

Story_telling

December 19, 2019

0.1 Story-telling:

0.2 Exploratory data analysis: YouTube videos from the US and UK

0.2.1 Introduction:

- YouTube is a video-sharing platform based in California, USA. , Three former employees of PayPal, Chad Hurley, Steve Chen, and Jawed Karim, devised this platform in February 2005. Google bought the site in November 2006. (source: Wikipedia)
- YouTube allows users to upload, view, rate, share, add to playlists, report, comment on videos, and subscribe to other users. It offers a wide variety of user-generated and corporate media videos. (source: Wikipedia)

0.2.2 Objectives:

- The objective of the project is to compare behavioral patterns of youtube viewers in the US and UK. Details of the trending YouTube videos from the US and the UK, for example, views, comments, likes, and dislikes will be analyzed.
- The project aims to find answers to the following questions: how are the top trending videos different in terms of views, categories, likes and dislikes; whether there exists any seasonal pattern in some specific categories; how are videos categories different in the US and UK.
- The project covers applications of data loading, data cleaning, exploratory data analysis, and inferential statistics.

0.2.3 Basic statistics of the number of YouTube views:

- We observe that distributions of both US and UK views are right-skewed.
- Few YouTube videos are virul with extermley high number of views.
- The mean number of views in the US is 939,102; while the mean number of views in the UK is 1,110,467.
- As there are some extreme number of views, standard deviation of both distributions are very high.
- The standard deviation of the US views is 2,147,691, which 2.3 times larger than its mean.

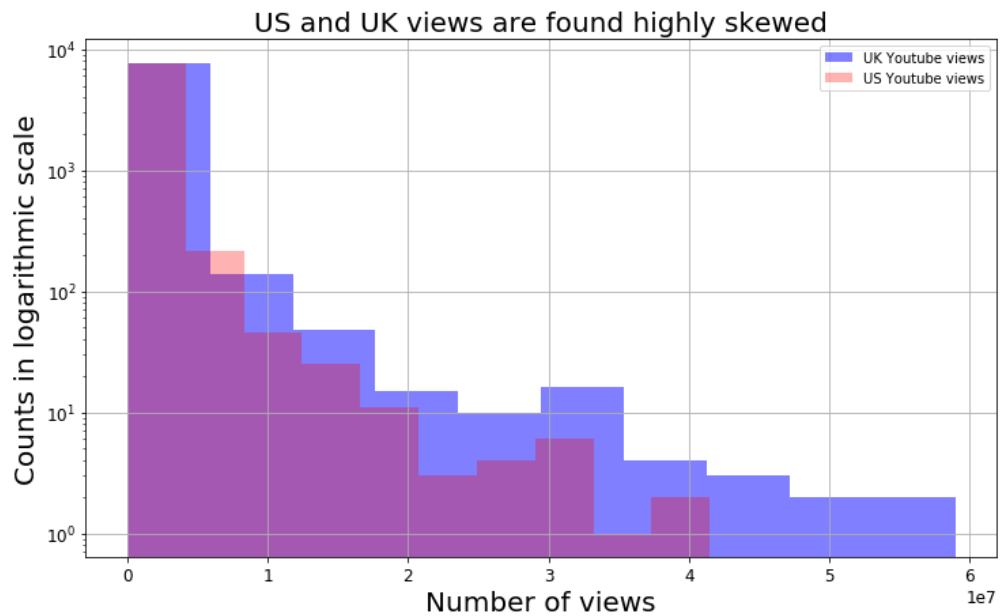
Basic statistics of the US YouTube views:

```
count    7.998000e+03
mean     9.391016e+05
std      2.147691e+06
min      0.000000e+00
25%      9.690000e+04
50%      3.086115e+05
75%      9.595132e+05
max      4.150067e+07
Name: views, dtype: float64
```

Basic statistics of the UK YouTube views:

