

```
In [1]: import pandas as pd
```

```
In [4]: data = pd.read_csv("C:\\Users\\makod\\Downloads\\Telegram Desktop\\googleplays
```

```
In [5]: type(data)
```

```
Out[5]: pandas.core.frame.DataFrame
```

In [6]: data.info

Out[6]: <bound method DataFrame.info of

App	Category \	
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN
1	Coloring book moana	ART_AND_DESIGN
2	U Launcher Lite - FREE Live Cool Themes, Hide ...	ART_AND_DESIGN
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
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	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
3	Teen	Art & Design	June 8, 2018
4	Everyone	Art & Design;Creativity	June 20, 2018
...
10836	Everyone	Education	July 25, 2017
10837	Everyone	Education	July 6, 2018
10838	Everyone	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	1.1	4.4 and up
...
10836	1.48	4.1 and up
10837	1.0	4.1 and up
10838	1.0	2.2 and up
10839	Varies with device	Varies with device
10840	Varies with device	Varies with device

[10841 rows x 13 columns]>

In [8]: `data.describe()`

Out[8]:

	Rating
count	8893.000000
mean	4.189542
std	0.545452
min	1.000000
25%	4.000000
50%	4.300000
75%	4.500000
max	19.000000

```
In [7]: data = data.drop_duplicates()
data
```

Out[7]:

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

10358 rows × 13 columns

```
In [9]: data.isnull()
```

```
Out[9]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
...
10836	False	False	False	False	False	False	False	False	False	False	False
10837	False	False	False	False	False	False	False	False	False	False	False
10838	False	False	True	False	False	False	False	False	False	False	False
10839	False	False	False	False	False	False	False	False	False	False	False
10840	False	False	False	False	False	False	False	False	False	False	False

10358 rows × 13 columns



```
In [11]: data.isnull().sum().sum()
```

```
Out[11]: 1478
```

```
In [12]: data2 = data.fillna(value = 0)
data2
```

Out[12]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0
...
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free	0
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100+	Free	0
10838	Parkinson Exercices FR	MEDICAL	0.0	3	9.5M	1,000+	Free	0
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0

10358 rows × 13 columns

```
In [13]: data3 = data.fillna(method = 'pad')
data3
```

Out[13]:

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

10358 rows × 13 columns

```
In [16]: data4 = data.fillna(method='bfill')
data4
```

Out[16]:

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

10358 rows × 13 columns


```
In [18]: import numpy as np
         from scipy import stats
```

```
In [19]: data2.columns
```

```
Out[19]: Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
               'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
               'Android Ver'],
              dtype='object')
```

```
In [20]: data2.drop(['App'],axis = 1,inplace = True)
         data2.columns
```

```
Out[20]: Index(['Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type', 'Price',
               'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
               'Android Ver'],
              dtype='object')
```

```
In [25]: Q1 = data2.quantile(0.25)
         Q3 = data2.quantile(0.75)
         IQR = Q3-Q1
         print(IQR)
```

```
Rating    0.8
dtype: float64
```

C:\Users\makod\AppData\Local\Temp\ipykernel_27776\4217053226.py:1: FutureWarning: The default value of numeric_only in DataFrame.quantile is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

```
Q1 = data2.quantile(0.25)
```

C:\Users\makod\AppData\Local\Temp\ipykernel_27776\4217053226.py:2: FutureWarning: The default value of numeric_only in DataFrame.quantile is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

```
Q3 = data2.quantile(0.75)
```

```
In [26]: data2 = data2[~((data2 < (Q1 - 1.5 * IQR)) | (data2 > (Q3 + 1.5 * IQR))).any(axis = 1)]
data2
```

C:\Users\makod\AppData\Local\Temp\ipykernel_27776\1137264789.py:1: FutureWarning: Automatic reindexing on DataFrame vs Series comparisons is deprecated and will raise ValueError in a future version. Do `left, right = left.align(right, axis=1, copy=False)` before e.g. `left == right`
 data2 = data2[~((data2 < (Q1 - 1.5 * IQR)) | (data2 > (Q3 + 1.5 * IQR))).any(axis = 1)]

Out[26]:

	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
0	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone
1	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone
2	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone
3	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen
4	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone
...
10834	FAMILY	4.0	7	2.6M	500+	Free	0	Everyone
10836	FAMILY	4.5	38	53M	5,000+	Free	0	Everyone
10837	FAMILY	5.0	4	3.6M	100+	Free	0	Everyone
10839	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0	Mature 17+
10840	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	Everyone

8744 rows × 12 columns



In [27]: data2.describe()

Out[27]:

	Rating
count	8744.000000
mean	4.224874
std	0.437870
min	2.600000
25%	4.000000
50%	4.300000
75%	4.500000
max	5.000000

In []: