

Python pandas learning EDA project 'YouTube channel Analysis'

1. Importing Libraries like pandas, numpy, seaborn.
2. Reading dataset and Loading in data variable.
3. Display All Rows Except the Last 5 rows Using Head Method.
4. Display All Rows Except the First 5 Rows Using Tail Method
5. Find the Shape of Our Dataset (Number of Rows And Number of Columns)
6. Get Information About Our Dataset Like Total Number Rows, Total Number of Columns, Datatypes of Each Column And Memory Requirement
7. Get Overall Statistics About The Dataframe
8. Data Cleaning (Replace '--' to NaN)
9. Check Null Values In The Dataset
10. Drop Null Values In The Dataset
11. Data Cleaning [Rank Column]
12. Data Cleaning [Video Uploads & Subscribers]
13. Data Cleaning [Grade Column]
14. Find Average Views For Each Channel
15. Find Out Top Five Channels With Maximum Number of Video Uploads
16. Find Correlation Matrix
17. Which Grade Has A Maximum Number of Video Uploads?
18. Which Grade Has The Highest Average Views?
19. Which Grade Has The Highest Number of Subscribers?
20. Which Grade Has The Highest Video Views?

```
In [ ]: # 1. Importing Libraries Like pandas, numpy, seaborn.  
import pandas as pd  
import numpy as np  
import seaborn as sns
```

```
In [ ]: # 2. Reading dataset and Loading in data variable.  
data = pd.read_csv("top-5000-youtube-channels.csv")  
data
```

| Out[]: | 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 × 6 columns

In []: # 3. Display ALL Rows Except the Last 5 rows Using Head Method.
data.head(-5)

| Out[]: | 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

In []: # 4. Display ALL Rows Except the First 5 Rows Using Tail Method
data.tail(-5)

| Out[]: | Rank | Grade | Channel name | Video Uploads | Subscribers | Video views |
|---------|-------------|---------|--------------|-----------------------|-------------|-------------|
| | 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 |

4995 rows × 6 columns

```
In [ ]: # 5.Find the Shape of Our Dataset (Number of Rows And Number of Columns)
data.shape
print("Rows",data.shape[0])
print("Columns",data.shape[1])
```

Rows 5000
Columns 6

```
In [ ]: # 6.Get Information About Our Dataset Like Total Number Rows, Total Number of Columns
data.info()
```

