```
In [100]: import pandas as pd
          import seaborn as sns
```

```
In [101]: data = pd.read_csv('top-5000-youtube-channels.csv')
```

```
In [102]: data.head()   #to see the first five rows
```

Out[102]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views
0	1st	A++	Zee TV	82757	18752951	20869786591
1	2nd	A++	T-Series	12661	61196302	47548839843
2	3rd	A++	Cocomelon - Nursery Rhymes	373	19238251	9793305082
3	4th	A++	SET India	27323	31180559	22675948293
4	5th	A++	WWE	36756	32852346	26273668433

```
In [103]: data.tail(-5)    # all rows except first five rows
```

Out[103]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views
5	6th	A++	Movieclips	30243	17149705	16618094724
6	7th	A++	netd müzik	8500	11373567	23898730764
7	8th	A++	ABS-CBN Entertainment	100147	12149206	17202609850
8	9th	A++	Ryan ToysReview	1140	16082927	24518098041
9	10th	A++	Zee Marathi	74607	2841811	2591830307
...	...	...	...	...	...	...
4995	4,996th	B+	Uras Benlioğlu	706	2072942	441202795
4996	4,997th	B+	HI-TECH MUSIC LTD	797	1055091	377331722
4997	4,998th	B+	Mastersaint	110	3265735	311758426
4998	4,999th	B+	Bruce McIntosh	3475	32990	14563764
4999	5,000th	B+	SehatAQUA	254	21172	73312511

4995 rows × 6 columns

```
In [104]: data.shape      #to print   numbers o frows and columns
```

Out[104]: (5000, 6)

```
In [105]: print("number of rows: ", data.shape[0])
          print("number of columns: ", data.shape[1])
```

```
number of rows:  5000
number of columns:  6
```

```
In [106]: data.info()    #use to  get iinfo about dataset like total no. of rows , columns , datatypes of each column and memory requirement
```

```
<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
```

```
In [107]: pd.set_option('display.float_format',  '{:,}'.format)  #creating exponential values to float values
```

In [108]: `data.describe()`

Out[108]:

	Video views
count	5,000.0
mean	1,071,449,400.15
std	2,003,843,972.1239882
min	75.0
25%	186,232,945.75
50%	482,054,780.0
75%	1,124,367,826.75
max	47,548,839,843.0

In [109]: `data.head(20)`

Out[109]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views
0	1st	A++	Zee TV	82757	18752951	20869786591
1	2nd	A++	T-Series	12661	61196302	47548839843
2	3rd	A++	Cocomelon - Nursery Rhymes	373	19238251	9793305082
3	4th	A++	SET India	27323	31180559	22675948293
4	5th	A++	WWE	36756	32852346	26273668433
5	6th	A++	Movieclips	30243	17149705	16618094724
6	7th	A++	netd müzik	8500	11373567	23898730764
7	8th	A++	ABS-CBN Entertainment	100147	12149206	17202609850
8	9th	A++	Ryan ToysReview	1140	16082927	24518098041
9	10th	A++	Zee Marathi	74607	2841811	2591830307
10	11th	A+	5-Minute Crafts	2085	33492951	8587520379
11	12th	A+	Canal KondZilla	822	39409726	19291034467
12	13th	A+	Like Nastya Vlog	150	7662886	2540099931
13	14th	A+	Ozuna	50	18824912	8727783225
14	15th	A+	Wave Music	16119	15899764	10989179147
15	16th	A+	Ch3Thailand	49239	11569723	9388600275
16	17th	A+	WORLDSTARHIPHOP	4778	15830098	11102158475
17	18th	A+	Vlad and Nikita	53	--	1428274554
18	19th	A+	Badabun	3060	23603062	5860444053
19	20th	A+	WorkpointOfficial	24287	17687229	14022189654

In [110]:
```python
import numpy as np
data = data.replace('--',np.nan,regex=True)   #replacing -- values with Nan
data.head(20)
```

