DATA FOLKZ

CATAPULT DATA LEADERS

CAPSTONE PROJECT

Unsupervised Machine Learning

Customer Segmentation

Problem: This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers.

Task: Based on this data you are required to build features and model on these characteristics of users to categorize users based on their transactions . Your model will be evaluated on these criteria: -

a. Feature Engineering (Variable Imputation)
b. Model Selection Criteria (Basis of choosing the final Technique)
c. Measurement Criteria (Comparison of Various Models)
d. Scope for improvement

Expected Outputs:-

1. The Final code (Python or R) along with proper comments
2. A summary file (PDF format) stating :
    a. Problem Statement
    b. Approach Taken
    c. Interpretation of Results
    d. Minimal but effective Storyboarding - Consider the final audience of this will be the Business team. You need to bring the actionable insights in this section.

Data Dictionary:-

- InvoiceNo: Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation.
- StockCode: Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.
- Description: Product (item) name
- Quantity: The quantities of each product (item) per transaction.
- InvoiceDate: Invice Date and time, the day and time when each transaction was generated.
- UnitPrice: Unit price, Product price per unit in sterling.
- CustomerID: Customer number, a 5-digit integral number uniquely assigned to each customer.
- Country: Country name, the name of the country where each customer resides.

How to Access the Data: Please use the below drive link to download the Data.

Click Here

Hints:

1.) Import the data in your System using Pandas using encoding= 'unicode_escape'
2.) Convert InvoiceDate Column to datetime format to extract necessary information.
3.) Extract Features from Description
4.) Apply Clustering Techniques

Evaluation: The project will be evaluated based on requirements and instructions provided above.

THANK YOU