| # | 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 [ ]: # 7.Get Overall Statistics About The Dataframe (These Functions performs only on numerical data)
data.describe()
```

Out[]:

| 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 |

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

Out[]:

| | 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 []: `# 8.Data Cleaning (Replace '--' to NaN)
data = data.replace("--", np.nan, regex=True)`In []: `data.head(20)`

| Out[]: | 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 []: # 9. Check Null Values In The Dataset
data.isnull().sum()

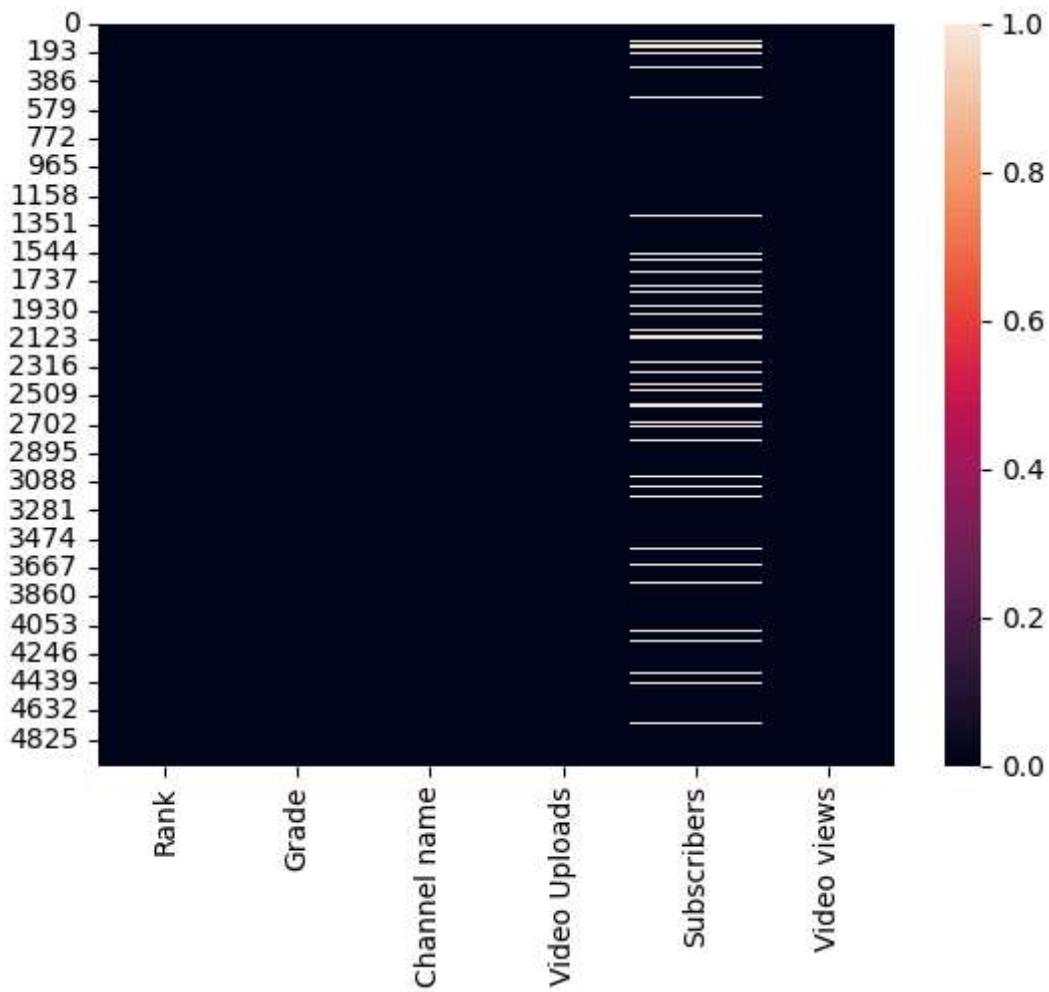
Out[]: Rank 0
Grade 0
Channel name 0
Video Uploads 6
Subscribers 387
Video views 0
dtype: int64

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

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

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

Out[]: <Axes: >



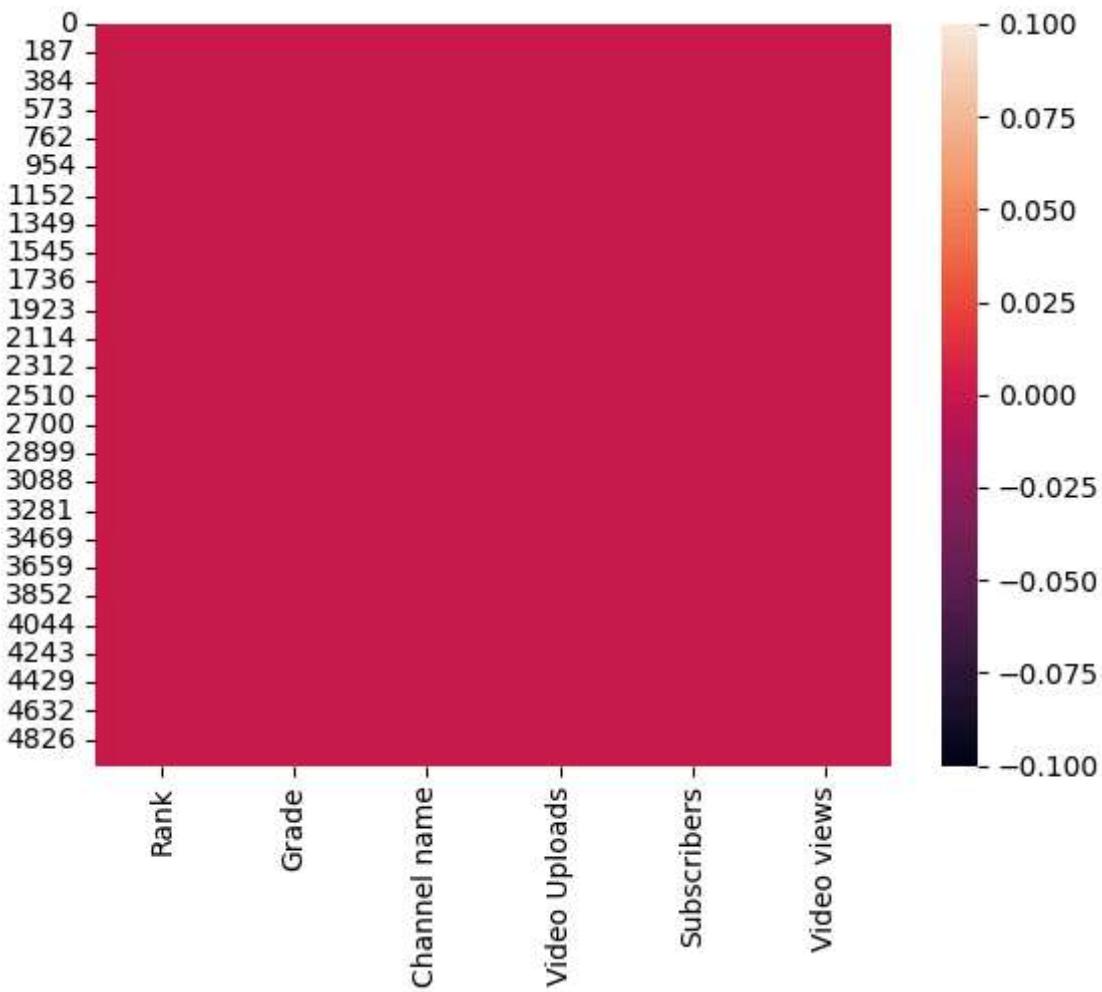
```
In [ ]: # 10. Check and Drop Null Values In The Dataset  
data.dropna(axis=0,inplace = True)
```

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

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

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

