CASE STUDY: FASTEST AIRLINES (PART 1)

1. Objectives

The purpose of this study is to become familiarize with Python to perform basic data analysis. There are many ways to assess the quality of an airline. You will learn how to assess airlines based on historical flight data. In the first part of the case, we will determine which airline is the fastest based on flight time calculations. In the second part, we will develop a predictive model to determine the likelihood of delays using what you've learned from the lecture.

The objectives of the first part of the case study is to:

- Read data using the *csv* package and output results
- Become familiarize with different data structures such as *lists* and *dictionaries*
- Use conditionals and loops to process data, fix anomalies and perform calculations
- Perform basic plotting in Python using the package *matplotlib*

2. How do we determine which airline is the "fastest"?

Consider the following scenario. Airline A says they will get you from LGA to DCA in 45 minutes, but took 60 minutes instead. On the same route, Airline B says it will only take 70 minutes but actually took 65 minutes. Which airline do you consider the fastest?

Most people will say Airline A since the trip took 60 minutes as opposed to 65. However, according to the government, Airline A is "late" since it took longer than 15 minutes to arrive. Because of this, most airlines pad their schedules by saying flight times are longer than they usually are. How do we accurately determine which airline is the fastest? An approach using historical data was proposed in [1]. In this case study, we will reproduce some of the results found in the study to determine the fastest airlines.

The dataset *FlightTime.csv* contains 743 observations of flights from ORD to LAX throughout November 2015. There are 10 variables.

Variable Number	Variable Name	Description
1	Flight Date	Date of the flight (mm/dd/yyyy)
2	Carrier	IATA Carrier Identification
3	Flight Number	Number assigned to the flight
4	Origin	Origin airport
5	Destination	Destination airport
6	Departure Time	Time of departure (hhmm)
7	Depature Delay	Difference between scheduled and actual departure time (minutes)
8	Arrival Time	Time of arrival (hhmm)
9	Arrival Delay	Difference between scheduled and actual arrival time (minutes)
10	Flight Time	Difference between actual departure and arrival time (minutes)

Reference: Bureau of Transportation Statistics

There are four types of time we are interested in:

(1) **Flight Time**: The difference between the departure time and arrival time.
(2) **Average Flight Time**: The average time an airline takes to complete the route.

(3) **Target Flight Time**: An estimate of how long a flight should take based on distance and direction of travel. This is calculated based on the spherical distance (great circle distance). There are many ways to compute this but we will use a simplified formula that does not include the complex variables like windspeed, jetstreams, etc. Let l_{ori} and l_{des} is the longitude of the origin and destination respectively, and d be the spherical distance between two points. Assuming a constant velocity and an average time of 20 minutes to runway time, define $TFT = 0.117 * d + .517 * (l_{ori} - l_{des}) + 20$.
(4) **Typical time**: Calculated by the target time plus the average delay associated with the origin and departure airport.

The measure of an airline's performance is determined by the difference between average flight time and typical time which is called **time added**. The lower the time added, the faster the airline is on that particular route.

Exercise 1: Using *FlightTime.csv*, write a Python program to perform the following.

(1) Read the dataset, ignoring any observations without a recorded departure or arrival time. Some times are recorded incorrectly resulting in incorrect flight time. For any flight time less than 230 minutes, delete the observations.

(2) Calculate the number of observations in your dataset.

(3) Calculate the *target flight time* with $d = 1741.16$ mi, $l_{ori} = -87.90°$ and $l_{des} = -118.41°$.

(4) Calculate the *typical time* of this route. You will need to get the average of the departure and arrival delays and add it to the target flight time.

(5) Calculate the time added **for each airline** and determine which airline have the lowest time added. Note, you will need to calculate the average flight time for each airline.

(6) Output the results of your calculations to a text file.

Exercise 2: Using *matplotlib*, create a bar graph for the time added of each airline. Be sure to label your axes and plot. You may use *PdfPages* from *matplotlib* to output the plot or simply print screen and save it.

Submit the output from Exercise 1, your plot and code to your lab section TA.

References

[1] "How We Found The Fastest Flights." *How We Found The Fastest Flights.* 11 Mar. 2015. Web. 1 Aug. 2015.