**University of Westminster**

School of Computer Science & Engineering

|  |  |
| --- | --- |
| **7BUIS010W Data Warehousing and Business Intelligence – Coursework (2023/4)** | |
| Module leader | Dr Panagiotis Chountas |
| Unit | Group Coursework-CWK2 |
| Weighting: | 50% |
| Qualifying mark | 40% |
| Description | The in-module assessment will consist of a group coursework report (up to 2500 words) requiring students to demonstrate knowledge of stablished concepts and apply them by selecting appropriate solutions to given real-world business data warehouse data for decision making. |
| Learning Outcomes Covered in this Assignment: | This assignment contributes towards the following Learning Outcomes (LOs):  LO4 Design and implement a BI solution that meets the needs of stakeholders and utilises relevant codes of practice and industry standards;  LO5 Evaluate the wider societal and environmental impact of the application of BI Technologies and Tools to solve complex organisational problems;  LO6 Work in a group setting and clearly and concisely communicate design and critically defend the work and design decisions by both written and oral means. |
| Handed Out: | 19TH March 2025 |
| Due Date | 29TH April 2025, Submission by 13:00 |
| Expected deliverables | Submit on Blackboard a single file containing the required documentation (either in docx or pdf format). All implemented codes should be included in your documentation together with the results/analysis. |
| Method of Submission: | Electronic submission on BB via a provided link close to the submission time. |
| Type of Feedback and Due Date: | Feedback will be provided on BB, on 20th May 2025 (appx.15 working days) |
| BCS CRITERIA MEETING IN THIS ASSIGNMENT | * 7.1.1 Critical review of literature * 7.1.2 Development of the self-directed learner * 7.1.3 Respond to opportunities for innovation * 7.1.6 Use appropriate processes * 7.1.7 Investigate and define a problem * 7.1.8 Apply principles of supporting disciplines * 8.1.1 Systematic understanding of knowledge of the domain with depth in particular areas * 8.1.2 Comprehensive understanding of essential principles and practices * 8.2.1 Produce work informed by research at the forefront * 9.1.1 Systematic understanding of knowledge at the forefront in development and implementation * of systems * 9.1.2 Comprehensive understanding of the state of the art techniques * 10.2.1 Critical awareness of current research issues, problems and/or insights |

Assessment regulations

Refer to section 4 of the “How you study” guide for undergraduate students for a clarification of how you are assessed, penalties and late submissions, what constitutes plagiarism etc.

Penalty for Late Submission

If you submit your coursework late but within 24 hours or one working day of the specified deadline, 10 marks will be deducted from the final mark, as a penalty for late submission, except for work which obtains a mark in the range 50 – 59%, in which case the mark will be capped at the pass mark (50%). If you submit your coursework more than 24 hours or more than one working day after the specified deadline you will be given a mark of zero for the work in question unless a claim of Mitigating Circumstances has been submitted and accepted as valid.

It is recognised that on occasion, illness or a personal crisis can mean that you fail to submit a piece of work on time. In such cases you must inform the Campus Office in writing on a mitigating circumstances form, giving the reason for your late or non-submission. You must provide relevant documentary evidence with the form. This information will be reported to the relevant Assessment Board that will decide whether the mark of zero shall stand. For more detailed information regarding University Assessment Regulations, please refer to the following website:<http://www.westminster.ac.uk/study/current-students/resources/academic-regulations>

Group work contract

Before starting the group work assignment, discuss the following areas together as a group. This exercise will enable you to plan your role/s during the project and may help avoid misunderstandings which could negatively affect the process and even your group’s mark.

Group Membership

Formulate groups of three persons (P1-P3). Register your group via Blackboard under the area “Assessment” by 26/03/2025, Failure to do so will result in being allocated randomly to a group the day after.

Tasks per Group Member

P1: Task A

P2: Task B

P3: Task C

Each member is responsible for the reporting of the overall results and achievements for each assigned task.

All members should be involved in the design and implementation of all Tasks A-C to ensure success. All group members will receive the same grade, which is the overall mark for Tasks A-C and the same feedback.

Group Responsibility

* Create a GitHub folder to allow group members to create, store, manage, and share their code and data. Make the folder public so it can be accessed by your tutors. Include the location of the GitHub folder in your group report.
* Submit the group report via Blackboard(BB)by 29/04/2025

Preparation

Ensure everyone is clear about their task and completes them before the group meetings

Absences

Do any group members anticipate being unavailable for several days during the group work period? Members should notify others if they are unable to attend any meetings. Record meetings if certain members are unable to attend.

Communication rules:

1. Be respectful at all times
2. Avoid using slang, acronyms or abbreviation which may not be known to everyone
3. Maintain professional language and attitude in group chats
4. Be mindful of your cultural differences
5. Ensure to engage and involve all group members in any discussion related to the task

Conflict management

E.g. in the event of a conflict between group members, all individuals agree to report this to the tutor

If one member doesn’t follow the above, doesn’t contribute, etc.

Consider what course of action the group will take if this happens. Write in group work log, report to tutor, etc.

Finally, remember to always treat each other with respect and good luck with your group work!

Bank Customer Segmentation

Most banks have a large customer base with different characteristics in terms of age, income, values, lifestyle, and more. Customer segmentation is the process of dividing a customer dataset into specific groups based on shared traits.

