Airlines Delay

Introduction

The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS) tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, canceled and diverted flights appears in DOT's monthly Air Travel Consumer Report, published about 30 days after the month's end, as well as in summary tables posted on this website. BTS began collecting details on the causes of flight delays in June 2003. Summary statistics and raw data are made available to the public at the time the Air Travel Consumer Report is released.

Problem Statement

The goal of this project was to build a Machine Learning Regression Model to analyze, and extract meaningful insights with respect to flight delays

Data Source

Year: 2008

Month: 1-12

DayofMonth: 1-31

DayOfWeek: 1 (Monday) - 7 (Sunday)

DepTime: actual departure time (local, hhmm)

CRSDepTime: scheduled departure time (local, hhmm)

ArrTime: actual arrival time (local, hhmm)

CRSArrTime: scheduled arrival time (local, hhmm)

UniqueCarrier: unique carrier code

FlightNum: flight number

TailNum: plane tail number

ActualElapsedTime: in minutes

CRSElapsedTime: in minutes

AirTime: in minutes

ArrDelay: arrival delay, in minutes

DepDelay: departure delay, in minutes

Origin: origin IATA airport code

Dest: destination IATA airport code

Distance: in miles

TaxiIn: taxi in time, in minutes

TaxiOut: taxi out time in minutes

Cancelled: was the flight cancelled?

CancellationCode: reason for cancellation (A = carrier, B = weather, C = NAS, D = security)

Diverted: 1 = yes, 0 = no

CarrierDelay: in minutes

WeatherDelay: in minutes

NASDelay: in minutes

SecurityDelay: in minutes

LateAircraftDelay: in minutes

Challenge

Using Anaconda Python:

1. Develop Simple Linear Regression Predictive Models to study the relationship between Arrival delay and other predictors.
2. Develop Multiple Linear Regression Predictive Models to study the relationship between Arrival delay and other predictors.
3. Identify the optimal model based on its accuracy and other metrics.