UOW MALAYSIA KDU PENANG UNIVERSITY COLLEGE

PART OF THE UNIVERSITY OF WOLLONGONG AUSTRALIA GLOBAL NETWORK

ASSIGNMENT COVER PAGE

Programme	**Course Code and Title**
Bachelor of Information Systems (Hons) Bachelor of Computer Science (Hons) Bachelor of Software Engineering (Hons)	CBD3034N Big Data Analysis
Student's name / student's id	**Lecturer's name**
	AP Dr J. Joshua Thomas

Date issued	**Submission Deadline**	**Indicative Weighting**
18th SEP 2024	25th OCT 2024	30%
Assignment [1] title	Meticulous Analysis of Accident Dashboard	

This assessment assesses the following course learning outcomes

# as in Course Guide	**UOWM KDU Penang University College Learning Outcome**
CLO1	Discuss the value of data science, big data landscape with examples of real world big data problems and approaches.
CLO2	
CLO3	Apply the data science components in the data science lifecycle and associated data flow in solving problems.
CLO4	
# as in Course Guide	**University of Lincoln Learning Outcome**
CLO1	Critically appraise and apply Big Data Analytics concepts, tools and techniques.
CLO2	Apply data science toolkits in a range of applications and solve real-world problems
CLO3	
CLO4	

Student's declaration

I certify that the work submitted for this assignment is my own and research sources are fully acknowledged.

Student's signature: Submission Date:

Dates and Mechanisms for Assessment Submission and Feedback

Mechanism for handout to students	OL LMS
Mechanism for submission of work by student	*Softcopy online submission via OL*
Date by which work, feedback and marks will be returned to students	30th OCT 2023
Mechanism for return of assignment work, feedback and marks to students	Feedback will be provided by a marking template. This will be available to students via OL. The discussions at the walkthroughs will also provide informal feedback

COURSEWORK SUBMISSION GENERAL INFORMATION

Academic Integrity Statement

You must adhere to the university college regulations on academic conduct. Formal inquiry proceedings will be instigated if there is any suspicion of plagiarism or any other form of misconduct in your work. Students must **NOT** collude with other groups of students or plagiarize their work. We practice zero tolerance towards plagiarism, and we use Turnitin to evaluate the similarity index. Your similarity index score must not exceed 20%.

Your tasks must be your own work. Unless the use of Artificial Intelligence (AI) is permitted in your assessment task, using AI to complete your assignment is a form of plagiarism.

Nature of the submission required

A softcopy of your assignment in **PDF version** should be submitted to lecturer, no later than the date and time stipulated on the cover sheet. In addition, an electronic copy of your work must be submitted to Turnitin. The first page of your report, immediately after the cover page, must be a page from Turnitin clearly showing your name and your Originality Score (Please refer to submission arrangement).

Diagrams may be used where they are helpful to support your arguments or description. If they are not your own work, the source must be referenced. Please help us to handle and mark your work efficiently.

Documentation guidelines

Student is required to submit a **SOFTCOPY** of the report and ensure that it use the following formatted styles: 1) Font type: **ARIAL**, 2) Font size: **11 pt**., 3) Line spacing: **Single spacing** and 4) Page layouts: **Justify**. Please make sure you have proper format alignment for all paragraphs, following standard writing style and use **HARVARD CITATION STYLE** for citation. Please include a **HEADER** with the following information: **Student ID, Student name, Course code and Assignment type**. Please also include a proper cover page for your submission which contains information about the students, assignment, course, and department with UOW Malaysia KDU Penang University College on top. Also include page number and list of references, which is shown in the last page.

Penalties for Late Submission

For late submission of this Assignment, a penalty of a reduction by 10% of the maximum mark may be applicable for each Calendar Day or part thereof that the submission is late. An Assignment submitted more than **TEN** Calendar Days after the deadline will have a mark of zero recorded for this Assignment.

Submission arrangement

1. Cover page
2. Turnitin similarity report
3. Table of Content
4. Main Report
5. Reference List or Bibliography List (whichever applicable)
6. Marking Rubric (in landscape orientation)

This assignment is weighted as **30%** of the assessed work for the course. The assignment is an **individual** work.

