



Customer Segmentation

Brazilian E-commerce Public Dataset by Olist

With R Programming
Maulana Yusuf

[Source data](#)

Summary

This dataset was generously provided by Olist, the largest department store in Brazilian marketplaces. This dataset total has 100k Rows and 54 Column.

RFM segmentation is one method for segmenting customers by measuring the level of recency, frequency and monetary

- The result of this project is :
- Getting information about monthly GMV
 - Gaining insight about potential sellers
 - Customer segments
 - Category Product Recommendation



DATA UNDERSTANDING

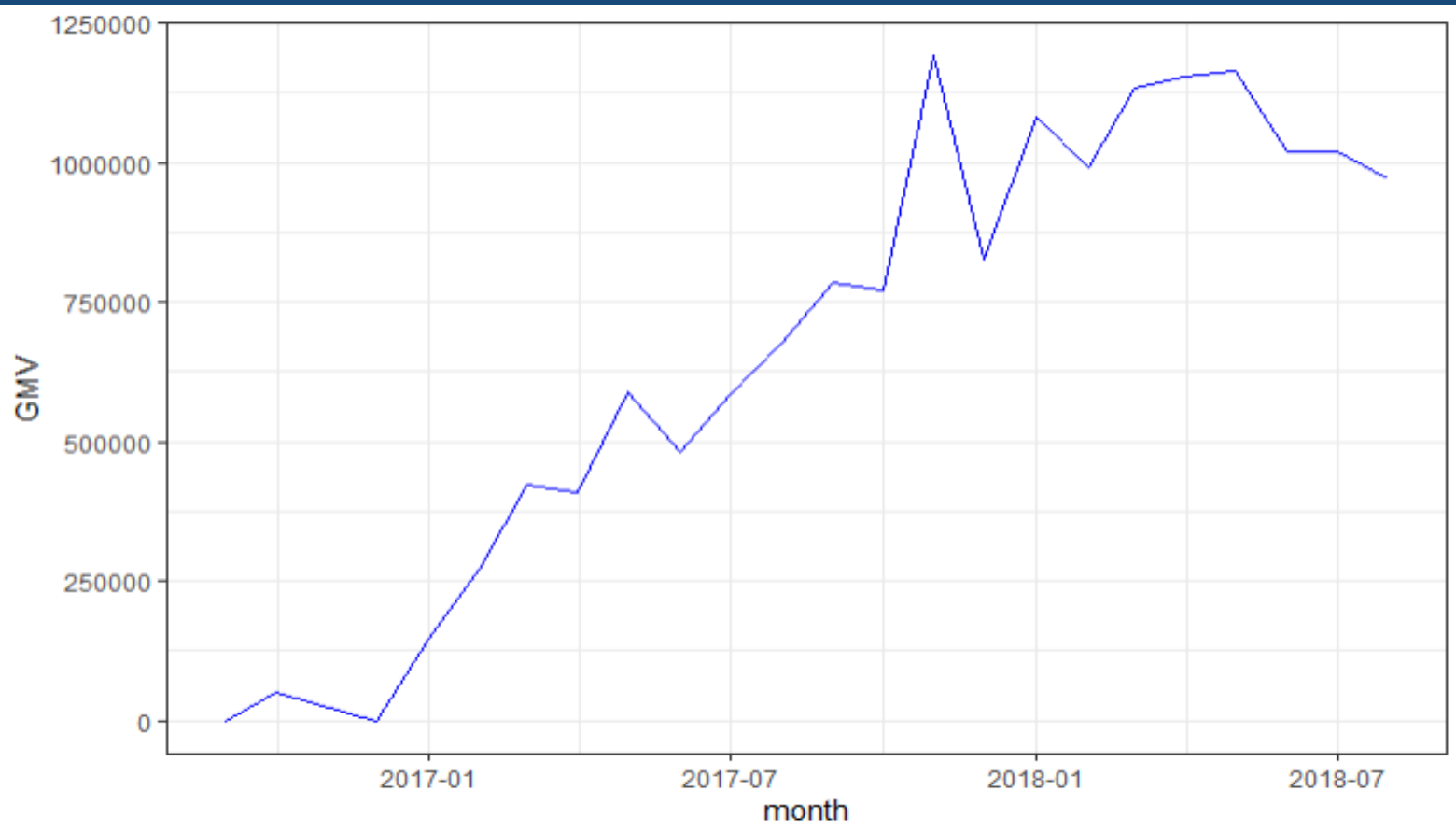
- The dataset has information of 100k orders (100k rows) from 2016 to 2018 made at multiple marketplaces in Brazil. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers. It also released a geolocation dataset that relates Brazilian zip codes to lat/lng coordinates (total has 54 column)

Column:

- | | |
|-----------------------------|------------------------------------|
| 1. Customer_unique_id | : unique identifier of a customer. |
| 2. Order_item_id | : number of items purchased |
| 3. Seller_id | : seller unique identifier |
| 4. Price | : item price |
| 5. Order_purchase_timestamp | : Shows the purchase timestamp. |



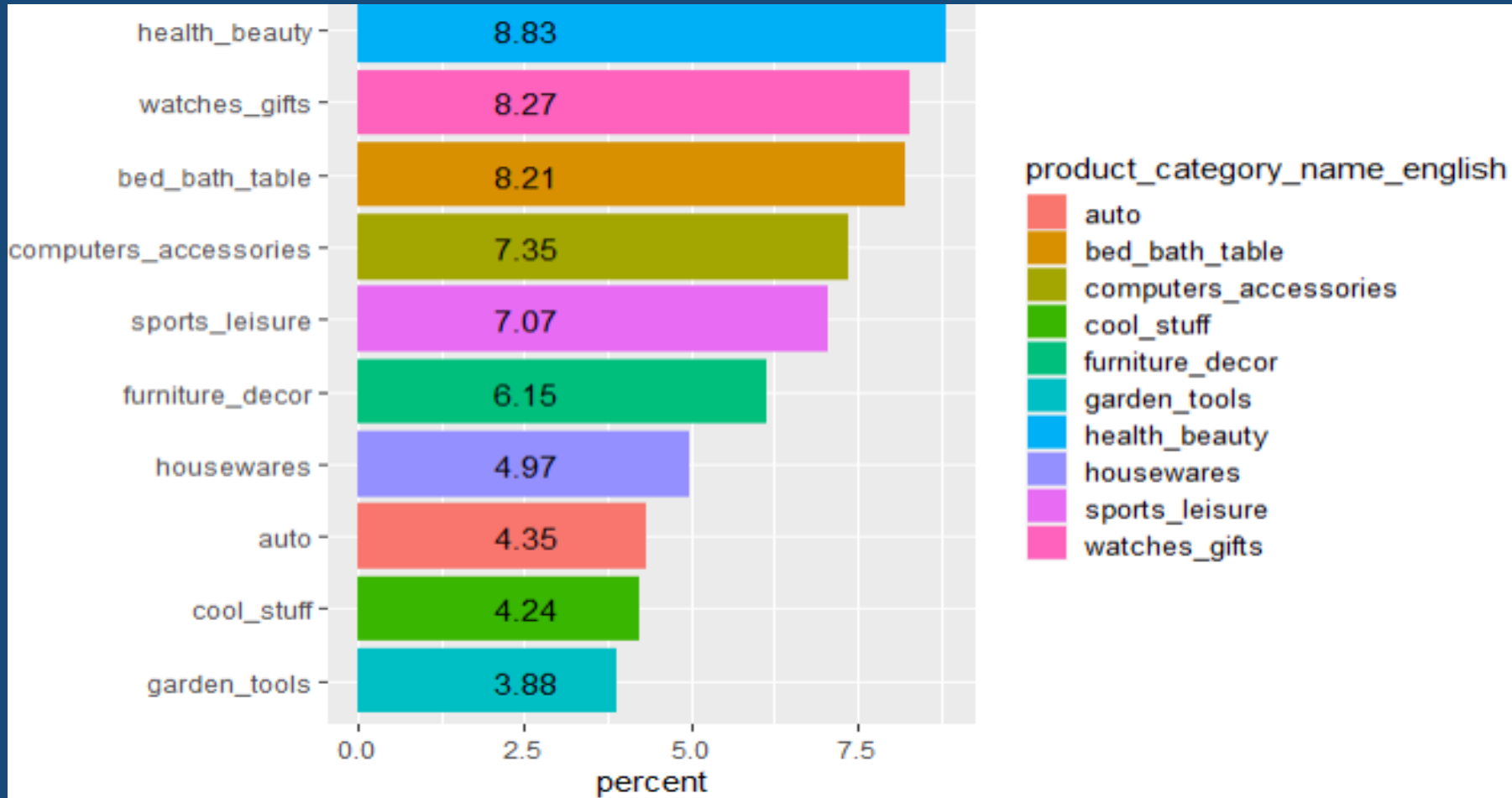
EXPLORATORY DATA ANALYSIS



- GMV increased rapidly from 01-01-2017
- Had a peak in sales in November 2017, up about 54% from the previous month. which indicates that the strategy carried out in that month or the previous month has proven effective



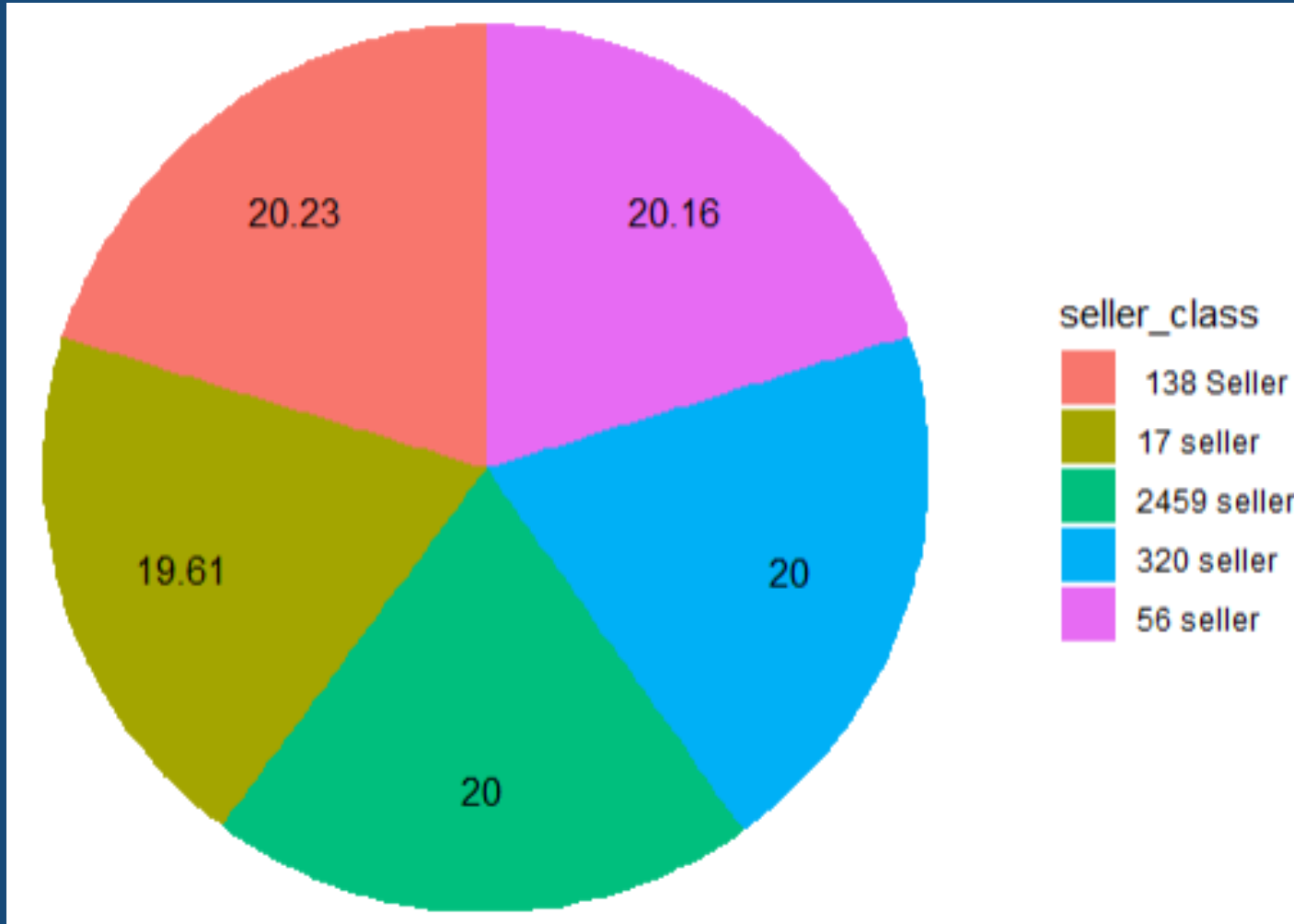
Gross Merchandise Value By Product Category



out of a total of 71 categories, these top 10 categories contribute about 63.37% GMV



Seller Sales



of a total of 2990 sellers:

#1-17 top sellers give about 20% of total GMV

#1-73 top sellers give around 40% of total GMV

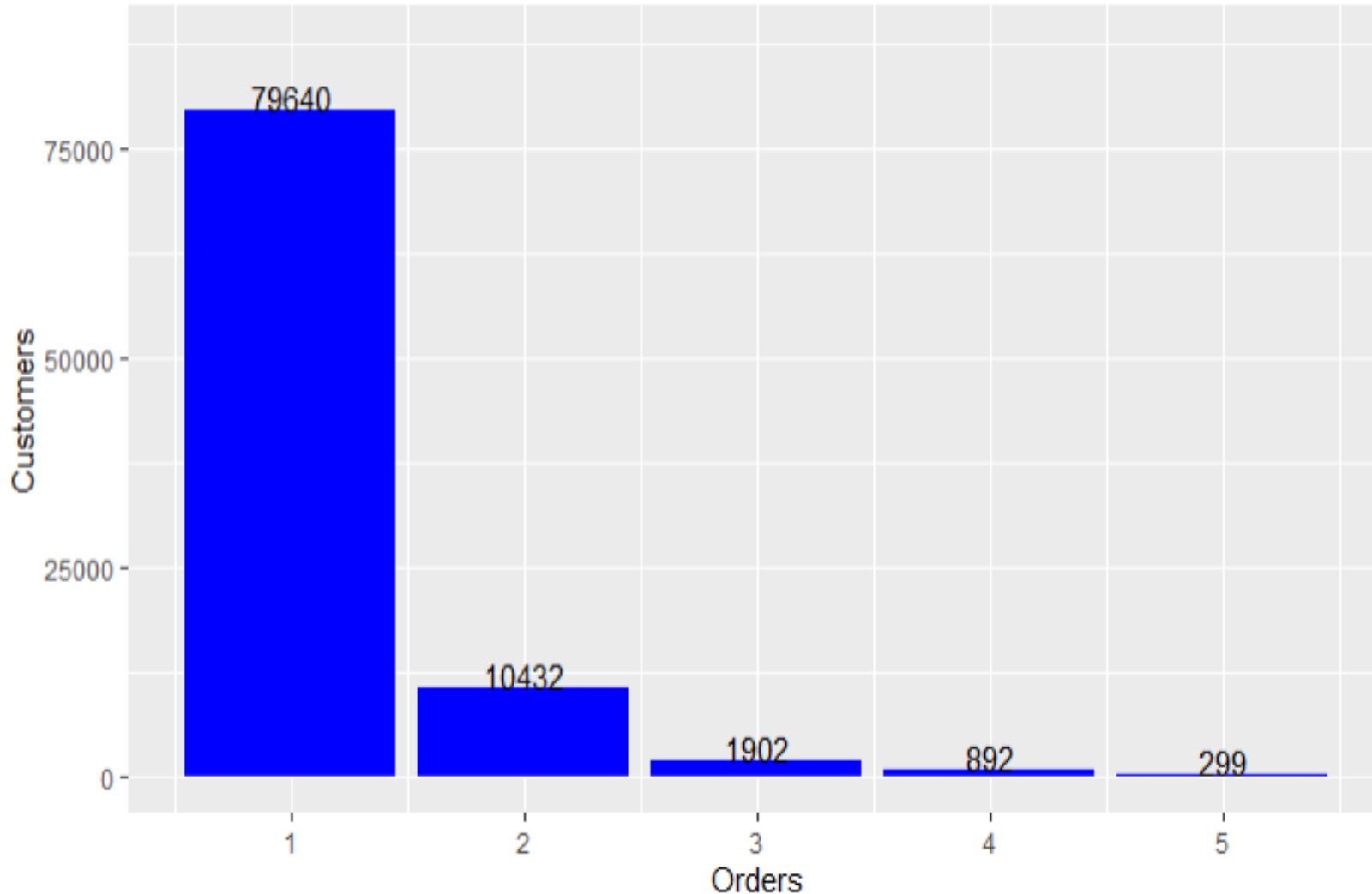
#1-211 top sellers give about 60% of total GMV

#1-531 top sellers give about 80% of total GMV

**20% of sellers earn 80% GMV,
while the other 80% of sellers
only earn 20% GMV**



Customers by Orders



Our main problem this time is that 85% of our customers only make 1x purchases. It indicates that our returning customers or retention rate is very low

From our total 93664 customers:

- 79640 or 85% made 1x orders
- 10432 or 11% customers made 2x orders
- 1902 or 2% customers made 3x orders

Customer Segmentation with RFM Analysis

- **Recency (R)** : the number of days since a customer made the last purchase
- **Frequency (F)** : how often customer bought the product in a given period.
- **Monetary (M)** : is the total amount of money a customer spent in that given period.



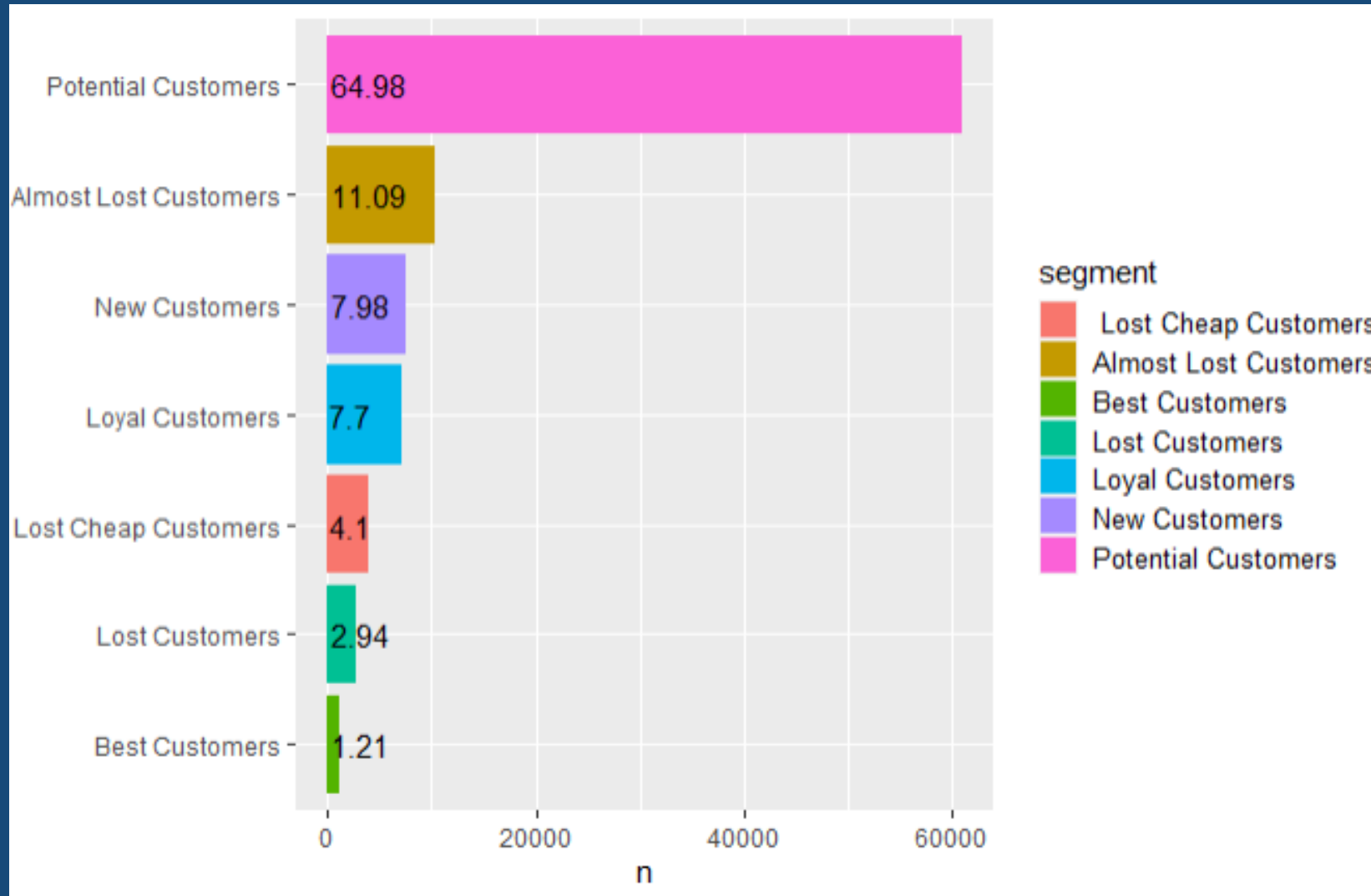
Customer Segmentation with RFM Analysis

- Assign a score criteria from 1 to 5 for each Recency, Frequency, and Monetary
- 5 is highest value and 1 is lowest value
- because the number of customers who make purchases more than 1x is only 15%, the F value is only 1 and 5

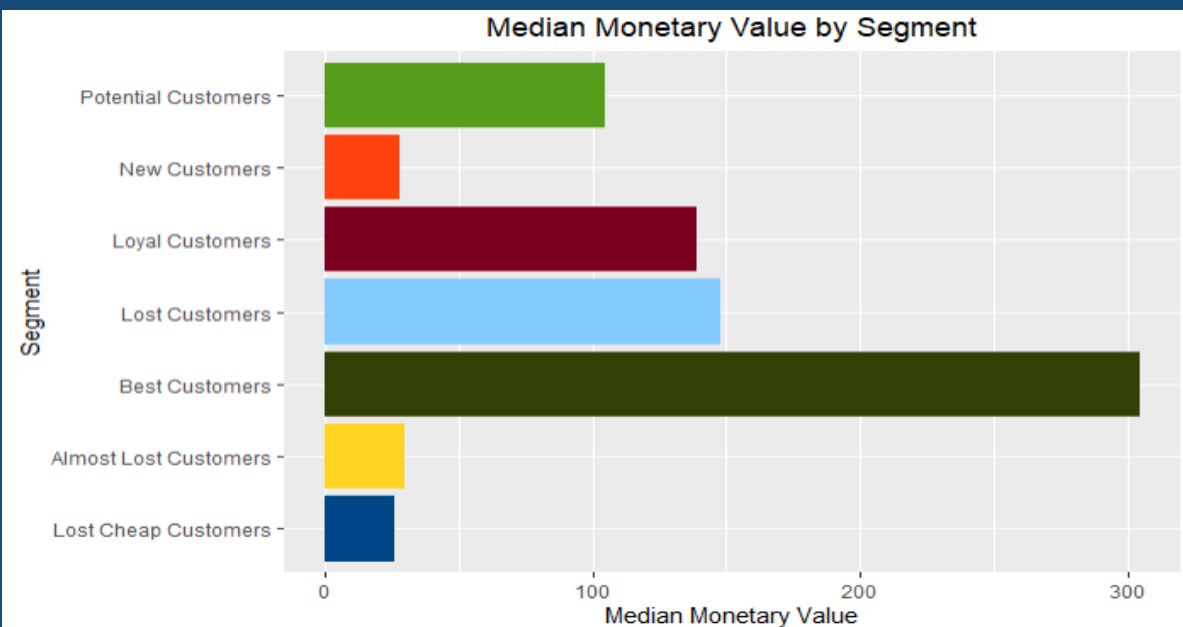
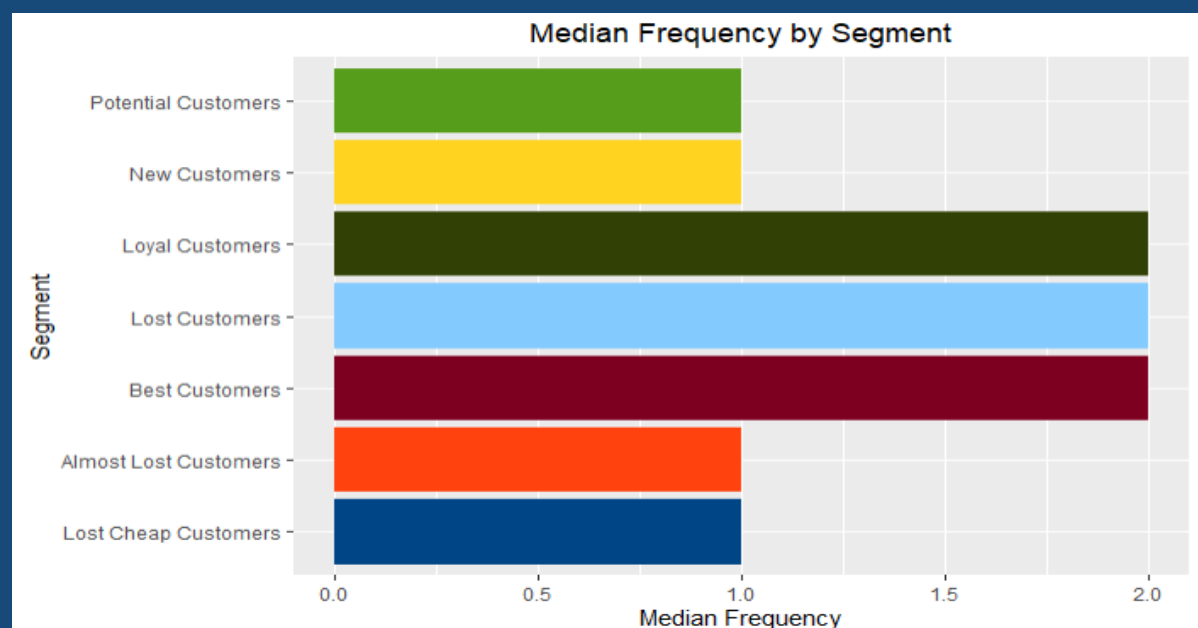
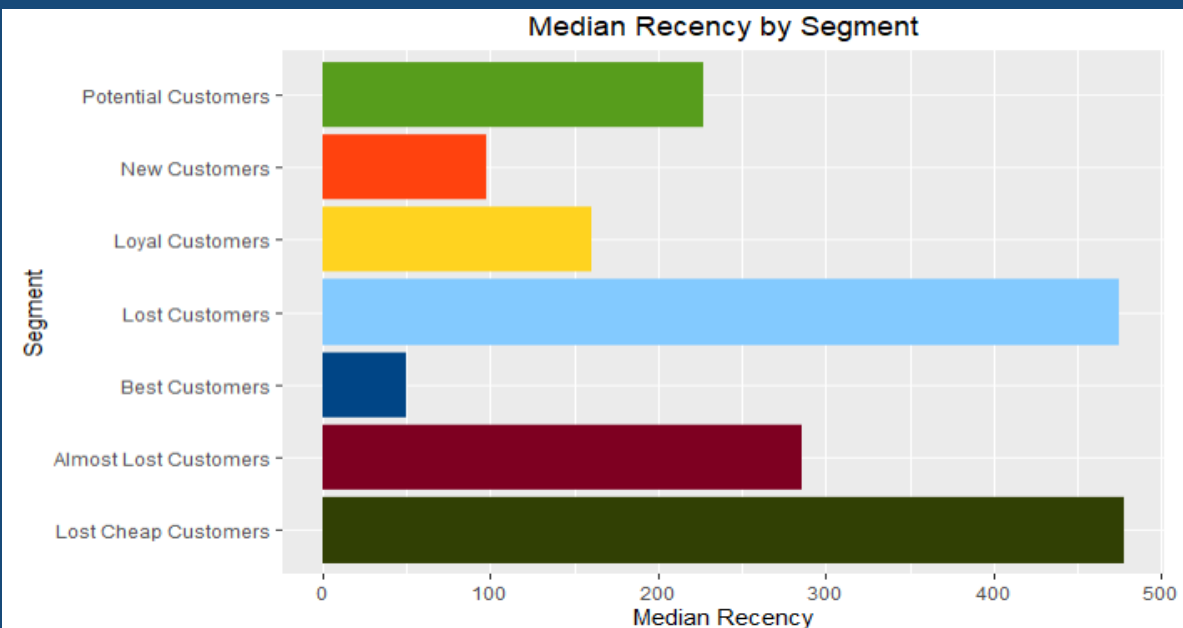
Segment	RFM Score Criteria
Best Customers	R=5 F=5 M=5
Loyal Customers	R= 3-5 and F= 5
Potential Customers	R=1-5 F= 1 M= 2-5
New Customers	R= 4-5 F= 1- 5 M= 1-5
Almost Lost Customers	R=2-3 F=1-5 M=1-5
Lost Customers	R= 1-2 F=5 M=1-5
Lost Cheap Customers	R=1 F=1 M=1



Customer Segmentation

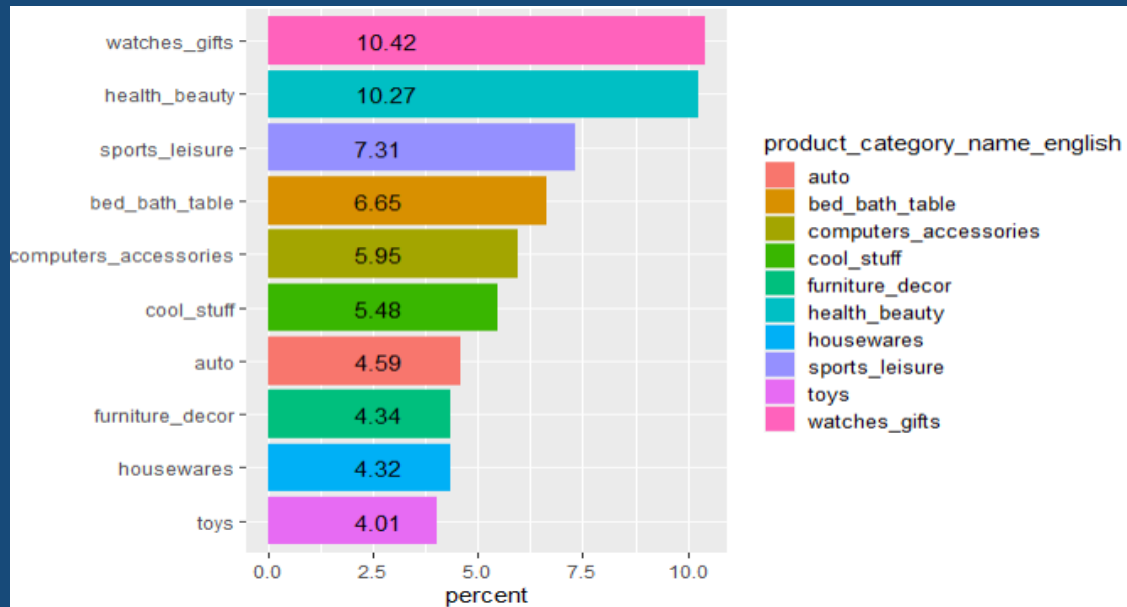


- Potential customers have a very large number, meaning that many customers can still be maximized
- Loyal customers and Best customers are very few in number

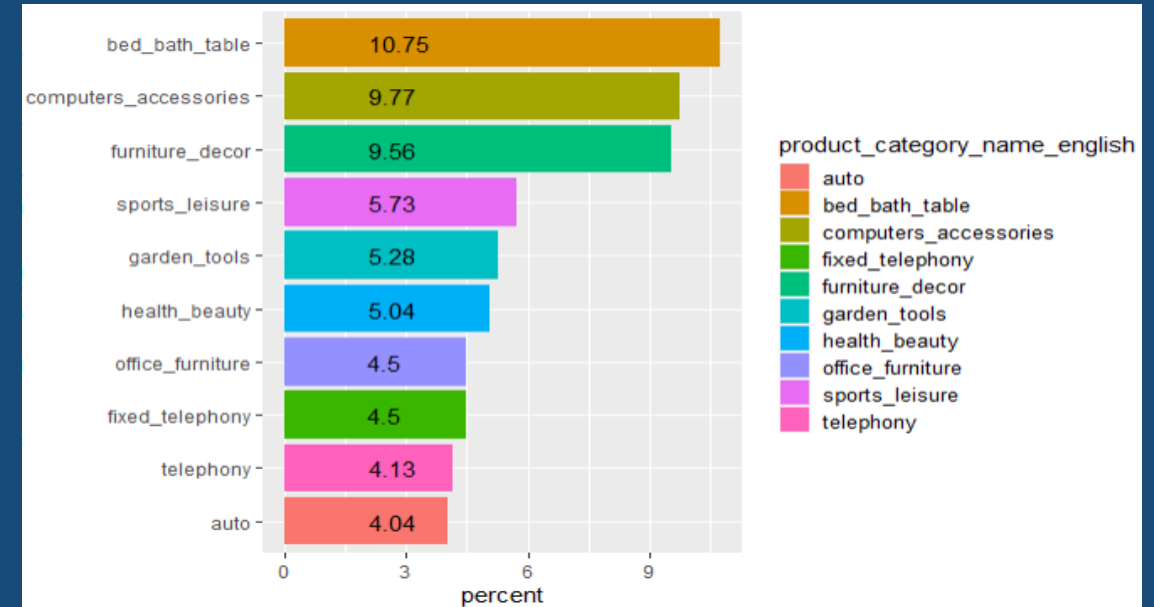


- potential customers have a fairly high median monetary value only from 1x purchase. So indeed, the potential customer segment can be maximized to become the best customer and loyal customer
- lost customers have a fairly high median monetary, but they haven't made a purchase in a long time, so we need to send notifications and give promo so that they become active buyers again

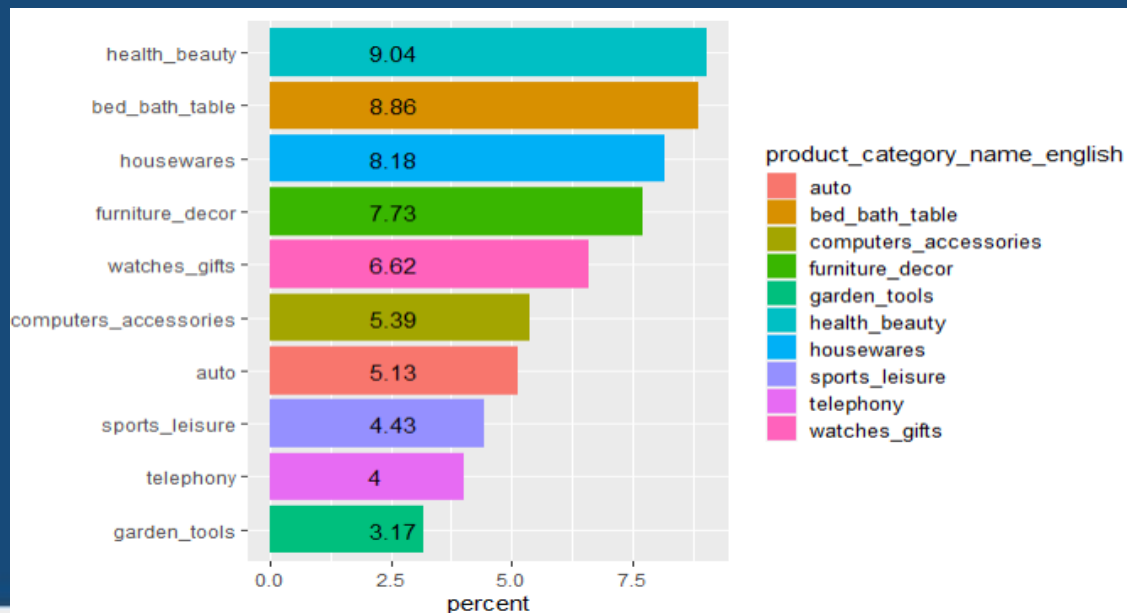
Potential Customer



Almost Lost Customers

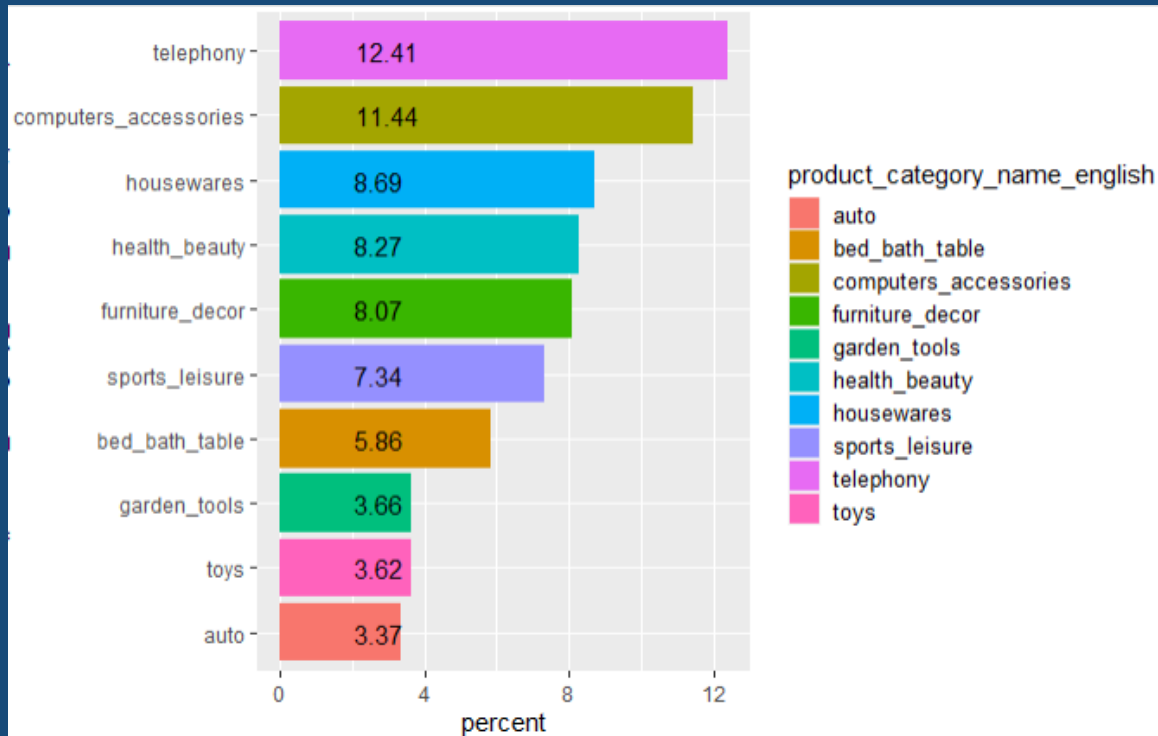


Best Customers

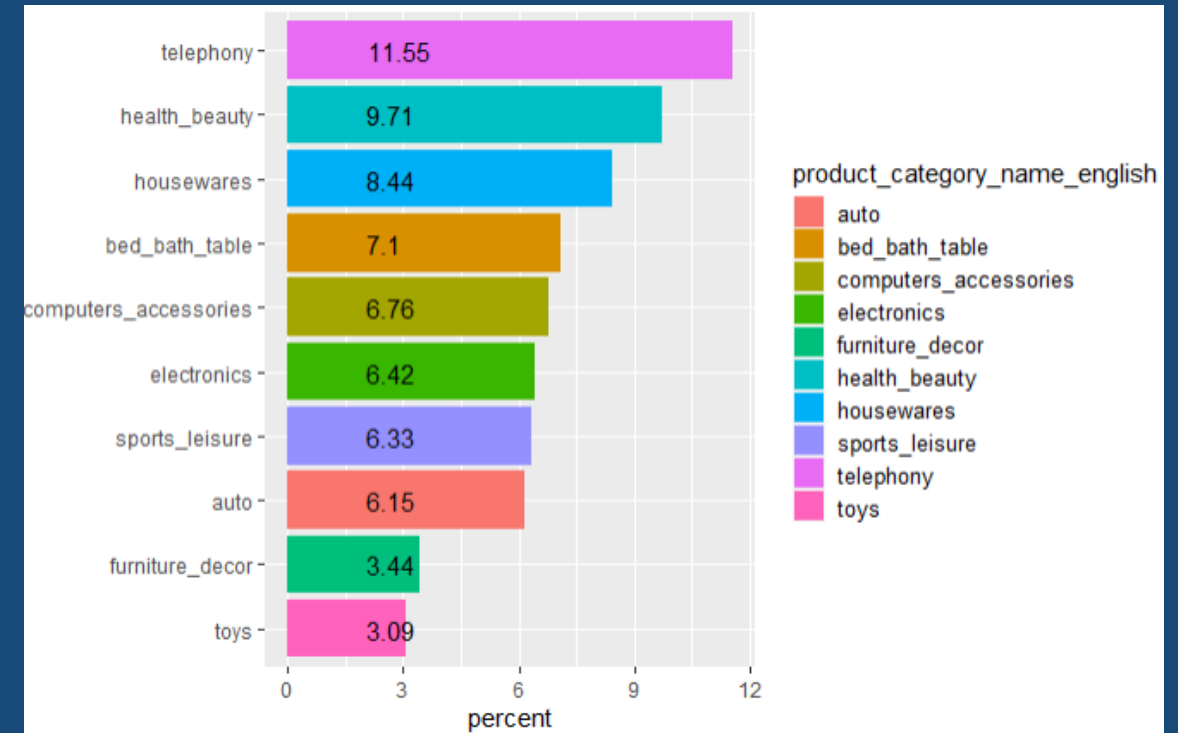


The best_customers segment has top categories: **bed_bath_table, housewares and furniture_decor** which may be used as bundling packages to attract customers

Lost Cheap Customers



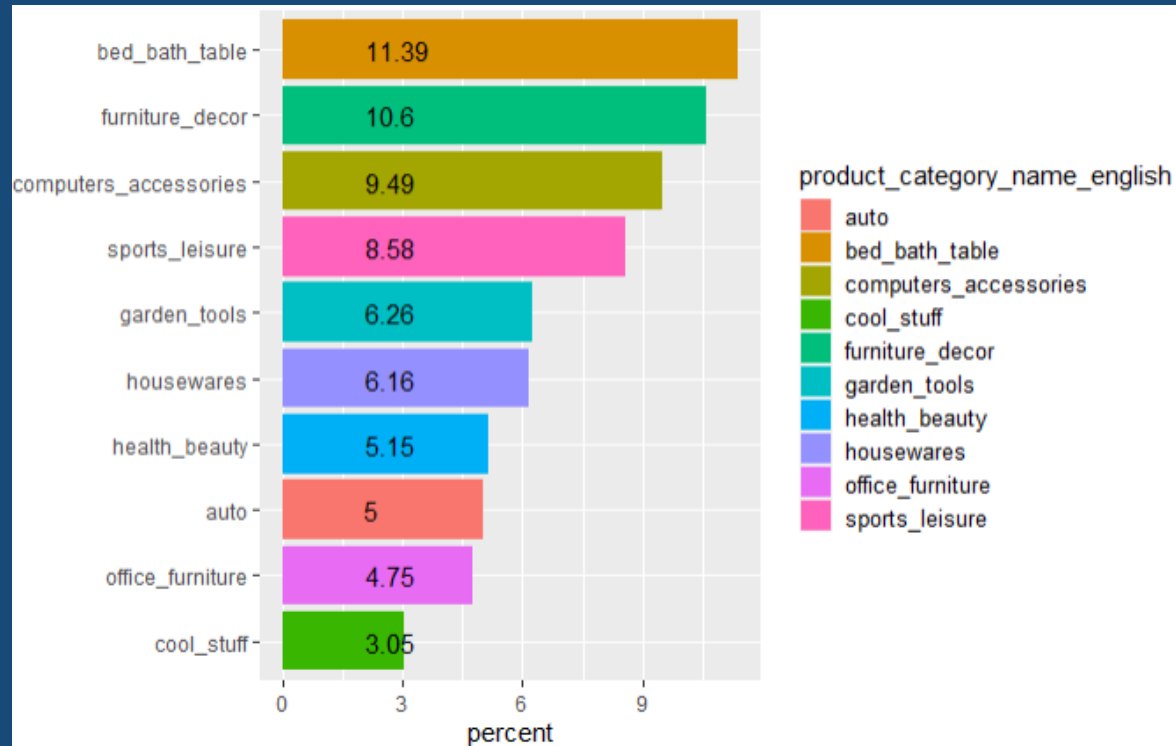
New Customers



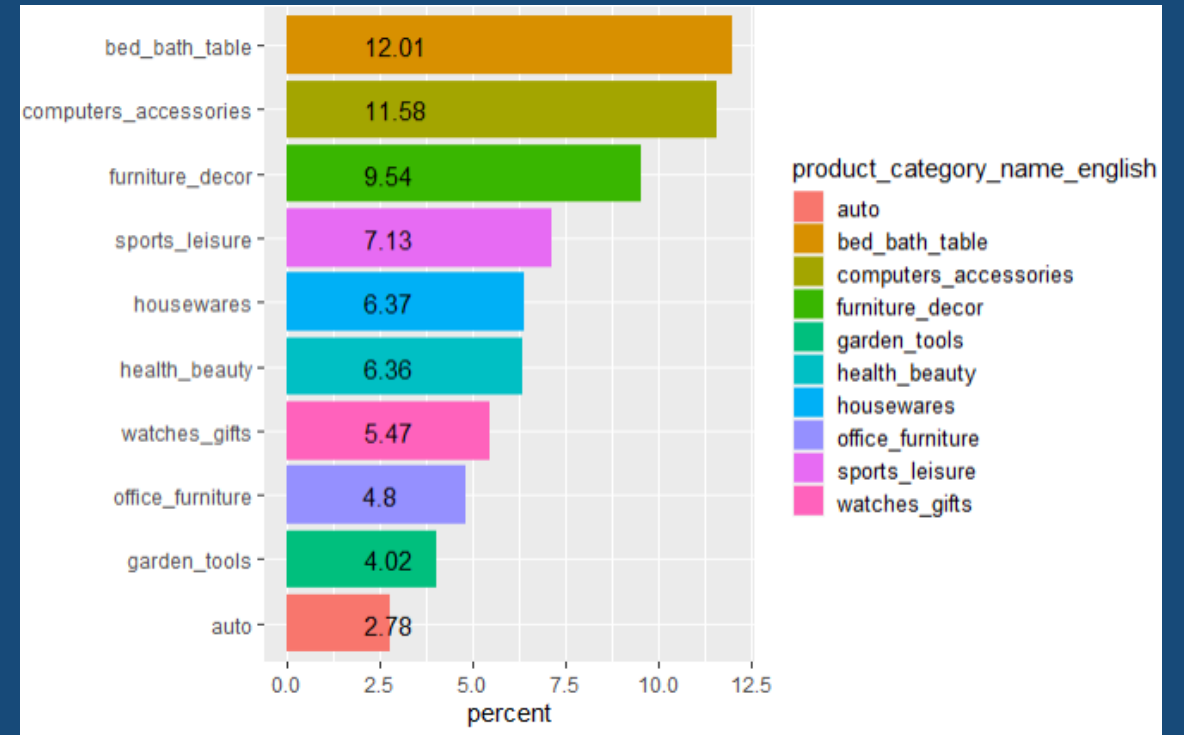
Although telephony category is a lot in lost cheap customers, it is also the best-selling category for new customers. **This indicates that the telephone category is effective for bringing in new customers**



Lost Customers



Loyal Customers



Categories bed_bath_table, furniture_decor, computer_accessories, and sport_leisure are best-seller categories in lost customer segments, but also the best-seller for loyal customers. **This indicates that these categories have the potential to be used as promos for lost customers so that they become active customers again**



RECOMMENDATION

- **Create a loyalty program/membership program** which customers will get rewards for every transaction
- **Send engaging emails to customers**, especially to potential customers
- **Send Notification** to potential, almost lost, and lost customers
- **To get new customers**, we can provide discount promos for the telephony and health_beauty category
- **Create bundling package**: bed_bath_table, housewares_and furniture_decor
- **Give promos**, especially to the top 10 product categories in potential customers
- Give rewards or make competition for sellers to increase their productivity



THANK YOU

