**All results should be considered as provisional, as they need to be checked by internal moderator and external examiner.**

**BI\_CW2\_2020**

|  |  |
| --- | --- |
| **Overall Mark:** | **63** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Q1** |  |  | **Mark** |
| **Import the dataset to MySQL** | **0-1** | **2-4** | **2** |
| Mistakes, or not showing all steps of the procedure | Full implementation and showing also how you added the various attributes of the dataset in the database. |
| **Perform queries via RMySQL/DBI package and display results** | **0-5** | **6-8** | **8** |
| Mistakes, or up to three queries using related to the packages commands. Show relevant outputs from R console. | Implement all five queries, by showing relevant outputs from R console |
| **Investigate dplyr package for similar style of queries and display results** | **0-4** | **5-10** | **10** |
| Mistakes, or up to two queries using related to the packages commands. Show relevant output from R console. | Implement all five queries, by showing relevant outputs from R console |
| **Provide requested visualisation schematics for this dataset** | **0-2** | **3-8** | **8** |
| Mistakes, or up to the first schematic. Show relevant outputs from R environment. | Implement all schematics (incl. 3rd with your response to the question) |
| **Total** | | | **28** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Q2** |  |  | **Mark** |
| **Generation of sales function** | **0-2** | **3-5** | **0** |
| Mistakes or simple initial attempt without all company’s products | Full implementation of the sales table function, with all products |
| **Revenue Cube Creation** | **0-1** | **2-5** | **1** |
| Mistakes, or just an attempt | Fully implementation |
| **OLAP operations (5 operations x 2 points)** | **0-6** | **7-10** | **7** |
| Mistakes, or implementation of three operations | Implementation of all operations |
| **Total** | | | **8** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Q3** |  |  | | | **Mark** |
| **Import dataset to R from MySQL** | **0-2** | | | | **2** |
| Mistakes, or fully import of the whole dataset (and its attributes) | | | |
| **Data exploration & Preparation:**  ***Statistics (4 marks), Visualisation (6 marks), Training/Testing dataset creation, missing information (2 marks)*** | **0-6** | | | **7-12** | **12** |
| Mistakes, or implementation the statistics and creation of training /testing dataset/missing information | | | Fully implementation, including the visualisation tasks |
| **C5.0 model *(7 marks)* + evaluation *(4 marks)*** | **0-7** | | | **8-11** | **6** |
| Implementation of the C5.0 model using the training set. Model testing via the testing set. Provide all R outputs in your report, as well as any graph to show the tree. Experiment with various parameters of the method, if needed. Check the size of tree. | | | The evaluation of the performance can be made via the tools: confusion matrix (CM), Area under the Curve (AUC) and F1 Score. Provide full detailed analysis of results using all these tools. |
| **CART model *(7 marks)* + evaluation *(4 marks)*** | **0-7** | | | **8-11** | **5** |
| Implementation of the CART model using the training set. Model testing via the testing set. Provide all R outputs in your report, as well as any graph to show the tree. Experiment with various parameters of the method, if needed. Check the size of tree. | | | The evaluation of the performance can be made via the tools: confusion matrix (CM), Area under the Curve (AUC) and F1 Score. Provide full detailed analysis of results using all these tools. |
| **Discussion, comparison of methods based on results** | **0-6** | | | | **0** |
| Compare the two methods, based on their results, discuss the suitability of CM and AUC, F1 as metrics (what info they provide) and suggest how these performances can be improved (mention issues such as feature selection) | | | |
| **Improvement of current models** | **0-4** | | **5-8** | | **2** |
| Implement one of the two requested schemes (either Adaptive Boosting (AB) or pruning) and evaluate its performance via the tools mentioned before. | | Implement the pruning scheme for CART and the AB for C5.0 and evaluate their performances via the tools mentioned before. | |
| **Total** | | | | | **27** |

|  |  |
| --- | --- |
| **Extra Comments** | Q2: no sales function is provided. Thus the remaining code is problematic |