In [1]:
```python
import pandas as pd
import seaborn as sns
```

In [2]:
```python
data = pd.read_csv('top-5000-youtube-channels.csv')
```

In [3]:
```python
data.head()
```

Out[3]:

	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

Display All Rows Except the Last 5 rows Using Head Method

In [4]:
```python
data.head(-5)
```

Out[4]:

	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
...	...	...	...	...	...	...
4990	4,991st	B+	Ho Ngoc Ha's Official Channel	208	--	127185704
4991	4,992nd	B+	Toys to Learn Colors	11	663114	141933264
4992	4,993rd	B+	KAZKA	25	131766	74304638
4993	4,994th	B+	United CUBE (CUBE Entertainment…	1055	1586835	371299166
4994	4,995th	B+	Wings Marathi	1735	1099659	346175699

4995 rows × 6 columns

Display All Rows Except the First 5 Rows Using Tail Method

In [5]:
```python
data.tail(-5)
```

Out[5]:

	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 [6]:
```python
data.shape
```

Out[6]:
```
(5000, 6)
```

In [7]:
```python
print('Number of Rows', data.shape[0])
print('Number of Columns', data.shape[1])
```

```
Number of Rows 5000
Number of Columns 6
```

In [8]:
```python
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
```

In [9]:
```python
pd.options.display.float_format = '{:.2f}'.format
```

In [10]:
```python
data.describe()
```

Out[10]:

	Video views
count	5000.00
mean	1071449400.15
std	2003843972.12
min	75.00
25%	186232945.75
50%	482054780.00
75%	1124367826.75
max	47548839843.00

Data Cleaning (Replace '--' to NaN)

In [11]:
```python
import numpy as np
```

In [12]:
```python
data=data.replace('--', np.nan, regex=True)
```

In [13]:
```python
data.head(20)
```

Out[13]:

	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 [14]:
```python
data.isnull().sum()
```

Out[14]:
```
Rank               0
Grade              0
Channel name       0
Video Uploads      6
Subscribers      387
Video views        0
dtype: int64
```

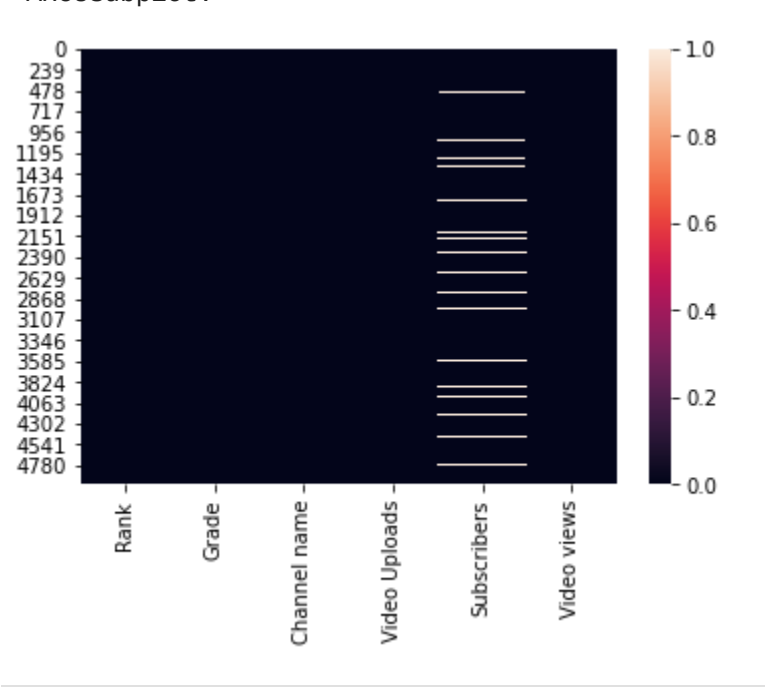
In [15]:
```python
per_missing = data.isnull().sum() *100 / len(data)
```

In [16]:
```python
per_missing
```

Out[16]:
```
Rank            0.00
Grade           0.00
Channel name    0.00
Video Uploads   0.12
Subscribers     7.74
Video views     0.00
dtype: float64
```

