

---

# ***BBC Arabic News Report***

---

## **Summarize:**

In this report, analyze the data to enhance the channel's performance. By collecting data using [YouTube API](#) and write the script by python, then we will understanding and prepare data to make insights by visualize it.

## **Data Section:**

### **BBC.csv**

This dataset includes several columns of data = 16 and over 4000 row

Column metadata:

1. categoryId : the YouTube video category
2. title: The title of video
3. description: the description about video content
4. video\_Link: The link if video
5. defaultAudioLanguage: The video audio set by author
6. defaultLanguage: The video language set by author
7. definition: The definition of video
8. viewCount: Number of views in the video
9. likeCount: Number of likes in the video
- 10.dislikeCount: Number of dislikes in the video
- 11.commentCount: Number of comments in the video
- 12.favoriteCount: Number of people make the video in favorite list
- 13.duration: The duration of video by seconds
- 14.projection: The type of video
- 15.publishedAt: The data of publish the video
- 16.tags: The tags in the video

## Methods Section:

Explore the data by pandas library method, analysis it by numpy, visualize it by matplotlib and seaborn

```
DF.dtypes
```

```
categoryId          int64
commentCount        float64
defaultAudioLanguage  object
defaultLanguage     object
definition          object
description          object
dislikeCount        float64
duration            int64
favoriteCount        int64
likeCount           float64
projection          object
publishedAt         object
tags               object
title              object
video_Link         object
viewCount           int64
dtype: object
```

From explore found favoriteCount, defaultAudioLanguage and projection is useless columns

The publishAt column should convert to Date object

We get new column likes (publish\_year, publish\_month, publish\_day, ..., ect)



The top tags in videos we see that BBC tags is the most and the Egypt, Nuqtat Hewar Program, Syria and Yemen.

	count
بي بي سي	3021
بي بي سي عربي	2722
بي بي سي العربية	2655
اخبار البي بي سي	2433
بي بي سي اخبار	2431
مصر	885
نقطة حوار	813
سوريا	654
اليمن	637
bbc	622
السعودية	619
العراق	584
فلسطين	536
تونس	496
لبنان	476
المغرب	471
ليبيا	446
السودان	440

## Observations:

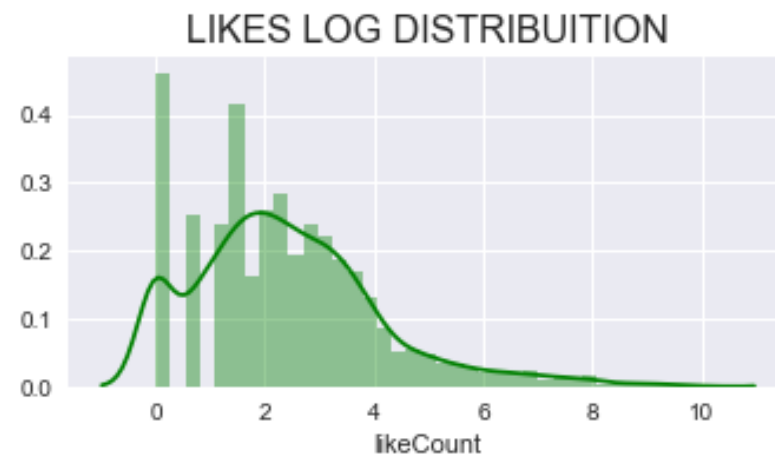
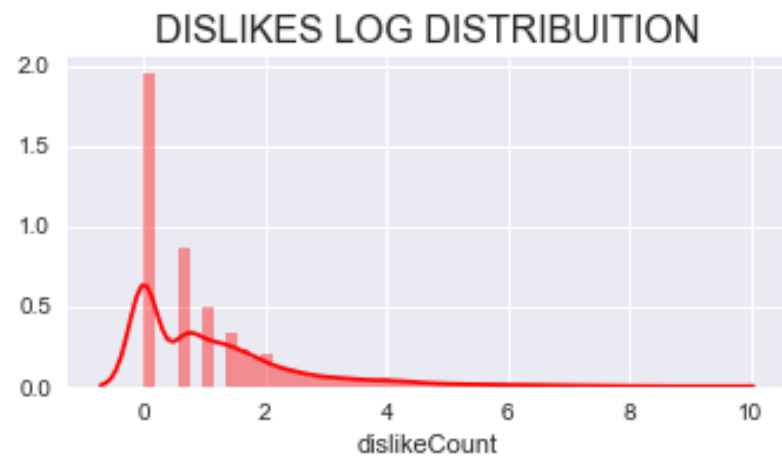
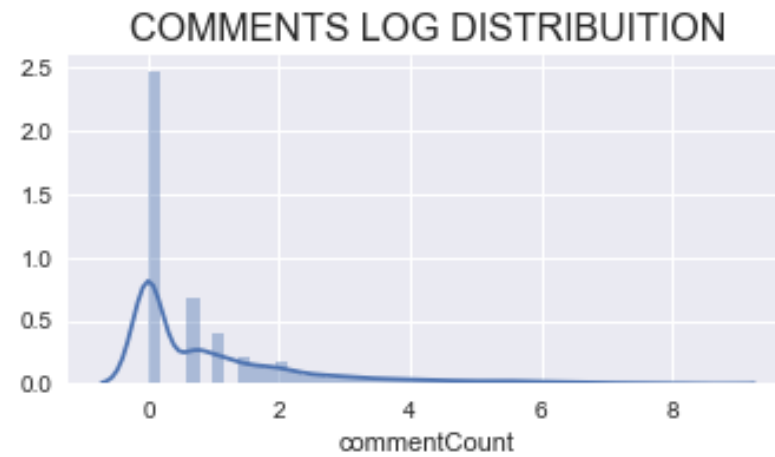
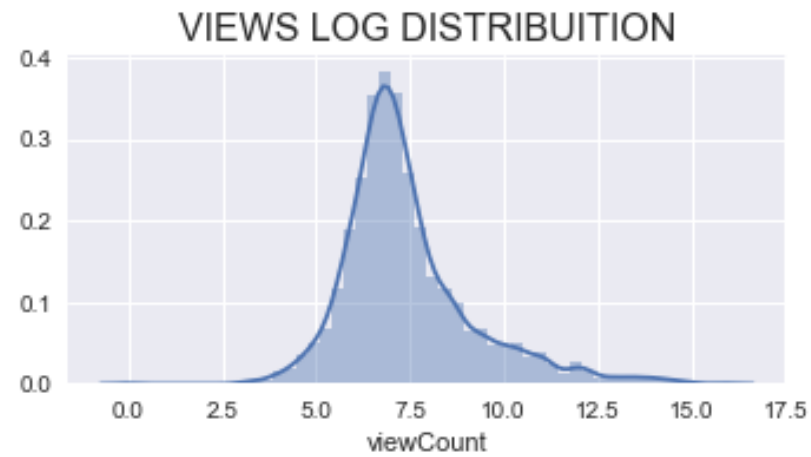
	categoryId	commentCount	dislikeCount	duration	favoriteCount	likeCount	viewCount
count	4337.000000	4169.000000	4017.000000	4337.000000	4337.0	4017.000000	4.337000e+03
mean	24.785797	34.910050	31.098083	1454.276458	0.0	133.063480	2.958303e+04
std	1.437505	204.832729	273.970155	1231.502876	0.0	842.566671	2.400191e+05
min	1.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000e+00
25%	25.000000	0.000000	0.000000	323.000000	0.0	3.000000	5.970000e+02
50%	25.000000	1.000000	1.000000	1093.000000	0.0	10.000000	1.170000e+03
75%	25.000000	4.000000	5.000000	3088.000000	0.0	29.000000	3.163000e+03
max	28.000000	5158.000000	11270.000000	6552.000000	0.0	21502.000000	8.081246e+06

The average number of views on BBC **29.5k** and the median number of views is **1.1k**. This suggests a very low average level of popularity of BBC.

On average, there are **35 comments** on every BBC video. Assuming the comments are constructive criticism, we can conclude that the BBC is lowly involved in discussions.

There is a **huge standard deviation** associated with the comments. In fact, it is even larger than the mean suggesting that the measures may be sensitive to outliers. We shall plot this to check the nature of the distribution.

