TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

IBIKUNLE PONMILE MUIDEEN

Agenda

1. Introduction
2. Data Exploration
3. Model Development
4. Interpretation

Introduction

Identify and Recommend Top Customer to Target from 1000 new customer Dataset

- Sprocket Central Pty Ltd is a medium-size bikes & cycling accessories organization.

- The marketing team want to boost the sales by targeted marketing to new customers.

- Using the 3 datasets provided, analyze and recommend high new value customers (at least 200) the Marketing team should target to achieve this.

Data Exploration

Data Quality Assessment, Cleaning and Mitigation

Summary

	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer Demography	DOB inaccurate	Blanks in Job title, Customer id Incomplete	Gender Inconsistency Inconsistent data type in DOB		Default Column is irrelevant	
Customer Address		Customer id incomplete	States Inconsistency			
Transactions	Profit not available		Product id inconsistency Inconsistent data type in list_price, standard cost, transcation date.			

Data Exploration

- **Various columns, such as the brand of a purchase, or job title, have empty values in certain records**

Mitigation: For transactions, less than 1% of transactions (totalling less than 0.1% of revenue) have missing fields. These records have been removed from the dataset. Job titles and DOB were changed to N/A.

- **Inconsistent values for the same attribute (e.g. Male being represented as "M" and "Male", Vicoria as "VIC" and "Victoria")**

Mitigation: Use regular expression to replaced extended values into abbreviations to ensure consistency across addresses.

- **Gender records having "U"**

Mitigation: The 'U' have been replaced based on the distribution from the dataset.

Data Exploration

- **Additional customer_ids in the 'Transactions table' and 'Customer Address table' but not in 'Customer Demographic'**

Mitigation: Only customers in the Customer Demography list will be used for the model.

- **Product_ids having ''0" in the transaction table**

Mitigation: Brand, Product specifications and list price were used to assign them to their appropriate product ids

- **Gender records having "U"**

Mitigation: The 'U' have been replaced based on the distribution from the dataset.

- **Inconsistent data type for the same attribute (e.g. numeric values for some fields and strings for others)**

Mitigation: Convert selected records in characters to numeric. Remove non-numeric characters from string.

Job Industry Distribution

▶ For new customer", Financial services at 24.31%, Manufacturing at 23.83% and Health at 18.3% accounted for about 66% of the customers.

▶ Telecommunications and Agriculture at 2.99% and 3.11% respectively have lowest numbers.

