```
import pandas as pd
```

```
df = pd.read_csv("googleplaystore.csv")
```

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

```
rating = df['Rating']
print (rating)
print ('---------------------------')
print (rating.value_counts())
```

```
0        4.1
1        3.9
2        4.7
3        4.5
4        4.3
        ...
10836    4.5
10837    5.0
10838    NaN
10839    4.5
10840    4.5
Name: Rating, Length: 10841, dtype: float64
---------------------------
Rating
4.4     1109
4.3     1076
4.5     1038
4.2      952
4.6      823
4.1      708
4.0      568
4.7      499
3.9      386
3.8      303
5.0      274
3.7      239
4.8      234
3.6      174
3.5      163
3.4      128
3.3      102
```

```
4.9       87
3.0       83
3.1       69
3.2       64
2.9       45
2.8       42
2.6       25
2.7       25
2.5       21
2.3       20
2.4       19
1.0       16
2.2       14
1.9       13
2.0       12
1.7        8
1.8        8
2.1        8
1.6        4
1.4        3
1.5        3
1.2        1
19.0       1
Name: count, dtype: int64
```

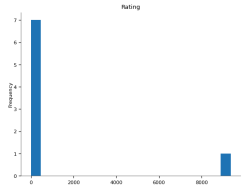
```
import pandas as pd
df = pd.read_csv("googleplaystore.csv")
current_ver = df['Current Ver']
print (current_ver)
print ('---------------------------')
print (current_ver.value_counts())
```

```
0                    1.0.0
1                    2.0.0
2                    1.2.4
3        Varies with device
4                      1.1
                ...
10836                 1.48
10837                  1.0
10838                  1.0
10839    Varies with device
10840    Varies with device
Name: Current Ver, Length: 10841, dtype: object
---------------------------
Current Ver
Varies with device    1459
1.0                    809
1.1                    264
1.2                    178
2.0                    151
                      ...
2.8.6                    1
1.25.4                   1
15                       1
1.022                    1
1.0.0.96                 1
Name: count, Length: 2832, dtype: int64
```

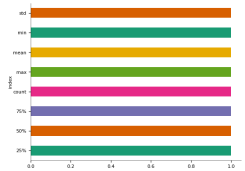
```
df.describe()
```

	Rating
count	9367.000000
mean	4.193338
std	0.537431
min	1.000000
25%	4.000000
50%	4.300000
75%	4.500000
max	19.000000

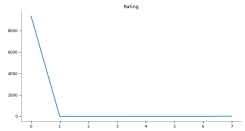
Distributions



Categorical distributions



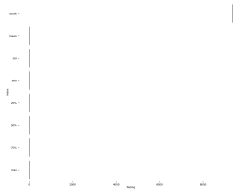
Values



Faceted distributions

```
<string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.
```



