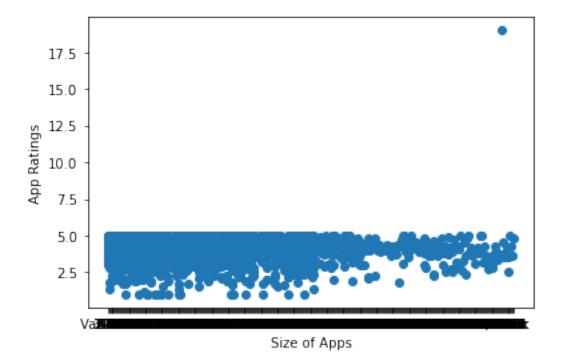
```
#Q7) 2- Make scatter plot/joinplot for Rating vs. Size

df.dropna(subset=["Rating"],axis=0,inplace=True) # Dropping the
missing values of Rating Column

df.dropna(subset=["Size"],axis=0,inplace=True) # Dropping the missing
values of size column

import matplotlib.pyplot as plt

x=df["Size"]
y=df["Rating"]
plt.xlabel("Size of Apps")
plt.ylabel("App Ratings")
plt.scatter(x,y)
<matplotlib.collections.PathCollection at 0x18eea1d2550>
```



df[["Size","Rating"]] # Clubing the Size and Rating Columns together

			Size	Rating
0			19M	4.1
9			3.1M	4.7
15			2.7M	3.2
17			17M	4.5
25			6.0M	4.8
10833			619k	4.8
10834			2.6M	4.0
10836			53M	4.5
10837			3.6M	5.0
10839	Varies	with	device	4.5

[2803 rows x 2 columns]

The heavier apps are not always rated better. From the above mentioned scatter plot, it is somewhat clear that app rating is fluctuating according to size. It is not very much clear from the scatter plot but from the above mentioned data frame it is clear that app ratings is not increasing wrt to size. For example, if the size of App is 6.0 M, the rating is 4.8, when it is 17M, the rating is 4.5. Heavier Apps are not always rated higher

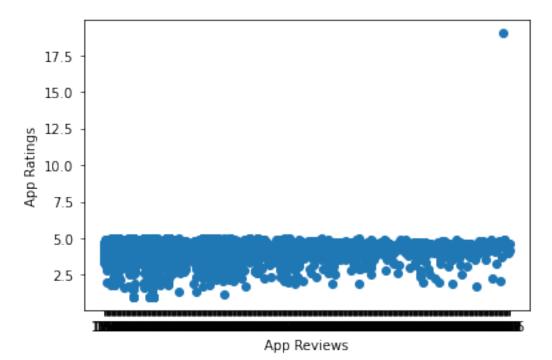
#Q7)3- Make scatter plot/joinplot for Rating vs. Reviews
df # The refined data frame

App Category \
0 Photo Editor & Candy Camera & Grid & ScrapBook

```
ART_AND_DESIGN
                         Kids Paint Free - Drawing Fun
ART_AND_DESIGN
                       Learn To Draw Kawaii Characters
15
ART AND DESIGN
17
                               350 Diy Room Decor Ideas
ART AND DESIGN
                            Harley Quinn wallpapers HD
ART AND DESIGN
. . .
10833
                                             Chemin (fr)
BOOKS AND REFERENCE
                                           FR Calculator
10834
FAMILY
10836
                                       Sya9a Maroc - FR
FAMILY
10837
                      Fr. Mike Schmitz Audio Teachings
FAMILY
                         The SCP Foundation DB fr nn5n
10839
BOOKS AND REFERENCE
                                                        Type Price
       Rating Reviews
                                       Size
                                              Installs
0
          4.1
                                        19M
                                                        Free
                   159
                                               10000.0
9
          4.7
                   121
                                       3.1M
                                               10000.0
                                                                  0
                                                        Free
15
          3.2
                                       2.7M
                    55
                                                5000.0
                                                        Free
                                                                  0
                    27
17
          4.5
                                        17M
                                               10000.0
                                                        Free
                                                                  0
25
          4.8
                   192
                                       6.0M
                                               10000.0
                                                        Free
                                                                  0
          . . .
                   . . .
                                         . . .
                                                         . . .
. . .
                                                   . . .
                                                                . . .
          4.8
                    44
                                       619k
                                                1000.0
                                                                  0
10833
                                                        Free
                    7
10834
          4.0
                                       2.6M
                                                500.0
                                                                  0
                                                        Free
          4.5
                    38
                                        53M
                                                5000.0
                                                                  0
10836
                                                        Free
          5.0
                                                 100.0
10837
                    4
                                       3.6M
                                                        Free
                                                                  0
10839
          4.5
                   114
                        Varies with device
                                                1000.0
                                                        Free
      Content Rating
                                         Genres
                                                      Last Updated
0
            Everyone
                                   Art & Design
                                                   January 7, 2018
                                                      July 3, 2018
9
            Everyone Art & Design; Creativity
15
                                   Art & Design
                                                      June 6, 2018
            Everyone
17
                                                  November 7, 2017
                                   Art & Design
            Everyone
25
            Everyone
                                   Art & Design
                                                    April 25, 2018
. . .
                                                    March 23, 2014
                              Books & Reference
10833
            Everyone
10834
            Everyone
                                      Education
                                                     June 18, 2017
                                                     July 25, 2017
                                      Education
10836
            Everyone
                                                      July 6, 2018
10837
            Everyone
                                      Education
          Mature 17+
                              Books & Reference
                                                  January 19, 2015
10839
               Current Ver
                                    Android Ver
0
                     1.0.0
                                   4.0.3 and up
```

```
9
                                  4.0.3 and up
                       2.8
15
                       NaN
                                     4.2 and up
17
                         1
                                     2.3 and up
25
                       1.5
                                     3.0 and up
                       . . .
                       0.8
10833
                                     2.2 and up
10834
                     1.0.0
                                     4.1 and up
10836
                                     4.1 and up
                      1.48
10837
                         1
                                     4.1 and up
       Varies with device Varies with device
10839
[2803 rows x 13 columns]
x=df["Reviews"]
y=df["Rating"]
import matplotlib.pyplot as plt
plt.xlabel("App Reviews")
plt.ylabel("App Ratings")
plt.scatter(x,y)
```

<matplotlib.collections.PathCollection at 0x18eeac57f70>



```
121
                      4.7
15
             55
                      3.2
17
             27
                      4.5
25
            192
                      4.8
            . . .
. . .
10833
             44
                      4.8
             7
                      4.0
10834
10836
                      4.5
             38
10837
              4
                      5.0
10839
            114
                      4.5
```

[2803 rows x 2 columns]

The above mentioned is the scatterplot of Reviews vs Ratings Column. The more review doesn't always mean a better ratings. It is not very much clear from the scatterplot, but from the above mentioned data frame it is clear that rating of app varies wrt to reviews. For example, when the rating is 4.1, the review count is 159. When the rating is 4.7 the review count is 121. When the review is 7, the rating is 4. When the review is 4, the rating is 5. Hence, one can conclude that app rating fluctuates wrt to review and thats why it is scattered

