

Statistical Programming Assignment Brief

Contents

Module Details	1
Assignment Description.....	2
Learning Outcomes	3
Advice and Guidance	3
How is this assessment marked?	4
Marking Criteria	6

Module Details			
Module code:	LDS7001M	Level of Study:	7
Module Leader(s):	Dr Anu Bala	Credits:	20
Assessment format:	Portfolio (Anonymous) - Statistical analysis of data with accompanying analytical report	Method of submission:	Turnitin within Moodle
Deadline or Assessment Period:	14 th Jan 2025, 12Noon	Feedback date and place:	5 th Feb 2025, Turnitin within Moodle submission
Assessment limits: length, load, word count, etc.	N/A	Component number:	1
Is this exempt from anonymous marking under the policy?	No	Component weighting:	100%

Assignment Description

The following scenario presents an open-ended, practical problem that challenges learners to apply their expertise in statistical programming and data analysis to a real-world dataset. The assignment requires using various data structuring, manipulation techniques, and effectively communicating results through visualizations and a report.

You work for an e-commerce company that wants to optimize its sales strategy by analysing historical transaction data. The company has a database containing details of customer transactions, including customer ID, transaction date, product ID, product category, quantity purchased, and total price.

Your task is to perform exploratory data analysis (EDA) on this data set to answer the following questions:

1. Analyse and compare the monthly fluctuations in total revenue and the number of transactions. Identify any significant anomalies or outliers.
2. Determine which product categories have the highest total revenue and demonstrate consistent revenue growth trends. Identify any categories with sustained increases in sales.
3. Explore the seasonal variations in sales for different product categories. Are there any categories that are sensitive to specific time periods?
4. Analyse shifts in customer purchasing behaviour across multiple transactions to identify recurring patterns or significant changes in preferences. Determine whether these trends can provide actionable insights for enhancing the company's marketing strategy.

To complete this assignment, you will primarily use **Python** and the **Pandas** library to read and manipulate the data. You may also utilize visualization libraries such as **Matplotlib** or **Seaborn** to create insightful graphical representations of the data. These tools will allow you to perform comprehensive exploratory data analysis, helping you to draw meaningful conclusions and clearly present your findings.

Your analysis will provide the company with actionable insights that can be used to optimize marketing and sales strategies based on customer behaviour and product performance.

Dataset:

The datasets are available to download from the VLE under the Statistical Programming Module as ***purchase_data.csv***

Assignment Description

Each plot in your assessment should be thoroughly explained to ensure clarity and relevance. Additionally, the plot itself should be clear enough to convey its insights independently. Support your code implementation with well-written documentation.

You should submit a report summarizing your findings, including tables and charts to support your analysis. Your report should also include a brief discussion of any limitations or caveats in your analysis.

Note:

- Convert the assessment report and code documentation into a **single PDF file** before submission.
- Submit the **Jupyter notebook code separately** in **.ipynb format**.

Learning Outcomes

You must successfully achieve the following Learning Outcomes to pass this assessment:

PLOs 7.1, 7.2, 7.3, 7.4, 7.9

7.1 Critically apply skills, techniques, and knowledge from a range of data analysis methods and algorithms for enhancing and solving problems in various domains.

7.2 Develop abstract thinking and design ability to analytically demonstrate concepts relating to data science.

7.3 Use research-based knowledge for the design of experiments, analysis, and interpretation of data to provide valid results.

7.4 Critically evaluate and analyse advanced data science topics, and concepts, and implement them in workplace.

7.9 Critique legal, social and ethical issues within the field of data science and applicable ancillary sectors, as applied to contemporary research and industry practice

Advice and Guidance

It is important that you present your own original work, and not just copies from others. It is acceptable to quote from others if you provide the author's name and use quotation marks or paraphrase. We will investigate your work for academic misconduct if you do not follow these instructions. Especially if your Turnitin similarity score is greater than 25% or your individual matches are greater than 6%, this is likely to occur.

Advice and Guidance

If you require support with your study skills, please visit <https://www.yorks.ac.uk/london-campus/student-support/student-support-and-guidance/>

Please refer to the York St John University [Code of Practice for Assessment and Academic-Related Matters](#).

We ask that you pay particular attention to the academic misconduct policy. Penalties will be applied where a student is found guilty of academic and/or ethical misconduct, including termination of programme ([Policy Link](#)).

You are required to keep to the word limit set for an assessment and to note that you may be subject to penalty if you exceed that limit. You are required to provide an accurate word count on the cover sheet for each piece of work you submit ([Policy Link](#)).

For late or non-submission of work by the published deadline or an approved extended deadline, a mark of 0NS will be recorded. Where a re-assessment opportunity exists, a student will normally be permitted only one attempt to be re-assessed for a capped mark ([Policy Link](#)).

