

Customer Segmentation – Purchasing Behavior Using Python

Problem Statement

An online retail store is trying to understand the various customer purchase patterns to gain insights that could benefit their business strategies. With Python, this project uses K-Means clustering to analyze the customer purchasing history and segment the customers based on their purchasing behavior.

Dataset Information

The dataset `online_retail.csv` contains 541,909 rows and 8 columns:

- **Invoice:** Invoice number
- **StockCode:** Product ID
- **Description:** Product Description
- **Quantity:** Quantity of the product
- **InvoiceDate:** Date of the invoice
- **Price:** Price of the product per unit
- **CustomerID:** Customer ID
- **Country:** Region of Purchase

Objectives

1. Derive useful insights from the customer purchasing history that can provide an added advantage to the online retailer.
2. Segment the customers based on their purchasing behavior using K-Means clustering.

Installation Instructions

Python Installation

To install Python, follow the instructions on the official Python website: [Install Python](#).

Required Libraries

Ensure you have the following libraries installed. You can install them using pip all at once using below command:

```
pip install datetime matplotlib numpy pandas plotly seaborn scikit-learn  
yellowbrick scipy
```

Installation Links for Required Packages (individually if required)

For datetime and warnings, these are part of the standard Python library and do not need separate installation.

Below are the links for details and commands (if required) to install the necessary Python packages:

- datetime: [Datetime Module](#)
- matplotlib: [Matplotlib Installation](#) : pip install matplotlib
- numpy: [NumPy Installation](#) : pip install numpy
- pandas: [Pandas Installation](#) : pip install pandas
- plotly: [Plotly Installation](#) : pip install plotly
- seaborn: [Seaborn Installation](#) : pip install seaborn
- warnings: [Warnings Module](#)
- mpl_toolkits.mplot3d: mpl_toolkits Installation (This is included with Matplotlib)
- scipy: [SciPy Installation](#) : pip install scipy
- scikit-learn: [Scikit-Learn Installation](#) : pip install scikit-learn
- yellowbrick: [Yellowbrick Installation](#) : pip install yellowbrick

Additional Resources

- [Installing Python Packages using pip](#)
- [Introduction to Python](#)

Usage

Clone the repository and navigate to the project directory:

```
git clone <repository_url>  
cd <repository_directory>
```

Run the script to perform the analysis:

python customer_segmentation.py

Insights and Analysis

The script performs the following steps:

1. **Data Preprocessing:** Cleans the data and handles missing values.
2. **Feature Engineering:** Creates new features to aid in clustering.
3. **Exploratory Data Analysis (EDA):** Visualizes data distributions and relationships.
4. **Customer Segmentation:** Uses K-Means clustering to segment customers based on their purchasing behavior.
5. **Cluster Evaluation:** Evaluates the clustering performance using metrics like Calinski-Harabasz Score, Davies-Bouldin Score, and Silhouette Score.

Conclusion

This project provides a comprehensive analysis of customer purchasing patterns and segments customers effectively to help the online retail store make data-driven decisions.

Contact

For any questions or suggestions, please contact:

Tharakh George Chacko

Email: tharakhgeorge@yahoo.co.in

LinkedIn: www.linkedin.com/in/tharakh-george-chacko-27a50051

This README file serves as a guide to understanding the project's objectives, installation requirements, and usage instructions. By following the steps outlined, users can replicate the analysis and gain insights into customer purchasing patterns.