Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

[Division Name] - [Engagement Manager], [Senior Consultant], [Junior Consultant]

Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

Agenda

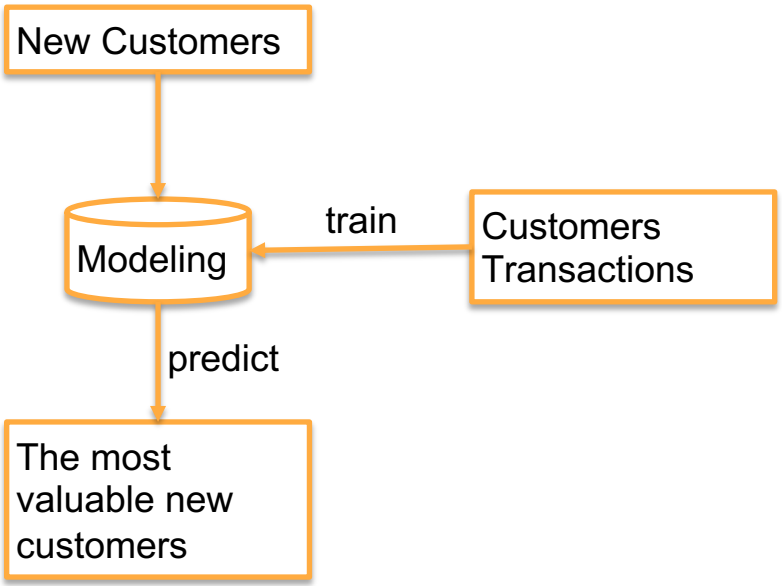
1. Introduction
2. Data Exploration
3. Model Development

Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

Introduction

Goal:

Recommend the most valuable customers from the 1000 new customers.



Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

Introduction

Two Tables:

