yt-channel-analysis

February 2, 2024

[4]:
```
import pandas as pd
import seaborn as sns
```

[3]:
```
data = pd.read_csv(r"E:\Youtube Channel\top-5000-youtube-channels.csv")
```

[7]:
```
data
```

[7]:

	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
…	…	…	…	…	…	…
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

[5000 rows x 6 columns]

Display All Rows Except the Last 5 rows Using Head Method

[8]:
```
data.head(-5)
```

[8]:
```
          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+                                    Z            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 Rows Except the First 5 Rows Using Tail Method

[9]:
```
data.tail(-5)
```

[9]:
```
          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]
```

Find Shape Of Our Dataset

[11]:
```
data.shape
```

[11]:
```
(5000, 6)
```

[12]:
```
print("No. Of Rows:",data.shape[0])
print("No. Of Columns:",data.shape[1])
```

```
No. Of Rows: 5000
No. Of Columns: 6
```

Get Informations

[13]:
```
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
```

Get Statistics

[15]:
```
pd.options.display.float_format = '{:.2f}'.format
```

[16]:
```
data.describe()
```

[16]:
```
          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)

[18]:
```
import numpy as np
```

[20]:
```
data = data.replace('--',np.nan,regex=True)
```

Check Null Values

[27]:
```
data.isnull().sum()
```

[27]:
```
Rank              0
Grade             0
Channel name      0
Video Uploads     6
Subscribers     387
Video views       0
dtype: int64
```

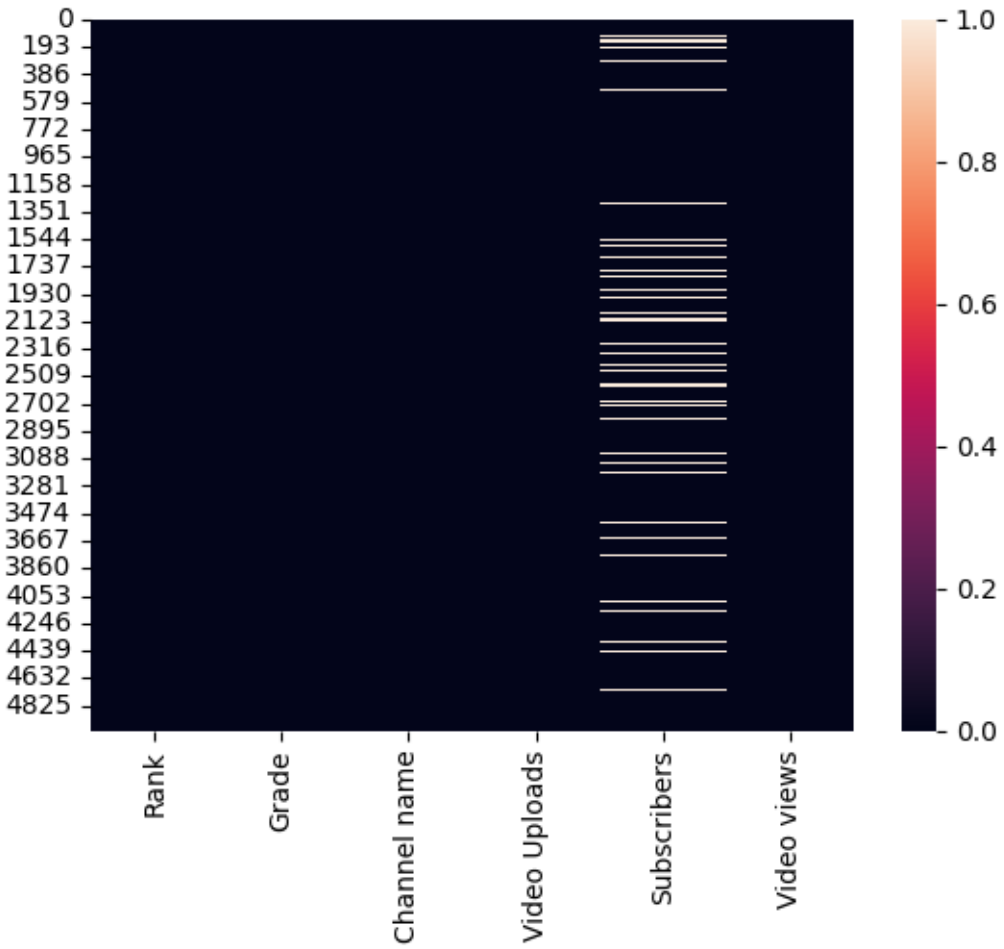
[28]:
```
per_missing = data.isnull().sum()*100/len(data)
```