```
Out[ ]: <Axes: >
```



In []: `data.head()`

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

| | 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 []: `# 11. Data Cleaning [Rank Column]
data["Rank"] = data["Rank"].str[0:-2]`

In []: `data.tail()`

| | 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 []: `data["Rank"] = data["Rank"].str.replace(",","").astype("int")`

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

| | 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 []: `# 12.Data Cleaning [Video Uploads & Subscribers]`
`data["Video Uploads"] = data["Video Uploads"].astype("int")`

In []: `data["Subscribers"] = data["Subscribers"].astype("int")`

In []: `# Shows Unique Values in Grade Column`
`data["Grade"].unique()`

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

In []: `# Declaring Dictionary keys and values pair`
`dic = {'A++ ':5, 'A+ ':4, 'A ':3, 'A- ':2, 'B+ ':1}`

In []: `# 13.Data Cleaning [Grade]`
`data["Grade"] = data["Grade"].map(dic)`

In []: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4610 entries, 0 to 4999
Data columns (total 6 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Rank             4610 non-null    int64  
 1   Grade            4610 non-null    int64  
 2   Channel name     4610 non-null    object  
 3   Video Uploads    4610 non-null    int64  
 4   Subscribers      4610 non-null    int64  
 5   Video views       4610 non-null    int64  
dtypes: int64(5), object(1)
memory usage: 252.1+ KB
```

In []: `data.head()`

| | 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 []: `# 14. Find Average Views For Each Channel`
`data["Avg_VIEWS"] = round(data["Video views"]/data["Video Uploads"])`

