Final Poster Project

**Project Proposal**

Sanchit Singhal

Fall 2018 – November 5, 2018

SDS 358 – Applied Regression Analysis w/ Dr. Michael Mahometa

Dept. of Statistics and Data Sciences

**Research Question**

In the US region, after controlling for category type and number of views, does the number of likes and number of dislikes account for a significantly greater proportion of variance in the number of comments on a trending YouTube video?

**Technique Used for Analysis**

Sequential Multiple Regression

* Requires quantitative dependent variable
* Requires quantitative/dichotomous independent variable

**Variables**

Outcome Variable:

* Number of comments – quantitative

Variables of Interest:

* Number of likes – quantitative
* Number of dislikes – quantitative

Nuisance Variables:

* Category type – categorical
* Number of views – quantitative

**Data Source**

Data will be taken from:

<https://www.kaggle.com/datasnaek/youtube-new/data>

There are several datafiles for different regions. I will be analyzing the data from the US region. The codebook for the category type is also provided.