To Fly
or
Not to Fly

MEET THE TEAM

RYAN

JANET

ERIC

NEESHA

The Story:

Delayed or canceled flights can ruin business trips, vacations, family events, and so much more. But what if you could predict whether or not a flight will be canceled before you even arrive at the airport? You could change your plans, find another way to get to your destination or call in on Zoom. The possibilities are endless.

Flight cancellations don't just impact travelers but can mean a loss in profits for airlines. Having an accurate estimation of flight cancellations is critical for airlines in order to increase both customer satisfaction and overall profit.

$28 Billion

FAA/Nextor estimated the annual costs of delays in 2018

Problem Statement:

Our project is about flight delays caused by inclement weather in May 2019. We covered all the airports in the US. The data gathered is limited to just one month. But we hope that it will give us a great direction on how weather plays a part in flight delays. In this project, the goal is to use exploratory analysis and build machine learning models to predict airline departure and arrival delays.

QUESTIONS:

How does weather impact flight cancellations?

Are there certain weather events (snow, rain, wind, etc.) that impact the decision to cancel at a higher rate than other weather events?

Are certain airlines more prone to canceling flights based on weather?

THE DATA:

WHERE DID THE DATA COME FROM?

The primary data for this project comes from a [Kaggle dataset](https://www.kaggle.com/datasets/ioanagheorghiu/historical-flight-and-weather-data) of millions of flight cancellations, the reason for the cancellation, and weather data from May through December of 2019. The data was accessed and downloaded with the use of Kaggle's API. This data was originally sourced from the United States Bureau of Transportation Statistics and the National Oceanic Atmospheric Administration.

2ND DATA SET

The secondary dataset for this project comes from [The Global Airport Database](https://www.partow.net/miscellaneous/airportdatabase/index.html) and includes latitude and longitude data for airports in the US. This will be joined to the primary dataset using PostgreSQL.

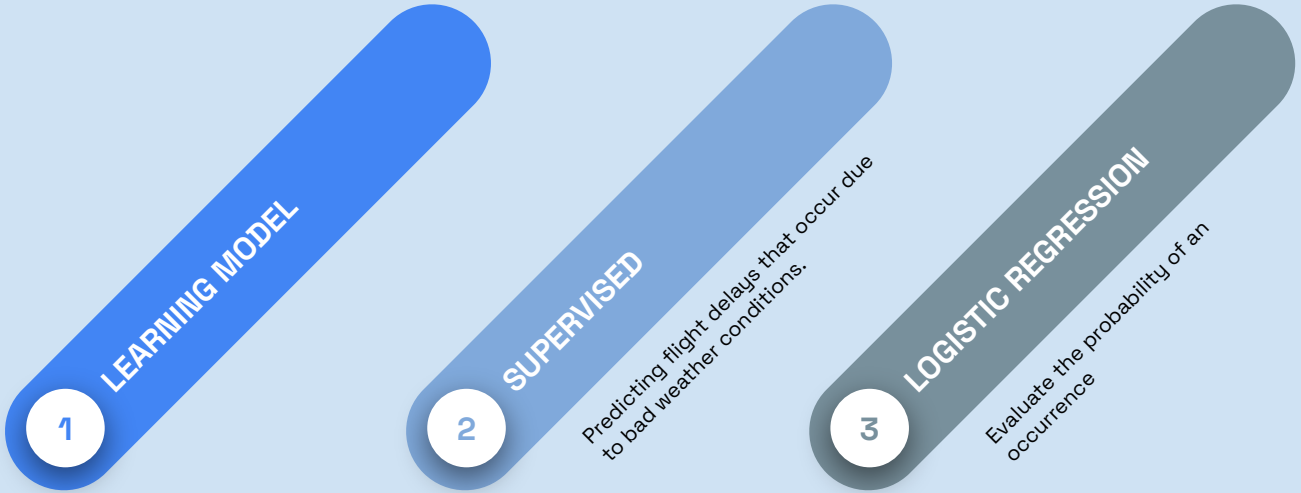
DATA EXPLORATION PROCESS:

The discovery process included researching Weather APIs, historical weather datasets, airport location datasets, and flight delay datasets.

DATA ANALYSIS PROCESS:

We will implement a machine learning model in Python. We will use Scikit-learn, a Python machine learning library. We will use supervised machine learning to perform tasks such as learning from data patterns and making predictions. Supervised Learning will help to predict, based on the airport location and the weather conditions whether a flight will be delayed or not. We will also explore whether certain airlines cancel flights at a higher rate than others.