In []: `data.head()`

| | Rank | Grade | Channel name | Video Uploads | Subscribers | Video views | Avg_VIEWS |
|----------|------|-------|----------------------------|---------------|-------------|-------------|------------|
| 0 | 1 | 5 | Zee TV | 82757 | 18752951 | 20869786591 | 252182.0 |
| 1 | 2 | 5 | T-Series | 12661 | 61196302 | 47548839843 | 3755536.0 |
| 2 | 3 | 5 | Cocomelon - Nursery Rhymes | 373 | 19238251 | 9793305082 | 26255510.0 |
| 3 | 4 | 5 | SET India | 27323 | 31180559 | 22675948293 | 829922.0 |
| 4 | 5 | 5 | WWE | 36756 | 32852346 | 26273668433 | 714813.0 |

In []: `# 15. Find Out Top Five Channels With Maximum Number of Video Uploads`
`data.sort_values(by="Video Uploads", ascending=False).head()`

| | Rank | Grade | Channel name | Video Uploads | Subscribers | Video views | Avg_VIEWS |
|-------------|------|-------|--------------|---------------|-------------|-------------|-----------|
| 3453 | 3454 | 1 | AP Archive | 422326 | 746325 | 548619569 | 1299.0 |
| 1149 | 1150 | 2 | YTN NEWS | 355996 | 820108 | 1640347646 | 4608.0 |
| 2223 | 2224 | 1 | SBS Drama | 335521 | 1418619 | 1565758044 | 4667.0 |
| 323 | 324 | 3 | GMA News | 269065 | 2599175 | 2786949164 | 10358.0 |
| 2956 | 2957 | 1 | MLB | 267649 | 1434206 | 1329206392 | 4966.0 |

```
In [ ]: # 16.Find Correlation Matrix
df = pd.DataFrame(data)
numeric_columns = df.select_dtypes(include="int")
Coorelation_matrix = numeric_columns.corr()
Coorelation_matrix
```

Out[]:

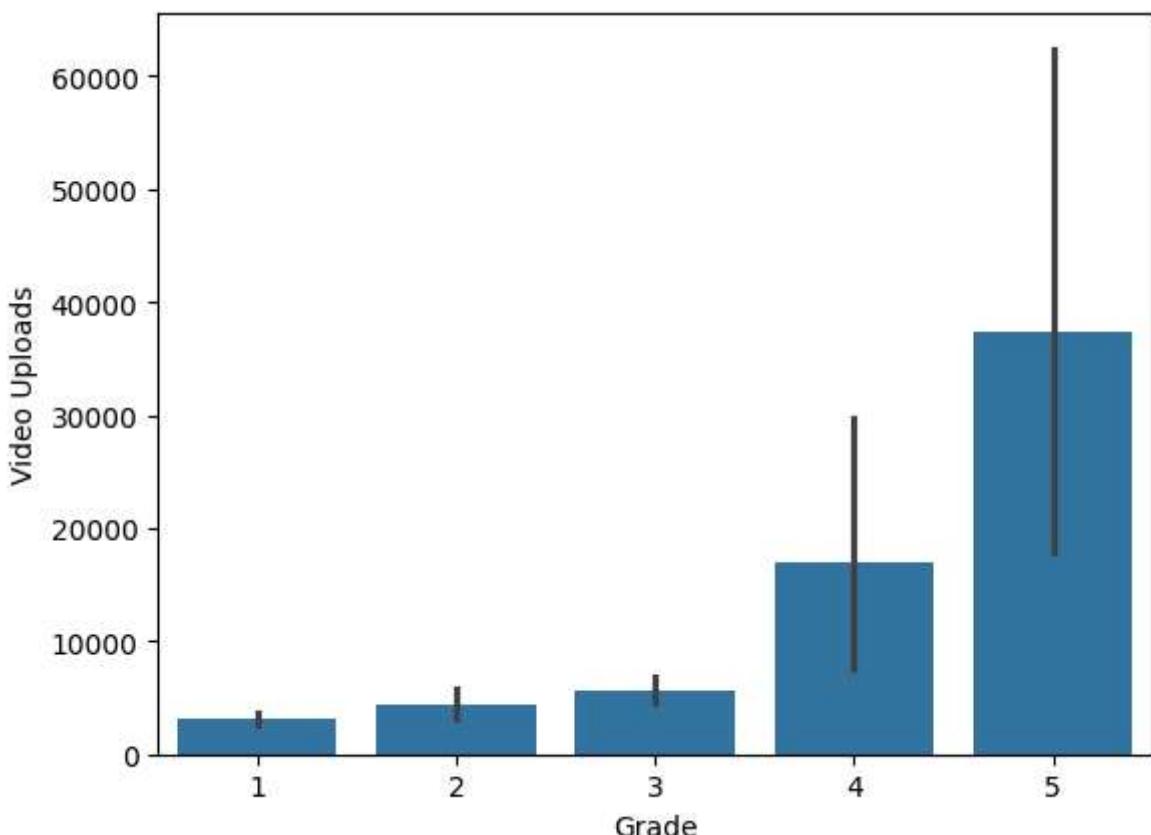
| | Rank | Grade | Video Uploads | Subscribers | Video views |
|---------------|-----------|-----------|---------------|-------------|-------------|
| Rank | 1.000000 | -0.865083 | -0.073180 | -0.383329 | -0.402873 |
| Grade | -0.865083 | 1.000000 | 0.088773 | 0.429213 | 0.477423 |
| Video Uploads | -0.073180 | 0.088773 | 1.000000 | 0.011362 | 0.087830 |
| Subscribers | -0.383329 | 0.429213 | 0.011362 | 1.000000 | 0.791241 |
| Video views | -0.402873 | 0.477423 | 0.087830 | 0.791241 | 1.000000 |