An extension to the published deadline may be granted to an individual student if they meet the eligibility criteria of the ([Policy Link](#)).

Please see the assessment criteria below.

How is this assessment marked?

Your work will be marked according to the assessment instructions provided within this document and the selected Learning Outcomes' (LOs) (see above).

Furthermore, this assessment is marked using the assessment marking criteria or a similar rubric that aligns with the University's Generic Assessment Descriptors (see below).¹ This is to ensure all assessment decisions are comparable regardless of the discipline or mode of assessment.

Please note that you **must** meet the required baseline standards (50 – 59%) which will include the LOs and minimum expectations of the assessment. Further still, you must ensure you meet the requirements of each grade boundary to progress to the next, i.e., you should demonstrate your learning through the standards of the Pass, Merit and Distinction to reach a Distinction (70 – 84%). These standards are designed to scaffold and build your learning to achieve your fullest potential in each criterion being assessed.

¹ A rubric is a type of scoring guide that markers use to set out specific components and expectations for an assignment for their students. It is then used to guide the marking they undertake.

Marking Criteria

Pass Grade Bands (100 – 50) (Learning Outcomes must be met)

Fail Grade Bands (49 – 0) (Learning Outcomes are not met)

Criteria	Deliverables	Marks
Research Skills	Learner should demonstrate good knowledge of statistical programming and data analysis to a real-world dataset. Learners should provide a rationale for the choices they have made.	5
Thinking skills & Creativity	Learners should provide detailed criticality and evidence of Data import and manipulation. These techniques should follow up with logical and sustained conclusions. Appropriate use of Pandas to read and manipulate the data <ul style="list-style-type: none"> Handling missing or invalid values, if applicable Applying appropriate data structuring techniques, such as filtering, grouping, and reshaping. 	20
Professional learning skills and practical skills	Learners should be able to demonstrate technical and professional Exploratory data analysis techniques. Using appropriate statistical methods to explore the data and answer the questions posed in the assignment.	30
	Producing informative and visually appealing plots and tables to support the analysis. <ul style="list-style-type: none"> Writing well-organised and well-documented code that follows best practices and is easy to read and understand Demonstrating proficiency with Python libraries and programming concepts. 	20
	Demonstrating critical thinking and creativity in identifying and interpreting trends and patterns in the data.	20

Communication & Academic Conventions	Citations are provided in Harvard's style and the report is well organised and presented. Producing a clear and concise written report that summarises the analysis and findings. Presenting the results in a visually appealing and informative way. Communicating technical concepts in a way that is accessible to a non-technical audience.	5
--------------------------------------	---	---

General Characteristics	Pass (50 – 59)	Merit (60 – 69)	Distinction (70 – 84)	Higher Distinction (85 – 100)	Borderline Fail (45 - 49) (Credits may be compensated)	Fail (30 - 44) (Credits may not be compensated)	Fail (0 - 29) (Credits may not be compensated)	Marks
Adjectives	Satisfactory	Good	Excellent	Exemplary	Not Successful	Unsuccessful	Fail	
Research Skills	Demonstrates basic understanding of statistical programming and data analysis, with some evidence of rationale and limited independent investigation.	Shows good knowledge and understanding of statistical programming with some independent research.	Excellent application with clear reasoning, supporting literature, and advanced analysis decisions.	Outstanding depth in understanding, innovative analysis, and problem-solving with solid research methodology.	Some misunderstanding of statistical programming concepts.	Limited understanding and misapplication of core concepts.	No evidence of statistical programming knowledge or rationale.	5
Thinking Skills & Creativity	Basic data structuring and manipulation using Pandas, with limited analysis and logical conclusions.	Good data handling, structuring, and manipulation, with some critical thinking.	Demonstrates excellent critical thinking, with insightful data analysis and evidence of creativity in problem-solving.	Highly creative and critical use of analytical techniques, offering original insights and well-sustained conclusions.	Limited critical engagement with the material; data analysis lacks depth.	Misapplies data techniques with little evidence of critical thought.	No manipulation of data or critical thinking provided.	20
Professional Learning & Practical Skills	Demonstrates a basic understanding of	Demonstrates a solid understanding of	Demonstrates strong competence in	Demonstrates exceptional mastery of	Limited understanding of EDA	Demonstrates little to no understanding of	No attempt at using EDA techniques is	30