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

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

```
Price             0
Content Rating    0
Genres            0
Last Updated      0
Current Ver       8
Android Ver       2
dtype: int64
```

```
df.head(30)
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up
5	Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	March 26, 2017	1.0	2.3 and up
6	Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	April 26, 2018	1.1	4.0.3 and up
7	Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1,000,000+	Free	0	Everyone	Art & Design	June 14, 2018	6.1.61.1	4.2 and up
8	Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1,000,000+	Free	0	Everyone	Art & Design	September 20, 2017	2.9.2	3.0 and up
9	Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Design;Creativity	July 3, 2018	2.8	4.0.3 and up
10	Text on Photo - Fonteee	ART_AND_DESIGN	4.4	13880	28M	1,000,000+	Free	0	Everyone	Art & Design	October 27, 2017	1.0.4	4.1 and up
11	Name Art Photo Editor - Focus n Filters	ART_AND_DESIGN	4.4	8788	12M	1,000,000+	Free	0	Everyone	Art & Design	July 31, 2018	1.0.15	4.0 and up
12	Tattoo Name On My Photo Editor	ART_AND_DESIGN	4.2	44829	20M	10,000,000+	Free	0	Teen	Art & Design	April 2, 2018	3.8	4.1 and up
13	Mandala Coloring Book	ART_AND_DESIGN	4.6	4326	21M	100,000+	Free	0	Everyone	Art & Design	June 26, 2018	1.0.4	4.4 and up
14	3D Color Pixel by Number - Sandbox Art Coloring	ART_AND_DESIGN	4.4	1518	37M	100,000+	Free	0	Everyone	Art & Design	August 3, 2018	1.2.3	2.3 and up
15	Learn To Draw Kawaii Characters	ART_AND_DESIGN	3.2	55	2.7M	5,000+	Free	0	Everyone	Art & Design	June 6, 2018	NaN	4.2 and up
16	Photo Designer - Write your name with shapes	ART_AND_DESIGN	4.7	3632	5.5M	500,000+	Free	0	Everyone	Art & Design	July 31, 2018	3.1	4.1 and up
17	350 Diy Room Decor Ideas	ART_AND_DESIGN	4.5	27	17M	10,000+	Free	0	Everyone	Art & Design	November 7, 2017	1.0	2.3 and up
18	FlipaClip - Cartoon animation	ART_AND_DESIGN	4.3	194216	39M	5,000,000+	Free	0	Everyone	Art & Design	August 3, 2018	2.2.5	4.0.3 and up
19	ibis Paint X	ART_AND_DESIGN	4.6	224399	31M	10,000,000+	Free	0	Everyone	Art & Design	July 30, 2018	5.5.4	4.1 and up
20	Logo Maker - Small Business	ART_AND_DESIGN	4.0	450	14M	100,000+	Free	0	Everyone	Art & Design	April 20, 2018	4.0	4.1 and up
21	Boys Photo Editor - Six Pack & Men's Suit	ART_AND_DESIGN	4.1	654	12M	100,000+	Free	0	Everyone	Art & Design	March 20, 2018	1.1	4.0.3 and up
22	Superheroes Wallpapers | 4K Backgrounds	ART_AND_DESIGN	4.7	7699	4.2M	500,000+	Free	0	Everyone 10+	Art & Design	July 12, 2018	2.2.6.2	4.0.3 and up
23	Mcqueen Coloring pages	ART_AND_DESIGN	NaN	61	7.0M	100,000+	Free	0	Everyone	Art & Design;Action & Adventure	March 7, 2018	1.0.0	4.1 and up
24	HD Mickey Minnie Wallpapers	ART_AND_DESIGN	4.7	118	23M	50,000+	Free	0	Everyone	Art & Design	July 7, 2018	1.1.3	4.1 and up
25	Harley Quinn wallpapers HD	ART_AND_DESIGN	4.8	192	6.0M	10,000+	Free	0	Everyone	Art & Design	April 25, 2018	1.5	3.0 and up
26	Colorfit - Drawing & Coloring	ART_AND_DESIGN	4.7	20260	25M	500,000+	Free	0	Everyone	Art & Design;Creativity	October 11, 2017	1.0.8	4.0.3 and up

```
df.drop_duplicates(subset=['App'], keep='last', inplace=True)
```

```
len(df)
```

```
9660
```

```
df.dropna(subset=['Type', 'Content Rating'], inplace=True)
rating_mean = df['Rating'].mean()
df['Rating'].fillna(rating_mean, inplace=True)
print(df.isnull().sum())
```

```
App               0
Category          0
Rating            0
Reviews           0
Size              0
Installs          0
Type              0
Price             0
Content Rating    0
Genres            0
Last Updated      0
Current Ver       8
Android Ver       2
dtype: int64
/tmp/ipython-input-3223281348.py:3: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the origin


  df['Rating'].fillna(rating_mean, inplace=True)
```

```
df['Reviews'] = pd.to_numeric(df['Reviews'])
print(df['Reviews'].dtype)
```

```
int64
```

```
df.dropna(subset=['Current Ver', 'Android Ver'], inplace=True)
print(df.isnull().sum())
```

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

