MAKERERE UNIVERSITY BUSINESS SCHOOL

PDG IN BUSINESS INTELLIGENCE AND DATA ANALYTICS

PGDBD 5106: PROGRAMMING FOR DATA ANALYTICS

MODULE 1, SEM 1 2026/2026

MODULE PROJECT

SUBMISSION DATE: 30th October 2026

Retail Branch Performance and Stock-Planning Analytics

Jupyter Notebook-based practical project

1. Project scenario

You are a junior business intelligence analyst supporting a retail business operating in Kampala, Jinja, Mbarara and Gulu. Management wants evidence about branch performance, product-category contribution, payment patterns and whether the available data can support a simple sales-planning model.

Main question: Which branch and product categories should management prioritise for stock planning in the next quarter, and what evidence supports your recommendation?

2. Objectives

i. Apply Python variables, data types, arithmetic, strings, formatted output and user input.
ii. Use lists, tuples, dictionaries and sets to represent business information.
iii. Use decisions, loops, break/continue, functions, modules, file handling and exception handling.
iv. Demonstrate classes, objects, attributes and methods through a branch-reporting class.
v. Load, inspect, select, filter, sort, clean and summarise a retail dataset with NumPy and pandas.
vi. Build useful charts with matplotlib and seaborn.
vii. Communicate defensible recommendations and explain assumptions and limitations.

3. Dataset

Use the supplied shop_sales.csv file. It contains 240 illustrative sales records from January to June 2026 and six original columns. The data covers four branches and includes deliberate missing values for cleaning practice.

Column	Meaning
date	Date of sale
branch	Kampala, Jinja, Mbarara or Gulu
category	Beverages, Grocery, Electronics or Stationery
units	Number of items sold; some values are missing
unit_price	Price per item in UGX
payment	Cash, MobileMoney or Card; some values are missing

The dataset is synthetic and intended only for teaching. Do not interpret it as real company performance.

4. Required notebook tasks

Part A — Programming foundations

1. Python basics: variables, int/float/string/Boolean, arithmetic operators, string methods and f-strings; calculate an illustrative VAT example.
2. Data structures: create and manipulate a list, tuple, dictionary and set; explain why each fits its purpose.
3. Control flow: implement if/elif/else, for and while loops, break, continue and nested loops.
4. Functions and modules: write calculate_revenue(), classify_performance() and branch_summary(); demonstrate return values and imports.
5. Object-oriented programming: create a BranchReport class with attributes and methods; instantiate objects and explain class/object/attribute/method.

Part B — Data programming

1. Workflow and toolkit: use Jupyter, NumPy, pandas, and matplotlib.
2. Load and explore: read CSV, inspect head/tail/shape/columns/dtypes/describe/sample; convert date to datetime.
3. Select/filter/sort: create revenue, filter records with Boolean masks, demonstrate loc and iloc, and sort results.
4. Clean/summarise: quantify missingness, justify imputation, check duplicates/invalid values, recalculate revenue and use groupby/agg.
5. Visualise: create at least four labelled charts — branch revenue bar chart, revenue histogram, category box plot and monthly revenue line chart.

Part C — Business communication

1. Write a 400–600 word management summary grounded in the notebook outputs.
2. Recommend stock-planning priorities while noting that revenue alone does not establish profitability, stockouts or demand.
3. Document cleaning choices, any limitations and further data needed.

5. Submission requirements

i. Completed Jupyter Notebook (.ipynb), executed from top to bottom.
ii. Original shop_sales.csv supplied with the project.
iii. Cleaned output dataset exported as shop_sales_cleaned.csv.
iv. Management summary in the notebook or as a one-page brief.
v. Optional branch_summary.txt created by your program.

6. Assessment rubric (100 marks)

Criterion	Marks	Evidence
Python fundamentals and data structures	15	Variables, types, strings, arithmetic, list/tuple/dict/set
Control flow, functions and modules	12	Conditions, loops, reusable functions and imports
OOP	15	Meaningful class and objects
Data loading and inspection	10	pandas load, shape, types, summaries and dates
Selection, filtering and sorting	10	Masks, new columns, loc/iloc and sorting
Cleaning and summarisation	15	Missing values, validation, groupby and descriptive statistics
Visualisation	12	Four appropriate labelled charts and interpretation
Business recommendations and communication	6	Evidence-based decisions, limitations and clear writing
Reproducibility and notebook quality	5	Runs top-to-bottom, readable sections and filenames
TOTAL	100	

7. Quality and interpretation expectations

i. Keep the original dataset unchanged and analyse a working copy.
ii. Explain why a cleaning method is appropriate; acknowledge that imputation can alter results.
iii. Every chart needs a title, axis labels, units and a short interpretation.
iv. Recommendations must distinguish observed results from assumptions.

8. Suggested filenames

PGDBD5106_Name_RetailAnalytics.ipynb
shop_sales_cleaned.csv
branch_summary.txt

Note: This work should be done individually. Each student should be able to explain the code they submit.