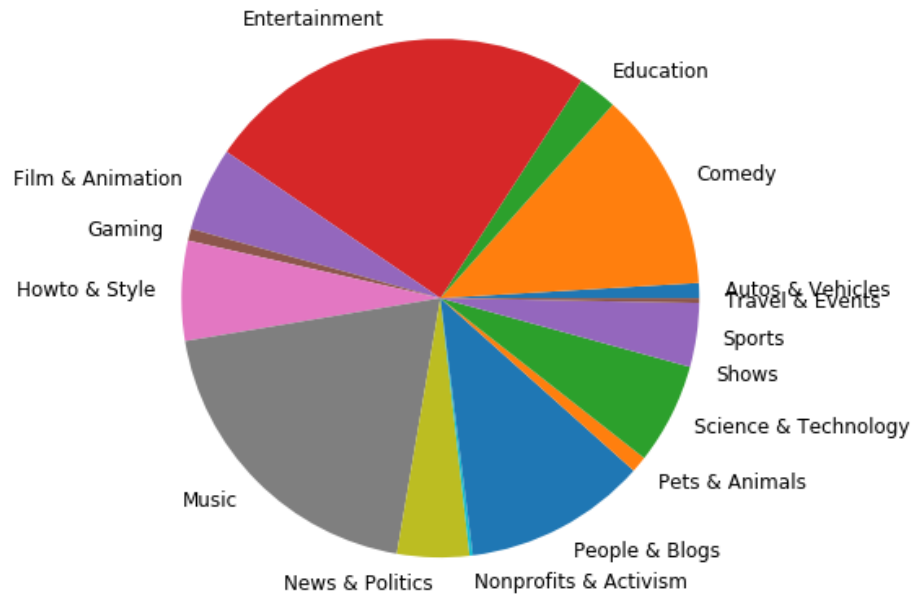
```
count    7.995000e+03
mean     1.110467e+06
std      3.048405e+06
min      0.000000e+00
25%      1.082980e+05
50%      3.150910e+05
75%      9.710280e+05
max      5.896141e+07
Name: views, dtype: float64
```

title



title

Proportions of the US YouTube viewers during Sep. and Oct., 2017



"Pie Chart"

0.2.4 Distribution of the total number of views in the US and UK:

- We observe that distributions of both US and UK views are right-skewed.
- Few YouTube videos are virul with extermley high number of views.
- The mean number of views in the US is 939,102; while the mean number of views in the UK is 1,110,467.
- As there are some extreme number of views, standard deviation of both distributions are very high.
- The standard deviation of the US views is 2,147,691, which 2.3 times larger than its mean.

0.2.5 Proportion of YouTube views by video categories in the US:

- It is observed that in the US, most viewed video catogiries are: Enetrainment, Music, Comedy, and People & Blogs

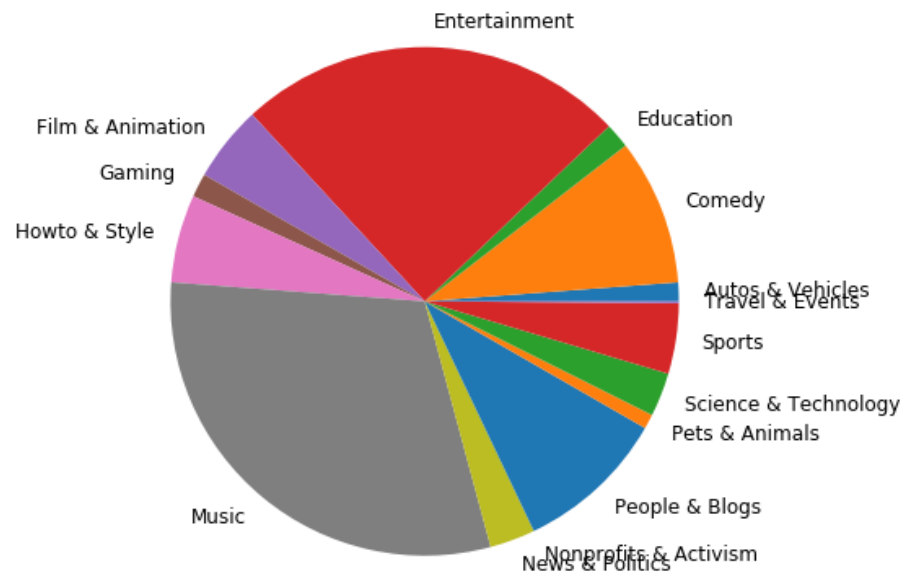
0.2.6 Proportion of YouTube views by video categories in the UK:

- In the UK, the most viewed video categories are: Music, Entertainment, Comdey, and People & Blogs

