



Sprocket Central Pty Ltd Data Analyze Project

Reported by Terry



Part 1: Data Cleaning



There are 197 missing values for brand, product line/class/size and standard cost in the Transaction table; and the null samples will be filtered out in the analysis that follows.

```
1 customer_list.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20000 entries, 0 to 19999
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   transaction_id         20000 non-null  int64  
1   product_id             20000 non-null  int64  
2   customer_id            20000 non-null  int64  
3   transaction_date       20000 non-null  datetime64[ns]
4   online_order           19640 non-null  float64 
5   order_status           20000 non-null  object  
6   brand                  19803 non-null  object  
7   product_line           19803 non-null  object  
8   product_class          19803 non-null  object  
9   product_size           19803 non-null  object  
10  list_price             20000 non-null  float64 
11  standard_cost          19803 non-null  float64 
12  product_first_sold_date 19803 non-null  datetime64[ns]
```



- There are around 500 missing values in occupation and industry variables in the Customer Demographics table.
- Default is not a relevant indicator, and hence its absence will be disregarded.
- There is an anomalous value in DOB; therefore, 1843 has been changed to 1943.
- The gender has been standardised to F-Female and M-Male.

#	Column	Non-Null Count	Dtype
0	customer_id	4000 non-null	int64
1	first_name	4000 non-null	object
2	last_name	3875 non-null	object
3	gender	4000 non-null	object
4	past_3_years_bike_related_purchases	4000 non-null	int64
5	DOB	3913 non-null	datetime64[ns]
6	job_title	3494 non-null	object
7	job_industry_category	3344 non-null	object
8	wealth_segment	4000 non-null	object
9	deceased_indicator	4000 non-null	object
10	default	3698 non-null	object
11	owns_car	4000 non-null	object
12	tenure	3913 non-null	float64



There is no missing value or outlier in
the customer address table

In State, 'New South Wales' is changed
to 'NSW'

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3999 entries, 0 to 3998
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   customer_id           3999 non-null   int64
1   address               3999 non-null   object
2   postcode              3999 non-null   int64
3   state                 3999 non-null   object
4   country               3999 non-null   object
5   property_valuation    3999 non-null   int64
dtypes: int64(3), object(3)
memory usage: 187.6+ KB
```



Part 2: Data Description

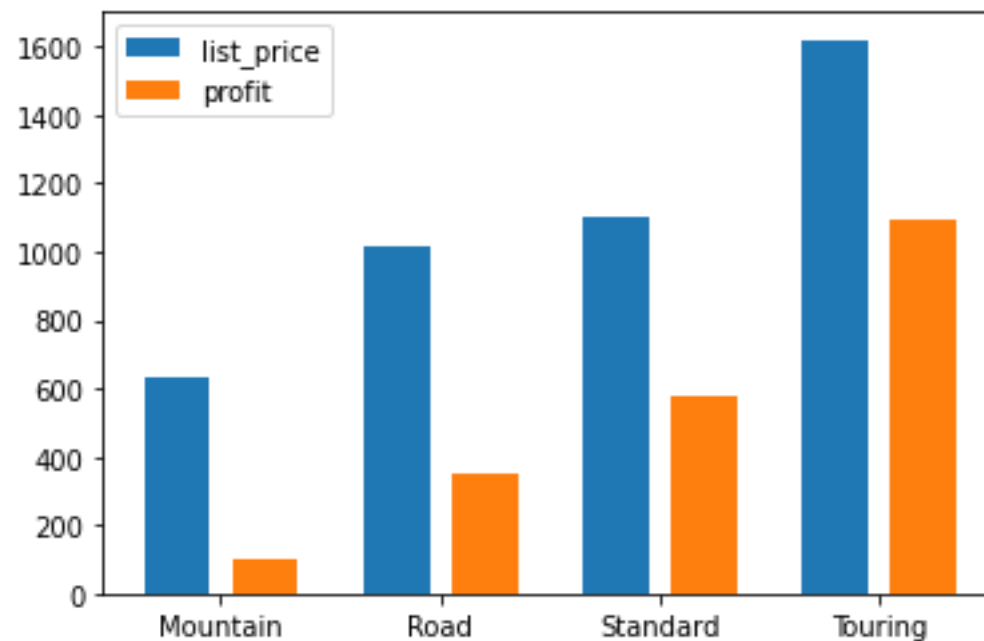


Product Line

The **touring product line** has the highest average profit, one of the fields worth focusing on.

The market of mountain bikes is the smallest and per value is the lowest, which is required to be updated

	list_price	profit	Amount
product_line			
Mountain	628.54	96.41	423
Road	1018.02	352.74	3970
Standard	1102.92	574.14	14176
Touring	1620.21	1094.13	1234



F-Value is 965.10, P-value is 0.00



| Brand

The right picture represents the TOP5 brand in each product line.

Since touring has the highest per value, Weare A2B is also the brand required to be placed in a primary partner brand

Solex and Giant are the basis of revenue of the company, and most customers choose these two brands.

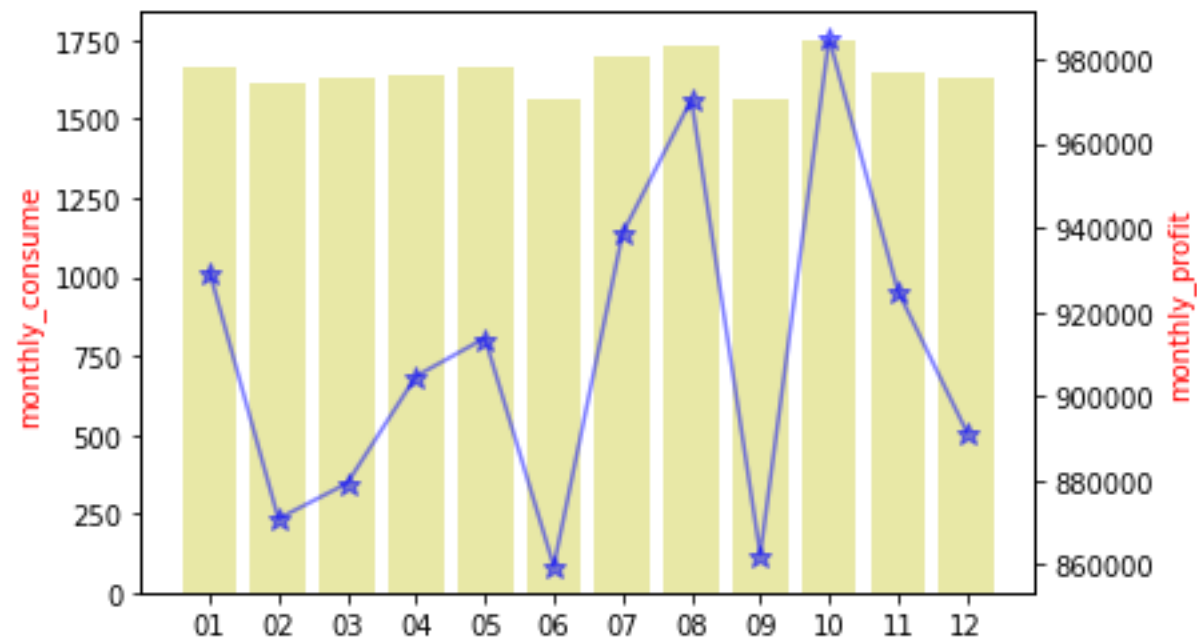
		Amount
product_line	brand	
Mountain	Trek Bicycles	223
