TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

[KPMG’s Lighthouse & Innovation Team][Shitab Tanzim]

Agenda

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

Introduction

Recommend Top Customers to Target from New Customer Dataset

Problem:

Sprocket Central Pty Ltd is a long-standing KPMG client who specialises in high-quality bikes and accessible cycling accessories to riders. Their marketing team is looking to boost business by analysing their existing customer dataset to determine customer trends and behaviour. They have given us a new list of 1000 potential customers with their demographics and attributes with no prior transaction history with the organisation.

Action:

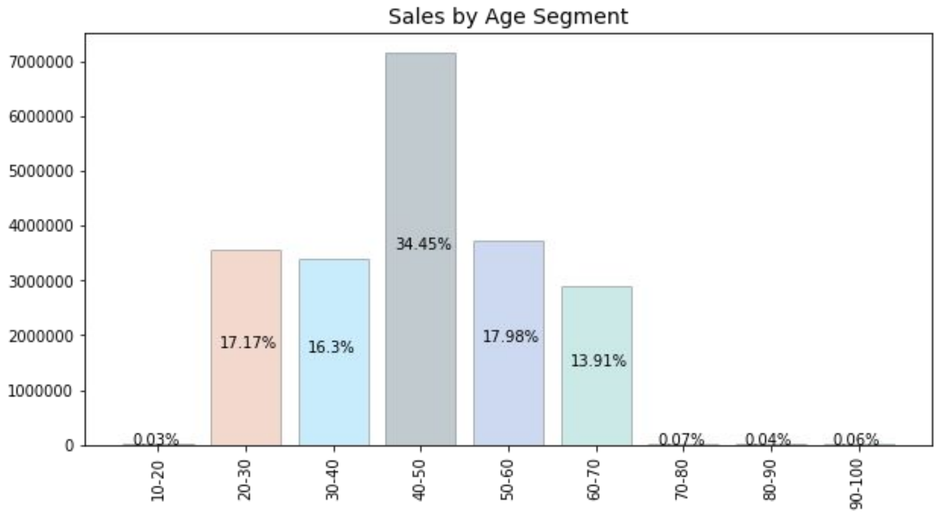
1. Analysis of the Transaction, Customer Demographic and Address dataset provided by Sprocket.
2. Find trends in sales volume by factors such as age, bike related purchase, job industry, wealth segment, car and property ownership,and state.
3. Calculate RFM score to recognise customer value.

Data Exploration

Age

<u>Insights</u>:

1. Customer with age between 40 and 50 years contribute to 34.45% of the sales.
2. Almost same sales came from 20-30, 30-40 and 50-60 age category.
3. Negligible sales from customer below 20 and above 70 years of age.

