TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

Priscilla Ng, Junior Consultant

Agenda

1. Introduction
2. Data Exploration
3. RFM Analysis
4. Model Development
5. Interpretation
6. Conclusions
7. Next Actions

Introduction

Introduction

Conduct customer analysis and identify potential new customers with bigger purchasing power to boost sales.

Background

- Sprocket Central specializes in high quality bike and accessories.
- The Marketing Team is looking to boost sales and request KPMG to recommend which of the 1000 new customers will bring the highest sales value.

Approach Adopted

- Data exploration
- Feature engineering
- RFM analysis
- Modelling
- Recommendations
- Future actions

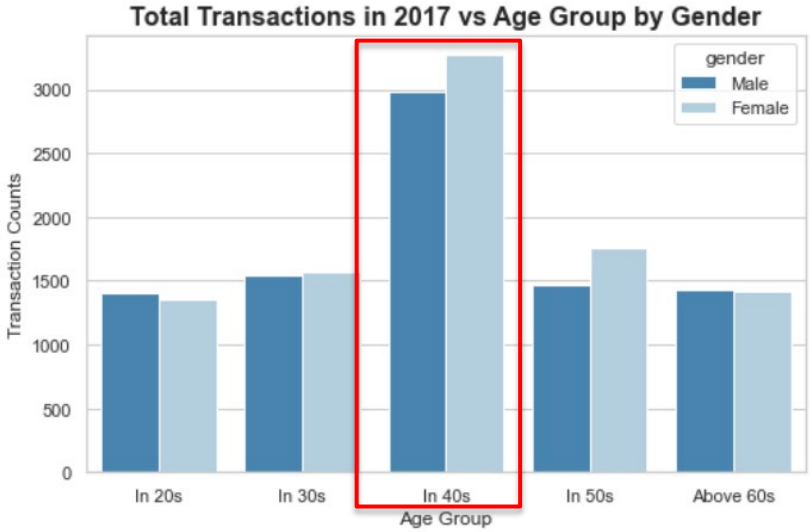
Data Exploration

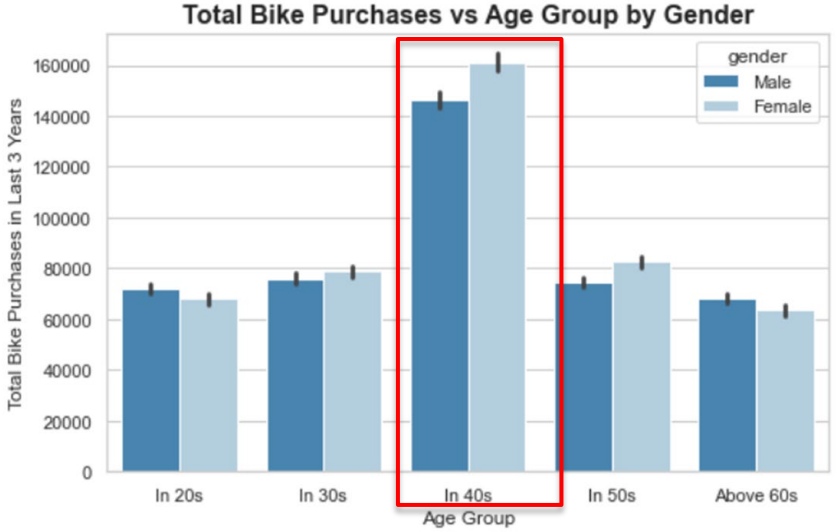
Transactions, Customer Demographics & Customer Address

Dataset	Records in Dataset	Data Quality Issue	Data Cleaning
Transactions	20,000 records from 1 Jan 2017 to 30 Dec 2017 3,494 unique customers	Missing data	Delete missing data < 2%
		Incorrect data type	Convert to correct datatype
Customer Demographic	4,000 unique customer records	Missing data	Delete missing data < 2%
		Data error & unreadable information	Delete these data
		Inconsistent values	Rectify the value
Customer Address	3,999 unique customer records	Inconsistent values	Rectify the value

Age Group By Gender

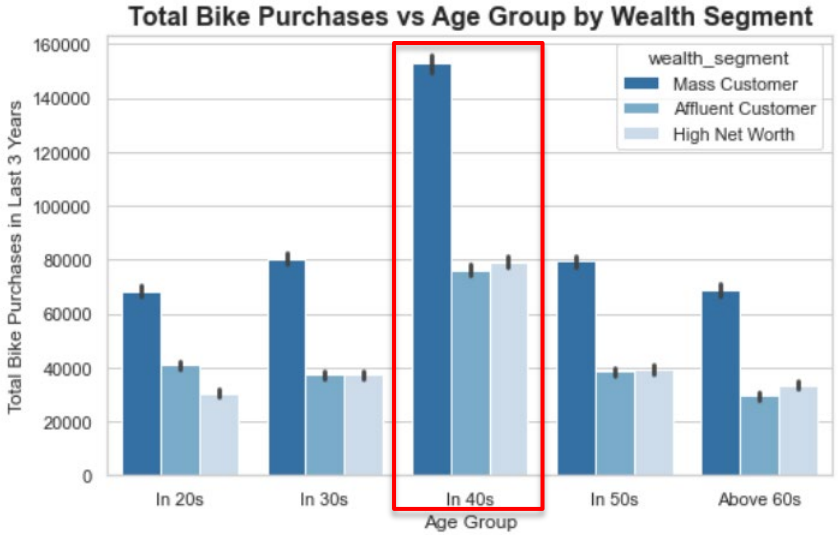
Age group 40s has highest transactions in 2017 & bike purchases in last 3 years. Female customers observed to be in higher ratio





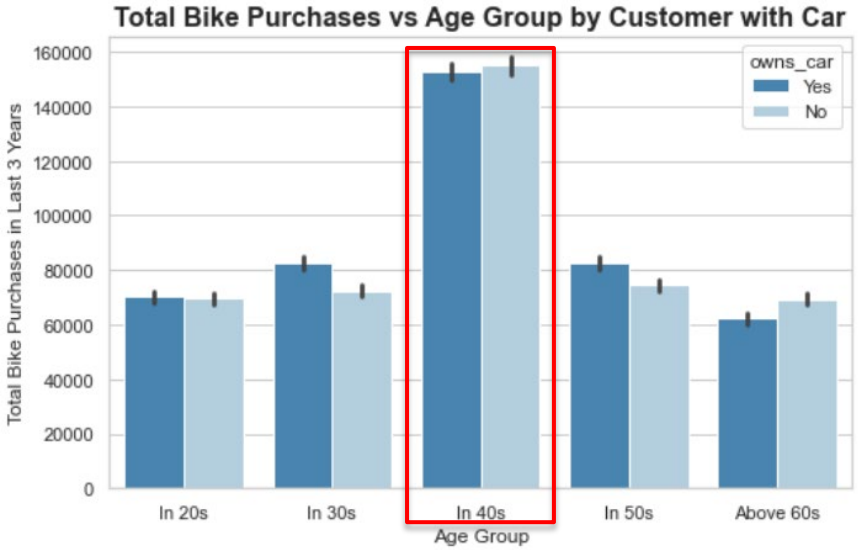
Age Group By Wealth Segment

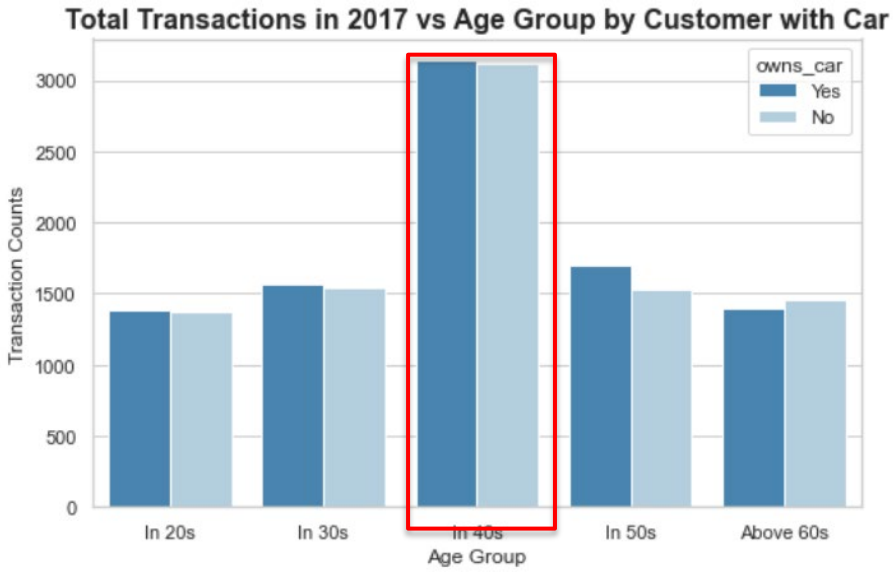
Mass customer wealth segment observed to be in higher ratio