*According to a report from Ernst & Young, “A more granular understanding of consumers is no longer a nice-to-have item, but a strategic and competitive imperative for banking providers. Customer understanding should be a living, breathing part of everyday business, with insights underpinning the full range of banking operations.*

About this Dataset

This dataset consists of 1 Million+ transactions by over 800K customers for a bank in South East Asia-India. The data contains information such as - customer age (DOB), location, gender, account balance at the time of the transaction, transaction details, transaction amount, time and date. The dataset can be downloaded from BB under the Assessment content area.

Interesting Analysis Ideas

The dataset can be used for different types of analysis, example -

1. Perform Customer Recency, Frequency, Monetary, (RFM)analysis.
2. Perform Clustering / Segmentation on the dataset and identify popular customer groups based on RFM features;
3. Perform Location-wise analysis to identify regional-social trends in India.

**Guidelines:**

You are required to deliver a group report of a maximum of 2500 words in length describing the methods adopted and the discussion of achieved results with reference to the tasks listed below. Assume that the report is targeted to a marketing strategist, who is interested in learning the business insights inferred in your analysis and receiving suggestions on how to take appropriate actions.

**Tasks**

**Data Understanding**: Useful as a preliminary step to capture the basic data properties

1. Identify and remove null values;
2. Identify and remove invalid transaction amount values;
3. Identify and remove invalid age values;
4. Display the **5 top Locations** where the maximum number of transactions occurred.

**[17 Marks]**

**Perform** **RFM Segmentation**: The second step is to build an RFM model to assign Recency, Frequency and Monetary values to each customer.

1. Write a query to define and calculate the RFM values per Customer;
2. Check the distribution of Recency, Frequency & Monetary Values;
3. Briefly discuss the issue of skewness and remove skew from the data.

**[17 Marks]**

1. **Customer segmentation with k-means:**

The second step is to divide the customer list into tiered groups for each of the three dimensions (R, F and M), using K-means clustering and discuss the profile of each found cluster (in terms of the properties that describe the properties of the customers of each cluster). The report should illustrate the following

* 1. **Brief discussion of the appropriateness of K-means as the** adopted clustering methodology.
  2. It is necessary to discuss the techniques applied to identify the best value of K-number of clusters.
  3. Implementation of K-Means using Python via Google Colab.

**[20 Marks]**

**Data Mart Desing:**

* 1. Based on your findings (Task B), suggest the main dimensions and metrics for designing a data mart for the analysis needs of the marketing department.

[**13 Marks]**

1. **Review Results and CRM as a driver for Sustainability:**
   1. Discuss briefly the business value for marketers of the specific clusters of customers and their behaviour **per Location** – in terms of maximum number of transactions occurred in the **5 top Locations** and cluster descriptive characteristics for RFM values. **[16 Marks]**
   2. Discuss how CRM tools contribute to sustainability for banking, highlighting the practical benefits and offering insights into how these systems can be used to create a more sustainable future for businesses. **[17 Marks]**

Total [100 Marks]

Marking Scheme

Due to the nature of the assessment candidates may come up with more than one equally, good solutions. Thus marks will be allocated as follows

**Tasks**

**Tasks**

**Data Understanding**: Useful as a preliminary step to capture the basic data properties

1. *Identify and remove null values;* ***[3 Marks]***
2. *Identify and remove invalid transaction amount values;* ***[4 Marks]***
3. *Identify and remove invalid age values;* ***[5 Marks]***
4. *Display the* ***5 top Locations*** *where maximum number of transactions occurred.* ***[5 Marks]***

**[17 Marks]**

**Perform** **RFM Segmentation**: The second step is to build an RFM model to assign Recency, Frequency and Monetary values to each customer.

1. *Write a query to define and calculate the RFM values per Customer;* ***[8 Marks]***
2. *Check the distribution of Recency, Frequency & Monetary Values;* ***[4 Marks]***
3. *Briefly discuss the issue of skewness and remove skew from the data.* ***[5 Marks]***

**[17 Marks]**

1. **Customer segmentation with k-means:**

The second step is to divide the customer list into tiered groups for each of the three dimensions (R, F and M), using K-means clustering and discuss the profile of each found cluster (in terms of the properties that describe the properties of the customers of each cluster). The report should illustrate the following

1. ***Brief discussion of the appropriateness of K-means as the*** *adopted clustering methodology.*

***[4 Marks]***

1. *It is necessary to discuss the techniques applied to identify the best value of K-number of clusters.*

***[7 Marks]***

1. *Implementation of K-Means using Python via Google Colab.*

***[9 Marks*]**

**[20 Marks]**

**Data Mart Desing:**

1. Based on your findings (Task B), suggest the main dimensions and metrics for designing a data mart for the analysis needs of the marketing department.

*Identification of business value customer segments/location in Python [7 Marks]*

*Correct Justification of their business value*  *[6 Marks]*

[13 Marks]

1. **Review Results and CRM as a driver for Sustainability:**
   1. Discuss briefly the business value for marketers of the specific clusters of customers and their behaviour **per Location** – in terms of maximum number of transactions occurred in the **5 top Locations** and cluster descriptive characteristics for RFM values.

*Identification of business value customer segments/location in Python [9 Marks]*

*Correct Justification of their business value [7 Marks]*

**[16 Marks]**

* 1. Discuss how CRM tools contribute to sustainability for banking, highlighting the practical benefits and offering insights into how these systems can be used to create a more sustainable future for businesses.

Present your findings for the above questions in your own words. Marks will be allocated for:

* ***Originality of the report***  ***[4 Marks]***
* ***Critical analysis***  ***[4 Marks]***
* ***Technical content [3 Marks]***
* ***Clarity of the paper [3 Marks]***
* ***Use of references***  ***[3 Marks]***

**[17 Marks]**

Total [100 Marks]