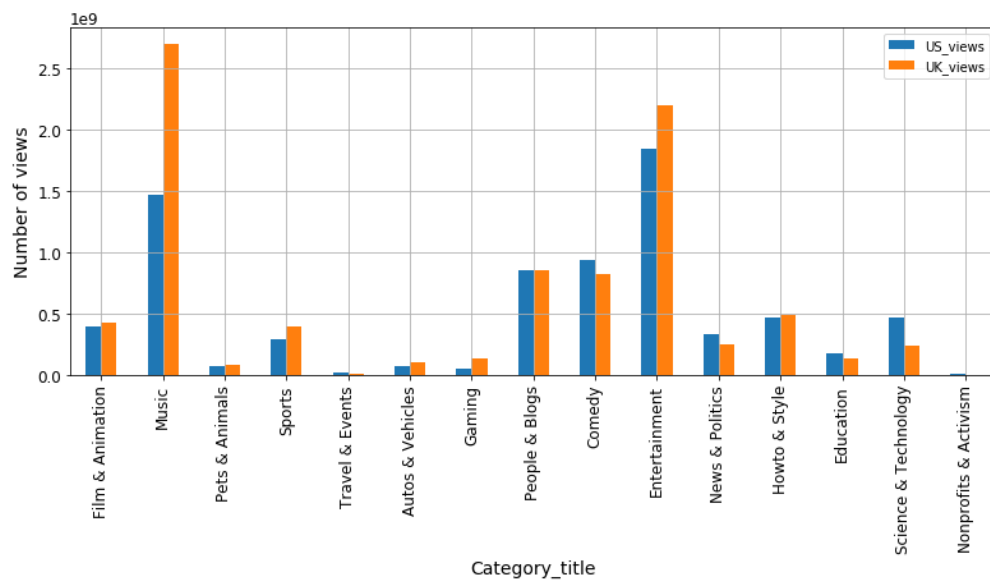
0.2.7 Comparison of the number of views by video categories in the US and UK:

- In the UK, there are more number views in the categories of Music, Entertainment, and Sports than the number of views in the US.
- However, in the US, there are more number of views in the categories of Comedy, News & Politics, Education, Science & Technology.

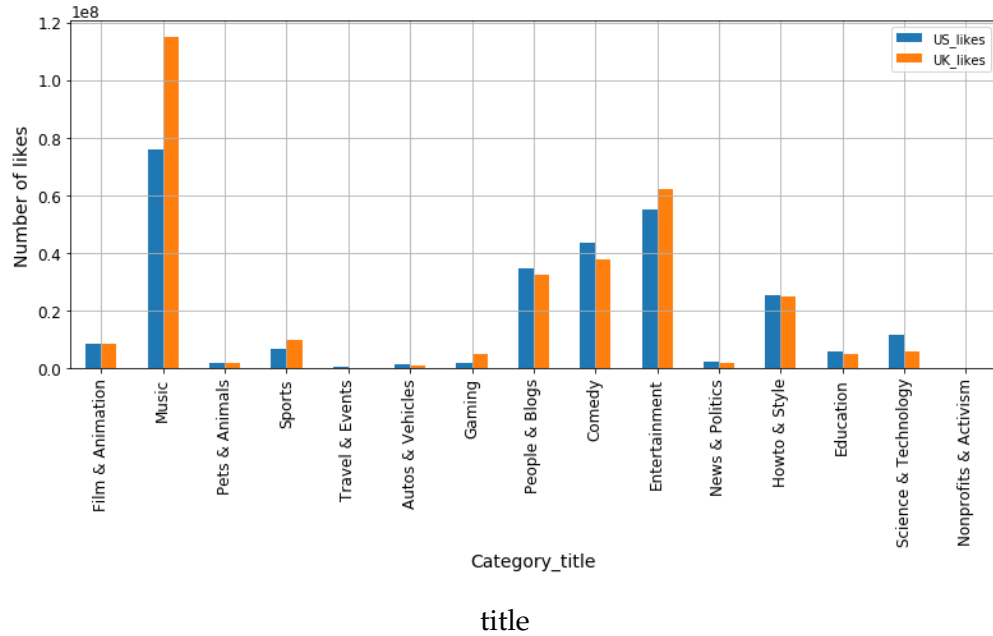
Proportions of the UK YouTube viewers during Sep. and Oct., 2017



"Pie Chart"



title



0.2.8 Comparison of the number of likes by categories in the US and UK:

- It is observed that the number of likes in the UK outnumbers the number of likes in the US by Music and Entertainment.
- But, in the US, in the categories of People & Blogs and Comedy, US likes outnumber that of in the UK.