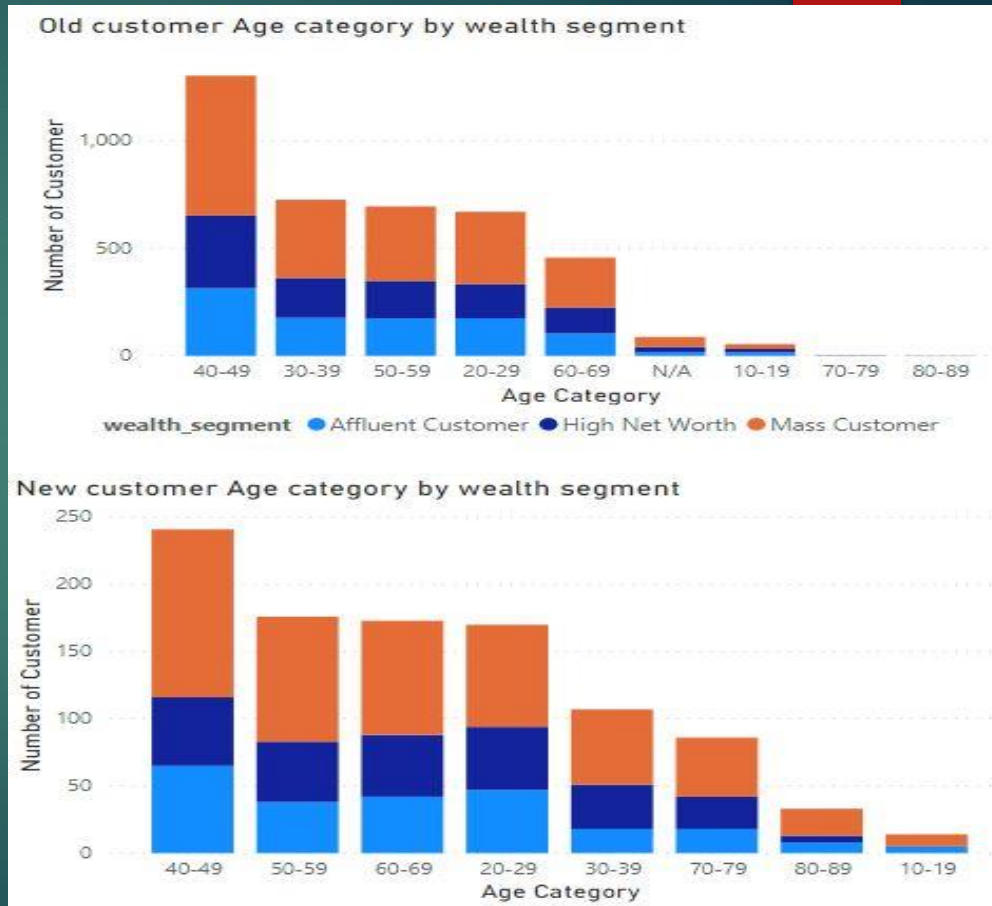▶ Similar pattern is observed in "Old Customers" with the three highest industry accounting for 65% of the customers .
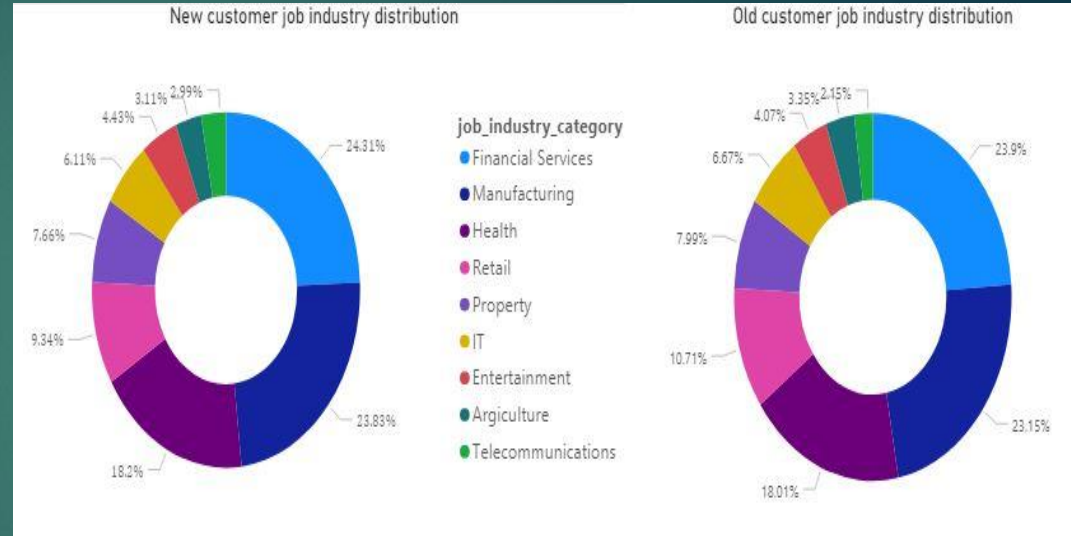
Age Distribution

▶ Most of the customers are aged between 40-49 in New customers as well as in Old customers.

▶ The highest age group for both is 80-89 while the lowest is 10-19

▶ The new customer list shows that the majority of the customers ages between 20 and 69.

▶ The old customer suggest 20-59.

