```
#import the libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns


#import the dataset
#Question-1 : Load a dataset in your IDE.
df=pd.read_csv("/content/googleplaystore_v2.csv")
df
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.000000	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.000000	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.000000	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25000.000000	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2800.000000	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	
...	...	...	...	...	...	...	...	...	...	...	...	
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53000.000000	5,000+	Free	0	Everyone	Education	July 25, 2017	
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.000000	100+	Free	0	Everyone	Education	July 6, 2018	
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9500.000000	1,000+	Free	0	Everyone	Medical	January 20, 2017	
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	21516.529524	1,000+	Free	0	Mature 17+	Books & Reference	January 19, 2015	
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.000000	10,000,000+	Free	0	Everyone	Lifestyle	July 25, 2018	

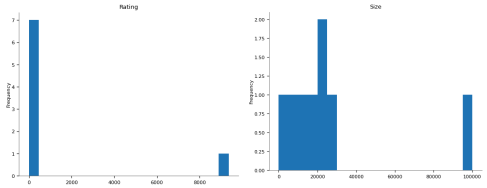
10841 rows × 13 columns

Next steps: Generate code with df View recommended plots

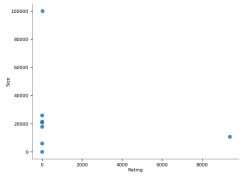
```
#Question 2 : Observe the statistics of all the features.
df.describe()
```

	Rating	Size
count	9367.000000	10841.000000
mean	4.193338	21516.529524
std	0.537431	20746.537567
min	1.000000	8.500000
25%	4.000000	5900.000000
50%	4.300000	18000.000000
75%	4.500000	26000.000000
max	19.000000	100000.000000

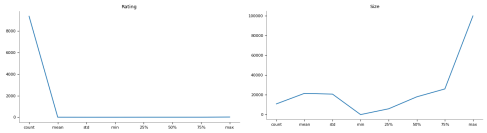
Distributions



2-d distributions



Values




```
#Questio 3 : Obtain the shape of the dataset
df.shape
```

```
(10841, 13)
```

```
#top 5 Value of the dataset
df.head()
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Cont Rat
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19000.0	10,000+	Free	0	Every
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14000.0	500,000+	Free	0	Every
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8700.0	5,000,000+	Free	0	Every
3	Sketch - Draw &	ART_AND_DESIGN	4.5	215644	25000.0	50,000,000+	Free	0	T

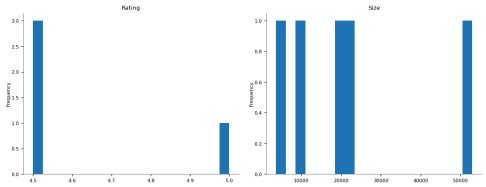
```
#Question 4 : Separate all the features
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   App             10841 non-null  object 
 1   Category        10841 non-null  object 
 2   Rating          9367 non-null   float64
 3   Reviews         10841 non-null  object 
 4   Size            10841 non-null  float64
 5   Installs        10841 non-null  object 
 6   Type            10840 non-null  object 
 7   Price           10841 non-null  object 
 8   Content Rating  10840 non-null  object 
 9   Genres          10841 non-null  object 
 10  Last Updated    10841 non-null  object 
 11  Current Ver     10833 non-null  object 
 12  Android Ver     10838 non-null  object 
dtypes: float64(2), object(11)
memory usage: 1.1+ MB
```

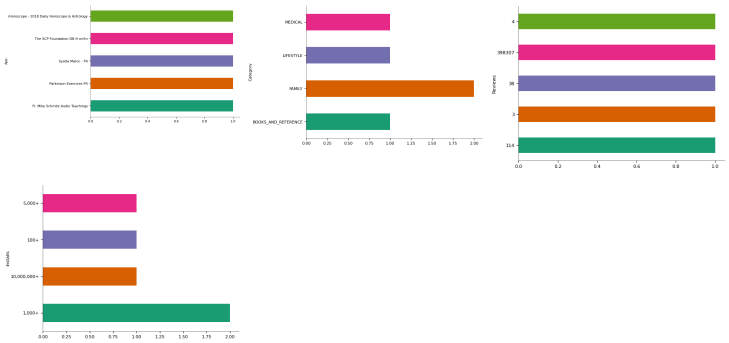
```
#below 5 values
df.tail()
```

	App	Category	Rating	Reviews	Size	Installs	T
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53000.000000	5,000+	F
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3600.000000	100+	F
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9500.000000	1,000+	F
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	21516.529524	1,000+	F
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19000.000000	10,000,000+	F

