**Coursework Assignment 1 Brief**

|  |  |
| --- | --- |
| **Academic year and term:** | 2021-2022 – Semester-2, Year 2 |
| **Module title:** | **Big Data Analytics** |
| **Module code:** | **QAC020N255S** |
| **Module Convener:** | Masum Billah |
| **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 23/03/2022 **by 4pm**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) |
| **Kind reminder:** *it is student’s full responsibility to ensure that all assignments are submitted on the correct link and on time (before 2pm). Failure to do so may result in CAPPED Resit and/or failure of the module.* | |

**Assignment 1**

**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. The suggested word count is 2500 words.

|  |
| --- |
| **How will your work be assessed?**  Your work will be assessed by a subject expert who will use the marking grid provided in this assessment brief. When you access your marked coursework, it is important that you reflect on the feedback so that you can use it to improve future assignments.  **How will feedback be provided?**  Students will have access to formative feedback on each task set in workshops, thereby helping them to refine their approach to the summative tasks that have been set. However, please note that this feedback is limited to recommendations on improving your work. Lecturers will not confirm any grades or marks.  The feedback can be one-to-one or in-group sessions. |

|  |
| --- |
| **Assignment support:**  Although you will be guided throughout the module by your lecturer, you can get extra support for your assignment, just make an appointment with the ACE team for any language, research and study skills issues and/or talk, email the Computing ACE expert for any advice on how to approach your assignment. *REMEMBER: they are not here to give you the answers!*  **Specific requirements for the assignment:** Tableau |

|  |
| --- |
| **Submission instructions:** Word Document via Turnitin.  A digital version of all assignment submissions. These must be submitted via Turnitin on the module’s Moodle site. They must be submitted as a Word file… and must not include scanned in text or text boxes. They must be submitted before 2pm on the given date.    For further general details on coursework preparation refer to the online information  <https://library.roehampton.ac.uk/studytoolkit>  For further details on how to submit your work, please click on the link below.  <https://roehamptonlearning.com/eLearningServices/?p=3654> |

**Word count**

**Don’t exceed the word count**

You need to state the word count at the end of their assignment. 10% over the stated word count is permitted without penalty.

If students go beyond this, then there is a penalty of 5 marks for every additional 10% beyond the word count with a maximum of a 15-mark penalty reduction.

There is no specific penalty for submitting a piece which is below the word count, but please note that shorter submissions are likely to attract poorer grades, particularly where they lack the necessary depth of analysis. Tables and figures (e.g. diagrams, graphs, photographs, etc) may be used as evidence to support academic argument. They are mostly used in report writing. However, it is important that tables and figures are used purposefully (i.e. with good reasons) and when appropriate. They should also be referenced correctly.

**How do you calculate the word count?**

The word count includes the Abstract or Executive Summary and all in-text citations. The word count does not include the Bibliography and Appendices.

*Please note that Appendices should only include supplementary information, not information critical to your work.*

|  |
| --- |
| **Use of pictures/images:**  It is **mandatory** that you add the following statement to your assignment: ‘disclosure: diagrams, pictures or any other images used in this assignment are for educational purposes only’.  Although images, diagrams and photographs etc..can be used in assignments, they MUST be referenced. However, you can access the following websites (copyright-friendly):   * [Pixabay](http://pixabay.com/)   Free photos for commercial and private use - no attribution necessary. Some photos may come from other websites, such as Shutterstock, which may not fall under public domain - check each photo to ensure that it is public domain.   * [Photos Public Domain](http://www.photos-public-domain.com/)   Search public domain clip art and photos by category. All photos on this site were taken by the author and released as public domain.   * [Picdrome](http://picdrome.com/)   The Picdrome gallery is divided by categories (e.g. Abstract, Food & Drink, Nature, Technology, etc.). You may also search for specific images, but the search will take you to a custom Google search of Picdrome, which makes viewing the photos a bit more difficult.   * [Ars Publik](http://www.arspublik.com/)   Browse images under categories such as medical, advertisements, people, sports, space, etc.   * [Public Domain Images](http://www.public-domain-image.com/)   Browse by category or search for images. Varied categories; the site also includes vintage photography and space images.   * [Unsplash](https://unsplash.com/)   "All photos published on Unsplash are licensed under Creative Commons Zero which means you can copy, modify, distribute and use the photos for free, including commercial purposes, without asking permission from or providing attribution to the photographer or Unsplash."   * [Shutterography](https://www.shutterography.com/)   Free images for personal and commercial use.  *(adapted from the University of Pennsylvania:* [*https://guides.library.upenn.edu/c.php?g=475958&p=3255323*](https://guides.library.upenn.edu/c.php?g=475958&p=3255323)*)* |

