Sprocket Central Pty Ltd

Data analytics approach

Sai Vamsi Krishna Kamatham | The Analytics Team

Introduction:

Identify and Recommend Most Valued Customers:

outline of problem:

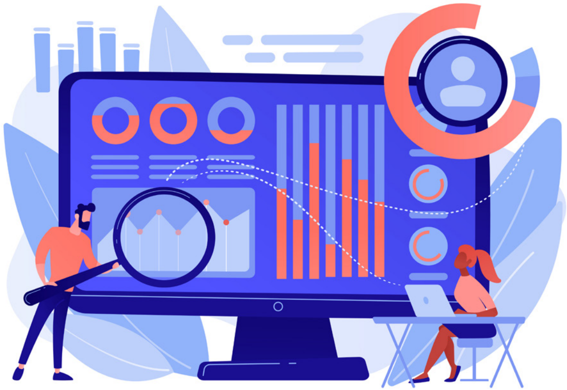
- sprocket central marketing team is looking to boost business sales by analyzing provided datasets.
- module 2: using the 3 datasets provided the aim is to analyze and recommend most valued old customers.
- module 3: is to train a ml model using interpreted data from old customers and test it on new customer data to identify high-value customers for targeted marketing.
- note: module 2 will involve three phases: data exploration, model development, and interpretation. additionally, the methodology for module 3 will be discussed as part of the preparation for module03.

Introduction:

Identify and Recommend Most Valued Customers:

contents of data analysis:

- distribution of orders by age group, gender, state, wealth segment, job industry category.
- past 3-year bike-related purchases by gender.
- car ownership distribution by gender.
- car ownership across different states.
- gender distribution by state.
- wealth segmentation by age group.
- age group vs state vs profit.
- wealth segment vs gender vs profit.
- rfm analysis and customer classification.



Data Exploration:

Data Quality Assessment and 'Clean Up':

key issues for data quality assessment:

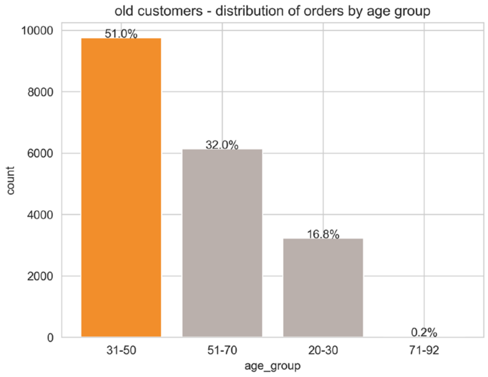
- **accuracy:** correct values
- **completeness:** data fields with values
- **consistency:** values free from contradiction
- **currency:** values up to date
- **relevancy:** data items with value meta-data
- **validity:** data containing allowable values
- **uniqueness:** records that are duplicated
- note: an in-depth analysis has been sent via email
- refer to data quality assessment report doc.

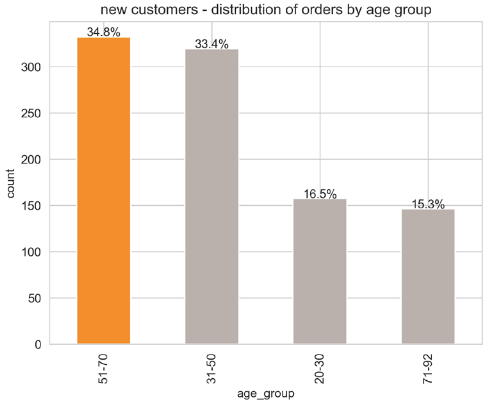
Table name	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer Demographic	DOB inaccurate, Age: missing.	Job title: blanks Customer id: incomplete	Gender: inconsistency	Diseased Customers: Filter out	Default Column: Delete	
Customer Address		Customer id: Incomplete	States: Inconsistency			
Transaction Data	Profit: missing	Customer id: incomplete Online Order: blanks Brand: blanks			Cancelled Status order: Filter out	List Price: Format Product sold date: Format

Data Exploration:

Distribution of Orders by Age Group :

- old: the individuals aged between 31-50 have placed the highest number of orders, accounting for 51% of the total orders. this suggests that the main contributors to the orders placed are individuals in their middle age.
- new: the data reveals that the majority of orders are placed by individuals between the ages of 31-50 and 51-70, together accounting for approximately 70% of the total orders. this indicates that middle-aged and older adults are the primary contributors to the orders placed.
- this insight can be used to inform marketing strategies targeted towards this age group and to tailor product offerings accordingly.