0.2.9 Comparison of the number of dislikes by categories in the US and UK:

- It is found that video categories with higher number of views and likes have also higher number of dislikes.

0.2.10 Category-wise scatterplots between number of views and number of likes in the US:

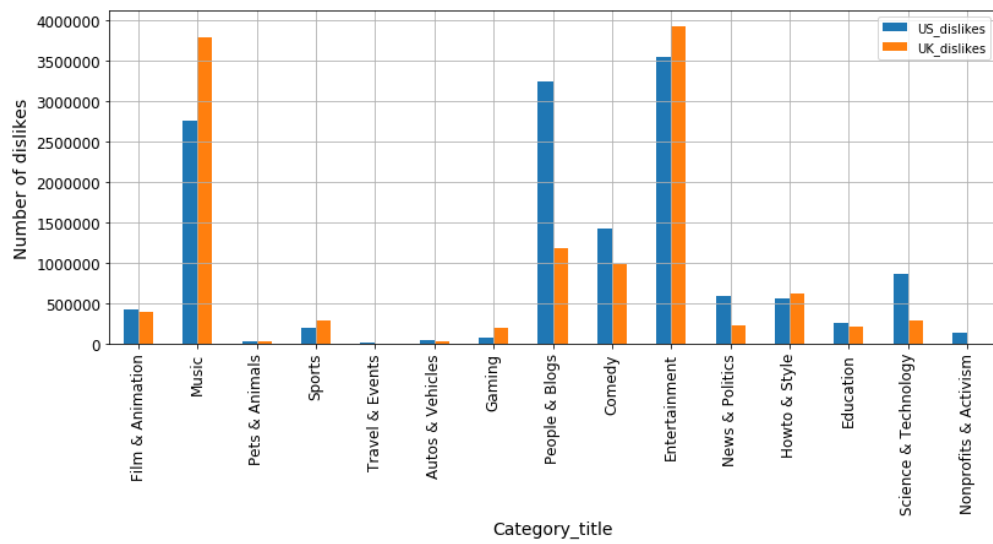
- Category wise scatterplots between number of views and number of likes show positive trends.

0.2.11 Scatterplot as a whole:

- it is observed that Entertainment and Music have higher number of views as well as higher number of likes

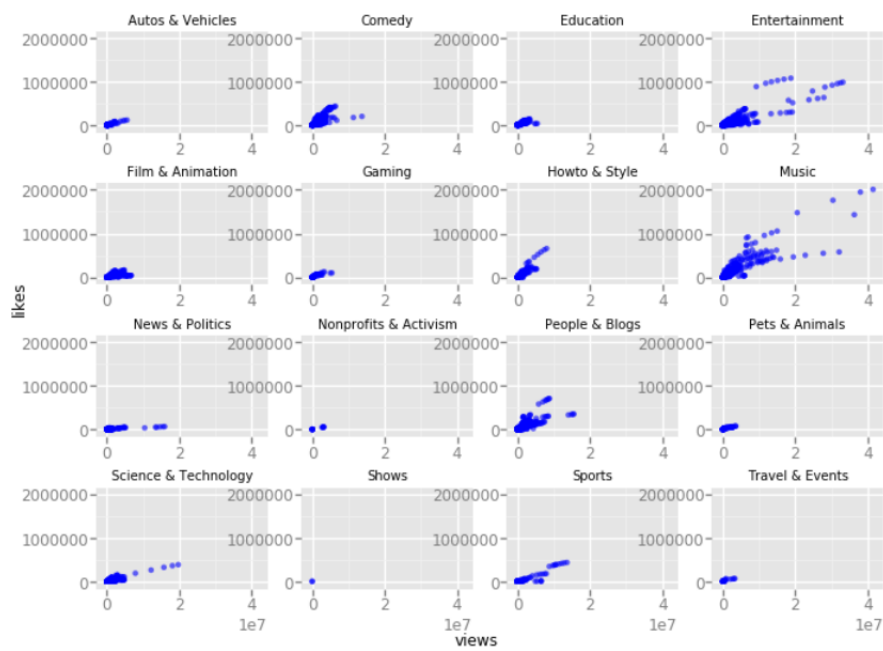
0.2.12 Correlations among number of views, likes, dislikes, and total number of comments