	Norco Bicycles	200
	Trek Bicycles	995
	Norco Bicycles	923
	OHM Cycles	779
	Giant Bicycles	579
	Solex	537
Standard	Solex	3508
	Giant Bicycles	2554
	WeareA2B	2520
	OHM Cycles	2035
	Norco Bicycles	1787
Touring	WeareA2B	618
	OHM Cycles	229
	Solex	208
	Giant Bicycles	179



Timeline

AUGUST/OCTOBER are high consumption seasons

The difference on profit is not significant and not pass the ANOVA test



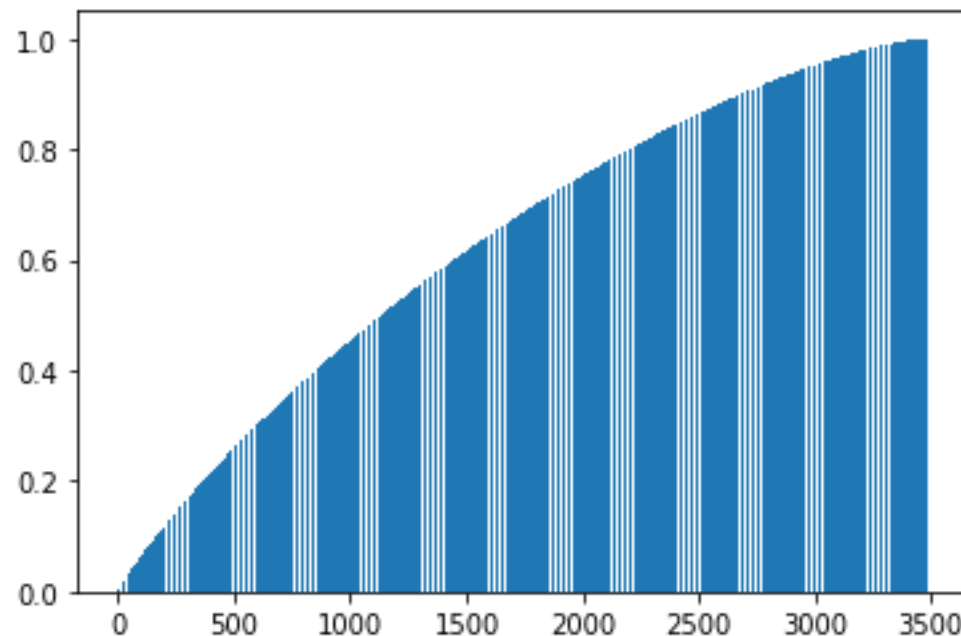
The F-value is [0.39393174] and P-value is [0.95911784]



| Consume Value

The order volume per capita is 5.668, the average order value is 1.106.225, and the value per customer is 6.279.817.

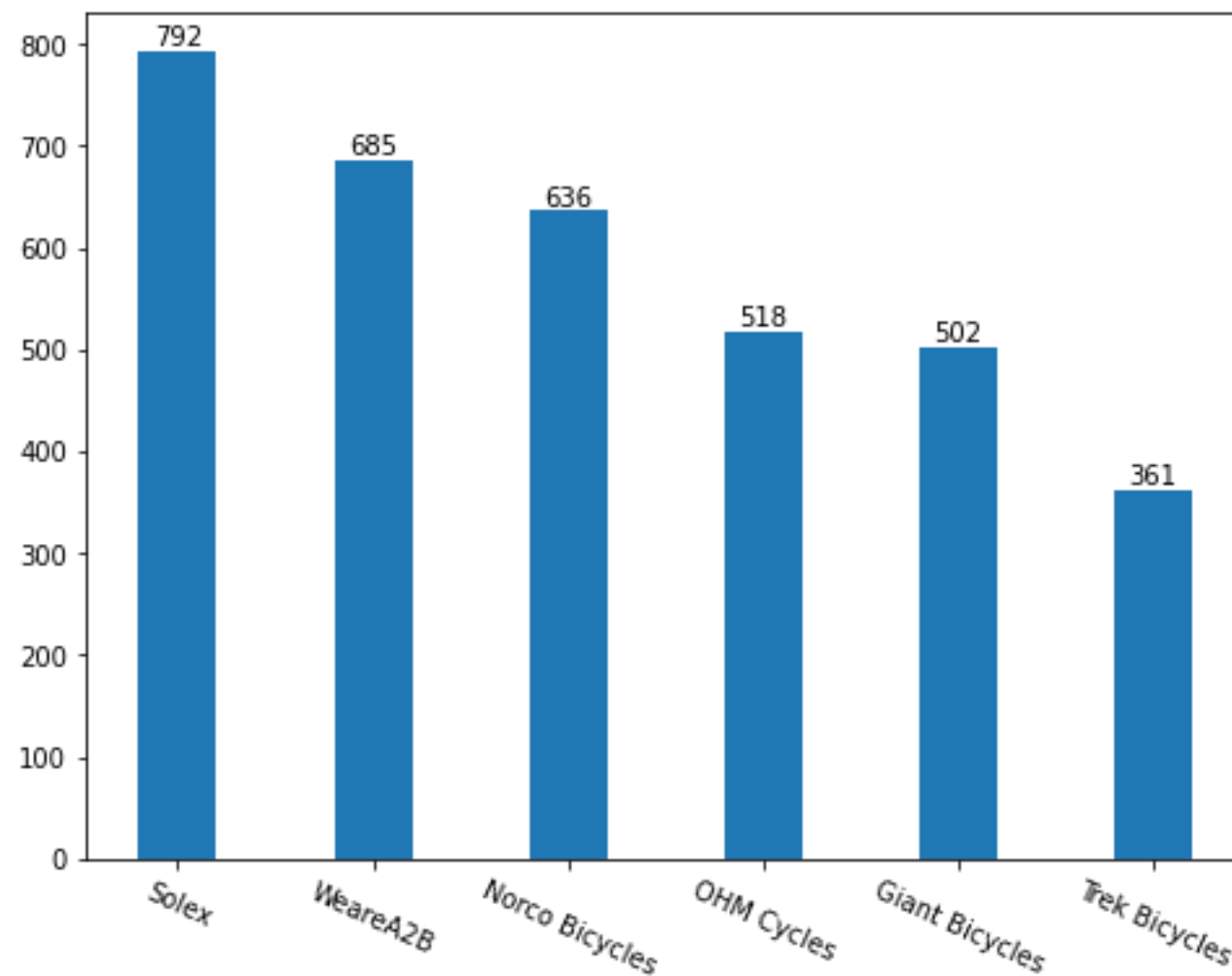
The right figure depicts the accumulated value of each customer; 62.96 % of customers contributed more than 80 % of revenue, while the TOP 20 percent of shoppers account for more than 30 % of orders and 34 % of revenue.





| First Bicycle

As the graphic on the right indicates, the majority of consumers have chosen Solex, WeareA2B, or Norco as their first bicycle; therefore, these three brands are suggested for beginners.



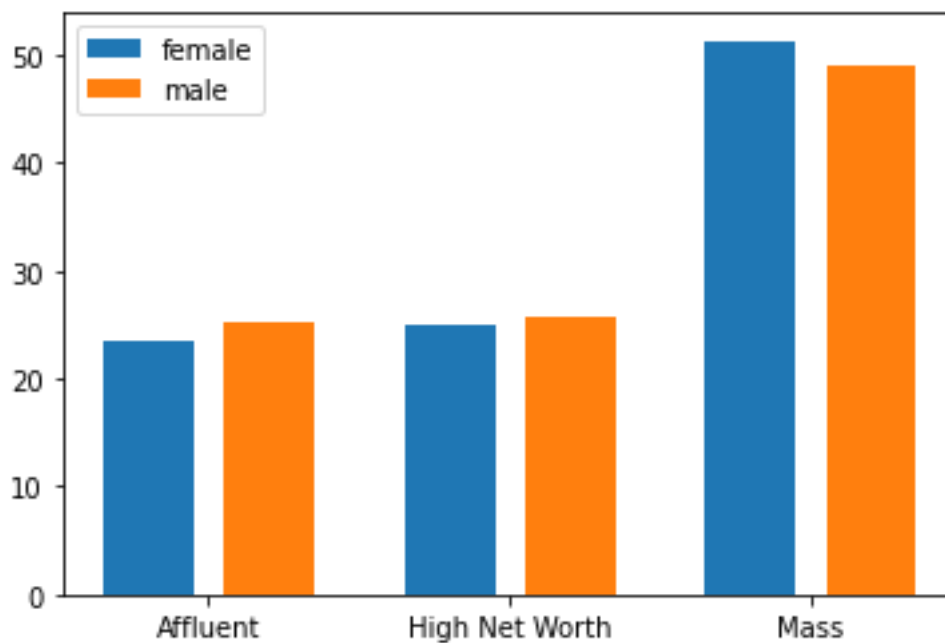


| Gender

According to the Chi-Square Test, there is a statistically significant difference between genders; hence, male customers may represent greater potential consumer value in the bicycle industry.

