

|  |
| --- |
| **COURSEWORK ASSESSMENT SPECIFICATION** |

|  |  |
| --- | --- |
| Module Title: | **Applied Data Science** |
| Module Number: | **LD7188** |
| Module Tutor Name(s): | Nitsa Herzog |
| Academic Year: | 2022-2023 |
| % Weighting (to the overall module): | 100% |
| Coursework Title: | Assignment |

Dates and Mechanisms for Assessment Submission and Feedback

|  |
| --- |
| Date of Handout to Students:  During week 1 |
| The mechanism for Handout to Students:  Via eLP, discussed during the Lecture |
| Date and Time of Submission by Student: **29 August 2023 16:00**  If you cannot submit for personal reasons by that date and feel you have grounds for requesting an extension, you should contact [nu.london@northumbria.ac.uk](mailto:nu.london@northumbria.ac.uk). |
| The mechanism for Submission of Work by Student:  Electronic submission on Turnitin via Blackboard. Please make sure the file you uploaded is named using this format **ID\_SurnameFirstname.** |
| Date by which Work, Feedback, and Marks will be returned to Students:  Within 20 working days after submission of this assignment |
| The mechanism for the return of assignment work, feedback and marks to students:  Formal feedback will occur following the completion of all reviews and the internal moderation of results. |

**Learning Outcomes Assessed in this assessment**

This assignment will assess the following learning outcomes:

* Demonstrate a critical understanding of data management, manipulation, and modelling methods and techniques applied in an organisational context.
* Critical appraisal of statistical and data science methods, techniques and tools applied for business intelligence.
* Systematically identify and critically analyse data-related problems and develop robust solutions using data management, manipulation and data modelling methods, techniques and tools.
* Critically demonstrate the use and application of statistical and data science techniques and tools in a problem scenario.

**General Information**

This assignment constitutes 100% towards the final mark for this module. Any queries relating to this assignment should be discussed with the module tutor:

**Type of the submission required**

This is an INDIVIDUAL piece of work contributing to the module assessment. Deliverables should be assembled into a single report document, which includes (critical appraisal, research, and snapshot evidence of tasks carried out and justification of technologies used). Submission will be in the form of an MS word report (4000 words).

**Section 1: Statistical Data Analysis**

The aim of the assignment is to apply different statistical methods to the dataset and analyse it through discovering new knowledge and data insight. You are required to analyse a data set “CarType” using SPSS or other statistical software, i.e., R / SAS (The dataset can be found in the Assessment folder). As a result of the analysis, you should answer the following business questions:

What type of car provides the most fuel economy?

What type of car is environmentally friendly?

Individual tasks which are required to be discussed in detail are presented below.

**Task one: Domain Understanding and Research Questions (10%)**

Describe the domain of research to which the dataset belongs. Evaluate the research area by using peer-reviewed academic articles, journals, and books (no websites). Explain what methods are described in the literature and what results are obtained by other researchers who studied a similar problem.

Formulate an appropriate hypothesis (NULL and alternative) from the research questions that can be tested via the different statistical methods.

Provide references used for this section in Harvard style.

**Task two: Dataset and Data Preparation (10%)**

Review and describe the dataset.

Appropriately prepare the data for analysis through data transformation and cleaning techniques. Give detailed justifications of what, why and how you have completed each stage of this process. Explain why you have chosen a particular method over others and what you got at the end of data pre-processing.

Support your explanation with screenshots.

**Task three: Data Analytical Methods (20%)**

Carefully consider which statistical methods are needed to answer your research questions and accept or reject the NULL hypothesis.

Apply appropriately chosen statistical methods which help to analyse your dataset and provide the results for further insight. For example, for descriptive statistics, use frequency, dispersion, central tendency and position measures such as mean, range, variance, standard deviation, percentiles, ranks, etc. You can add methods for correlation and regression analysis.

Present your results clearly and concisely. You might need to provide a correlation matrix to support your ideas. For acceptance or rejection of the NULL hypothesis, you have to calculate the p-value and explain the obtained results.

**Task four: Evaluation and Conclusion (10%)**

You have to make a conclusion about the analyzed dataset and formulate the answers to the research questions.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Section 2: Data Modelling**

Scenario:

Kangaroo is an online delivery company, which is looking for the development an effective Relational Database Management System (RDBMS) and data warehouse to satisfy the needs of the growing business. As a data analyst, you have been asked to develop a required system for Kangaroo.

**Section 1**

Develop the system that satisfies the following business information requirements:

* The database should contain information about Customers, Items, Restaurants, Orders, Drivers and their Vehicle. For payroll, the National Insurance (NI) number of the drivers is recorded.
* For each customer, the database should store the Customer ID, Last Name, First

Name, Email, and the Phone Number.

* For each driver, their name, salary, email address and their manager are recorded

as well as details of their Driving License such as Driving License Number, Issue

Date, country of issue and Expiry Date.

* Each driver is assigned a motorbike when they start with the company and they

normally keep it during the duration of their contract. Details of the motorbike are

registered such as Registration Number, colour, date of purchase, engine size etc.