#Q7)4- Make boxplot for Rating vs. Content Rating #Q7)Is there any difference in the ratings? Are some types liked better?

import pandas as pd

df=pd.read_csv("C:/Users/shiva/Desktop/Shivani/SimpliLearn/Python/
googleplay/googleplaystore.csv")

df # The data frame

```
App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART AND DESIGN
                                      Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                    Sketch - Draw & Paint
ART AND DESIGN
                   Pixel Draw - Number Art Coloring Book
ART AND DESIGN
. . .
                                                       . . .
                                         Sya9a Maroc - FR
10836
FAMILY
10837
                         Fr. Mike Schmitz Audio Teachings
FAMILY
                                   Parkinson Exercices FR
10838
```

MEDICAL The SCP Foundation DB fr nn5n 10839 BOOKS_AND_REFERENCE iHoroscope - 2018 Daily Horoscope & Astrology 10840 LIFESTYLE

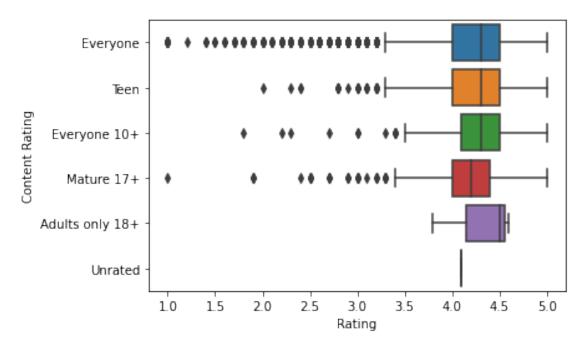
\

0 1 2 3 4	Rating H 4.1 3.9 4.7 4.5 4.3	159 967 87510 215644 967			14 8.7 25 2.8	9M 4M 7M 5M 8M	1 50, 5,00	stalls 0,000+ 0,000+ 0,000+ 0,000+ 0,000+	Free Free Free Free	Price 0 0 0 0	
10836 10837 10838 10839 10840	4.5 5.0 NaN 4.5 4.5	38 4 3 114 398307	Varies	with	53 3.6 9.5 device	5M ce		5,000+ 100+ 1,000+ 1,000+ 0,000+	Free Free Free Free Free	 0 0 0 0	
0 1 2 3 4	Eve Eve	eryone	Art & D Art &		Art 8	S De end S De S De	Play sign sign	Janua Januar Augu Ju	st Upo ry 7, y 15, st 1, ne 8, e 20,	2018 2018 2018 2018	\
10836 10837 10838 10839 10840	Evo Evo Matu	eryone eryone eryone re 17+ eryone		Books	Ed s & Re	duca Med efer	tion tion ical ence tyle	Ju Januar Januar		2018 2017 2015	
0 1 2 3 4 10836 10837 10838 10839 10840	Varies v	2. 1. with dev	0.0 0.0 2.4 ice 1.1 .48 1 1 ice Va	4 4 4 ries v	Andros .0.3 a .0.3 a .0.3 a 4.2 a 4.1 a 4.1 a 2.2 a with a	and and and and and and and and devi	up up up up up up up ce				
[10841	rows x	13 colum	ns]								

import seaborn as sns

x=df["Rating"] # Placing the rating on x -axis

y=df["Content Rating"] # Placing the Content Rating on Y axis
sns.boxplot(x="Rating",y="Content Rating",data=df)
<AxesSubplot:xlabel='Rating', ylabel='Content Rating'>



df[["Rating","Content Rating"]] # Clubing the Rating and Content
Rating together

	Rating	Content Rating
0	4.1	Everyone
1	3.9	Everyone
2	4.7	Everyone
3	4.5	Teen
4	4.3	Everyone
10836	4.5	Everyone
10837	5.0	Everyone
10838	NaN	Everyone
10839	4.5	Mature 17+
10840	4.5	Everyone

[10841 rows x 2 columns]

In the App rating Prediction, the ratings are basically in the range of 1.1 to 5. These ratings are visible in the Box Plot. From the Box Plot one can conclude that when rating is within the range of 3.0 to 3.5, It is liked by 17+, Everyone (10+), the Teenagers and viewers of all age groups including senior citizens. The distribution of data points are visible inside the Box Plot.

```
#Make boxplot for Ratings vs. Category
#Which genre has the best ratings?
```

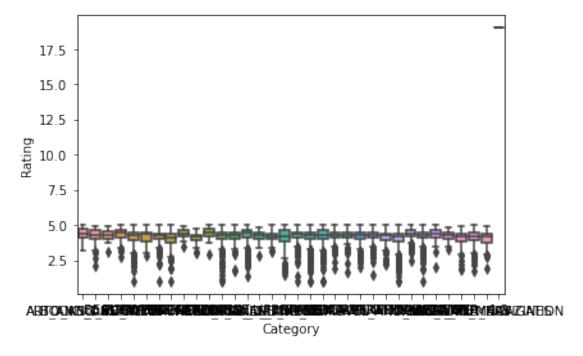
import seaborn as sns

x=df["Category"]

y=df["Rating"]

sns.boxplot(x="Category",y="Rating",data=df) # Box Plot of Rating vs
Category

<AxesSubplot:xlabel='Category', ylabel='Rating'>



df[["Category","Rating"]] # Clubing the Category and Rating Columns
together

	Category	Rating
0	ART_AND_DESIGN	4.1
1	ART_AND_DESIGN	3.9
2	ART_AND_DESIGN	4.7
3	ART_AND_DESIGN	4.5
4	ART_AND_DESIGN	4.3
10836	FAMILY	4.5
10837	FAMILY	5.0
10838	MEDICAL	NaN
10839	BOOKS AND REFERENCE	4.5
10840	LIFESTYLE	4.5

[10841 rows x 2 columns]

From the above mentioned dataframe and box plot,I can analyze that Categories such as Art and Design,Family,Books and Reference and Lifestyle has best ratings ie 4.5,4.7 and 5

#Q8) Data preprocessing #For the steps below, create a copy of the dataframe to make all the edits. Name it inpl.

#8)1-Reviews and Install have some values that are still relatively very high. Before building a linear regression model, #you need to reduce the skew. Apply log transformation (np.log1p) to Reviews and Installs.

df # The data frame

10838

10839

10840

NaN

4.5

4.5 398307

3

114 Varies with device

Арр
Category \ Photo Editor & Candy Camera & Grid & ScrapBook
RT_AND_DESIGN Coloring book moana
RT_AND_DESIGN U Launcher Lite — FREE Live Cool Themes, Hide
RT_AND_DESIGN Sketch - Draw & Paint
RT_AND_DESIGN Pixel Draw - Number Art Coloring Book BT AND DESIGN
RT_AND_DESIGN
 0836
AMILY 0837 Fr. Mike Schmitz Audio Teachings
AMILY .0838 Parkinson Exercices FR
IEDICAL .0839 The SCP Foundation DB fr nn5n
00KS_AND_REFERENCE .0840 iHoroscope - 2018 Daily Horoscope & Astrology .IFESTYLE
Rating Reviews Size Installs Type Price \ 4.1 159 19M 10,000+ Free 0 3.9 967 14M 500,000+ Free 0 4.7 87510 8.7M 5,000,000+ Free 0 4.5 215644 25M 50,000,000+ Free 0 4.3 967 2.8M 100,000+ Free 0