Age Distribution of New Customers

Age Distribution of Old Customer
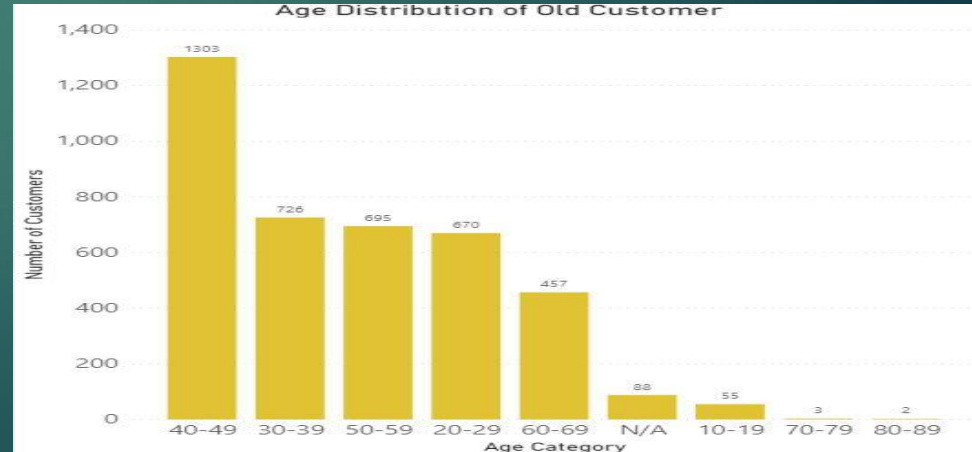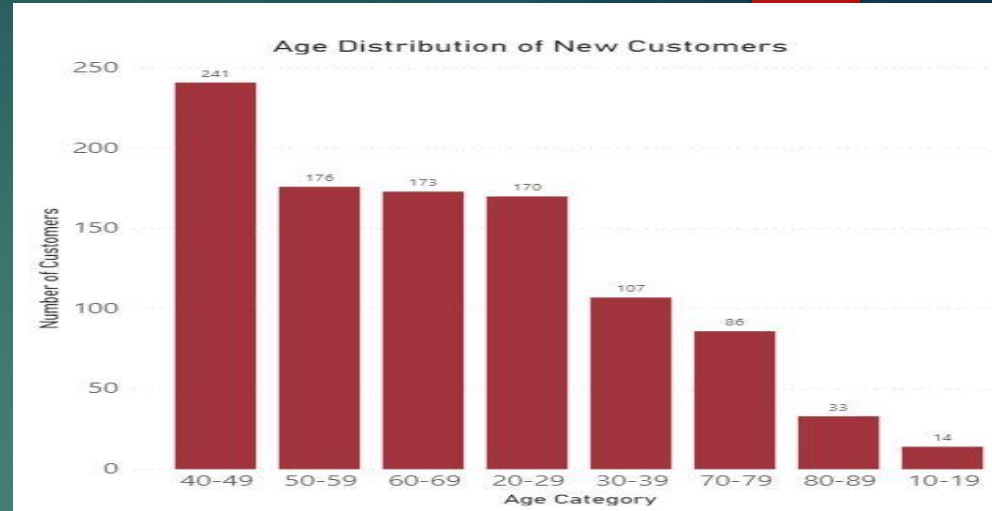
Wealth Segmentation

▶ Mass Customer category dominates every age group in both old and new customers.

▶ The next segment with majority of customer is High net worth.

▶ Affluent customers are the least in all age group except 20-29.

Old customer Age category by wealth segment

wealth_segment ● Affluent Customer ● High Net Worth ● Mass Customer

New customer Age category by wealth segment

Model Development

- RFM analysis is a marketing technique used to determine which customers are the best to target to increase sales

- RFM analysis is a marketing technique used to determine which customers are the best to target to increase sales

- The RFM model (Recency, Frequency and Monetary) is a data driven customer behaviour segmentation technique

Model Development

- The 0th quartile, 1st quartile up to the 4th quartile of Recency, Frequency and Monetary are calculated and assigned a scale 0-4 as appropriate.