Age Group By Customers with Car

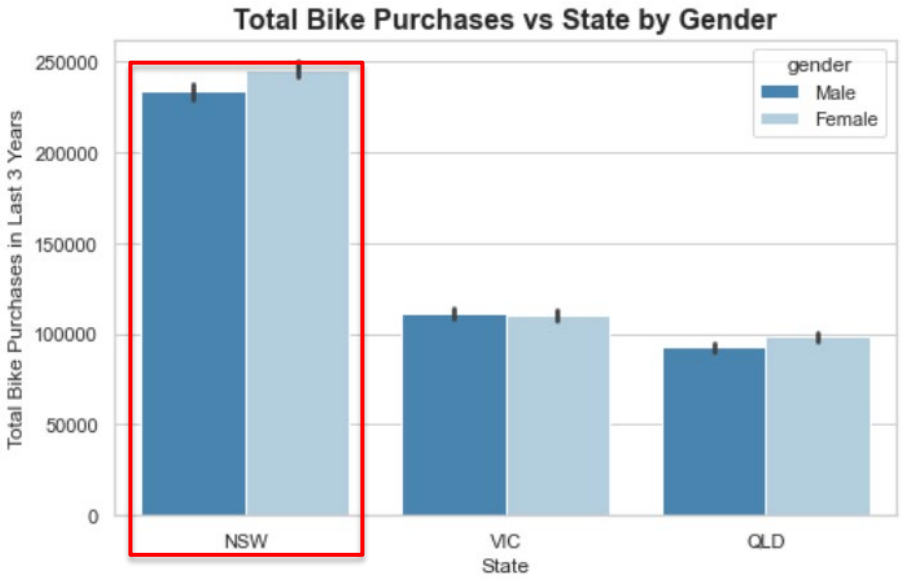
No significant different observed

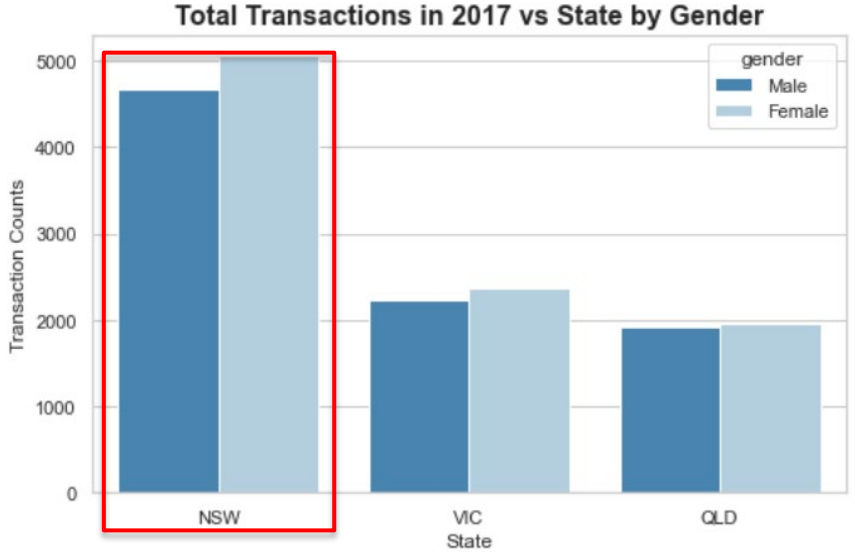




State By Gender

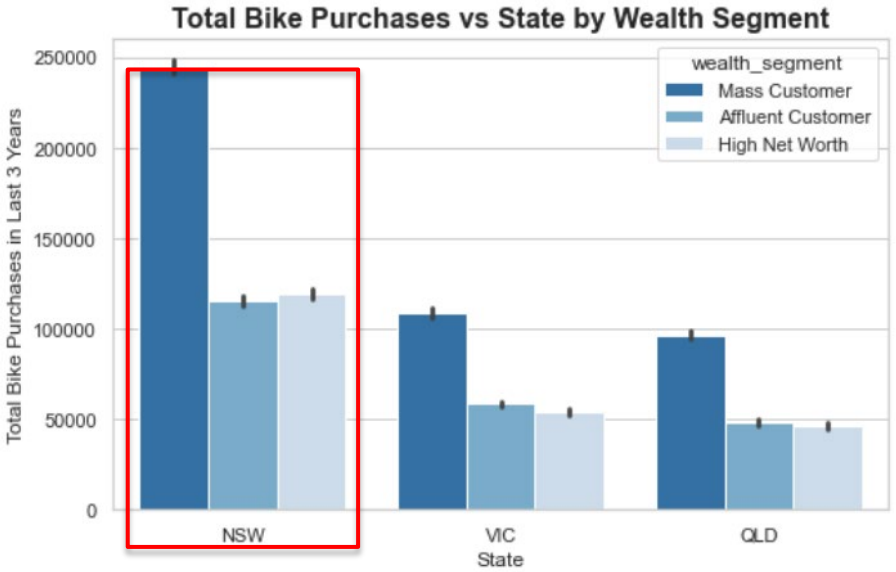
New South Wales has highest transactions in 2017 & bike purchases in last 3 years. Female customers observed to be in higher ratio

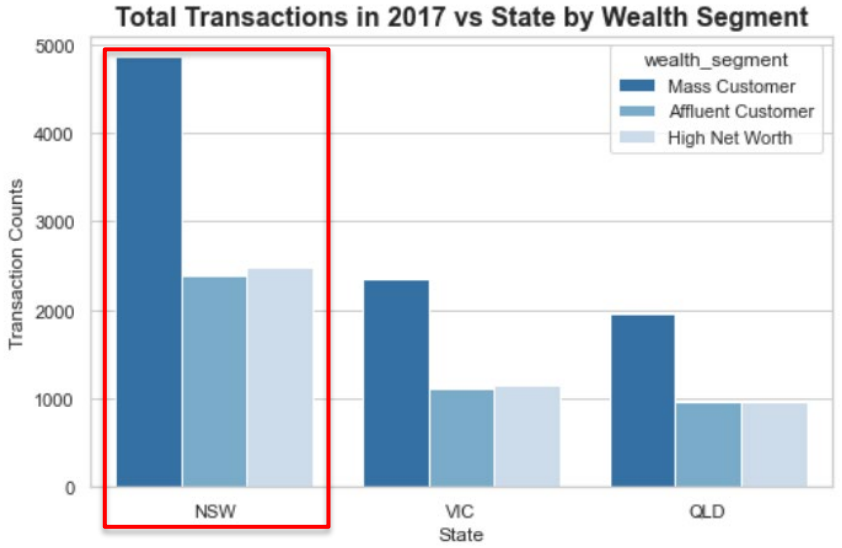




State By Wealth Segment

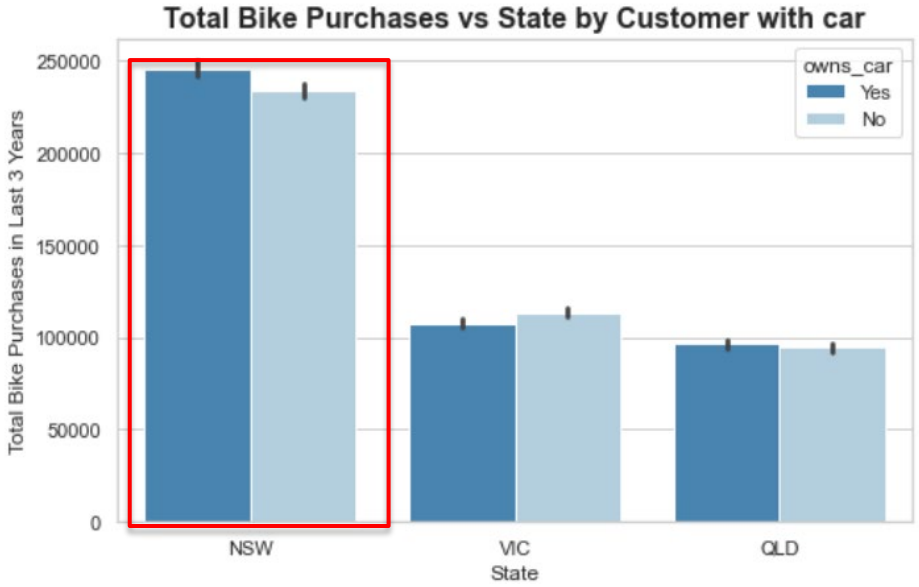
Mass customer wealth segment observed to be in higher ratio