9.5M

19M

1,000+

1,000+

10,000,000+

Free

Free

Free

0

0

```
Content Rating
                                          Genres
                                                       Last Updated \
                                                   January 7, 2018
0
            Everyone
                                    Art & Design
1
            Everyone Art & Design; Pretend Play
                                                   January 15, 2018
2
            Everyone
                                    Art & Design
                                                     August 1, 2018
3
                                                       June 8, 2018
                Teen
                                    Art & Design
4
                       Art & Design; Creativity
                                                      June 20, 2018
            Everyone
                                                      July 25, 2017
10836
            Everyone
                                       Education
                                                       July 6, 2018
10837
            Everyone
                                       Education
10838
            Everyone
                                         Medical
                                                   January 20, 2017
                                                   January 19, 2015
10839
          Mature 17+
                               Books & Reference
10840
                                                      July 25, 2018
            Everyone
                                       Lifestyle
              Current Ver
                                   Android Ver
0
                    1.0.0
                                  4.0.3 and up
1
                    2.0.0
                                  4.0.3 and up
2
                    1.2.4
                                  4.0.3 and up
3
       Varies with device
                                    4.2 and up
4
                       1.1
                                    4.4 and up
. . .
                       . . .
                                    4.1 and up
10836
                      1.48
10837
                                    4.1 and up
                         1
10838
                         1
                                    2.2 and up
       Varies with device
10839
                           Varies with device
10840
       Varies with device Varies with device
[10841 rows x 13 columns]
inpl=df.copy() # Creating a copy of data frame
inpl # Copy of original data frame
                                                       App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART AND_DESIGN
                                      Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                    Sketch - Draw & Paint
ART AND DESIGN
                   Pixel Draw - Number Art Coloring Book
ART_AND_DESIGN
. . .
10836
                                         Sya9a Maroc - FR
FAMILY
                        Fr. Mike Schmitz Audio Teachings
10837
```

```
FAMILY
                                     Parkinson Exercices FR
10838
MEDICAL
10839
                             The SCP Foundation DB fr nn5n
BOOKS AND REFERENCE
10840
            iHoroscope - 2018 Daily Horoscope & Astrology
LIFESTYLE
       Rating Reviews
                                        Size
                                                  Installs
                                                             Type Price
0
           4.1
                   159
                                         19M
                                                   10,000+
                                                             Free
                                                                       0
1
          3.9
                   967
                                         14M
                                                  500,000+
                                                             Free
                                                                       0
2
          4.7
                                                5,000,000+
                 87510
                                        8.7M
                                                             Free
                                                                       0
3
          4.5
                215644
                                         25M
                                              50,000,000+
                                                             Free
                                                                       0
                                                  100,000+
4
          4.3
                   967
                                        2.8M
                                                             Free
                                                                       0
                                         . . .
                                                              . . .
           . . .
                                                    5,000+
          4.5
                    38
                                                                      0
10836
                                         53M
                                                             Free
10837
          5.0
                     4
                                        3.6M
                                                      100+
                                                             Free
                                                                       0
          NaN
                     3
                                        9.5M
                                                                       0
10838
                                                    1,000+
                                                             Free
                                                    1,000+
                                                                       0
10839
          4.5
                   114
                         Varies with device
                                                             Free
10840
          4.5
                398307
                                         19M
                                              10,000,000+
                                                             Free
                                                                       0
      Content Rating
                                            Genres
                                                         Last Updated
0
                                      Art & Design
                                                      January 7, 2018
             Everyone
1
             Everyone
                       Art & Design; Pretend Play
                                                     January 15, 2018
2
                                      Art & Design
                                                       August 1, 2018
             Everyone
3
                                                         June 8, 2018
                 Teen
                                      Art & Design
4
                                                        June 20, 2018
                          Art & Design; Creativity
             Everyone
. . .
             Everyone
                                         Education
                                                        July 25, 2017
10836
10837
             Everyone
                                         Education
                                                         July 6, 2018
                                                     January 20, 2017
10838
             Everyone
                                           Medical
10839
          Mature 17+
                                Books & Reference
                                                     January 19, 2015
                                                        July 25, 2018
10840
             Everyone
                                         Lifestyle
               Current Ver
                                     Android Ver
0
                     1.0.0
                                    4.0.3 and up
1
                     2.0.0
                                    4.0.3 and up
2
                     1.2.4
                                    4.0.3 and up
                                      4.2 and up
3
       Varies with device
4
                                      4.4 and up
                        1.1
                        . . .
                                      4.1 and up
10836
                       1.48
                          1
                                      4.1 and up
10837
10838
                                      2.2 and up
                          1
10839
       Varies with device
                             Varies with device
10840
       Varies with device
                             Varies with device
```