In [17]:
```python
sns.heatmap(data.isnull())
```

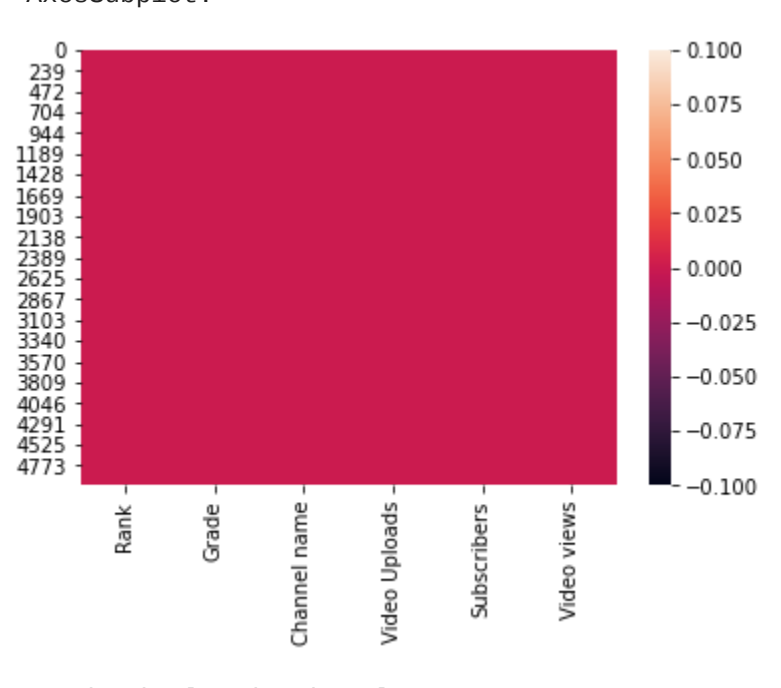
Out[17]:
```
<AxesSubplot:>
```



In [19]:
```python
data.dropna(axis=0, inplace=True)
```

In [20]:
```python
sns.heatmap(data.isnull())
```

Out[20]:
```
<AxesSubplot:>
```



Data Cleaning [Rank Column]

In [22]:
```python
data.head()
```

Out[22]:

	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 [23]:
```python
data.dtypes
```

Out[23]:
```
Rank             object
Grade            object
Channel name     object
Video Uploads    object
Subscribers      object
Video views       int64
dtype: object
```

In [25]:
```python
data['Rank']=data['Rank'].str[0:-2]
```

In [27]:
```python
data.tail()
```

Out[27]:

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

In [31]:
```python
data['Rank']=data['Rank'].str.replace(',', '').astype('int')
```

In [32]:
```python
data.tail()
```

Out[32]:

	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 [33]:
```python
data.dtypes
```

Out[33]:
```
Rank              int32
Grade            object
Channel name     object
Video Uploads    object
Subscribers      object
Video views       int64
dtype: object
```

Data Cleaning [Video Uploads & Subscribers]

In [34]:
```python
data.head()
```

Out[34]:

	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 [35]:
```python
data.dtypes
```

Out[35]:
```
Rank              int32
Grade            object
Channel name     object
Video Uploads    object
Subscribers      object
Video views       int64
dtype: object
```

In [36]:
```python
data['Video Uploads'] = data['Video Uploads'].astype('int')
```

In [37]:
```python
data['Subscribers'] = data['Subscribers'].astype('int')
```

In [38]:
```python
data.dtypes
```

Out[38]:
```
Rank              int32
Grade            object
Channel name     object
Video Uploads     int32
Subscribers       int32
Video views       int64
dtype: object
```

Data Cleaning [Grade Column]

In [39]:
```python
data.head()
```

Out[39]:

	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 [41]:
```python
data['Grade'].unique()
```

Out[41]:
```
array(['A++ ', 'A+ ', 'A ', 'A- ', 'B+ '], dtype=object)
```

In [43]:
```python
data['Grade'] = data['Grade'].map({'A++ ': 5, 'A+ ': 4, 'A ': 3, 'A- ': 2, 'B+ ':1})
```

In [44]:
```python
data.dtypes
```

Out[44]:
```
Rank              int32
Grade             int64
Channel name     object
Video Uploads     int32
Subscribers       int32
Video views       int64
dtype: object
```

Find Average Views For Each Channel

In [45]:
```python
data.columns
```

Out[45]:
```
Index(['Rank', 'Grade', 'Channel name', 'Video Uploads', 'Subscribers',
       'Video views'],
      dtype='object')
```

In [48]:
```python
data['Avg_views']=data['Video views']/data['Video Uploads']
```