- The resulting score is combined to form the RFM values used for creating the model ranking

|	Min of Recency	Frequency	Monetary
MIN	0	1	15.08
Q1	17	4	1841.73
MEDIAN	44	6	2862.33
Q3	86	7	4183.81
MAX	353	14	11668.95

Scatter plot based on RFM Analysis

- The chart shows that the customers who purchased recently (between the past 50 days) generated more revenue .

- Also the customers that visited between 50-100 days generated a good amount of revenue.

- The customers that haven't visited a while gave a very low revenue.


Recency against Monetary
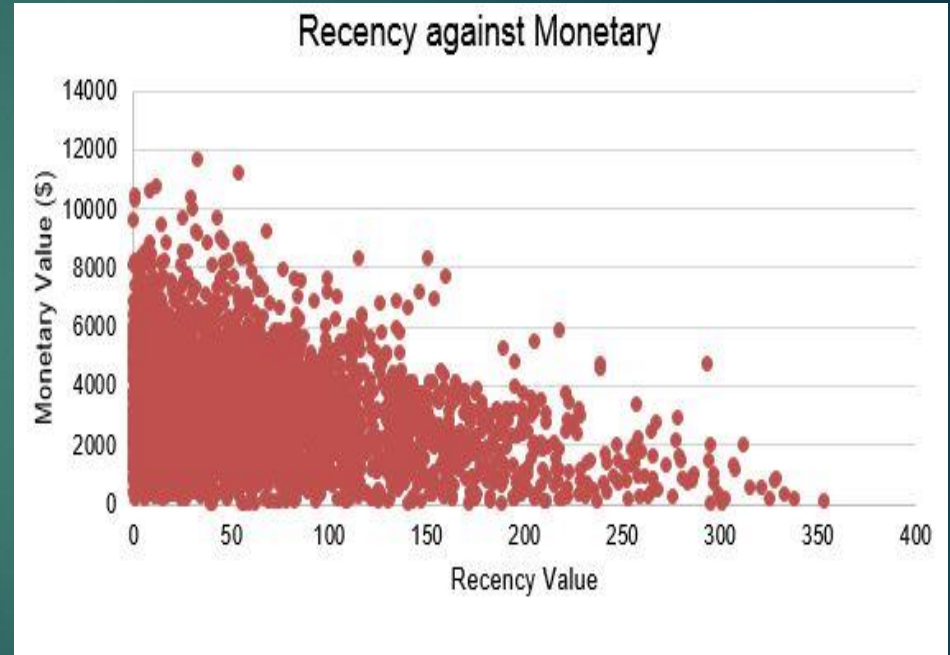
Scatter plot based on RFM Analysis

- The chart shows the frequency in negative relationship with recency.

- The recent customers are the most frequent customers

- Customers that visited between 0-100 days have a higher chance of visiting more frequently

Recency against Frequency
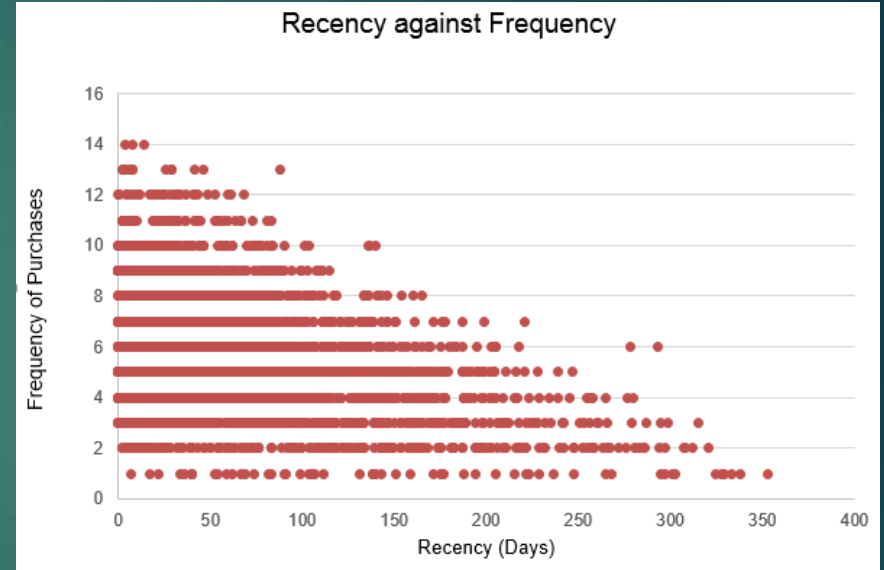
Scatter plot based on RFM Analysis