[10841 rows x 13 columns]

```
inp1[["Reviews","Installs"]] # Clubing the Reviews and Installs
column in inpl
      Reviews
                  Installs
                   10,000+
0
          159
1
          967
                  500,000+
2
        87510
                5,000,000+
3
       215644 50,000,000+
                  100,000+
4
          967
                    5,000+
10836
           38
                      100+
10837
            4
10838
            3
                    1,000+
10839
                    1,000+
          114
10840
      398307
               10,000,000+
[10841 rows x 2 columns]
inp1["Reviews"]=pd.to numeric(inp1["Reviews"],errors="coerce")
inp1["Reviews"].replace(to_replace=["3.0M"],value="3000000")
0
            159.0
1
            967.0
2
          87510.0
3
         215644.0
            967.0
             38.0
10836
10837
              4.0
10838
              3.0
10839
            114.0
10840
         398307.0
Name: Reviews, Length: 10841, dtype: float64
# Build The Linear Regression Model of Reviews and Installs
import pandas as pd
import numpy as np
import statistics as sp
import sklearn
from sklearn.linear_model import LinearRegression
lm=LinearRegression() # Make the constructor of Linear Regression
df.dropna(subset=["Reviews"],axis=0,inplace=True) # Dropping the
Missing Values of Reviews Column
```

```
df.dropna(subset=["Installs"],axis=0,inplace=True) # Dropping the
Missing Values of Installs Column
df['Installs'] = df['Installs'].str.replace('+',' ',regex=True)
df['Installs'] = df['Installs'].str.replace(',',','',regex=True)
df["Installs"]=pd.to numeric(df["Installs"],errors="coerce")
df["Installs"] # Converting Installs to Numeric
0
            10000.0
1
           500000.0
2
          5000000.0
3
         50000000.0
           100000.0
10836
             5000.0
10837
              100.0
10838
             1000.0
10839
             1000.0
10840
         10000000.0
Name: Installs, Length: 10841, dtype: float64
x=df[["Reviews"]]
y=df[["Installs"]]
lm.fit(x,y)
LinearRegression()
z=lm.predict(x)
Z
array([[ 7171489.7096363 ],
       [7186581.40839971],
       [ 8803018.14859996],
       [ 7168575.96581564],
       [7170649.20661111],
       [14608036.34237667]])
lm.intercept # Finding the intercept
array([7168519.93228063])
lm.coef_ # Finding the coefficient
array([[18.677845]])
# Log Transformations to Review Column
```

```
df["Reviews"]
0
             159
1
             967
2
          87510
3
         215644
4
             967
              38
10836
10837
               4
10838
               3
             114
10839
10840
         398307
Name: Reviews, Length: 10841, dtype: object
Reviews column in the dataframe have lot of Numeric values. Here, I am taking few values
of Review column and apply log1p method to it
import numpy as np
reviews array=[159,967,87510,215644,38,4,3,114,398307] # Placing Some
values of Review column in list
out array=np.log1p(reviews array)
print("The Log Transformation of Reviews Column", out array)
The Log Transformation of Reviews Column [ 5.07517382 6.87523209
11.37951978 12.28138882 3.66356165 1.60943791
  1.38629436 4.74493213 12.894980851
df["Installs"] # Data Frame of Installs Column
0
             10000.0
1
            500000.0
2
          5000000.0
3
         50000000.0
4
            100000.0
10836
              5000.0
10837
               100.0
10838
              1000.0
10839
              1000.0
10840
         10000000.0
Name: Installs, Length: 10841, dtype: float64
Installs Column of the data frame has lots of Numeric Values.Here,I am taking few numeric
values of Install Column and will apply log1p method to it
import numpy as np
myinstall=[10000,500000,5000000,50000000,100000,5000,1000,1000,1000,1000]
```

```
out array=np.log1p(myinstall)
print("Log Transformation of Install Column is:",out array)
Log Transformation of Install Column is: [ 9.21044037 13.12236538
15.42494867 17.72753358 11.51293546 8.51739317
  4.61512052 6.90875478 6.90875478 16.11809575]
#8)2--Drop columns App, Last Updated, Current Ver, and Android Ver.
#These variables are not useful for our task.
inp1=df.copy()
inp1 #Duplicated Data frame
                                                       App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART_AND_DESIGN
                                      Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                    Sketch - Draw & Paint
ART AND DESIGN
                   Pixel Draw - Number Art Coloring Book
ART AND DESIGN
10836
                                         Sya9a Maroc - FR
FAMILY
                        Fr. Mike Schmitz Audio Teachings
10837
FAMILY
10838
                                   Parkinson Exercices FR
MEDICAL
                            The SCP Foundation DB fr nn5n
10839
BOOKS AND REFERENCE
           iHoroscope - 2018 Daily Horoscope & Astrology
10840
LIFESTYLE
       Rating Reviews
                                      Size
                                              Installs
                                                         Type Price \
0
          4.1
                  159
                                       19M
                                                10000.0
                                                         Free
                                                                  0
1
          3.9
                  967
                                       14M
                                              500000.0
                                                        Free
                                                                  0
2
          4.7
                87510
                                             5000000.0
                                                                  0
                                      8.7M
                                                         Free
3
          4.5
               215644
                                       25M
                                            50000000.0
                                                         Free
                                                                  0
          4.3
                                              100000.0
4
                  967
                                      2.8M
                                                         Free
                                                                  0
                   . . .
          4.5
10836
                   38
                                       53M
                                                 5000.0
                                                         Free
                                                                  0
10837
          5.0
                    4
                                      3.6M
                                                  100.0
                                                         Free
                                                                  0
10838
          NaN
                    3
                                      9.5M
                                                 1000.0
                                                         Free
                                                                  0
10839
          4.5
                  114 Varies with device
                                                1000.0
                                                         Free
                                                                  0
```

1 2

3

4

Content Rating

Everyone

Everyone

Everyone

Teen

10836 Evervone Everyone Education July 6, 2018 10837 10838 January 20, 2017 Everyone Medical January 19, 2015 10839 Mature 17+ Books & Reference July 25, 2018 10840 Everyone Lifestyle Current Ver Android Ver

0 1.0.0 4.0.3 and up 1 4.0.3 and up 2.0.0 2 4.0.3 and up 1.2.4 3 Varies with device 4.2 and up 4 4.4 and up 1.1 10836 1.48 4.1 and up 10837 1 4.1 and up 10838 1 2.2 and up 10839 Varies with device Varies with device Varies with device Varies with device 10840

[10841 rows x 13 columns]

inpl.drop(["App","Last Updated","Current Ver","Android Ver"],axis=1) # Dropping the Columns from duplicate data frame

T 4 - 11	Category	Rating	Reviews	Size
Installs \ 0 10000.0	ART_AND_DESIGN	4.1	159	19M
1	ART_AND_DESIGN	3.9	967	14M
500000.0 2 5000000.0	ART_AND_DESIGN	4.7	87510	8.7M
3 50000000.0	ART_AND_DESIGN	4.5	215644	25M
4 100000.0	ART_AND_DESIGN	4.3	967	2.8M
 10836 5000.0	FAMILY	4.5	38	53M
10837	FAMILY	5.0	4	3.6M

100.0				
10838	MEDICAL	NaN	3	9.5M
1000.0				
10839	BOOKS AND REFERENCE	4.5	114	Varies with device
1000.0				
10840	LIFESTYLE	4.5	398307	19M
1000000	00.0			
				_

	Type	Price	Content Rating	Genres
0	Free	0	Everyone	Art & Design
1	Free	0	Everyone	Art & Design;Pretend Play
2	Free	0	Everyone	Art & Design
3	Free	0	Teen	Art & Design
4	Free	0	Everyone	Art & Design;Creativity
10836	Free	0	Everyone	Education
10837	Free	0	Everyone	Education
10838	Free	0	Everyone	Medical
10839	Free	0	Mature 17+	Books & Reference
10840	Free	0	Everyone	Lifestyle

[10841 rows x 9 columns]

#Q8)3-Get dummy columns for Category, Genres, and Content Rating. This needs to be done

#as the models do not understand categorical data, and all data should be numeric. Dummy encoding is one way to convert character fields to numeric.

#Name of dataframe should be inp2.

df # The data frame

	Арр
Category \ 0 Photo Editor 8	& Candy Camera & Grid & ScrapBook
ART_AND_DESIGN 1	Coloring book moana
ART_AND_DESIGN 2 U Launcher Lite - ART AND DESIGN	- FREE Live Cool Themes, Hide
3 ART AND DESIGN	Sketch - Draw & Paint
	l Draw - Number Art Coloring Book
• • •	•••
10836 FAMILY	Sya9a Maroc - FR
10837 FAMILY	Fr. Mike Schmitz Audio Teachings
10838	Parkinson Exercices FR

MEDICAL
10839 The SCP Foundation DB fr nn5n
B00KS_AND_REFERENCE
10840 iHoroscope - 2018 Daily Horoscope & Astrology
LIFESTYLE

