E-Commerce Data:

Customer Segmentation Capstone

By Shane McCallum

March, 2021

Problem Statement:

Problem statement worksheet

In order to develop a consumer focused market strategy, how can our client's customer data be categorized for forecasting what they will purchase over the next year using an RFM categorization strategy?

Context:

Our client is a UK-based retailer that primarily sells unique all-occasion gifts. Many customers of the retailer are wholesalers. Our client has provided the previous year's (01/12/2010 - 09/12/2011) transaction history of its customers. The client wants a model which identifies customers based on recency, frequency and monetary value for targeted marketing strategies such as reward programs, VIP memberships, and mailing lists. All of this will allow our client to focus on customer retention across the entire customer base.

Criteria for success:

The model most properly categorize:

- the recency of customers (time between purchases)
- the frequency of a customer's purchase history
- the monetary value of a customer (how much they spend)
- and accurately categorize future customers.

Scope of solution space:

Focus will be on the past year's transactional history for all customers provided by the client. After identifying key features of the data, a model can be developed based on the key features that properly categorizes the customers.

Constraints within solution space:

- Customers that have no ID (likely customers that purchased as a guest) will need to be dropped as they cannot be consistently measured
- Customers that are other large retail companies will likely be outliers
- The model will be based on the recent trends and cannot accurately measure developing trends for the next year (think fidget spinners and pop culture).

Stakeholders to provide key insight:

Our largest stakeholders are our client's:

- CEO
- Head of Marketing
- Marketing Strategist
- PR department
- Highest RFM customers

Key data sources:

The past years transaction data provided in the csv file labeled "data.csv."

Data Science Approach:

The idea behind this capstone is to implement the Recency, Frequency and Monetary Value Classification method (RFM) for targeting a company's customer base. This allows for targeted marketing approaches such as an email mailing list, coupons regarding the items customers look at but hesitate in purchasing, what products to promote, and much more.

My approach is to use the KMeans classification method in order to properly sort and classify the customers based on RFM. I intend to develop most of the visualizations with Seaborn and perhaps Tableau depending on how well the data lends itself to interesting visualizations. For the predictive model I anticipate a Random Forest or Gradient boosting model will provide the highest accuracy as they often perform the best with categorical data like this.

The RFM classification model will answer the clients questions about the nature of their customers and the predictive model will ideally aid in classifying future customers over the next year.

However, this requires proper implementation of the models by the client and keeping the models up to date with incoming customer transactions in order to remain relevant to emerging trends.

For further reading regarding the way RFM can increase profits and reduce expenditure please look at this resource. In order to understand the shortcomings of RFM and why it might not work as expected please look into this resource. It should be noted that RFM ***cannot*** predict how successful a future marketing campaign will be. It ***can*** tell you which customers are worth developing a market campaign for and what they are interested in purchasing from you.

Deliverables:

- All code I developed (in a Jupyter notebook)
- A final written report highlighting the methods used and why
- A non-technical slide deck presentation to the client aimed at highlighting the key takeaways
- All hosted on the GitHub repository