The minimum number of comments on a video is **0** and the maximum is **5158**.





---

*How to get more views and interact ?!!*

---

```
DF['total_interaction'] = DF['likeCount'] + DF['dislikeCount'] + DF['commentCount']
```

```
DF['total_interaction'].mean()
```

```
197.26683804627248
```

```
DF['interaction_percent'] = (DF['total_interaction'] / DF['viewCount']) * 100
```

```
DF['interaction_percent'].mean()
```

```
1.2192115362251783
```

```
DF['interaction_percent'].max()
```

```
11.594202898550725
```

```
DF['dislike_percentage'] = (DF['dislikeCount'] / (DF['dislikeCount'] + DF['likeCount']) * 100)
```

```
DF['dislike_percentage'].mean()
```

```
15.774550654854831
```

From this queries we will found that interact with BBC videos is very weak the average of (like, dislike and comment) is 197 from the viewers by percentage 1.2% from all views.

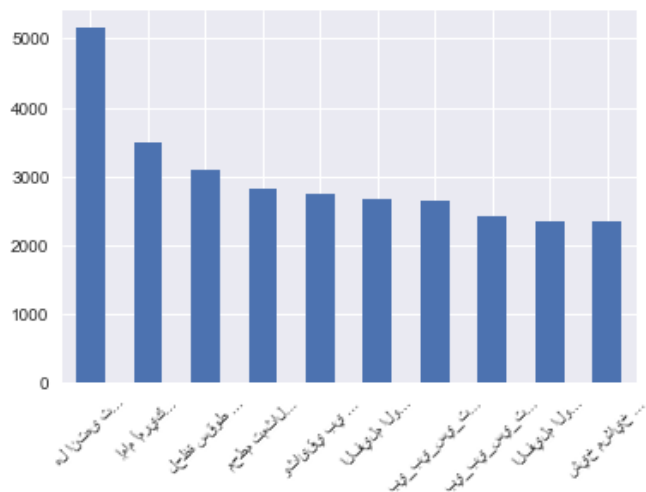
	viewCount	likeCount	dislikeCount	commentCount
viewCount	1.000000	0.656276	0.468667	0.457238
likeCount	0.656276	1.000000	0.546461	0.616247
dislikeCount	0.468667	0.546461	1.000000	0.649403
commentCount	0.457238	0.616247	0.649403	1.000000

From correlation, we found that a strong positive correlation between like and views, positive correlation between like and comments and positive correlation between dislike and comments

## The most videos get comments

### Most comments in controversial topics

- 1 ) هل انتهى تنظيم الدولة الإسلامية فعلا ؟
- 2 ) إمام أمريكي متلى يزوج المسلمين المتليين
- 3 ) لحظة سقوط الرئيس اللبناني ميشال عون أرضا في القمة العربية
- 4 ) محطم تمثال صدام حسين نادم على فعلته
- 5 ) "وثائقي بي بي سي" أمراء آل سعود المخطوفون
- 6 ) الفيلم الوثائقي: جنود إسرائيل العرب
- 7 ) بي\_بي\_سي\_ترندينغ: إمام مسجد أردني يصلي التراويح وحده بعد مغادرة المصلين في منتصف الصلاة #الأردن
- 8 ) بي\_بي\_سي\_ترندينغ | غضب في #السودان من مسلسل "بلوك غشمرة" الكويتي #الكويت
- 9 ) الفيلم الوثائقي: حصار مكة
- 10 ) شيخ مشايخ قبيلة شمر حميدي دهمام الجريا - بلا قيود