0 1 2 3 4 10836 10837 10838 10839 10840	Rating Reviews 4.1 159 3.9 967 4.7 87510 4.5 215644 4.3 967 4.5 38 5.0 4 NaN 3 4.5 114 4.5 398307		19M 10 14M 500 8.7M 5000 25M 50000 2.8M 100 53M 5 3.6M 9.5M	stalls 0000.0 0000.0 0000.0 0000.0 5000.0 100.0 1000.0 1000.0	Type Free Free Free Free Free Free Free Fr	Price 0 0 0 0 0 0 0 0 0 0 0 0 0	\
0 1 2 3 4 10836 10837 10838 10839 10840	Content Rating Everyone	Art & Design;Pr Ar Ar Art & Design;	Genres Tt & Design Tetend Play Tt & Design Tt & Design	Janua Janua Aug Ju Ju Janua Janua	ast Up ary 7, ry 15, ust 1, une 8, ne 20, ly 25, uly 6, ry 20, ry 19, ly 25,	odated 2018 2018 2018 2018 2018 2017 2018 2017 2015	\
0 1 2 3 4	2. 1. Varies with dev	9.0 4.0. 9.0 4.0. 2.4 4.0. ice 4.	roid Ver 3 and up 3 and up 3 and up 2 and up 4 and up				
10836 10837 10838 10839 10840		1 4. 1 2. ice Varies wit					
[10041	rove v 12 colum	ac 1					

[10841 rows x 13 columns]

inp2=pd.get_dummies(df[["Category","Genres","Content Rating"]]) #
Dummy Column for Category, Genres and Content Rating

inp2 # The Updated Data Frame with one hot encoding

Catago	Category_1.9 Ca	tegory_ART_AND_DESIGN	
0	ry_AUTO_AND_VEHICI 0	1	
0 1	0	1	
0	0	1	
0 3	0	1	
0 4	Θ	1	
0		• • •	
10836	Θ	0	
0 10837	Θ	0	
0 10838	Θ	Θ	
0 10839	Θ	0	
0 10840 0	0	0	
ŭ	Category BEAUTY	Category BOOKS AND REFERENCE	
Catego	ry_BUSINESS \	Category_BOOKS_AND_REFERENCE	
Catego 0 0	ry_BUSINESS \ 0	Θ	
Catego 0 0 1	ry_BUSINESS \ 0	0	
Catego 0 0 1 0 2	ry_BUSINESS \ 0 0	0 0 0	
Catego 0 0 1 0 2 0 3	ry_BUSINESS \	9 9 9	
Catego 0 0 1 0 2 0 3	ry_BUSINESS \ 0 0	0 0 0	
Catego 0 0 1 0 2 0 3 0 4 0	ry_BUSINESS \	0 0 0 0	
Catego 0 0 1 0 2 0 3 0 4 0 	ry_BUSINESS \		
Catego 0 0 1 0 2 0 3 0 4 0 10836 0 10837	ry_BUSINESS \	0 0 0 0 	
Catego 0 0 1 0 2 0 3 0 4 0 10836 0 10837 0 10838	ry_BUSINESS \		
Catego 0 0 1 0 2 0 3 0 4 0 10836 0 10837 0 10838	ry_BUSINESS \	0 0 0 0 	

```
Category_COMICS
                           Category_COMMUNICATION
                                                      Category_DATING
0
                        0
                                                   0
1
                                                                       0
2
                        0
                                                   0
                                                                       0
3
                        0
                                                   0
                                                                       0
4
                        0
                                                   0
                                                                       0
10836
                        0
                                                   0
                                                                       0
                        0
                                                   0
10837
                                                                       0
10838
                        0
                                                   0
                                                                       0
10839
                        0
                                                   0
                                                                       0
                        0
                                                   0
                                                                       0
10840
                               ... Genres_Video Players &
        Category_EDUCATION
Editors; Creativity \
                           0
                               . . .
0
1
                           0
0
2
                           0
0
3
                           0
0
4
                           0
0
. . .
10836
                           0
10837
                           0
0
10838
                           0
10839
                           0
0
10840
                           0
0
        Genres_Video Players & Editors;Music & Video
                                                             Genres_Weather
0
                                                                            0
1
                                                         0
                                                                            0
2
                                                         0
                                                                            0
3
                                                         0
                                                                            0
4
                                                         0
                                                                            0
                                                         0
10836
                                                                           0
10837
                                                         0
                                                                            0
10838
                                                         0
                                                                            0
```

```
10840
                                                        0
       Genres_Word Content Rating_Adults only 18+ Content
Rating_Everyone \
                   0
                                                       0
1
1
                                                       0
                   0
1
2
                   0
                                                       0
1
3
                   0
                                                       0
0
4
                   0
                                                       0
1
. . .
10836
                   0
                                                       0
10837
                   0
                                                       0
10838
                                                       0
1
10839
                   0
                                                       0
10840
                   0
                                                       0
1
       Content Rating_Everyone 10+ Content Rating_Mature 17+
0
1
                                     0
                                                                   0
2
                                     0
                                                                   0
3
                                     0
                                                                   0
4
                                     0
                                                                   0
10836
                                     0
                                                                   0
                                     0
                                                                   0
10837
10838
                                     0
                                                                   0
10839
                                     0
                                                                   1
10840
                                     0
       Content Rating_Teen Content Rating_Unrated
0
                                                       0
1
                            0
2
                            0
                                                       0
3
                            1
4
                            0
                                                       0
10836
                            0
                                                       0
10837
                            0
                                                       0
```

10838	0	0
10839	0	0
10840	0	0

[10841 rows x 160 columns]

#09) Train test split and apply 70-30 split. Name the new dataframes df_{train} and df_{test} .

import sklearn

from sklearn.model_selection import train_test_split
df

Арр
Category \ 0 Photo Editor & Candy Camera & Grid & ScrapBook ART AND DESIGN
1 Coloring book moana
ART_AND_DESIGN 2
3 Sketch - Draw & Paint
ART_AND_DESIGN 4 Pixel Draw - Number Art Coloring Book ART AND DESIGN
10836 Sya9a Maroc - FR
10837 Fr. Mike Schmitz Audio Teachings
10838 Parkinson Exercices FR MEDICAL
10839 The SCP Foundation DB fr nn5n BOOKS AND REFERENCE
10840 iHoroscope - 2018 Daily Horoscope & Astrology LIFESTYLE

	Rating	Reviews		Size	Installs	Type	Price	\
0	4.1	159		19M	10,000+	Free	0	
1	3.9	967		14M	500,000+	Free	0	
2	4.7	87510		8.7M	5,000,000+	Free	0	
3	4.5	215644		25M	50,000,000+	Free	0	
4	4.3	967		2.8M	100,000+	Free	0	
10836	4.5	38		53M	5,000+	Free	0	
10837	5.0	4		3.6M	100+	Free	0	
10838	NaN	3		9.5M	1,000+	Free	0	
10839	4.5	114	Varies with	device	1,000+	Free	0	

```
10840
          4.5 398307
                                             10,000,000+ Free
                                       19M
                                                                    0
      Content Rating
                                           Genres
                                                       Last Updated
0
            Everyone
                                    Art & Design
                                                    January 7, 2018
                                                   January 15, 2018
1
            Everyone Art & Design; Pretend Play
2
                                    Art & Design
                                                     August 1, 2018
            Everyone
3
                Teen
                                    Art & Design
                                                       June 8, 2018
                                                      June 20, 2018
4
            Everyone
                         Art & Design;Creativity
            Everyone
                                                      July 25, 2017
10836
                                        Education
                                        Education
                                                       July 6, 2018
10837
            Everyone
                                                   January 20, 2017
10838
            Everyone
                                          Medical
                                                   January 19, 2015
10839
                               Books & Reference
          Mature 17+
10840
            Everyone
                                       Lifestyle
                                                      July 25, 2018
              Current Ver
                                   Android Ver
0
                     1.0.0
                                  4.0.3 and up
1
                     2.0.0
                                  4.0.3 and up
2
                     1.2.4
                                  4.0.3 and up
3
       Varies with device
                                    4.2 and up
4
                                    4.4 and up
                       1.1
10836
                      1.48
                                    4.1 and up
10837
                         1
                                    4.1 and up
10838
                         1
                                    2.2 and up
10839
       Varies with device
                           Varies with device
       Varies with device
10840
                           Varies with device
[10841 rows x 13 columns]
x=df["Reviews"]
y=df["Installs"]
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.30,rand
om state=0)
x train
5191
           25627
7157
3184
           18039
1916
         1083571
6423
              52
4859
          384602
3264
           28250
9845
          398746
```