[29]:
```
per_missing
```

[29]:
```
Rank            0.00
Grade           0.00
Channel name    0.00
Video Uploads   0.12
Subscribers     7.74
Video views     0.00
dtype: float64
```

[31]:
```
sns.heatmap(data.isnull())
```

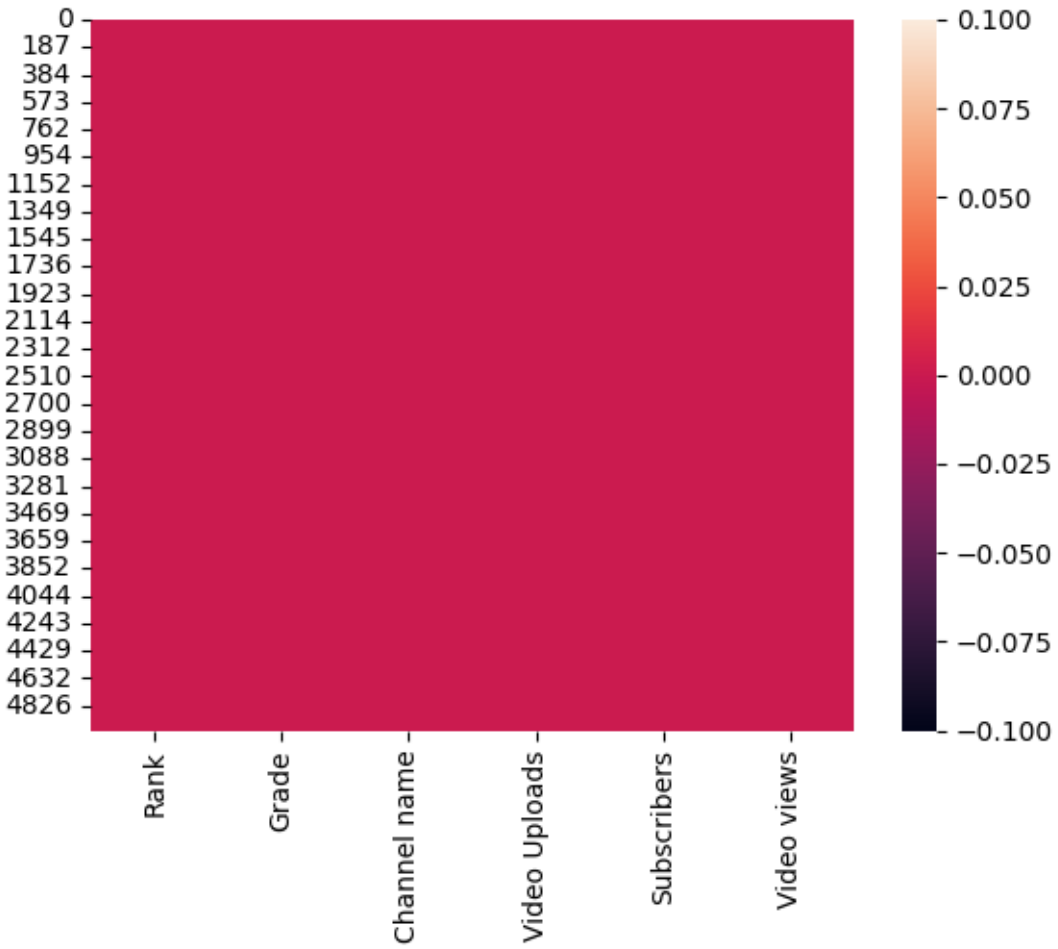
[31]:
```
<Axes: >
```



[32]:
```
data.dropna(axis=0,inplace=True)
```

[33]:
```
sns.heatmap(data.isnull())
```

[33]: <Axes: >



Data Cleaning [Rank Column]