* Each Manager manages at least one driver, and each driver is associated with one

restaurant only, but one restaurant employs many drivers.

* For each of the Restaurant, the Restaurant ID, Restaurant Name and Address are

recorded.

* For each Item, the Item ID, the Item Name and Item Price should be recorded. The

items are divided into four categories such as Starter, Main Course, Deserts and

Drinks. Prices for each item/Product may vary in different branches. For example,

the Pizza Hut Croydon Branch sells Garlic Bread at £3.50 but the Oxford Street

Branch charges £4.00 for the same item.

* For each order, it is required to store the Order ID, Order date and the Products that

have been ordered. A customer must order at least one item per order.

* A customer can have one or more orders from the same or different restaurants.
* A driver can deliver more than one order, but one driver delivers a particular order

only.

**Section 2**

**Task One:**

Design a relational database using either Chen’s notation or Crow Foot notation, capable of supporting the given business scenario. Your design should include Relationships and any participation constraints. **[15 marks]**

**Task Two:**

Write MySQL code to implement your database design. You should document your code and use constraints, default values, ON DELETE clauses, etc., as appropriate for the business scenario. The use of wizards is prohibited.

Populate all the tables in the database you created with some data (At least 10 records in each table). The data should be meaningful but does not need to be extensive.

Support your work with screenshots. **[20 Marks]**

**Task Three**: Explain how database design will support the business scenario. **[15 Marks]**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Grading Guidance**

**Distinction:**

Excellent in-depth application and critical research on the processes and user requirements. Provide in-depth knowledge of how to design and evaluate a complete database and use appropriate Machine Learning tools to provide valuable insights and knowledge in a dataset.

**Commendation:**

Shows good research on the processes and user requirements. Provides good knowledge of how to design databases and apply Machine Learning algorithms to real-life datasets.

**Pass:**

Provides a basic understanding of the deliverables. Provides end-to-end design, and all requirements are met. The report has some errors and lacks adequate explanation. The robustness and correctness of Database and data analytics are not explained thoroughly. Evidence of design is shown but inadequate explanation using tableau.

**Fail:**

Provides an incomplete attempt or lack of substantial parts of the deliverables. Fails to demonstrate understanding of the concepts required to implement deliverables. Work lacks profound clarity and detail. There are several errors in the report.

**Academic Integrity Statement**:

You must adhere to the university 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. Refer to the University’s Assessment Regulations for Northumbria Awards if you are unclear about the meaning of these terms. The latest copy is available on [Assessment Regulations and Taught Awards](https://northumbria-cdn.azureedge.net/about-us/university-services/student-library-and-academic-services/quality-and-teaching-excellence/assessment/guidance-for-students/) (last accessed on the 17th of August 2022).

**Formative Feedback**

There will be an opportunity for formative feedback during the semester. You are advised to start working on this assignment as early as possible to seek clarification from the module tutor regarding any questions you might have during the semester. Note that tutors will not predict your grade, and you should not take the lack of comment on any aspect of your work as indicating that it is correct. It would help if you made every effort to take advantage of formative feedback, as tutors will not comment on draft work at other times. Remember that you will get more helpful feedback from us by asking specific questions than just presenting us with your documentation and asking, ‘Is this right?’

## Penalties for Exceeding Word Limits:

## The following penalties will be applied after any reductions in a mark due to late submission have been made; penalties will be applied as defined in the University Policy on [Word Limit Policy](https://northumbria-cdn.azureedge.net/-/media/services/academic-registry/documents/qte/assessment/guidance-for-students/pl013-v002-word-limits-policy.pdf?modified=20200803200335) (last accessed on the 17th of August 2022).

The actual word count is to be declared on the front of the assessment submission.

|  |
| --- |
| Graphical user interface, text, application, chat or text message  Description automatically generated |
| Graphical user interface, text, application  Description automatically generated |

**Late Submission Policy:**

For coursework submitted up to 1 working day (24 hours) after the published hand-in deadline without approval, 10% of the total marks available for the assessment (i.e.,100%) shall be deducted from the assessment mark. Penalties will be applied as defined in the University Policy on the [Late Submission](https://northumbria-cdn.azureedge.net/about-us/university-services/student-library-and-academic-services/quality-and-teaching-excellence/assessment/guidance-for-students/) (last accessed on the 17th of August 2022).

**For clarity:** a late piece of work that would have scored 65%, 55%, or 45% had it been handed in on time will be awarded 55%, 45% or 35%, respectively, as 10% of the total available marks will have been deducted.

**Failure to submit:** The University requires all students to submit assessed coursework by the deadline stated in the assessment brief. Where coursework is submitted without approval after the published hand-in deadline, penalties will be applied as defined in the University Policy on the [Guidance for Students](https://northumbria-cdn.azureedge.net/about-us/university-services/student-library-and-academic-services/quality-and-teaching-excellence/assessment/guidance-for-students/) (last accessed on the 17th of August 2022).