Twitter Airline Sentiment Analysis

**OVERVIEW**

* First classify positive, negative, and neutral tweets, followed by categorizing negative reasons (such as “late flight” or “rude service”).
* Raw number of tweets, segmented by Sentiment.
* Breakdown of negative sentiment, by airline, by reason.
* Geotags of the tweets.

**DATA**

* This Twitter data was scraped from February of 2015 and released to the public in January 2016 by Twitter with 55000 rows.
* Attributes in this Data Set are tweet\_id, airline\_sentiment, airline\_sentiment\_confidence, negativereason, negativereason\_confidence, airline, airline\_sentiment\_gold, name, negativereason\_gold etc…
* 570306133677760000 (5.70306E+17 ) - tweet\_id
* Virgin America - airline
* Bad Flight - negative\_reason
* @VirginAmerica I flew from NYC - text
* to SFO last week and couldn't fully
* sit in my seat due to two large
* Gentleman on either side of me. HELP!
* Etc…

**MODELS TO SOLVE**

* Descriptive Data Mining techniques
* Association for raw number of tweets segmented by sentiment and breakdown of negative reason, by airline, by reason.
* Clustering for Geo tweets.

**Project Team Members and Roles:**

Sai karthik Napa - Data Collection

Akhila Katkam – Testing

Manepalli Mallikarjuna rakesh – Data analysis

Avinash chowdary koganti – API development (Python), documentation