- There is a direct relationship between frequency and monetary value for the business.

- The customers that frequent the store gives the highest cumulative revenue
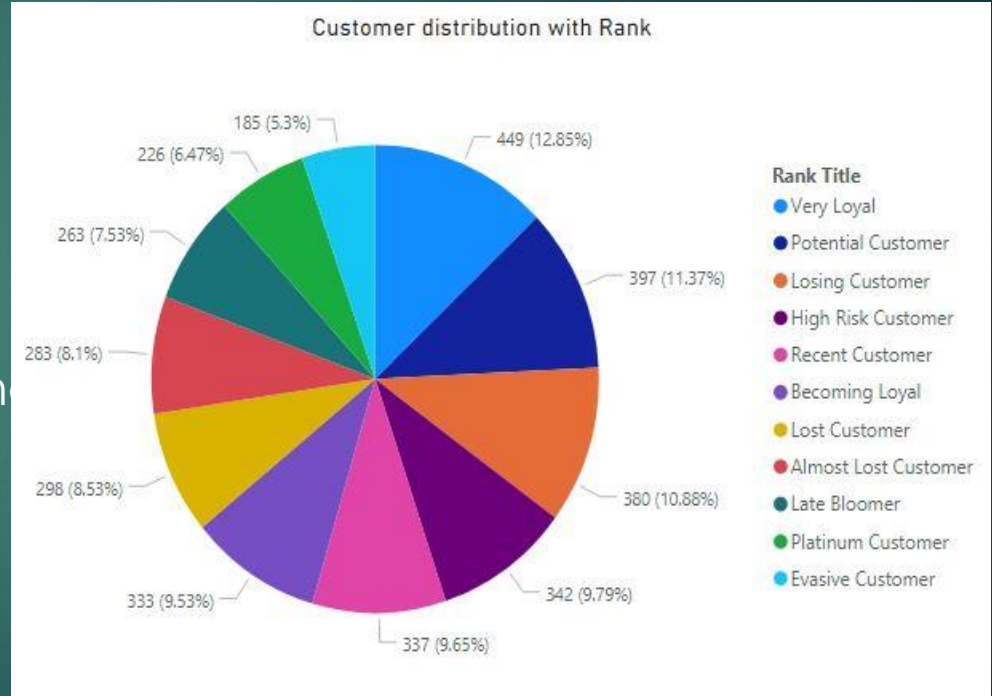

Monetary against Frequency
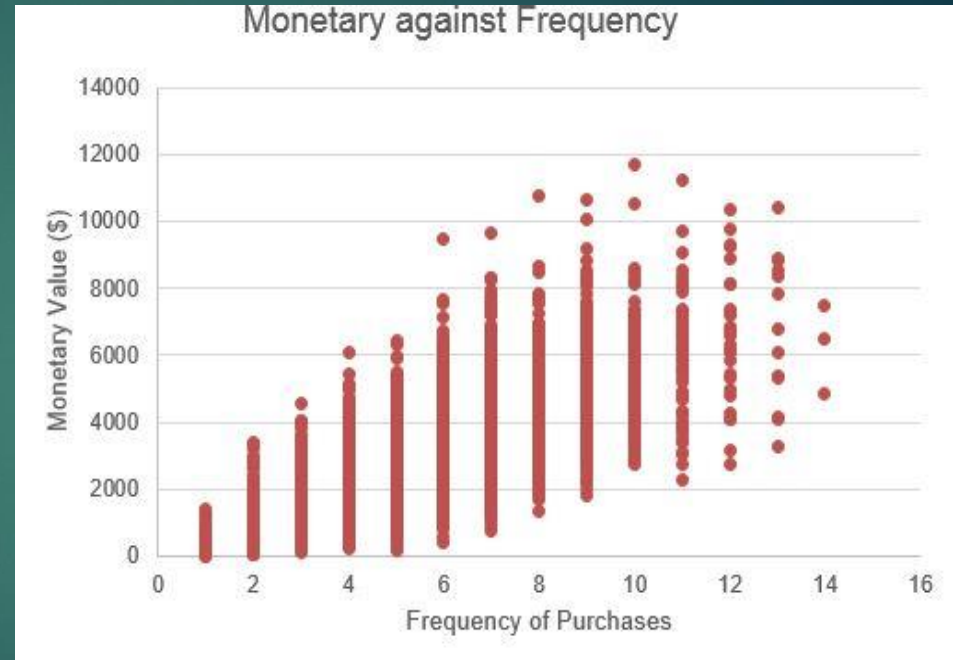
Customer Rank with assigned RFM values

RANK	Customer Rank Title	Description	RFM Value
1	Platinum Customer	Buys recently, buys often, most spent	40
2	Very Loyal	Most recent, buys often, spends large amount of money	37
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	33
4	Recent Customer	Bought recently, not very often, average money spent	30
5	Potential Customer	Recent shoppers, but haven't spent much	27
6	Late Bloomers	No purchases recently but RFM value is large than average	24
7	Losing Customer	Below average RFM, Will lose them if not reactivated	22
8	High Risk Customer	Spent Big money, purchased long time ago, frequency is high	19
9	Almost Lost Customer	Very low recency, Very low frequency, but spent high amount of moiney	16
10	Evasive Customer	Very low recency, Very low frequency, but spent small amount of money	13
11	Lost Customer	Lowest RFM	10

Old Customer rank distribution

- About 13% of Old customers are "very loyal"

- 10.88% are regarded as losing Customers

- The lowest set of customer are the ones with low recency, low frequency
and spend little

Customer distribution with Rank

185 (5.3%)
226 (6.47%)
263 (7.53%)
283 (8.1%)
298 (8.53%)
333 (9.53%)
337 (9.65%)
342 (9.79%)
380 (10.88%)
397 (11.37%)
449 (12.85%)

Rank Title
- Very Loyal
- Potential Customer
- Losing Customer
- High Risk Customer
- Recent Customer
- Becoming Loyal
- Lost Customer
- Almost Lost Customer
- Late Bloomer
- Platinum Customer
- Evasive Customer

Summary Table of New Customer rank distribution

RANK	Customer Rank Title	Description	Number of Customer
1	Platinum Customer	Buys recently,buys often,most spent	229
2	Very Loyal	Most recent, buys often, spends large amount of money	446
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	325
4	Recent Customer	Bought recently, not very often, average money spent	0
5	Potential Customer	Recent shoppers, but haven't spent much	0
6	Late Bloomers	No purchases recently but RFM value is large than average	0
7	Losing Customer	Below average RFM, Will lose them if not reactivated	0
8	High Risk Customer	Spent Big money, purchased long time ago, frequency is high	0
9	Almost Lost Customer	Very low recency, Very low frequency, but spent high amount of moiney	0
10	Evasive Customer	Very low recency, Very low frequency, but spent small amount of money	0
11	Lost Customer	Lowest RFM	0

Interpretation

The Platinum Customer will be the Target customers

- Run the new customer list on the model by assigning the Rank conditions and title to the table

- The 1000 new customer is categorized to the rank title using the developed model

- The top 229 customers would have bought recently, buys often and spent a huge amount of money

THANK YOU