- **Transactions**
- Customers

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 19445 entries, 0 to 19999
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   transaction_id          19445 non-null  int64
 1   product_id              19445 non-null  int64
 2   customer_id             19445 non-null  int64
 3   transaction_date        19445 non-null  datetime64[ns]
 4   online_order            19445 non-null  float64
 5   order_status            19445 non-null  object
 6   brand                   19445 non-null  object
 7   product_line            19445 non-null  object
 8   product_class           19445 non-null  object
 9   product_size            19445 non-null  object
 10  list_price              19445 non-null  float64
 11  standard_cost           19445 non-null  float64
 12  product_first_sold_date 19445 non-null  datetime64[ns]
 13  margins                 19445 non-null  float64
dtypes: datetime64[ns](2), float64(4), int64(3), object(5)
memory usage: 2.2+ MB
```

Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

Introduction

Two Tables:

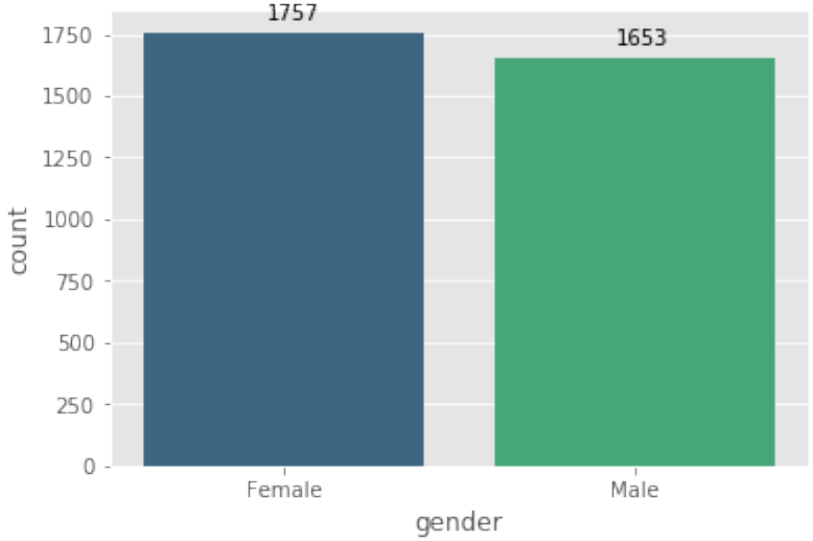
- Transactions
- **Customers**

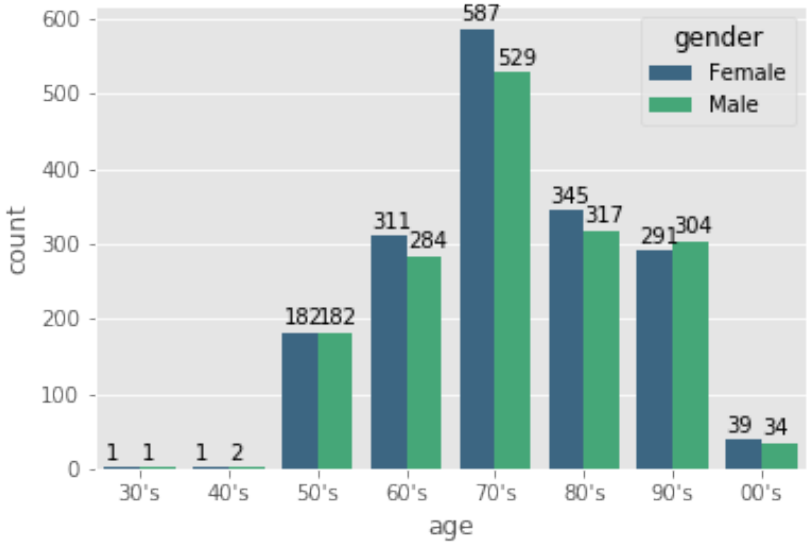
#	Column	Non-Null Count	Dtype
0	customer_id	3908 non-null	int64
1	gender	3908 non-null	object
2	past_3_years_bike_related_purchases	3908 non-null	int64
3	wealth_segment	3908 non-null	object
4	owns_car	3908 non-null	object
5	tenure	3908 non-null	int64
6	address	3908 non-null	object
7	postcode	3908 non-null	float64
8	state	3908 non-null	object
9	property_valuation	3908 non-null	float64
10	age	3908 non-null	category

Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

Data Exploration