Out[110]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views
0	1st	A++	Zee TV	82757	18752951	20869786591
1	2nd	A++	T-Series	12661	61196302	47548839843
2	3rd	A++	Cocomelon - Nursery Rhymes	373	19238251	9793305082
3	4th	A++	SET India	27323	31180559	22675948293
4	5th	A++	WWE	36756	32852346	26273668433
5	6th	A++	Movieclips	30243	17149705	16618094724
6	7th	A++	netd müzik	8500	11373567	23898730764
7	8th	A++	ABS-CBN Entertainment	100147	12149206	17202609850
8	9th	A++	Ryan ToysReview	1140	16082927	24518098041
9	10th	A++	Zee Marathi	74607	2841811	2591830307
10	11th	A+	5-Minute Crafts	2085	33492951	8587520379
11	12th	A+	Canal KondZilla	822	39409726	19291034467
12	13th	A+	Like Nastya Vlog	150	7662886	2540099931
13	14th	A+	Ozuna	50	18824912	8727783225
14	15th	A+	Wave Music	16119	15899764	10989179147
15	16th	A+	Ch3Thailand	49239	11569723	9388600275
16	17th	A+	WORLDSTARHIPHOP	4778	15830098	11102158475
17	18th	A+	Vlad and Nikita	53	NaN	1428274554
18	19th	A+	Badabun	3060	23603062	5860444053
19	20th	A+	WorkpointOfficial	24287	17687229	14022189654

In [111]:
```python
data.head(25)
```

Out[111]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views
0	1st	A++	Zee TV	82757	18752951	20869786591
1	2nd	A++	T-Series	12661	61196302	47548839843
2	3rd	A++	Cocomelon - Nursery Rhymes	373	19238251	9793305082
3	4th	A++	SET India	27323	31180559	22675948293
4	5th	A++	WWE	36756	32852346	26273668433
5	6th	A++	Movieclips	30243	17149705	16618094724
6	7th	A++	netd müzik	8500	11373567	23898730764
7	8th	A++	ABS-CBN Entertainment	100147	12149206	17202609850
8	9th	A++	Ryan ToysReview	1140	16082927	24518098041
9	10th	A++	Zee Marathi	74607	2841811	2591830307
10	11th	A+	5-Minute Crafts	2085	33492951	8587520379
11	12th	A+	Canal KondZilla	822	39409726	19291034467
12	13th	A+	Like Nastya Vlog	150	7662886	2540099931
13	14th	A+	Ozuna	50	18824912	8727783225
14	15th	A+	Wave Music	16119	15899764	10989179147
15	16th	A+	Ch3Thailand	49239	11569723	9388600275
16	17th	A+	WORLDSTARHIPHOP	4778	15830098	11102158475
17	18th	A+	Vlad and Nikita	53	NaN	1428274554
18	19th	A+	Badabun	3060	23603062	5860444053
19	20th	A+	WorkpointOfficial	24287	17687229	14022189654
20	21st	A+	xxxtentacion	48	11593509	2029279311
21	22nd	A+	Zee Music Company	3403	23440427	11666909155
22	23rd	A+	SAB TV	19358	11347420	9572788450
23	24th	A+	DeadloxMC	1833	2137907	245437560
24	25th	A+	ETV Andhra Pradesh	186194	712562	349096736

In [112]: `data.isnull().sum()`   *#checking null values*

Out[112]: Rank   0
           Grade   0
           Channel name   0
           Video Uploads   6
           Subscribers   387
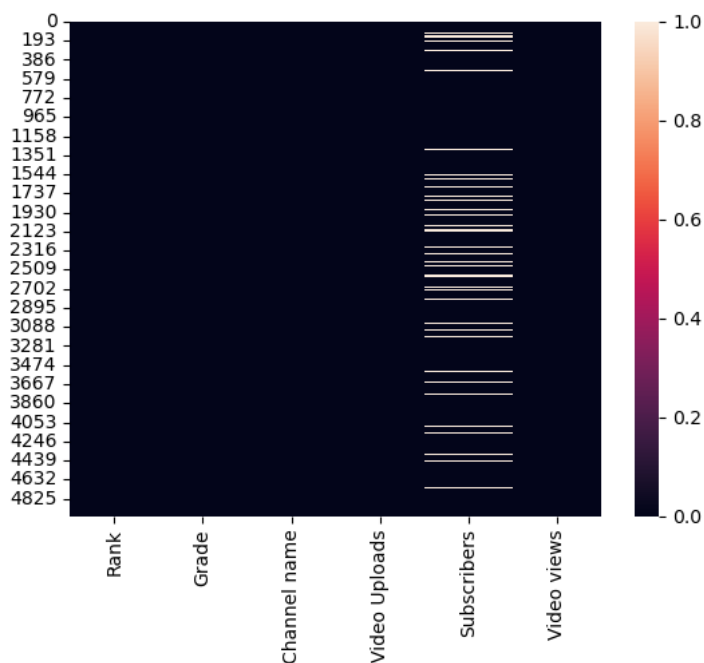   Video views   0
           dtype: int64