2732

2036

Name: Reviews, Length: 7588, dtype: object

```
x_test
7487
           144545
5963
                10
8654
            25370
7789
                10
9702
                18
3998
               873
7360
              4928
10035
              9699
9433
                85
2550
         17014787
Name: Reviews, Length: 3253, dtype: object
y_train
5191
          1,000,000+
7157
                 100+
3184
          1,000,000+
1916
         50,000,000+
6423
              10,000+
4859
         10,000,000+
3264
         10,000,000+
9845
         10,000,000+
10799
             100,000+
2732
         50,000,000+
Name: Installs, Length: 7588, dtype: object
y_test
7487
           5,000,000+
5963
                5,000+
           1,000,000+
8654
7789
                1,000+
                1,000+
9702
3998
              50,000+
               50,000+
7360
              100,000+
10035
9433
               10,000+
2550
         500,000,000+
Name: Installs, Length: 3253, dtype: object
df_train=pd.DataFrame(x_train,columns=["Reviews"])
df_train
       Reviews
5191
         25627
7157
              6
```

```
3184
         18039
1916
       1083571
6423
            52
. . .
            . . .
        384602
4859
3264
         28250
9845
        398746
10799
          2036
2732
       1370749
[7588 rows \times 1 columns]
df_test=pd.DataFrame(x_test,columns=["Reviews"])
df test
        Reviews
7487
         144545
5963
              10
          25370
8654
7789
              10
9702
              18
. . .
             . . .
3998
            873
7360
           4928
10035
           9699
9433
             85
2550
       17014787
[3253 rows x 1 columns]
#Q10) Separate the dataframes into X_train, y_train, X_test, and
y test.
import sklearn
from sklearn.model selection import train test split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.30,rand
om_state=0)
x train
5191
           25627
7157
3184
           18039
1916
         1083571
6423
               52
4859
          384602
3264
           28250
9845
          398746
```

```
10799
            2036
2732
         1370749
Name: Reviews, Length: 7588, dtype: object
x test
7487
           144545
5963
               10
8654
            25370
7789
               10
9702
               18
3998
              873
7360
             4928
10035
             9699
9433
               85
2550
         17014787
Name: Reviews, Length: 3253, dtype: object
y_train
          1,000,000+
5191
7157
                100+
3184
          1,000,000+
1916
         50,000,000+
6423
             10,000+
4859
         10,000,000+
3264
         10,000,000+
9845
         10,000,000+
10799
            100,000+
2732
         50,000,000+
Name: Installs, Length: 7588, dtype: object
y_test
7487
           5,000,000+
               5,000+
5963
8654
           1,000,000+
7789
               1,000+
9702
               1,000+
3998
              50,000+
7360
              50,000+
10035
             100,000+
9433
              10,000+
2550
         500,000,000+
Name: Installs, Length: 3253, dtype: object
#Q11)-- Model building
#- Use linear regression as the technique
```

```
import seaborn as sns
import sklearn
from sklearn.linear model import LinearRegression
lm=LinearRegression() # Create the Constructor
df # The data frame
                                                        App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART AND DESIGN
                                       Coloring book moana
ART AND DESIGN
       U Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                     Sketch - Draw & Paint
ART AND DESIGN
                    Pixel Draw - Number Art Coloring Book
ART AND DESIGN
. . .
                                          Sya9a Maroc - FR
10836
FAMILY
                         Fr. Mike Schmitz Audio Teachings
10837
FAMILY
10838
                                   Parkinson Exercices FR
MEDICAL
                            The SCP Foundation DB fr nn5n
10839
BOOKS AND REFERENCE
10840
           iHoroscope - 2018 Daily Horoscope & Astrology
LIFESTYLE
       Rating Reviews
                                       Size
                                                Installs
                                                           Type Price
                                                 10,000+
                                                           Free
0
          4.1
                   159
                                        19M
                                                                    0
                                                500,000+
1
          3.9
                   967
                                        14M
                                                           Free
                                                                    0
2
          4.7
                                              5,000,000+
                                                           Free
                87510
                                       8.7M
                                                                    0
          4.5
                                             50,000,000+
3
               215644
                                        25M
                                                           Free
                                                                    0
4
          4.3
                   967
                                       2.8M
                                                100,000+
                                                           Free
                                                                    0
. . .
10836
          4.5
                    38
                                        53M
                                                  5,000+
                                                           Free
                                                                    0
          5.0
                     4
                                                                    0
10837
                                       3.6M
                                                     100+
                                                           Free
                     3
10838
          NaN
                                       9.5M
                                                  1,000+
                                                           Free
                                                                    0
          4.5
                        Varies with device
                                                  1,000+
                                                                    0
10839
                   114
                                                           Free
10840
          4.5
               398307
                                        19M
                                             10,000,000+
                                                           Free
                                                                    0
      Content Rating
                                           Genres
                                                        Last Updated \
0
            Everyone
                                     Art & Design
                                                    January 7, 2018
```