**Scenario:**

In this scenario, you are hired as a data scientist from a reputable 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%]**

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

#### The suggested word count for this task is between 1200-1500 words. (6x5=30 marks)

1. Data Lakes

#### In-memory Databases

#### Streaming Analytics

#### Edge Computing

#### Artificial Intelligence

#### Apache Spark

**Task 2: [60%]**

1. You must use **Tableau** to analyze and disclose various associations within the data set provided. Using Tableau, find answers to the following questions. You need to provide a rationale and justification for your solution(s).

*Note: each question must have a separate worksheet with a chart, graph or a table that shows your findings.*

The suggested 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 lowest quantity of models.

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

1. Find out the highest average fuel economy for urban and motorway driving from the given data set.
2. Find out the 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 and whose engine power is more than 3.5.
2. Critically evaluate the strengths and weaknesses of data analytics, using Tableau and other recommended 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 Criteria**

|  |  |  |
| --- | --- | --- |
| **Task** | **Guidance/Criteria** | **Marks** |
| **Task 1** | You have explained in detail how Big Data technologies (**e.g. Data Lakes,In-memory Databases,Streaming Analytics,Edge Computing.Artificial intelligence and Apache Spark)** 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(s), 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** |

**Check list**

|  |  |
| --- | --- |
| **TASKS COMPLETED** | Related imagetick |
| 1. Cover sheet completed |  |
| 1. Table of contents |  |
| 1. In-text citations correctly written |  |
| 1. Task 1: [30%] |  |
| 1. Task 2: [60%] |  |
| 1. [10%] Reference list on a separate page and in alphabetical order, completed and in the correct format |  |

**Appendix:**

Car Fuel Data Set



***Assignment Preparation and Guidelines***

* All components of the assignment (text, diagrams. code etc.) must be submitted in ….**one-word file** (hand-written text or hand drawn diagrams are not acceptable), any other accompanied materials such as simulation file, code, etc. should be attached in appendices.
* Standard and commonly used fonts such as Arial or Calibri should be used, font size must be within the range of 10 to 15 points including the headings, body text and any texts within diagrams,
* Spacing should not be less than 1.5
* Pay attention to the Assessment criteria / Marking scheme, the work is to be concise and technical. Try to analyse, compare and evaluate rather than simply describe.
* All figures, screenshots, graphs and tables must be numbered and labelled.
* Tables and figures (e.g. diagrams, graphs, photographs, etc) may be used as evidence to support academic argument. They are mostly used in report writing. However, it is important that tables and figures are used purposefully (i.e. with good reasons) and when appropriate. They should also be referenced correctly.
* Screenshots without description / discussion does not constitute understanding and maybe assumed irrelevant.
* The assignment should be logically structured, the core of the report may start by defining the problem / requirements, followed by the proposed solution including a detailed discussion, analysis and evaluation, leading to implementation and testing stage, finally a conclusion and/or personal reflection on learning.
* Please access your Turnitin Test Page via Dashboard or My modules to learn more about Turnitin and to make a test submission and to check your similarity score before uploading your final version
* You will have opportunity to submit as many times to your module pages as you want up until the deadline.
* Make sure to make backup of your work to avoid distress for loss or damage of your original work, use multiple storage media (memory stick, cloud and personal computer).
* Please note file size limitation might apply. You work must be under 50MB.

***Assignment support:***

* During the delivery of the module, you will have several opportunities to get formative feedback on your assessment during tutorials.
* Although you will be guided throughout the module by your lecturer, you can get extra support for your assignment, just make an appointment with the ACE team for any language, research and study skills issues and/or talk, email the Computing ACE expert for any advice on how to approach your assignment. REMEMBER: they are not here to give you the answers!
* Students will have access to formative feedback on each task set in workshops, thereby helping them to refine their approach to the summative tasks that have been set.
* However, please note that this feedback is limited to recommendations on improving your work. Lecturers will not confirm any grades or marks.
* The feedback can be one-to-one or in-group sessions.

