

TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

Analytics, Information & Modelling team - Kathleen,
Jr. Consultant – Deep Patel.

Agenda

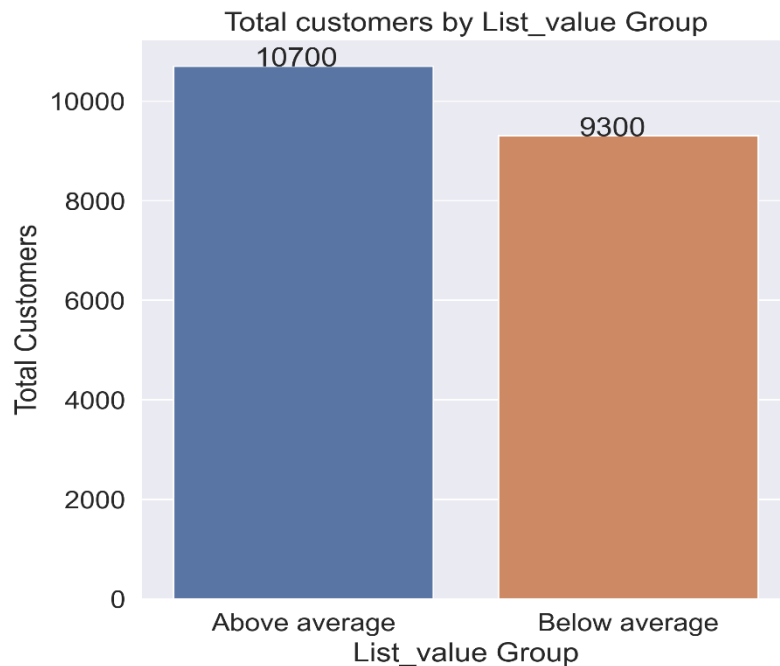
1. Introduction
2. Data Exploration
3. Model Development
4. Interpretation

Introduction

Focus Group: high-valued past customers

Out of the past customers data, we will only focus on observing the upper-half group based on Listed Price of their purchases.

We will only be focusing on 10,700 considering we do not want to spend on marketing, if couldn't get more back than average.



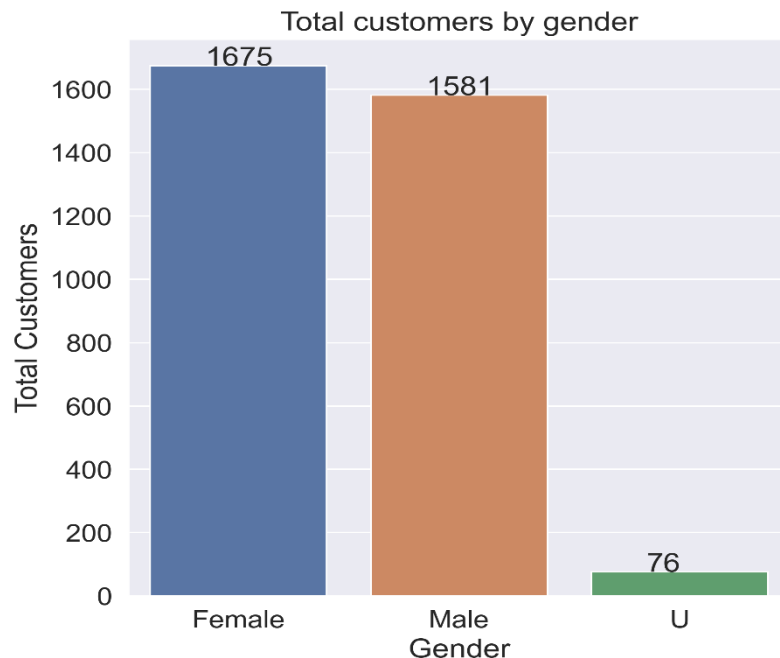
Data Exploration

Explore basic trends within focused group based on:

1. Gender

Focus more on female for advertising in general, but will have to check for reverse behavior in specific age group if any?