	Affluent	High Net Worth	Mass
female	23.59	25.10	51.32
male	25.23	25.88	48.89

Chi_Square value=12.1754,
p-value=0.0023, degree of freedom=2

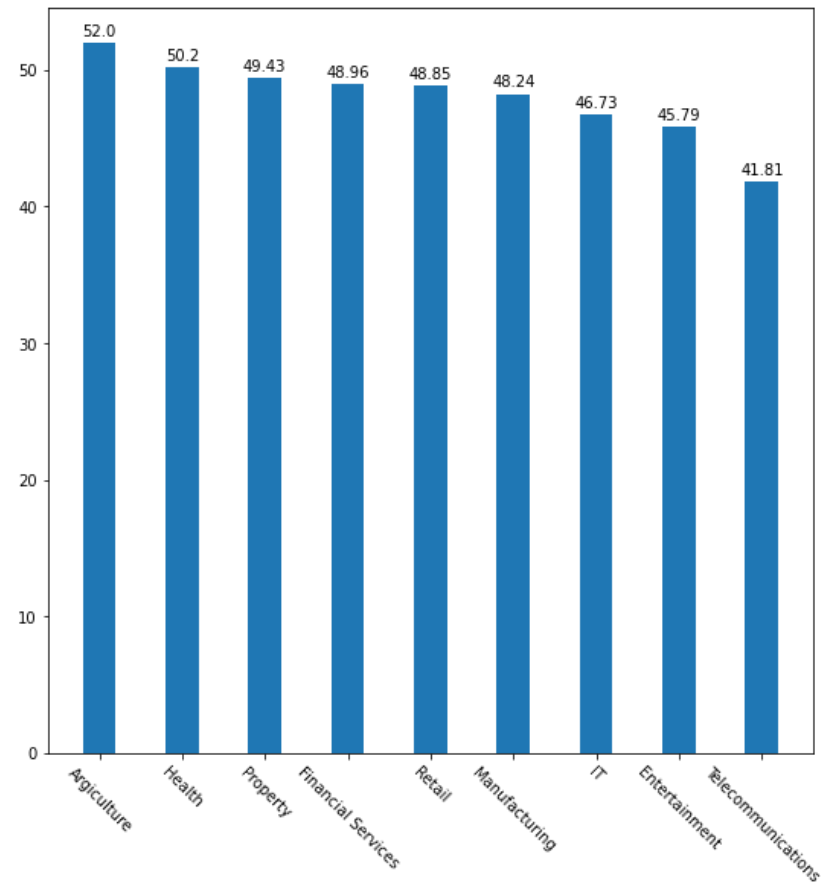
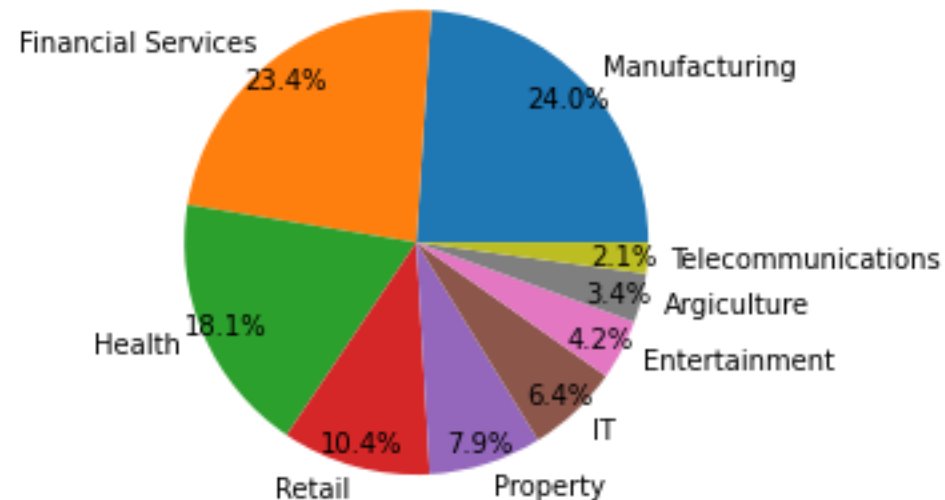




Job And Industry

As seen in the pie chart, manufacturing and financial service customers account for more than 20 percent of revenue.