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

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

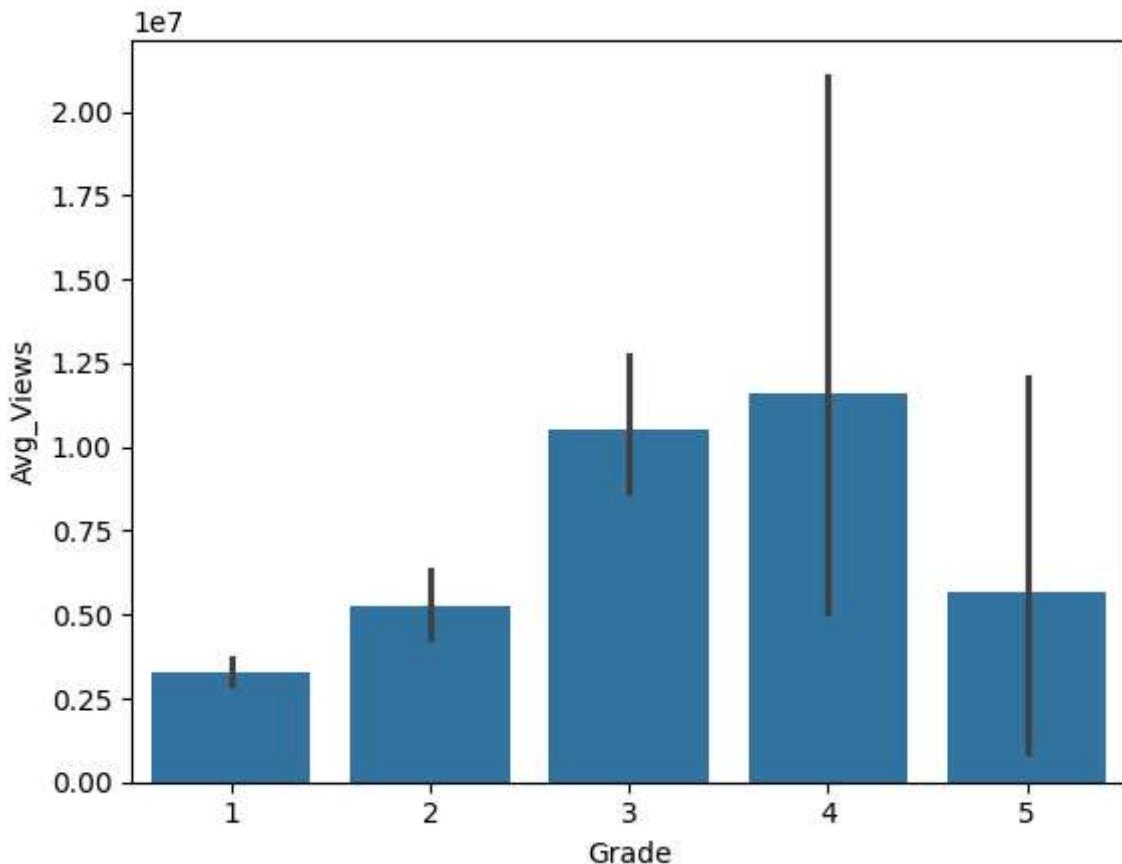
```
In [ ]: # 17.Which Grade Has A Maximum Number of Video Uploads?
sns.barplot(x='Grade',y='Video Uploads',data=data)
```

