Exploring customer segmentation and customer lifetime value for sales forecasting

Eliza Ganguly

Highlights

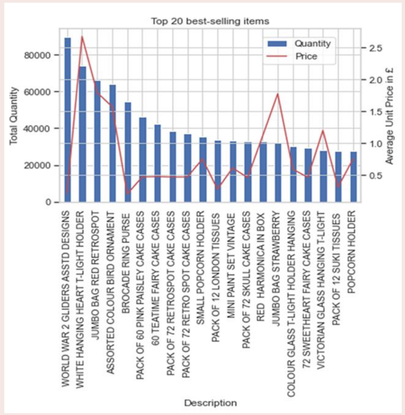
- Customer segmentation is invaluable in **targeted marketing**. Segmentation can be based on location, demographics, shopping behavior etc.
- A retail company may target athletic customers with promotions on ski jackets while showing ads for healthcare products to their elderly customers.
- **Forecasting sales** helps understand the demand, order right amount of inventory thereby avoiding excess stock or stockouts.

Background

A UK-based company that supplies unique all-occasion gifts wants to better understand the customer data available and historical sales, to **identify an opportunity** to build targeted segments to market and to forecast sales.

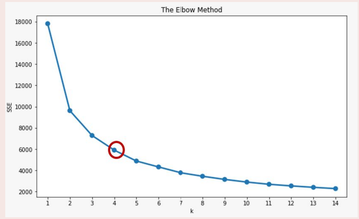
Data

The data consists of over 1 million transactions between 2009 and 2011. the 8 columns consist of invoice ID, stock code, product description, quantity, invoice dae, price, customer ID and country. Over 90% of customers were from UK.

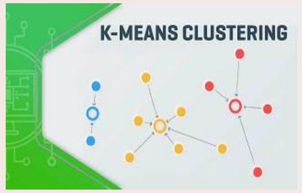


Model

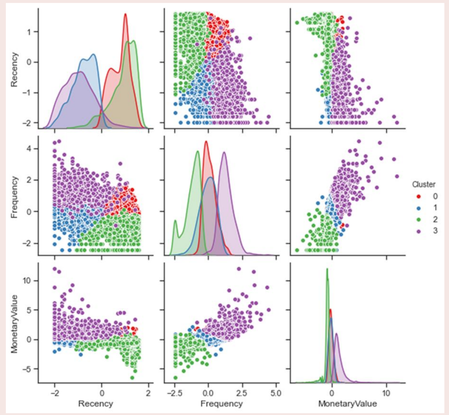
As the data focuses on shopping behavior, a K-means clustering model was implemented using recency (R), frequency (F) and monetary value (M) of purchases. The limitation of the model is that a value of K (optimal number of clusters in the data) is needed to be provided by the user.



The elbow method was employed to find the optimal value of K. At **K=4**, the clusters are found to be distinct and well separated.



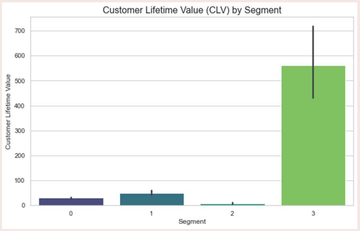
Working of a K-means algorithm.



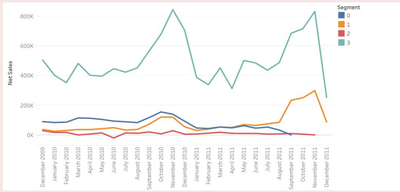
Cluster Interpretation

- cluster 0: at-risk customers
- cluster 1: new customers
- cluster 2: lost customers
- cluster 3: loyal/ high-value customers

Loyal customer segment is found to have the highest customer lifetime value (CLV)

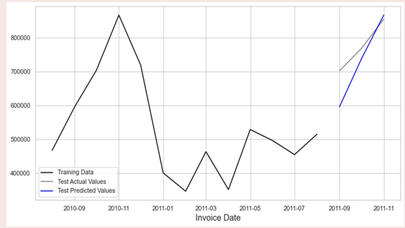


Sales Forecasting - regression analysis



- **'Net Sales'** prediction for loyal customers (segment 3)
- Cancellations are disregarded as we only want to predict profitable sales
- Sales show a cyclical trend (high sales during festive months) but no upward trend over time.

- Monthly sales predicted for the **last 3 months** in the data
- Model is trained on data from December 2009 – August 2011.
- **XGBoost** gives the best result with a **look-back period of 8 months**.



The 4 primary customer types were identified from the transaction data. At-risk and new customers may need different marketing strategies to ensure high retention rate.