School of Arts, Humanities and Social Science

Module title and code: Data Visualisation CMP020L013

Title of coursework(s): Artefact and Written Report

Learning outcomes:	**LO1:** Critically analyse the effectiveness of a given visualisation technique for a particular task. **LO2:** Critically evaluate the appropriate visualisation technique for specific problems. **LO3:** Develop visualisations to report on studies of both quantitative and qualitative evaluations.
Assessment weighting	100%
Maximum mark	100%
Submission link	https://moodle.roehampton.ac.uk/mod/assign/view.php?id=2131744
Word limit	4,000
Date set	
Deadline	21 July 25, 4 pm
Feedback and marks	See the assessment criteria table.
Assessment setter's name	Dr Mohammad Javaheri

Academic Misconduct:

"Academic integrity and honesty are fundamental to the academic work you produce at the University of Roehampton. You are expected to complete coursework which is your own and which is referenced appropriately. The university has in place measures to detect academic dishonesty in all its forms. If you are found to be cheating or attempting to gain an unfair advantage over other students in any way, this is considered academic misconduct, and you will be penalised accordingly."

Further details about "Student Code of Conduct" and "Disciplinary Regulations" can be found at: https://www.roehampton.ac.uk/corporate-information/policies/

Data Visualisation Assessment Brief (CMP020L013)

The summative assessment for this module comprises two mandatory components: (1) an artefact (Python code implementation) and (2) a written report, collectively accounting for 100% of the final mark.

1. Introduction

Understanding customer purchasing behaviour and product performance is crucial for organisations aiming to optimise sales strategies, improve customer engagement, and enhance profitability. Businesses rely on data-driven insights to make informed decisions about product offerings, pricing, marketing, and customer retention. Analysing sales transactions at the customer level enables businesses to identify high-value segments, track purchasing patterns, and assess the effectiveness of sales interventions. In this context, a comprehensive sales dataset can support strategic decisions across departments, including marketing, operations, and finance.

2. Dataset Information

The `Prodcut_Sales.csv` dataset provides a detailed snapshot of product sales at the customer level. It includes records of individual transactions, capturing key attributes such as customer ID, product ID, quantity sold, unit price, date of purchase, and product category. The dataset also includes customer-specific features that can support segmentation and targeted analysis. This data enables detailed exploration of customer-product interactions and supports use cases such as market basket analysis, sales forecasting, and customer lifetime value estimation. The metadata of the dataset is available on the Moodle page (`Metadata_Prodcut_Sales.txt`).

3. Tasks

3.1 Data Preparation and Exploratory Data Analysis (EDA) (20 Marks)

- Clean and preprocess the dataset by handling missing values and outliers
- Perform statistical analysis to understand data distribution
- Document all data cleaning steps and transformations

3.2 Research Questions and Integration of Domain Knowledge (30 Marks)

- Formulate three specific research questions based on the dataset
- Integrate domain-specific knowledge to formulate your questions
- Ensure questions are answerable with available data
- Justify the relevance and importance of each question

3.3 Data Visualisation (40 Marks)

- Create appropriate visualisations relevant to previous tasks and research questions
- Apply appropriate colour schemes, visual encoding, labels, and legends

- Optimise visualisations for clarity, impact taking into account your potential audience

3.4 Report Compilation

Compile a 4,000-word report (excluding references and appendices) using the provided template on Moodle (`Report Submission Template`) that synthesises all aspects of your analysis and visualisations into a cohesive narrative. The report should cover:

- Detailed interpretation of each visualisation
- Identify key patterns, trends, and relationships
- Support findings with statistical evidence
- Draw meaningful conclusions from the visualisations
- Address potential limitations and biases
- Structure report with clear sections and flow
- Include proper citations following the IEEE format
- Create an engaging narrative connecting visualisations
- Provide a summary of findings and limitations

4. Deliverables

The assessment requires the mandatory submission of two components:

1. Artefact: Python script (`.py`) or Jupyter notebook (`.ipynb`), named with your student number (`ABC12345678.py/.ipynb`)

2. Report: A written report named with your student number, preferably in PDF format (`ABC12345678.pdf`)

- Use the provided report template for your submission

4.1 Technical Requirements

- The Python implementation file (`.py/.ipynb`) and the dataset (`Prodcut_Sales.csv`) must be collocated in the working directory (same folder) to ensure proper execution. This single-directory configuration is mandatory for the assessment.
- The Python code must successfully load the dataset from the working directory on university machines running a Python `3.10+` runtime environment and generate the reported output without issues as evidence of analysis and visualisations; otherwise, you will fail the assessment.
- You must submit both parts for a valid submission; otherwise, you will fail the assessment.
- While coding quality and style are not directly assessed, clear and well-commented code is expected.

