

Finding Potential Customer Opportunities

- Analyzing the Evolution of Power Purchasing

CADS Capstone Project 2022

Group 4 Cohort 6:

Asyraaf

Aqil

Najwa

Humaira

Aisyah

Analytics Wisdom

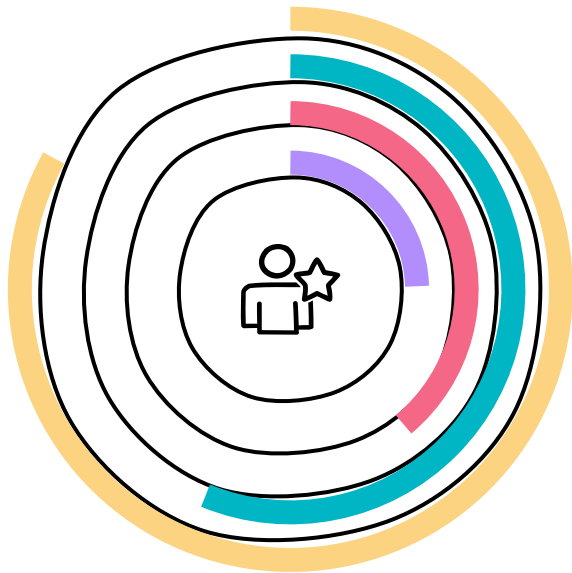
01	Overview	<ul style="list-style-type: none">• What the project is about?
02	Project Niche	<ul style="list-style-type: none">• Objective & Dataset• Flow Chart Machine Learning in Python
03	Problem Statement	<ul style="list-style-type: none">• The Concern of Project
04	Deep Analysis	<ul style="list-style-type: none">• Algorithms / Findings / Insights• Solution
05	Upswing	<ul style="list-style-type: none">• Recommendation• Q&A• Improvements

PROJECT CONTENT NOTIFIER

01

Encapsulation

Project Overview



Encapsulation



The growth of the human population, in line with the rapid pace development of technology in this world, require humans to have assets as a necessity of life. One of them is the ability to obtain shelters, which are houses.



Supposedly, the real estate business should gain benefits from this sub-conscious occasion.



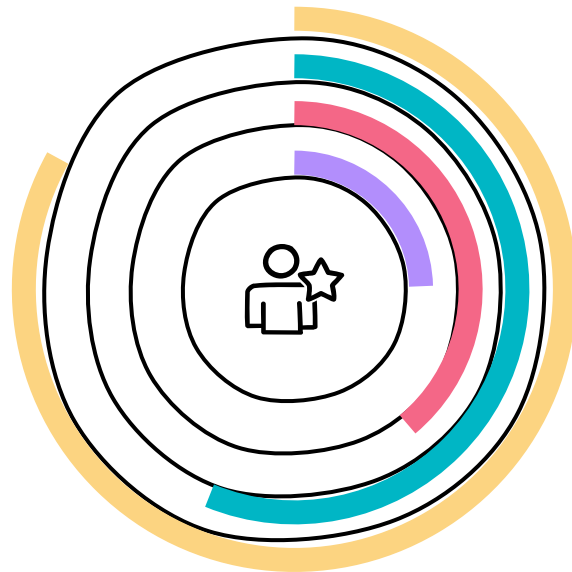
Thus, to enhance the company's constant business progress, the **analysis of acquiring potential customers as buyers of property** is subjected to **detailed scrutiny**.

PROJECT CONTENT NOTIFIER

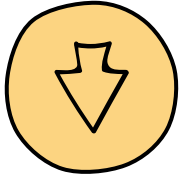
02

Project Niche

Objective, Dataset, Flow



Project Niche



Objective

Company is looking to gain understanding and increase potential customers from their customers' purchase history