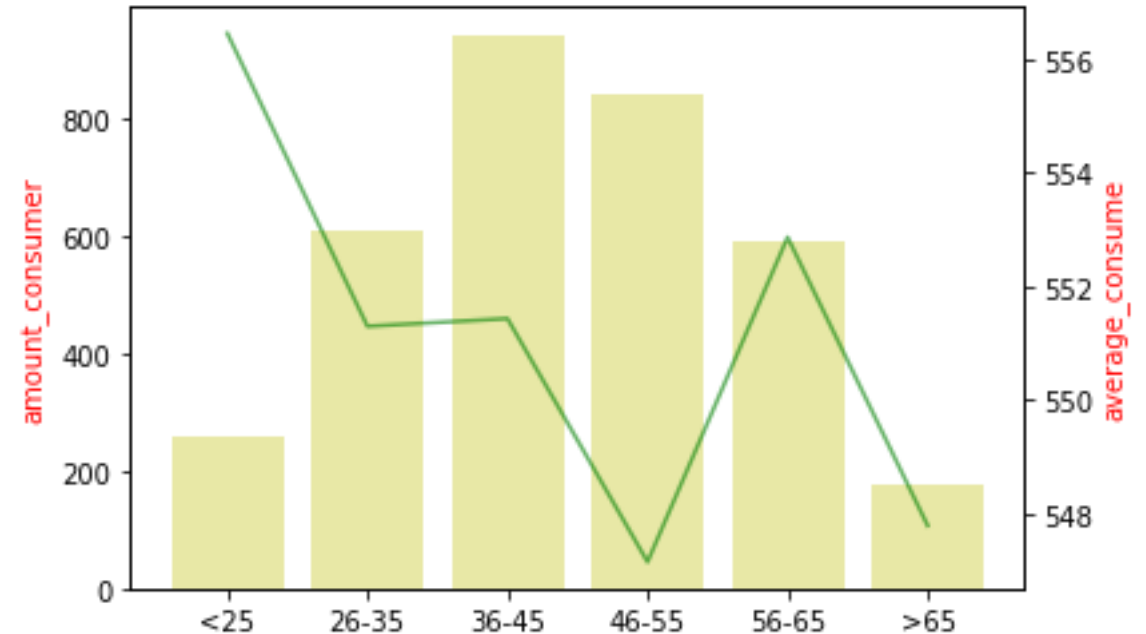
The bar chart demonstrates that Agriculture and Health practitioners have greater demands in high-end bicycle service





| Age

The mass customers are aged between 36-45 and the younger generation represents strong purchasing capability



The F-value is [0.11190573] and P-value is [0.9897797]



| Location

NSW has the largest per-suburb market
with the highest average consumer

From the Top 5 suburbs in each state,
the remote area have stronger demand
for bicycles

state	consumer_amount	
	sum	mean
NSW	1866	4.936508
QLD	743	3.952128
VIC	880	3.271375

state	postcode	consumer_amount
NSW	2153	28
	2170	28
	2145	27
	2155	26
	2770	24
QLD	4300	16
	4503	15
	4350	15
	4670	13
	4211	12