In [49]:
```python
data.head()
```

Out[49]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views	Avg_views
0	1	5	Zee TV	82757	18752951	20869786591	252181.53
1	2	5	T-Series	12661	61196302	47548839843	3755535.89
2	3	5	Cocomelon - Nursery Rhymes	373	19238251	9793305082	26255509.60
3	4	5	SET India	27323	31180559	22675948293	829921.62
4	5	5	WWE	36756	32852346	26273668433	714813.05

Find Out Top Five Channels With Maximum Number of Video Uploads

In [50]:
```python
data.columns
```

Out[50]:
```
Index(['Rank', 'Grade', 'Channel name', 'Video Uploads', 'Subscribers',
       'Video views', 'Avg_views'],
      dtype='object')
```

In [53]:
```python
data.sort_values(by='Video Uploads', ascending=False).head()
```

Out[53]:

	Rank	Grade	Channel name	Video Uploads	Subscribers	Video views	Avg_views
3453	3454	1	AP Archive	422326	746325	548619569	1299.04
1149	1150	2	YTN NEWS	355996	820108	1640347646	4607.77
2223	2224	1	SBS Drama	335521	1418619	1565758044	4666.65
323	324	3	GMA News	269065	2599175	2786949164	10357.90
2956	2957	1	MLB	267649	1434206	1329206392	4966.23

Find Correlation Matrix

In [55]:
```python
data.corr()
```

Out[55]:

	Rank	Grade	Video Uploads	Subscribers	Video views	Avg_views
Rank	1.00	-0.87	-0.07	-0.38	-0.40	-0.15
Grade	-0.87	1.00	0.09	0.43	0.48	0.16
Video Uploads	-0.07	0.09	1.00	0.01	0.09	-0.06
Subscribers	-0.38	0.43	0.01	1.00	0.79	0.29
Video views	-0.40	0.48	0.09	0.79	1.00	0.29
Avg_views	-0.15	0.16	-0.06	0.29	0.29	1.00

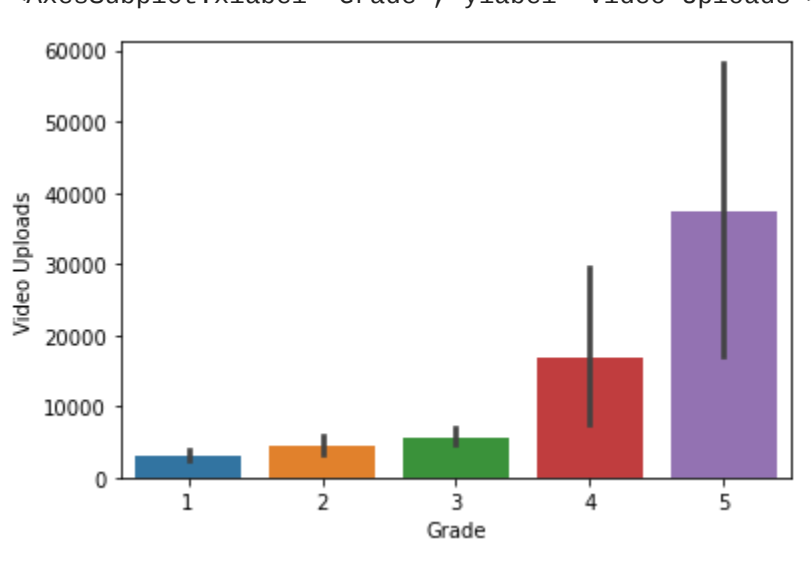
Which Grade Has A Maximum Number of Video Uploads?

In [56]:
```python
data.columns
```

Out[56]:
```
Index(['Rank', 'Grade', 'Channel name', 'Video Uploads', 'Subscribers',
       'Video views', 'Avg_views'],
      dtype='object')
```

