

Importing the Required Libraries & Dataset

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

data=pd.read_csv("Top-5000-Youtube-Channels.csv")
```

Display all the Rows except last 5 Rows

```
data.head(-5)
```

	Rank	Grade	Channel name	Video Uploads \
0	1st	A++	Zee TV	82757
1	2nd	A++	T-Series	12661
2	3rd	A++	Cocomelon - Nursery Rhymes	373
3	4th	A++	SET India	27323
4	5th	A++	WWE	36756
...	...	...	...	...
4990	4,991st	B+	Ho Ngoc Ha's Official Channel	208
4991	4,992nd	B+	Toys to Learn Colors	11
4992	4,993rd	B+	KAZKA	25
4993	4,994th	B+	United CUBE (CUBE Entertainment...	1055
4994	4,995th	B+	Wings Marathi	1735

	Subscribers	Video views
0	18752951	20869786591
1	61196302	47548839843
2	19238251	9793305082
3	31180559	22675948293
4	32852346	26273668433
...	...	...
4990	--	127185704
4991	663114	141933264
4992	131766	74304638
4993	1586835	371299166
4994	1099659	346175699

```
[4995 rows x 6 columns]
```

Display all the rows except First 5 Rows

```
data.tail(-5)
```

	Rank	Grade	Channel name	Video Uploads
Subscribers \				
5	6th	A++	Movieclips	30243 17149705

6	7th	A++	netd müzik	8500	11373567
7	8th	A++	ABS-CBN Entertainment	100147	12149206
8	9th	A++	Ryan ToysReview	1140	16082927
9	10th	A++	Zee Marathi	74607	2841811
...	...	...	...	...	...
4995	4,996th	B+	Uras Benlioğlu	706	2072942
4996	4,997th	B+	HI-TECH MUSIC LTD	797	1055091
4997	4,998th	B+	Mastersaint	110	3265735
4998	4,999th	B+	Bruce McIntosh	3475	32990
4999	5,000th	B+	SehatAQUA	254	21172

	Video views
5	16618094724
6	23898730764
7	17202609850
8	24518098041
9	2591830307
...	...
4995	441202795
4996	377331722
4997	311758426
4998	14563764
4999	73312511

[4995 rows x 6 columns]

Finding the Shape of the Dataset[No. of Rows & Columns]

```
print("The No. of the Rows:",data.shape[0])
print("The No. of the Columns:",data.shape[1])
```

The No. of the Rows: 5000
The No. of the Columns: 6

Getting all the Info of the Dataset

```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
```

```
Data columns (total 6 columns):
#      Column      Non-Null Count  Dtype
---  -
0     Rank         5000 non-null    object
1     Grade         5000 non-null    object
2     Channel name   5000 non-null    object
3     Video Uploads  5000 non-null    object
4     Subscribers    5000 non-null    object
5     Video views     5000 non-null    int64
dtypes: int64(1), object(5)
memory usage: 234.5+ KB
```

Getting the Overall Statistics of the Dataset

```
data.describe()

           Video views
count  5.000000e+03
mean   1.071449e+09
std    2.003844e+09
min    7.500000e+01
25%    1.862329e+08
50%    4.820548e+08
75%    1.124368e+09
max    4.754884e+10
```

Checking the Null Values in the Dataset

```
data.isnull().sum()

Rank         0
Grade        0
Channel name  0
Video Uploads  0
Subscribers   0
Video views   0
dtype: int64
```

Data Cleaning[Replace '--' to NaN]

```
data=data.replace("--",np.nan,regex=True)
```

Data Cleaning[Rank Column]