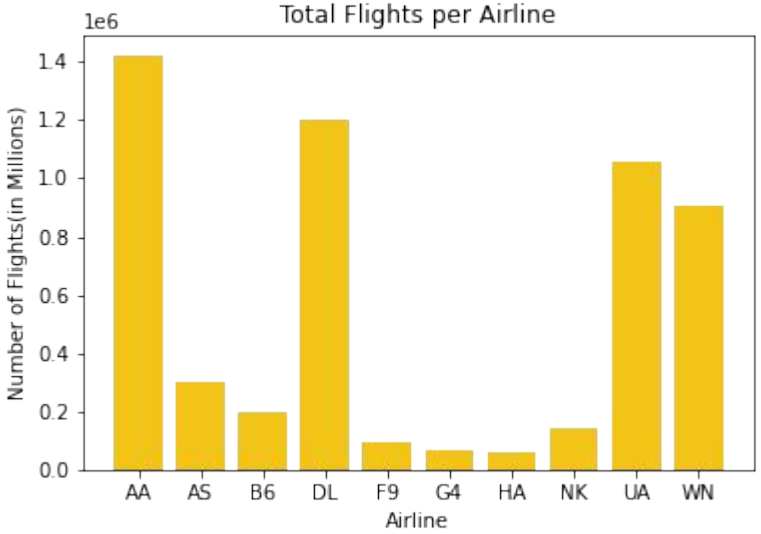
Our process

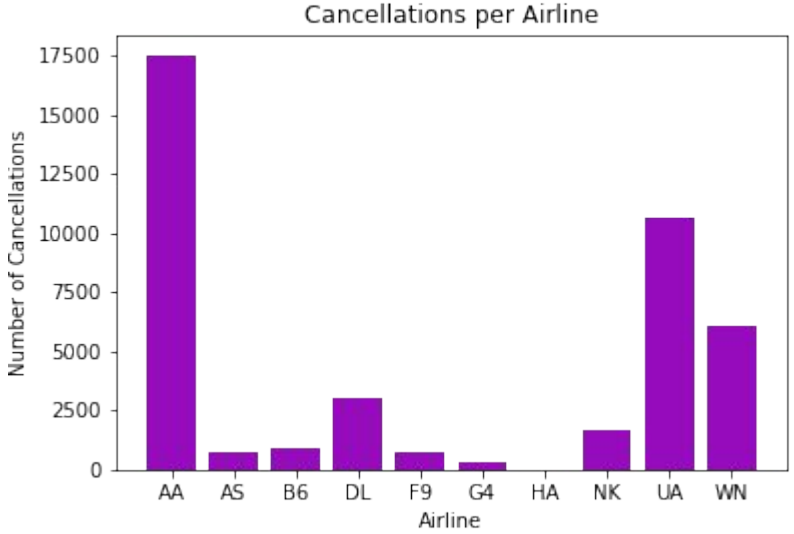

1
LEARNING MODEL
2
SUPERVISED
Predicting flight delays that occur due to bad weather conditions.
3
LOGISTIC REGRESSION
Evaluate the probability of an occurrence

Descriptions of Dashboard

Our Dashboard will be built in Google Slides and our visualizations will made with matplotlib, D3, and Leaflet.

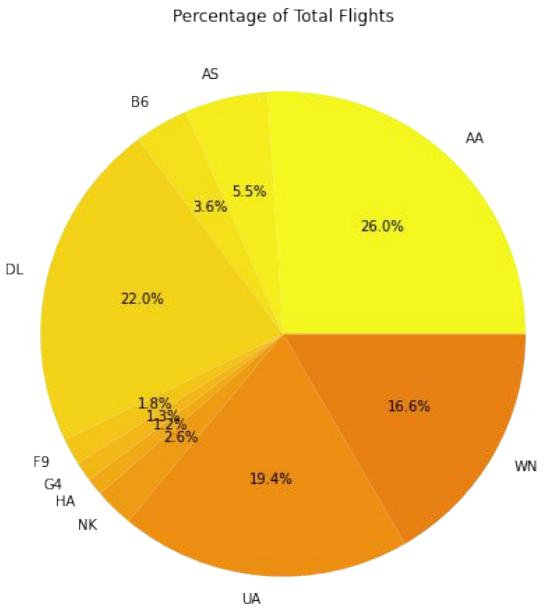
Airline Observations

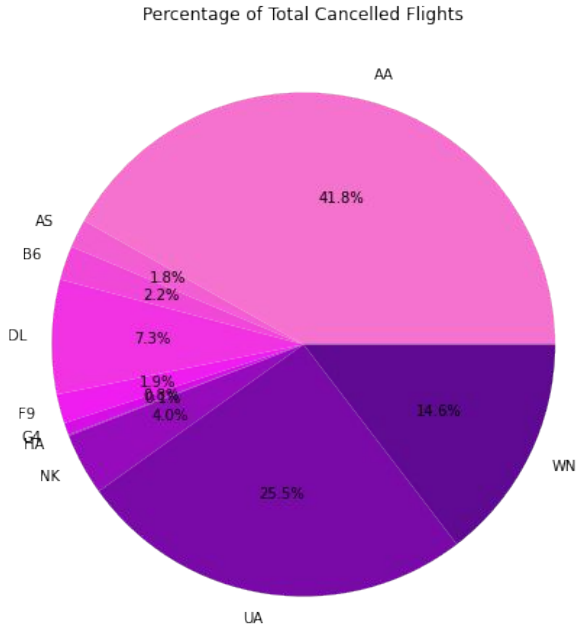




We noticed that although Delta Airlines (DL) is the second highest carrier by volume of total flights, they are fourth in number of cancellations. We wanted to look into this further, so we looked at percentages of total flights vs. percentages of cancellations (next slide).

Airline Observations





American Airlines (AA) accounts for about 26% of total flight volume, but about 42% of cancellations due to weather, whereas Delta (DL) accounts for about 22% of total flights but only about 7% of cancelled flights due to weather.

Further Visualizations

In the next segment we will create:

1. Line chart by month
2. Possibly an interactive map of airports around the country
3. Possibly a chart of cancellations by flight route
4. Possibly any visualizations from interesting findings when we run the ML model