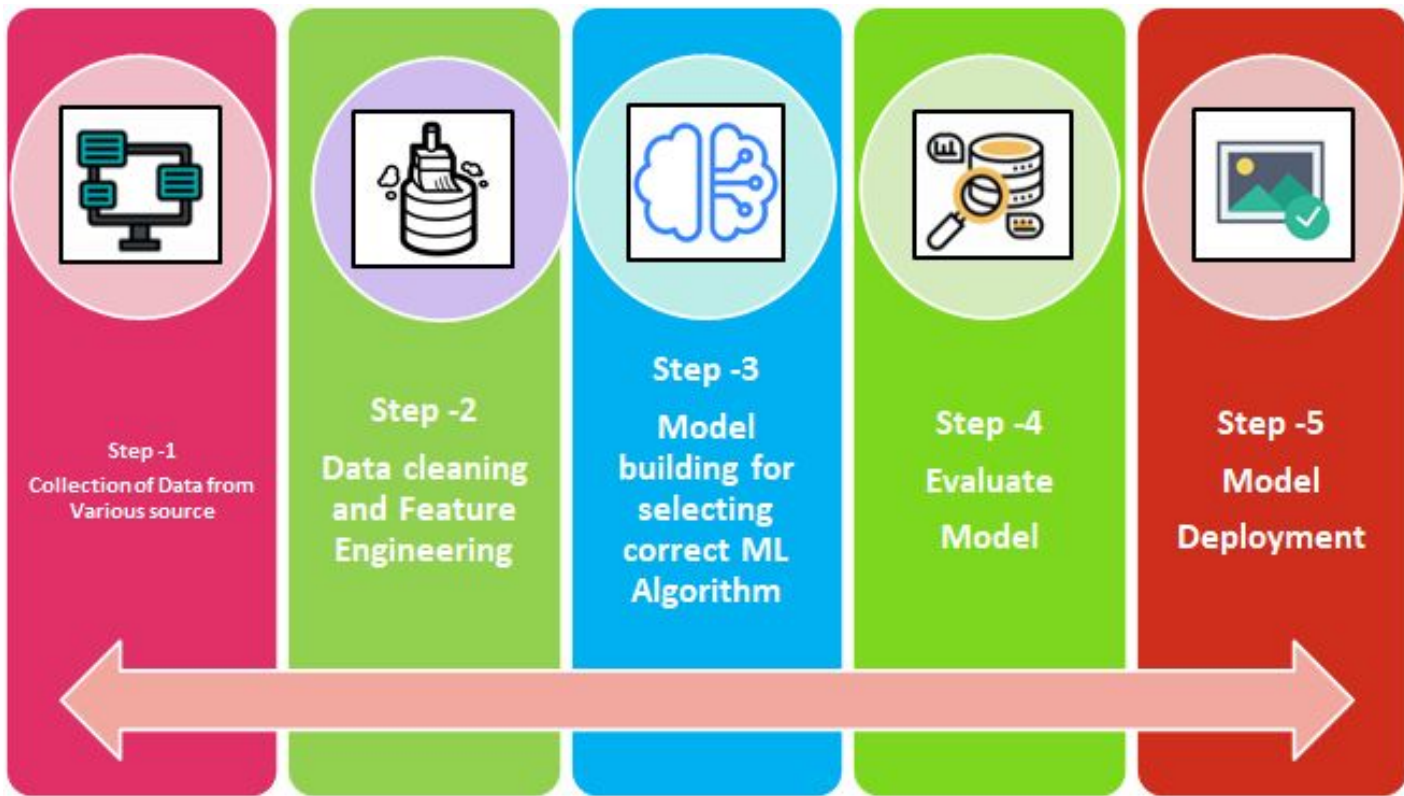


Dataset

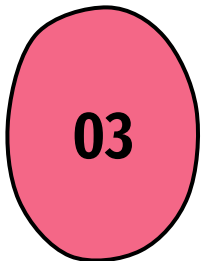
To analyse the customers' purchase history data based on their features, and to focus on their tendency as potential customer.