```
data.head(20)
```

	Rank	Grade	Channel name	Video Uploads	Subscribers \
0	1st	A++	Zee TV	82757	18752951

1	2nd	A++	T-Series	12661	61196302
2	3rd	A++	Cocomelon - Nursery Rhymes	373	19238251
3	4th	A++	SET India	27323	31180559
4	5th	A++	WWE	36756	32852346
5	6th	A++	Movieclips	30243	17149705
6	7th	A++	netd müzik	8500	11373567
7	8th	A++	ABS-CBN Entertainment	100147	12149206
8	9th	A++	Ryan ToysReview	1140	16082927
9	10th	A++	Zee Marathi	74607	2841811
10	11th	A+	5-Minute Crafts	2085	33492951
11	12th	A+	Canal KondZilla	822	39409726
12	13th	A+	Like Nastya Vlog	150	7662886
13	14th	A+	Ozuna	50	18824912
14	15th	A+	Wave Music	16119	15899764
15	16th	A+	Ch3Thailand	49239	11569723
16	17th	A+	WORLDSTARHIPHOP	4778	15830098
17	18th	A+	Vlad and Nikita	53	NaN
18	19th	A+	Badabun	3060	23603062
19	20th	A+	WorkpointOfficial	24287	17687229
Video views					
0	20869786591				
1	47548839843				
2	9793305082				
3	22675948293				
4	26273668433				
5	16618094724				
6	23898730764				
7	17202609850				
8	24518098041				
9	2591830307				
10	8587520379				

```

11 19291034467
12 2540099931
13 8727783225
14 10989179147
15 9388600275
16 11102158475
17 1428274554
18 5860444053
19 14022189654

```

```
data["Rank"]=data["Rank"].str[:-2]
```

```
data.head(20)
```

	Rank	Grade	Channel name	Video Uploads	Subscribers	\
0	1	A++	Zee TV	82757	18752951	
1	2	A++	T-Series	12661	61196302	
2	3	A++	Cocomelon - Nursery Rhymes	373	19238251	
3	4	A++	SET India	27323	31180559	
4	5	A++	WWE	36756	32852346	
5	6	A++	Movieclips	30243	17149705	
6	7	A++	netd müzik	8500	11373567	
7	8	A++	ABS-CBN Entertainment	100147	12149206	
8	9	A++	Ryan ToysReview	1140	16082927	
9	10	A++	Zee Marathi	74607	2841811	
10	11	A+	5-Minute Crafts	2085	33492951	
11	12	A+	Canal KondZilla	822	39409726	
12	13	A+	Like Nastya Vlog	150	7662886	
13	14	A+	Ozuna	50	18824912	
14	15	A+	Wave Music	16119	15899764	
15	16	A+	Ch3Thailand	49239	11569723	
16	17	A+	WORLDSTARHIPHOP	4778	15830098	
17	18	A+	Vlad and Nikita	53	NaN	
18	19	A+	Badabun	3060	23603062	
19	20	A+	WorkpointOfficial	24287	17687229	

```

Video views
0 20869786591
1 47548839843
2 9793305082
3 22675948293
4 26273668433
5 16618094724
6 23898730764
7 17202609850
8 24518098041
9 2591830307
10 8587520379
11 19291034467
12 2540099931

```

```

13 8727783225
14 10989179147
15 9388600275
16 11102158475
17 1428274554
18 5860444053
19 14022189654

```

```
data["Rank"]=data["Rank"].str.replace(",","")
```

```
data["Rank"]=data["Rank"].astype('int')
```

```
data.dtypes
```

```

Rank          int32
Grade         object
Channel name  object
Video Uploads object
Subscribers   object
Video views   int64
dtype: object

```

Data Cleaning[Video Uploads & Subscribers]

```
data["Video Uploads"]=data["Video Uploads"].astype('float')
```

```
data["Subscribers"]=data["Subscribers"].astype('float')
```

Data Cleaning[Grade column]

```
data["Grade"].value_counts()
```

```

Grade
B+      2956
A-      1024
A        963
A+       41
A++      10
        6

```

```
Name: count, dtype: int64
```

```

data["Grade"] = data["Grade"].str.strip()
x= data["Grade"].map({"A++": "1", "A+": "2", "A": "3", "A-": "4",
"B+": "5"})
data.insert(2,"Grade_Values",x)

```