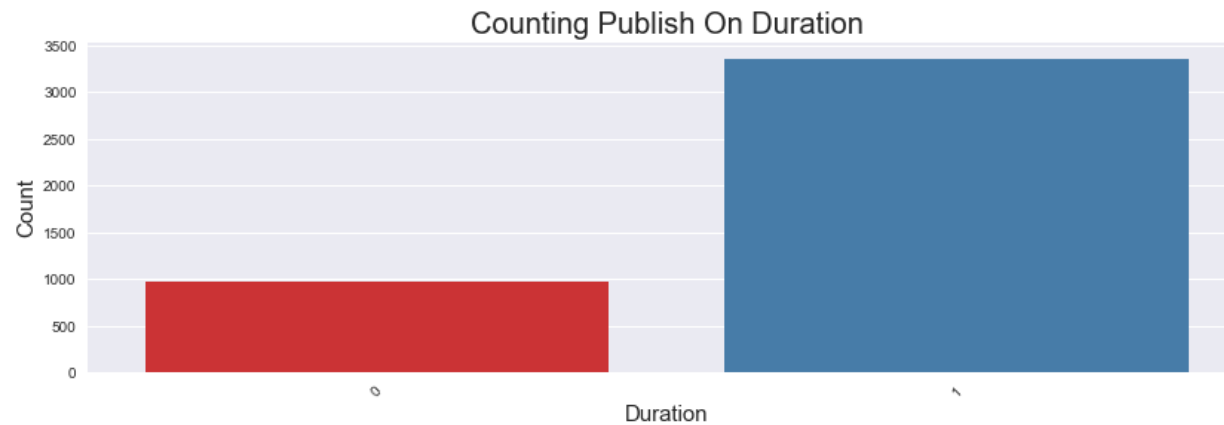


---

***BBC must focus in interactive content controversial and emotional***

---

The rate of publish the videos at AM (0) or PM (1)



BBC publish most videos at PM but I found that the average of PM videos 30k view however AM videos 26K view

```
publish_durationCnt = DF['publish_duration'].value_counts()
```

```
PMview = DF[DF['publish_duration'] == 1]['viewCount'].sum()
```

```
AMview = DF[DF['publish_duration'] == 0]['viewCount'].sum()
```

```
PMview / publish_durationCnt[1]
```

```
30607.26531826294
```

```
AMview / publish_durationCnt[0]
```

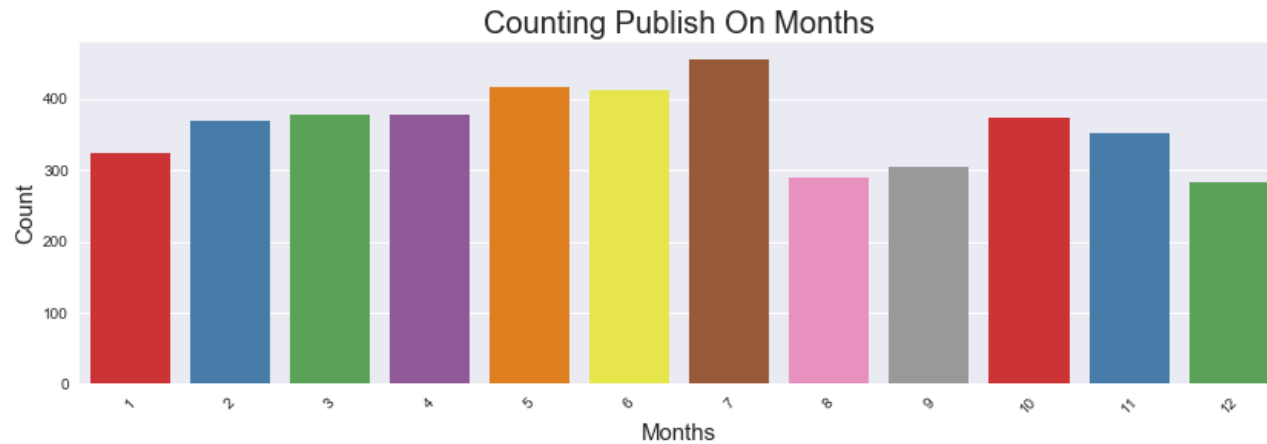
```
26051.244102564102
```

---

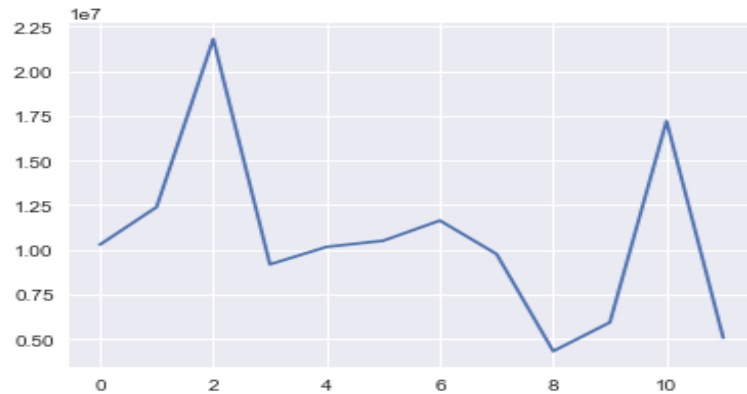
***BBC should publish more videos at AM***

