Noroff University College

Assessment Cover Sheet
2022/2023

Course Code:	Course Title:	Lecturer:
UC2SAT102	Statistical Analysis Tools & Techniques	Isah Lawal
Assignment No: 3	**Total number of pages in this document:** 4	**Maximum Word Count:** NA
Assignment Title: Statistical Data Analysis final course work		
Date Set: 04.05.2023	**Submission Date:** 23:59 Sunday, 14.05.2023	**Feedback Date:** Within 3 weeks afterward

Section 1: Submission

Record of Submission and Plagiarism Declaration

In submitting completed work for assessment, you declare that this submission is your own work and that any collaborative work or reuse of existing material has been explicitly indicated in the text. When submitting completed work, you also agree that it may be submitted for plagiarism testing.

Work should be submitted by the above deadline:

- Electronically Via Moodle in the first case. If problems are encountered, then via email to the course staff.
- Late submissions will result in a ZERO GRADE being awarded for this assessment, unless agreed with course leader in advance.
- Ensure that you read and follow all instructions for this assignment and any additional instructions provided by the course staff. Failure to do so may adversely affect your grade.

IT IS YOUR RESPONSIBILITY TO KEEP A COPY OF ALL SUBMITTED WORK.

Section 2: Important Notice

Note:

- This assignment aimed at assessing your ability to complete a statistical data analysis task in a business setting as an independent aspiring data scientist. Thus, the tasks specified in this assignment requires you to make independent decisions to achieve optimal solutions for the problems presented in the questions. Provide cogent justification for your choices accordingly.

- Generating good and actionable insight needs demonstration of good understanding of all the aspect of the question asked.

- If a modelling method that is not covered in this course is employed your modelling process, then the burden is on you to provide a detail description of its parameters and how the method work.

- Code developed for the analysis is required to be submitted along with the report. But this is not graded. Only the information in the report will be reviewed and graded. Any analysis in the code that is not discussed in the report is not considered.

- Grade will be awarded based on the soundness of your analysis, correctness of your answers, and comprehensiveness of your report.

Deliverables:

- Submit a single pdf file of your work using the Final Assessment submission link provided in the course page in Moodle.
- Submit the code for your analyses using the Artifact/code link on Moodle.

Grading:

The general breakdown of the grade is as follows:

A: Outstanding effort. Student understands and knows how to apply concepts, can put the results into a cogent argument, can identify weaknesses in the argument, and can clearly communicate the results to others.

B: Good effort. Student understands most of the concepts, puts together an adequate argument, identifies some weaknesses of their argument, and communicates most results clearly to others.

C: Passing effort. Student has some understanding of concepts, has some trouble putting results together in a cogent argument, and communication of results is somewhat clear.

D: Struggling effort. Student is making some effort but has some misunderstanding of some of the concepts and is unable to put together a cogent argument. Communication of results is sometimes unclear.

E: Student is not making a sufficient effort. Demonstrates marginal understanding of the concepts and unable to choose appropriate techniques for a given task. The communication of the results is mostly unclear.

Section 3: Statistical Data Analysis (100%)

Introduction:

Customer churn is when a customer leaves a company entirely. Customer churn is a good indicator of the number of clients a company loses over a given period. Customer churn analysis is vital for many companies as it helps them to understand and be able to predict which of the customers is likely to leave because it costs more to acquire new customers than to maintain existing ones. Thus, retaining loyal customers for years makes it much easier to grow and weather financial hardship than spending money to acquire new customers to replace those who have left. Also, successful customer retention increases the customer's average lifetime value, making all future sales more valuable and improving unit margins. In this analysis, you will answer some business questions and perform a predictive analysis of the bank customer churn dataset to help the bank management make informative business decisions.

Data:

The dataset contains several explanatory variables used to assess customer churn, including customer ID, credit score, location, gender, age, tenure, bank balance, credit cards, status (active or not), complaints, satisfaction scores, exited or not, etc. You can find the dataset file and complete description of the variables on the course page. A sample of the original data is provided for you for your analysis. The sample records are divided into two files **Main Sample** and **New Sample**.

Description of Tasks:

Task A: (40%)

- Knowing some statistics about the churn dataset can help to improve the banking business. Thus, analyse the dataset provided in the **Main Sample** file to answer the following business questions.

i. What is the proportion of the customers that are still using the banking services compared to those that have left in the period covered in the dataset? Is there a significant difference in the proportion that the bank authority should be worried about?

ii. What is the relationship between the number of complaints received by the bank authorities and the number of exited customers?

iii. What are the characteristics and statistics (in terms of gender, age groups, and tenure etc,) of the customers that are more likely to complain? Provide an informative profile description of those type of customers.

iv. Is there a significant difference between the credit scores of all the customers that have complained and those who have not in the period covered in the dataset?

v. Do the satisfaction scores on complain resolution provide indication of the customers' likelihood of exiting the bank?

vi. The bank has a reward system where the customers earn points when they use their Diamond, Gold, Silver, and Platinum bank card. Determine if there is a significant difference in the average points earned by the different groups of customers.

Task B: (40%)

- Develop a model to predict whether a customer will complain or not given the historical customer records in the **Main Sample** file. Evaluate the developed model using appropriate metrics and report its performances accordingly. Remember to also discuss the implications of the performance (error or accuracy) of the model with respect to the banking business.

Task C: (10%)

- Once your model is finalised, use it to predict whether the bank customers included in **New Sample** file, will complain about the banking services based on their profile information in the dataset. Tabulate the predicated results in your report, in the order the records are arranged in the file. This will be compared with the expected outcomes for assessing the goodness of your model.

Task D: (10%)

- Report your data analysis and modelling process. The aim of the report is to demonstrate that you are proficient at interpreting and presenting results of your statistical data analysis professionally. Emphasis should be placed on clarity and comprehensiveness. For instance, loads of python code snapshots would, in general, not be acceptable. Properly labelled charts and well-organized tables are good things to have in the report. All relevant sections of a scientific report as we have learned in the course are recommended for this report.

Section 3: Marking and Assessment	
This assignment will be marked out of:	**100%**
Contribution to the final course mark:	**60%**
Estimated number of hours to be spent on this assignment:	**25hr**