# University of Westminster

# Department of Computer Science

|  |  |
| --- | --- |
| **6BUIS017C CRM & CM With Business Intelligence** | |
| Module leader | Panagiotis Chountas |
| Unit | Coursework 1 |
| Weighting: | 50% |
| Qualifying mark | 30% |
| Description | Students will be given a large data set to perform RFM analysis using different techniques (OLAP vs Clustering) in Python. This will involve data modelling and system design to benefit the development of thoughts, supporting arguments and awareness of limitations. The students will be expected to produce a word project report on their analysis of the data set resulting from applying their own implementation of (OLAP vs Clustering) for the purposes of CRM. |
| Learning Outcomes Covered in this Assignment: | This assignment contributes towards the following Learning Outcomes (LOs):   * Apply data mining (DM) methods to derive statistical models for given data and then interpret their results; * Demonstrate competence in using BI Technologies and Tools on business data for the purposes of CRM and CM |
| Handed Out: | 22th October 2024 |
| Due Date | 28th November 2024  Submission by 13:00 hours |
| Expected deliverables | Submit on Blackboard a zip 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 13th December 2024 |
| BCS CRITERIA MEETING IN THIS ASSIGNMENT | * 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 * 8.2.2 Tackling a significant technical problem * 10.1.2 Comprehensive understanding of the scientific techniques |

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 40 – 49%, in which case the mark will be capped at the pass mark (40%). 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>

**Coursework One**

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+ transaction 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 trends in India.

**Guidelines:**

You are required to deliver a report (max 15 pages including all figures) 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 to learn the business insights inferred in your analysis and to receive suggestions on how to take appropriate actions therefore.

**Tasks**

1. **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 maximum number of transactions occurred.

**[10 Marks]**

1. **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.

**[10 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 clustering using K-means 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.

**[10 Marks]**

1. **Review of Results:** 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.

**[10 Marks]**

1. **Data Mart Design:** Based on your findings (Tasks (2,3)) and conclusions (Task 4), suggest the main dimensions and metrics for designing a data mart for the analysis needs of the marketing department.

[10 Marks]

Total [50 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**

1. **Data Understanding**: useful as a preliminary step to capture the basic data properties
   1. Identify and remove null values; **[2 Marks]**
   2. Identify and remove invalid transaction amount values; **[2 Marks]**
   3. Identify and remove invalid age values; **[3 Marks]**
   4. Display the 5 top Locations where maximum number of transactions occurred.

**[3 Marks]**

**[10 Marks]**

1. **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; **[5 Marks]**
   2. Check the distribution of Recency, Frequency & Monetary values; **[3 Marks]**
   3. Briefly discuss the issue of skewness and remove skew from the data. **[2 Marks]**

**[10 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 clustering using K-means 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 Marks]**
   2. it is necessary to briefly discuss the techniques applied to identify the best value of K-number of clusters; **[3 Marks]**
   3. Implementation of K-Means using Python via Google Colab or Jupiter Notebook.

**[5 Marks]**

**[10 Marks]**

1. **Review of Results:** 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.
   1. Identification of business value customer segments/location in Python [5 Marks]
   2. Correct Justification of their business value [5 Marks]

**[10 Marks]**

1. **Data Mart Design:** Based on your findings (Tasks (2,3)) and conclusions (Task 4), suggest the main dimensions and metrics for designing a data mart for the analysis needs of the marketing department.
   1. Identification of Dimensions [3 Marks]
   2. Justification of Selected Dimensions [3 Marks]
   3. Identification of Measures [2 Marks]
   4. Justification of Selected Dimensions [2 Marks]

[10 Marks]

Total [50 Marks]