

```
In [10]: import pandas as pd
import numpy as np
```

```
In [30]: # 1. import the dataset
df = pd.read_csv('youtube.csv')
```

```
In [31]: df.head()
```

```
Out[31]:
```

	Rank	Grade	Ch_name	Uploads	Subscriptions	Views
0	1st	A++	T-Series	14,297	135M	104,724,369,854
1	2nd	A++	Cocomelon - Nursery Rhymes	517	78.2M	57,054,290,512
2	3rd	A++	🌟 Kids Diana Show	691	50.9M	24,157,678,368
3	4th	A++	Like Nastya	400	52.2M	30,591,257,306
4	5th	A++	SET India	37,017	69.3M	52,149,505,781

2. print the data types

```
In [32]: #converting columns to numeric type
df["Subscriptions"]=df["Subscriptions"].str.extract(r'(\d+)')
df["Subscriptions"]=pd.to_numeric(df["Subscriptions"])
df["Uploads"]=df["Uploads"].str.replace(',','')
df["Uploads"]=pd.to_numeric(df["Uploads"])
df["Views"]=df["Views"].str.replace(',','')
df["Views"]=df["Views"].str.replace('--','')
df["Rank"]=df["Rank"].str.extract(r'(\d+)')
df["Views"]=pd.to_numeric(df["Views"])
```

3. print first 10 records

```
In [14]: df.head(3)
```

```
Out[14]:
```

	Rank	Grade	Ch_name	Uploads	Subscriptions	Views
0	1st	A++	T-Series	14297.0	135	1.047244e+11
1	2nd	A++	Cocomelon - Nursery Rhymes	517.0	78	5.705429e+10
2	3rd	A++	🌟 Kids Diana Show	691.0	50	2.415768e+10

4. print names of the columns

```
In [16]: df.columns
```

```
Out[16]: Index(['Rank', 'Grade', 'Ch_name', 'Uploads', 'Subscriptions', 'Views'], dtype='object')
```

5. print details of the channels T-Series,SAB TV and Zee TV

```
In [23]: df[(df["Ch_name"]=="T-Series") | (df["Ch_name"]=="SAB TV") | (df["Ch_name"]=="Zee TV") ]
```

```
Out[23]:
```

	Rank	Grade	Ch_name	Uploads	Subscriptions	Views
0	1st	A++	T-Series	14297.0	135	1.047244e+11
7	8th	A++	Zee TV	98621.0	43	4.154426e+10
10	11th	A++	SAB TV	23812.0	29	2.559749e+10

6. print the channels where subscription is between 50 and 100 million

```
In [24]: df[(df["Subscriptions"] >= 50) & (df["Subscriptions"]<=100)]
```

Out[24]:

	Rank	Grade	Ch_name	Uploads	Subscriptions	Views
1	2nd	A++	Cocomelon - Nursery Rhymes	517.0	78	5.705429e+10
2	3rd	A++	🌟 Kids Diana Show	691.0	50	2.415768e+10
3	4th	A++	Like Nastya	400.0	52	3.059126e+10
4	5th	A++	SET India	37017.0	69	5.214951e+10
18	19th	A+	WWE	47028.0	57	4.129950e+10
25	26th	A+	Zee Music Company	4693.0	53	2.507553e+10
75	76th	A	Canal KondZilla	1387.0	56	2.941463e+10
100	101st	A	Go Turkey	405.0	52	5.505536e+08
218	219th	A	FlyntofRWBY	64.0	94	1.041827e+08
220	221st	A	5-Minute Crafts	4011.0	65	1.748778e+10
240	241st	A	Walls Thailand	117.0	55	1.024819e+08
266	267th	A	PhonePe	160.0	90	1.513958e+09
390	391st	A	Dude Perfect	229.0	50	1.031170e+10
400	401st	A	뮤지컬웨딩MusicalWedding	3055.0	56	7.379514e+07
488	489th	A	lester villegas	34.0	72	6.370011e+07

7.What are the top 10 youtube channels as ordered by ‘subscriptions in million’?**

In [26]:

```
df.sort_values("Subscriptions", ascending=False).head(10)
```

Out[26]:

	Rank	Grade	Ch_name	Uploads	Subscriptions	Views
176	177th	A	بيجي بالعربي	113.0	991	1.418034e+08
401	402nd	A	Телеканал Звезда	2663.0	990	1.816934e+08
308	309th	A	3D Music India	245.0	960	8.575662e+07
339	340th	A	VSRAP	147.0	947	8.403346e+07
257	258th	A	Siyah Giyen Genç	599.0	947	2.127583e+08
61	62nd	A	MUSIC BANGLA TV	328.0	943	2.270196e+08
149	150th	A	SO LY DA	27.0	908	2.270388e+08
265	266th	A	BillieEilishVEVO	56.0	876	4.369175e+09
38	39th	A+	Odia E News	1466.0	838	2.650261e+08
437	438th	A	tvN D CLASSIC	9343.0	825	1.396809e+09

8. Generate a table which only contains data for channels which have less than 5000 video uploads.
Sort this table by Rank.**

```
In [33]: (df[df["Uploads"]<5000]).sort_values("Rank",ascending = True)
```

Out[33]:

	Rank	Grade	Ch_name	Uploads	Subscriptions	Views
9	10	A++	Vlad and Nikita	219.0	37	1.807406e+10
99	100	A	LooLoo Kids TV	32.0	738	1.840053e+08
100	101	A	Go Turkey	405.0	52	5.505536e+08
102	103	A	DisneyMusicVEVO	1162.0	18	1.196514e+10
103	104	A	ABCKids Misa	47.0	3	6.491976e+08
...
90	91	A	BabyBus - Canciones Infantiles & C...	574.0	11	4.841513e+09
91	92	A	Venus	4216.0	21	7.778258e+09
93	94	A	shfa show India	90.0	3	1.197781e+09
96	97	A	Jkk Entertainment	127.0	15	5.655917e+09
97	98	A	LETRAS RD	2.0	250	1.803838e+08

373 rows × 6 columns

9. Print the average number of subscribers per grade type

```
In [41]: df1= df.groupby(["Grade"])[ "Subscriptions"].mean()
df1
```

```
Out[41]: Grade
A      75.404598
A+     40.061224
A++    48.818182
Name: Subscriptions, dtype: float64
```

10. What are the top 5 channels which belong to grade A++ based on Video_views?

```
In [34]: df[df["Grade"]=="A++"].sort_values("Views",ascending=False).head(5)
```

Out[34]:

	Rank	Grade	Ch_name	Uploads	Subscriptions	Views
0	1	A++	T-Series	14297.0	135	1.047244e+11
1	2	A++	Cocomelon - Nursery Rhymes	517.0	78	5.705429e+10
4	5	A++	SET India	37017.0	69	5.214951e+10
7	8	A++	Zee TV	98621.0	43	4.154426e+10
6	7	A++	Movieclips	35226.0	36	3.505581e+10