Assignment Specification

|  |  |  |  |
| --- | --- | --- | --- |
| Module: Data 9910 – Working with Data | | | |
| Title: | Lecturer: Jack O’Neill | | Class Group: DT9231/1 |
| Name: Working with Data - Databases | | Worth:  50% | Due Date/Time: 21/12/2019 21:00 |
| Submission Mechanism: Brightspace | | Late Submission Penalty: 3% per day  No submissions after 7 days | |
| Description of Task: **Section A: ER Diagram (25%)**  The consultancy agency, *WeConsult,* is designing a new database backend for an in-house HR and Account Management system which is currently under construction. The company has multiple offices throughout the country, and each office will be responsible for keeping its employee and customer data up-to-date. Head-office is responsible for keeping data on “area manager” employees.  Employee roles are as follows (in order of most to least senior): “area manager”, “office manager”, “senior consultant”, “junior consultant”, “intern”.  Employee details will be stored in the database, including first name, last name, PPS number, job role, salary and the office at which they are based. Employees who are “Area Managers” are not based at any office. No employee may be based at more than one office.  A system of supervision is in place, where each employee above the rank of “intern” supervises up to three employees of a lesser rank.  Customers are identified by a unique customer code. The application will need to display the name and address of each company as well as contact details (name, email, phone number) for the company. Larger customers may have multiple contacts, but each company must have at least one.  Customers are categorized into one of a number of “Industry areas” (e.g. finance, software, legal etc.). Each customer has exactly one account manager. An account manager is an employee of *WeConsult*. The account manager must have as one of his/her areas of expertise the “industry area” of the customer he/she is supervising.  Once a customer has been with *WeConsult* for over a year the account may be manually marked as trusted. This allows employees to begin work on contracts based on a verbal agreement, before a PO has come through. It is important that no customer with less than a year’s history of working with *WeConsult* is ever marked as trusted.  Using the above scenario, complete the following   * Complete ER diagram, using Oracle Data Modeler * ER model to include primary and foreign keys, mandatory attributes, etc * Include all additional indexes and other physical database features in the Physical model * Generate the DDL script and save as a file. * Run the DDL script in your Oracle Schema (show evidence for this) * Test your tables by adding a minimum of 5 records for each. * Provide some SQL queries, and their results, to demonstrate using the tables and their data. | | | |
| **Section B: Data Audit Report Using PL/SQL and SQL Statistical Functions (50%)**  Using the attached customer churn data, design a data model for all the data files that are part of this data set. Generate the DDL script and run this script to create the tables in the database. Load the data set into the table. The deliverables for this part of the assignment are:   * ER Diagram for Customer Churn Data Set. * DDL script used to create the tables in the Database. You can use Oracle Data Modeler for this. * Details of how the data set was loaded into the tables. * Design of the database view for the analytical record, with descriptions of each attribute in the view and any transformations needed. * PL/SQL code used to perform the data audit. * Output from the data audit table formatted to make it readable | | | |
| **Section C: Machine Learning using SQL (25%)**  For this part of the assignment you will create three machine learning models using the in-database machine learning features. Use the analytical record create in Part B for the Telco Customer Churn data set.  NOTE : Decision Trees machine learning algorithm should not be used. If used, no marks will be allocated to its use.  You may need to create additional tables, views, and used sampling to prepare the training and test data sets. You may need to create a view that contains the predicted values for the testing data set.  Complete this process by evaluating the accuracy of the models.  Write a PL/SQL program to combine the accuracy measures from the various models and to present them the user. For example, you can use DBMS\_OUTPUT function to display the results | | | |
| Submission Requirement: A single Zip file containing each of the deliverables outlined above. | | | |
| Demo Requirement: None. However, the code will be executed by the examiner and should run without error. | | | |