Hausarbeit im Modul „Data Science und Machine Learning“ WS22/23 - Teil 3

The shopping data set

A retail company has hired you. Over the last three years, the retailer has collected data about its customers, the revenue generated for different products, and the number of purchases in its different channels. Your job as a data scientist is to perform cluster analysis and identify customer clusters that can be addressed by different marketing activities (e.g., advertising, coupons).

The following data is given:

Customer data

- *Year_of_Birth*: Customer's birth year
- Education: Customer's highest level of education
- *Marital_Status*: Customer's marital status
- *Estimated_Income*: Customer's yearly estimated household income
- *Kids_home*: Number of children in customer's household
- *Teen_home*: Number of teenagers in customer's household
- *Registration_date*: Customer’s registration date at the retailer
- Recency: Number of days since customer's last purchase
- Exist_Complain: status whether the customer filed a complaint within the last year

Revenue for each category

- *MntWines:* Amount spent on wine in the last 2 years
- *MntFruits:* Amount spent on fruits in the last 2 years
- *MntMeatProducts:* Amount spent on meat in the last 2 years
- *MntFishProducts:* Amount spent on fish in the last 2 years
- *MntSweetProducts:* Amount spent on sweets in last 2 years
- *MntGoldProds:* Amount spent on gold in the last 2 years

Number of purchases made on different channels

- *No._WebPurchases*: Number of purchases made through the company’s website
- *No._CatalogPurchases*: Number of purchases made using a catalog
- *No._StorePurchases*: Number of purchases made directly in stores

Tasks

1. Cluster the existing data based on appropriate procedures. Justify how you would cluster the data set based on the results.

2. Describe the clusters found based on the characteristics of the attributes of the customers and the cluster size.
3. What recommendation could you give the marketing department based on your analysis?