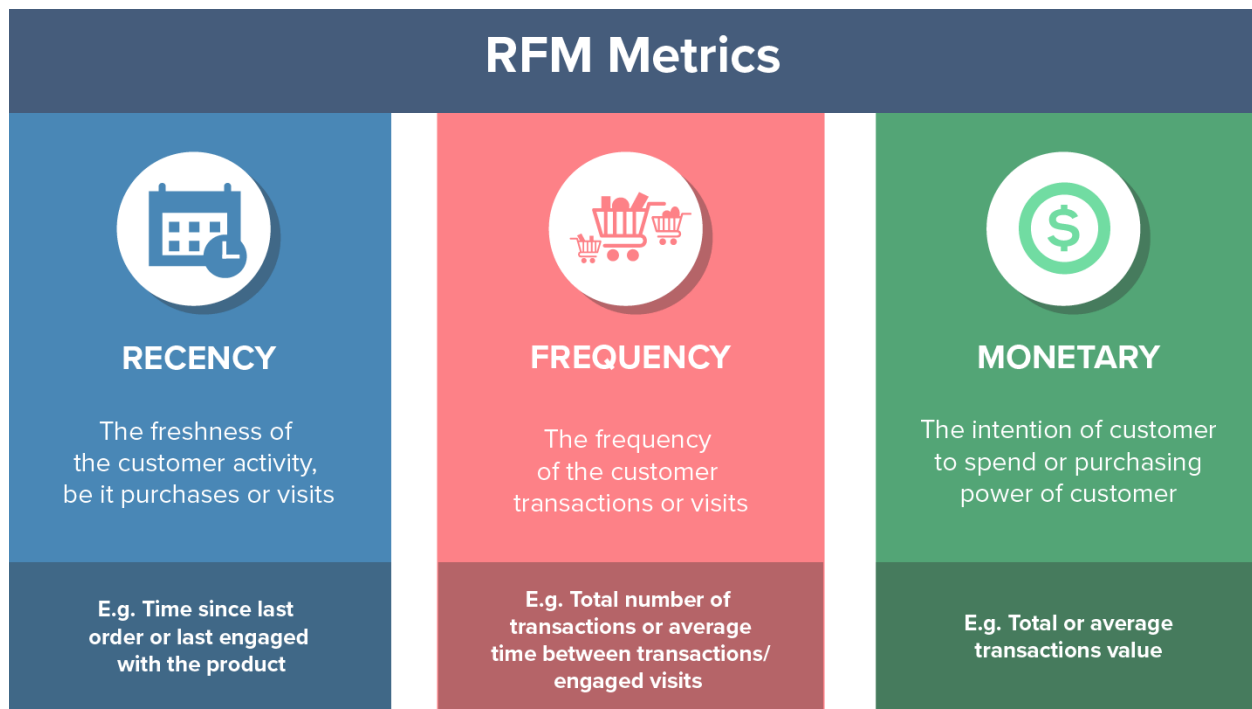**RFM Customer Segmentation**

Segmentation of customers in an online retail database combining RFM analysis and clustering.

Smart marketers understand the importance of "know thy customer." Instead of analyzing the entire customer base as a whole, it's better to segment them into homogeneous groups, understand the traits of each group, and engage them with relevant campaigns rather than segmenting on just customer age or geography.

One of the most popular, easy-to-use, and effective segmentation methods to enable marketers to analyze customer behavior is RFM analysis.

Create Recency Frequency Monetary (RFM) table

RFM is a basic customer segmentation algorithm based on their purchasing behavior. The behavior is identified by using only three customer data points:

Recency: the recency of purchase: 'How many days ago was their last purchase?'

Frequency: the frequency (total number) of purchases: 'How many times has the customer purchased from our store?'

Monetary: the mean monetary value of purchases: 'How much has this customer on average spent?' Again limit to last two years — or take all time

The RFM Analysis will help the businesses to segment their customer base into different homogenous groups so that they can engage with each group with different targeted marketing strategies. Sometime RMF is also used to identify the High-Value Customers (HVCs).

Data

This data set contains all of the transactions recorded for an online retailer based and registered in the UK between 2009–12–01 and 2011–12–09. The retailer specializes in all-occasion gift items.

Column Descriptions

Column Name	Description	Data Type
InvoiceNo	Invoice number.If this code starts with letter 'c', it indicates a cancellation.	Nominal, a 6-digit integral number uniquely assigned to each transaction
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	Invice 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 reside

After some investigation, we found that the data needed some cleaning up, so we have done some filtering to get better output. Here's what we've done:

- Filtered out all transactions where we don't have a proper CustomerID.
- Filtered out transactions where the Quantity or Unit price is zero or less.

Then we aggregated sales by customer and created a RFM table (using SQL):

```
CREATE VIEW salesByInvoice AS
SELECT CustomerId,
       InvoiceNo,
       MAX(InvoiceDate) AS InvoiceDate,
       SUM(Quantity * UnitPrice) as TotalMonetaryValue
FROM Sales
GROUP BY CustomerId, InvoiceNo

SELECT CustomerId,
       MIN(EXTRACT (DAY FROM ('2011-12-10' - InvoiceDate))) as Recency,
       COUNT(InvoiceNo) as Frequency,
       AVG(TotalMonetaryValue)::BIGINT as MonetaryValue
```

```
INTO CustomerRFM
FROM salesByInvoice
GROUP BY CustomerId
```

The table *CustomerRFM* has been exported to the *Online Retail RFM.csv* file.

Data transformation
To manage the skewness of the variables, we transformed the data into a more symmetrical form. We tried both log and cubic root transformations. The latter gives better results.

Clustering with K-means algorithm
We used the Orange K-Means algorithm. The silhouette score is near optimal for 4 clusters.

We compare the clusters based on the mean values of recency, frequency and monetary statistics of cluster data.

Cluster	Recency	Frequency	Money	Size (%)
1	222	1,6	242	28%
2	37	2,7	266	42%
3	90	2,7	1417	12%
4	16	13	417	18%

By interpreting RFM values, we can characterize the customers in the clusters:

Cluster	Type of customers	RFM interpretation
1 (28%)	Lost customers	Customers who purchased a very long time ago and who have made few purchases
2 (42%)	About to sleep customers	Customers who purchased a while ago and who have made few purchases. Will loose them if not reactivated.
3 (12%)	Hibernating customers	Customers who purchased a long time ago and who have made few but large purchases in the past. May be lost.
4 (18%)	Loyal customers	Frequent customers with high monetary spending and recent purchases