In [113]: `per_missing = data.isnull().sum()*100/len(data)`

In [114]: `per_missing`

Out[114]: Rank   0.0
           Grade   0.0
           Channel name   0.0
           Video Uploads   0.12
           Subscribers   7.74
           Video views   0.0
           dtype: float64

In [115]: `sns.heatmap(data.isnull())`   *#using seaborn to visualise missing values*
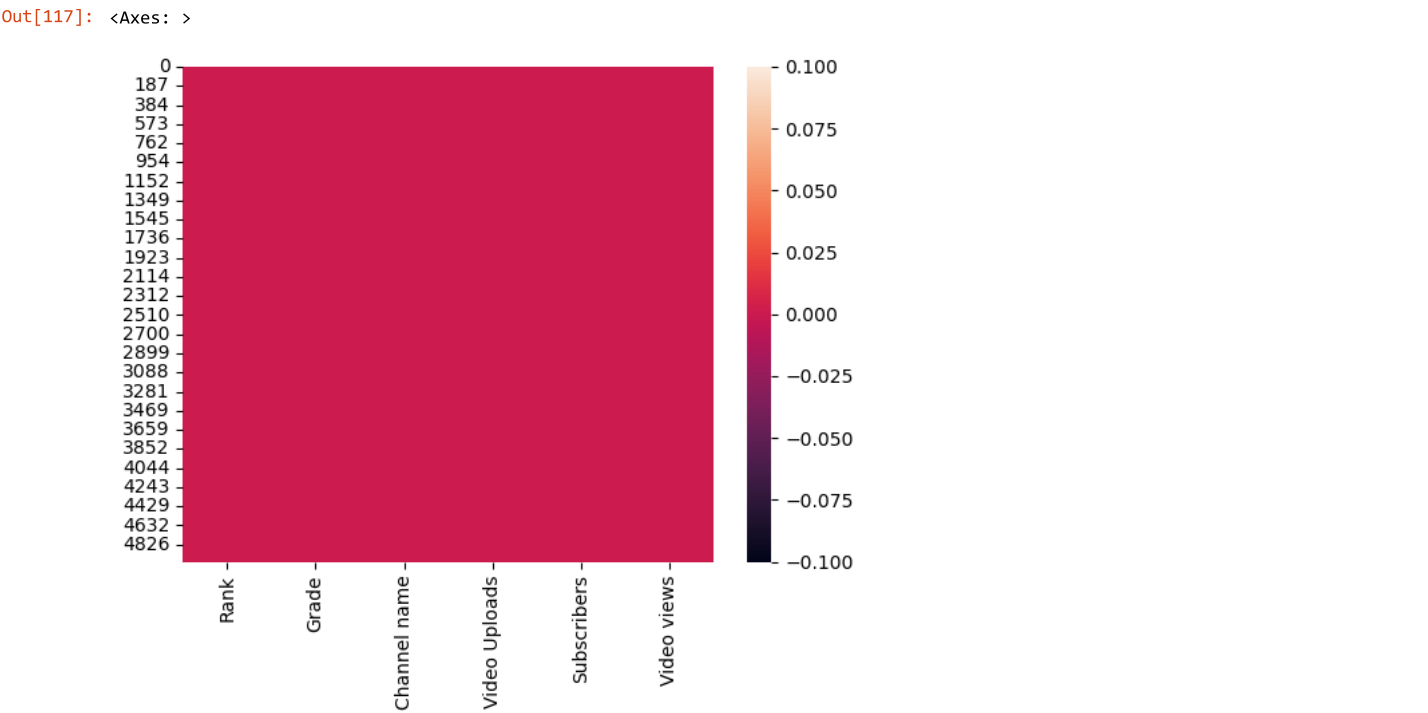
Out[115]: <Axes: >

In [116]: `data.dropna(axis=0,inplace=True)`

In [117]: `sns.heatmap(data.isnull())`

Out[117]: <Axes: >

In [118]: `data.head()`

Out[118]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views
0	1st	A++	Zee TV	82757	18752951	20869786591
1	2nd	A++	T-Series	12661	61196302	47548839843
2	3rd	A++	Cocomelon - Nursery Rhymes	373	19238251	9793305082
3	4th	A++	SET India	27323	31180559	22675948293
4	5th	A++	WWE	36756	32852346	26273668433

