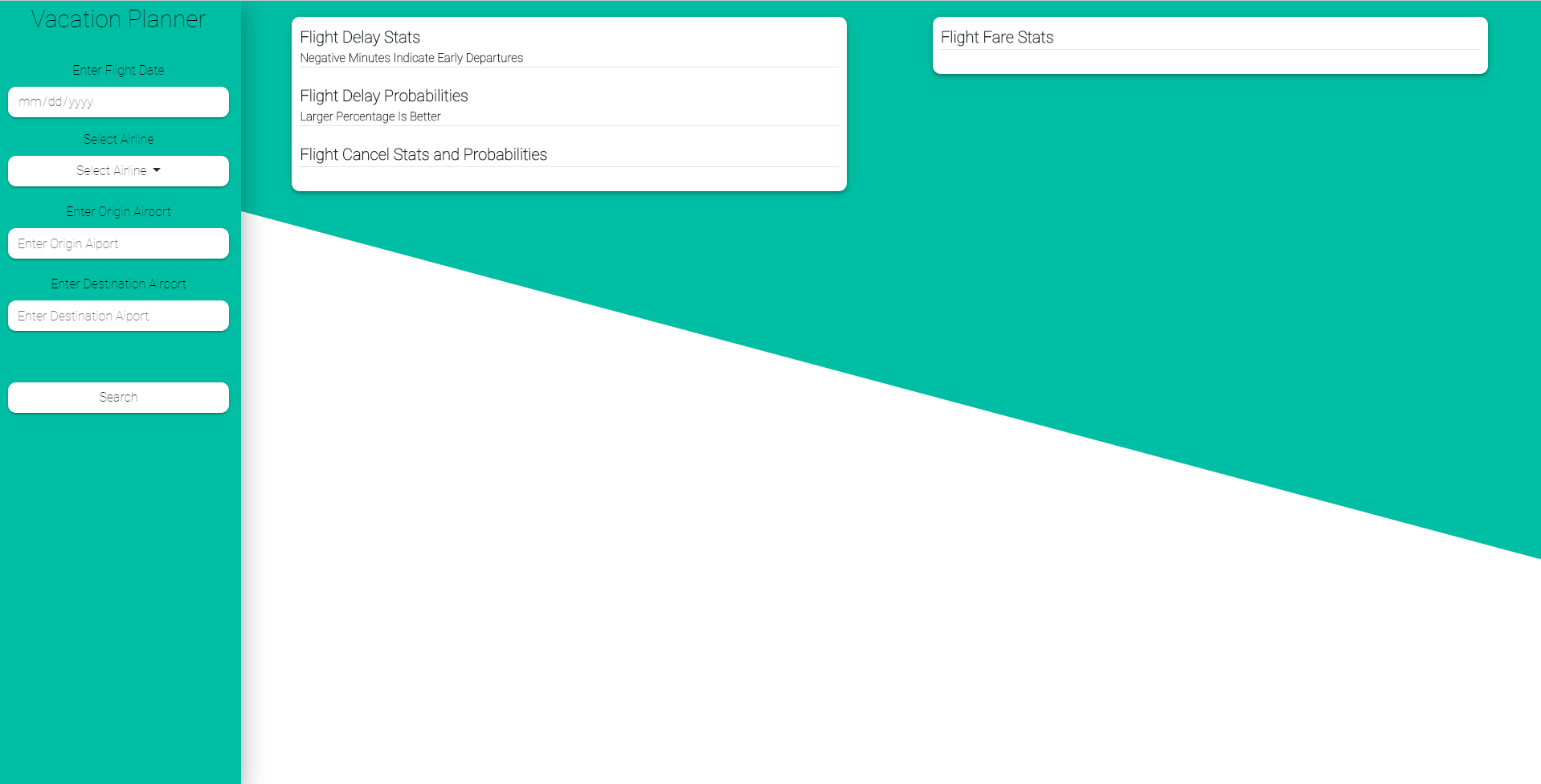
**Vacation Planner**

A Statistical Tool for Assisting in Planning Your Vacation

Cameron Stark

MA-412 Probability & Statistics Final Project

04/21/19

**Introduction**

The purpose of this report and the tool itself is to analyze the rate and probability at which flights get delayed between two different airports and the likely that the flight will get delayed based on previous years of information. The tool itself is built as a website, with Angular JavaScript as the core with an advanced array manipulation library called LoDash, and two math and statistics libraries called Simple Statistics and JStat.

The tool takes in 4 parameters from the user, being the date of the flight, the airline/carrier (which is consisted of all U.S based airlines from 2015 – 2018), the airport of origin and the destination airport. With this the tool outputs the stats for all of the previous flights and also provides probabilities that the flight will get delay on the chosen date, aswell as giving the average air fare price and the probability that it will increase or decrease.

**Data Collection**

The source of the data was off the Bureau of Transportation Statistics website, specifically under the Airlines and Airports data section. For the Delay and Cancel data I combined the data for every airline for the past 4 years, which totaled out at over 700,000 items (in which each item counts as a flight) for each month. The Delay and Cancel has 6 columns of data being the flight date, the ICAO airline code, origin airport code, destination airport code, delay minutes (with positive being the delay minutes and negative being an early departure) and cancel column with 0 being no cancel and 1 being a cancelled flight. The flight fare data is a combination of the average fare of tickets for each quarter for each airport from the years of 2012-2018, as well as a combination of fuel prices and airline expenses to create a model of the airline revenue and profit prediction.

**Data Summary**

Delay and Cancel Data

The data itself appears to be normally or evenly distributed around the mean delay minutes for each month, however there is a slight right skew to the delay side because the average delay for each month is positive and the majority of flights are in fact a delay rather than an on time or even an early departure

**Statistical Analysis**

**Conclusion**