5. Assessment Criteria

Criteria	0-49% (Fail)	50-59% (Pass)	60-69%	70-79%	80-100%
Data Preparation and Exploratory Data Analysis (EDA) **(20 Marks)**	Minimal or ineffective data cleaning; critical issues remain unaddressed, leading to unreliable analysis. Lacks critical analysis or insight; limited or superficial examination of data characteristics. Python code fails to load the dataset from the working directory; data cleaning/EDA calculations fail to execute; the reported visualisations cannot be generated.	Basic data cleaning attempted, addressing some issues but leaving gaps; limited rigor in handling outliers. Basic EDA with limited insights; lacks depth in identifying patterns, trends, or relationships. Python code successfully loads the dataset from the working directory; some of the data cleaning/EDA calculations are executed, and only some of the visualisations are generated successfully.	Adequate data preparation with minor issues; most missing values and inconsistencies addressed. Competent EDA showing good understanding of data characteristics, with some insightful observations. Python code successfully loads the dataset from the working directory; most data cleaning/EDA calculations are executed, and most visualis ations are generated as reported.	Comprehensive data cleaning with minimal adjustments needed; data is thoroughly prepared for analysis. Detailed EDA with strong critical analysis; insightful observations on trends, distributions, and relationships. Python code successfully loads the dataset from the working directory; all reported data cleaning/EDA calculations are executed, and all reported visualisations are generated as reported.	Exemplary data preparation with rigorous handling of all issues; data is fully ready for robust analysis. Advanced EDA demonstrating exceptional critical insight; provides in-depth understanding of complex patterns, trends, and relationships. Python code successfully loads the dataset from the working directory; all reported data cleaning/EDA calculations are executed and well-commented, and all reported visualisations are generated as reported.
Research Questions and Integration of Domain Knowledge **(30 Marks)**	Absent or irrelevant; lacks focus or relevance, providing minimal guidance for analysis; no integration of domain knowledge;	Basic questions that are somewhat relevant but lack critical depth and focus; minimal integration of domain knowledge	Relevant and clear research questions that guide meaningful exploration, showing moderate insight, and relevant domain	Well-developed questions that drive a focused, critical exploration of the data, revealing clear insights, domain knowledge is well-integrated	Highly insightful, relevant, and in-depth questions that guide complex analysis; demonstrate critical understanding; expertly integration of domain

Criteria	0-49% (Fail)	50-59% (Pass)	60-69%	70-79%	80-100%
	minimal contextualisation.		knowledge but lack depth.	to enhance understanding.	knowledge, and current research, adding significant context and depth.
Data Visualisation **(40 Marks)**	Visualisations are poorly designed, unclear, or fail to support the analysis meaningfully.	Some visualisations present but lack clarity, relevance, or impact; limited support for analysis; limited application of visual encoding and colour.	Clear and mostly relevant visualisations that support analysis; Effective visualisations that are mostly clear and relevant; appropriate use of visual encoding and colour	well-crafted visualisations that enhance clarity and depth of analysis. Highly effective visualisations; well-designed application of visual encoding and colour for clarity and insight.	Expertly designed visualisations that are clear, highly relevant, and demonstrate critical insight; add significant depth to the analysis. Exceptional visuals are expertly crafted with excellent visual encoding and colour that enhance insight and clarity.
Structure, Writing Style, Storytelling with Data and Referencing **(10 Marks)**	Poor organisation, lacking logical flow; frequent language errors; lacks a coherent narrative, missing or incorrect citations; and inconsistent IEEE referencing.	Basic structure with limited coherence; some language issues; basic narrative with limited insights; citations used but with IEEE referencing errors.	Well-organized report with clear explanations; logical flow; mostly, narrative conveys some insights; accurate IEEE referencing with minor citation issues.	Strong structure and clear, precise language; strong narrative with critical insights; accurate citations; IEEE referencing used consistently.	Exceptionally organised, cohesive, and engaging; precise language; compelling, well-rounded narrative demonstrating exceptional understanding; critical insight and in-depth, thorough and accurate use of IEEE referencing; citations well-integrated.