**TASK 1: SQL for Data Scientists**

**A relationship with relational databases**

**Background Information:**

Standard Bank is embracing the digital transformation wave and intends to use new and exciting technologies to give their customers a complete set of services from the convenience of their mobile devices.

As Africa’s biggest lender by assets, the bank aims to improve the current process in which potential borrowers apply for a home loan. The current process involves loan officers having to manually process home loan applications. This process takes 2 to 3 days to process, upon which the applicant will receive communication on whether or not they have been granted the loan for the requested amount.

To improve the process, Standard Bank wants to make use of machine learning to assess the creditworthiness of an applicant by implementing a model that will predict if the potential borrower will default on his/her loan or not, and do this such that they receive a response immediately after completing their application.

You will be required to follow the data science lifecycle to fulfil the objective. The data science lifecycle includes the following:

- Business Understanding

- Data Understanding

- Data Preparation

- Modeling

- Evaluation

- Deployment.

You will be required to read about and understand this project lifecycle, which is crucial to the success of any data science project. You will also get to experience how data is identified, collect, stored, and handled when working as a work data scientist using SQL.

In this task, you will be required to complete multiple-choice questions. This multiple choice consists of daily tasks data scientists complete using SQL.

**Here is your task:**

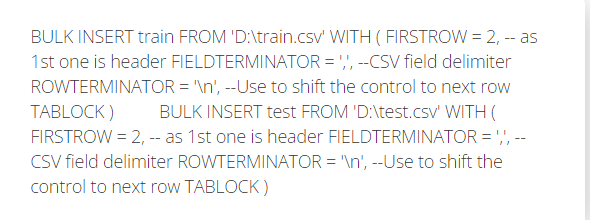
The beginning of any data science project involves understanding the business and the data. Business understanding is focused on setting the business objectives, project assessment, defining the success criteria and project planning. Understanding the business is essential to any data science project. After this, the next phase is Data Understanding. To understand the data, we have to identify, collect, analyze and verify data quality from one or multiple data sources to accomplish the business goals. These tasks can be done in SQL. SQL is a crucial skill to have as a data scientist, as you will be required to work with structured data stored in relational databases.

To complete this task, answer the multiple-choice quiz. This multiple choice consists of daily tasks data scientists complete using SQL. Start the quiz by clicking 'Click here to start the task' in section 5 below. Please note there are 5 multiple choice questions to complete in this task. Please be patient as each question loads.

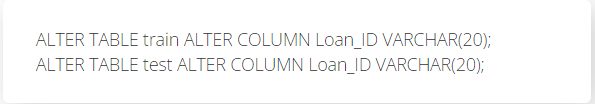
**Q1/5: The home loans department manager provides you with two files (CSV’s) that you are to store securely in a table in SQL. However, the home loans department does not have a database. Which SQL statement creates the database for the home loans department?**

CREATE DATABASE HOMELOANSDB

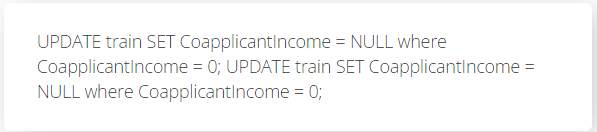
## Q2/5: After creating the database, one of your team members created two empty tables for you to insert data from the respective files. The files have many records. Which SQL statements do this?



## Q3/5: Your Team Lead asks you to change the data type from integer to string of the Loan\_ID column. Which two SQL statements are correct in achieving this?



## Q4/5: The home loans department manager has informed the team that the CoapplicantIncome column is populated with 0 where there isn’t a co-applicant or value is missing. She has asked us to replace the 0’s accordingly. Which SQL statements achieve this?



## Q5/5: The home loans department manager has informed the team that the LoanAmount column is in thousands. In other words, the data capturers insert a loan amount as 30 instead of 30 000. She has asked us to deal with it accordingly so as to not cause confusion. Which SQL statements are correct?