Dataset obtained contains such variables ;

Understanding Machine Learning in Python

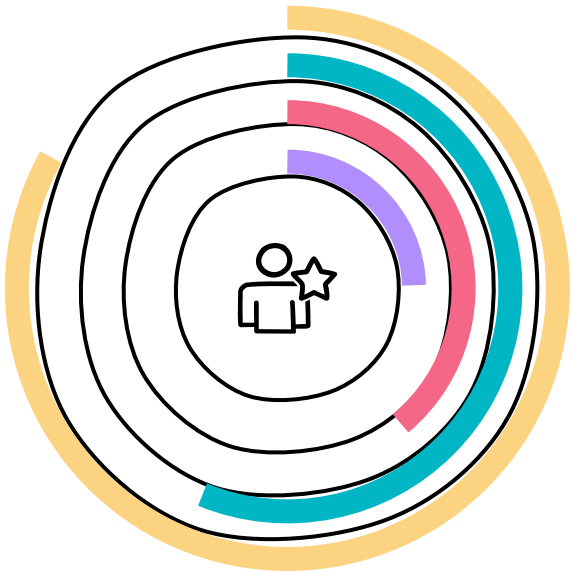


PROJECT CONTENT NOTIFIER



Problem Statement

Related to Project Focus



Problem Statement



To ensure company secure or gain potential customers, the company should **analyse the engagement of customers to their marketing** campaign towards their **customer purchase potential**.



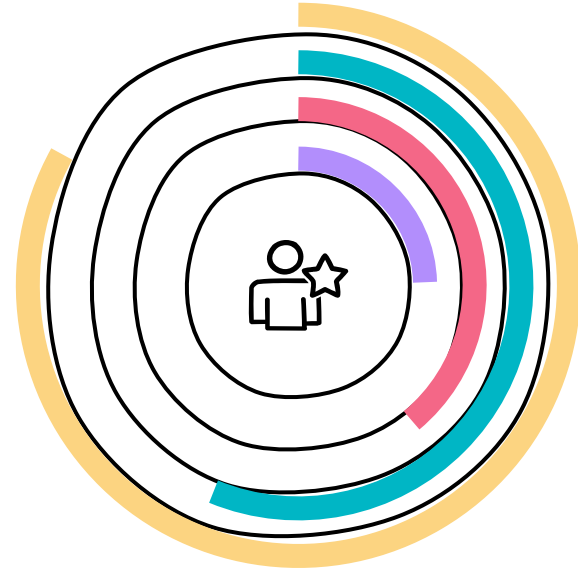
To **analyse** the **company's potential customer** based on the **significant features** of the customers.

PROJECT CONTENT NOTIFIER

04

Deep Analysis

Algorithms, Insights, Solutions



Data Wrangling, Cleaning & Merging Feature Engineering & Selection

Dataset Summary



```
[5] cust_data.shape
```

```
(4469, 25)
```

```
[6] cust_data.info()
```

```
##Features == Column = 25
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4469 entries, 0 to 4468
Data columns (total 25 columns):
 #   Column                Non-Null Count  Dtype  
---  --
 0   Potential_Customer    4469 non-null   int64  
 1   C_ID                  4469 non-null   int64  
 2   Cust_Last_Purchase    2149 non-null   object  
 3   Pur_3_years           4469 non-null   int64  
 4   Pur_5_years           4469 non-null   int64  
 5   Pur_3_years_Indirect  4469 non-null   int64  
 6   Pur_5_years_Indirect  4469 non-null   int64  
 7   Pur_latest            4469 non-null   object  
 8   Pur_3_years_Avg       4469 non-null   object  
 9   Pur_5_years_Avg       4469 non-null   object  
10   Pur_3_years_Avg_Indirect 3642 non-null   object  
11   InAct_Last            4469 non-null   int64  
12   InAct_First           4469 non-null   int64  
13   Ad_Res_1_year         4469 non-null   int64  
14   Ad_Res_3_Year         4469 non-null   int64  
15   Ad_Res_5_Year         4469 non-null   int64  
16   Ad_Res_Ind_1_Year     4469 non-null   int64  
17   Ad_Res_Ind_3_Year     4469 non-null   int64  
18   Ad_Res_Ind_5_Year     4469 non-null   int64  
19   Status_Cust           4469 non-null   object  
20   Status_Latest_Ad      4469 non-null   int64  
21   Age                   3510 non-null   float64 
22   Gender                4469 non-null   object  
23   Cust_Prop             4469 non-null   object  
24   Cust_Ann_Income       4469 non-null   object  
dtypes: float64(1), int64(15), object(9)
memory usage: 873.0+ KB
```