```
data.head(20)
```

	Rank	Grade	Grade_Values	Channel name	Video Uploads
0	1	A++	1	Zee TV	82757.0

1	2	A++	1	T-Series	12661.0
2	3	A++	1	Cocomelon - Nursery Rhymes	373.0
3	4	A++	1	SET India	27323.0
4	5	A++	1	WWE	36756.0
5	6	A++	1	Movieclips	30243.0
6	7	A++	1	netd müzik	8500.0
7	8	A++	1	ABS-CBN Entertainment	100147.0
8	9	A++	1	Ryan ToysReview	1140.0
9	10	A++	1	Zee Marathi	74607.0
10	11	A+	2	5-Minute Crafts	2085.0
11	12	A+	2	Canal KondZilla	822.0
12	13	A+	2	Like Nastya Vlog	150.0
13	14	A+	2	Ozuna	50.0
14	15	A+	2	Wave Music	16119.0
15	16	A+	2	Ch3Thailand	49239.0
16	17	A+	2	WORLDSTARHIPHOP	4778.0
17	18	A+	2	Vlad and Nikita	53.0
18	19	A+	2	Badabun	3060.0
19	20	A+	2	WorkpointOfficial	24287.0
	Subscribers	Video views			
0	18752951.0	20869786591			
1	61196302.0	47548839843			
2	19238251.0	9793305082			
3	31180559.0	22675948293			
4	32852346.0	26273668433			
5	17149705.0	16618094724			
6	11373567.0	23898730764			
7	12149206.0	17202609850			
8	16082927.0	24518098041			
9	2841811.0	2591830307			

```

10  33492951.0    8587520379
11  39409726.0   19291034467
12   7662886.0    2540099931
13  18824912.0    8727783225
14  15899764.0   10989179147
15  11569723.0    9388600275
16  15830098.0   11102158475
17         NaN    1428274554
18  23603062.0    5860444053
19  17687229.0   14022189654

```

```
data["Grade_Values"]=data["Grade_Values"].astype('float')
```

Find the Average Views for each Channel

```
data.dtypes
```

```

Rank          int32
Grade          object
Grade_Values   float64
Channel name   object
Video Uploads  float64
Subscribers    float64
Video views    int64
dtype: object

```

```
data["Avg_Views"]=data["Video views"]/data["Video Uploads"]
```

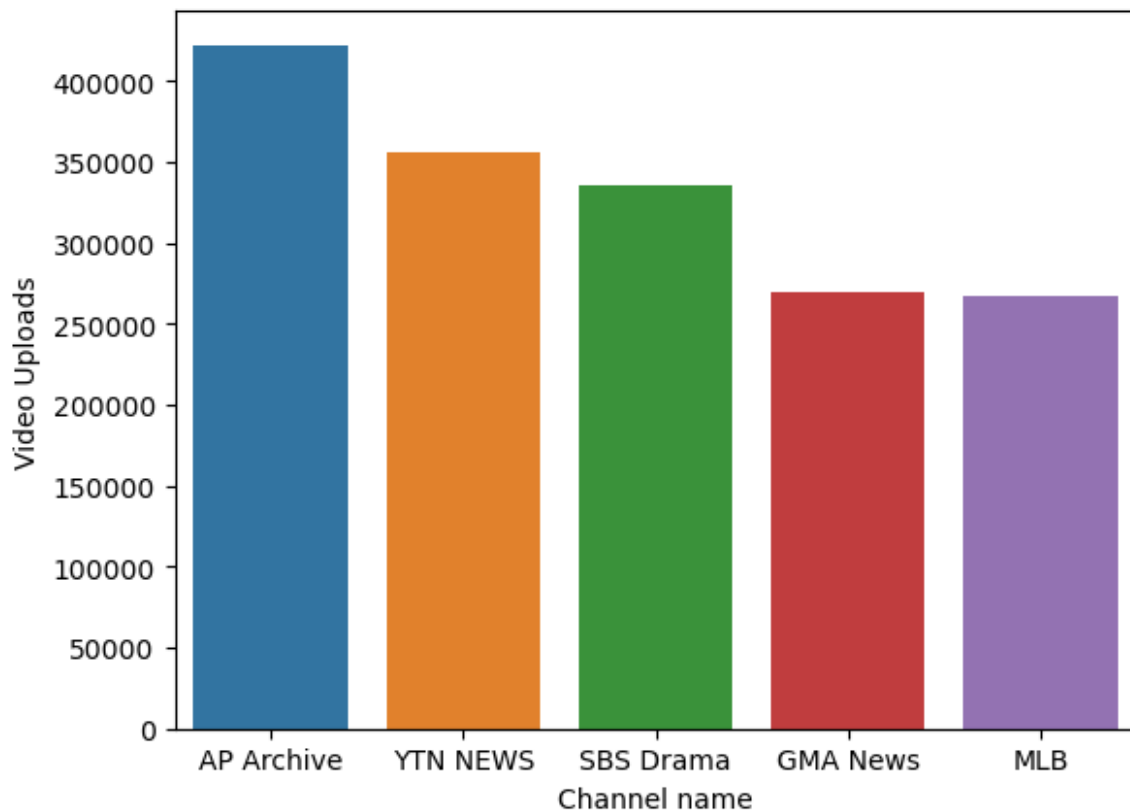
```
data.head()
```

	Rank	Grade	Grade_Values	Channel name	Video Uploads
0	1	A++	1.0	Zee TV	82757.0
1	2	A++	1.0	T-Series	12661.0
2	3	A++	1.0	Cocomelon - Nursery Rhymes	373.0
3	4	A++	1.0	SET India	27323.0
4	5	A++	1.0	WWE	36756.0

	Subscribers	Video views	Avg_Views
0	18752951.0	20869786591	2.521815e+05
1	61196302.0	47548839843	3.755536e+06
2	19238251.0	9793305082	2.625551e+07
3	31180559.0	22675948293	8.299216e+05
4	32852346.0	26273668433	7.148130e+05

Find the top 5 Channels with the maximum no. of Video Uploads