Everyone Art & Design; Pretend Play January 15, 2018

```
2
            Everyone
                                     Art & Design
                                                     August 1, 2018
3
                Teen
                                     Art & Design
                                                        June 8, 2018
4
            Everyone
                         Art & Design;Creativity
                                                       June 20, 2018
                  . . .
                                        Education
                                                       July 25, 2017
10836
            Everyone
                                                        July 6, 2018
10837
            Everyone
                                        Education
            Everyone
10838
                                          Medical
                                                   January 20, 2017
                                                   January 19, 2015
10839
          Mature 17+
                               Books & Reference
10840
            Everyone
                                        Lifestyle
                                                       July 25, 2018
              Current Ver
                                   Android Ver
0
                     1.0.0
                                  4.0.3 and up
1
                     2.0.0
                                  4.0.3 and up
2
                                  4.0.3 and up
                     1.2.4
3
       Varies with device
                                     4.2 and up
4
                       1.1
                                     4.4 and up
. . .
                       . . .
10836
                      1.48
                                     4.1 and up
10837
                         1
                                     4.1 and up
10838
                         1
                                     2.2 and up
       Varies with device
10839
                           Varies with device
10840
       Varies with device Varies with device
[10841 rows x 13 columns]
df.dtypes
                    object
App
Category
                    object
Rating
                   float64
Reviews
                    object
Size
                    object
Installs
                    object
Type
                    object
Price
                    object
Content Rating
                    object
Genres
                    object
Last Updated
                    object
Current Ver
                    object
Android Ver
                    object
dtype: object
df['Price'] = df['Price'].str.replace('$','',regex=True)
df["Price"]
0
         0
1
         0
2
         0
3
         0
4
         0
```

```
10836
         0
10837
         0
10838
         0
10839
         0
10840
Name: Price, Length: 10841, dtype: object
df["Price"].replace(to replace=["Everyone"], value="0")
         0
0
1
         0
2
         0
3
         0
4
         0
10836
         0
10837
         0
10838
         0
10839
         0
10840
         0
Name: Price, Length: 10841, dtype: object
df["Price"]
0
         0
1
         0
2
         0
3
         0
4
         0
10836
         0
10837
         0
10838
         0
10839
         0
10840
Name: Price, Length: 10841, dtype: object
df.dropna(subset=["Price"],axis=0,inplace=True) # Drop the Missing
Values of Price Column
df.dropna(subset=["Rating"],axis=0,inplace=True) #Dropping the
Missing Values of Rating Column
df["Price"].replace(to replace=["Everyone"], value="0")
         0
0
1
         0
2
         0
3
         0
4
         0
```

```
10834
         0
10836
         0
10837
         0
10839
         0
10840
Name: Price, Length: 9367, dtype: object
df["Price"]=pd.to numeric(df["Price"],errors="coerce")
x=df[["Rating"]] # Placing the Rating on X axis
y=df["Price"] # Placing the App Price on Y axis
lm.fit(x,y)
LinearRegression()
lm.intercept
3.7794838025024253
lm.coef
array([-0.67240436])
Linear Regression basically means one independent variable to make a prediction. Here, I
am placing the independent variable Rating on X axis and dependent variable price on Y
axis. Hence, The Relationship between App Ratings and Price is 3.777-0.67
#Q11)B- Report the R2 on the train set
Here, R2 is Ridge Regression. Ridge Regression is use to prevent overfitting
import sklearn
from sklearn.linear model import Ridge
RigeModel=Ridge(alpha=0.1)
x=df[["Rating"]] # Placing the Rating on X axis
from sklearn.model selection import train test split
from sklearn.model selection import cross val predict
x=df["Rating"]
y=df["Price"]
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=3,random_
state=0)
x train
```

```
5378
         4.7
3673
         3.7
7136
         5.0
4910
         2.5
6592
         3.9
8923
         4.1
10635
         4.6
5073
         4.3
3405
         4.3
2871
         4.1
Name: Rating, Length: 9363, dtype: float64
x_test
        4.2
1792
4729
        4.2
8177
        4.3
Name: Rating, dtype: float64
df test=pd.DataFrame(x test,columns=["Rating"])
inp3test = pd.DataFrame(y test, columns=["Price"])
refinedr2=RigeModel.fit(df_test,inp3test)
refinedr2
Ridge(alpha=0.1)
#Q12) Make predictions on test set and report R2.
import numpy as np
import statistics
import sklearn
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_predict
from sklearn.model_selection import cross_val_score
df # The data frame
                                                      App
Category
          Photo Editor & Candy Camera & Grid & ScrapBook
ART_AND_DESIGN
                                      Coloring book moana
ART_AND_DESIGN
```

```
U Launcher Lite - FREE Live Cool Themes, Hide ...
ART AND DESIGN
                                     Sketch - Draw & Paint
ART AND DESIGN
                    Pixel Draw - Number Art Coloring Book
ART AND_DESIGN
                                                         . . .
. . .
10834
                                              FR Calculator
FAMILY
                                           Sva9a Maroc - FR
10836
FAMILY
10837
                         Fr. Mike Schmitz Audio Teachings
FAMILY
10839
                            The SCP Foundation DB fr nn5n
BOOKS AND REFERENCE
           iHoroscope - 2018 Daily Horoscope & Astrology
LIFESTYLE
       Rating Reviews
                                                 Installs
                                                            Type
                                       Size
                                                                  Price \
          4.1
                   159
                                        19M
                                                  10,000+
                                                            Free
                                                                     0.0
          3.9
                   967
1
                                        14M
                                                 500,000+
                                                            Free
                                                                     0.0
2
          4.7
                                               5,000,000+
                 87510
                                       8.7M
                                                            Free
                                                                     0.0
          4.5
                                              50,000,000+
3
                215644
                                        25M
                                                            Free
                                                                     0.0
4
          4.3
                   967
                                       2.8M
                                                 100,000+
                                                            Free
                                                                    0.0
           . . .
                   . . .
                                         . . .
10834
          4.0
                    7
                                       2.6M
                                                      500+
                                                            Free
                                                                    0.0
          4.5
                    38
                                                   5,000+
10836
                                         53M
                                                            Free
                                                                     0.0
          5.0
10837
                     4
                                       3.6M
                                                      100+
                                                            Free
                                                                     0.0
          4.5
                        Varies with device
10839
                   114
                                                   1,000+
                                                            Free
                                                                    0.0
10840
          4.5
               398307
                                         19M
                                              10,000,000+
                                                            Free
                                                                     0.0
      Content Rating
                                                         Last Updated
                                            Genres
                                     Art & Design
                                                     January 7, 2018
0
            Everyone
                                                    January 15, 2018
1
            Everyone Art & Design; Pretend Play
2
            Everyone
                                     Art & Design
                                                      August 1, 2018
3
                 Teen
                                     Art & Design
                                                         June 8, 2018
4
            Everyone
                       Art & Design; Creativity
                                                        June 20, 2018
                                                        June 18, 2017
            Everyone
                                         Education
10834
10836
            Everyone
                                         Education
                                                        July 25, 2017
                                                         July 6, 2018
10837
            Everyone
                                         Education
                                Books & Reference
                                                    January 19, 2015
10839
          Mature 17+
10840
            Everyone
                                         Lifestyle
                                                       July 25, 2018
               Current Ver
                                    Android Ver
0
                     1.0.0
                                   4.0.3 and up
1
                     2.0.0
                                   4.0.3 and up
2
                     1.2.4
                                   4.0.3 and up
3
       Varies with device
                                     4.2 and up
```

```
4
                       1.1
                                    4.4 and up
. . .
                       . . .
10834
                     1.0.0
                                     4.1 and up
10836
                      1.48
                                     4.1 and up
10837
                         1
                                     4.1 and up
10839 Varies with device Varies with device
10840 Varies with device Varies with device
[9366 rows x 13 columns]
x=df[["Rating"]]
y=df["Price"]
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.30,rand
om state=0)
lm.fit(x_test,y_test)
LinearRegression()
x_{train}
       Rating
3926
          4.3
7683
          3.9
          5.0
640
5536
          4.7
2139
          4.6
          4.1
8923
          4.6
10635
          4.3
5073
          4.3
3405
2871
          4.1
[6556 rows x 1 columns]
x test
      Rating
1792
         4.2
4729
         4.2
8177
         4.3
5378
         4.7
3673
         3.7
         . . .
4495
         4.4
         4.6
3422
3982
         4.8
1031
         4.5
6921
         4.5
```

```
[2810 rows x 1 columns]
y_train
3926
         0.00
7683
         0.00
640
         0.00
5536
         0.00
         0.00
2139
8923
         0.00
10635
         0.00
5073
         0.00
3405
         0.99
2871
         0.00
Name: Price, Length: 6556, dtype: float64
y_test
1792
        0.0
4729
        0.0
8177
        0.0
5378
        0.0
3673
        0.0
4495
        0.0
3422
        0.0
3982
        0.0
1031
        0.0
6921
        0.0
Name: Price, Length: 2810, dtype: float64
lm.fit(x_test,y_test)
LinearRegression()
lm.intercept
4.438301713071732
lm.coef_
array([-0.85464264])
import sklearn
from sklearn.linear_model import Ridge
RigeModel=Ridge(alpha=0.1)
RigeModel.fit(x_test,y_test)
```

```
Ridge(alpha=0.1)
Ridge(alpha=0.1)
Ridge(alpha=0.1)
RigeModel.predict(x test)
array([0.84880482, 0.84880482, 0.7633514 , ..., 0.3360843 ,
0.59244456,
       0.592444561)
RigeModel=Ridge(alpha=0.6)
RigeModel.fit(x test,y test)
Ridge(alpha=0.6)
Ridge(alpha=0.6)
Ridge(alpha=0.6)
RigeModel.predict(x test)
array([0.84881581, 0.84881581, 0.76341656, ..., 0.33642032,
0.59261807,
       0.592618071)
RigeModel=Ridge(alpha=0.9)
RigeModel.fit(x test,y test)
Ridge(alpha=0.9)
Ridge(alpha=0.9)
Ridge(alpha=0.9)
RigeModel.predict(x_test)
array([0.8488224 , 0.8488224 , 0.76345562, ..., 0.33662174,
0.59272207,
       0.59272207])
In RidgeModel, the concept of trained models are used. Here I am placing the Ratings of App
on x axis and app prices on y axis. When I am increasing the value of alpha,the values in
array are fluctuating a bit
pip install nbconvert
Note: you may need to restart the kernel to use updated
packages.Requirement already satisfied: nbconvert in c:\users\shiva\
anaconda3\lib\site-packages (6.1.0)
```

```
Requirement already satisfied: jupyterlab-pygments in c:\users\shiva\
anaconda3\lib\site-packages (from nbconvert) (0.1.2)
Requirement already satisfied: pygments>=2.4.1 in c:\users\shiva\
anaconda3\lib\site-packages (from nbconvert) (2.10.0)
Requirement already satisfied: traitlets>=5.0 in c:\users\shiva\
anaconda3\lib\site-packages (from nbconvert) (5.1.0)
Requirement already satisfied: jupyter-core in c:\users\shiva\
anaconda3\lib\site-packages (from nbconvert) (4.8.1)
Requirement already satisfied: jinja2>=2.4 in c:\users\shiva\
anaconda3\lib\site-packages (from nbconvert) (2.11.3)
Requirement already satisfied: entrypoints>=0.2.2 in c:\users\shiva\
anaconda3\lib\site-packages (from nbconvert) (0.3)
Requirement already satisfied: testpath in c:\users\shiva\anaconda3\
lib\site-packages (from nbconvert) (0.5.0)
Requirement already satisfied: mistune<2,>=0.8.1 in c:\users\shiva\
anaconda3\lib\site-packages (from nbconvert) (0.8.4)
Requirement already satisfied: bleach in c:\users\shiva\anaconda3\lib\
site-packages (from nbconvert) (4.0.0)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\shiva\
anaconda3\lib\site-packages (from nbconvert) (1.4.3)
Requirement already satisfied: defusedxml in c:\users\shiva\anaconda3\
lib\site-packages (from nbconvert) (0.7.1)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in c:\users\
shiva\anaconda3\lib\site-packages (from nbconvert) (0.5.3)
Requirement already satisfied: nbformat>=4.4 in c:\users\shiva\
anaconda3\lib\site-packages (from nbconvert) (5.1.3)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\shiva\
anaconda3\lib\site-packages (from jinja2>=2.4->nbconvert) (1.1.1)
Requirement already satisfied: jupyter-client>=6.1.5 in c:\users\
shiva\anaconda3\lib\site-packages (from nbclient<0.6.0,>=0.5.0-
>nbconvert) (6.1.12)
Requirement already satisfied: async-generator in c:\users\shiva\
anaconda3\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert)
Requirement already satisfied: nest-asyncio in c:\users\shiva\
anaconda3\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert)
(1.5.1)
Requirement already satisfied: tornado>=4.1 in c:\users\shiva\
anaconda3\lib\site-packages (from jupyter-client>=6.1.5-
>nbclient<0.6.0,>=0.5.0->nbconvert) (6.1)
Reguirement already satisfied: pyzmg>=13 in c:\users\shiva\anaconda3\
lib\site-packages (from jupyter-client>=6.1.5->nbclient<0.6.0,>=0.5.0-
>nbconvert) (22.2.1)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\shiva\
anaconda3\lib\site-packages (from jupyter-client>=6.1.5-
>nbclient<0.6.0,>=0.5.0->nbconvert) (2.8.2)
Reguirement already satisfied: pywin32>=1.0 in c:\users\shiva\
anaconda3\lib\site-packages (from jupyter-core->nbconvert) (228)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in c:\users\
shiva\anaconda3\lib\site-packages (from nbformat>=4.4->nbconvert)
```