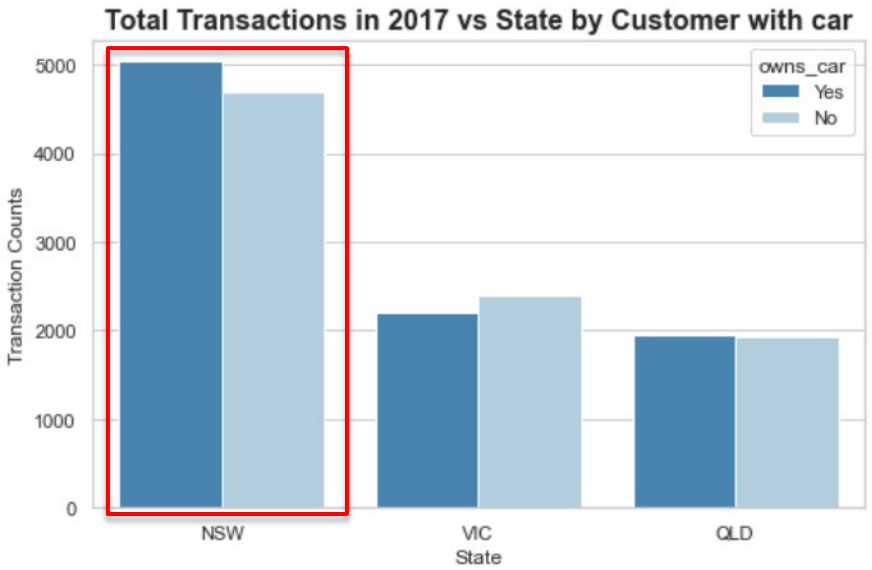




Location By Customers with Car

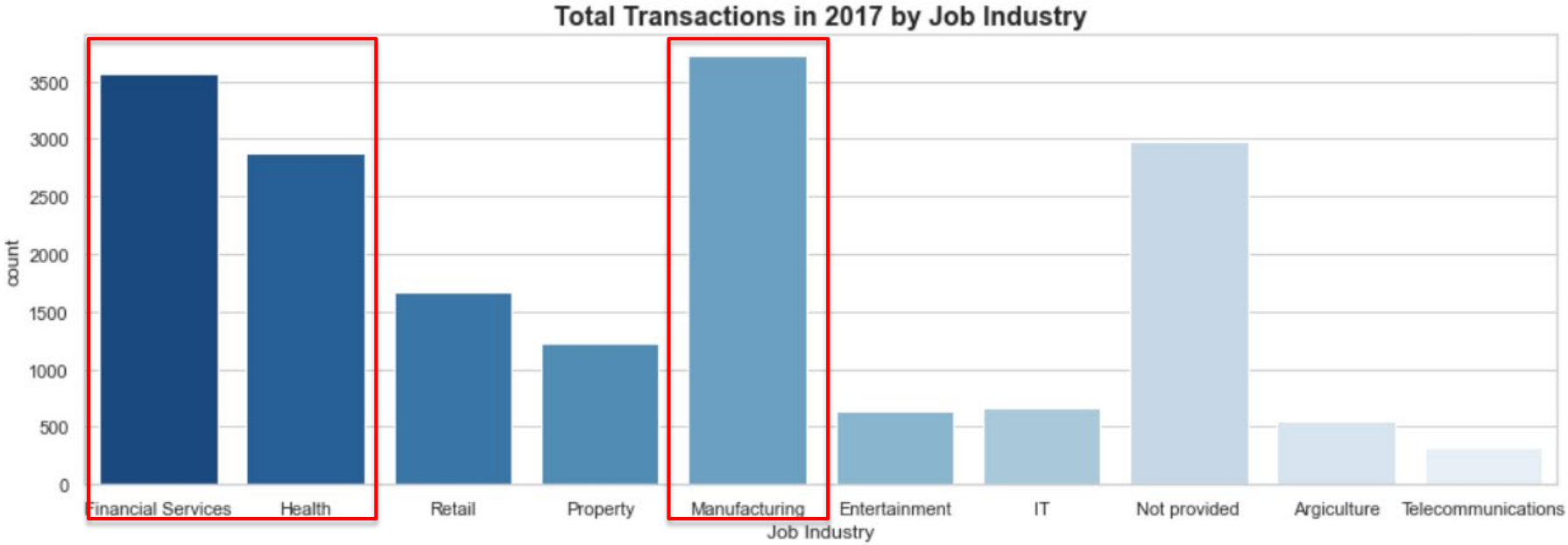
NSW customers with car observed to be in slightly higher ratio