Is data skewed by gender or age?

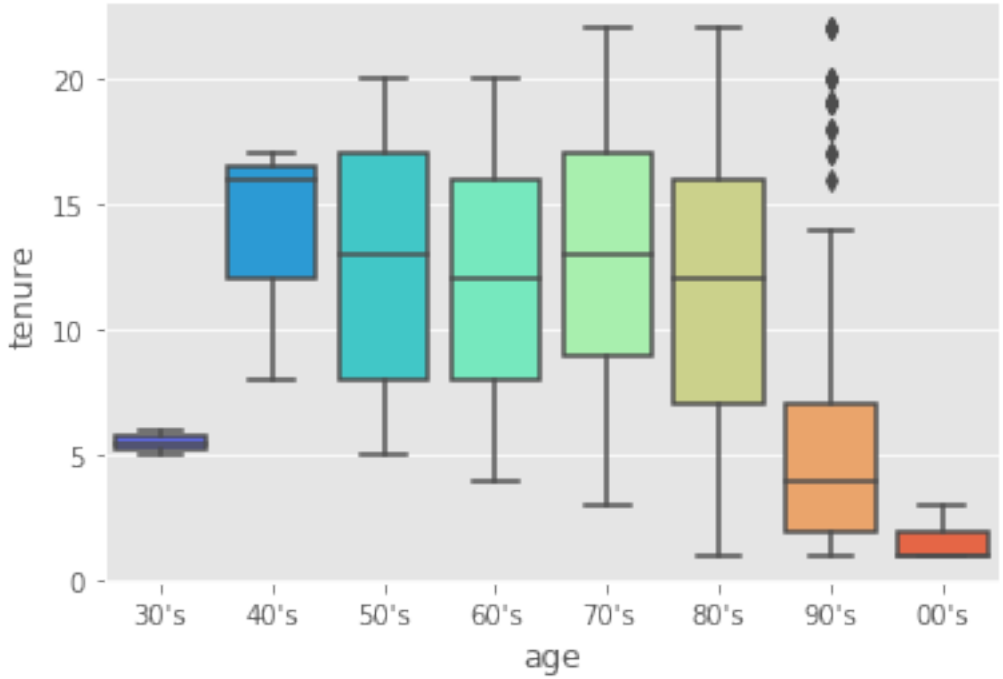




Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

Data Exploration

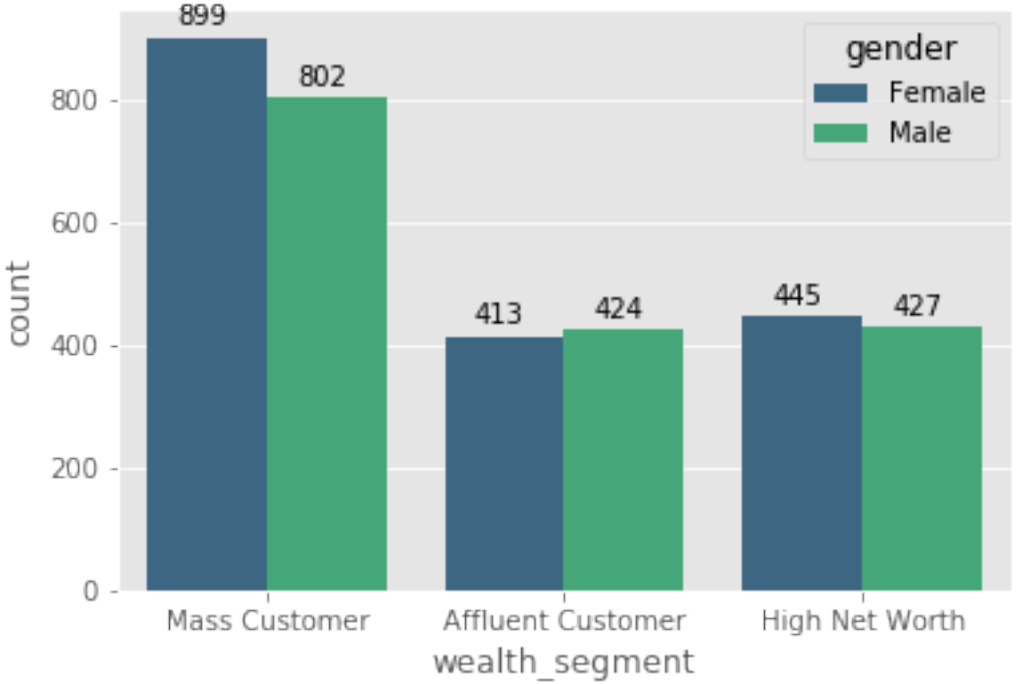
tenure and age



Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

Data Exploration

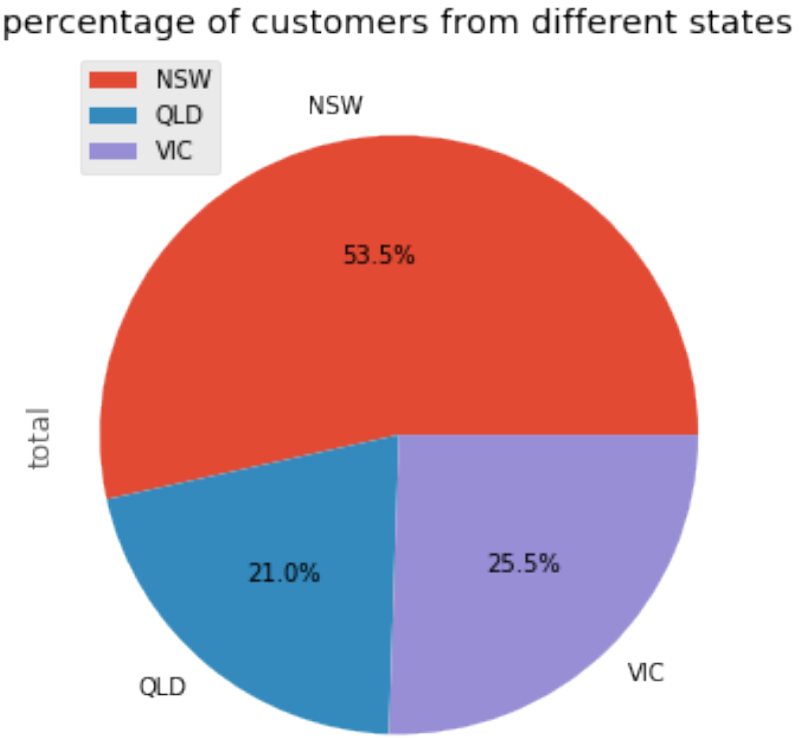
Segment of wealth



Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

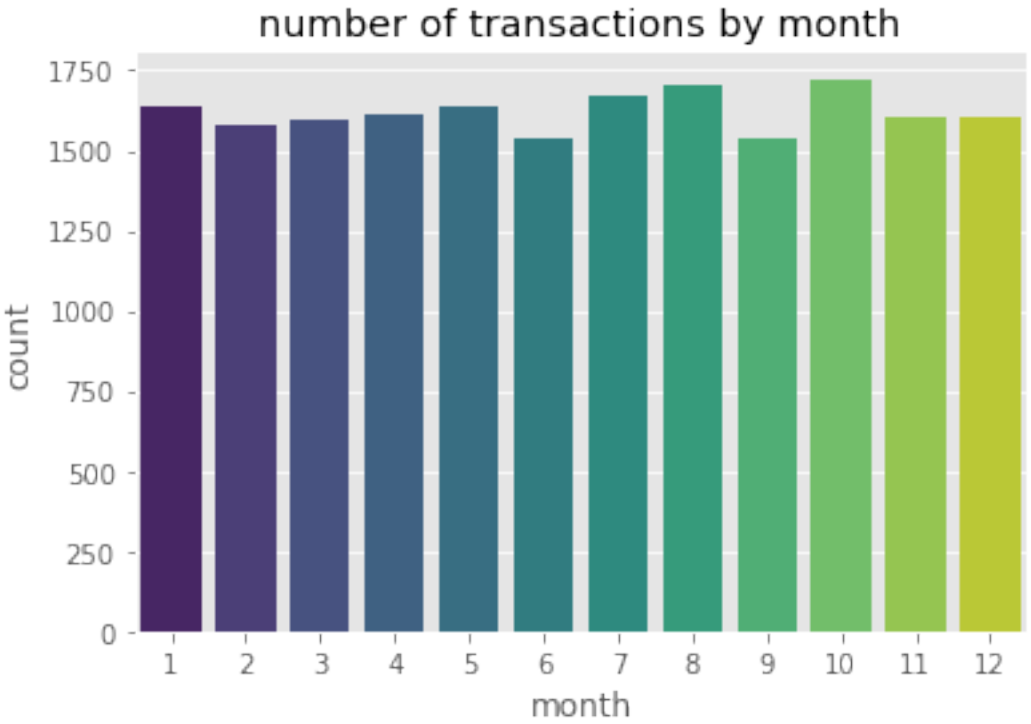
Data Exploration

States where people from



Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

Data Exploration

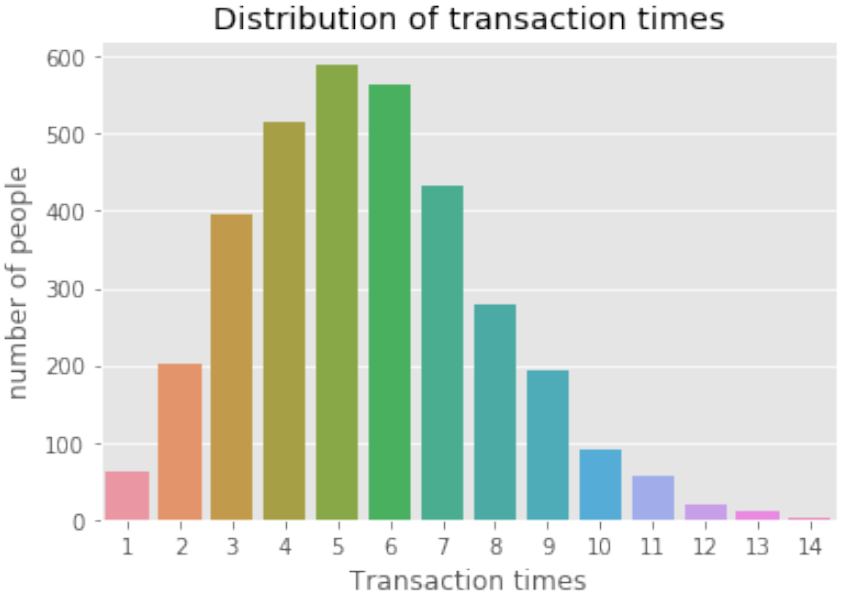


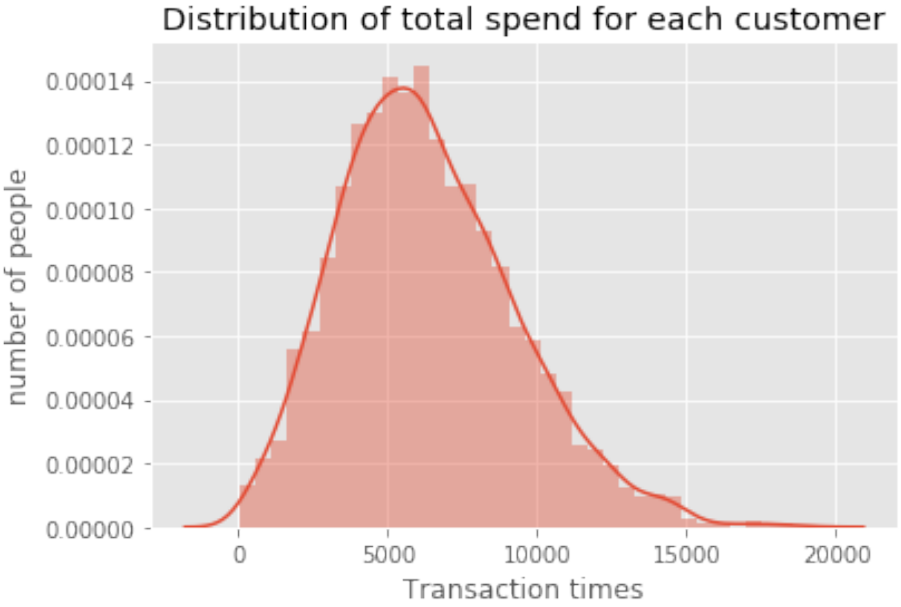
Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

Model Development

Labels

valuable: transaction time >= 6 and total spend >=6000





Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