---

The rate of publish the videos at Months of the year



We see that BBC publish more videos at (7, 5, 6) but the awesome I found that most months get views (3, 11, 2) and (10, 4) that have many publish get less views



	views	Days
2	21802566	3
10	17213622	11
1	12394119	2
6	11639803	7
5	10509202	6
0	10289295	1
4	10165289	5
7	9764053	8
3	9184531	4
9	5935839	10
11	5076343	12
8	4326927	9

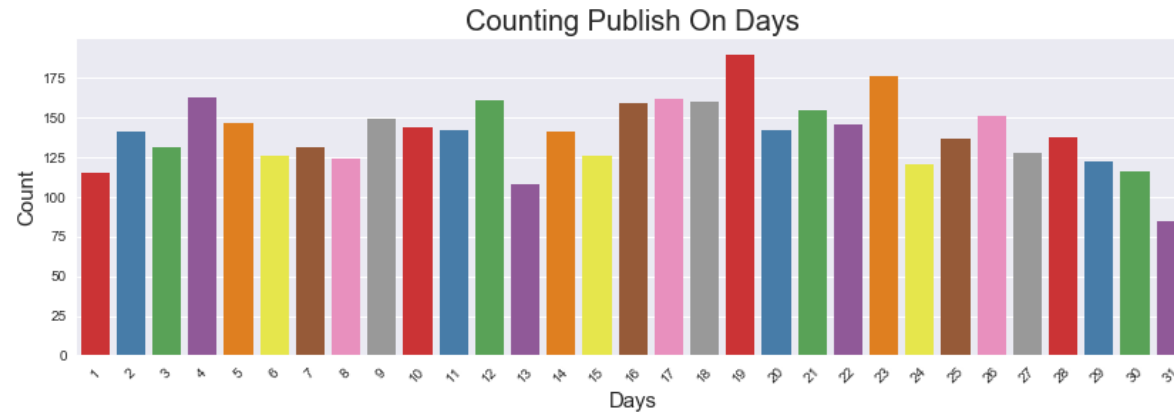


---

*BBC should publish more videos at months (3, 11, 2, and 1)*

---

## The rate of publish the videos at Days of the month



We see that BBC publish more videos at (19, 23, 4, 12) but the awesome I found that most days get views (15, 14, 6, 17) and (19, 4, 12) that have many publish get less views.



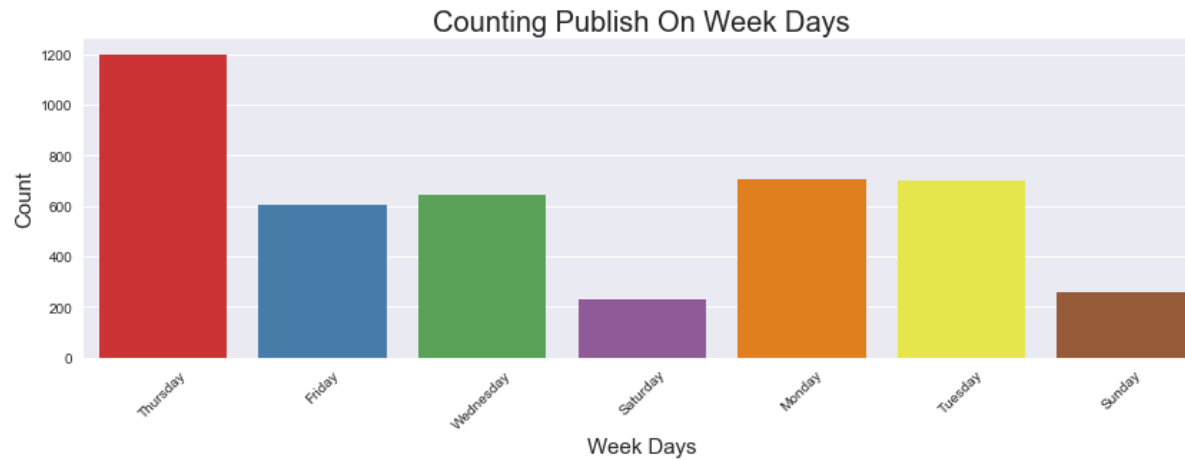
views Days		
14	12997231	15
13	10119756	14
5	8725385	6
16	6683349	17
29	6305460	30
20	6113746	21
17	5674287	18
22	5589487	23
26	5396505	27
18	5116092	19
9	5018402	10
4	4903899	5
25	4868131	26
0	4641105	1
28	4538910	29
6	3862139	7
10	3046040	11
15	2931606	16
3	2894083	4
11	2731083	12
21	2124411	22
23	2123647	24
27	2032508	28
7	1814719	8
19	1813763	20
8	1404931	9
30	1323515	31
1	923347	2
24	883859	25
2	855740	3
12	844453	13

