YT-Thumbnail selector

Week 1

Data Preparation Stage.

Action Item:

* Collecting thumbnails from youtube
* Extracting data from youtube API
  + Title
  + Views

Process

* Explored 4 different web scraping libraries
  + Beautiful Soup
  + Selenium
  + Scrappy
  + Autoscapper
* Went with selenium because its web driver where you can interact with web pages. It will allow us to fill form.
* Discovered <https://thumbnailphoto.net/> allows to download thumbnail by providing url.
* We Went for web scrapping with selenium for given above website.
* We Were able to interact the website with selenium but ended up with HTML error status code 403 (forbidden) when trying to download thumbnail with urllib library.
* Considered going for alternative plan to download thumbnail.
* We got url that allows us to directly download from youtube by just providing video id.
* Created simple script with urllib to download thumbnail.
* Tested OK for 207 videos.
* Created another script to pull data from youtube API. Tested OK.

Week 2:

Action Item

* Discussion with teammates on
  + Research Paper.
  + Deciding upon number of videos that we need for training the model.
  + Deciding on which categories to be included.