

ITECH1103- Big Data and Analytics Group Assignment – Semester 1, 2019

Weight – 20%

Due Date: Analytic Report -Due Week 11 Friday 16:00

Analytic Report: Learning Outcomes Assessed: A3, K3, K6, and S2:

Purpose: The purpose of this task is to provide students with practical experience in working in teams to write a data analytical report to provide useful insights, pattern and trends in the chosen/given dataset. This activity will give students the opportunity to show innovation and creativity in applying SAS Analytics, and designing useful visualization and predictive solutions for various analytics problems.

Project Details:

This is a group assignment and you will complete the task with your team. Your team will be made up of at most 3 members who are all enrolled in the same laboratory – the teams will be allocated by your tutor. It is expected that each team member will contribute equally to the project.

Your team will use an analytical tool (i.e SAS Visual Analytics) to explore, analyze and visualize the dataset provided. You will receive feedback on the draft about presentation choices, content, analysis, and style.

The aim is to use the data set allocated to provide interesting insights, trends and patterns amongst the data. Your intended audience is the CEO and middle management of the Federal Aviation Administration who are responsible for overseeing the airline industry in America.

In addition, each individual team member will write a short reflection as part of the report on their individual experience on working on the project.

Tasks

- **Task 1- Background information** - Write a description of the dataset and project, and its importance for the organization. Discuss the main benefits of using visual analytics to explore big data. In this you should include a justification for using the visualizations that you will use and how they have been successful in other similar projects. This discussion should be suitable for a general audience. Information must come from at least 6 appropriate sources (2 per student) be appropriately referenced. [2 to 3 pages]
- **Task 2 – Reporting / Dashboards** - For your project, perform the relevant data analysis tasks by answering the guided questions provided (see Appendix for questions and dataset) and, identify the visualization you need to develop.

Note: remove any missing data points from your visualizations where possible/suitable

- **Task 3 – Additional Visualizations** - In addition to the guided questions, it is expected that each student will provide at least two other visualizations of the data (i.e. for a group of 3 students this is 6 extra visualizations). These additional visualizations will be judged in terms of quality of the findings and complexity of analysis.
- **Task 4 – Justification** -Justify why these visualizations are chosen in Task 2 and 3. Note: To ensure that you discuss this task properly, you must include visual samples of the reports you produce (i.e. the screenshots of the BI report/dashboard must be presented and explained in the written report; use 'Snipping tool'), and also include any assumptions that you may have made about the analysis in your Task 2 (i.e. the report to the operational team of the company).[1 to 2 pages]
- **Task 5 – Discussion of findings** – using the visualizations created discuss the findings from the data set. In this discussion you should explain what each visualization shows. Then summarize the main findings. [3 to 4 pages]
- **Task 6 – Executive Summary** – summary of the data analysis including a brief introduction, methods used and

a list of the key findings [1 page only]

- **Task 7 - The Reflection (Individual Task)** - each team member is expected to write a brief reflection about this project in terms of challenges, learning and contribution. [1 to 2 pages]

Report Submission:

Each member of the group is to submit an electronic copy of the completed group assignment and their own individual reflection via Moodle.

The report will be approximately 8 to 12 pages in length (not counting cover page and references). The report will include the following in the order provided below:

- A cover page including the names and student id of all team members
- Table of Contents
- Table of Figures / Tables
- Executive Summary
- Background
- The body of the report including reports, insights, justifications and visuals
- Discussion of findings
- Conclusion
- References
- Appendices

All references used in your report must be from peer-reviewed sources. Include any and all sources of information including any person(s) you interviewed for this project.

Your report should be formatted according to the ***“Assignment Layout and Appearance Guidelines”*** and your references presented using the **APA** referencing style; information is available:

<https://federation.edu.au/library/student-resources/help-with-study-skills2/guides-to-your-assessments>

You are reminded to read the “Plagiarism” section of the course description.

A passing grade will be awarded to assignments adequately addressing all assessment criteria. Higher grades require better quality and more effort. For example, a minimum is set on the wider reading required. A student reading vastly more than this minimum will be better prepared to discuss the issues in depth and consequently their report is likely to be of a higher quality. So before submitting, please read through the assessment criteria very carefully.

ITECH1103- Big Data and Analytics - Assignment 1- Data Analysis- Report Marking Guide

Tasks	Marks	Awarded	Comments
1 – Background of the Project: Description of Project, Datasets and organisation. The importance of project for the organisation. Benefits and examples provided for the use of data visualisation using peer reviewed sources.	15		
2 - Dashboard/Reports utilisation of appropriate data analysis tools, visualizations and dashboards developed for the report	30		
3 - Additional Visualizations The quality and complexity of additional visualisations provided other than the guided questions.	5		
4 - Justification - Each analysis/dashboard and report explanation with relevant research papers, complexity and in-depth of the justification.	5		
5 – Discussion of findings - Key data insights, recommendations to achieve organisational objectives with theoretical justifications with proper references.	15		
6- Executive summary - Format, key findings and recommendation	10		
7 – The Reflection (Individual Task)	10		
Presentation of Report - Report is well-written and presented professionally, containing all requirements	10		
Total Marks	100		
Total Marks out of 20	20%		

Appendix 1: Data Set 1 and Guided Questions

Data Set 1

- Teradata – SAS Visual Analytics Data Source – BIRDSTRIKES_DATA

Guided Questions

1. GROUP TASK: Create a data dictionary for the data source for use by the group.
2. How many different species (distinct) have been involved in bird strikes?
3. Where the data has been recorded, what percentage of the time was the pilot warned of birds or wildlife?
4. What are the top five airline/operators involved in bird strikes?
5. What proportion (percentage) of bird strikes involved medium and large wildlife?
6. In which phase of flight would the aircraft be most likely to experience a bird strike?
7. Has a bird strike ever occurred whilst the aircraft was parked? If so, in how many instances?
8. Is a bird strike more likely to happen when there is **no cloud** as compared with **overcast** conditions?
9. In what precipitation conditions is a bird strike most likely to happen?
10. Which species of wildlife had the largest total cost? How much was this?
11. For the most costly species of wildlife, which phase of flight was the most common for a bird strike?
12. For the species **Bald Eagle** and **Canadian Goose**, examine the time of day that bird strikes occur to see if there are any differences between the two species. Hint: you will need to use a filter on the species of wildlife and also use two categories.
13. Investigate using a **crosstab** for the data values **Pilot warned of bird strike**, **Altitude bin**, and **the total cost**.
14. Create a geomap of the bird strike data.
15. Using the sum of the number of records, investigate the bird strike count by airport. Hint: remove any missing data and use a sort.
16. Investigate the bird strike cost by airport. Hint: remove any missing data and use a sort on the cost.
17. Compare the answers of the two questions above. What does this indicate?
18. Perform a cluster analysis on the two variables **time out of service** and **total cost**. Use the cluster matrix graph to describe how the four clusters differ from one another in terms of the two variables used.