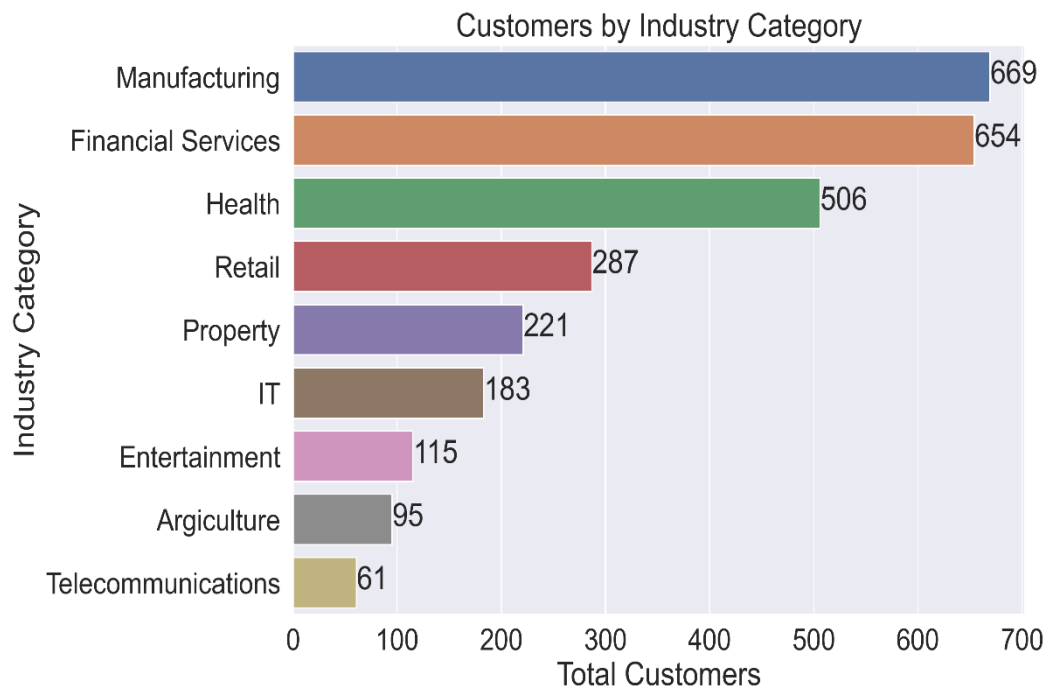
Impute gender in 'U' group maintaining the proportion present in the rest of the dataset.



Data Exploration

2. Industrial Category

Only targeting customers from 3 industry categories: Manufacturing, Financial Services, and Health for marketing will cover the major proportion of the customers

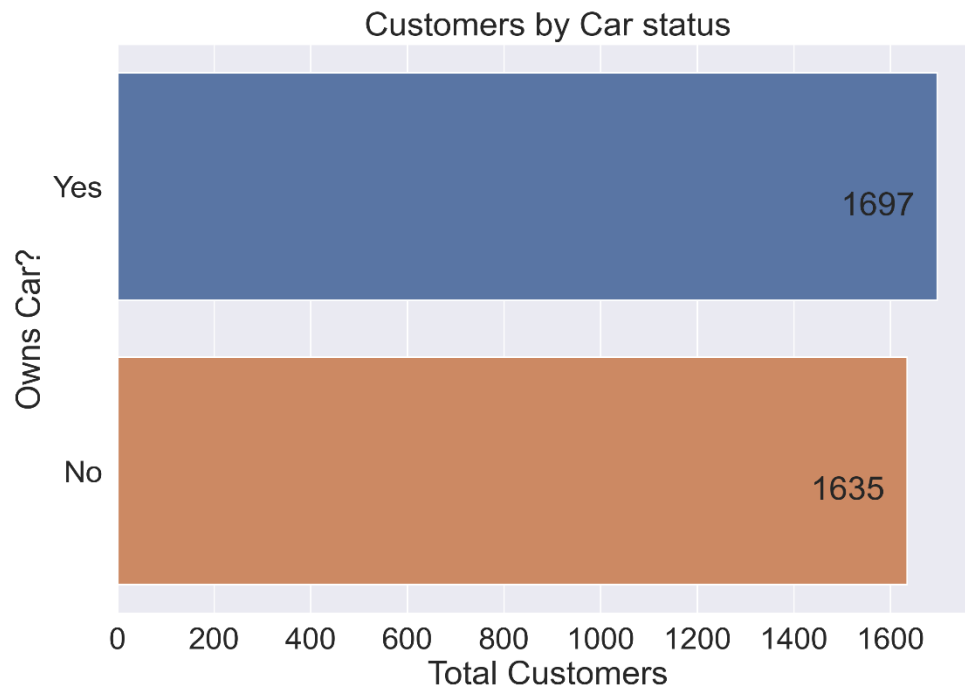


Data Exploration

3. Owning a car

Hypothesis: Customers having 2 or more cars are less likely to buy bikes and accessories parts.

Well, from it is opposite than what we thought. Owning a car or not merely makes a difference.



Data Exploration

4. State customers live in

We should target the customers from a particular state offering highest value to business for ensuring success of marketing efforts.