Model Development

Features:

'gender', 'past_3_years_bike_related_purchases', 'wealth_segment', 'owns_car', 'tenure', 'state', 'property_valuation', 'age'

19.6% missing
14.5% missing

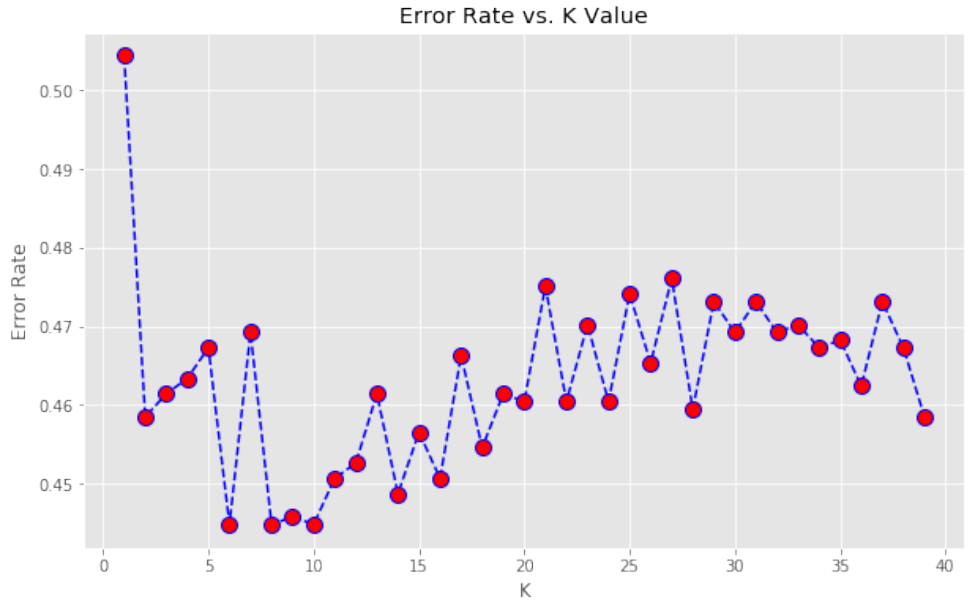
customer_id	0
first_name	0
last_name	125
gender	0
past_3_years_bike_related_purchases	0
DOB	87
job_title	506
job_industry_category	656
wealth_segment	0
deceased_indicator	0
owns_car	0
tenure	87
address	4
postcode	4
state	4
country	4
property_valuation	4

dtype: int64

Further work:
1) job_title and job_industry_category
2) Remoteness of areas

Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

Model Development



WITH K=6

```
[[472 113]
 [342  96]]
```

	precision	recall	f1-score	support
0	0.58	0.81	0.67	585
1	0.46	0.22	0.30	438
accuracy			0.56	1023
macro avg	0.52	0.51	0.49	1023
weighted avg	0.53	0.56	0.51	1023