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

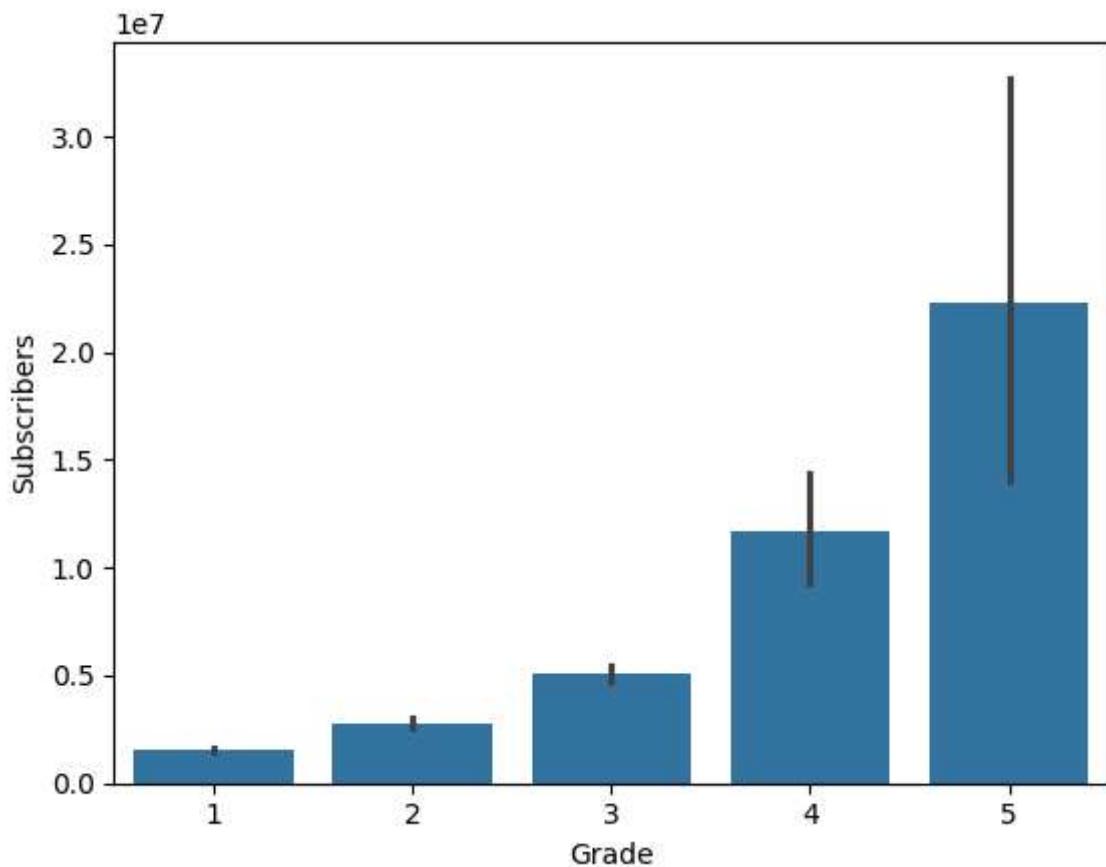


```
In [ ]: # 18.Which Grade Has The Highest Average Views?
sns.barplot(x='Grade',y='Avg_VIEWS',data=data)
```