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

[34]:

	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

[35]:
```
data.tail()
```

[35]:

	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

[37]:
```
data['Rank'] = data['Rank'].str[0:-2]
```

[45]:
```
data['Rank']=data['Rank'].str.replace(',','').astype('int')
```

[46]:
```
data.dtypes
```

[46]:
```
Rank             int32
Grade           object
Channel name    object
Video Uploads   object
Subscribers     object
Video views      int64
dtype: object
```

Data Cleaning [Video Uploads and Subscribers]

[48]:
```
data['Video Uploads'] = data['Video Uploads'].astype('int')
data['Subscribers'] = data['Subscribers'].astype('int')
```

[49]:
```
data.dtypes
```

[49]:
```
Rank             int32
Grade           object
Channel name    object
Video Uploads    int32
Subscribers      int32
Video views      int64
dtype: object
```

Data Cleaning [Grade Column]

[50]:
```
data['Grade'].unique()
```

[50]:
```
array(['A++ ', 'A+ ', 'A ', 'A- ', 'B+ '], dtype=object)
```

[53]:
```
data['Grade']=data['Grade'].map({'A++ ':5,'A+ ':4,'A ':3,'A- ':2,'B+ ':1})
```

[54]:
```
data.dtypes
```

[54]:
```
Rank              int32
Grade             int64
Channel name     object
Video Uploads     int32
Subscribers       int32
Video views       int64
dtype: object
```

Find Avg Views For Each Channel

[56]:
```
data['Avg_Views'] = data['Video views']/data['Video Uploads']
```

[57]:
```
data['Avg_Views']
```

[57]:
```
0        252181.53
1       3755535.89
2      26255509.60
3        829921.62
4        714813.05
           ...
4995     624933.14
4996     473440.05
4997    2834167.51
4998       4191.01
4999     288631.93
Name: Avg_Views, Length: 4610, dtype: float64
```

Find Out Top Five Chaneels With Maximum Number of Video Uploads

[59]:
```
data.sort_values(by='Video Uploads', ascending=False).head()
```

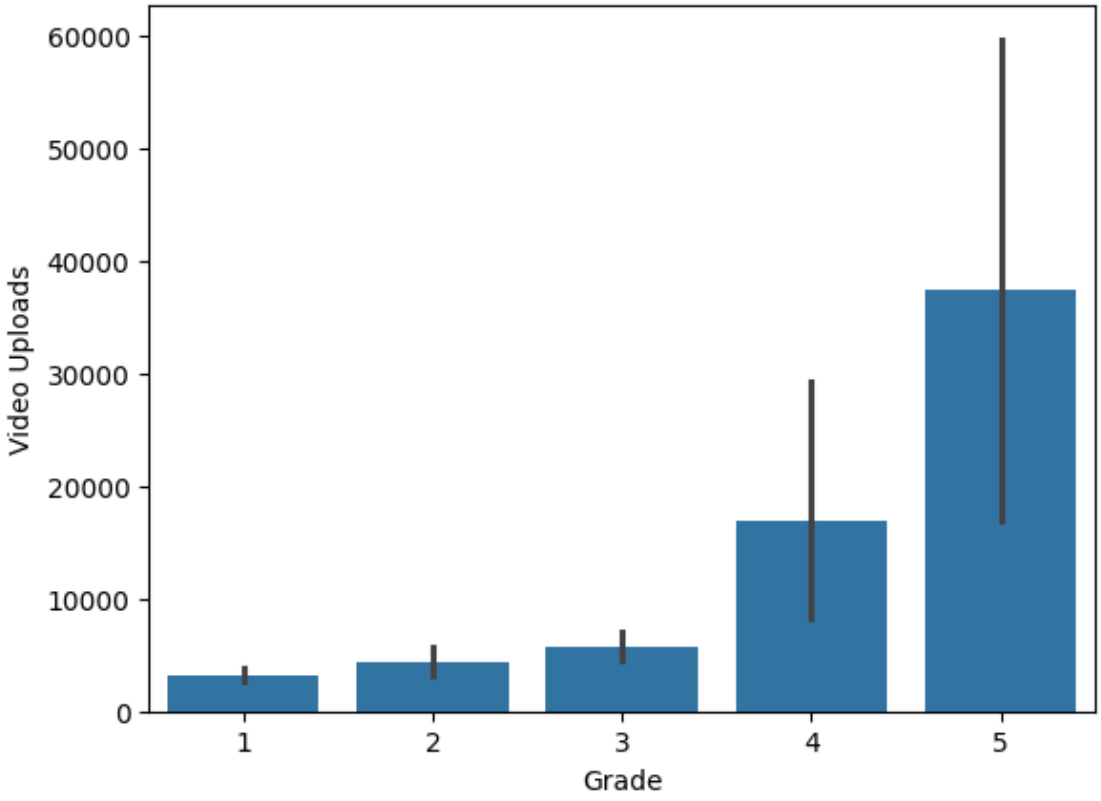
[59]:

	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

Which Grade Has a Maximum Number of Video Uploads

[66]:
```
sns.barplot(x='Grade', y='Video Uploads', data=data)
```

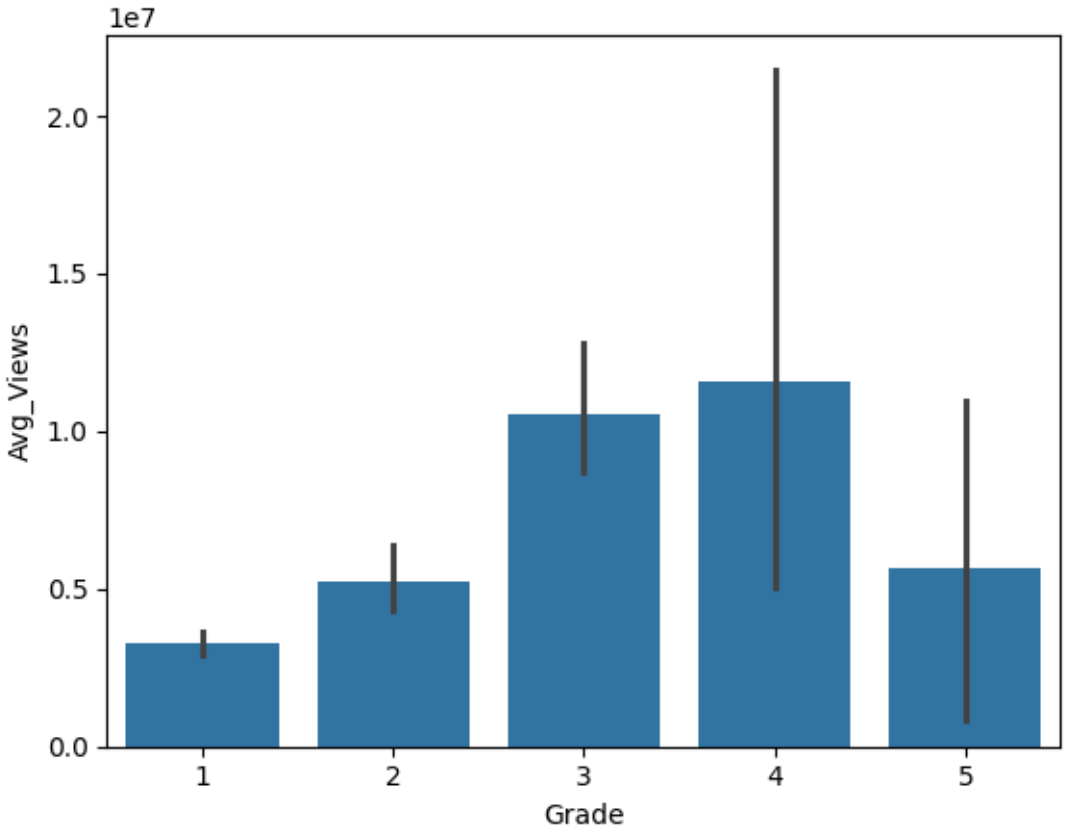
[66]: <Axes: xlabel='Grade', ylabel='Video Uploads'>



Which Grade Has A Highest Average Views?

[67]:
```
sns.barplot(x='Grade', y='Avg_Views', data=data)
```

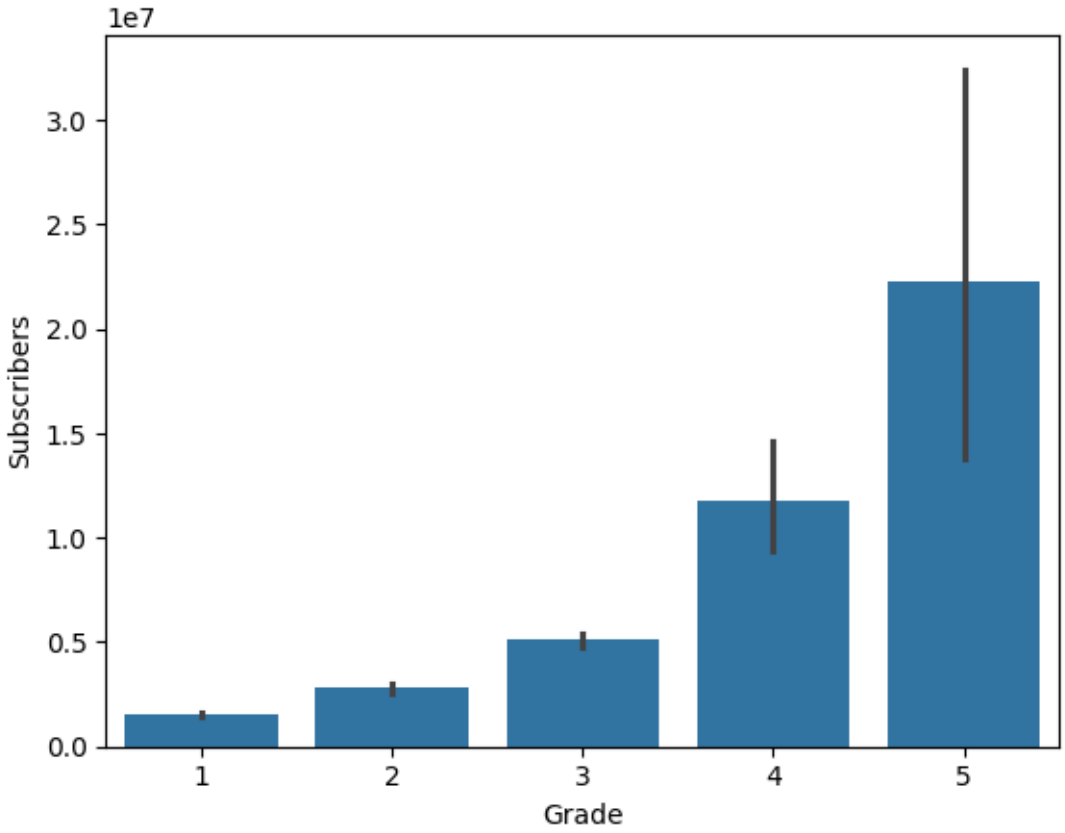
[67]: <Axes: xlabel='Grade', ylabel='Avg_Views'>



Which Grade Has the Highest Numer of Subscribers?

[68]:
```
sns.barplot(x='Grade', y='Subscribers', data=data)
```

[68]:
```
<Axes: xlabel='Grade', ylabel='Subscribers'>
```



Which Grade Has The Highest Video Views?

[77]:
```
data.columns
```

[77]:
```
Index(['Rank', 'Grade', 'Channel name', 'Video Uploads', 'Subscribers',
       'Video views', 'Avg_Views'],
      dtype='object')
```

[83]:
```
data.groupby('Grade')['Video views'].sum()
```

[83]:
```
Grade
1    1511210409992
2    1037405476053
3    2240681735353
4     246749670909
5     211990911928
Name: Video views, dtype: int64
```