VIC	3977	22
	3630	13
	3023	12
	3175	11
	3046	11

Each state Top5 biggest market

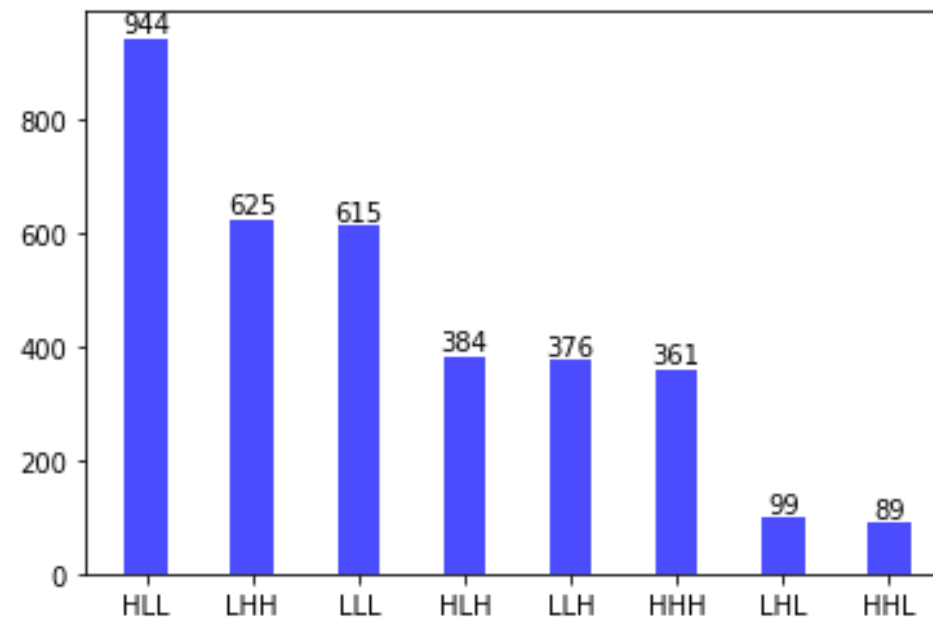
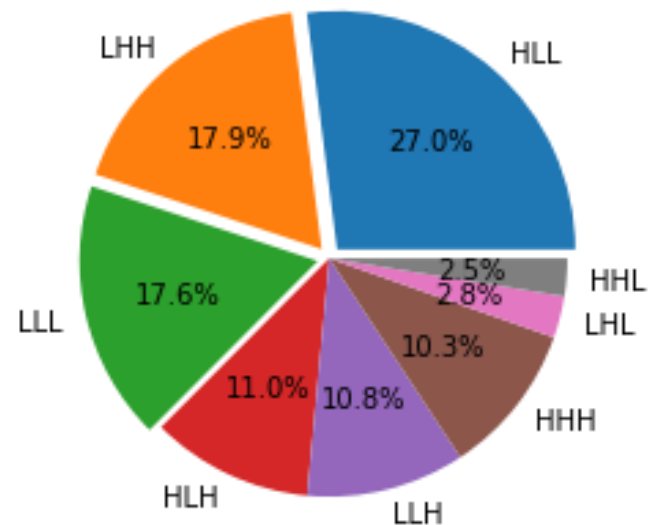


Part 3: RFM Modelling



| RFM Category

- From the perspectives of Recency/ Frequency/ Monthly to measure, categorise customers and employ the appropriate marketing strategies.
- More than a quarter of clients are HLL customers, who represent the company's primary revenue source.
- LHH customers are significant to the firm as a whole due to their second-largest market size and relatively high customer value.

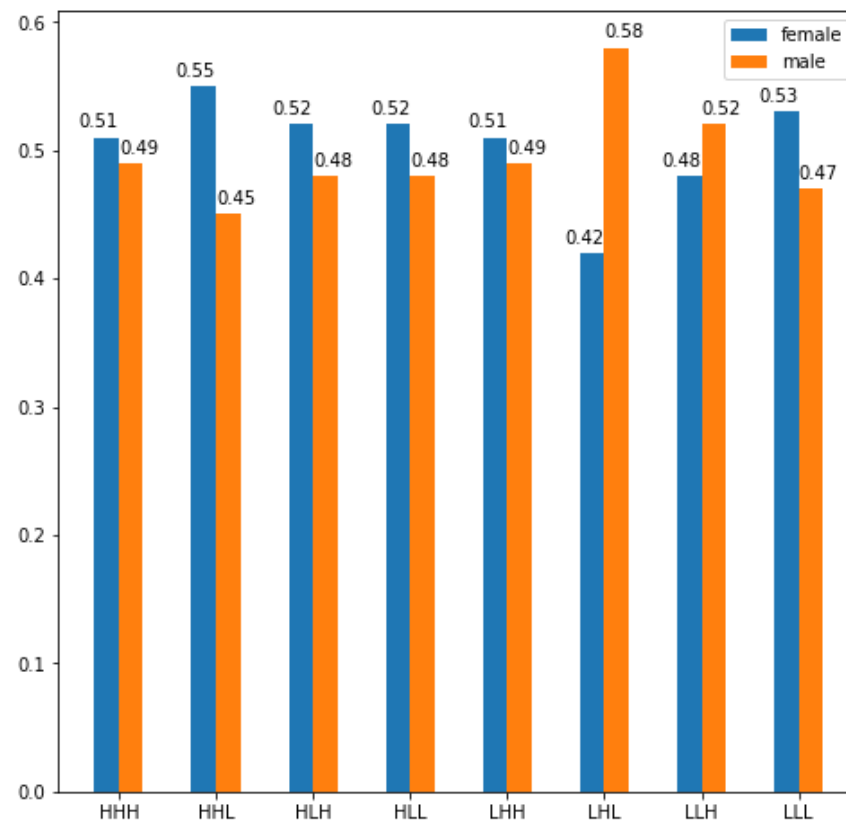




| Gender and RFM

In LHL and LLH, the proportion of male consumers is substantially larger than in other categories.