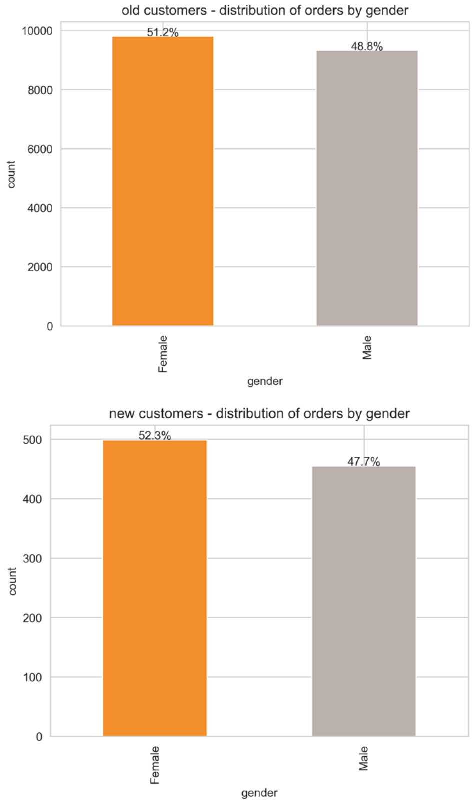




Data Exploration:

Distribution of Orders by Gender :

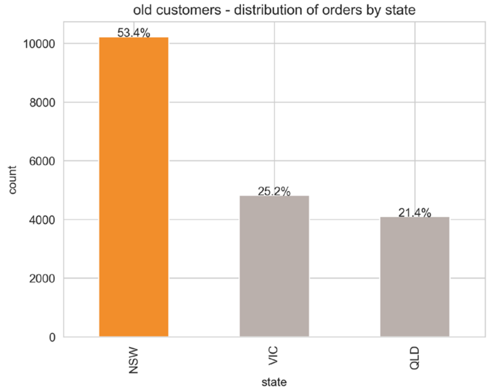
- old: there is a relatively equal distribution of orders between genders, with female customers placing approximately 3% more orders than male customers.
- new: there is a relatively equal distribution of orders between genders, with female customers placing approximately 5% more orders than male customers. this indicates that women make up the majority of the orders compared to men, and there is a slightly higher demand from women compared to men.
- this information can be used to develop marketing strategies and promotions that target women specifically.

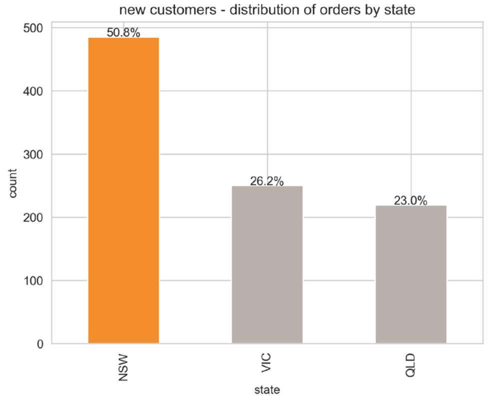


Data Exploration:

Distribution of Orders by State :

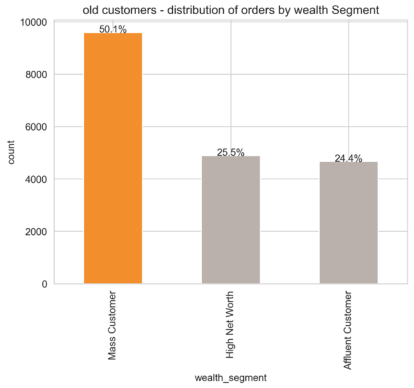
- old: customers from new south wales have placed the highest number of orders, accounting for approximately 53% of the total orders.
- new: customers from new south wales have placed the highest number of orders, accounting for approximately 50% of the total orders. this indicates that the majority of orders are made in the state of new south wales in australia.
- this information can be used to tailor marketing campaigns and promotional activities specifically for customers in new south wales.