Original Data

Consists of 4469 entries and 25 columns

Features Description

S.No	Variable	Description
1	Potential_Customer	Response Variable
2	C_ID	Customer Identification Number
3	Cust_Last_Purchase	Amount purchased in \$ (most latest / recent purchase)
4	Pur_3_years	No of purchases made in the recent 3 years
5	Pur_5_years	No of purchases made in the last five years
6	Pur_3_years_Indirect	No of purchases made in 3 years through link from other websites (indirect buys)
7	Pur_5_years_Indirect	No of purchases made in the last five years through link from other websites
8	Pur_latest	The latest purchase amount (in thousands)
9	Pur_3_years_Avg	Average Purchase over the last 24 months
10	Pur_5_years_Avg	Average Purchase over the last 5 years
11	Pur_3_years_Avg_Indirect	Average Indirect Purchase through link from other sources for the last 24 months
12	InAct_Last	Inactive no of months since the customers made the last purchase
13	InAct_First	Inactive no of months since the customers made the first purchase
14	Ad_Res_1_year	No of Promotional Ads by MyPurchase responded by the customer online in the last 1 year
15	Ad_Res_3_Year	No of Promotional Ads responded by the customer online in the last 3 years
16	Ad_Res_5_Year	No of Promotional Ads responded by the customer online in the last 5 years
17	Ad_Res_Ind_1_Year	No of Ads responded to the other sources (indirect) which directed to MyPurchase in the last 1 year
18	Ad_Res_Ind_3_Year	No of Calls made by References to the individual for the last 36 months
19	Ad_Res_Ind_5_Year	No of Calls made by References to the individual over the period of few years
20	Status_Cust	A if active buyer, S if star buyer, N if new buyer, E if inactive buyer, F if first time buyer, L if lapsing buyer
21	Status_Latest_Ad	1 if individual has purchased in response to the last promotional sale, 0 if not
22	Age	Age of the individual
23	Cust_Prop	Owns a House H- Owner / U-Unknown
24	Gender	Sex of the individual
25	Cust_Ann_Income	Customer Annual Income

Dataset Summary



Data Cleaning and Merging

- Removed duplicated data
- Filled missing values
- Changed datatype for analysis
- Combined and drop columns and selected important features

Remove duplicate rows

Hint: Use `df=df.drop_duplicates()`

```
[15] # Remove duplicates
```

```
cust_data=cust_data.drop_duplicates()
```

```
cust_data
```

Import necessary Packages

```
# Importing relevant packages
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings("ignore")

#Your Code Here
from IPython.display import display
```

```
[150] cust_data_new = cust_data_new.drop(['Pur_3_years', 'Pur_5_years', 'Pur_3_years_Indirect', 'Pur_5_years_Indirect', 'Ad_Res_1_year', 'Ad_Res_3_Year', 'Ad_Res_5_Year', 'Ad_Res_Ind_1_Year', 'Ad_Res_Ind_3_Year', 'Ad_Res_Ind_5_Year'], axis = 1)
```