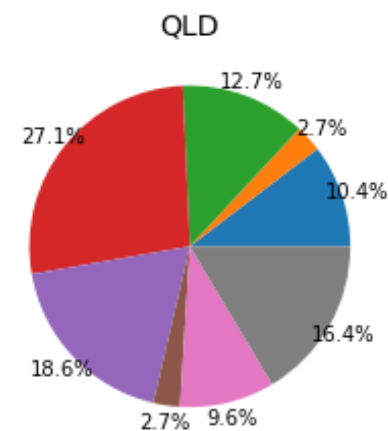
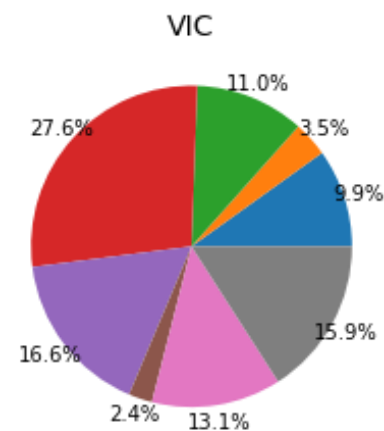
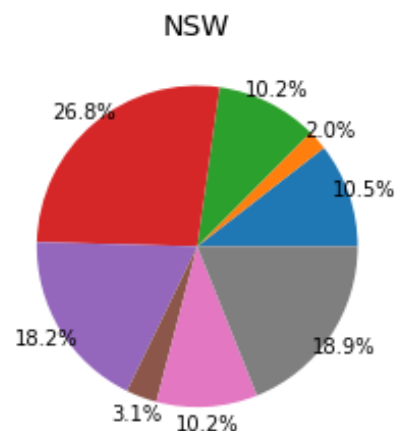
Customers of HHL are devoted to our company despite their low value, and this segment is dominated by female users





| State and RFM

Among three states, NSW has the largest proportion of HHH clients; Queensland has the highest proportion of LHH clients, while Victoria has the biggest proportion of HLL customers





| Brand and RFM

For LHH customers, though it is not the largest customer group, it has the largest segment market, where is indicated at Solex is the business's primary revenue driver and attractant for LHH customers.

		amount
category	brand	
HHH	Solex	674
	Trek Bicycles	483
HHL	Solex	126
	Norco Bicycles	118
HLH	WeareA2B	408
	Solex	398
HLL	Solex	789
	Giant Bicycles	588
LHH	Solex	1123
	Giant Bicycles	907
LHL	Solex	164
	OHM Cycles	139
LLH	WeareA2B	448
	Solex	430
LLL	Solex	544
	Giant Bicycles	470



Part 4: Conclusion



- According to the RFM model, HLL consumers are the company's greatest market; therefore, it is recommended that the recommendation algorithm be applied to them in order to increase their following customer value.
- Customers of LHL and LLH are predominantly male; therefore, it is advantageous to generate advertising on the masculinity channel with a discount incentive to encourage repurchase.
- It is proposed that premium products be offered to this demographic in order to boost per-unit customer value among HHL, a business-loyal and female-dominated group.
- LHH is the second-largest customer segment with a high churn risk and high value, and QLD has the biggest proportion of this segment; hence, recall methods should be adopted first in QLD.
- Solex is the most popular brand in the Australian market and holds a key position within the LHH group; as a result, recall tactics may be more effective if Solex is promoted.