```
top=data.sort_values(by="Video Uploads",ascending=False).head()
sns.barplot(x="Channel name",y="Video Uploads",data=top)
<Axes: xlabel='Channel name', ylabel='Video Uploads'>
```



Find the Correlation Matrix

```
numeric_data = data.select_dtypes(include='number')
correlation_matrix = numeric_data.corr()
correlation_matrix
```

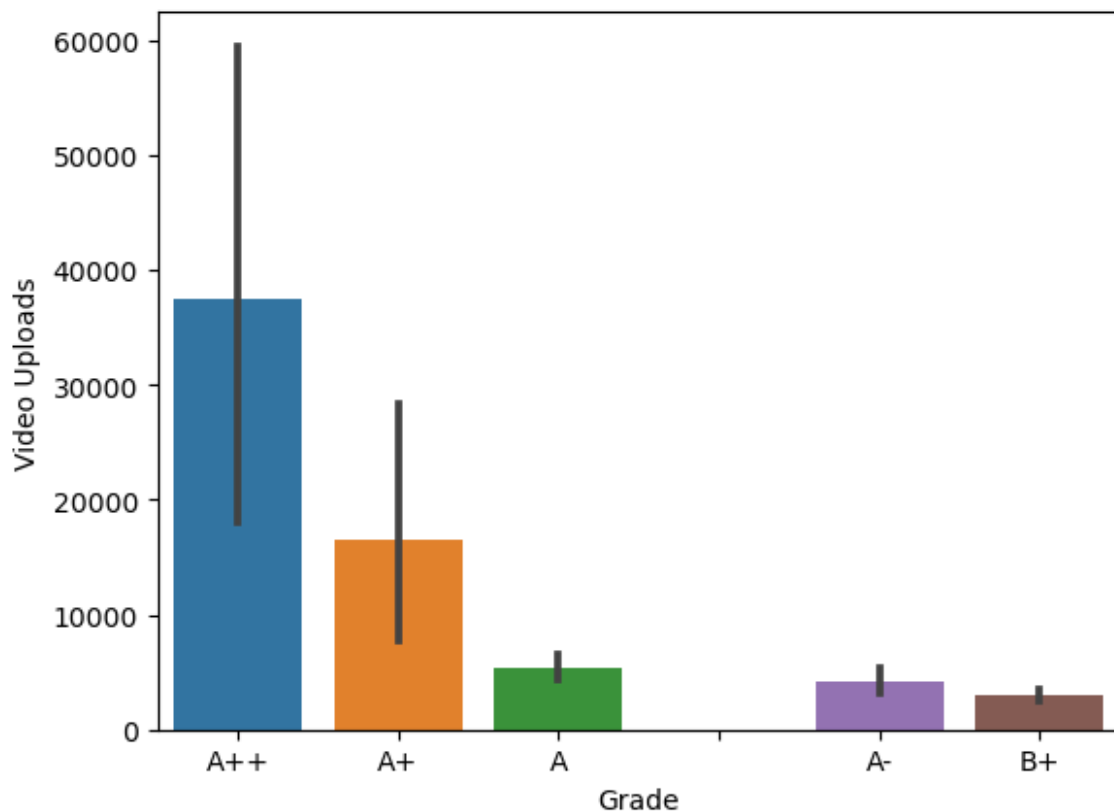
	Rank	Grade_Values	Video Uploads	Subscribers	\
Rank	1.000000	0.863952	-0.069139	-0.383417	
Grade_Values	0.863952	1.000000	-0.085451	-0.429213	
Video Uploads	-0.069139	-0.085451	1.000000	0.011362	
Subscribers	-0.383417	-0.429213	0.011362	1.000000	
Video views	-0.390028	-0.463834	0.092436	0.791281	
Avg_Views	-0.150529	-0.154399	-0.062063	0.289386	
	Video views	Avg_Views			
Rank	-0.390028	-0.150529			
Grade_Values	-0.463834	-0.154399			

Video Uploads	0.092436	-0.062063
Subscribers	0.791281	0.289386
Video views	1.000000	0.297096
Avg_Views	0.297096	1.000000

Which Grade has the Highest Video Uploads

