**National University of Singapore**
**School of Computing**

**BT2102 Data Management and Visualisation**

**Assignment 3:**
**Airline Delay and Utilization Analysis**

**Total marks: 30**

## BACKGROUND INFORMATION

The airline industry has grown into one of the world's important industries having a direct impact on the world economy. However, the airline industry is also characterized by the presence of strong competition among the carriers. In addition, the industry faces rapid changes due to technology innovations, increased costs due to fuel price hikes and increased regulatory requirements. In such a competitive environment, airlines that have optimum aircraft utilization and on time performance get an advantage on the airline's economic perspective and customer satisfaction.

A recent industry analysis in the USA has indicated that the worst airline delays occur during the summer travel season. Each year, a record number of travelers turn up at the airports for leisure travel during the summer season. It was reported that the airlines in the USA operated around 1 million flights between June and August 2015, according to the U.S. Department of Transportation's Bureau of Transportation Statistics (BTS). As a result of this increased demand, airlines are prone to delays during this season due to various reasons.

For an airline, an aircraft is a very expensive asset. State of the art aircrafts cost several million dollars a month in lease payments alone. As a result, airlines tend to operate with the least number of aircrafts with each aircraft utilized optimally with higher on-air times and lower turnaround times at the airport. Since the airlines operate with the minimum number of aircrafts, a delay in one flight segment can have a propagation effect of delaying several other flights that utilize the same aircraft. In case of a delay in departure, the airlines do not have freedom of flying at will due to increased air traffic and increased fuel costs due to the increased speed of flight.

### The Pilot Study

Due to these reasons, it is intended to conduct a pilot analysis on the operations of aircrafts and airports by two competing airlines in the USA. An analytics team has been appointed to conduct the analysis.

In this pilot study, two airlines are selected. The first airline is Alpha Airlines. It is one of the largest airlines in the USA that has significant efficiency issues. The second airline is Beta Airlines which has been nominated as the top performing U.S. airline by several travel magazines. The names of the airlines are anonymized due to regulatory reasons. A total of 160,695 flight records are available for the two airlines. A flight is considered on time if the flight arrives exactly at the expected time or earlier than the expected time. If the **arrival time exceeds the expected time**, it is considered as a **delayed flight**.

The flights are profiled into four groups:
- Group 1: Alpha Airlines + On-time
- Group 2: Alpha Airlines + Delayed
- Group 3: Beta Airlines + On-time
- Group 4: Beta Airlines + Delayed

The initial estimate on the population, divided into these 4 groups, was:

**Table 1: Delay Profile**

|  | On-time | Delayed |
|---|---|---|
| **Alpha Airlines** | Group 1: 62.46% | Group 2: 37.54% |
| **Beta Airlines** | Group 3: 69.32% | Group 4: 30.68% |

In addition, arrival delays are categorized as follows:
- Minor Delay: Arrival delay less than 15 minutes (delay <15)
- Significant Delay: Arrival delay greater than or equal to 15 minutes and less than 45 minutes(15 <= delay < 45)
- Severe Delay: Arrival delay greater than or equal to 45 minutes (delay >= 45)

A **significant** or **severe** delay needs additional explanations by the flight crew and will be categorized as carrier delay, weather delay, national air system delay, security delay and late aircraft delay.

## PROBLEM DOMAIN INFORMATION
The enclosed file contains the data from a screening of flight details for the month August 2015 obtained from the U.S. Department of Transportation's Bureau of Transportation Statistics (BTS).

These are the fields in the file:

| | |
|---|---|
| **DAY_OF_WEEK** | Day of Week (1-Monday, 2-Tuesday, 3-Wednesday, 4-Thursday, 5-Friday, 6-Saturday, 7-Sunday) |
| **FL_DATE** | Flight Date (yyyy-mm-dd) |
| **AIRLINE** | Unique Carrier Code (AA-Alpha Airlines, BA-Beta Airlines) |
| **TAIL_NUMBER** | Tail Number. The tail number uniquely identifies each aircraft |
| **FLIGHT_NUMBER** | Flight Number |
| **ORIGIN_AIRPORT** | Origin Airport |
| **ORIGIN_STATE_CODE** | Origin Airport, State Code |
| **DESTINATION_AIRPORT** | Destination Airport |
| **DESTINATION_STATE_CODE** | Destination Airport, State Code |
| **DEPARTURE_TIME** | Actual Departure Time (local time: hhmm) |
| **DEPARTURE_DELAY** | Difference in minutes between scheduled and actual departure time. Early departures show negative numbers. |
| **TAXI_OUT** | Taxi Out Time, in Minutes |
| **TAXI_IN** | Taxi In Time, in Minutes |
| **ARRIVAL_TIME** | Actual Arrival Time (local time: hhmm) |
| **ARRIVAL_DELAY** | Difference in minutes between scheduled and actual arrival time. Early arrivals show negative numbers. |
| **CANCELLED** | Cancelled Flight Indicator (1=Yes) |
| **DIVERTED** | Diverted Flight Indicator (1=Yes) |
| **AIR_TIME** | Flight Time, in Minutes |
| **DISTANCE** | Distance between airports (miles) |
| **AIRLINE_DELAY** | Carrier Delay, in Minutes |
| **WEATHER_DELAY** | Weather Delay, in Minutes |
| **NAS_DELAY** | National Air System Delay, in Minutes |
| **SECURITY_DELAY** | Security Delay, in Minutes |
| **LATE_AIRCRAFT_DELAY** | Late Aircraft Delay, in Minutes |

**WHAT YOU ARE REQUIRED TO DO:**

A dataset (**flights.xlsx**) is provided for your analysis.

The analysis is intended to compare and contrast between Alpha Airlines and Beta Airlines. Since Beta Airlines has been identified as the best airline in the USA, it is intended to use the airline as comparison benchmark.

The Analytics team has been charged with creating visualisations that help airline executives better understand the data. Since this programme will be extended to the other American airlines, several hundreds of thousands of more data points are expected to come in as the study expands. Executives would like to see key overall results, at a glance, and be able to dig deeper into them to identify insights on airline utilization and delays.

1. Show the **overall flight profile** based on the available data, where the population is classified into the 4 groups (see Table 1). Verify whether the initial estimate in Table 1 is correct.
2. Show with illustrations how the **average delay time** of all delayed flights vary with the day of the week (Monday, Tuesday, etc.) for Alpha airlines. Compare the results with Beta airlines on the same illustration.
3. It was identified that flights between the following airports have high fluctuations. Compare the air time fluctuations between the two airlines.  Use appropriate illustrations (such as box plot etc.) to **demonstrate the air time fluctuations**.

| Origin | Destination |
|--------|-------------|
| ORD | ATL |
| SLC | DFW |
| JFK | MCO |
| MCO | JFK |
| JFK | MIA |

4. Identify the top 20 **most utilized aircrafts** for Alpha Airlines. An aircraft with highest airtime for the month is considered as most utilized i.e. the aircraft is idling a lesser time on the ground. Make illustrations to compare the results with Beta Airlines. For Alpha Airlines, identify which origin and destination airports are the most common airports where these aircrafts operate.
5. Similarly, identify the top 10 **most underutilized aircrafts** for Alpha Airlines. Make illustrations to compare the results with Beta Airlines.
6. Use suitable illustrations to demonstrate whether the amount of departure delay, amount of taxi out time and departure time of the day has an impact on the arrival delay for Alpha Airlines. Compare the results with Beta Airlines.
7. Use suitable illustrations to demonstrate the **cause of significant and severe delays** for Alpha Airlines. Compare with Beta Airlines.
8. To understand reasons for aircraft delays, what figures would you like to show to the airline executives from the information provided from the data?

**EVALUATION**

1. Submit a PowerPoint file containing the presentation to the 'Assignment 3 Submission' folder in Luminus->Files.

2. **YOU ARE NOT REQUIRED** to make a presentation using the PowerPoint file submitted to explain **the findings through data visualizations** and the **design decisions for the representation and presentation used for the visualizations**.

3. You are free to make reasonable assumptions regarding the data requirements for this initiative. These assumptions should be documented in your submission.

4. Please be reminded that plagiarism is a serious offence. Disciplinary actions will be taken against those caught cheating.