

Capstone Project 3

Introduction:

Customer segmentation using the Online Retail dataset.

I would like to use customer segmentation, to find generalized characteristics to group customers. Once they are characterized into groups we can address their demands by

- Targeted marketing activities to specific groups
- Launch of features aligning with the customer demand
- Development of the product roadmap

Data Used:

The Online Retail dataset from UCI machine learning repository is a transnational data set which contains all the transactions 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.

Link: <https://archive.ics.uci.edu/ml/datasets/Online+Retail#>

Attribute Information:

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. Nominal.

Quantity: The quantities of each product (item) per transaction. Numeric.

InvoiceDate: Invoice Date and time. Numeric, the day and time when each transaction was generated.

UnitPrice: Unit price. Numeric, Product price per unit in sterling.

CustomerID: Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.

Country: Country name. Nominal, the name of the country where each customer resides.

Records: 541909

Approach:

I will attempt a customer segmentation with this dataset. Customer segmentation is the problem of uncovering information about a firm's customer base, based on their

interactions with the business. In most cases this interaction is in terms of their purchase behavior and patterns. I will try to explore some of the ways in which this can be used.

Since our dataset is limited to the sales records, and does not include any other information about our customers, I will use a **RFM,(Recency, Frequency and Monetary Value)**, based model of customer value for finding our customer segments.

Once I create our customer value dataset, I will perform some preprocessing on the data. For the clustering, I will be using the K-means clustering algorithm. The method is as simple as collecting as much data about the customers as possible in the form of features or attributes and then finding out the different clusters that can be obtained from that data.

I will use use elbow/Gap/Silhouette analysis on K-Means clustering plots to determine the best number of clusters to calculate;

I will attempt to find the traits of customer segments by analyzing the characteristics of the clusters.

I will build an item based recommendation system to recommend items to users by using the similarities between users.

I will be using the same dataset to build a recommendation system using the Surprise package.

Finally, I will perform Market Basket Analysis to predict the affinity between two products, defining which products are good to be sold together. I will use association rules to determine the affinity scores based on cosine similarity measures.

Deliverables

- All Python code I will develop
- A written final report
- A presentation slide deck