# 2015 Flight Delay and Cancellation Data Analysis

## Context/Abstract/Background:

The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, cancelled, and diverted flights is published in DOT's monthly Air Travel Consumer Report and in this dataset of 2015 flight delays and cancellations.

## Business Scenario:

* Why this data might have collected?
  + As a passenger, aircraft operator, airport authorities the possibility of knowing in advance that there could be a possible delay with flight operation could be very useful.
  + Passenger:
    - They could avoid such flight.
    - They can plan accordingly on schedules if they know that there will be some delay.
  + Aircraft Operator:
    - They would try their best to avoid it, as they have to pay extra charges for the time plane in on ground.
    - If they are known reasons on a particular route where flight get delayed, they can be changed in advance.
  + Airport Authority:
    - They would be prepared for any accommodation of flights in the taxi way or at the docks.
    - They can reduce traffic at the airport both for flights and passengers.
    - This information can be helpful to identify what all airports are with maximum delays and new process could be put in place to reduce such events.
* Research and reading on what this is data can be about.
  + What factors can contribute to flight delays.
    - Air Carrier: The cause of the cancellation or delay was due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.).
    - Extreme Weather: Significant meteorological conditions (actual or forecasted) that, in the judgment of the carrier, delays or prevents the operation of a flight such as tornado, blizzard or hurricane.
    - National Aviation System (NAS): Delays and cancellations attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control.
    - Late-arriving aircraft: A previous flight with same aircraft arrived late, causing the present flight to depart late.
    - Security: Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.
* Possible use-cases that be addressed using that data.
  + Where your analysis can provide value?
    - Revenue Increase
      * If flights operate without any delays then airport can accommodate more passengers and as such make more revenue.
    - Reduce Revenue losses
      * If flight operate on time, they wouldn’t have to pay extra fee’s to airport parking or other miscellaneous fees.
    - Factor like reduce time-taken or efforts.
      * This could reduce waiting time for passengers at airports.

## Business Problem:

* If you have a client or it’s a take away assignment, understand business problem.
* Hypothetically, put yourself as a stakeholder and think of problem that you’re facing and

See if it can be solved by this data. Or you can gain some insight why this problem might be happening?

To understand the possible factors involved in the flight delays and cancellations. Getting some insights on what are the main causes and if there are any patterns in such occurrences.

Finally, provide some data-driven recommendations to minimize such events.

## Data:

There are 3 files containing the data information:

1. Airport Data
   1. Add a brief summary on data
2. Airlines Data
   1. Add a brief summary on data
3. Flights Data
   1. Add a brief summary on data

### Data Dictionary:

1. Airport Data:
   1. IATA\_CODE:
      1. The **location code** 3-digit code for all airports.
      2. There are 322 unique airports are available in the data.
      3. This is a String value.
   2. AIRPORT:
      1. This is the name of the airport.
      2. This a string value
   3. CITY:
      1. This is a name of city.
      2. This is a string value
   4. STATE:
      1. This is a 2-digit state code in US.
      2. This is a string value.
   5. COUNTRY
      1. This is a country code, we have data only for USA.
      2. This is a string value.
   6. LATITUDE:
      1. This contains geographical latitude of location.
      2. This is a float value
   7. LONGITUDE:
      1. This contains geographical longitude of location.
      2. This is a float value
2. Airline Data:
3. Flights Data:

## Analysis:

### Descriptive Analytics:

To understand what has happened in Data or to provide a summary on the data.

### Exploratory Analytics:

To understand why something happened in Data or to understand relation between events and possible cause of it.

You try to answer question what has happened and factors between them?

Question: Use framework of design thinking to collect all possible questions that you might have.

List down around 10 (sufficient number of questions) different question which adds different perspective.

Visualization

* 1 Why the data was collected?
* 2 Who is your stakeholder or to whom you want to present the data?
* 3 What is the size of your data?
* 4 What kind of data they are numerical or categorical data, date-time element?
* 5 You’ve to judge that your plot gives good clarity and representation to your data?

See the 5 points for visualisation and different types of charts that you should use to present your analytics.

Analysis points:

Univariate Analysis

* Solving analytical question and representing them with appropriate charts and plots.

Bi-variate Analysis

* Solving analytical question and representing them with appropriate charts and plots.

### Predictive Analytics:

To Predict what will happen in future by understanding occurrences of the past. We create a model based on our data to predict future events.

### Prescriptive Analytics:

To prescribe or to advise what course of action should be taken based on previous Exploratory and Predictive Analytics. Here try to give recommendation based of our insights from data and help client to make data-driven decision.

List down around 10 (sufficient number of questions) different question which adds different perspective.

1. How many numbers of times a flight has delayed in a year?
2. Which month maximum, minimum delay and cancellation occurred?
3. Which airlines is having maximum delayed flights?
4. For which destination and source delay is maximum?
5. What is average amount of delays?
6. How many cancelled flights are there?
   1. Observation
      1. \* The percentage of flights cancelled are less than 2%.
      2. \* Numbers of flight cancelled is 89884.
7. Which factors are impacting flights and what is the percentage responsible on that factor.
8. What is max, min, std, mean for taxi time airports and in both source and destination?
9. Which geographical regions are more prone to delays.
10. Which state has maximum and minimum delay.
11. Which flight has maximum air time?
12. Which time of the day we see more delays?