# ***Marking and feedback process***

Between you handing in your work and then receiving your feedback and marks within 20 working days, there are a number of quality assurance processes that we go through to ensure that students receive marks which reflects their work. Please note that the summative feedback and the grades remain provisional until approval from the exam board.

A brief summary is provided below.

* **Step One** – The module and marking team meet to agree standards, expectations and how feedback will be provided.
* **Step Two** – A subject expert will mark your work using the criteria provided in the assessment brief.
* **Step Three** – A moderation meeting takes place where all members of the teaching and marking team will review the marking of others to confirm whether they agree with the mark and feedback
* **Step Four** – Work at Levels 5 and 6 then goes to an external examiner who will review a sample of work to confirm that the marking between different staff is consistent and fair
* **Step Five** – Your mark and feedback is processed by the Office and made available to you.

***Plagiarism and Collusion***

* **Academic Integrity** is a matter that is taken very seriously at the university and student should endeavour to enforce it to all their assignments. In other words, plagiarism, collusion (working and copying from another student) and ghost writing will not be tolerated and will result in sanctions eg: capped resit, suspension and/or withdrawal. Correct referencing demonstrates your academic and professional skill. It also reflects your academic honesty and thus to some degree protects you from cases of plagiarism.
* You must write your assignment in your own words to demonstrate your understanding of the subject.
* Material from external sources must be properly referenced and cited within the text using the Harvard referencing system,
* You are required to follow the**Roehampton Harvard referencing System**. Please refer to Moodle for the latest version of the **Roehampton Harvard referencing System** or ask the library.
* An accompanying list of references (on a **separate page and in alphabetical order**) must also be provided as part of your work.
* It is mandatory that you add the following statement to your assignment: ‘disclosure: diagrams, pictures or any other images used in this assignment are for educational purposes only’.
* Plagiarism: occurs when you present somebody else’s work as your own, whether that work is an idea, graphs, figure, illustration or a pure text. Be it available in web, textbooks, reports or otherwise.
* Wholesale use of text and diagrams from websites is considered as plagiarism when not acknowledged.
* Plagiarism will be dealt with firmly and can lead to serious consequences and disciplinary procedures.
* Collusion: occurs when copying another student’s report (Text, Figures, Illustration etc..) and submitting it as your own.

***Submission and Late submission***

* Students must ensure that their work is satisfactory and fit for purpose, both academically and free from any plagiarism.
* Students must use an appropriate coversheet, which must include the subject, assignment title, student ID and date-time.
* Tutors, lecturers and module convenors do not have the authority to extend the submission deadlines nor the exam time/date. In case of any mitigating circumstances, students should fill in the relevant mitigating circumstances form(s) available at qahe.uormitigatingcircumstances@qa.com
* The marking of the assignment will be capped at 40 if the assignment is submitted within first seven (7) days after the deadline, any submission late than 7 days will be ignored.
* The Submission File should be appropriate to the topic/title of the assignment and contain the Student ID, (Student ID-Assignment title)
* All coursework related material must be attached as an appendix in the final coursework/assignment document, including any computer-generated document, software/ code, simulation file etc..

# ***Resit submission date: TBC***

For students, who are offered a resit, you are required to improve and resubmit your original work. You must resubmit your work using the specific resit Turnitin link on Moodle. Please check the Moodle page and your emails for more information.

You should:

1. Review your previously submitted work and read carefully the feedback given by the marker.
2. Use this feedback to help you revisit and rewrite your work, improving it in the areas identified as weak in the original marking process.

If you did not submit work at the first opportunity you cannot reflect on your feedback. However, you are still required to submit your work. Please note that a non-submission will result in a Qualified-Fail. The original marking criteria will still apply, please refer to the marking criteria.

# ***Mitigating circumstances/what to do if you cannot submit a piece of work or attend your presentation***

The University Mitigating Circumstances Policy can be found on the University website [-](http://www.roehampton.ac.uk/uploadedFiles/Pages_Assets/PDFs_and_Word_Docs/Quality_and_Standards/Mitigat) [Mitigating Circumstances Policy](https://www.roehampton.ac.uk/globalassets/documents/corporate-information/policies/mitigating-circumstances-policy-2019.pdf). Please contact the MITIGATING CIRCUMSTANCES TEAM ON: [QAHE.uormitigatingcircumstances@qa.com](mailto:QAHE.uormitigatingcircumstances@qa.com)