	EDA techniques, though the analysis may be incomplete or not fully correct. Some questions are answered, but gaps in the analysis or errors in applying statistical methods are present.	EDA techniques, applying statistical methods to explore data and answer most of the questions. Some analysis may lack depth or thoroughness, but key trends and patterns are correctly identified.	applying EDA techniques, using appropriate statistical methods to thoroughly explore the data. Answers most questions posed in the assignment with accuracy and detail. Analysis is clear, insightful, and demonstrates a high level of understanding.	advanced EDA techniques, using sophisticated statistical methods to rigorously explore datasets and answer all assignment questions with precision. The analysis is not only accurate but also insightful, uncovering patterns and trends that provide deep understanding.	techniques is evident. Statistical methods are either incorrectly applied or incomplete, resulting in a failure to fully explore the data. Key questions posed by the assignment remain unanswered or poorly addressed.	EDA techniques. Statistical methods are misapplied or not used, resulting in an incomplete or incorrect analysis. Key questions are not addressed.	evident. The analysis is missing or entirely incorrect, with no application of statistical methods.	
	Produces basic plots and tables that provide minimal support for the analysis. Visuals are functional but lack depth or clarity, and do not significantly enhance understanding. Demonstrates basic proficiency with Python libraries and programming concepts. The	Produces functional and relevant plots and tables, though they may lack sophistication or depth in conveying insights. Visualizations support the analysis but may not fully utilize their potential for data exploration. Demonstrates	Produces clear, relevant, and visually appealing plots and tables that enhance the analysis. Visuals are well-designed and provide meaningful insights. Communicates ideas clearly to both technical and non-technical audiences. The code is well-organised and functional, adhering to best practices. It is	Produces highly polished, visually engaging, and informative plots and tables. Visuals are used to significantly enhance the analysis, making complex data easily understandable. The work communicates effectively to both technical and non-technical audiences. The code is exceptionally well-	Produces unclear or poorly designed plots and tables that do not significantly contribute to the analysis. Visuals may be irrelevant or misleading. Limited proficiency with Python libraries and programming concepts. The code is functional but inefficient and	Plots and tables are missing or completely ineffective, providing no value to the analysis. Visuals may be irrelevant or incorrect. Shows little to no proficiency with Python libraries or programming concepts. Code may not be functional, and the use of libraries is	No plots or tables are provided, or those that are present are irrelevant and add no value to the analysis. Demonstrates no proficiency with Python libraries or programming concepts. Code is non-functional, with no meaningful use of libraries or techniques, disorganised, and	20

code works but may be inefficient or lack advanced techniques. While readable, it is disorganised and does not fully adhere to best practices. Documentation is present but incomplete, making the code difficult to follow or understand in places.	good proficiency with Python libraries and concepts. The code is functional and well-organised, following best practices, but could be more efficient or optimised with advanced techniques. Documentation is present but may lack thoroughness or clarity in some areas.	mostly efficient, though there may be minor opportunities for optimisation. Documentation is thorough but may have minor gaps.	organised, clean, and follows industry best practices. It is well-documented, easy to follow, and demonstrates a highly efficient use of Python libraries and programming concepts.	poorly organised or lacking important functionality. Documentation is minimal or missing, making the code difficult to understand or follow.	either incorrect or inadequate. Code is highly disorganised and is poorly documented, making it hard to understand or work with.	lacks any documentation. It fails to meet the basic requirements of the assignment.	
Demonstrates limited critical thinking, with interpretations that are largely conventional and lack depth. Some trends and patterns are identified, but the analysis is superficial and lacks creativity.	Shows good critical thinking, with reasonable interpretations of trends and patterns. Creativity is present but limited, with more conventional solutions being used. Analysis is sound but may not show significant	Demonstrates strong critical thinking skills, offering well-supported interpretations of data trends and patterns. Some creative approaches are used to solve problems, though they may not be entirely novel. Shows a solid understanding of key issues and provides	Shows outstanding critical thinking, applying original and creative solutions to data interpretation. The work goes beyond standard expectations, proposing new hypotheses or insights that push the boundaries of existing knowledge in the field. The analysis shows a professional level of judgment and interpretation.	Demonstrates minimal critical thinking, with interpretations that are either incorrect or overly simplistic. Creativity is absent, and the analysis lacks depth or originality.	Shows no evidence of critical thinking or creativity. Trends and patterns in the data are either not identified or are completely misinterpreted. The work lacks any meaningful insight or originality.	No critical thinking or creativity is demonstrated. The analysis is incomplete, incorrect, or entirely missing, with no meaningful attempt to interpret the data.	20

		originality or depth of insight.	thoughtful insights.					
Communication & Professional Conventions	Basic report structure with minimal references, lacking clarity or organisation.	Well-structured report with appropriate use of academic conventions and citations.	Excellent report organisation, clear communication, and good use of academic conventions.	Exceptional report, with clear, concise language, and professional language, adhering to academic standards.	Disorganised report, lacking in clarity or proper citations.	Unclear and poorly structured report, lacking academic rigor.	No clear communication or academic conventions followed.	5