



Regression Project Proposal

Background:

- The sponsoring company is a company that sponsors YouTube celebrities to market the company's products through YouTube channels.
- The value for the company is to predict the greatest growth in subscribers and views to know the best YouTube channel to sponsor.

Data Description:

- Web scraping to extract data of top five YouTube channels pages from [SocialBlade](#) website, and using the [web.Archive](#) to get old data from [SocialBlade](#).
- We have extracted five datasets, most of them from 2016 to 2020.
- Each dataset contains five columns, a minimum of ~1000 rows, and a maximum of ~1100 rows, the sum of all datasets rows is ~5300.
- Columns: date, dailySubscribers, totalSubscribers, dailyVideoViews and totalVideoViews.

Tools:

- Technologies: Python, Jupyter notebook
- Libraries: Numpy, Pandas, Matplotlib, Seaborn, Sklearn, Plotly, Prophet, Requests, BeautifulSoup, Regex