Dataset Summary



Results

- 3618 unique entries
- Combined features
 - Total Purchase for 3 Year = Direct and Indirect Purchase for 3 years
 - Total Purchase for 5 Year = Direct and Indirect Purchase for 5 years
 - Total Ads Response for particular year range = Direct and Indirect Ads Response for particular year range

```
cust_data_new.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 3618 entries, 0 to 3617
Data columns (total 21 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Potential_Customer                    3618 non-null   object
1   C_ID                                  3618 non-null   int64
2   Cust_Last_Purchase                    1736 non-null   float64
3   Pur_latest                            3618 non-null   float64
4   Pur_3_years_Avg                       3618 non-null   float64
5   Pur_5_years_Avg                       3618 non-null   float64
6   Pur_3_years_Avg_Indirect              2956 non-null   float64
7   InAct_Last                            3618 non-null   int64
8   InAct_First                           3618 non-null   int64
9   Status_Cust                           3618 non-null   string
10  Status_Latest_Ad                      3618 non-null   int64
11  Age                                    2825 non-null   float64
12  Gender                                3618 non-null   category
13  Cust_Prop                              3618 non-null   category
14  Cust_Ann_Income                       3618 non-null   float64
15  Status_CustN                           3618 non-null   object
16  Total_Pur_3_years                      3618 non-null   int64
17  Total_Pur_5_years                      3618 non-null   int64
18  Total_Ad_Res_1_year                   3618 non-null   int64
19  Total_Ad_Res_3_year                   3618 non-null   int64
20  Total_Ad_Res_5_year                   3618 non-null   int64
dtypes: category(2), float64(7), int64(9), object(2), string(1)
memory usage: 732.6+ KB
```

EDA

Analysis Distribution

Findings



Marketing Department need **ASAP** action!



Total of Non-Potential Customers is **higher** compared to the **Potential Customers**.

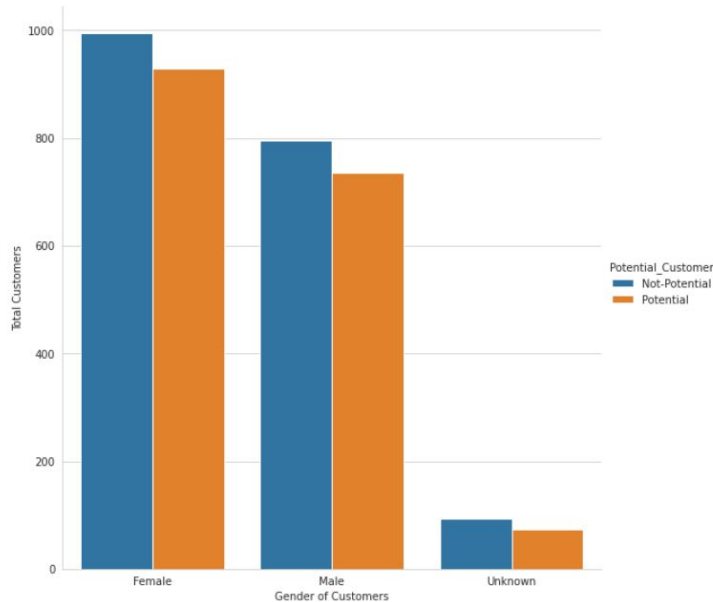
Current potential customer does **not meet the company's profiting goal**.

Findings



FEMALE CUSTOMER **conquer** as Non-Potential Customer

Relationship of Gender and Potential Customer

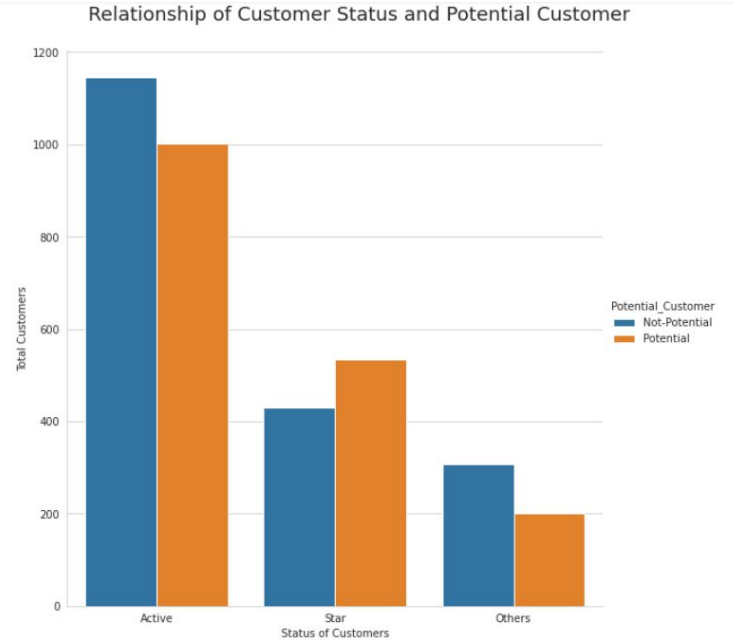


Female customer as **major potential** and **non-potential** customers compared to male gender.