In [57]:
```python
sns.barplot(x='Grade', y='Video Uploads', data=data)
```

Out[57]:
```
<AxesSubplot:xlabel='Grade', ylabel='Video Uploads'>
```



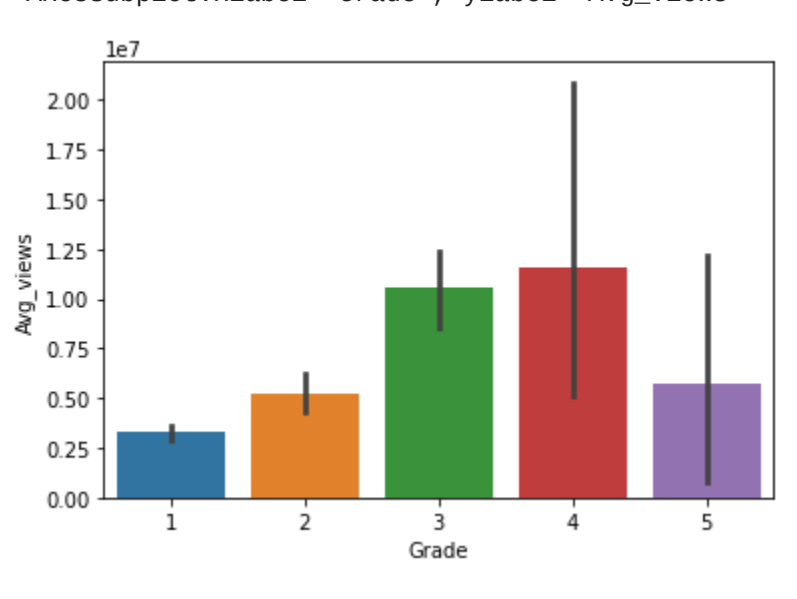
Which Grade Has The Highest Average Views?

In [58]:
```python
data.columns
```

Out[58]:
```
Index(['Rank', 'Grade', 'Channel name', 'Video Uploads', 'Subscribers',
       'Video views', 'Avg_views'],
      dtype='object')
```

In [59]:
```python
sns.barplot(x='Grade', y='Avg_views', data=data)
```

Out[59]:
```
<AxesSubplot:xlabel='Grade', ylabel='Avg_views'>
```



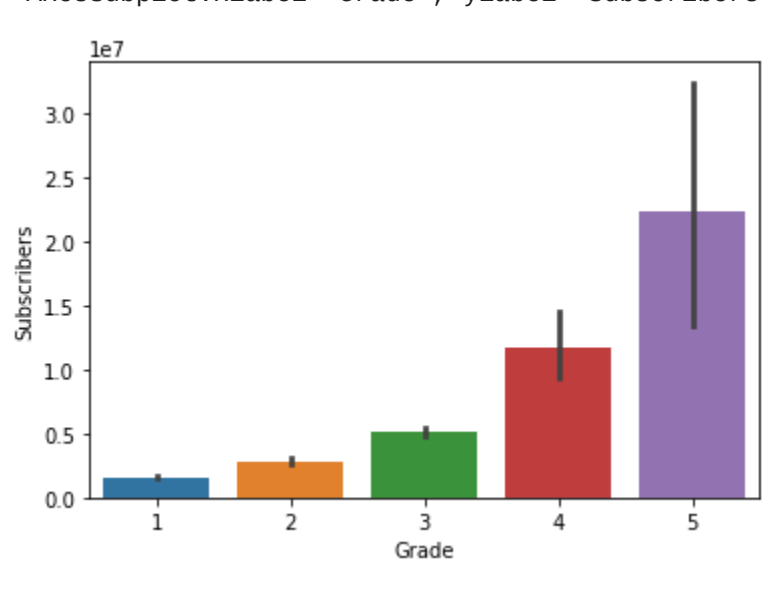
Which Grade Has The Highest Number of Subscribers?

In [60]:
```python
data.columns
```

Out[60]:
```
Index(['Rank', 'Grade', 'Channel name', 'Video Uploads', 'Subscribers',
       'Video views', 'Avg_views'],
      dtype='object')
```

In [61]:
```python
sns.barplot(x='Grade', y='Subscribers', data=data)
```

Out[61]:
```
<AxesSubplot:xlabel='Grade', ylabel='Subscribers'>
```



Which Grade Has The Highest Video Views?

In [65]:
```python
data.groupby('Grade').mean()
```

Out[65]:

Grade	Rank	Video Uploads	Subscribers	Video views	Avg_views
1	3520.54	3136.16	1535207.95	555183839.09	3280380.88
2	1533.99	4382.58	2798520.38	1102450027.69	5254804.04
3	534.29	5709.86	5107136.29	2497972949.11	10540908.45
4	31.32	16960.30	11726947.47	6168741772.73	11577080.32
5	5.50	37450.70	22281762.50	21199091192.80	5688267.96

In [ ]:

In [ ]: