

## **Project 2 – Project Proposal**

### **The team:**

Kevin Li (kl553) Aaron Chao (ac958) Charles Yu (ccy27)

**Crowdsourcing system/ project :** Going through tweets about airborne travel experiences. Scraping the departure and arrival locations as well as the sentiment of the tweet to determine which locations have the most travel issues.

**Method:** We will assemble a list of common/major US Airports to consider in our analysis. This can be as simple as getting the airport codes of the top 50 most trafficked airports in the US from Wikipedia. Then, we will do simple string matches on the tweets against the list of airport codes to determine source and destination. We aggregate positive/negative sentiments for locations with a +1 for positive tweet and -1 with negative tweet and normalize by the total number of location mentions at the end to come up with a score between -1 and 1 that describes how pleasant it is to travel to/from a certain airport.

**Question:** Which airports are the most difficult to travel in or out of (based on Feb 2015 Twitter data)? What are the most common issues for each airport?

**Data collection:** Data provided by Kaggle -  
<https://www.kaggle.com/crowdflower/twitter-airline-sentiment/version/2>

**Presentation:** We will present 1-2 charts showing the results of our data analysis.