

# Stars and Stats

This is a python script for my project of Big Data Science

## Installation of Libraries

You will first have to install the below libraries:

```
pip install google-api-python-client
```

```
pip install pandas
```

```
pip install ipython
```

```
pip install vaderSentiment
```

```
pip install numpy
```

```
pip install scikit-learn
```

```
pip install imbalanced-learn
```

```
pip install lime
```

```
pip install matplotlib
```

Or use below code in your terminal to install all these libraries at once.

```
pip install google-api-python-client pandas ipython vaderSentiment numpy scikit-learn  
imbalanced-learn lime matplotlib
```

## API Key

I am sharing my API key that I have exclusively used for this project. It is a free one. Either you can create a new API key by heading over to <https://console.cloud.google.com> or just use the Key that I have shared. You will have to run the command to Define the API Key. After this, proceed to run the remaining commands.

## Part 1

Run all the commands in this section. This section is for retrieving statistical data to be used in Parts 2 and 3 of this script.

## Do Not Run Multiple Times

There is a section of code in Part 1 that says DO NOT RUN AGAIN AND AGAIN AS LIMIT MIGHT CROSS. Please do not run this cell more than once, as it might cross the daily quota allocated by Google. Each request to the API consumes a certain number of quota

points, and Google imposes a daily limit on the number of quota points you can use. And this part of the code will take a lot of time to run.

## **Part 2**

After Part 1, head over to Part 2 and run all cells consecutively.

## **Part 3**

After Part 2 proceed to Part 3.

## **LIME for one comment index 0**

After Part 3 run this section which is just testing if LIME is working when I give just 1 comment as input.

## **LIME category based**

Run this section. Here I am giving a video ID and categorizing the comments into 5 sections. And then showing the interpretation of one comment from each category.

## **LIME for trend forecasting**

This is continuation of part 3.

## **Results**

This is the final section of my script.

## **NOTE**

The results of LIME analysis will only be visible after you run the code. And the final results might be different from those reported in the project document as we are using real time YouTube data.