We will also check if we get any other advantage for targeting the particular state.



Model Development

1. Basic cleaning

Observing the 'Deceased Indicator' being same for all 1000 customers, will not play any role in our analysis.

With our decision to target customers from state 'NSW', we do not need customer's addresses.

There are 5 columns having no details regarding their values and are Unnamed. Until n unless we do not know about them, we should not include them.

```
df["deceased_indicator"].value_counts()
```

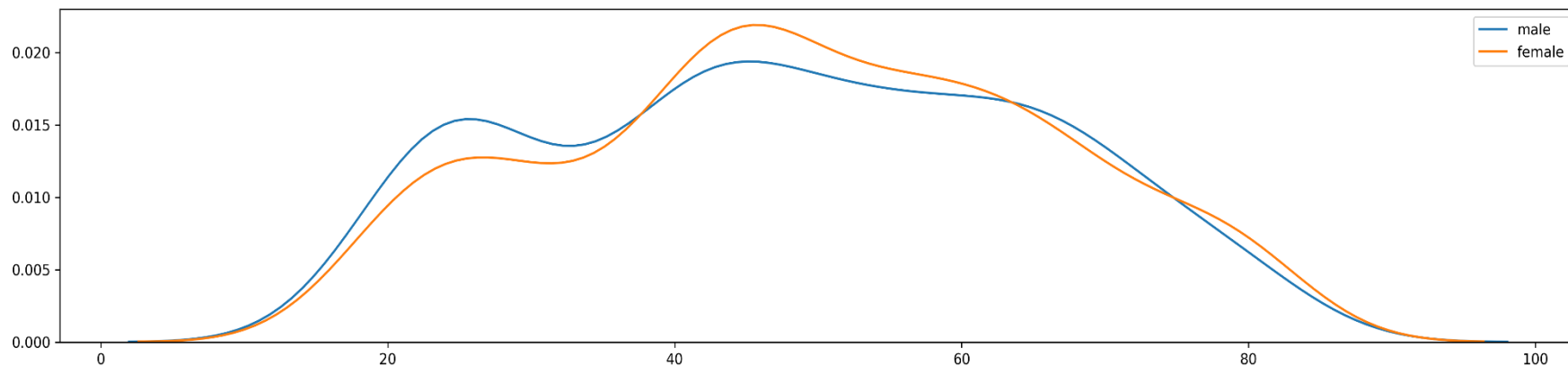
```
N      1000  
Name: deceased_indicator, dtype: int64
```

Unnamed: 16	Unnamed: 17	Unnamed: 18	Unnamed: 19	Unnamed: 20
0.56	0.7000	0.8750	0.743750	1
0.89	0.8900	1.1125	0.945625	1
1.01	1.0100	1.0100	1.010000	1

Model Development

2. Counting age and proper imputing against missing DOBs

Counted age for the customers and found the age group '25-35' having lesser observations. To maintain the distribution, we will impute ages for 17 customers with missing DOBs such that this age group receives justification.



Model Development

3. Data Imputation for Gender

Surprising Facts:

- Customer records having missing DOB and Gender are the same.
- ABS: Male and female participation rates were similar, except in the 25-34 age group where participation rates were higher for males (67%) than females (61%).

Utilization:

- We have imputed values for age for missing DOB to help the age group '25-34'. Now, drop 'DOB' column.
- We will impute gender against 'U' in 'gender' column maintaining the approximate 70:30 ratio for Male: Female.

	gender	age
59	U	NaN
226	U	NaN
324	U	NaN
358	U	NaN
360	U	NaN
374	U	NaN
434	U	NaN
439	U	NaN
574	U	NaN
598	U	NaN
664	U	NaN
751	U	NaN
775	U	NaN
835	U	NaN
883	U	NaN
904	U	NaN
984	U	NaN

Model Development

4. Industry Category & Job Title

- 165 records miss Job Industry Category and 106 miss Job Titles of the customers
- First, we impute the Job Industry Category by maintaining the ratio present in the remaining of the dataset.
- Second, for job titles, we impute titles based on the ratio of job titles present in the remaining of the dataset corresponding to the same Industry Category as of customer with missing title.

Interpretation

Based on our findings in Data exploration, we will eliminate the customers not fitting our targeting criteria and get the list of customers to be targeted.

We will create a separate customer list out of the 1000 new customers applying following criteria:

- Age range: 15-65
- More Males for age group '25-34'
- Should be living in NSW
- Industry category should be Manufacturing or Health of Financial services
- Should be a 'Mass' customer