

EDA on Google Playstore

```
In [1]: ▶ import pandas as pd
```

```
In [2]: ▶ df=pd.read_csv(r"C:\Users\hp\Desktop\pandas projecrs\googleplaystore.csv")
```

```
In [4]: ▶ #1 Display the top 5 records .
df.head(5)
```

Out[4]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Al
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	
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	
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	
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
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

In [5]: **#2 Display the last 5 records**
`df.tail(5)`


Out[5]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Cu
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free	0	Everyone	Education	July 25, 2017	
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100+	Free	0	Everyone	Education	July 6, 2018	
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9.5M	1,000+	Free	0	Everyone	Medical	January 20, 2017	
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0	Mature 17+	Books & Reference	January 19, 2015	V de
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	Everyone	Lifestyle	July 25, 2018	V de

In [7]: **#3 Find the shape of the dataset (columns and rows)**
`df.shape`

```
print ("Number of Rows : ",df.shape[0])
print ("Number of Columns : ",df.shape[1])
```

Number of Rows : 10841
 Number of Columns : 13

In [9]:  *#4 Get the information of the data*
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
```

In [11]: **#5 Find the statistics of the data**
`df.describe (include='all')`

Out[11]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
count	10841	10841	9367.000000	10841	10841	10841	10840	10841	10840	10841	10841	10833	10838
unique	9660	34	NaN	6002	462	22	3	93	6	120	1378	2832	33
top	ROBLOX	FAMILY	NaN	0	Varies with device	1,000,000+	Free	0	Everyone	Tools	August 3, 2018	Varies with device	4.1 and up
freq	9	1972	NaN	596	1695	1579	10039	10040	8714	842	326	1459	2451
mean	NaN	NaN	4.193338	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	0.537431	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
min	NaN	NaN	1.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25%	NaN	NaN	4.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
50%	NaN	NaN	4.300000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	4.500000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
max	NaN	NaN	19.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [13]: **#6 Find the total number of App titles contain Astrology .**
`df.columns`

Out[13]: Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
 'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
 'Android Ver'],
 dtype='object')

```
In [15]: df['App']
```

```
Out[15]: 0          Photo Editor & Candy Camera & Grid & ScrapBook
1              Coloring book moana
2      U Launcher Lite - FREE Live Cool Themes, Hide ...
3              Sketch - Draw & Paint
4      Pixel Draw - Number Art Coloring Book
...
10836          Sya9a Maroc - FR
10837      Fr. Mike Schmitz Audio Teachings
10838      Parkinson Exercices FR
10839      The SCP Foundation DB fr nn5n
10840      iHoroscope - 2018 Daily Horoscope & Astrology
Name: App, Length: 10841, dtype: object
```

```
In [23]: df[df['App'].str.contains('Astrology')]
```

```
Out[23]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
1570	Horoscopes – Daily Zodiac Horoscope and Astrology	LIFESTYLE	4.6	161143	11M	10,000,000+	Free	0	Everyone 10+	Lifestyle	June 25, 2018	5.2.4(881)	4.0.3 and up
1592	آسٲٲٲٲ Astrology - Min Thein Kha BayDin	LIFESTYLE	4.7	2225	15M	100,000+	Free	0	Everyone	Lifestyle	July 26, 2018	4.2.1	4.0.3 and up
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	Everyone	Lifestyle	July 25, 2018	Varies with device	Varies with device

```
In [24]: len(df[df['App'].str.contains('Astrology')])
```

```
Out[24]: 3
```

```
In [27]: ▶ #7 Find Average App rating  
df['Rating'].mean()
```

```
Out[27]: 4.193338315362443
```

```
In [32]: ▶ #8 . Find total number of Unique category  
df['Category'].nunique()
```

```
Out[32]: 34
```

```
In [37]: ▶ #9 .Which category getting the highest average rating .  
df.groupby('Category')['Rating'].mean().sort_values(ascending=False).head(5)
```

```
Out[37]: Category  
1.9          19.000000  
EVENTS       4.435556  
EDUCATION    4.389032  
ART_AND_DESIGN 4.358065  
BOOKS_AND_REFERENCE 4.346067  
Name: Rating, dtype: float64
```

```
In [41]: ▶ #10 . Total Number of apps having 5 star rating .  
len(df[df['Rating']==5.0])
```

```
Out[41]: 274
```

```
In [45]: ▶ #11. Find average value of reviews .  
df['Reviews'].dtype
```

```
Out[45]: dtype('O')
```

```
In [47]: ▶ df['Reviews'].astype('float')
```

ValueError

Traceback (most recent call last)

