National University of Singapore
School of Computing

BT2102 Data Management and Visualisation

Assignment 2 (Group Project):
Data Visualisation and Storytelling

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 more than 1 million flights between July and August 2019, 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 as a result of increased speed of flight.

DATASET PROVIDED

You are provided with a dataset "**TwoAirlineFlights.xlsx**" on flight data of two airlines for July and August 2019. The two airlines are Acorn Airlines (AA) and Berry Airlines (BA).

Dataset Details

There are a total of 25 columns in the dataset. They are described here:

FIELD	DESCRIPTION
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-Acorn Airlines, BA-Berry Airlines)
TAIL_NUMBER	Tail Number. The tail number uniquely identifies each aircraft
FLIGHT_NUMBER	Flight Number
ORIGIN_AIRPORT	Origin Airport
DESTINATION_AIRPORT	Destination Airport
CRS_DEPARTURE_TIME	Scheduled Departure Time (local time: hhmm)
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
CRS_ ARRIVAL _TIME	Scheduled Arrival Time (local time: hhmm)
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)
CANCELLATION_CODE	A – Airlines/Carriers; B – Weather; C – National Air Systems; D – Security
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 ARE YOU REQUIRED TO DO?

Your team is the data analytics team of Acorn Airlines and is required to make a data-storytelling presentation to the executive management of Acorn Airlines. Among the questions they have in their mind include the followings and they are here to seek answers from your team:

- In what areas is Berry Airlines performing better than Acorn Airlines?
- What are the factors contributing to arrival delays for Acorn Airlines and how do they compare against Berry Airlines?
- What course of actions should Acorn Airlines take to better its performance against Berry Airlines?

You may approach this presentation in this manner:

1. Consider the level-1 questions that need answering from the data. You may add in more level-1 questions that are of interest to your audience.
2. Add in any further sub-level questions for producing the appropriate insights for your story.
3. Conduct exploratory analysis of the data to have an initial understanding of the data.
4. Clean the data where appropriate.
5. Produce data visualisations featuring appropriate *representation* and *presentation* of data**.**
6. Draw insights from the data visualisations.
7. Craft your presentation and produce a **coherent presentation** for your story.

Note that data insights lead to the production of story, but they are not stories themselves. The provided dataset should be sufficient for the questions your audience have; however, if necessary, your group may seek out other ***relevant*** data to supplement the currently available dataset.

DELIVERABLES

1. A PowerPoint (or similar) storytelling presentation.
    a. There is no restriction on the number of slides but note that your group has only 10 minutes for the presentation.

SUBMISSION

1. Zip the PowerPoint file and name it as "**Team_XX_AS2.zip**" (for example, file name for team 9 is "Team_09_AS2.zip" and for team 19 it is "Team_19_AS2.zip"). **PLEASE FOLLOW THIS NAMING CONVENTION ACCORDINGLY.**
2. Upload the zip file to "Assignment 2" folder in Canvas->Assignment 2 by the assignment due date.
3. **Each member of the team must submit a copy of the same team zip file to the canvas assignment folder.**
4. Please be reminded that plagiarism is a serious offence. Disciplinary actions will be taken against those caught cheating.

PROJECT CONSULTATION

- Due to the large class size, Ms Kong Guoting (my graduate tutor) and I will be happy to answer any questions you may have regarding this project. Contact/email Ms Kong (e0546217@u.nus.edu) or me (dannypoo@nus.edu.sg) as a team should your team requires consultation.

EVALUATION

1. Each group is given 10 minutes to present and 5 minutes Questions and Answers during the allocated 15 minutes. Evaluation timeslot will be scheduled nearer week 12.
2. Your group will be evaluated on the data visualisation and storytelling presentation.
3. Each member in the group is required to present, even for a minute.
4. Evaluation criteria:
    a. **Preparation (5 marks):**
        i. Understanding of dataset.
        ii. Data exploration.
        iii. Data cleaning (where necessary).
    b. **Content (9 marks):**
        i. Clear explanation of insights through data visualisations.
        ii. Appropriate use of chart, colours, fonts, labels, annotations, interactivity.
    c. **Story (9 marks):**
        i. A convincing, coherent story crafted from data insights.
    d. **Presentation style (3 marks):**
        i. Professional presentation by presenters.
        ii. Appropriate dressing e.g., no singlet or sleeveless T-shirts.
    e. **Questions and Answers (4 marks):**
        i. Can answer questions well.