(3.2.0)Requirement already satisfied: ipython-genutils in c:\users\shiva\ anaconda3\lib\site-packages (from nbformat>=4.4->nbconvert) (0.2.0) Requirement already satisfied: setuptools in c:\users\shiva\anaconda3\ lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (58.0.4) Requirement already satisfied: pyrsistent>=0.14.0 in c:\users\shiva\ anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (0.18.0) Requirement already satisfied: six>=1.11.0 in c:\users\shiva\ anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (1.16.0) Requirement already satisfied: attrs>=17.4.0 in c:\users\shiva\ anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (21.2.0) Requirement already satisfied: webencodings in c:\users\shiva\ anaconda3\lib\site-packages (from bleach->nbconvert) (0.5.1) Requirement already satisfied: packaging in c:\users\shiva\anaconda3\ lib\site-packages (from bleach->nbconvert) (21.0) Requirement already satisfied: pyparsing>=2.0.2 in c:\users\shiva\ anaconda3\lib\site-packages (from packaging->bleach->nbconvert) (3.0.4)

pip install pandoc

Requirement already satisfied: pandoc in c:\users\shiva\anaconda3\lib\site-packages (2.3)
Requirement already satisfied: plumbum in c:\users\shiva\anaconda3\lib\site-packages (from pandoc) (1.8.1)
Requirement already satisfied: ply in c:\users\shiva\anaconda3\lib\site-packages (from pandoc) (3.11)Note: you may need to restart the kernel to use updated packages.

Requirement already satisfied: pywin32 in c:\users\shiva\anaconda3\lib\site-packages (from plumbum->pandoc) (228)