Data analysts: Ryan Cramer, Leah Ikenberry, Lisa Macera, Leslie Mayeux, Krystin McKee

Group 2

Module 20: Deliverable 1

### topic and reasoning

We have decided to examine a data set of flights during the COVID-19 pandemic (January to June 2020). We will be examining the amount of flight delays, and the primary causes of those delays (staffing shortages, mechanical issues, adverse weather, and others). We have all experienced these delays personally or close family and friends have been delayed and forced to book a new flight or forced to deal with an unexpected delay.

### data set

COVID-19 Airline Flight Delays and Cancelations data set from Kaggle.

[link] <https://www.kaggle.com/akulbahl/covid19-airline-flight-delays-and-cancellations?select=jantojun2020.csv>

This data set contains approximately 11 million flights and contains detailed flight delay and cancellation data. The United States Department of Transportation Bureau of Transportation Statistics tracks on time performance of domestic flights operated by large air carriers. This data contains relevant flight information for our analysis; on-time, delayed, canceled and diverted flights.

### questions we will answer

We are attempting to predict the percentage likelihood of a flight being canceled by training a machine learning model with our data set. We are also looking to determine the largest causation of delay or cancelation with in our data set. We are also looking to determine the ideal time and day of the week to schedule a flight to avoid as many delays as possible.