Data Pipeline Implementation for YouTube Data

Andrew Wong
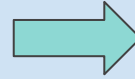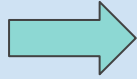
Goals:

- **Ingest YouTube Trending Video Data into Database**

- **Build clean processing Pipeline**

- **Perform EDA to find interesting data insights**

- **Deploy interactive web application**

Workflow:

Data:

- 500,000+ (and counting) trending YouTube Videos from the past 9 months
- 11 Countries (India, USA, Great Britain, Germany, Canada, France, Russia, Brazil, Mexico, South Korea, and, Japan) in separate tables
- 50,000+ Rows per country

- Updated daily with 200 trending videos

- 16 feature columns (Channel Title, Views, Category, Trending Date, Likes and Dislikes)

Data Cleaning:

- **Description column**

  - **YouTube allows up to 5,000 characters (1-2 pages!)**

  - **Information not very useful, no common format/convention**

  - **Removing Description = 70% file size reduction!**

Design: Web Application

- Built user-friendly, interactive web application via Streamlit

- Algorithms and filtered aggregation with a few clicks!

- Allows for user input/selection (Country, Category, Date)
  - Returns visualizations
  - Can benefit marketing/advertising teams
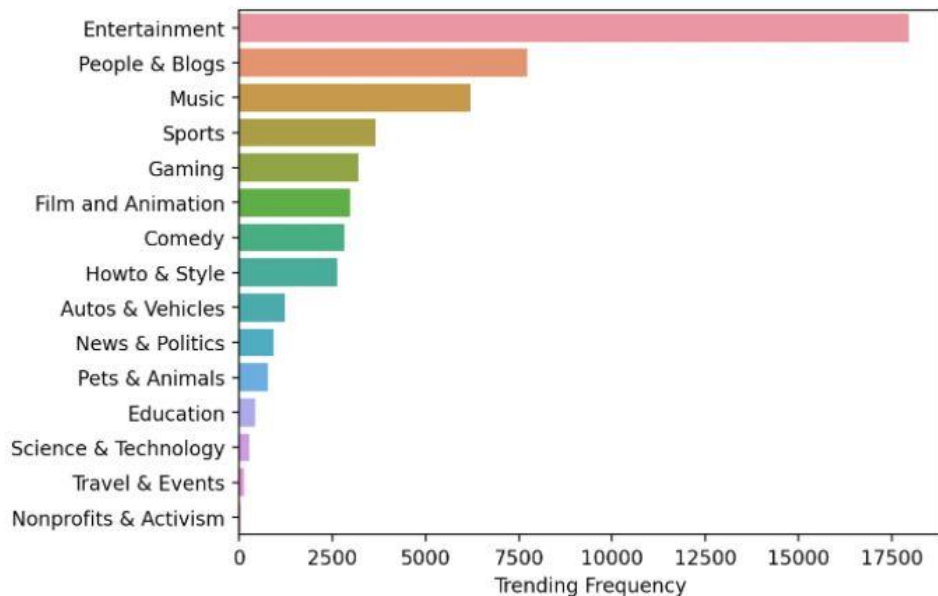
YouTube Trending Video Data By Country

Top Trending Video Categories

Top Trending Video Categories in JP

Filters

Pick a country to show some info about

JP

Pick a Video Category

ALL
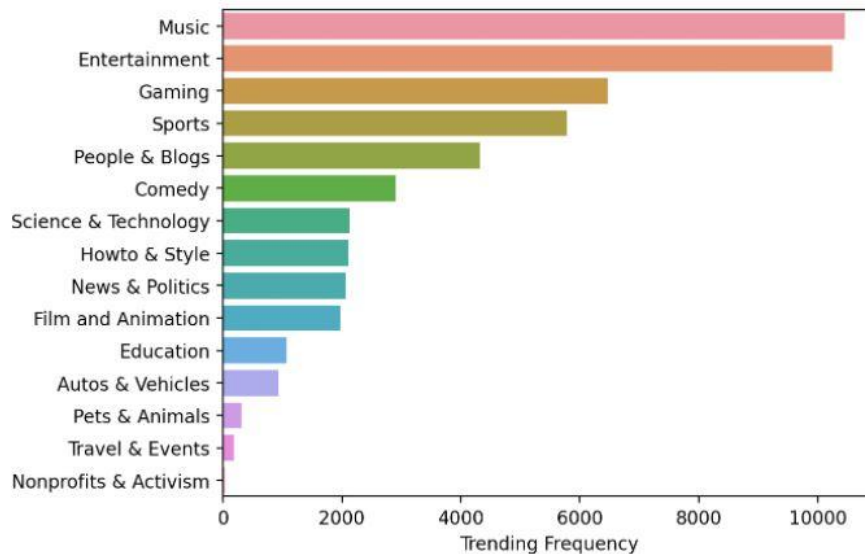
Pick a date range

Start date

2020/08/12

End date

2021/04/27

Filters

Pick a country to show some info about

US

Pick a Video Category

ALL

Pick a date range

Start date

2020/08/12

End date

2021/04/27

YouTube Trending Video Data By Country

Top Trending Video Categories

Top Trending Video Categories in US

YouTube Trending Video Data By Country

Filters

Pick a country to show some info about

US ▾

Pick a Video Category

Music ▾
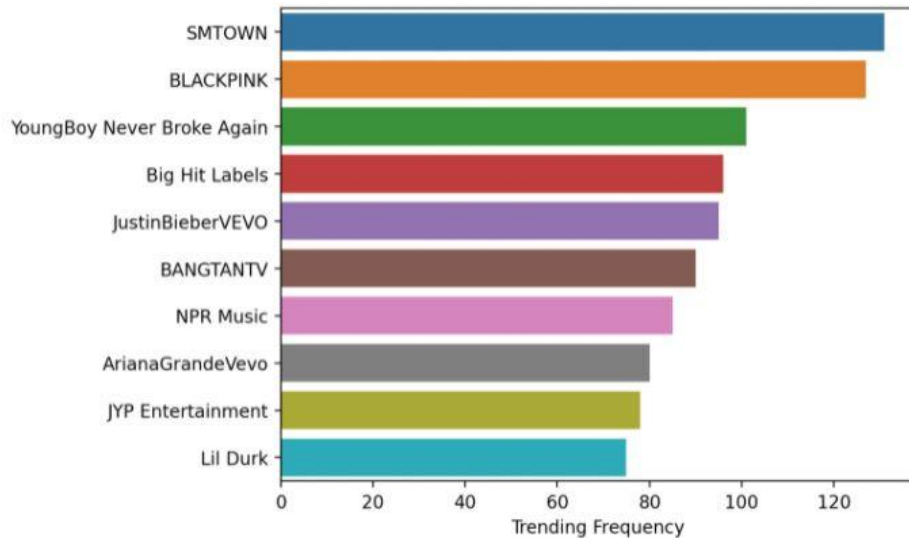
Pick a date range

Start date

2020/08/12

End date

2021/04/27

Top Trending Video Categories

Top Channels By Category

Top Music Channels

Data Insights:

- **Differences in media consumption between countries**

  - **Most popular YouTube channels**

- **Case Study: KPOP**

  - **Every country has a KPOP channel in Top 10 Music Channels, except for India**

Future Work:

- Heroku to deploy Web App and DB

- Utilize YouTube API to request more specific data

- Automation to download updated dataset daily