FAKULTÄT FÜR
INFORMATIK

Customer Behavioural Analytics in the Retail Sector

Team CuBA: Rutuja Shivraj Pawar, Nadiia Honcharenko, Shivani Jadhav, Sumit Kundu

Under the Guidance of M.Sc. Uli Niemann

rutuja.pawar@ovgu.de, nadiia.honcharenko@st.ovgu.de, shivani.jadhav@st.ovgu.de, sumit.kundu@st.ovgu.de, uli.niemann@ovgu.de

Data Science with R (DataSciR) - Final Presentation

January 18, 2019

Agenda

1. Motivation
2. Project Objective
3. Data Source
4. Research Questions
5. Conclusion and Scope for Future Work

Motivation

- Customer as a key-centric factor in a business
- Cost of retaining an existing customer is far less than acquiring a new one
- Customer Behavioural Analytics plays an important role in leveraging data analytics to find meaningful behavioural patterns in the customer-specific business data

"A satisfied customer is the best business strategy of all"
- Michael LeBoeuf

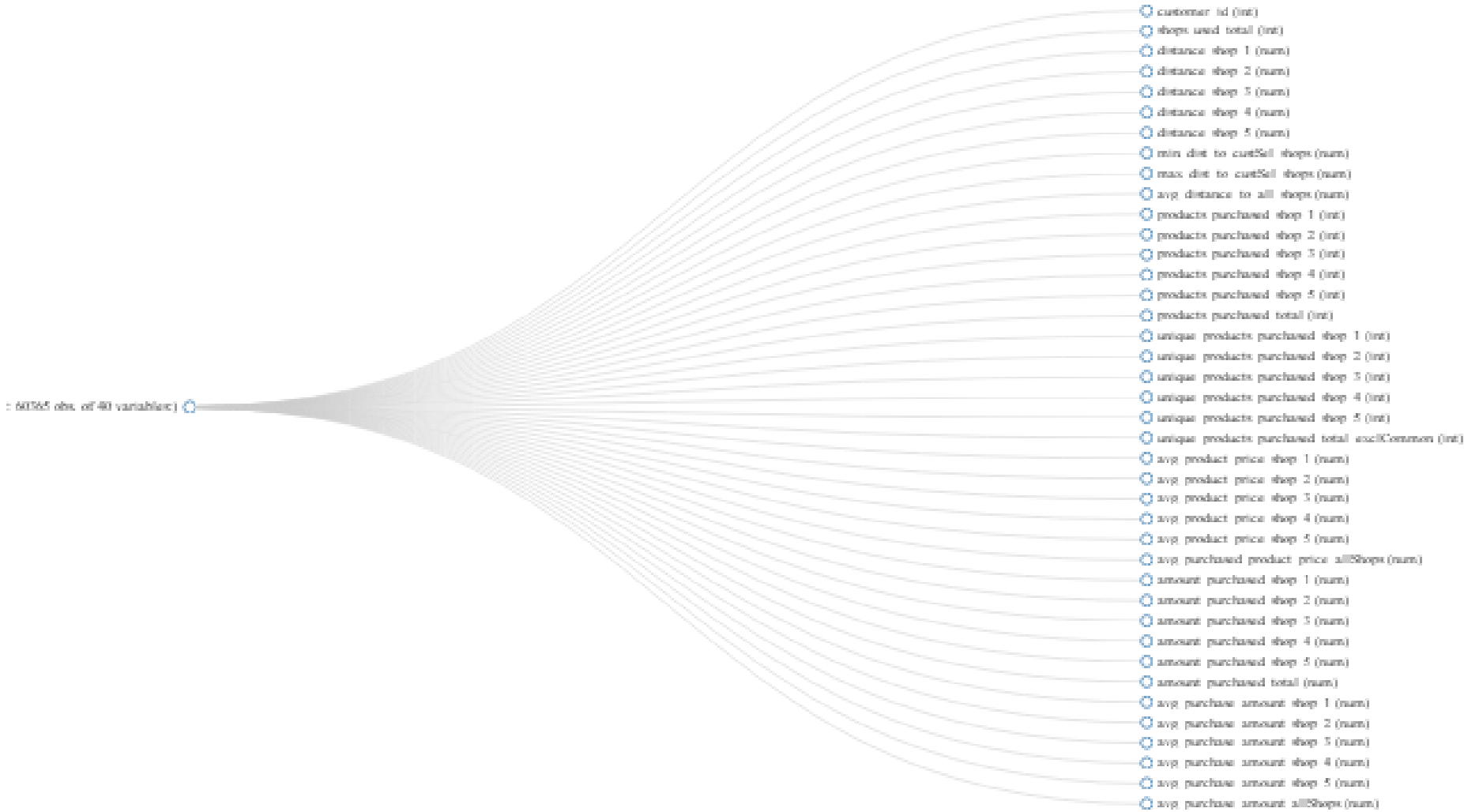
Project Objective

- Aim: To understand the purchase behaviour of customers in the retail sector specifically of an Italian retail distribution company *Coop* in a single Italian city

- Intends to discover different analytical insights through answering different formulated Research Questions (RQ)

Data Source

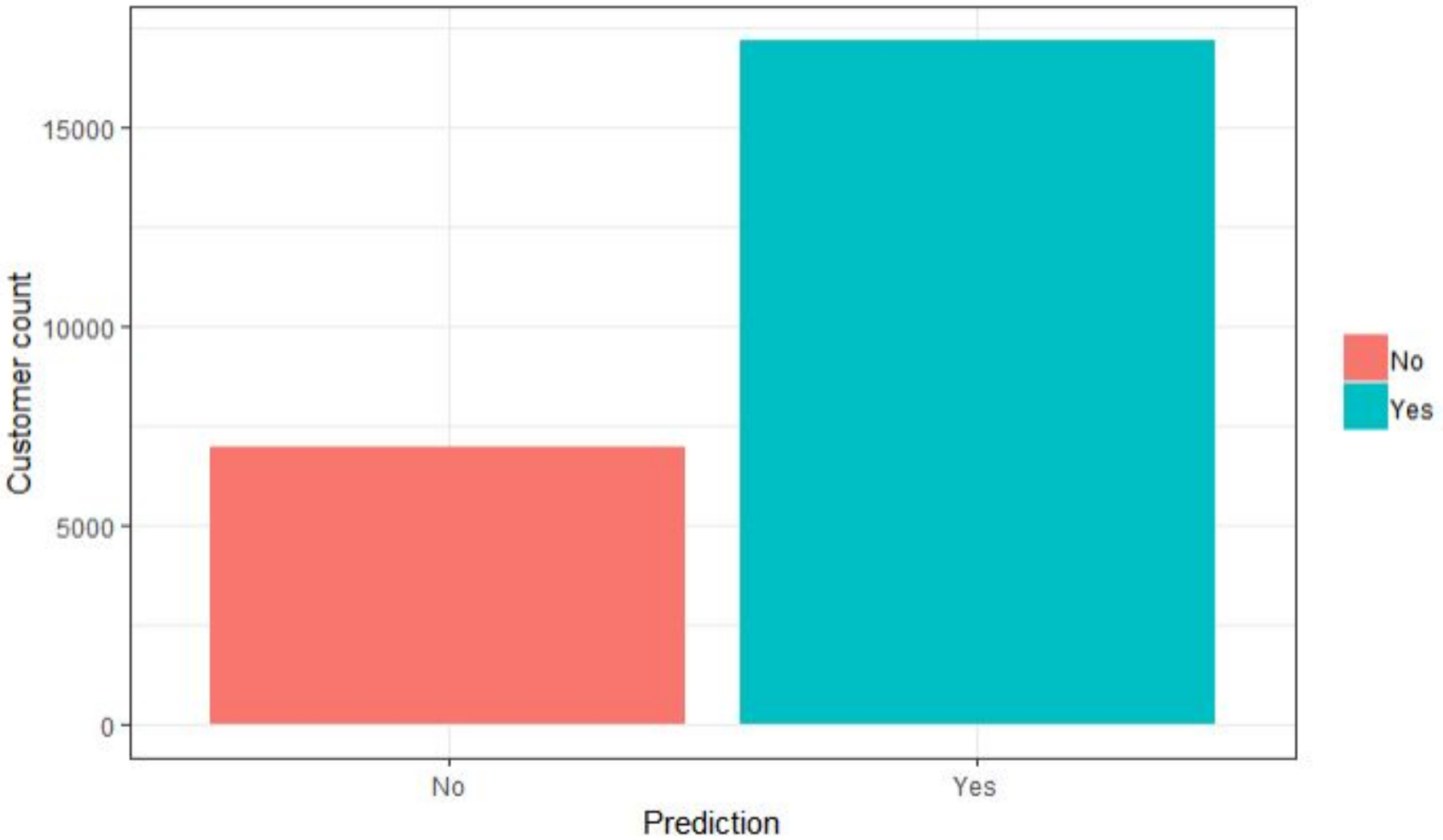
- Retail market data of one of the largest Italian retail distribution company called *Coop* for a single Italian city [2]
- Contains data aggregated from the original datasets [1] [3] and mapped to new columns
- The dataset contains 40 features with 60,366 instances and is approximately 14.0 MB in size


customer_id (int)
shops_used_total (int)
distance_shop_1 (num)
distance_shop_2 (num)
distance_shop_3 (num)
distance_shop_4 (num)
distance_shop_5 (num)
min_dist_to_custSel_shops (num)
max_dist_to_custSel_shops (num)
avg_distance_to_all_shops (num)
products_purchased_shop_1 (int)
products_purchased_shop_2 (int)
products_purchased_shop_3 (int)
products_purchased_shop_4 (int)
products_purchased_shop_5 (int)
products_purchased_total (int)
unique_products_purchased_shop_1 (int)
unique_products_purchased_shop_2 (int)
unique_products_purchased_shop_3 (int)
unique_products_purchased_shop_4 (int)
unique_products_purchased_shop_5 (int)
unique_products_purchased_total_exclCommon (int)
avg_product_price_shop_1 (num)
avg_product_price_shop_2 (num)
avg_product_price_shop_3 (num)
avg_product_price_shop_4 (num)
avg_product_price_shop_5 (num)
avg_purchased_product_price_all5shops (num)
amount_purchased_shop_1 (num)
amount_purchased_shop_2 (num)
amount_purchased_shop_3 (num)
amount_purchased_shop_4 (num)
amount_purchased_shop_5 (num)
amount_purchased_total (num)
avg_purchase_amount_shop_1 (num)
avg_purchase_amount_shop_2 (num)
avg_purchase_amount_shop_3 (num)
avg_purchase_amount_shop_4 (num)
avg_purchase_amount_shop_5 (num)
avg_purchase_amount_all5shops (num)

RQ1: Are customers willing to travel long distances to purchase products?

Approach: Classifying the data based on distances and other decision factors

Algorithms Used: Support Vector Machine, K-nearest neighbour, Random Forest


15000
10000
5000
0
Customer count
No
Yes
Prediction
No
Yes

Insights for RQ1

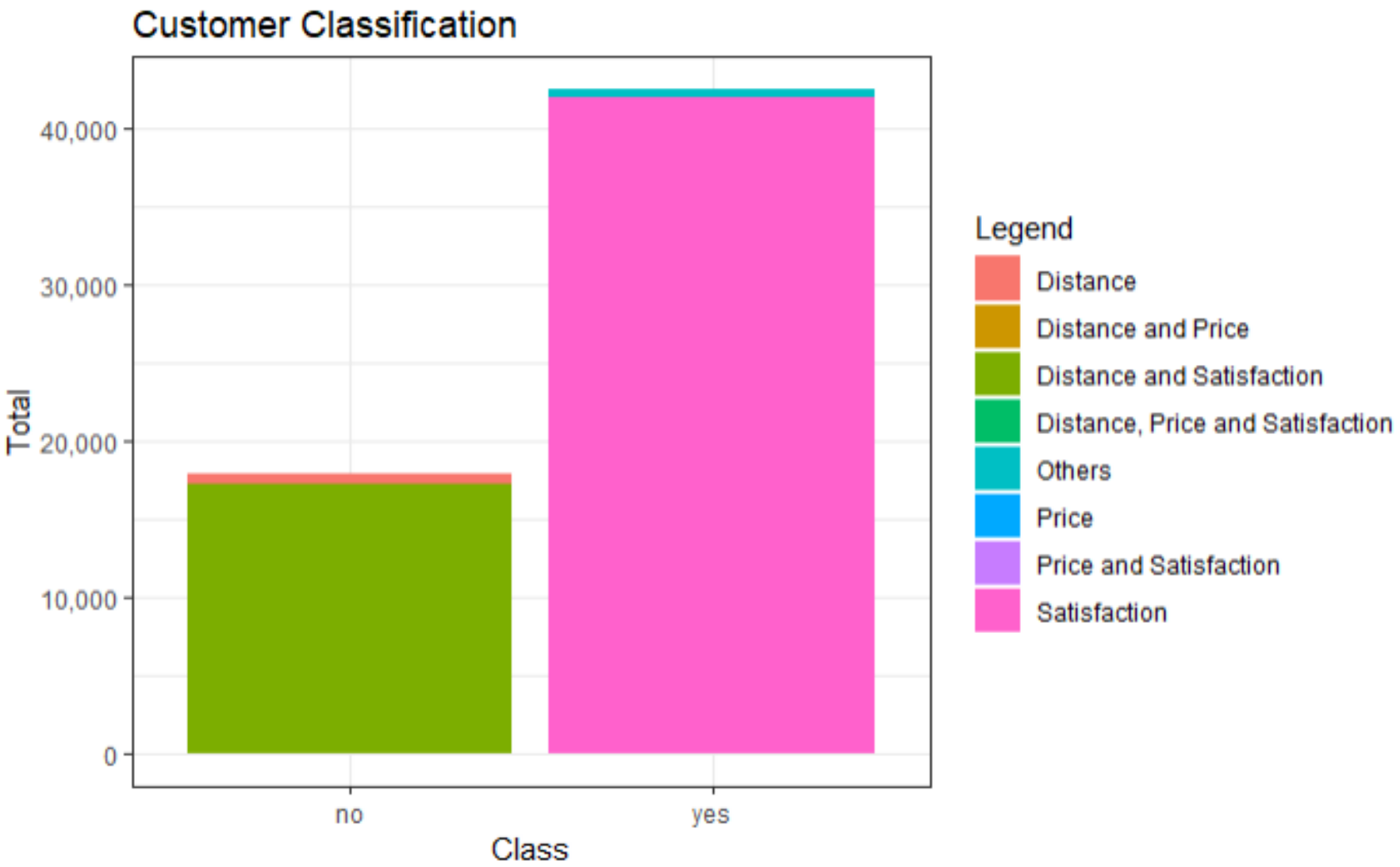
- **Observations:** It was found out that the majority of the customers are ready to travel long distances to purchase products and this decision is affected by certain factors

- **Applications for the Business:**
  - Understand the behaviour trend of the majority of the customers related to long distance travel to purchase products
  - Understand the reasons behind such majority trends
  - Devise strategies in the context of the store locations, enhance the factors influencing such trends, generate more revenue and increase customer satisfaction

RQ2: What are the factors that contribute towards the long distance travel of the customer to purchase products?

Approach: Identifying the factors for the long-distance travel of the majority of the customers

Algorithm Used: Custom Algorithm and as a follow-up investigation for RQ1


Customer Classification
40,000
30,000
20,000
10,000
0
Total
no
yes
Class
Legend
Distance
Distance and Price
Distance and Satisfaction
Distance, Price and Satisfaction
Others
Price
Price and Satisfaction
Satisfaction

Insights for RQ2

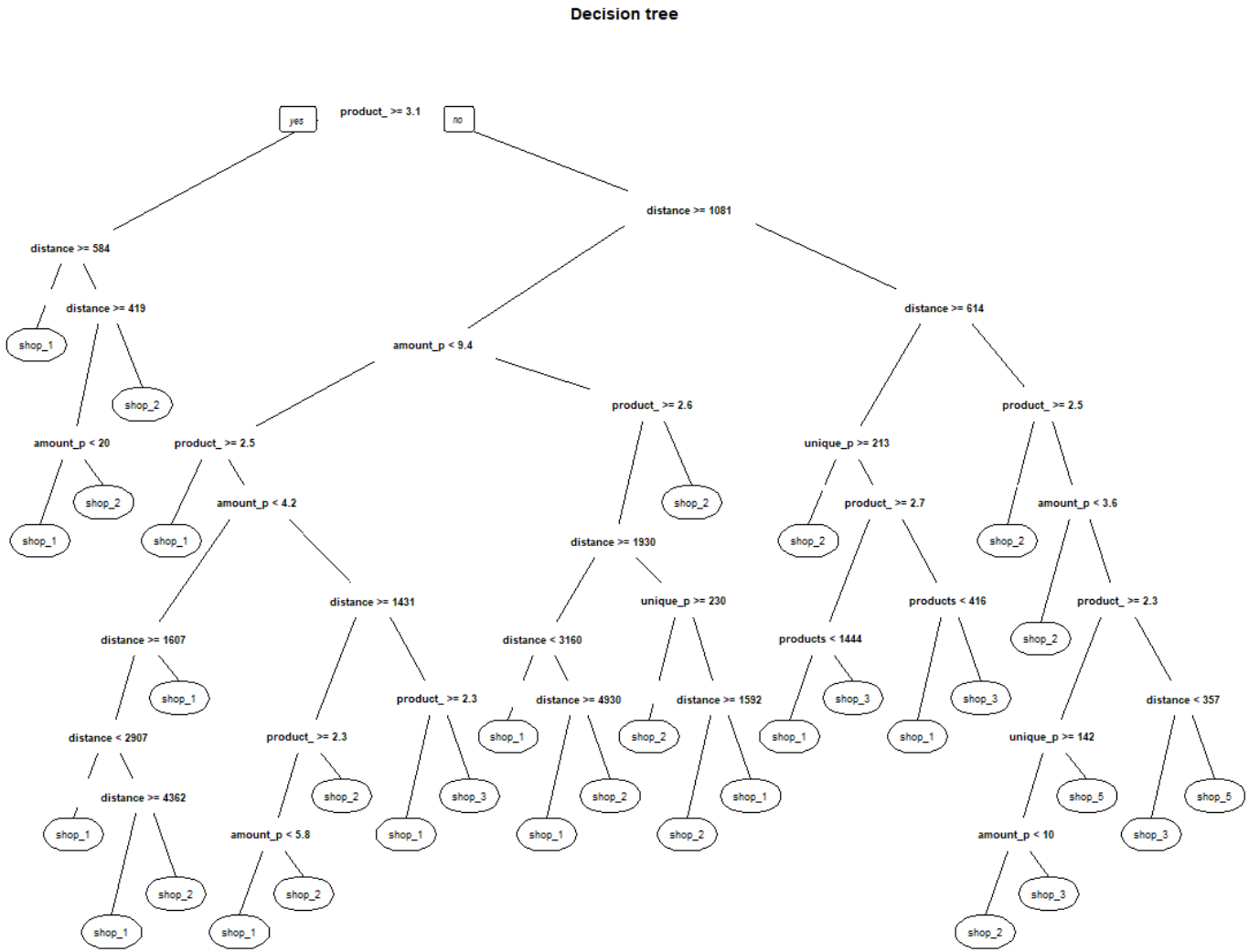
- **Observations:** Responsible factors for the majority of the customer trend towards long-distance travel were determined and It was observed that 'Satisfaction' is a key role factor affecting a customer's decision-making process.

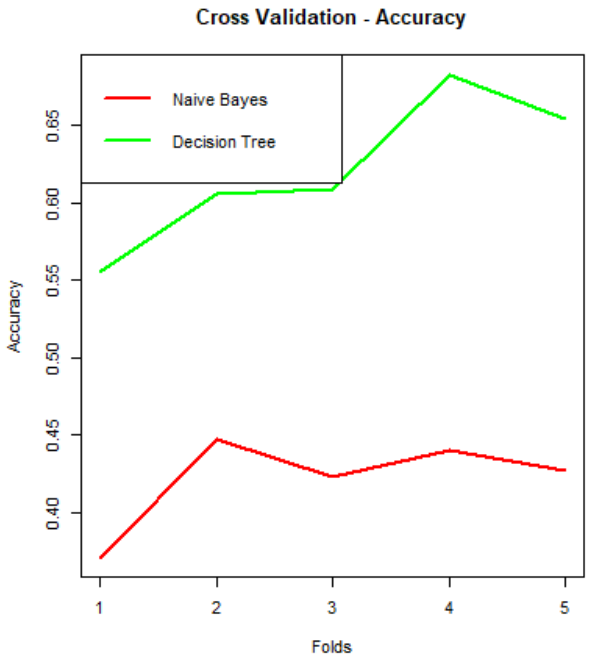
- **Applications for the Business:**
  - Devise strategies to enhance the observed most important factors facilitating customer satisfaction retention
  - Further paves way for a steady business growth

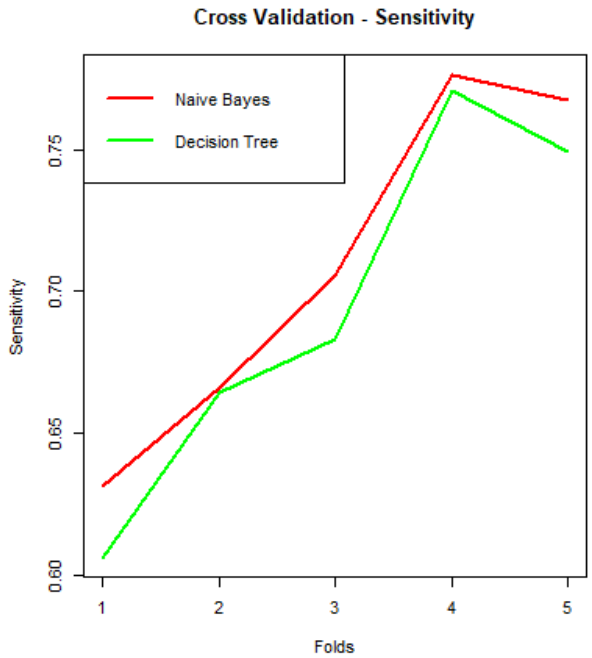
RQ3: What is the maximum likelihood of a customer to select a particular shop?

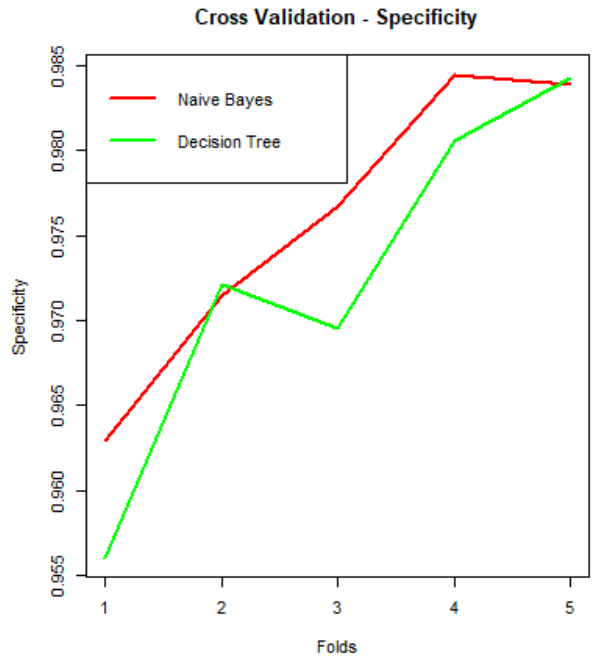
Approach: Predicting the shop most likely to be selected by a new customer

Algorithms Used: Naive Bayes, Decision Tree


Decision tree
yes
product_ >= 3.1
no
distance >= 584
distance >= 1081
distance >= 419
amount_p < 9.4
distance >= 614
shop_1
shop_2
amount_p < 20
product_ >= 2.5
product_ >= 2.6
unique_p >= 213
product_ >= 2.5
shop_1
shop_2
amount_p < 4.2
distance >= 1930
product_ >= 2.7
amount_p < 3.6
distance >= 1607
distance >= 1431
distance < 3160
unique_p >= 230
products < 1444
products < 416
product_ >= 2.3
distance < 2907
product_ >= 2.3
product_ >= 2.3
distance >= 4930
distance >= 1592
distance < 357
distance >= 4362
amount_p < 5.8
unique_p >= 142
shop_5
amount_p < 10
shop_3


Cross Validation - Accuracy
Naive Bayes
Decision Tree
Accuracy
0.40
0.45
0.50
0.55
0.60
0.65
1
2
3
4
5
Folds


Cross Validation - Sensitivity
Naive Bayes
Decision Tree
Sensitivity
0.60
0.65
0.70
0.75
1
2
3
4
5
Folds


Cross Validation - Specificity
Naive Bayes
Decision Tree
Specificity
0.955
0.960
0.965
0.970
0.975
0.980
0.985
1
2
3
4
5
Folds

Insights for RQ3

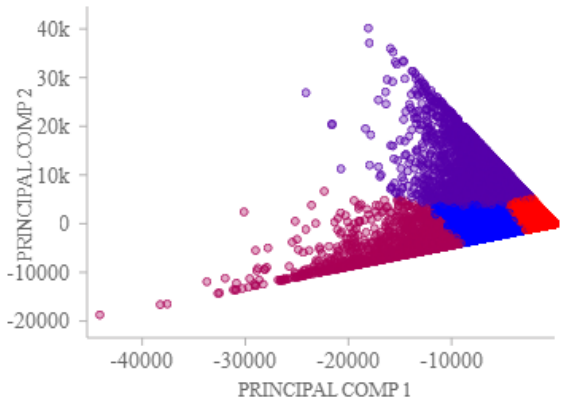
- **Observations:** The predictive model based on Decision tree will output better results in predicting the maximum likelihood of a new customer to select a particular shop.

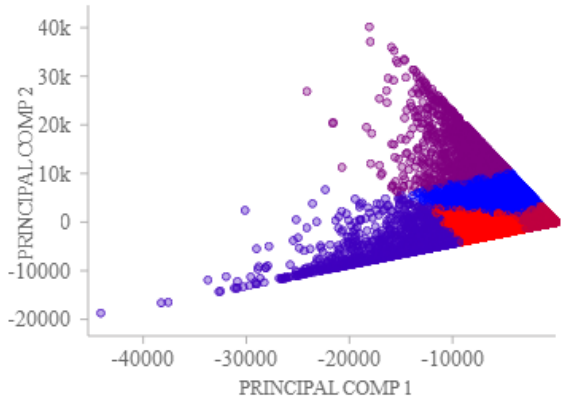
- **Applications for the Business:**
  - Understand which shops in the retail chain are most likely to be preferred by new customers
  - Facilitates towards better stock management to meet the increasing customer demands
  - Strategies to increase profit and attract new customers in different shops

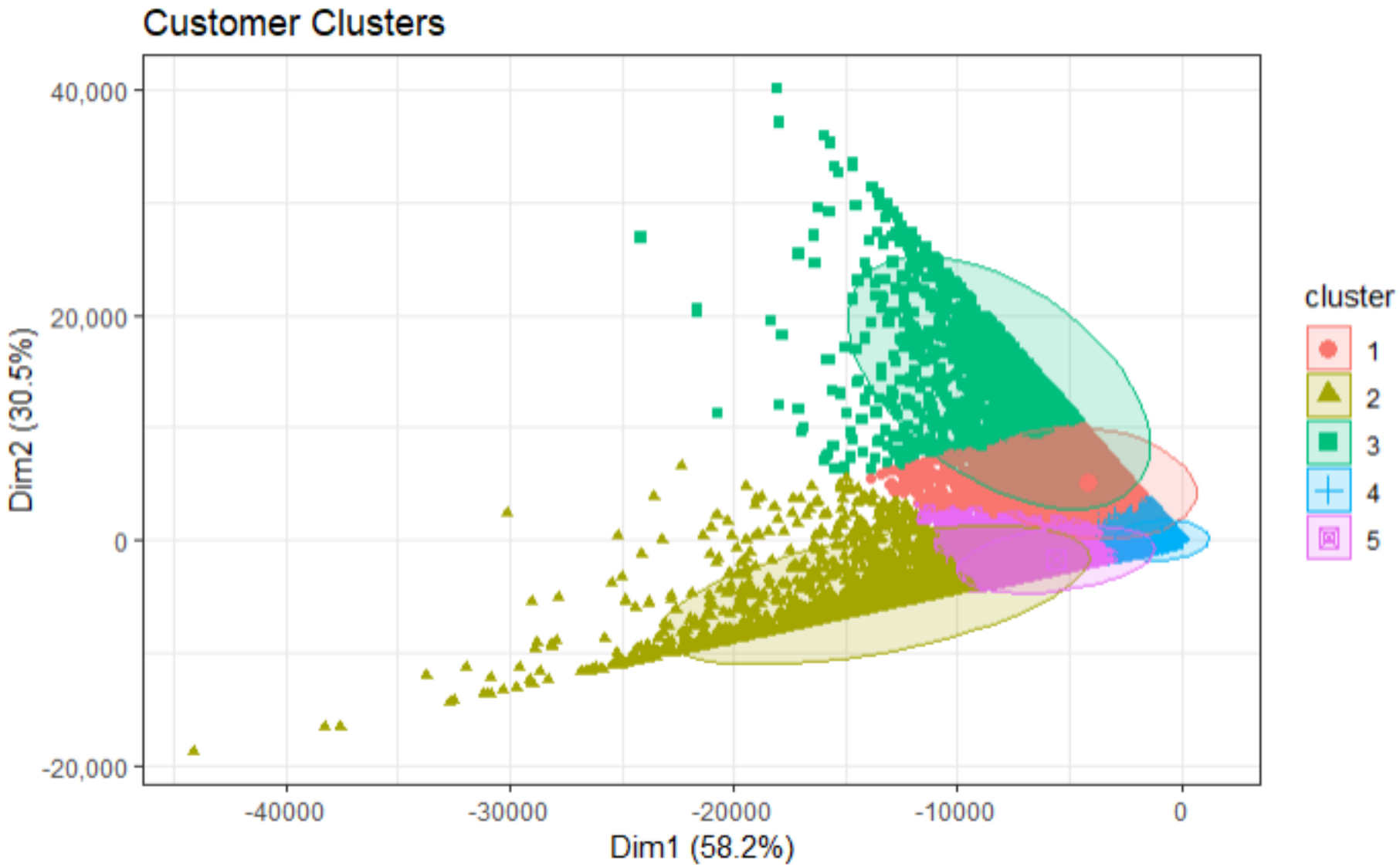
RQ4: What are the different customer segments based on their purchase behaviour?

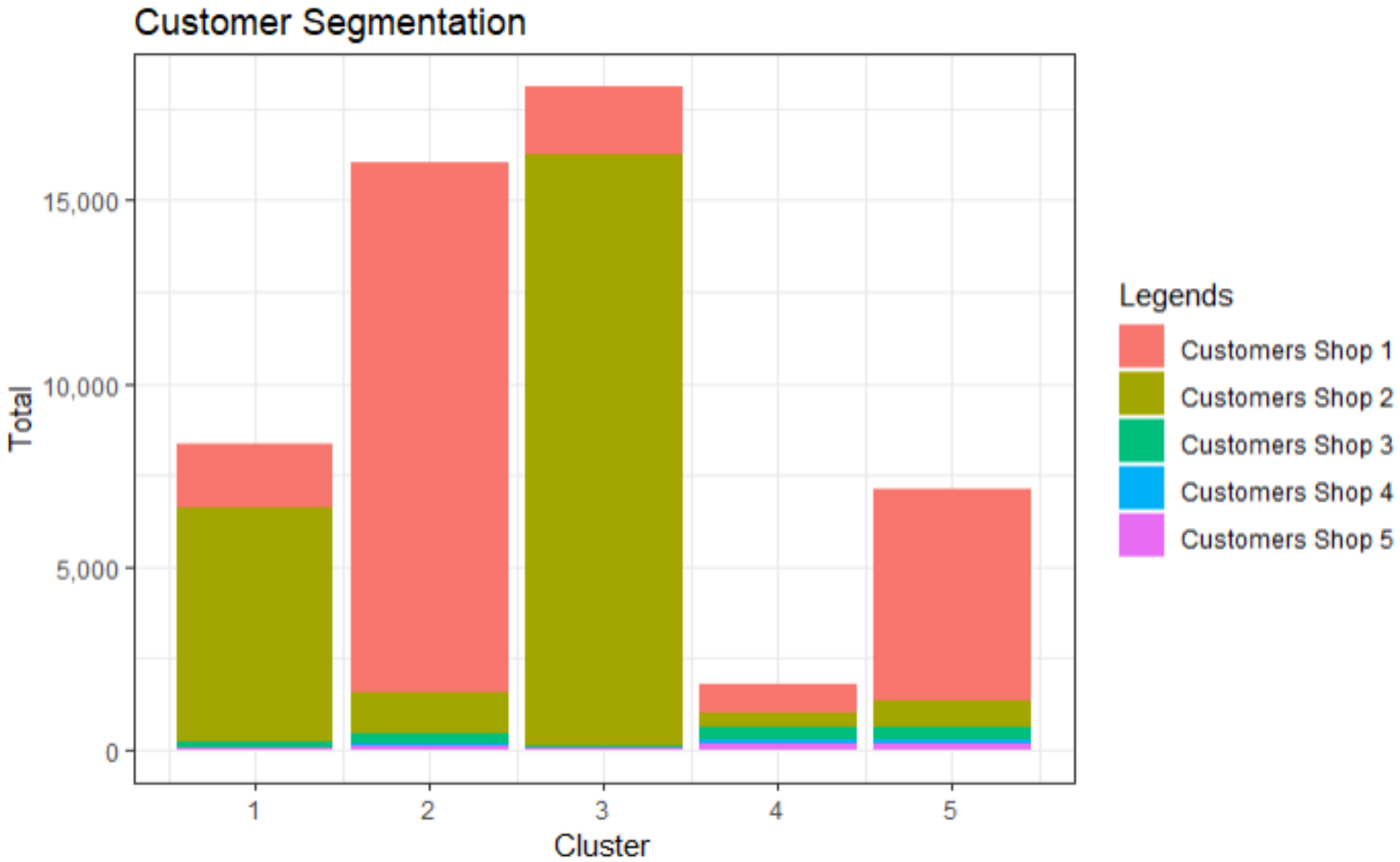
Approach: Clustering the data based on the shops customers shop the most

Algorithms Used: K-means, Principal Component Analysis






Customer Clusters
40,000
20,000
0
-20,000
Dim2 (30.5%)
-40000
-30000
-20000
-10000
0
Dim1 (58.2%)
cluster
1
2
3
4
5


Customer Segmentation
Total
15,000
10,000
5,000
0
1
2
3
4
5
Cluster
Legends
Customers Shop 1
Customers Shop 2
Customers Shop 3
Customers Shop 4
Customers Shop 5

Insights for RQ4

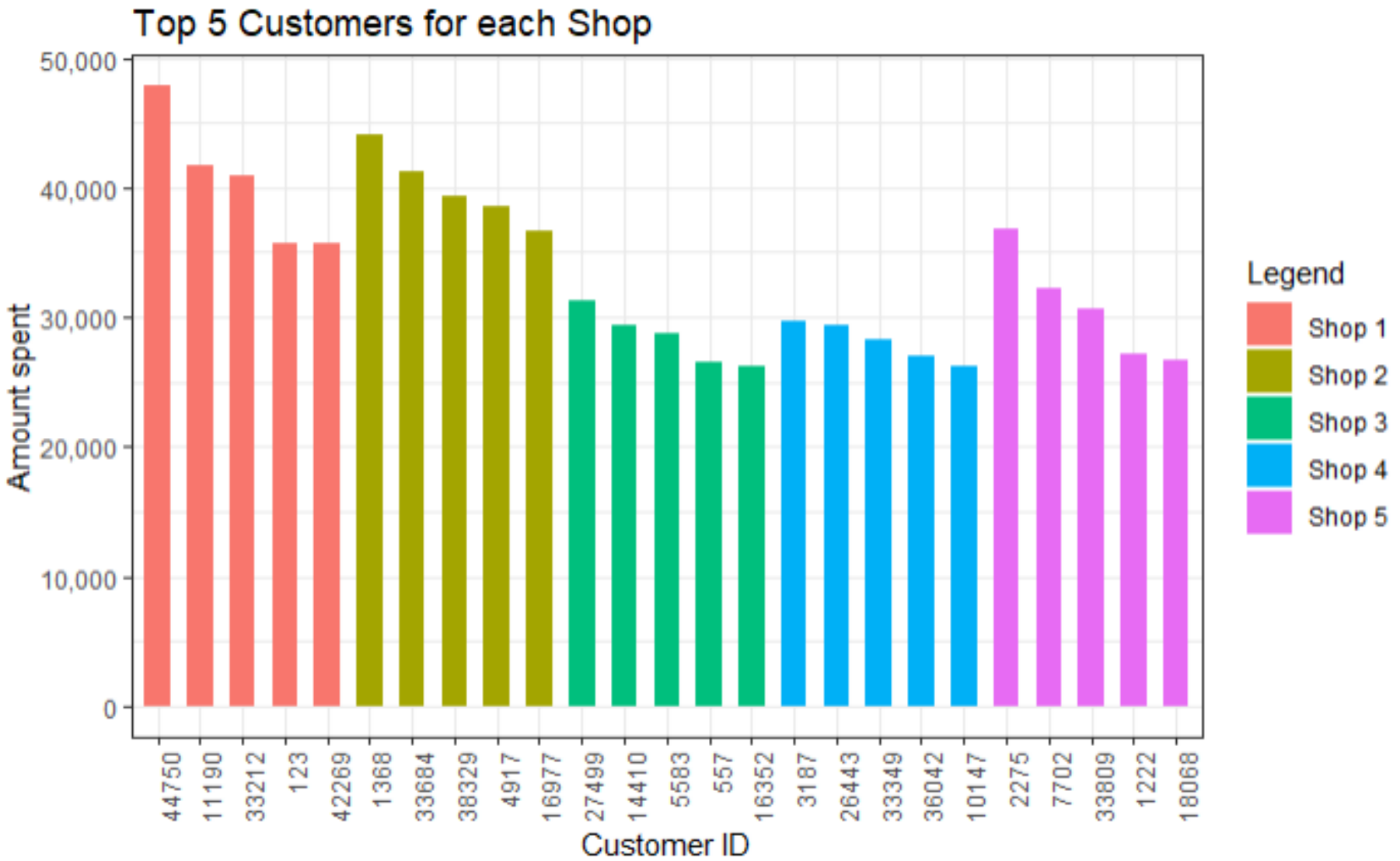
- **Observations:** Five customer segments were detected and further partitioning these segments revealed the specific customers belonging to the five different shops

- **Applications for the Business:**
  - Develop a specific strategy for each cluster base
  - Understand the purchase behaviour of customers
  - Focus marketing efforts on the right customers
  - Cut-down the marketing costs, generate more revenue and increase customer satisfaction

RQ5: Which are the Top 100 customers that are most profitable in terms of revenue generation for each shop?

Approach: Ranking the Top 100 profitable customers for each shop

Algorithms Used: Custom Ranking algorithm


Top 5 Customers for each Shop
Amount spent
50,000
40,000
30,000
20,000
10,000
0
44750
11190
33212
123
42269
1368
33684
38329
4917
16977
27499
14410
5583
557
16352
3187
26443
33349
36042
10147
2275
7702
33809
1222
18068
Customer ID
Legend
Shop 1
Shop 2
Shop 3
Shop 4
Shop 5

Insights for RQ5

- **Observations:** Top 100 customers that spend the most amount of money in each of the 5 shops based on their loyalty score were ranked and determined. Further, Top Ranked 5 customers for each shop were visualized.

- **Applications for the Business:**
  - Identify top profitable customers for each shop
  - Formulate reward schemes to retain the high-value customer base
  - Paves way to transform satisfied loyal customers as advocates for the business

Conclusion and Scope for Future Work

- Meaningful insights discovered will help the business to implement customer-centric strategies leading to an increased revenue through customer satisfaction
- Long distances travel and its factors, Customer shop selection prediction, Customer segments, Top 100 profitable customers
- Formulation of different RQs, Selection of different data source, Extension of analysis to other business sectors can further help to *decode the customer behaviour through analytics leading to a sustainable business growth*

Thank you all for your attention!

Any Questions?

- Dataset. https://bigml.com/user/czuriaga/gallery/dataset/5559c2c6200d5a6570000084
- Datasets. http://www.michelecoscia.com/?pageid=379
- Project Objective. [Image] *The Importance of Customer Insight*. https://www.klbdkosher.org/news-and-articles/the-importance-of-customer-insight/
- Data Source. [Image] *Coop Logo*. https://en.wikipedia.org/wiki/Coop(Italy)