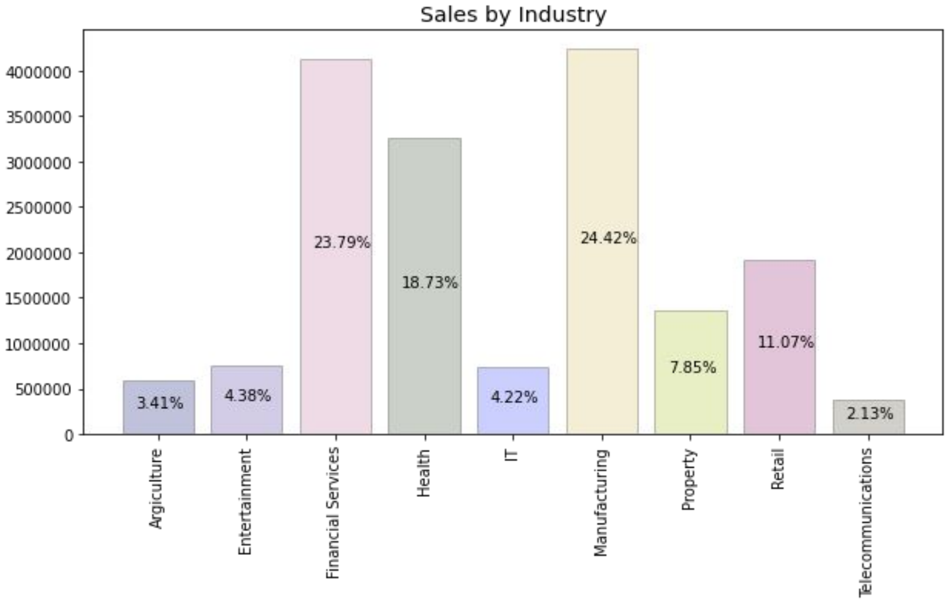


Data Exploration

Industry

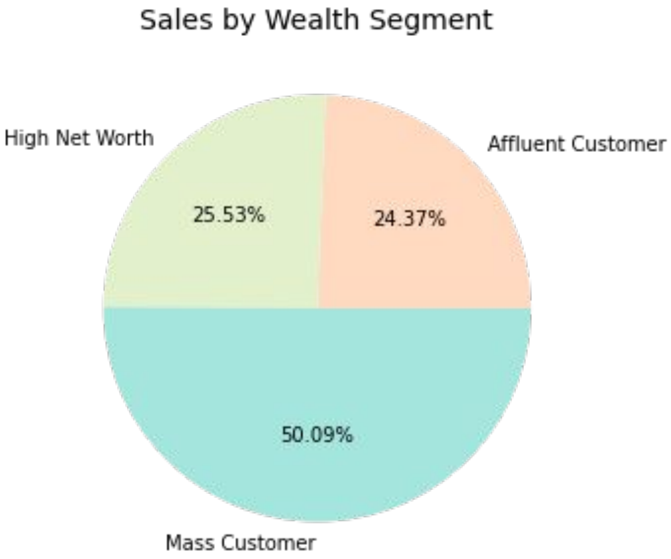
<u>Insights</u>:

1. Customer from Financial Service and Manufacturing contributed most; with 23.79% and 24.42% of the sales.
2. Health sector follows with 18.73% of the sales.
3. More than 2/3rd of the sales came from these 3 industry.



Data Exploration

Wealth Segment



Insights:

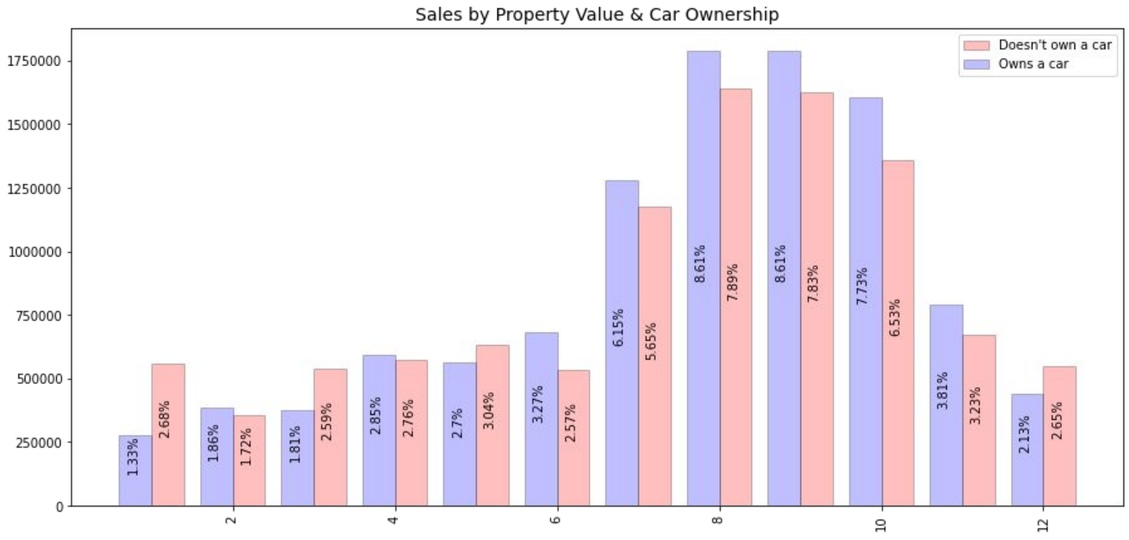
1. Half of the sales came from Mass Customers
2. High Net Worth and Affluent Customers contributed nearly equal.

Data Exploration

Property Value and Car Ownership

<u>Insights</u>:

1. Most sales came from customer with property value between 7 and 10
2. Almost 60% sales came from these customers
3. Customer who owns a car are slightly more profitable.

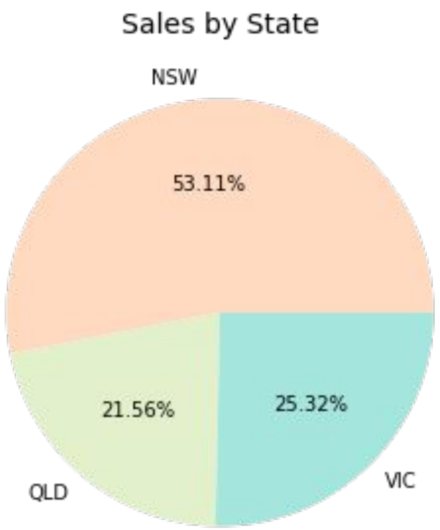


Data Exploration

State

<u>Insights</u>:

1. More than half of the sales are from New South Wales.
2. According to Australian Bureau of Statistics population of NSW, VIC, and QLD are 8M, 6.6M, and 5.2M. Therefore, population is not the main factor in sales per state.



