

# S&P 500 - Youtube Trends



{‘Team’: [‘Jen Darrouzet’, ‘Chase Inguva’, ‘Praba Santhanakrishnan’]}

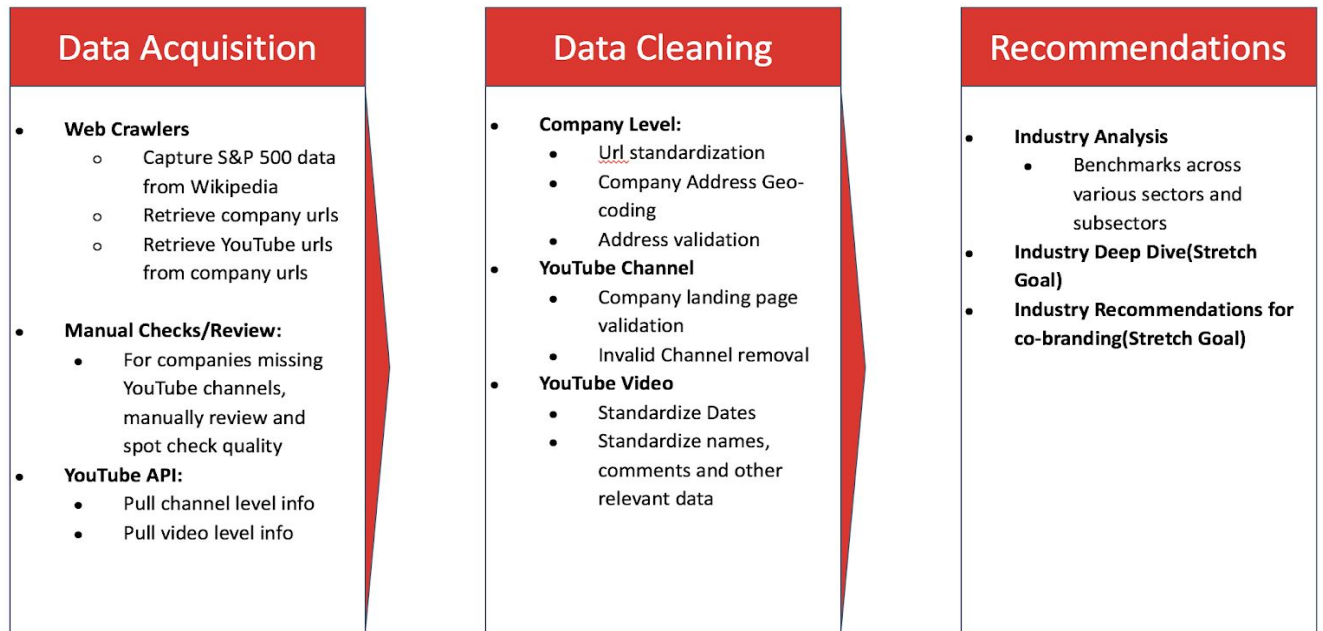
**Project Goal:**

- Understand how S&P 500 companies are using Youtube as marketing and communications channel to engage with customers, investors, and the public
- Few key outcomes of research study:
  - Understand Industry level video publishing on YouTube. Example Question: Does Consumer packaged goods have a stronger (more committed, more methodical, more successful) Youtube strategy vs. Technology companies?
  - Describe video message content at a high level. For instance, how often do Companies release product launch videos on youtube? How often do they post recruiting videos on YouTube?
  - Which sectors/industries should watch out for their competitors?
  - Youtube marketing strategy/Execution tactics:
    - How frequently do companies use Youtube to communicate to their users
    - Is this part of their ongoing marketing tactics/and or actively managed?
    - When do companies stop broadcasting or when do they produce videos?
  - Generate Interesting League Tables:
    - Video Productions
    - Channel Subscribers
    - Video Views
- Stretch Goals:
  - Deep dive into specific industry (example: Hospitality) and identify issues/recommendations
  - Recommend Youtube emerging channels to companies to promote their products