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

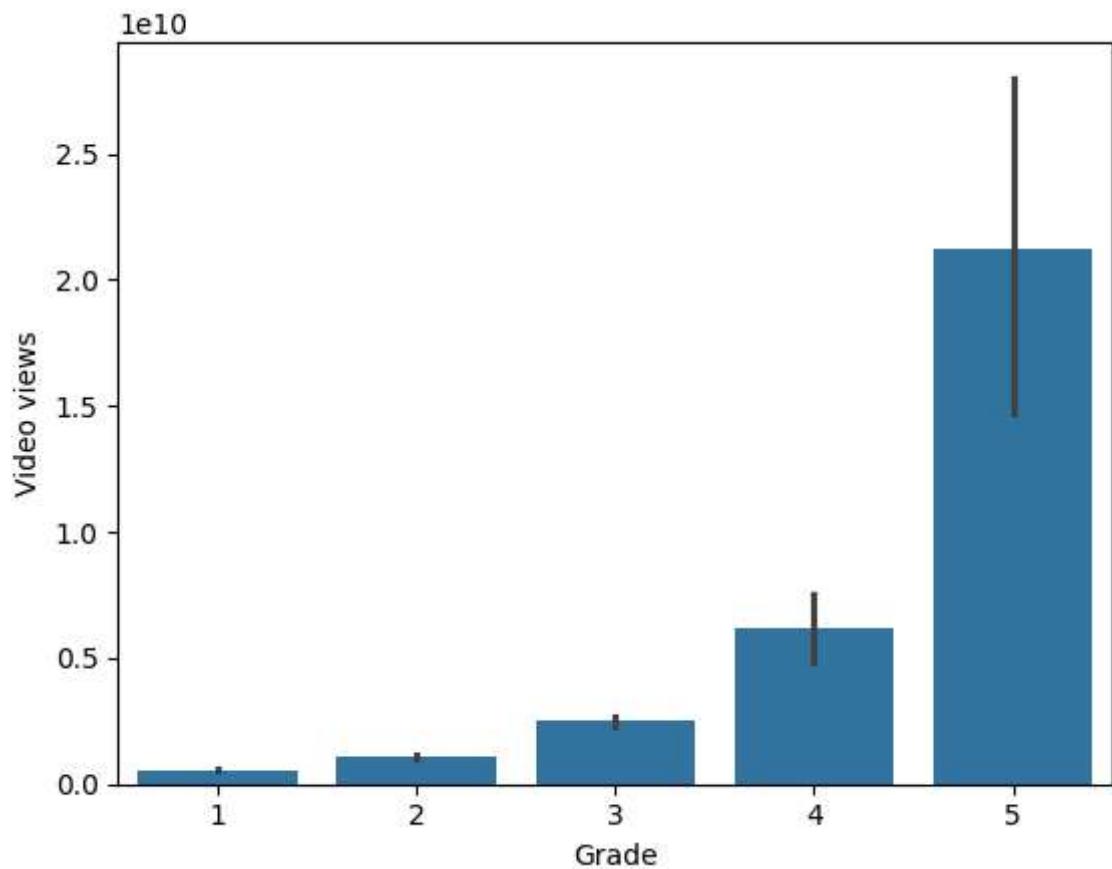


```
In [ ]: # 19.Which Grade Has The Highest Number of Subscribers?  
sns.barplot(x='Grade',y='Subscribers',data=data);
```



```
In [ ]: # 20.Which Grade Has The Highest Video Views?  
sns.barplot(x='Grade',y='Video views',data=data)
```

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



In []: