



# MEMORANDUM

**To:** Madeleine Udell, ORIE 4741 Staff

**From:** Alan Lee (aml326), Jason Miao (jjm467)

**Subject:** Data Analysis Proposal: Predicting Airline Satisfaction

**Date:** Monday, September 30, 2019

---

## 1 Executive Summary

The aviation industry is intrinsic to the functioning of the global economy. Yet, in recent times airline profit margins have become increasingly slim. With that, more and more airlines are failing. 2019 alone has already seen high profile bankruptcies of key airlines such as Thomas Cook and WOW Air. At the same time, it has become increasingly difficult for new airlines to enter the market. This has major implications on the options for customers as well as the aviation industry as a whole.

In such a time, there is little margin for error within the aviation industry. As such, we seek to determine: *Can we predict airline customer satisfaction? And, if presented with several options for a route, which airline would a customer prefer to take and why?* We plan to do so by performing a data analysis on a data set of all customer reviews from Skytrax, an airline airport review and ranking site. Ultimately, an airline is driven by its customer base. Thus, through this data analysis, we hope to improve the airline industry by learning both what is working and what is not.

## 2 Dataset

The dataset we selected is accessible here:

<https://github.com/quankiquanki/skytrax-reviews-dataset>.

This data set is a web scraping of all customer reviews on the Skytrax site as of August 2nd, 2015. We plan to focus on efforts on the airline reviews which account for 41,396 entries with 20 features. Features include the airline name, review text, overall rating, seat comfort rating, cabin staff rating, food and beverages rating, etc.

Critically, the data set includes a feature for whether or not the reviewer recommends the airline. This is invaluable to our efforts in predicting customer satisfaction. Furthermore, though we have identified massive amounts of missing data entries for features such as aircraft and route, we have noticed that the majority of reviews mention this information in the text body. Since aircraft model types and airport codes are rather standardized, we are hopeful to pull this important information from the customer reviews themselves.