U.S Domestic Flight Delays

Project Plan

Team:

Aden Ramirez

Data:

Source: Bureau of Transportation Statistics

About:

The data being analyzed is sourced from the U.S Bureau of Transportation Statistics, part of the Department of Transportation. This data is collected by the Bureau is sourced from each U.S based airline self-reports each month. This includes each flight details that would be seen at any airport, its operation time on the ground, air, and deviation from scheduled times. This also includes some data beyond the flight itself such as the cause of delay and how long each is contributing to the overall total. In the event of a diversion or cancellation that is also observed and coded with reasons, and its destination. There is a lot of terms used that will need to be well defined in the project to assist in the readers understanding but is well documented by the organization.

Organization Details:

Primary Company Details:

Address:

Bureau of Transportation Statistics (BTS)

1200 New Jersey Avenue, SE

Washington, DC 20590

United States

Company Communication:

Website: https://www.bts.gov/

Phone: 800-853-1351

Alt Phone: 202-366-3282

Business Hours:

8:30am-5:00pm ET, M-F

Key Leaders:

Ms. Patricia S. Hu: Director of the Bureau of Transportation Statistics

Dr. Rolf R. Schmitt: Deputy Director of the Bureau of Transportation Statistics

Research Motivation:

The U.S air travel system is a convoluted complex system, this past year we have seen the effects of natural disasters effecting holiday travel. Even after we witnessed a meltdown of Southwest airlines that was completely operational. Airlines have huge amounts of data available to them, and the U.S. Department of Transportation collects huge amounts from them. Data on delays, cancellations and diversions we can look more closely at the fragile system and areas that have the most effect of these studders. With data like this, airlines can focus changes on specific areas, or be able to plan around common weather conditions in areas of the country.

Research Questions:

Research Question 1:

Are flight delays preventable through airline operations?

Research Question 2:

What role do airports play in flight delays?

Research Question 3:

Does airport congestion make the air travel system more prone to delays?

Hypotheses:

Hypothesis 1:

Looking at different categories of flight delays and difference in actual air time vs. actual flight time we can observe the effect of factors that are less directly related to flying the plan, but the business surrounding it. Seeing airline meltdowns outside of a winter storm as this past year, there are more problems than unpredictable weather. I intend to identify, quantify, and predict where these problems can lead to delays in flights.

Hypothesis 2:

When travelling often delays given have a wide range of reasons, many of which are owned by the airport property itself. This can be delays stemming from waiting for an available gate, spacing out the large amount of flights traveling through for safety, or checks and repairs on

the grounds. The airports are the start and endpoints we wait in and play some sort of role when it comes to you missing your arrival by minutes or hours.

Hypothesis 3:

Over the years some of the same flights have gotten scheduled longer even as planes have gotten faster and more aerodynamic. There have also been more and more flights being added to airlines rosters. Are these behaviors reflected from an expectation of delays? Flying from two busy airports such as LAX-JFK there is more time on the ground to maintain safe separation and avoid incident.

Methodology:

For research question one, we are looking at where delays are coming from. I want to use the data to make regression models (ANCOVA or Multiple Linear Regression) that can predict delays based on several factors given in the data and possibly built upon with some classification such as airport traffic flow class or if the flight is flown hub and spoke or point to point.

For research question two, I want to investigate the role airports play in delays, looking a lot at the time on the ground. I want to use some classification model looking at flights where the time on the ground during an airport operation such as taxing, waiting for a gate may occur or if the delay stems from airborne operations or nature.

For research question three I want to look at time on the ground, this information I can classify if the flight was delayed from the ground operation or during the flight. I would like to

use clustering to see similar causes of the delay and see where it is concentrated in the country. This could also show flaws in the two main airline models of hub and spoke or point to point flying.

References:

https://www.transtats.bts.gov/Fields.asp?gnoyr_VQ=FGK

https://www.bts.gov/learn-about-bts-and-our-work/contact-us

https://www.bts.gov/about-BTS