20 Days Project Challenge

**Unlocking Insights from Retail Data**

Retail businesses generate vast amounts of data daily, and within this data lies the key to understanding customer behavior, optimizing operations, and driving revenue growth. For this exercise, we will explore a dataset containing retail transaction data spanning two time periods: 2009-2010 and 2010-2011.

This dataset includes information such as:

•⁠ ⁠Invoice Number: A unique identifier for each transaction.

•⁠ ⁠Stock Code: Unique codes for each product sold.

•⁠ ⁠Description: Detailed information about the products.

•⁠ ⁠Quantity: The number of items purchased in each transaction.

•⁠ ⁠Invoice Date: The date and time of the transaction.

•⁠ ⁠Unit Price: Price per unit of each product.

•⁠ ⁠Customer ID: An anonymized identifier for each customer.

•⁠ ⁠Country: The country where the transaction took place.

**Task Instructions for Mentees**

**For Data Analysts:**

Your goal is to uncover detailed insights from this data. Consider exploring the following:  
1.⁠ ⁠*Sales Trends*: Analyze how sales vary across months, quarters, and years.  
2.⁠ ⁠*Customer Segmentation*: Investigate customer purchasing behavior—identify high-value customers, repeat buyers, and seasonal shoppers.  
3.⁠ ⁠*Product Performance*: Determine which products are best-sellers and which ones lag in performance.  
4.⁠ ⁠*Geographic Insights*: Identify countries with the highest sales volume and growth potential.

Present your findings using clear visualizations, summaries, and actionable insights.

**For Data Scientists:**

Your task is to perform a clustering analysis on this dataset to group customers or transactions into meaningful segments. Considerations:  
•⁠ ⁠Use features every possible feature to cluster customers based on purchasing patterns.  
•⁠ ⁠Explore clustering algorithms such as K-Means, DBSCAN, or Hierarchical Clustering.  
•⁠ ⁠Visualize the clusters and interpret the results to understand customer behavior.

**Submission Guidelines**

Submit your findings and insights to *info@oaorogun.co.uk*. Be sure to include:  
•⁠ ⁠For data analysts: A comprehensive report with charts and insights.  
•⁠ ⁠For data scientists: A detailed explanation of your clustering methodology, code snippets, and visualizations.

**Final Note**

This exercise is designed to bridge the gap between data analysis and data science. Whether you’re diving deep into sales trends or exploring clustering techniques, this dataset offers the opportunity to develop critical skills for real-world applications.

Good luck, and happy analyzing!