# Aggregating Youtube Search and Video Data and Metadata

This project is to pull Youtube data, organize and analyze Youtube search data as well as video metadata and comments.

## Modules

- data: Parent Class
- search: Class for search related code
- video: Class for video related code
- analyze: Class for data analysis related code

## Variables

For the code to work, create a variables.yaml file in the root directory. These variables must be present: - api_key
- dir_results
- youtube_api_name
- youtube_version
- save_file_results_separator
- save_file_whitespace_substitute
- query_terms (list)
- number_of_search_results
- number_of_comment_threads
- category (list)
- theme (list)

# Recreating everything

## Extract data from Youtube Data API

Create folder "search_results" if not created already
Create folder "generated_tables" if not created already
Run-

```
extract.py
```

Find the files video_info.csv and comment_threads.csv in the above folder

## Anonymize

Create folder "unique_id_map" if not created already
Run the entire notebook-

```
anonymize.ipynb
```

Find the mappings and the anonymized datasets in the "unique_id_map" folder.

## LDA Analysis

See-

```
analysis_lda.ipynb
```

## Descriptive statistical analysis

See-

```
descriptive_analysis.ipynb
```