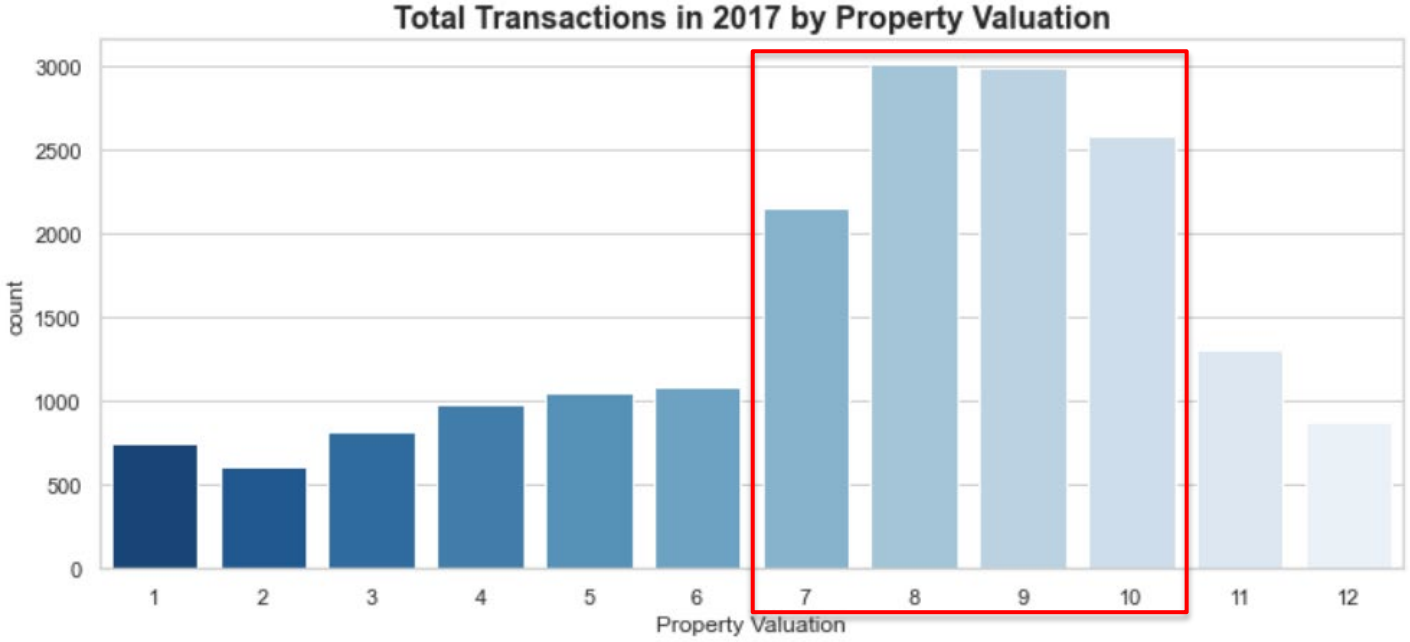
Top 3 Job Industry

Manufacturing, Financial Services, and Health are the top three industries with sales