Cell In[47], line 1

```
----> 1 df['Reviews'].astype('float')
```

File ~\anaconda3\lib\site-packages\pandas\core\generic.py:6240, in NDFrame.astype(self, dtype, copy, errors)

```

6233     results = [
6234         self.iloc[:, i].astype(dtype, copy=copy)
6235         for i in range(len(self.columns))
6236     ]
6237 else:
6238     # else, only a single dtype is given
-> 6240     new_data = self._mgr.astype(dtype=dtype, copy=copy, errors=errors)
6241     return self._constructor(new_data).__finalize__(self, method="astype")
6242 # GH 33113: handle empty frame or series

```

File ~\anaconda3\lib\site-packages\pandas\core\internals\managers.py:448, in BaseBlockManager.astype(self, dtype, copy, errors)

```

447 def astype(self: T, dtype, copy: bool = False, errors: str = "raise") -> T:
--> 448     return self.apply("astype", dtype=dtype, copy=copy, errors=errors)

```

File ~\anaconda3\lib\site-packages\pandas\core\internals\managers.py:352, in BaseBlockManager.apply(self, f, align_keys, ignore_failures, **kwargs)

```

350     applied = b.apply(f, **kwargs)
351     else:
--> 352     applied = getattr(b, f)(**kwargs)
353 except (TypeError, NotImplementedError):
354     if not ignore_failures:

```

File ~\anaconda3\lib\site-packages\pandas\core\internals\blocks.py:526, in Block.astype(self, dtype, copy, errors)

```

508 """
509 Coerce to the new dtype.
510 (...)
522 Block
523 """
524 values = self.values
--> 526 new_values = astype_array_safe(values, dtype, copy=copy, errors=errors)
528 new_values = maybe_coerce_values(new_values)
529 newb = self.make_block(new_values)

```



```
File ~\anaconda3\lib\site-packages\pandas\core\dtypes\astype.py:299, in astype_array_safe(values, dtype, copy, errors)
    296     return values.copy()
    298 try:
--> 299     new_values = astype_array(values, dtype, copy=copy)
    300 except (ValueError, TypeError):
```

```
    301     # e.g. astype_nansafe can fail on object-dtype of strings
    302     # trying to convert to float
    303     if errors == "ignore":
```

```
File ~\anaconda3\lib\site-packages\pandas\core\dtypes\astype.py:230, in astype_array(values, dtype, copy)
```

```
    227     values = values.astype(dtype, copy=copy)
    229 else:
--> 230     values = astype_nansafe(values, dtype, copy=copy)
    232 # in pandas we don't store numpy str dtypes, so convert to object
    233 if isinstance(dtype, np.dtype) and issubclass(values.dtype.type, str):
```

```
File ~\anaconda3\lib\site-packages\pandas\core\dtypes\astype.py:170, in astype_nansafe(arr, dtype, copy, skipna)
```

```
    166     raise ValueError(msg)
    168 if copy or is_object_dtype(arr.dtype) or is_object_dtype(dtype):
    169     # Explicit copy, or required since NumPy can't view from / to object.
--> 170     return arr.astype(dtype, copy=True)
    172 return arr.astype(dtype, copy=copy)
```

ValueError: could not convert string to float: '3.0M'

```
In [48]: ▶ df[df['Reviews']=='3.0M']
```

Out[48]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
10472	Life Made WI-Fi Touchscreen Photo Frame	1.9	19.0	3.0M	1,000+	Free	0	Everyone	NaN	February 11, 2018	1.0.19	4.0 and up	NaN

```
In [53]: ▶ df['Reviews']=df['Reviews'].replace('3.0M',3.0)
```

```
In [54]: ▶ df['Reviews']=df['Reviews'].astype('float')
```

```
In [55]: ▶ df['Reviews'].dtype
```

```
Out[55]: dtype('float64')
```

```
In [56]: ▶ df['Reviews'].mean()
```

```
Out[56]: 444111.9265750392
```

```
In [63]: ▶ #12. Total Number of free and paid apps
df['Type'].value_counts()
print("total number of Free apps : ",df['Type'].value_counts()[0])
print("total number of Paid apps : ",df['Type'].value_counts()[1])
```

```
total number of Free apps : 10039
total number of Paid apps : 800
```

```
In [75]: ▶ #13. Which apps has Maximum reviews .
df[df['Reviews'].max()==df['Reviews']]
```

```
Out[75]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
2544	Facebook	SOCIAL	4.1	78158306.0	Varies with device	1,000,000,000+	Free	0	Teen	Social	August 3, 2018	Varies with device	Varies with device

```
In [ ]: ▶
```

```
In [ ]: ▶
```

```
In [ ]: ▶
```

In []: ▶

In []: ▶

In []: ▶

In []: ▶

In []: ▶

In []: ▶

In [96]: ▶

In []: ▶

In []: ▶