**Data Sets:**

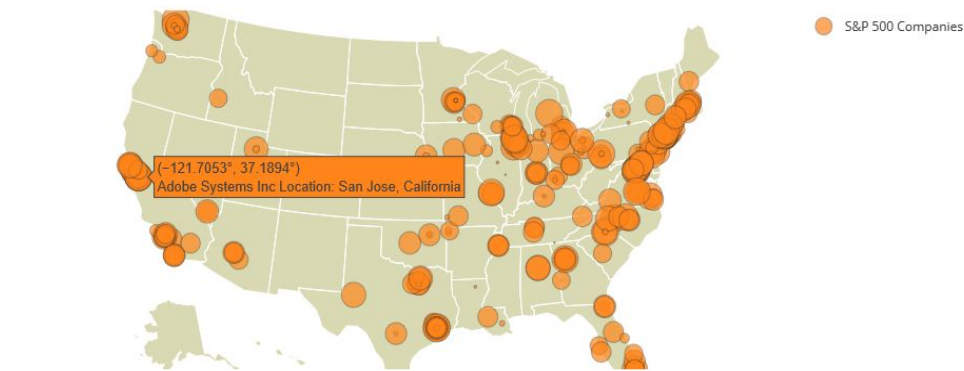
- Unfortunately the data is not readily available and we will be using a custom pipeline the team designed to build the data sets. Our intent is to use S&P 500 companies and retrieve the following information:
  - Company Table:
    - Primary Key: Ticker symbol
    - Company Name
    - Address
    - Industry
    - Sub Industry
    - Primary youtube channel
    - Secondary youtube channels listed
  - Youtube Channel Table:
    - Primary Key: Channel ID of default YouTube channel
    - Channel Name
    - Subscriber Count
    - Views in Total
    - Start Date
    - Number of videos
  - YouTube Video:
    - Primary Key: Video ID
    - Video Name
    - Video Description
    - Publishing Date
    - Video Duration
    - Cumulative View Count
    - Likes vs Dislikes Counts

**Research Methodology:** Our research methodology is outlined below

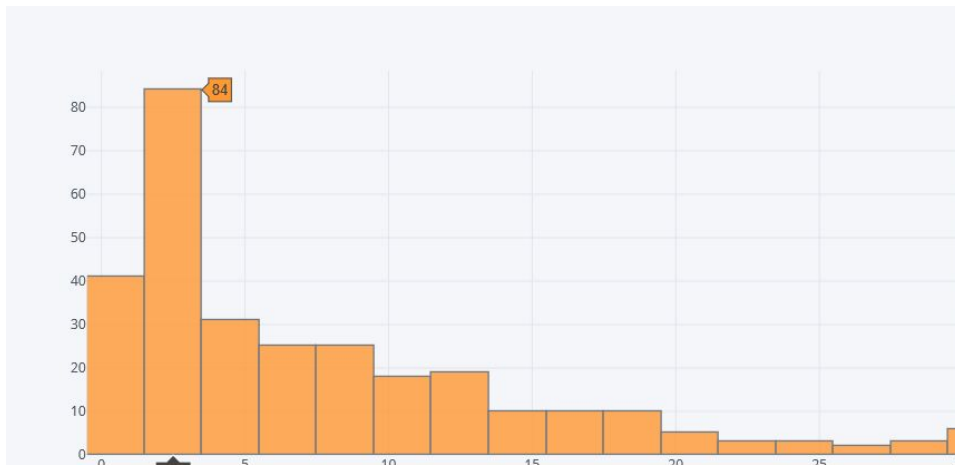


### Visualizations:

- We will also be able to provide the visualizations of the various analysis in the form of charts and interactive graphs (using plotly package)
- Example 1 : Location all the S&P 500 companies.



- Example 2: Histogram on # of channels groups:



### Project Risks:

- API limits
  - We are able to pull all the channels for the companies. We need to get videos for each channel and there is a risk that the full data set could exceed our allotted API usage quota (in which case we may need to use a partial data set). To mitigate this risk, we intend to proceed industry by industry, so that any partial set still has value in and of itself.

### Appendix & links:

- Youtube API:
  - <https://developers.google.com/youtube/v3/getting-started>
  - <https://medium.com/greyatom/youtube-data-in-python-6147160c5833>
- Web scraping using BeautifulSoup
  - <https://www.dataquest.io/blog/web-scraping-beautifulsoup/>

## Initial Data Model:

Companies	(sample)	YT Channels	(sample)	Videos	(sample)
<b>**ticker**</b>	AAPL	<b>**ticker**</b>	AAPL	<b>**ch_id**</b>	UCE_M8A5yxnLfw0KghEeajjw
Co Name	Apple	<b>**ch_id**</b>	UCE_M8A5yxnLfw0KghEeajjw	<b>**video_id**</b>	lprmiOa2zH8
Homepage		Ch Name	Apple	video_title	IPad Homework (Full Version)
Default YT chanr	UCE_M8A5yxnLfw0KghEeajjw	ch_viewcount	456521914	Vid Description	Product launch, etc.
Industry		ch_subcount	6178621	Duration	2min, 20sec
Sub Industry		ch_vidcount	242	video_pub_date	3/27/2018
Location?		ch_commcount	73	Production value	HD
		ch_subs_hidden	FALSE	Vid_viewcount	900000
		<b>id_kind (QC)</b>	youtube#channel	Vid_likes	25000
				Vid_dislikes	2100
		<b>??published at</b>		Vid_comments	Hidden
		<b>??description</b>		<b>id_kind (QC)</b>	