

# **Classification of customers of a company according to their cooperation characteristics**

## **Introduction**

The data set used in this analysis is from Online Retail Data Set. 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. Data set contains following variables:

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.

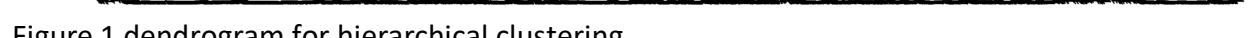
The primary aim of the analysis is to identify whether this company's consumers can be segmented meaningfully by the recency and duration of their interaction with the company, the frequency of their purchases/orders, and the amount they spent (amounts are in Sterling pounds). Further segmentation by country would also be very helpful for marketing purposes.

## **Analysis**

Table 1 lists summary statistics for variables of Online Retail data set, after data cleaning. 9 purchase variables are listed for 4286 customers from 36 different countries on a 12-month period.

In order to find a possible purchase groups among customers, hierarchical clustering method was performed on data. Pairwise distances between individual customers were calculated using daisy() function in R, which takes into account both categorical and numerical variables. Also, data points were standardized for computing pairwise distances. Figure 1 illustrates resulted dendrogram. Since the dendrogram consists of 4286 leaves, we cannot get useful visual insight by considering the dendrogram.

```
> summary(eaggr)
```



Cluster	1	2	3	4	5
---------	---	---	---	---	---

[illegible]

Instead we cut the dendrogram such that 5 clusters are obtained. Table 2 summarize the number of customers in each one of 5 clusters. For each cluster, variables associated with each member of the cluster were assembled in a data frame. Tables 3-7 list the summary statistics for each variable of members of clusters in Table 2. Using the variable Amount.per.Month, one can distinguish between clusters. Mean values for this variable are 299, 367, 709, 77184, and 14057 (sterling per month) for clusters 2, 3, 4, 1, and 5, respectively. Clusters 2 and 3 differ in terms of Month duration when a customer cooperate with the company. While customers in cluster 2 had been cooperating for 8 months on average, those in cluster 3 only cooperated with the company for a short period of time, i.e. 2 months on average. Although Amount.per.Month for cluster 4 is close to that of cluster 2 and 3, Month duration is quite different. Cluster 4 specifies customers who are permanent subscriber to the company. That is, they have a 12-month cooperation with the company.

In clusters 1 and 5 each with only one customer, the value of Amount.per.Month is extremely higher than that of other groups. Clusters 1 and 5 also differ in terms of Month duration and the first and last month of their purchases. The customer in cluster 1, buy gifts only on February, probably for Valentine's day, however, customer in cluster 5, buy gifts at the last 4 months of the year when people send out gifts for congratulating the new year.

Looking at the nationality of customers, one may conclude that almost all of the customers who made major purchases and the permanent subscribers to the company were located in UK.

However, foreign customers cooperate with the company in a limited amount of transactions.

Table 3 Summary statistics for members of cluster 1 with 1 customer.

Country	Amount	FirstMth	LastMth	Months	Purchases
United Kingdom:1	Min. :77184	Min. :2	Min. :2	Min. :1	Min. :1
Australia :0	1st Qu.:77184	1st Qu.:2	1st Qu.:2	1st Qu.:1	1st Qu.:1
Austria :0	Median :77184	Median :2	Median :2	Median :1	Median :1
Bahrain :0	Mean :77184	Mean :2	Mean :2	Mean :1	Mean :1
Belgium :0	3rd Qu.:77184	3rd Qu.:2	3rd Qu.:2	3rd Qu.:1	3rd Qu.:1
Brazil :0	Max. :77184	Max. :2	Max. :2	Max. :1	Max. :1
(Other) :0					
Amount.per.Purchase	Purchases.per.Month	Amount.per.Month			
Min. :77184	Min. :1	Min. :77184			
1st Qu.:77184	1st Qu.:1	1st Qu.:77184			
Median :77184	Median :1	Median :77184			
Mean :77184	Mean :1	Mean :77184			
3rd Qu.:77184	3rd Qu.:1	3rd Qu.:77184			
Max. :77184	Max. :1	Max. :77184			

Table 4 Summary statistics for members of cluster 2 with 1319 customers

Country	Amount	FirstMth	LastMth	Months
United Kingdom:1015	Min. : 51.56	Min. : 1.000	Min. : 1.00	Min. : 1.000
Germany : 93	1st Qu.: 693.33	1st Qu.: 2.000	1st Qu.:10.00	1st Qu.: 7.000
France : 61	Median : 1378.40	Median : 3.000	Median :11.00	Median : 9.000
Spain : 27	Mean : 2323.35	Mean : 3.409	Mean :10.73	Mean : 8.318
Belgium : 22	3rd Qu.: 2649.53	3rd Qu.: 4.000	3rd Qu.:12.00	3rd Qu.:10.000
Portugal : 11	Max. :124914.53	Max. :12.000	Max. :12.00	Max. :12.000
(Other) : 90				
Purchases	Amount.per.Purchase	Purchases.per.Month	Amount.per.Month	
Min. : 1.0	Min. : 2.673	Min. : 0.1818	Min. : 9.406	
1st Qu.: 35.0	1st Qu.: 14.415	1st Qu.: 5.2500	1st Qu.: 99.246	
Median : 78.0	Median : 18.869	Median : 10.7778	Median : 183.681	
Mean : 114.4	Mean : 33.011	Mean : 14.1823	Mean : 299.056	
3rd Qu.: 145.0	3rd Qu.: 25.642	3rd Qu.: 18.0417	3rd Qu.: 339.121	
Max. :1602.0	Max. :2027.860	Max. :149.0000	Max. :11355.866	

Table 5 Summary statistics for members of cluster 3 with 2557 customers.