```
sns.barplot(x="Grade",y="Video Uploads",data=data)
```

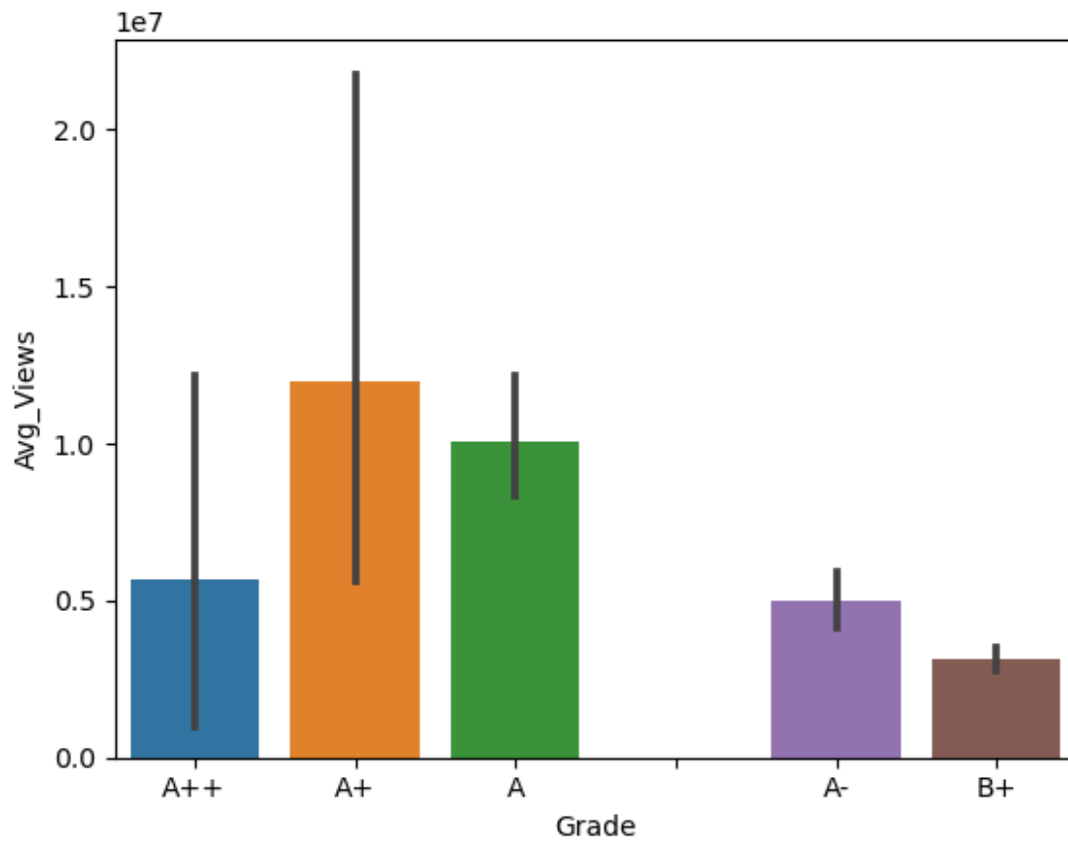
```
<Axes: xlabel='Grade', ylabel='Video Uploads'>
```



Which Grade has the highest Average Views