```
df['Installs'] = df['Installs'].astype(str).str.replace('[,+]', '', regex=True)
df['Installs'] = df['Installs'].astype(int)
print("Tipe data 'Installs' sekarang:", df['Installs'].dtype)
```

```
Tipe data 'Installs' sekarang: int64
```

```
df['Price'] = df['Price'].astype(str).str.replace('$', '', regex=False)
df['Price'] = pd.to_numeric(df['Price'])
print(df['Price'].dtype)
```

```
float64
```

```
def convert_size(size):
    if isinstance(size, str):
        if 'M' in size:
            return float(size.replace('M', ''))
        elif 'k' in size:
            return float(size.replace('k', '')) / 1024
        elif 'Varies with device' in size:
            return pd.NA
    return size
df['Size'] = df['Size'].apply(convert_size)

df['Size'] = pd.to_numeric(df['Size'])
print("Tipe data 'Size' sekarang:", df['Size'].dtype)

size_median = df['Size'].median()
df['Size'] = df['Size'].fillna(size_median)
```

```
Tipe data 'Size' sekarang: float64
```

```
import pandas as pd
df = pd.read_csv("googleplaystore.csv")
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  object 
 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(1), object(12)
memory usage: 1.1+ MB
```

```
df.drop_duplicates(subset=['App'], keep='last', inplace=True)
```

```
len(df)
```

```
9660
```

```
df.dropna(subset=['Type', 'Content Rating'], inplace=True)
rating_mean = df['Rating'].mean()
df['Rating'].fillna(rating_mean, inplace=True)
print(df.isnull().sum())
```

```
App              0
Category         0
Rating           0
Reviews          0
Size             0
Installs         0
Type             0
Price            0
Content Rating   0
Genres           0
Last Updated     0
Current Ver      8
Android Ver      2
dtype: int64
/tmp/ipython-input-3223281348.py:3: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the origin


  df['Rating'].fillna(rating_mean, inplace=True)
```

```
df.dropna(subset=['Current Ver', 'Android Ver'], inplace=True)
print(df.isnull().sum())
```

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

```
df['Installs'] = df['Installs'].astype(str).str.replace('[,+]', '', regex=True)
df['Installs'] = df['Installs'].astype(int)
print("Tipe data 'Installs' sekarang:", df['Installs'].dtype)
```

```
Tipe data 'Installs' sekarang: int64
```

```
df['Price'] = df['Price'].astype(str).str.replace('$', '', regex=False)
df['Price'] = pd.to_numeric(df['Price'])
print(df['Price'].dtype)
```

```
float64
```

```
def convert_size(size):
    if isinstance(size, str):
        if 'M' in size:
            return float(size.replace('M', ''))
        elif 'k' in size:
            return float(size.replace('k', '')) / 1024
        elif 'Varies with device' in size:
            return pd.NA
    return size
df['Size'] = df['Size'].apply(convert_size)
```

```
df['Size'] = pd.to_numeric(df['Size'])
print("Tipe data 'Size' sekarang:", df['Size'].dtype)

size_median = df['Size'].median()
df['Size'] = df['Size'].fillna(size_median)
```

```
Tipe data 'Size' sekarang: float64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 9658 entries, 0 to 10840
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   App             9658 non-null   object 
 1   Category        9658 non-null   object 
 2   Rating          9658 non-null   float64
 3   Reviews         9658 non-null   int64  
 4   Size            9658 non-null   object 
 5   Installs        9658 non-null   int64  
 6   Type            9658 non-null   object 
 7   Price           9658 non-null   float64
 8   Content Rating  9658 non-null   object 
 9   Genres          9658 non-null   object 
 10  Last Updated    9658 non-null   object 
 11  Current Ver     9650 non-null   object 
 12  Android Ver     9656 non-null   object 
dtypes: float64(2), int64(2), object(9)
memory usage: 1.0+ MB
```

```
df.to_csv('googleplaystore_clean.csv', index=False)
```