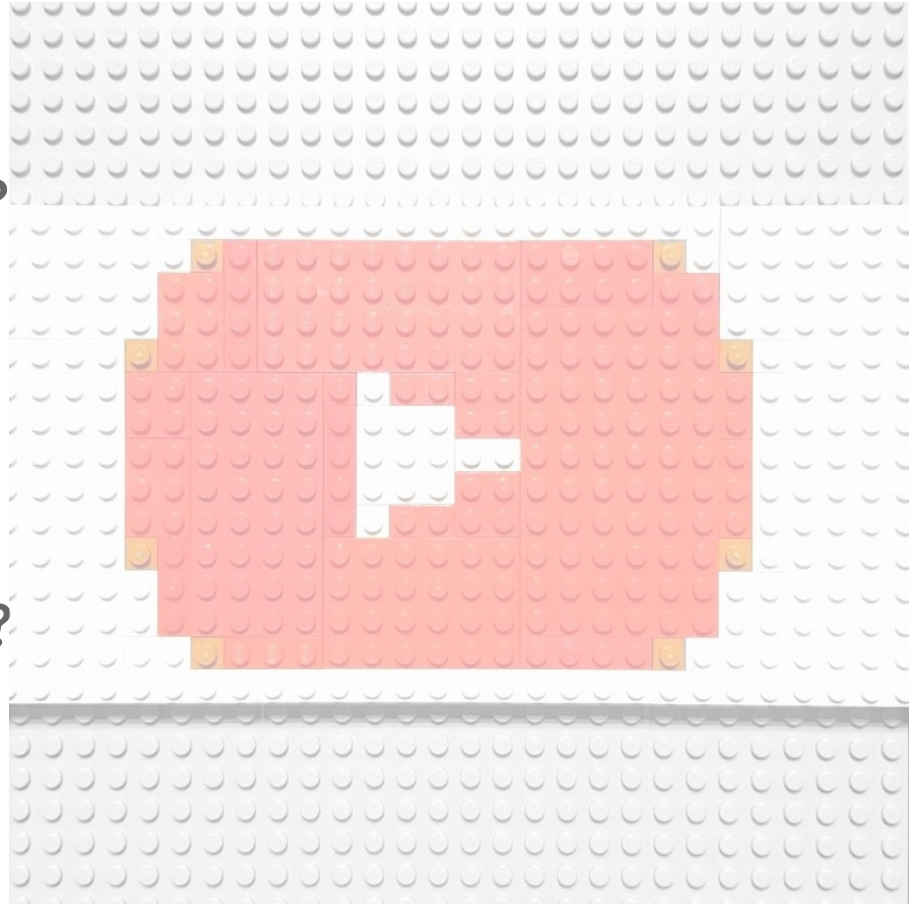
Building the Optimal Youtube Tech Channel

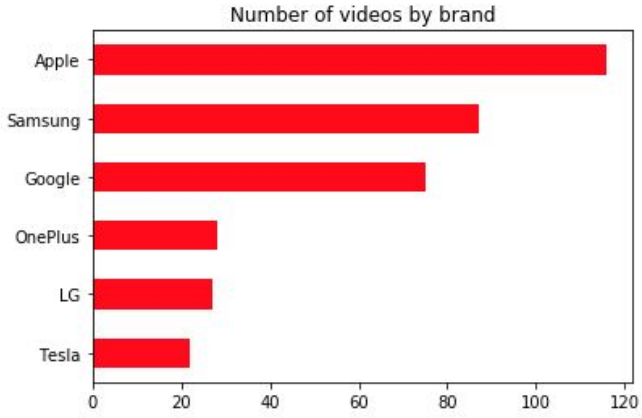
A guide by Jack Tann & Carter Bouley

Youtube Video Building Blocks

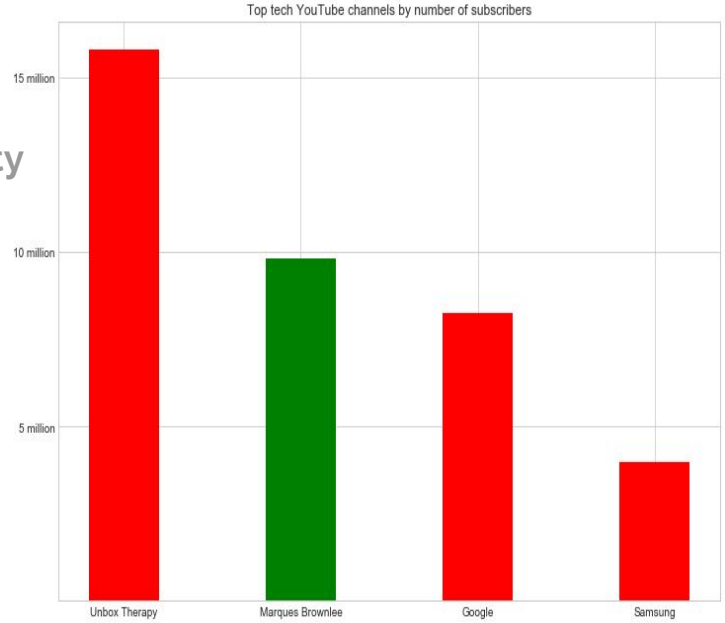
- How long should the video be?
- Which brands should you discuss?
- How many tags is too many tags?
- How often should you post?
- When should you post?
- Should you encourage comments?