In [119]: `data.tail()`

Out[119]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views
4995	4,996th	B+	Uras Benlioğlu	706	2072942	441202795
4996	4,997th	B+	HI-TECH MUSIC LTD	797	1055091	377331722
4997	4,998th	B+	Mastersaint	110	3265735	311758426
4998	4,999th	B+	Bruce McIntosh	3475	32990	14563764
4999	5,000th	B+	SehatAQUA	254	21172	73312511

In [120]: `data.dtypes`

Out[120]:
```
Rank            object
Grade           object
Channel name    object
Video Uploads   object
Subscribers     object
Video views      int64
dtype: object
```

In [121]:
```python
data['Rank'] = data['Rank'].str[0:-2]
data.head()
```

Out[121]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views
0	1	A++	Zee TV	82757	18752951	20869786591
1	2	A++	T-Series	12661	61196302	47548839843
2	3	A++	Cocomelon - Nursery Rhymes	373	19238251	9793305082
3	4	A++	SET India	27323	31180559	22675948293
4	5	A++	WWE	36756	32852346	26273668433

In [122]:
```python
data['Rank'] = data['Rank'].str.replace(',','')  #replaced , by empty string
data.tail()
```

Out[122]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views
4995	4996	B+	Uras Benlioğlu	706	2072942	441202795
4996	4997	B+	HI-TECH MUSIC LTD	797	1055091	377331722
4997	4998	B+	Mastersaint	110	3265735	311758426
4998	4999	B+	Bruce McIntosh	3475	32990	14563764
4999	5000	B+	SehatAQUA	254	21172	73312511

In [123]:
```python
data['Rank'] = data['Rank'].str.replace(',','').astype('int')    #rank columns data type has been changed to integer
```

In [124]:
```python
data['Video Uploads'] = data['Video Uploads'].astype('int')    #video upload columns data type has been changed to integer
```

In [125]:
```python
data['Subscribers'] = data['Subscribers'].astype('int')    #susbcriber columns data type has been changed to integer
```

In [126]:
```python
data['Grade'].unique()
```

Out[126]: array(['A++ ', 'A+ ', 'A ', 'A- ', 'B+ '], dtype=object)

In [127]:
```python
data['Grade'] = data['Grade'].map({'A++ ':5,'A+ ':4,'A ':3,'A- ':2,'B+ ':1})
data.head()
```

Out[127]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views
0	1	5	Zee TV	82757	18752951	20869786591
1	2	5	T-Series	12661	61196302	47548839843
2	3	5	Cocomelon - Nursery Rhymes	373	19238251	9793305082
3	4	5	SET India	27323	31180559	22675948293
4	5	5	WWE	36756	32852346	26273668433

In [133]:
```python
data['Avg_views'] = data['Video views']/data['Video Uploads']
```

Out[133]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views	Avg	Avg_views
0	1	5	Zee TV	82757	18752951	20869786591	252,181.52652947788	252,181.52652947788
1	2	5	T-Series	12661	61196302	47548839843	3,755,535.885238133	3,755,535.885238133
2	3	5	Cocomelon - Nursery Rhymes	373	19238251	9793305082	26,255,509.60321716	26,255,509.60321716
3	4	5	SET India	27323	31180559	22675948293	829,921.6152325879	829,921.6152325879
4	5	5	WWE	36756	32852346	26273668433	714,813.0491076286	714,813.0491076286

In [134]:
```python
data.drop(['Avg'], axis=1)
```

Out[134]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views	Avg_views
0	1	5	Zee TV	82757	18752951	20869786591	252,181.52652947788
1	2	5	T-Series	12661	61196302	47548839843	3,755,535.885238133
2	3	5	Cocomelon - Nursery Rhymes	373	19238251	9793305082	26,255,509.60321716
3	4	5	SET India	27323	31180559	22675948293	829,921.6152325879
4	5	5	WWE	36756	32852346	26273668433	714,813.0491076286
...	...	...	...	...	...	...	...
4995	4996	1	Uras Benlioğlu	706	2072942	441202795	624,933.1373937677
4996	4997	1	HI-TECH MUSIC LTD	797	1055091	377331722	473,440.05269761605
4997	4998	1	Mastersaint	110	3265735	311758426	2,834,167.5090909093
4998	4999	1	Bruce McIntosh	3475	32990	14563764	4,191.011223021583
4999	5000	1	SehatAQUA	254	21172	73312511	288,631.93307086616

4610 rows × 7 columns

In [135]:
```python
data.sort_values(by='Video Uploads',ascending = False)    #so here we can see the top five channels as per video uploads
```

