2019/07/01 Homework

**(a)**

Trending YouTube Video Statistics

**Daily statistics for trending YouTube videos**

[Mitchell J](https://www.kaggle.com/datasnaek)  updated a month ago (Version 115)

### Context

YouTube (the world-famous video sharing website) maintains a list of the [top trending videos](https://www.youtube.com/feed/trending) on the platform. [According to Variety magazine](http://variety.com/2017/digital/news/youtube-2017-top-trending-videos-music-videos-1202631416/), “To determine the year’s top-trending videos, YouTube uses a combination of factors including measuring users interactions (number of views, shares, comments and likes). Note that they’re not the most-viewed videos overall for the calendar year”. Top performers on the YouTube trending list are music videos (such as the famously virile “Gangam Style”), celebrity and/or reality TV performances, and the random dude-with-a-camera viral videos that YouTube is well-known for.

This dataset is a daily record of the top trending YouTube videos.

Note that this dataset is a structurally improved version of [this dataset](https://www.kaggle.com/datasnaek/youtube).

### Content

This dataset includes several months (and counting) of data on daily trending YouTube videos. Data is included for the US, GB, DE, CA, and FR regions (USA, Great Britain, Germany, Canada, and France, respectively), with up to 200 listed trending videos per day.

EDIT: Now includes data from RU, MX, KR, JP and IN regions (Russia, Mexico, South Korea, Japan and India respectively) over the same time period.

Each region’s data is in a separate file. Data includes the video title, channel title, publish time, tags, views, likes and dislikes, description, and comment count.

The data also includes a category\_id field, which varies between regions. To retrieve the categories for a specific video, find it in the associated JSON. One such file is included for each of the five regions in the dataset.

For more information on specific columns in the dataset refer to the [column metadata](https://www.kaggle.com/datasnaek/youtube-new/data).

### Acknowledgements

This dataset was collected using the YouTube API.

* 問題定義

關於youtube上的熱門影片做數據分析，分析出最後歡迎的影片種類及各國家不同文化差異而喜愛的影片種類。

* 潛在問題

近期興起的youtuber職業，會讓影片的種類出現難以分類的情形、各國間也會因為影片語言的差異，使得分析較為困難。

* 分析與預測難度

本篇分析的數據來源由YouTube AP所收集，在收集的當下僅以點閱率為主、其中有不確定的點開而未感興趣、甚至是dislike的情形，必定會造成一些誤差。

除此之外，隨著潮流的發展，將結果視為未來預測，短時間或許準確、長時間就會出現明顯預測錯誤！

* 價值

可以藉由喜愛影片，提供廣告業者製作廣告的方式吸引觀眾、也可以提供youtube公司加載廣告，可以選擇較為受喜愛影片進行穿插廣告、藉此使得廣告收益達到最大效益！

**(b)**

我就讀化學碩；可以藉由從系辦查得成大化學所的畢業人數及畢業流向、進而統計資料。

可以從這些數據得到下列結果或預測:

1. 就讀化學碩士班大致上有多少人會延畢及順利畢業
2. 畢業後的工作流向、進而推導出近期產業結構，哪些比較需要化學人才
3. 提供學校作為演講、生涯規劃、課外技能的額外資訊，可以多舉辦揖謝對於學生有所幫助的課程
4. 可以試著預測，化學所之中，哪些老師最容易讓學生延畢，提供系上做秘密調查。