YouTube Trending Prediction

Technology Reviews

Bingwen Hu, Xinran Wang, Dau Cheng, Rui Wang

Project Overview

YouTube shows the most popular and currently trending topics all over the world which are refreshed every day.

It is worthwhile to track these data every day if you need prepare for taking new films, select recommended videos or evaluate the value of videos before business collaboration.

We are mainly targeting YouTube creators, operators and advertisers.

Requirement addressed by Technologies

Category Bubble Chart for trending topics

- Bubble Chart will display top 20 trending categories.

Word Cloud for top #hashtags

- Trending tags will be highlighted and their size is based on relevant features.

Line Chart for tracking individual videos

- To display top 10 popular videos' tendency during a period of time.

Techonologies Overview

Tableau

D3.js

Matpotlib

Shiny

Plotly

Bokeh

Seaborn

Altair

Choice of Technologies

D3.js

D3.js can be used for the Bubble Chart and the Word Cloud, offering the final interactive presentation.

Tableau

Tableau is a visual analytics platform transforming the way we use data to solve problems—empowering people and organizations to make the most of their data.

Matpotlib

In our project, Matplotlib is mostly used for intermediate tests when developing our application. We can easily import this package and plot some figures inside notebooks or PyCharm, which can help us check the data, etc.

Appeal of choice

Tableau	D3.js	Matplotlib
Intuitive and user-friendly interface that are easy to use	Easy to customize, available for animations and interactivity plots	Easy to install and use inside Python, works fine with data science packages
Support full functionality for mobile user.	Can be easily hosted on the webacross browers	Accessible and detailed documentation, bunch of references
Fast and reliable even on big data	Fast and light on system resources	Good for back-end intermediate tests and plots

Drawbacks of choice

Tableau	D3.js	Matplotlib
No automatic refreshing	Has a steep initial learning curve	The basic plots are very ugly, mostly scientifically good but aesthetically bad
High cost (free for UW student)	Requires web development experience	Low level and lack of web interaction functions
Not smooth when embedded on web or integrated with other tech	Slow when dealing with large datasets	Not good for time series data

YouTube Trending Prediction

Standup

Bingwen Hu, Xinran Wang, Dau Cheng, Rui Wang

Project Plan

Progress in this period:

- Select technologies to satisfy requirement of our project.
- Data Mining and Cleaning

Deliverables for next period:

- Load and apply ML to hashtags and
- Visualize data leveraging packages