```
sns.barplot(x="Grade",y="Avg_Views",data=data)
```

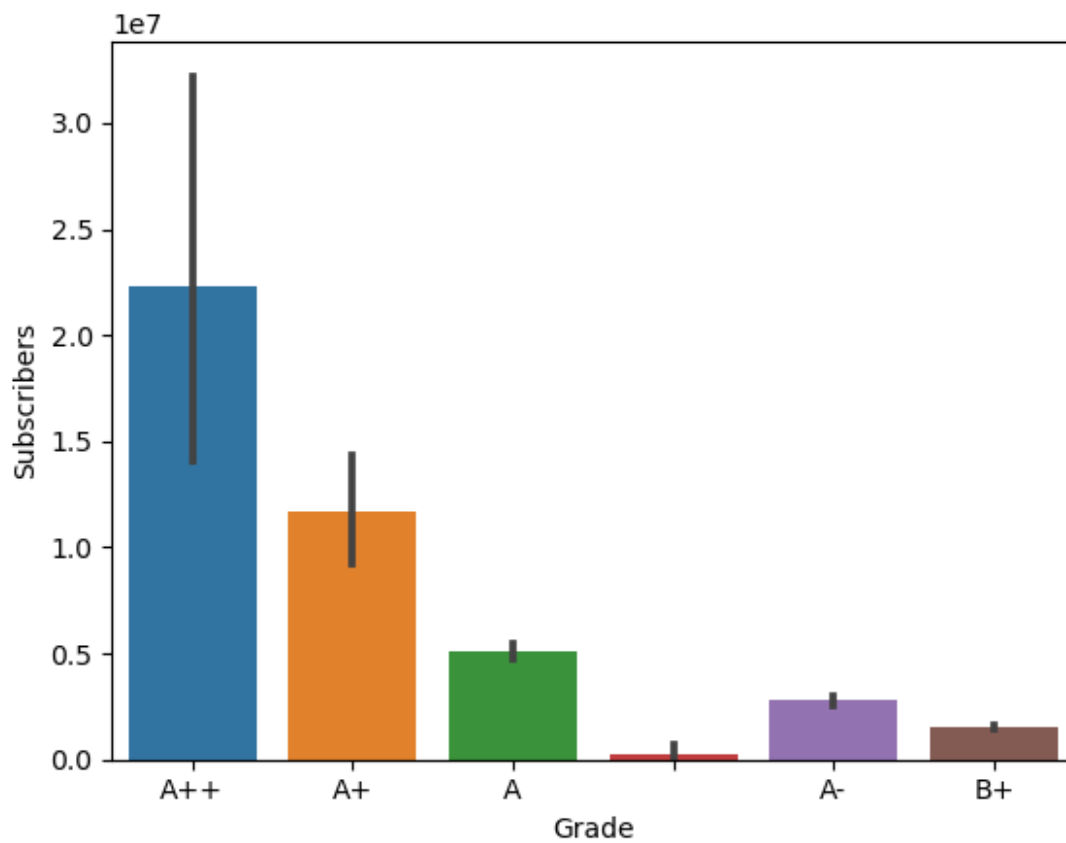
```
<Axes: xlabel='Grade', ylabel='Avg_Views'>
```



Which grade has the highest no. of the Subscribers

```
sns.barplot(x="Grade",y="Subscribers",data=data)
```

```
<Axes: xlabel='Grade', ylabel='Subscribers'>
```



---Alhamdulillah---