- It is found that -
 - Correlation coefficient between number of likes and number of views is 0.83
 - Correlation coefficient between number of dislikes and number of views is 0.54
 - Correlation coefficient between number of comments and number of views is 0.72



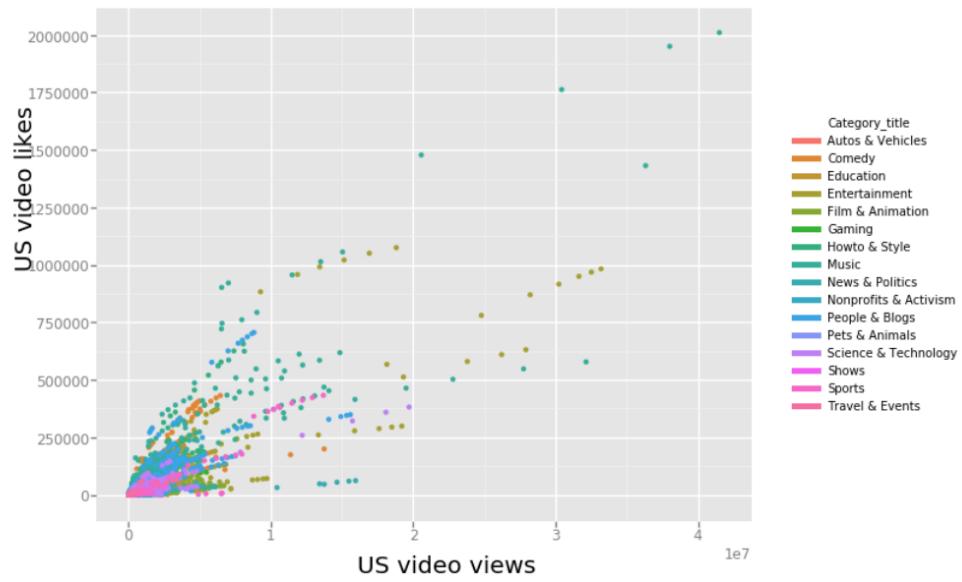
title

US views vs. likes by video category

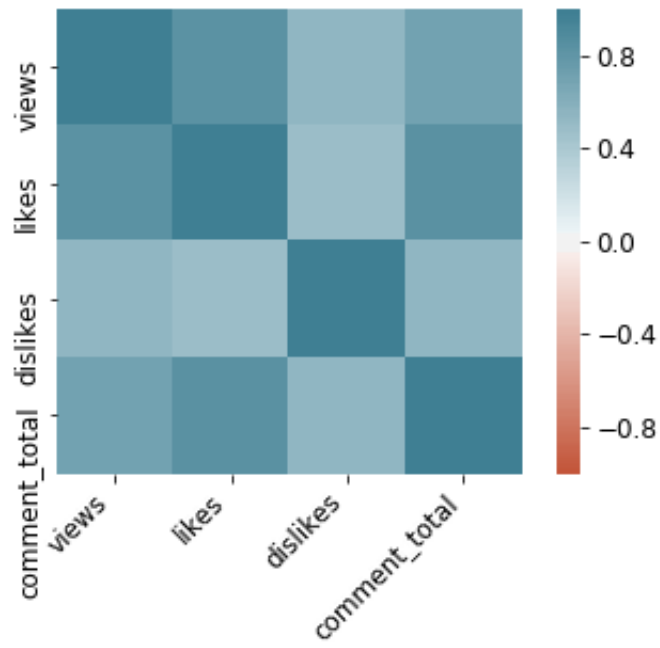


title

US views vs. likes by video category



title



title



0.2.13 Trending YouTube video views in the US:

- The above figure shows the number of views of particular videos that were high on September 13, 2017, declines rapidly by September 19, 2017.
- This shows that the number of views declines quickly over a week.

0.2.14 Conclusions:

- This is an exploratory data analysis of comparison of number of YouTube views, likes, dislikes in the US and UK.
- The dataset is collected from www.kaggle.com
- It is found that the most viewed video categories are Entertainment and Music.
- The least video categories are Pets & Animal, Travel & Events
- The number of views and number of likes are highly correlated. It is quite natural that videos with higher number of views have higher number of likes.
- Distributions of total number of views are right-skewed. Some viral videos have extreme number of views.
- Statistical inference analysis shows that the mean of distributions are significantly different at 5% confidence level.

In []: