**Coursework Assignment 1 Brief**

|  |  |
| --- | --- |
| **Academic year and term:** | 2019 – Semester-2, Year 2 |
| **Module title:** | **Big Data Analytics** |
| **Learning outcomes assessed within this piece of work as agreed at the programme level meeting** | *On successful completion of this module students will be able to*   1. Demonstrate knowledge and understanding of the Big Data technology landscape, terminology and choices available 2. Identify and communicate the challenges of scalability of databases and data analysis 3. Collect and synthesise information needed to select the optimum solution for a problem domain involving big data 4. Analyse and evaluate a range of information comparing alternative methods and techniques used in the analysis of big data including methods to visualise and learn from large volumes of data 5. Implement and test solutions using big data technologies |
| **Type of assessment:** | Group Project Report (2500 words) |
| **Assessed Learning outcomes** | 1, 2, 5 |
| **Weighting** | 60% of the total module marks |
| **Assessment 1 deadline:** | Group project report should be submitted on or before **28/02/2019 by 2pm**This should be submitted via Moodle as a word file which should contain screen shots of the code as well as sample output (see the deliverables required in this assignment ) |

**Tasks / Deliverables of the Assignment 1**

This is a group-work project and marks are awarded for producing a properly documented system that meets the requirements as specified below:

**Scenario:**

In this scenario, you are hired as a data scientist from a reputed marketing firm. Your client is an Automobile research company which is launching a new research project, based on given data sets and fuel economy for 2015 model cars.

The company would like to analyse fuel economy between different car manufacturers.

To achieve this goal, you will use publicly available data sets that have been prepared for you and are available under the appendix section.

The dataset has 729 entries and 19 attributes, which provides the following information on each car:

|  |  |
| --- | --- |
| 1. **Model Year** 2. **Mfr Name** 3. **Division** 4. **Carline** 5. **Engine Displacement** 6. **# Cylinders** 7. **Transmission** 8. **City FE** 9. **Highway FE** 10. **Combined FE** | 1. **Air Aspiration Method** 2. **Transmission Description** 3. **# Gears** 4. **Drive Desc** 5. **Carline Class Desc** 6. **Release Date** 7. **City CO2** 8. **Highway CO2** 9. **Combined CO2** |

**Task 1: [30%]**

#### Review and evaluate the following Big Data technologies and explain how these can be helpful for the Automobile research company to analyse the data according to their requirements. You should state any assumptions clearly.

#### The word count for this task is 1200 words. (7.5x4=30 marks)

1. Predictive Analytics

#### NoSQL Databases

#### Hadoop Ecosystem

#### Blockchain

#### Automobile research

**Task 2: [60%]**

1. You must use **Tableau** to analyse and disclose various associations within the data set provided. Using Tableau, find answers to the following questions based on your analysis. You need to provide a rationale and justification for your solution. Each question must have a separate worksheet with a chart, graph or a table that shows your findings.

The word count for this task is 1300 words. (12 x 5= 60 marks)

**Note:**

*(You will find the car fuel data set excel sheet in the appendix)*

1. Find out which car manufacturer produces the highest quantity of models

e.g BMW 3 Series and BMW 5 Series are different models**.**

1. Find out the highest average fuel economy for city and highway driving from the given data set.
2. Find out high and low average fuel economy from all transmission types.

1. Find out which car manufacturers have 4WD (4 wheel drive) and 2WD (2 wheel drive) models, whose engine power is more than 3.5.
2. Critically evaluate the strengths and weaknesses of data analytics, using Tableau and other recommend tools (up to 3 tools) which are used for data analytics.

**Presentation, Report Layout and References: [10 marks]**

You are required to use an appropriate report layout and formatting style (see the guidelines below), as well as academic citations and a reference list. Your report should be free from grammatical and spelling errors.

**Marking Guidance**

|  |  |  |
| --- | --- | --- |
| **Task** | **Guidance/Criteria** | **Marks** |
| **Task 1** | You have explained in detail how Big Data technologies (e.g**. Predictive analytics, NoSQL Databases, Hadoop Ecosystem, Blockchain**) could be used by the Automobile research **company** to analyse the data according to the requirements. Your explanation is thorough and supported withacademic literature, as well as examples. | 30 |
| **Task 2** | You have correctly loaded the given data sets using Tableau and produced the required output for the problems in this task.  You have provided snippets of all outputs.  You have provided a rationale and justification for your solution, and strategies used to produce the desired output.  You have thoroughly commented on your code.  You have used academic literature to support your arguments. | 60 |
| **Presentation, Report Layout and References** | Your report is well laid out and formatted according to the given requirements. Your report is free from grammatical and spelling errors. Harvard references style has been used to cite the work where necessary and a list of references is also provided. | 10 |
| **Total** | | **100** |

**Preparation Guidelines for Assignment Report**

1. All components of the assignment report must be word processed (hand written text or hand drawn diagrams are not acceptable), font size must be within the range of 12 point to 14 point including the headings, body text and any texts within diagrams.
2. Standard and commonly used fonts such as Times New Roman, Arial or Calibri should be used.
3. Your document must be aligned left or justified with a line spacing of 1.5.
4. All figures, graphs and tables must be numbered and labelled.
5. Material from external sources must be properly refereed and cited within the text using the **Harvard** referencing system.
6. All components of the assignment (text, diagrams. Code etc.) must be submitted in one pdf or word file.

**Appendix:**

Car Fuel Data Set