Country	Amount	FirstMth	LastMth	Months
United Kingdom:2465	Min. : 0.0	Min. : 1.000	Min. : 1.000	Min. :1.000
France : 25	1st Qu.: 205.3	1st Qu.: 4.000	1st Qu.: 6.000	1st Qu.:1.000
Switzerland : 10	Median : 376.6	Median : 8.000	Median : 9.000	Median :1.000
Portugal : 8	Mean : 663.2	Mean : 7.213	Mean : 8.421	Mean :2.208
Norway : 7	3rd Qu.: 741.3	3rd Qu.:11.000	3rd Qu.:12.000	3rd Qu.:3.000
Finland : 6	Max. :44534.3	Max. :12.000	Max. :12.000	Max. :8.000
(Other) : 36				
Purchases	Amount.per.Purchase	Purchases.per.Month	Amount.per.Month	
Min. : 1.00	Min. : 0.00	Min. : 0.250	Min. : 0.0	
1st Qu.: 11.00	1st Qu.: 11.50	1st Qu.: 7.286	1st Qu.: 138.3	
Median : 24.00	Median : 17.05	Median : 15.000	Median : 239.3	
Mean : 40.81	Mean : 38.25	Mean : 22.648	Mean : 367.9	
3rd Qu.: 51.00	3rd Qu.: 23.62	3rd Qu.: 27.333	3rd Qu.: 397.0	
Max. :853.00	Max. :13305.50	Max. :366.000	Max. :39916.5	

Table 6 Summary statistics for members of cluster 4 with 408 customers.

Country	Amount	FirstMth	LastMth	Months	Purchases
United Kingdom:405	Min. : 36.56	Min. :1	Min. :12	Min. :12	Min. : 4.0
EIRE : 2	1st Qu.: 1469.98	1st Qu.:1	1st Qu.:12	1st Qu.:12	1st Qu.: 88.5
Netherlands : 1	Median : 3087.32	Median :1	Median :12	Median :12	Median : 164.5
Australia : 0	Mean : 8511.96	Mean :1	Mean :12	Mean :12	Mean : 293.2
Austria : 0	3rd Qu.: 5944.07	3rd Qu.:1	3rd Qu.:12	3rd Qu.:12	3rd Qu.: 312.5
Bahrain : 0	Max. :268478.00	Max. :1	Max. :12	Max. :12	Max. :7376.0
(Other) : 0					
Amount.per.Purchase	Purchases.per.Month	Amount.per.Month			
Min. : 2.151	Min. : 0.3333	Min. : 3.047			
1st Qu.: 10.717	1st Qu.: 7.3750	1st Qu.: 122.498			
Median : 18.795	Median : 13.7083	Median : 257.277			
Mean : 37.551	Mean : 24.4365	Mean : 709.330			
3rd Qu.: 30.123	3rd Qu.: 26.0417	3rd Qu.: 495.340			
Max. :855.935	Max. :614.6667	Max. :22373.167			

Table 7 Summary statistics for members of cluster 5 with 1 customer.

Country	Amount	FirstMth	LastMth	Months	Purchases
United Kingdom:1	Min. :56229	Min. :9	Min. :12	Min. :4	Min. :4582
Australia :0	1st Qu.:56229	1st Qu.:9	1st Qu.:12	1st Qu.:4	1st Qu.:4582
Austria :0	Median :56229	Median :9	Median :12	Median :4	Median :4582
Bahrain :0	Mean :56229	Mean :9	Mean :12	Mean :4	Mean :4582
Belgium :0	3rd Qu.:56229	3rd Qu.:9	3rd Qu.:12	3rd Qu.:4	3rd Qu.:4582
Brazil :0	Max. :56229	Max. :9	Max. :12	Max. :4	Max. :4582
(Other) :0					
Amount.per.Purchase	Purchases.per.Month	Amount.per.Month			
Min. :12.27	Min. :1146	Min. :14057			
1st Qu.:12.27	1st Qu.:1146	1st Qu.:14057			
Median :12.27	Median :1146	Median :14057			
Mean :12.27	Mean :1146	Mean :14057			
3rd Qu.:12.27	3rd Qu.:1146	3rd Qu.:14057			
Max. :12.27	Max. :1146	Max. :14057			

## Conclusion

In the Online Retailer data set, hierarchical clustering revealed that customers of the company can be clustered into different groups based on Amount.per.Month, Month and Country variables. However, determining the true classes of customers requires a large number of trial and errors assessments, which require further analyses.