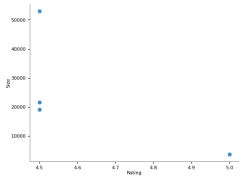
Distributions



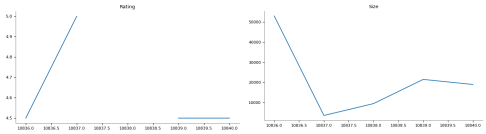
Categorical distributions



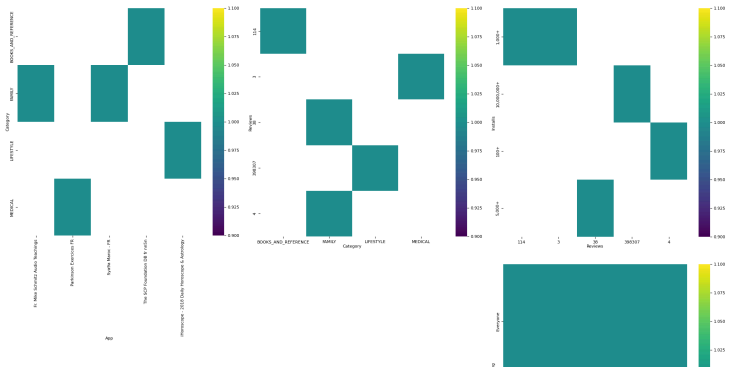
2-d distributions

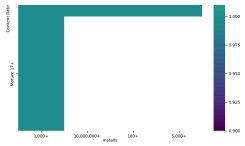


Values



2-d categorical distributions

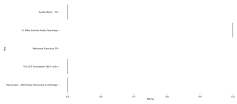




Faceted distributions

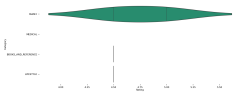
```
<string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0.
```



```
<string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0.
```



```
<string>:5: FutureWarning:
```

```
#Question 5 : Fill the missing values, if any, using the statistically relevant value
df.isnull().sum()
```

```
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=df.drop(df[df["Rating"].isnull()].index)
```

```
df.shape
```

```
(9367, 13)
```

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

```
App             0
Category        0
Rating          0
Reviews         0
Size            0
Installs        0
Type            0
Price           0
Content Rating  1
Genres          0
Last Updated    0
Current Ver     4
Android Ver     3
dtype: int64
```

```
df['Android Ver'].fillna(value='4.1 and up',inplace=True)
```

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

```
App             0
Category        0
Rating          0
Reviews         0
Size            0
Installs        0
Type            0
Price           0
```

```
Content Rating    1
Genres            0
Last Updated      0
Current Ver       4
Android Ver       0
dtype: int64
```

```
df['Android Ver'].value_counts()
```

```
Android Ver
4.1 and up            2062
Varies with device    1319
4.0.3 and up          1240
4.0 and up            1131
4.4 and up             875
2.3 and up             582
5.0 and up             535
4.2 and up             338
2.3.3 and up           240
3.0 and up             211
2.2 and up             208
4.3 and up             207
2.1 and up             113
1.6 and up              87
6.0 and up              48
7.0 and up              41
3.2 and up              31
2.0 and up              27
5.1 and up              18
1.5 and up              16
3.1 and up               8
2.0.1 and up             7
4.4W and up              6
8.0 and up               5
7.1 and up               3
4.0.3 - 7.1.1            2
5.0 - 8.0                2
1.0 and up               2
7.0 - 7.1.1              1
4.1 - 7.1.1              1
5.0 - 6.0                1
Name: count, dtype: int64
```

```
#Question 6 : Observe the Box-Plot of each feature
%matplotlib inline
plt.boxplot(df["Rating"])
```

```
{'whiskers': [<matplotlib.lines.Line2D at 0x7c8b42304700>,
  <matplotlib.lines.Line2D at 0x7c8b42306050>],
 'caps': [<matplotlib.lines.Line2D at 0x7c8b423049a0>,
  <matplotlib.lines.Line2D at 0x7c8b42304970>],
 'boxes': [<matplotlib.lines.Line2D at 0x7c8b42305990>],
 'medians': [<matplotlib.lines.Line2D at 0x7c8b42306d10>],
 'fliers': [<matplotlib.lines.Line2D at 0x7c8b42306fe0>],
 'means': []}
```