Why Marques Brownlee?



- All videos are tech related
- Consistent title formatting
- Sample size of over 1,000
- Upward trajectory in popularity
- Diversity in brand coverage
- Over 1.5bn views
- 9.8M subscribers



Methodology

➔ **Extract**
We navigated the YouTube API system to extract the video parameters we wanted

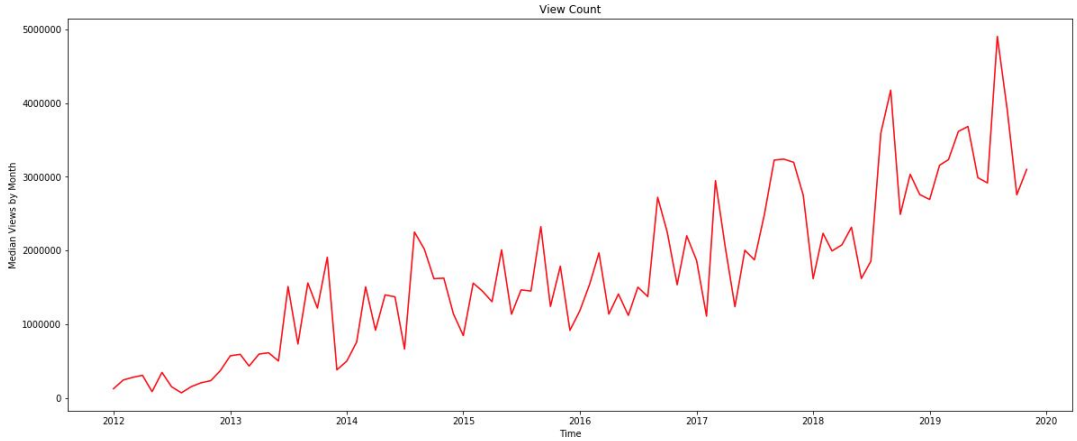
➔ **Transform**
Converted our data into numerical and categorical formats which could be passed into a Machine Learner, for example:

- Extracting brand names from titles
- Extracting seasons from upload date

➔ **Regression Model**
Optimized our model based on features and strength of relationship to predict view count.

➔ **Time Series Analysis**

Predicting future views based on historical data



Baseline Model

Our baseline model included all our predictors

- The R^2 for the train set is 0.3401
- The R^2 for the test set is -0.6796

As expected, our model overfitted with strong bias towards our training set

Top 4 coefficients:

- Duration
- Comment Count
- Tag Count
- Google Topic

Business Recommendations:

- Produces many, shorter videos
- Encourage comments
- Use lots of tags
- Discuss Google products

First Iteration

Our model included all possible combinations of interaction terms

- The R^2 for the train set is -3.235
- The R^2 for the test set is -10.085

Adding interaction terms reduced the effectiveness of the model for the test data

Top coefficients:

- Interactions with comment count
- Interactions with Tesla

Business Recommendations:

- Consider other video attributes along with comment count for best results

Final Model

To deal with the overfitting in our first two models, we will implement a number of feature selection techniques.

These include:

- Variance threshold method
- Recursive feature elimination
- L1 Regularisation

Top 4 coefficients:

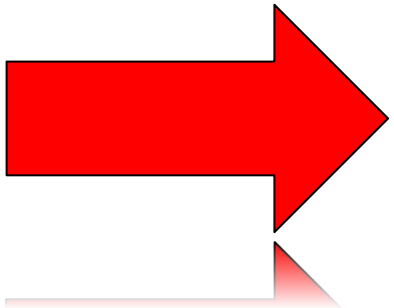
- TBC

Business Recommendations:

- TBC

Limitations

- Video quality is hard to quantify using meta-attributes
- Virality is unpredictable
- Not accounting for popularity growth over time
- Heavily skewed distribution of view count
- Limited numerical variables



Time Series Analysis