---

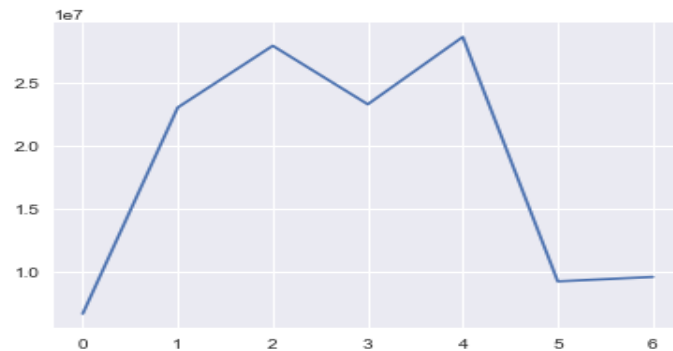
*BBC should publish more videos at days (15, 14, 6, 17, 30, 21, 18, 23, 27, and 19)*

---

## The rate of publish the videos at Week Days of the week



We see that BBC publish more videos at (Thursday, Tuesday, Monday) but the awesome I found that Saturday get more veiws however it has small number of video publish.



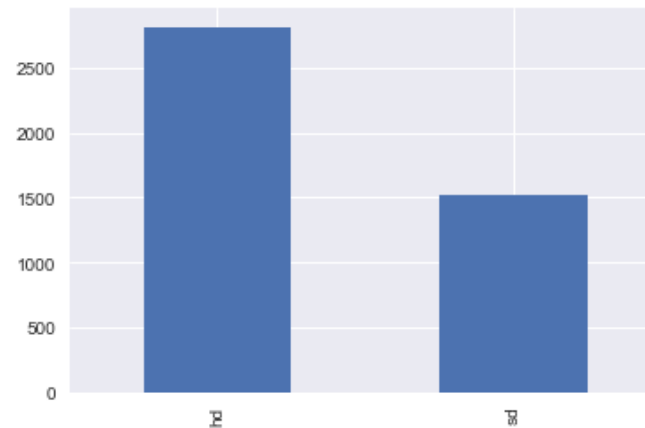
	views	week_days
4	28651778	Thursday
2	27950188	Tuesday
3	23291546	Wednesday
1	23036696	Monday
6	9549683	Saturday
5	9196169	Friday
0	6625529	Sunday

---

*BBC should publish more videos at weekday (Saturday and Wednesday)*

---

The rate of publish the videos by quality definition



BBC publish most videos at definition HD quality but I found that the average of HD quality videos 21k view however SD quality videos 44K view

```
definitionCnt = DF['definition'].value_counts()
```

```
hdViews = DF[DF['definition'] == 'hd']['viewCount'].sum()
```

```
sdViews = DF[DF['definition'] == 'sd']['viewCount'].sum()
```

```
hdViews / definitionCnt['hd']
```

```
21700.954174067494
```

```
sdViews / definitionCnt['sd']
```

```
44161.237187910643
```

---

***BBC should publish more videos by quality definition SD***

---