Online Video Task

Goal

Company XYZ is an online video streaming company, just like YouTube or Dailymotion.

The company has a website, in which they are publishing videos. In the website, video can be placed in some spots, which can be vary in their exposure. So, there are some spots (e.g. top page) which consider more “hot spots” than the other (e.g. bottom page)

This company is interested in knowing whether a video is "hot" , stable or going down. Understanding this would allow to optimize the videos promoted on the home-page and, therefore, maximize ads revenue.

Challenge Description

The Head of Product has identified as a major problem for the site a very high home page drop-off rate. That is, users come to the home-page and then leave the site without taking any action or watching any video. Since customer acquisition costs are very high, this is a huge problem: the company is spending a lot of money to acquire users who don't generate any revenue.

Currently, the videos shown on the home page to new users are manually chosen. The Head of Product had this idea of creating a new recommended video section on the home page.

She asked you the following:

- Classify each video into one these 3 categories:
    - "Hot"
    - "Stable and Popular"
    - "Everything else"
- What are the main characteristics of the "hot videos"?
- After having identified the characteristics of the hot videos, how would you use this information from a product standpoint?

Data

The 2 tables are:

video_count - provides information about how many times each video was seen by day

Columns:

- **video_id** : video id, unique by video and joinable to the video id in the other table
- **count** : total count of views for each video
- **date** : on which day that video was watched

video_features - characteristics of the video.

Columns:

- **video_id** : video id, unique by video and joinable to the video id in the other table
- **video_length** : length of the video in seconds
- **video_language** : language of the video, as selected by the user when she uploaded the video
- **video_upload_date** : when the video was uploaded
- **video_quality** : quality of the video. It can be [240p, 360p, 480p, 720p, 1080p]

Example

Let's check one video: how many times was it seen on a given day?

head (video_count, 1)

Column Name	Value	Description
video_id	2303	id of the video
count	22	it was seen just 22 times.
date	2015-01-07	on Jan, 7th.

Let's now check the characteristics of that video 2303:

subset(video_features, video_id == 2303)

Column Name	Value	Description
video_id	2303	It is the video we care about. Same as above
video_length	1071	the video lasts almost 18 min (1071 seconds)
video_language	Cn	the video is in Chinese
video_upload_date	2014-12-10	was uploaded on Dec, 10
video_quality	1080p	video quality is 1080p, i.e. very high