

Problem Statement:

I try to define this target: why a video on youtube gets a dislike.

Data Collection:

I found my file's dataset on Kaggle that contains 12 columns format (csv) and I extract the informations using pandas function: video_id, title, channel_id, channel_title, published_at, view_count, likes, dislike, comment_count et tags.

Data Exploration:

Data is cleaned because it doesn't contain duplicated value there is no lack on observations.

Hypothesis:

- Maybe video gets dislike due to few comments or bad title or it doesn't contain a tags.

Modeling:

I realize a model that explain statistics on my jupyter notebook and it prouved that hypothesis is true: "bad title".