The Analytics team

Sprocket Central Pty Ltd

Data Analytics Approach

TABLE OF CONTENTS

01 Introduction

Introduction

- In this presentation. Details will be provided about **three datasets** which are Transaction dataset, Customer Address dataset and lastly the Customer Dempgraphics dataset.

- **Three processes** are being followed which are the data exploration, model development, and finally the intrepretation.

- All of which to reach the **final goal** of the marketing team. Which is reaching the optimal resource allocation and boosting revenues, by understanding the customer behaviour and trends for target marketing.

02 Data Exploration

Data Exploration

- Data exploration is done by performing **EDA** which is exploratory data analysis. EDA is an approach for analysing datasets to summarize their main characteristics using statistical methods and approaches.

- EDA is used in **exploring record** sets to apprehend their predominant traits, discover patterns, locate outliers, and identify relationships between variables.

- It allows us to dive deeper into the data and draw primarily insights.

- The insights drawn from the EDA can be used in **model building, hypothesis testing, and further analysis**.

Data Exploration

The first Dataset **(CustomerAddress)** contains 3999 rows and 6 columns.

- The columns are the customers' address, id, postcode, state, country, and property valuation.
- The data in the property valuation is negatively skewed with a value of -0.63.
- There is no correlation among the variables in this dataset.

The Second Dataset (**CustomerDemographics**) contains 4000 rows and 13 columns.

- The columns are the customers' id, First name, Last name, gender, past 3 years bike related purchases, date of birth, job title, job industry, category, wealth segment, deceased indicator, default, own cars, tenure.
- The default column will be excluded as it is irrelevant to the model in addition to its inclusion of unspecified characters.
- This dataset is not skewed and there is no correlation among its numeric variables.

The Third Dataset (**Transaction**) contains 2000 rows and 13 columns.

- The columns are the transaction id, product id, customer id, transaction date, online order, order status, brand, product line, product class, product size, list price, standard cost, product first sold date.
- The standard cost column is positively skewed with a value of 0.86.
- There is no correlation among the variables of this dataset.

Data Exploration

The **limitations** of those datasets are:

- The nulls, as two of the available datasets contains nulls. Those nulls are less than 25% of the dataset so they will be dropped for better analysis.

- The non-numeric values need to be encoded to enter the model.

- The D.O.B column will be converted into age groups. Furthermore, the undefined columns will be excluded from the analysis like the default column in the Customer Demographics dataset.

If there was **more time**:

- A detailed study would be built to complete the nulls in the dataset.

- Further data would be extracted from different sources to be included in the analysis for better and more accurate model and assumptions.

03 Model Development

Model Development

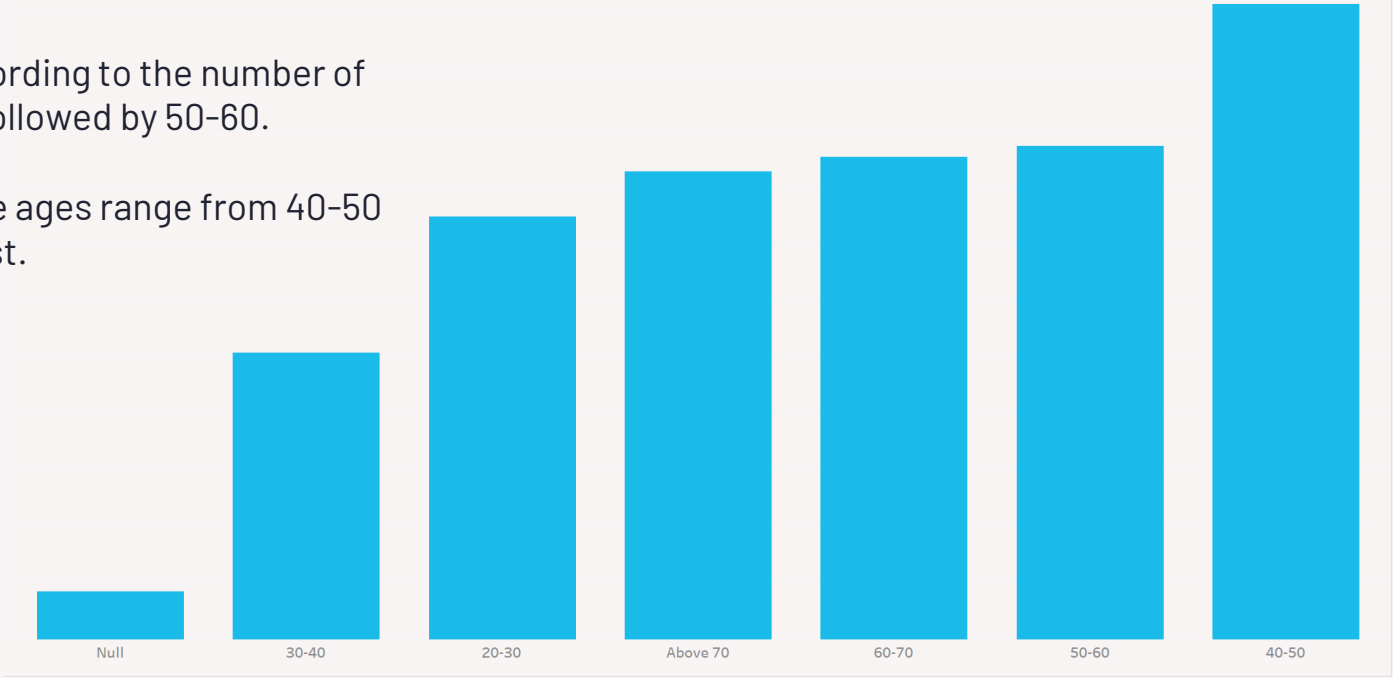
According to the EDA, the following Customers will be targeted as they showed the highest numbers thus, they are considered high value customers:

- Customers in the age group **40-50**.
- Customers who are **females**.
- Customers lying in the **Mass Customer** Wealth Segment.
- **Financial Services** and **Manufacturing job** categories shows the highest customers among the other job categories.
- Customer who **do not own cars** and from **South Wales State**.

04 Interpretation

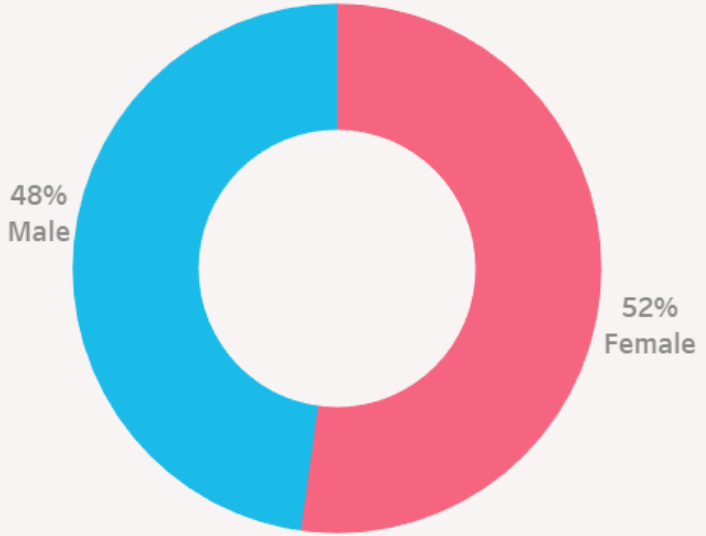
EDA > Age Groups

- According to the following data the D.O.B was converted into **6 age groups**.
- The highest age group according to the number of customers is from **40-50** followed by 50-60.
- Thus, the customers whose ages range from 40-50 should be targeted the most.



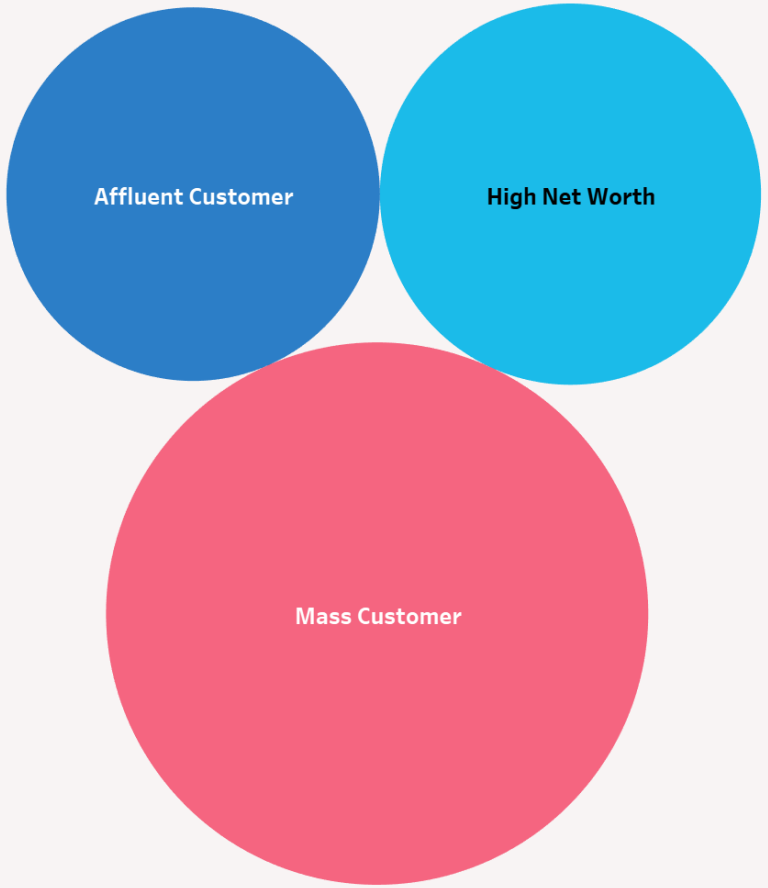
EDA > Gender

- According to the following graph the **females** are more likely to purchase bikes than males.
- Thus, they are to be targeted more than male customers.



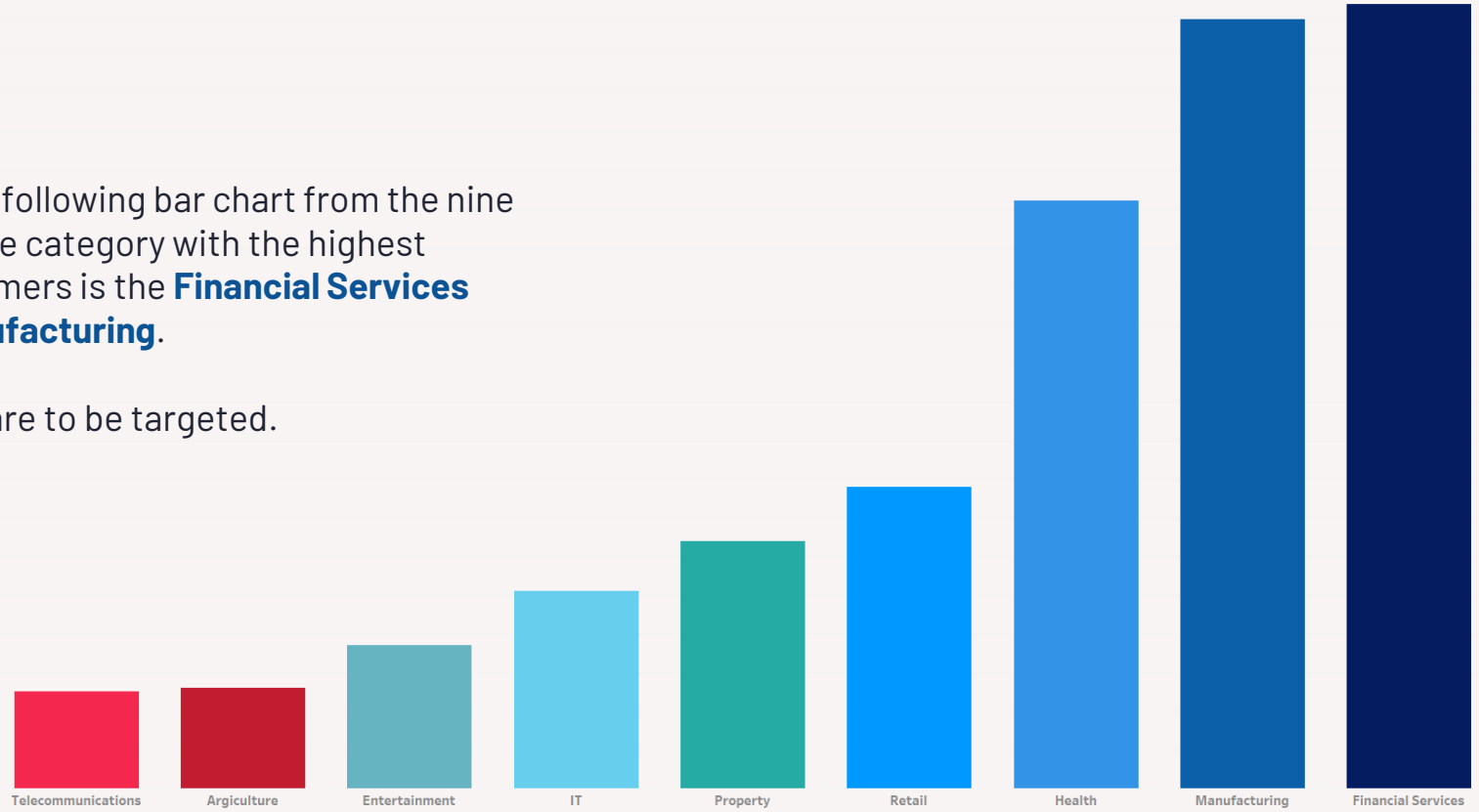
EDA > Wealth Segment

- According to the following bubble chart from the three wealth segments the customers with the highest number is the **Mass Customer**.
- Therefore, they are to be targeted.



EDA > Job Category

- According to the following bar chart from the nine job categories the category with the highest number of customers is the **Financial Services** followed by **Manufacturing**.
- Therefore, they are to be targeted.



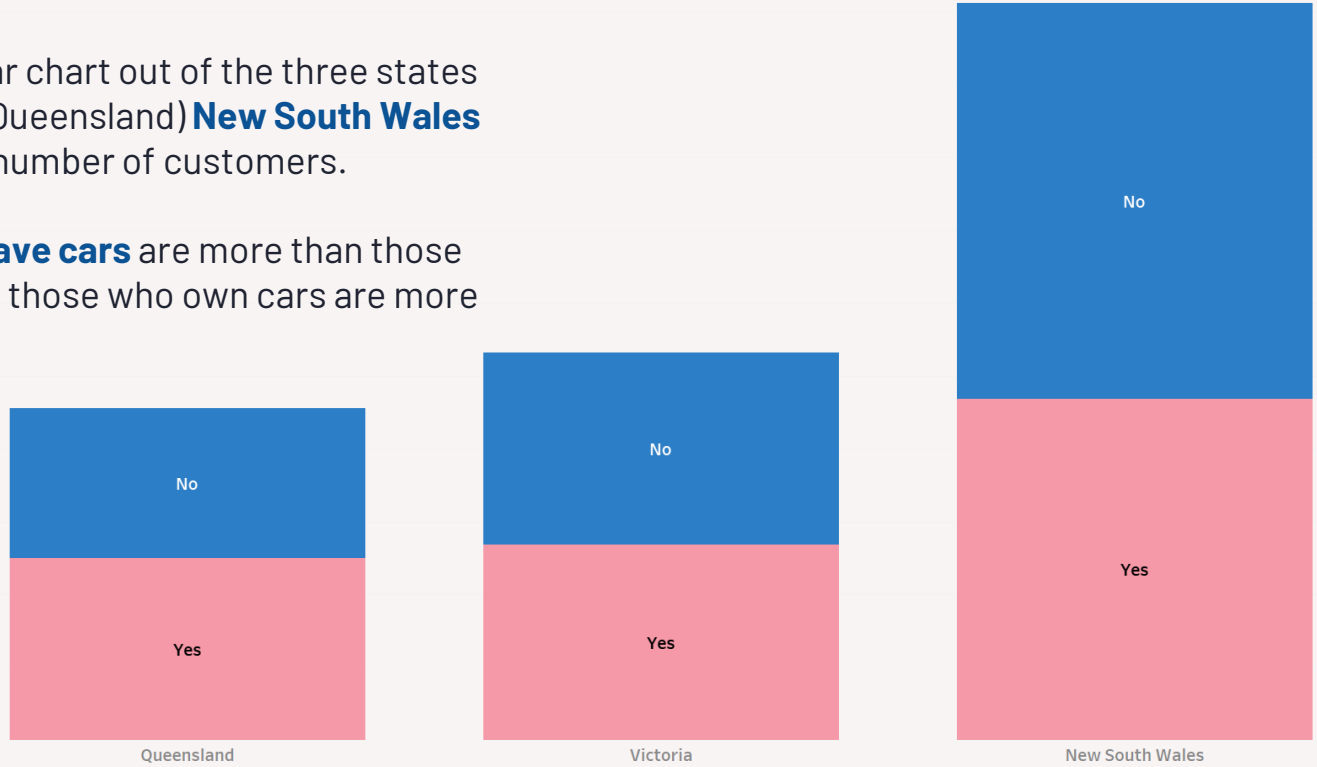
EDA > Location

- According to the following map from the **three states** (Queensland, New South Wales, Victoria) in Australia **New South Wales** is the highest state in customers.

- Therefore, customers from New South Wales are to be targeted.

EDA > Car Ownership

- According to the following bar chart out of the three states (New South Wales, Victoria, Queensland) **New South Wales** is the state with the highest number of customers.
- The customers who **do not have cars** are more than those who own cars in the NSW but those who own cars are more in the other two States.



THANKS!