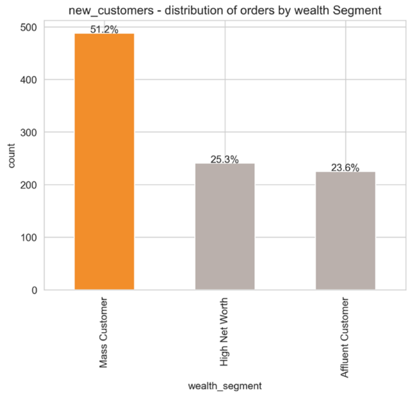




Data Exploration:

Distribution of Orders by Wealth Segment :



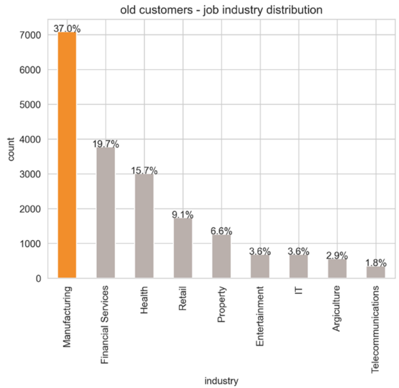


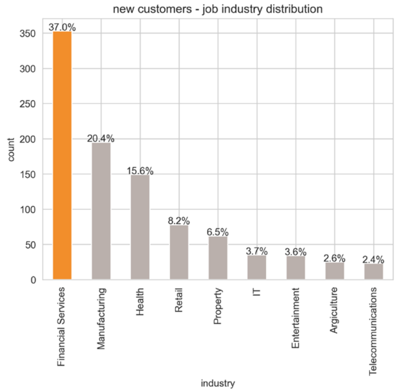
- old: mass customers have placed the highest number of orders, accounting for approximately 50% of the total orders.
- new: mass customers have placed the highest number of orders, accounting for approximately 50% of the total orders. this indicates that the largest segment placing orders is the average income customers (mass customers).
- understanding the behavior and preferences of mass customers can be valuable in creating products and services that cater to their needs and preferences.

Data Exploration:

Distribution of Orders by Job Industry :

- old: customers from the **manufacturing, financial services, and health sectors** have placed the highest number of orders, collectively accounting for approximately 72% of the total orders.
- new: customers from the **financial services, manufacturing, and health sectors** have placed the highest number of orders, collectively accounting for approximately 72% of the total orders. this indicates that most of our customers work in the manufacturing sector, followed by the financial services and health sectors.
- this information can be used to develop targeted marketing and promotional activities to further expand the customer base and drive sales.

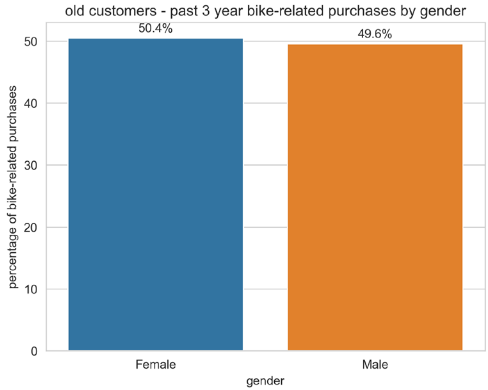


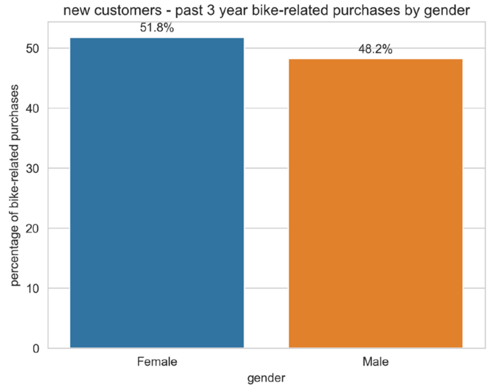


Data Exploration:

Past 3-Year Bike-Related Purchases by Gender :

- old: there is a relatively equal distribution of past 3-year bike-related purchases between genders.
- new: there is a relatively equal distribution of past 3-year bike-related purchases between genders, with female customers purchasing approximately 4% more bikes than male customers. this indicates that women make up a slightly larger proportion of the orders than men, suggesting a slightly higher demand for bikes from women compared to men.
- understanding the differences in purchasing behavior between male and female customers can be useful in designing and marketing products that cater to their specific preferences and needs.

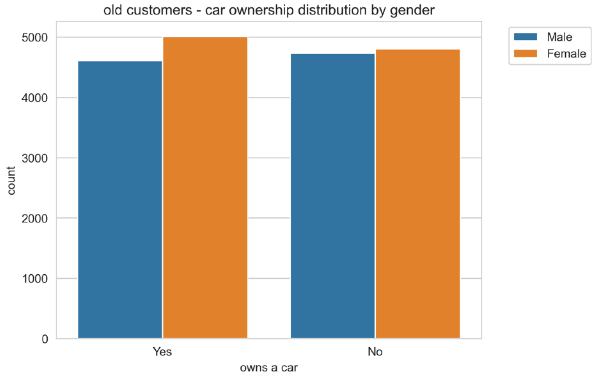


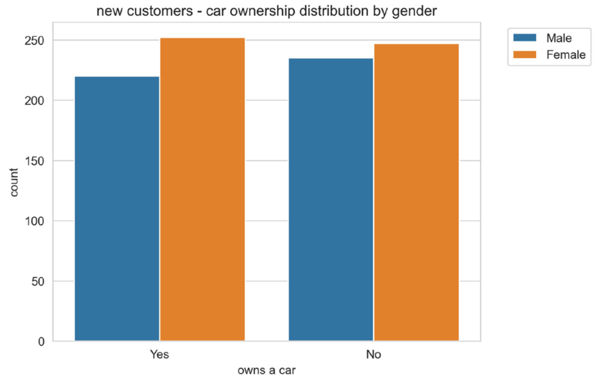


Data Exploration:

Car Ownership Distribution by Gender:

- old: women who own a car tend to make more purchases compared to those who don't own a car.
- new: women who own a car tend to make more purchases compared to those who don't own a car.
- this suggests that having a car may be associated with greater mobility and access to shopping opportunities, which can influence purchase behavior. understanding the factors that influence customer behavior can help businesses develop strategies to attract and retain customers.





Data Exploration:

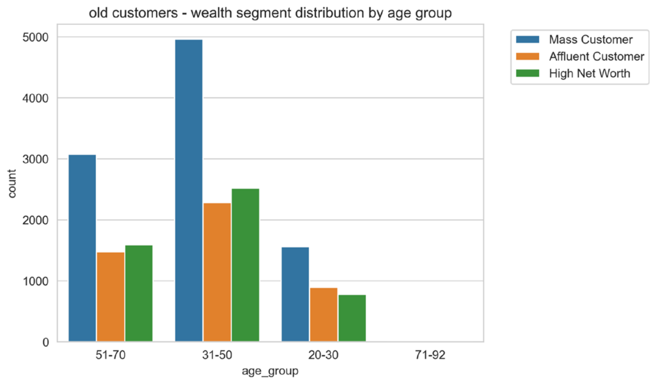
Gender Distribution by State:

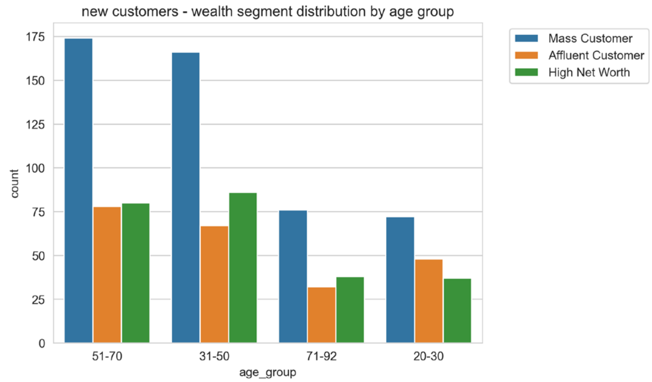
- old: women make more purchases than men in most states.
- new: women make more purchases than men in most states. however, in nsw, both genders have relatively equal buying patterns.
- understanding the differences in purchasing behavior between different demographic groups and geographic regions can help businesses develop targeted marketing strategies and tailor their product offerings to better meet the needs and preferences of their customers.

Data Exploration:

Wealth Segmentation by Age Group:

- old: the largest number of customers in all age categories are classified as 'mass customer'.
- new: the largest number of customers in all age categories are classified as 'mass customer'. this indicates that the majority of our customers fall into the average income category.
- understanding the demographic and economic characteristics of our customer base can help businesses tailor their marketing strategies and product offerings to better meet the needs and preferences of their target audience.





Data Exploration:

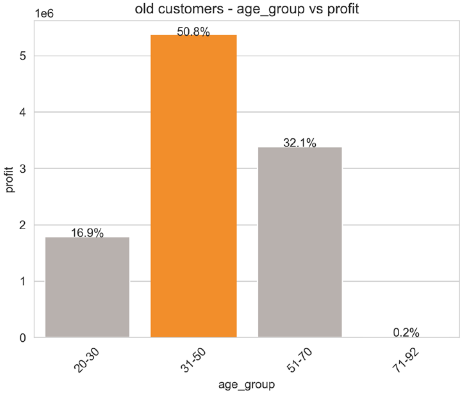
Car Ownership across different States:

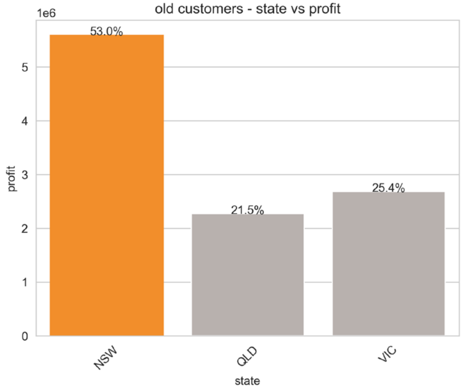
- old: nsw has the highest proportion of people who own a car, and it also appears to have the largest sample size as well.
- new: nsw has the highest proportion of people who do not own a car, and it also appears to have the largest sample size as well.
- this suggests that there is a wide range of transportation patterns among our customers in nsw, and businesses may need to tailor their marketing strategies and product offerings accordingly to meet the needs of different customer segments.

Data Exploration:

Age Group vs State vs Profit:

- old: individuals aged between 31-50 and 51-70 have contributed the highest number of profit, accounting for approximately 70% of the total profit. this suggests that middle-aged and older adults are the primary contributors to the profit gained.
- this indicates that the company should focus on targeting middle-aged and older adults to continue generating high levels of profit.
- old: customers from new south wales have contributed the highest number of profit, accounting for approximately 50% of the total profit. this indicates that a significant portion of the profit comes from the state of nsw in australia.
- this indicates that the company should focus on maximizing sales and profit in nsw, which appears to be a high potential market.

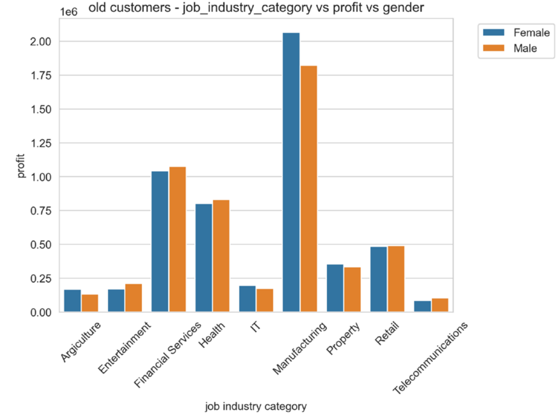




Data Exploration:

Wealth Segment vs Gender vs Profit:

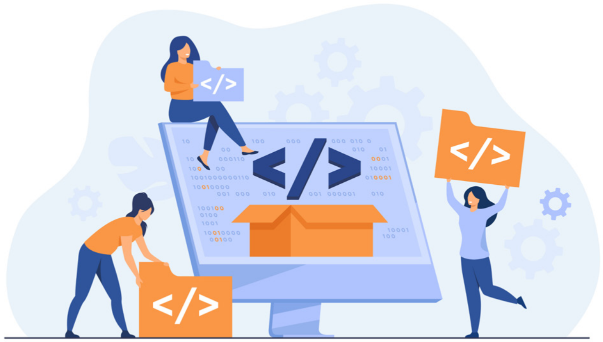
- old: majority of the profit has been generated by mass customers, with female customers contributing slightly more to the total profit than male customers.
- old: female customers in the manufacturing industry generated the highest profit compared to other industries.
- this suggests that marketing strategy that targets female mass customers would lead to more profits. additionally, investing in the manufacturing industry to better serve this customer segment could lead to increased profits



Data Exploration:

RFM Analysis and Customer Segmentation - Old Customers:

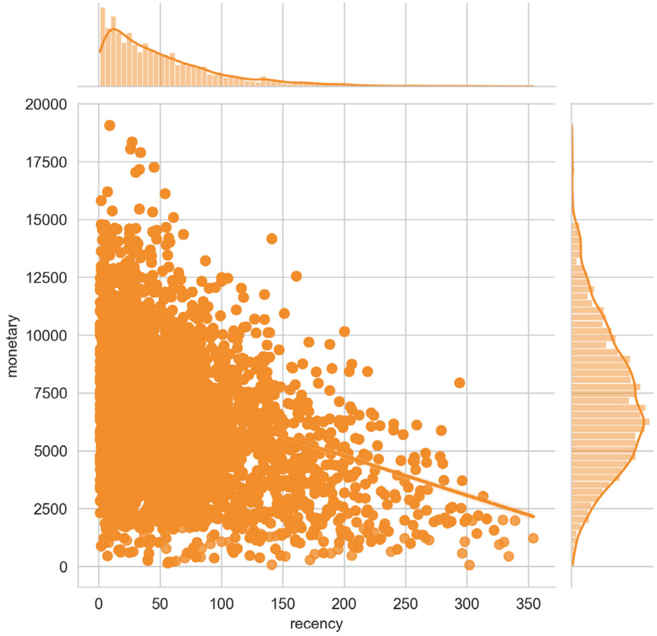
- rfm analysis is used to determine **which customers a business should target to increase its revenue and value**.
- the rfm (recency, frequency, and monetary) model shows customers that have displayed high level of engagement with the business in the three categories mentioned.
- **recency** refers to how recently a customer has made a purchase, **frequency** refers to how often they make purchases, and **monetary** refers to how much revenue they generate for the company.



Data Exploration:

Scatter-Plot based off RFM Analysis:

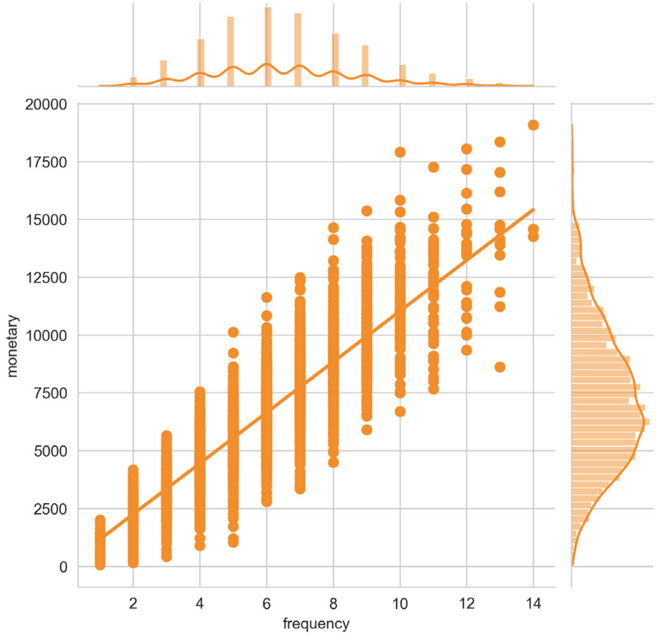
- the scatter plot illustrates that customers who made purchases more recently have generated higher revenue compared to those who visited a while ago.
- customers who made purchases within the past 50-100 days have generated a moderate amount of revenue.
- on the other hand, customers who visited more than 200 days ago have generated relatively low revenue.
- the plot reveals a clear and strong negative correlation between recency and monetary value, indicating that customers who visit more recently tend to spend more than those who visited a while ago.



Data Exploration:

Scatter-Plot based off RFM Analysis:

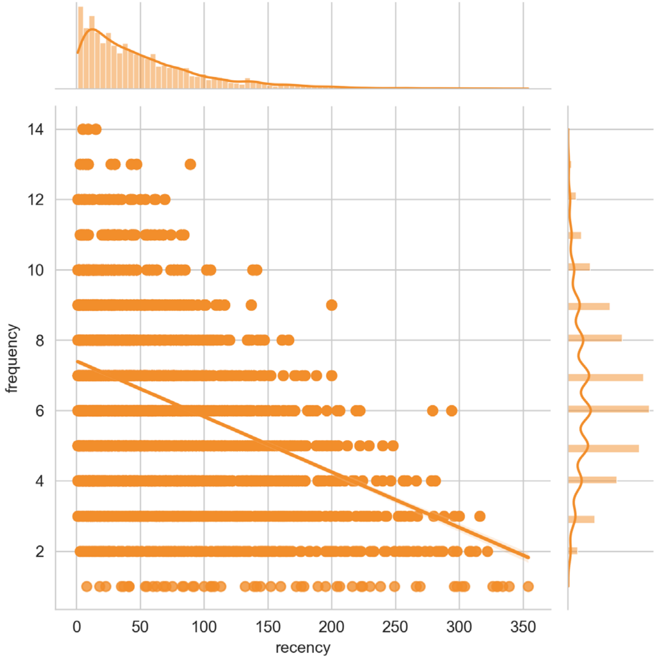
- the scatter plot highlights that **customers classified as "platinum customer," "very loyal," and "becoming loyal" tend to visit frequently**, which is correlated with increased revenue for the business.
- this suggests that customers who visit more often tend to spend more money.
- additionally, the scatter plot demonstrates a **clear positive relationship** between frequency and monetary gain for the business. customers who visit more frequently tend to generate higher revenue for the business.



Data Exploration:

Scatter-Plot based off RFM Analysis:

- the data suggests that customers with very low visit frequency (0-2) tend to have high recency values, meaning they visited the business more than 250 days ago.
- on the other hand, customers who have visited more recently (0-50 days) have a higher chance of being frequent customers (6 or more visits).
- furthermore, the data shows a negative relationship between frequency and recency values, indicating that very recent customers are more likely to be frequent customers.



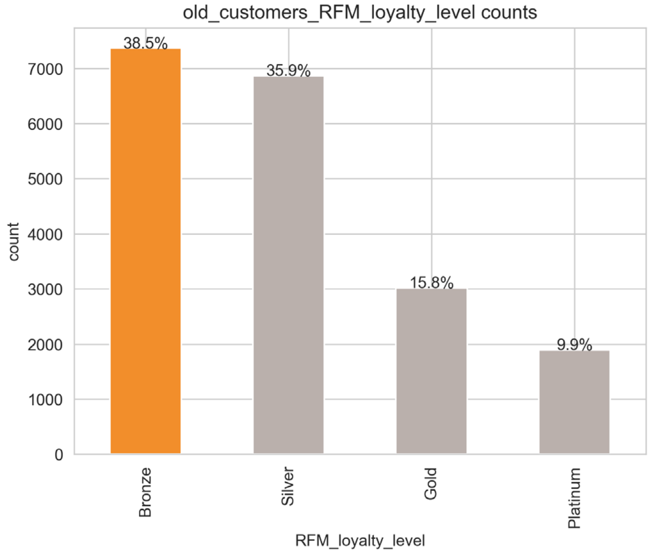
Data Exploration:

Old Customer Title Definition list with RFM values assigned:

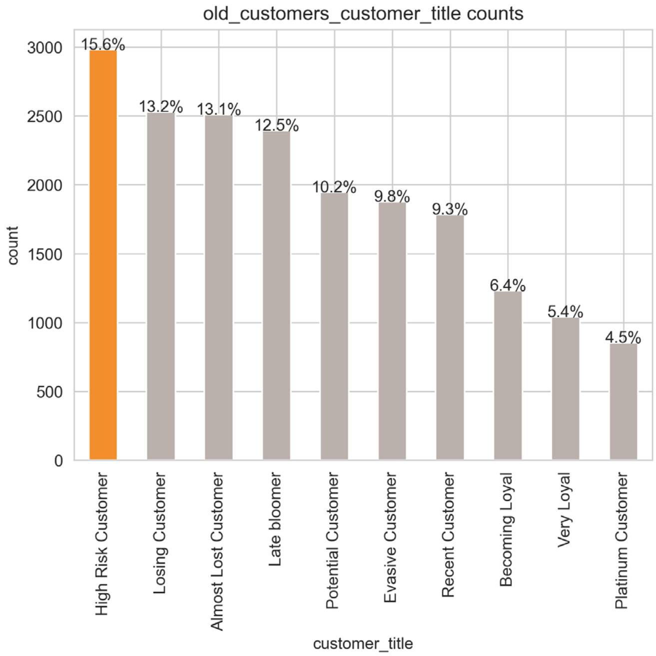
Rank	Customer Title	Description	RFM Value
1	Platinum Customer	Most recent buy, buys often, most spent	444
2	Very Loyal	Most recent, buys often, spends large amount of money	433
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	421
4	Recent Customer	Bought recently, not very often, average money spent	344
5	Potential Customer	Bought recently, never bought before, spent small amount	323
6	Late bloomer	No purchases recently, but RFM value is larger than average	311
7	Losing Customer	Purchases was a while ago, below average RFM value	224
8	High Risk Customer	Purchase was long time ago, frequency is quite high, amount spent is high	212
9	Almost Lost Customer	Very low recency, low frequency, but high amount spent	124
10	Evasive Customer	Very low recency, Very low frequency, small amount spent	112
11	Last customer	Very low RFM	111

Data Exploration:

Old Customer Title Distribution:



old customers RFM loyalty levels



Interpretation:

Old Customer Title Definition list with RFM values assigned:

Rank	Customer Title	Description	RFM Value	Cumulative
1	Platinum Customer	Most recent buy, buys often, most spent	444	178
2	Very Loyal	Most recent, buys often, spends large amount of money	433	360
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	421	704
4	Recent Customer	Bought recently, not very often, average money spent	344	1072
5	Potential Customer	Bought recently, never bought before, spent small amount	323	1427
6	Late bloomer	No purchases recently, but RFM value is larger than average	311	1760
7	Losing Customer	Purchases was a while ago, below average RFM value	224	2116
8	High Risk Customer	Purchase was long time ago, frequency is quite high, amount spent is high	212	2478
9	Almost Lost Customer	Very low recency, low frequency, but high amount spent	124	2802
10	Evasive Customer	Very low recency, Very low frequency, small amount spent	112	3203
11	Last customer	Very low RFM	111	3493

Interpretation:

New Customer Target and Methodology:

- in module #03, the client provided an additional dataset called "**new customers list**" comprising 1000 records of customers **who haven't purchased any products**. they need help identifying which customers to target with marketing campaigns based on this new dataset.

- we'll use a **machine learning classification model** trained on the old customer dataset, which includes rfm segmentations, to predict the most probable segment for each new customer. this approach will guide us in making informed decisions on which marketing campaigns to focus on.

- the model can help us predict which customers are likely to make a purchase in the future, which ones are at risk of leaving, and which ones we should focus on for retention efforts. this will ultimately lead to increased revenue and better customer satisfaction.

Appendix