Out[135]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views	Avg	Avg_views
3453	3454	1	AP Archive	422326	746325	548619569	1,299.04284604784	1,299.04284604784
1149	1150	2	YTN NEWS	355996	820108	1640347646	4,607.769879436848	4,607.769879436848
2223	2224	1	SBS Drama	335521	1418619	1565758044	4,666.646928210157	4,666.646928210157
323	324	3	GMA News	269065	2599175	2786949164	10,357.902975117537	10,357.902975117537
2956	2957	1	MLB	267649	1434206	1329206392	4,966.229621631315	4,966.229621631315
...	...	...	...	...	...	...	...	...
4100	4101	1	DenzelCurryVEVO	1	8012	25878252	25,878,252.0	25,878,252.0
2404	2405	1	ItsHARSH007 Gaming	1	23499	75	75.0	75.0
4474	4475	1	hyghhopes	1	15325	14842098	14,842,098.0	14,842,098.0
628	629	3	cocoz toon	1	2162992	333012122	333,012,122.0	333,012,122.0
4975	4976	1	Funny Kids	1	150091	21933876	21,933,876.0	21,933,876.0

4610 rows × 8 columns

In [137]:
```python
data.corr()    #to see the coorelation matrix
```
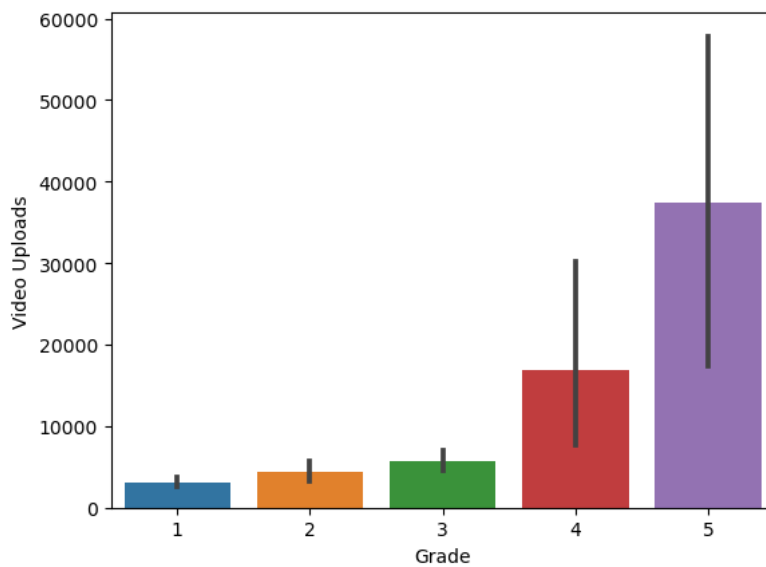
```
C:\Users\Lenovo\AppData\Local\Temp\ipykernel_2804\2627137660.py:1: FutureWarning: The default value of numeric_only in DataFram
e.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_
only to silence this warning.
  data.corr()
```

Out[137]:

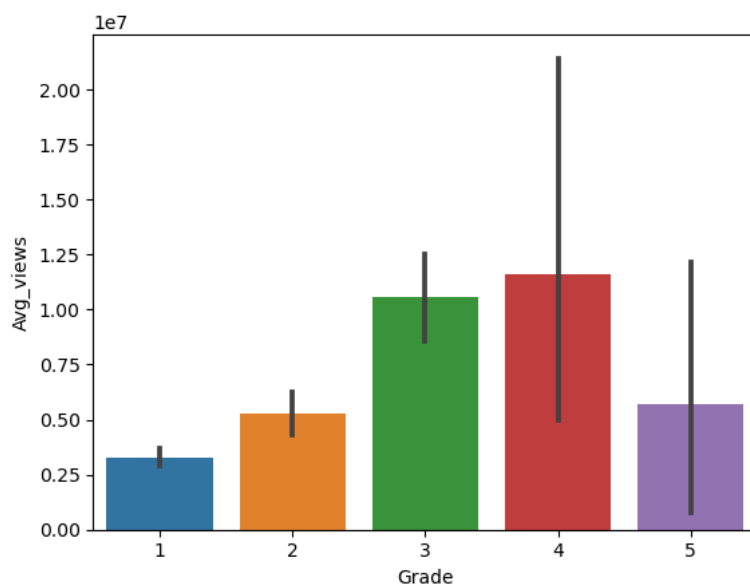|	Rank	Grade	Video Uploads	Subscribers	Video views	Avg	
Rank	1.0	-0.8650831909022979	-0.07317993953028674	-0.38332851942555984	-0.40287300698548156	-0.15367008750257183	-0.15367008
Grade	-0.8650831909022979	1.0	0.08877344223749478	0.42921251332828675	0.47742343608776905	0.15523143918832177	0.15523143
Video Uploads	-0.07317993953028674	0.08877344223749478	1.0	0.011362046412882852	0.08783041184561236	-0.06440844862474529	-0.06440844
Subscribers	-0.38332851942555984	0.42921251332828675	0.011362046412882852	1.0	0.7912411514219967	0.28938596191979615	0.28938596
Video views	-0.40287300698548156	0.47742343608776905	0.08783041184561236	0.7912411514219967	1.0	0.2944215695038217	0.2944215
Avg	-0.15367008750257183	0.15523143918832177	-0.06440844862474529	0.28938596191979615	0.2944215695038217	1.0	
Avg_views	-0.15367008750257183	0.15523143918832177	-0.06440844862474529	0.28938596191979615	0.2944215695038217	1.0	

