原有features：

1. Channel frequency: According to the dataset, during 213 days, different channels have different number of video posting. We calculate the video posting frequency and categorize it into three categories: low, moderate and high.
2. Facebook link: This is a boolean categorical feature that demonstrates whether the description of the video contains facebook links and this features has true or false as values.
3. Instagram link: Same as the facebook link feature. It lists true or false for whether the video contains instagram link.
4. Length of title: This column contains integer values, indicates the length of words of the title.
5. Number of tags: This column contains integer values, indicates the number of tags used.
6. Number in title: This column contains if the title contains number, this feature has true or false as values.
7. Special characters in title: This column contains the title contains special characters(such as “!” “?”), this feature has true or false as values.
8. Days between the published date and data collection date: This column contains how long it is from the published date of the video to the date that we get the data.
9. Daily views/likes/dislikes/comments: This column contains the count of daily views/likes/dislikes/comments.
10. Classification of the daily views: In this column, we divide daily views  into 3 tags “small”, ”middle”, ” large”, according to the values, each of tag has ⅓ of the total data.

新加Features：(灰色表示已完成)

1. 标题动词位置
2. 视频长度
3. 视频品质
4. 标题加入括号
5. 标题加入年份 （所有都没有年份）
6. 有没有description
7. 缩略图
8. 经常上传视频也是影响排名很重要的指标：上传频率低中高
9. 标题中stop words的个数
10. Tags中stop words的个数
11. Description，title，tag是否包含某个重要的关键字？
12. 合適或熱門的關鍵字
13. 早上/中午/晚上发布的视频 -> 发布在哪个小时（24小时制）
14. 标题长度分三档，低中高，哪种对提高view 有正面影响？哪种有负面影响？
15. Tags数量分三档，低中高，哪种对提高view有正面影响？哪种有负面影响？

Insights：

會有這樣的轉變可能和YouTube已經能夠透過語音判讀影片內容有關，也就是說，YouTube不必再只透過標題、影片描述、關鍵字以及影片中的字幕來瞭解影片的內容，可以透過影片中的聲音來辨識影像的內容。從另一個研究的調查結果也可以發現，將關鍵字放在影片的標題中，其不必然和排名有高度相關：