**MAKERERE UNIVERSITY**



**COLLEGE OF COMPUTING AND INFORMATION SCIENCES.**

**SCHOOL OF COMPUTING AND INFORMATICS TECHNOLOGY.**

**DEPARTMENT OF NETWORKS.**

**BACHELOR OF SCIENCE IN SOFTWARE ENGINEERING.**

**COURSE UNIT: BSSE YEAR 2 RECESS TERM.**

**GROUP N:**

|  |  |  |
| --- | --- | --- |
| **NAME.** | **REGISTRATION NUMBER.** | **STUDENT NUMBER.** |
| Mukoza Duncan Mwesigwa | 17/U/6502/PS | 217001851 |
| Makwasi Chrispus Arnold | 17/U/5910/PS | 217011259 |
| Ssenono Fransis Xavier | 17/U/10247/PS | 217005992 |
| Lyazi Marvin | 17/U/5850/EVE | 217011257 |

**Sales Prediction Systems Requirements Specification Document.**

**Document Description.**

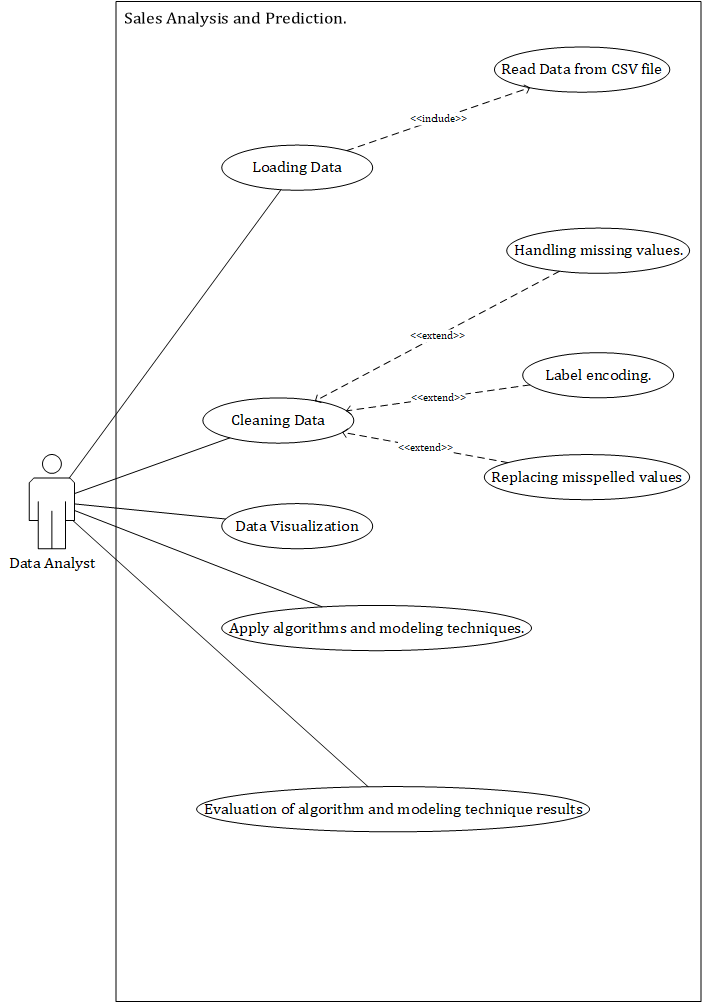
This document contains a detailed use case (and its description) of the sales analysis and prediction data pipeline through which the data will pass when being analyzed. It states the primary users of the pipeline and also how they will be interacting with it in order to get the desired outcome from the data analysis process.

**Primary User.**

The data analyst is the primary user of the data pipeline whose use case is described below.

**Use Case Diagram.**

Below is the use case diagram of the sale analysis and prediction data pipeline.



**Description of the Use Cases in the Use Case Diagram.**

Below are the descriptions of the key use cases in the use case diagram.

1. **Loading Data.**

In this module, the data, which is stored in a csv file, will be copied and loaded into the system. This will be done so that an analysis can be done on the copied data. In this use case, there is a sub-module of **Read Data from CSV file** in which the data will be copied from a csv file into the system.

**Outcome.**

At the end of this module, the data which is to be analyzed will have been copied from the csv file in which it is stored and loaded into the system for analysis and evaluation.

1. **Cleaning Data.**

In this module, the data which will have been loaded into the system will be cleaned and prepared before the analysis process can be carried out on the data. In this module, there will be three sub-modules, **Handling missing values**, **Replacing misspelled values** and **Label encoding** in which the missing values in the data will be filled, all the misspelled in the data will be corrected and label encoding will be carried out on the loaded data respectively.

**Outcome.**

At the end of this module, the loaded data will have been converted into a state in which an analysis can be comfortably carried out on the data without having any errors or wrong results.

1. **Data Visualization.**

In this module, the data which will have been cleaned will be visualized using the selected visualization tools and graphs/plots so as to enable the analyst to get some of the required deductions from the data.

**Outcome.**

The end result of this module will be a set of graphs and plots generated from the data which will enable the analyst to acquire some of the required deductions concerning the data analysis process.

1. **Apply algorithms and modeling techniques.**

In this module, the necessary data analysis algorithms will be applied to the data so as to identify which product properties affect the volumes of sales of the specific product most as well as predict the sales of that product. The model to be used in this module is a **Linear Regression Model**.

**Outcome.**

The end result of this module will be the predicted volume of sales of each product as well as a summary of which product properties affect the volumes of sales of all the products the most.

1. **Evaluation of algorithm and modeling technique results.**

In this module, an evaluation of the results of the model used will be done so as to establish its level of accuracy.

**The New Github Repository.**

Because the link to the first github repository did not work, the link to the new git hub repository to which the documents are being uploaded is below.

https://github.com/Data-Science1998/RecessDataScience.git