In [141]: `sns.barplot(x='Grade',y='Video Uploads',data = data)`   *#which grade has maximum number of views*

Out[141]: `<Axes: xlabel='Grade', ylabel='Video Uploads'>`
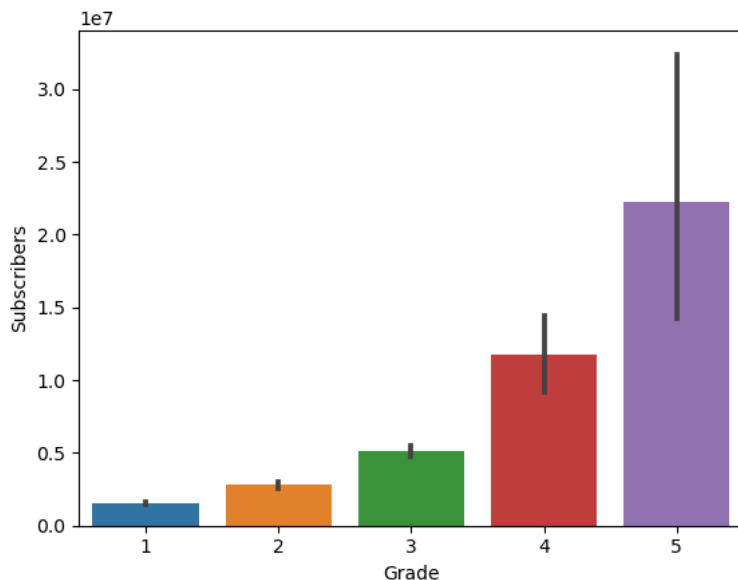

In [143]: `sns.barplot(x='Grade',y='Avg_views',data = data)`   *#which grade has highest averager of views*

Out[143]: `<Axes: xlabel='Grade', ylabel='Avg_views'>`

In [144]: `sns.barplot(x='Grade',y='Subscribers',data = data)`   *#which grade has highest subsciber*

Out[144]: `<Axes: xlabel='Grade', ylabel='Subscribers'>`

In [145]: `data.groupby('Grade').mean()`   *#which grade has high video views*

```
C:\Users\Lenovo\AppData\Local\Temp\ipykernel_2804\3670195322.py:1: FutureWarning: The default value of numeric_only in DataFram
eGroupBy.mean is deprecated. In a future version, numeric_only will default to False. Either specify numeric_only or select onl
y columns which should be valid for the function.
  data.groupby('Grade').mean()
```

Out[145]:

Grade	Rank	Video Uploads	Subscribers	Video views	Avg	Avg_views
1	3,520.537839823659	3,136.1572373254958	1,535,207.9470977222	555,183,839.0859662	3,280,380.877526468	3,280,380.877526468
2	1,533.9883103081827	4,382.582359192348	2,798,520.3751328373	1,102,450,027.6865036	5,254,804.043742598	5,254,804.043742598
3	534.2853957636567	5,709.855072463768	5,107,136.294314382	2,497,972,949.1114826	10,540,908.454617342	10,540,908.454617342
4	31.325	16,960.3	11,726,947.475	6,168,741,772.725	11,577,080.321918543	11,577,080.321918543
5	5.5	37,450.7	22,281,762.5	21,199,091,192.8	5,688,267.963020266	5,688,267.963020266

In [146]: `sns.barplot(x='Grade',y='Video views',data = data)`   *#which grade has highest video views*

Out[146]: `<Axes: xlabel='Grade', ylabel='Video views'>`