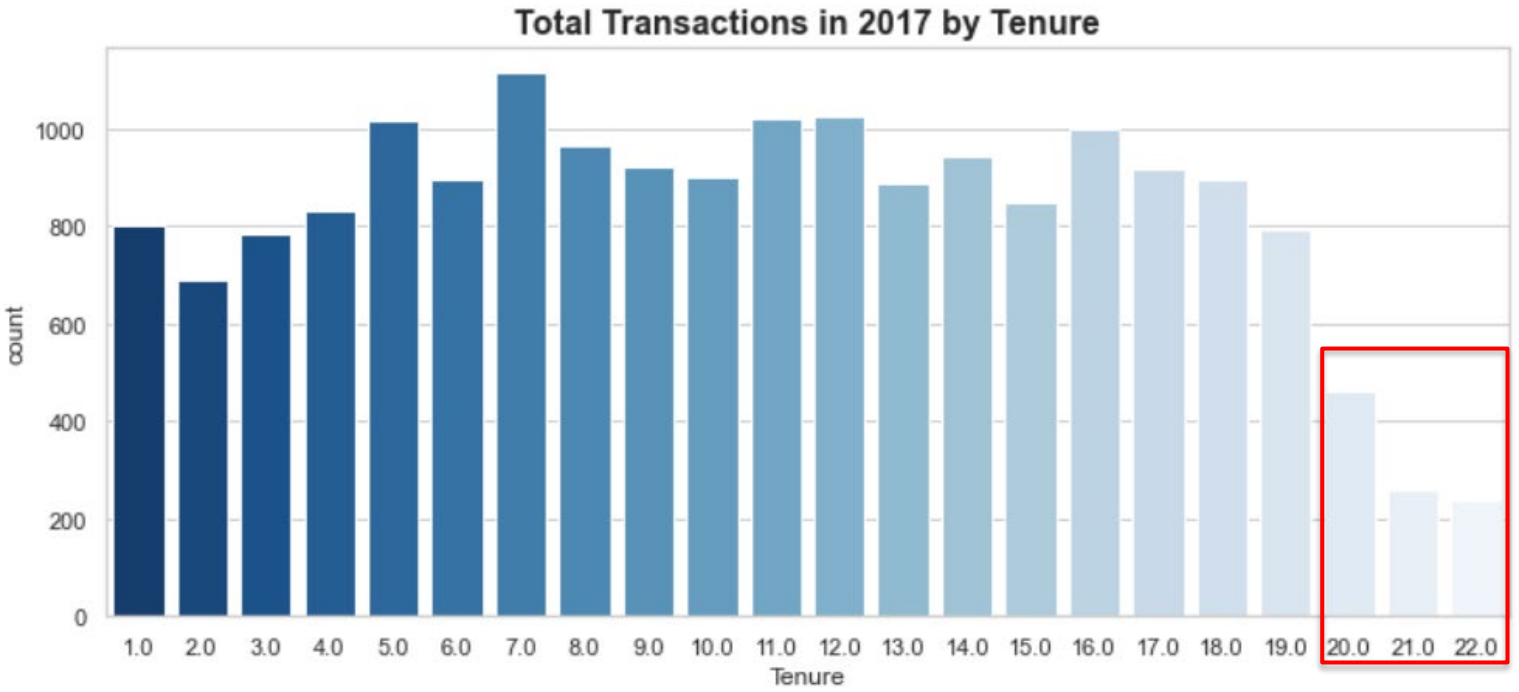
Property Valuation

Customers with property valuation between 7 to 10 generates more sales



Tenure

Customers with tenure between 20 t0 22 generates less sales



RFM Analysis

RFM Analysis

Recency (R)

- freshness of the customer's activity (purchase or visit)

Frequency (F)

- frequency of the customer's transactions or visits

Monetary value (M)

- intention of the customer to spend or the purchasing power of customer

➔ Calculate **RFM Score** for each customer

Modelling

Model Development

Identify best customer using RFM score

Target	Counts
Best Customer (1)	804
Other Customer (0)	2495

- Dataset is imbalanced
- Adopt SMOTE to create synthetic samples for sample balancing
- Models used: Logistics Regression, Random Forest, Gradient Boost, XgBoost
- Evaluation metric adopted: ROC-AUC score

Interpretation

Random Forest or Gradient Boost show better performance

All models have low ROC-AUC scores. The features available in the datasets may not be sufficient to train the model.

Model	ROC-AUC Score	Accuracy
Logistics Regression (Baseline)	0.4905	0.4945
Random Forest	0.5080	0.7200
Gradient Boost	0.4985	0.7515
XgBoost	0.4954	0.6933

Conclusions & Next Actions

Conclusions

- Current high value customers are:

- Aged between 40 – 50
- currently living in New South Wales
- Females
- Mass customers
- Working in Manufacturing, Financial Service, and Health

- The models developed have low ROC-AUC score.

Next Actions

- Consider to add in new features (e.g. distance from home to office) to improve the training of the model
- Try other sampling methods to handle the imbalanced dataset
- Perform hyper-tuning to optimize the parameters to improve the score

Thank you