Students should NOT collude with other students or plagiarize their work. Appropriate action will be taken, according to UOW Malaysia KDU Penang University College regulations, if collusion or plagiarism is suspected. Evidence of academic misconduct will be taken seriously and University College regulations followed. You are advised to be familiar with the University College definitions of plagiarism and collusion. (Refer Handbook) Please do not include references to lecture notes.

You are expected to produce a word-processed answer to this assignment. Please use Arial font with font size of 11.

Background:

In response to the alarming rise of road accidents in our modern world, a pioneering initiative has been launched to comprehensively analyze and understand this pressing issue. This endeavor utilizes advanced data analysis techniques to delve into the statistical landscape of traffic incidents, aiming to uncover crucial patterns and insights. The primary objective of this assignment is to shed light on the factors contributing to the increasing frequency of road accidents. By examining key variables such as accident location, date and time, specific areas, and vehicle types involved, this initiative seeks to provide valuable information to policymakers, transportation authorities, and communities. The ultimate goal is to empower these stakeholders with actionable insights, enabling them to implement effective measures and create safer roads for all.

You will provide some working data; however, if you believe it will improve your analysis, you are free to use supplementary data sets.

Students will need to establish a research question based on the data and results and say a convincing' data tale,' demonstrating the knowledge gained so far during the program. We encourage creativity!

You may want to focus on some specific attributes from the data provided etc. Consider different ways to group data in order to generate the most compelling insights, or you could perform longitudinal analysis.

Example insight: Show a line chart depicting the trend in total accidents over the two-year period.

Dataset Explanation:
You will provide multiple datasets related to the assignment task.
Dataset 1: Accident Analysis.csv
Dataset 2: Day and Time (you can extract from the original data)

Task 1: Large data Dashboard (70%)

Traffic accidents in urban areas not only compromise the quality of life but also place significant strain on city infrastructures. In the era of smart cities, an abundance of data has become available, enabling a more profound analysis of this issue. The process of data fusion draws from diverse information sources, including road accident data, weather conditions, local authority reports, traffic data, and insights from the fire brigade.

Modern IoT systems play a pivotal role in collecting sensor data, which is then processed using ArcGIS kernel density analysis. This approach allows for a comprehensive examination of various factors, including human behavior, environmental conditions, and circumstantial elements that may impact the severity of accidents.

The designed dashboard serves as a powerful tool for visualizing and analyzing these insights. It is intended for use by expert committees responsible for managing urban areas. The Road Accident Dashboard provides a holistic view, offering trend analysis and resource management insights. Its purpose is to empower policymakers, law enforcement agencies, and researchers with the information needed to enhance road safety, plan effective interventions, and ultimately reduce accidents, injuries, and fatalities on city roads.

You will be encouraged to blend dataset and use a variety of visualization formats. Each visualization should have an accompanying insight framed using *Power BI*. You need to prepare and present your work using the software tool **(**Azure and PowerBI) with useful data dashboards based on the questions.

A sample screen capture of useful dash board consists of various storyboard interpreting the datasets is shown in Figure 1.

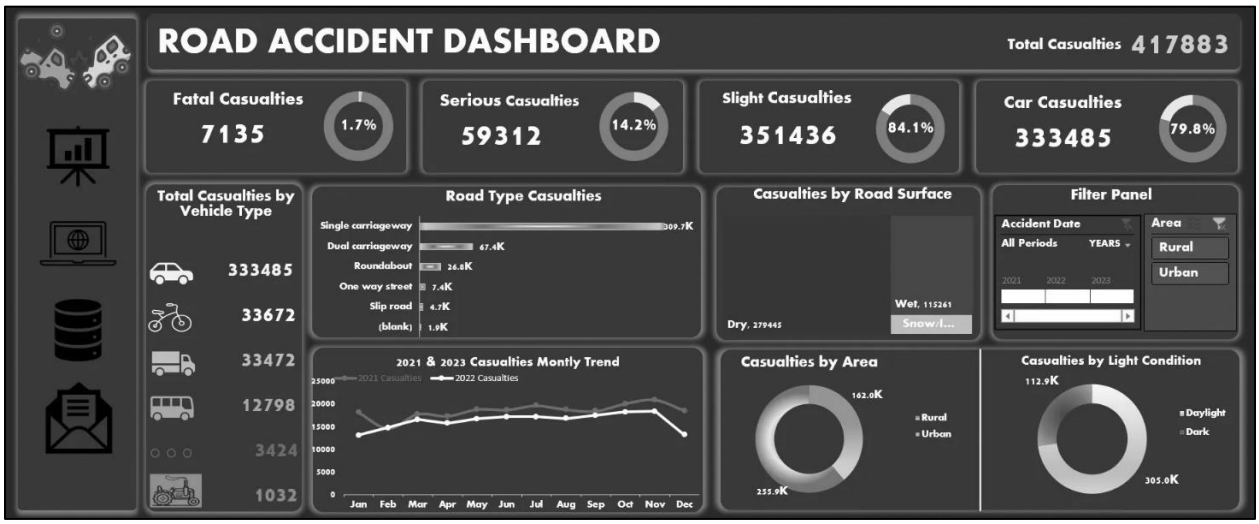

Figure 1. A Sample Road Accident Dashboard

The inspiration of questions for storyboard:
You need to identify the dashboard offers several benefits by answering the questions/stories.

1. Explore a complete picture of road accidents in various cities, aiding evidence-based decisions and targeted safety actions. (20 Marks)

2. Identifies important trends in road accident data, enabling preventative safety measures. (10 Marks)
3. Present the accident hotspots on tree maps which highlighting high-risk zones and accident, death occurrence. (10 Marks)
4. Assists in the best use of resources by examining police actions and accident severity, improving responses. (10 Marks)
5. Calculate and Predict (Hint: Use DAX functions) (20 Marks)
    - Total Causalities from current year and year to growth
    - Total Accidents for current year and year on year growth.

Task 2: Critique Dashboard (30%)
You need to write an A4 page report based on the following criteria about Task 1.

1. Was the data preprocessing for the dashboard suitable?
2. Did your visualization help you to identify interesting data points for further inquiry?
3. Did you pick the most appropriate chat for the data?
4. Did your data insights help you to make a strong argument?

CBD3034: Big Data Analysis MARKING RUBRIC ASSIGNMENT [1] Data Visualization and Dashboard (30%)							
Learning Outcome	**Criteria**	**Fail (0-49)**	**3rd Class (50-59)**	**2nd Lower Class (60-69)**	**2nd Upper Class (70-79)**	**1st Class (80-100)**	**Marks Obtained**
LO1: Discuss the value of data science, big data landscape with examples of real world big data problems and approaches.	Task1: Create a Dashboards Based on 5 questions asked (70%)	You have load the data to Power BI, the data model you have created has no major relationship, not using calculation measures. Calculated. No clear indication of predict results in water quality issues are available.	You have imported the dataset to Power BI. You have addressed questions however the selection of attributes form the dataset is not appropriate. Questions 1,4 are indicating repeating values. However the selection of the chart is not appropriate to extract the good insights of the data. .	You have imported the dataset and explore the repetitive patterns of water quality in the same area with suitable chart. Investigation issues in question 2 is not addressed correctly or missed components. Good representation of question 4 and good selection of attributes. There is a room of improvement in predicting water quality issues. Overall the developmental storyboard are suitable.	You have addressed all the questions with suitable indication of repetitive patterns Selection of chart and colour are appropriate to bring the insights from the data items. The data has tell the data analyst whether to upgrading the water quality with suitable insight result. Good and suitable prediction areas are awarded appropriately.	Importing the dataset has been done after selecting of the columns to Power BI. The multiple storyboard brings the overall and sub divisional insights from the data which is beneficial for the govt or the authorities. Clear representation. Prediction of the issues in certain areas are addressed with suitable insight and interactive data	**/70**
LO2: Apply the data science components in the data science lifecycle and associated data flow in solving problems	Task 2 Critique Data boards Based on 4 criteria for the A4 Size Report. (30%)	You have tried out the visualization no clear use of chart.	You have created correctly and reported with suitable charts. However the insights and the internal connection you have created is not suitable enough.	You have created correctly and reported with screenshots four of the visualizations as suggested in the you have addressed all the critique on storyboards. .	You have created correctly and reported with screenshots all four of the visualizations the prediction results and insights are suitable and well addressed you have defend with extra argument with research articles.	You have created correctly and reported with screenshots all four of the visualizations the prediction results and insights are suitable and well addressed you have defend with extra argument with research articles..	**/30**
					TOTAL (Task 1& 2)		**/100**
					Percentage (Weighted 30%)		**/30**
	Comments:						