Model Development

RFM Scoring System

RFM stands for Recency, Frequency, and Monetary value, each corresponding to some key customer trait. The RFM score is a numerical score that helps you recognize all types of customers, from the best to the worst.

Predicting the RFM score of New Customers based on their demographic can give us an idea about how valuable these customers can be.

	customer_id	recency	frequency	monetary	r_score	f_score	m_score	rfm_class	rfm_score	rfm_label
0	1	7	93.0	9084.45	4	4	4	444	12	Can't Loose Them
1	2	128	81.0	4149.07	1	4	2	142	7	Loyal
2	4	195	33.0	1047.72	1	2	1	121	4	Needs Attention
3	5	16	56.0	5903.20	4	3	3	433	10	Can't Loose Them
4	6	64	35.0	5931.69	2	2	3	223	7	Loyal
...	...	...	...	...	...	...	...	...	...	...
3402	3496	256	99.0	4725.38	1	4	2	142	7	Loyal
3403	3497	52	73.0	3744.07	2	3	1	231	6	Potential
3404	3498	127	28.0	5177.06	1	2	2	122	5	Promising
3405	3499	51	29.0	7673.48	2	2	3	223	7	Loyal
3406	3500	144	71.0	4922.41	1	3	2	132	6	Potential

Model Development

Build RFM prediction model by Machine Learning.

- To predict the RFM score of the New Customers, we are going to build a ML Model.
- Age, gender, no. of bike related purchases, and all the other factors explored before will be used as an input variable (X).
- RFM score will be used as the output variable (Y) to perform supervised learning.

Interpretation

Customer to target

-After predicting the rfm score of the New Customer and sorting based on the same we can get the best customers to target.

-We can take the top 100 (or how many Sprocket wants) from the sorted table

-Moreover we can filter the best in each categories (job industry, state, etc.) to further narrow down our target customers.

- The order of predicted rfm score and the filtering of categories will give us the best results.

rfm_pred
9.966295
9.906200
9.881624
9.864768
9.835748

Interpretation

Customer to target

<u>Glimpse of the top 15 customers to target:</u>

	first_name	last_name	gender	past_3_years_bi	Age	job_industry_cat	wealth_segment	owns_car	state	property_valuati	rfm_pred
2	Davie	Blay	Male	94	36	Financial Servic	Mass Customer	No	NSW	7	9.429494527
3	Bessie	Roscow	Female	78	28	Financial Servic	Mass Customer	No	NSW	10	9.422510055
4	Sybilla	MacCart	Female	88	35	Financial Servic	Mass Customer	Yes	NSW	7	9.414306151
5	Darlleen	Shalcras	Female	77	41	Health	Mass Customer	No	NSW	10	9.372660955
6	Dorian	Stollen	Male	78	42	Financial Servic	Mass Customer	Yes	NSW	11	9.365327324
7	Noami	Cokly	Female	74	59	Manufacturing	Mass Customer	Yes	NSW	11	9.342068306
8	Theresa	Cowper	Female	99	45	Manufacturing	Mass Customer	No	NSW	10	9.335894949
9	Maximilian	Geffen	Male	96	67	Manufacturing	Mass Customer	Yes	NSW	8	9.325213632
10	Engracia	Dobbs	Female	84	63	Health	Mass Customer	No	NSW	8	9.264102327
11	Sammy	Borsi	Female	99	50	Financial Servic	Mass Customer	No	NSW	7	9.256804759
12	Ellwood	Budden	Male	82	24	Health	Mass Customer	Yes	NSW	10	9.231887261
13	Dorian	Emery	Female	94	23	Manufacturing	Mass Customer	Yes	NSW	8	9.231822272
14	Geoff	Sitford	Male	97	57	Financial Servic	Mass Customer	Yes	NSW	8	9.201736676
15	Wylie	Huntingdon	Male	99	56	Financial Servic	Mass Customer	No	NSW	8	9.112378698
16	Guss	Karim	Male	95	53	Manufacturing	Mass Customer	No	NSW	11	9.090264227

Appendix

Appendix

1. RFM Score: https://www.datacamp.com/community/tutorials/introduction-customer-segmentation-python
2. EDA and Model Training: https://colab.research.google.com/drive/1TFOaymyi6WyVfvMy88KLPlQ8Y870juPt?usp=sharing
3. Catboost: https://catboost.ai/
4. Data Quality Assessment: https://docs.google.com/document/d/1ds-pXY8mok3ksSyu4P5i8TN6IDMyu9xx7Ne1twBkKbg/edit
5. Top 100 customers: https://docs.google.com/spreadsheets/d/1CJw594Ec3KFs1Im-Hwrre2gOSHkzfAfKpu_kQS5Jtgg/edit#gid=299829157