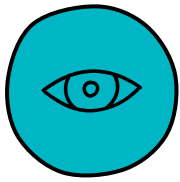
Findings



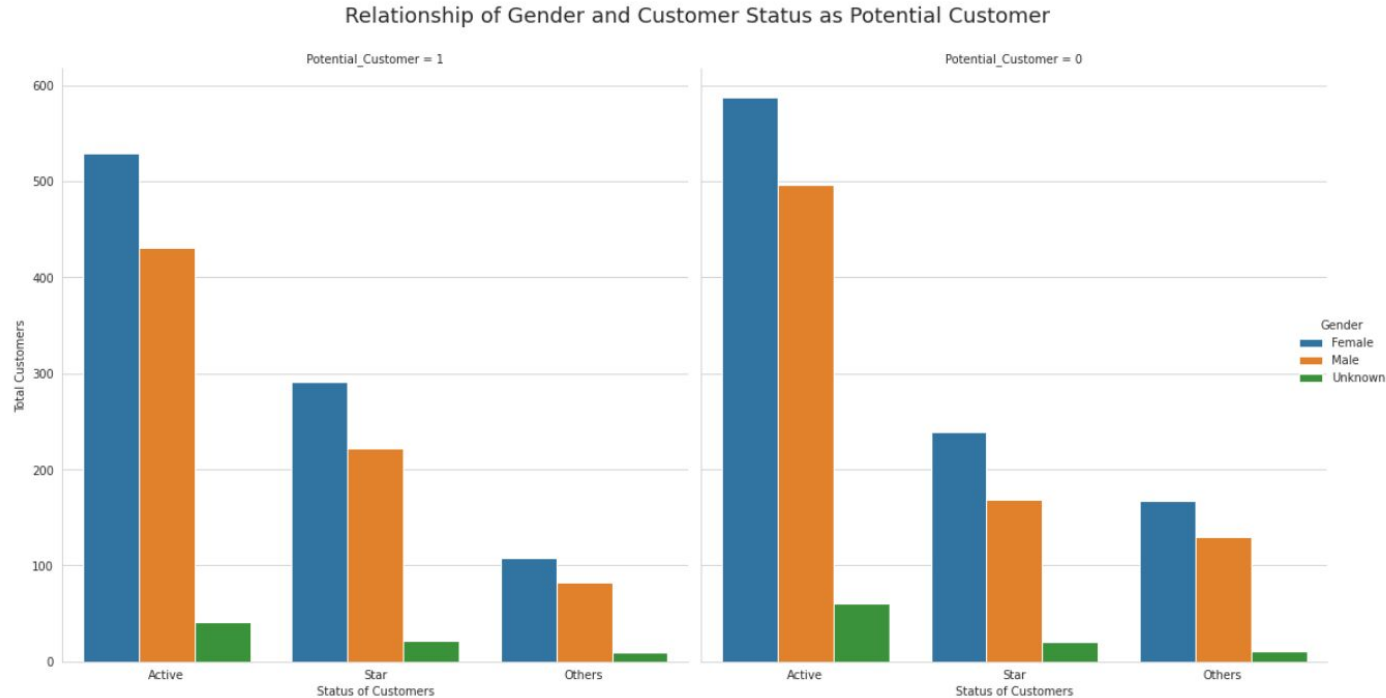
Company should **offers** more **Special benefits** to **Active Buyer** to be **Potential**



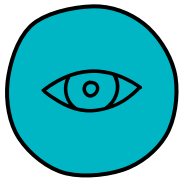
Most active customers is a Not-Potential Customer but star customers for potential customer is higher than potential customers.



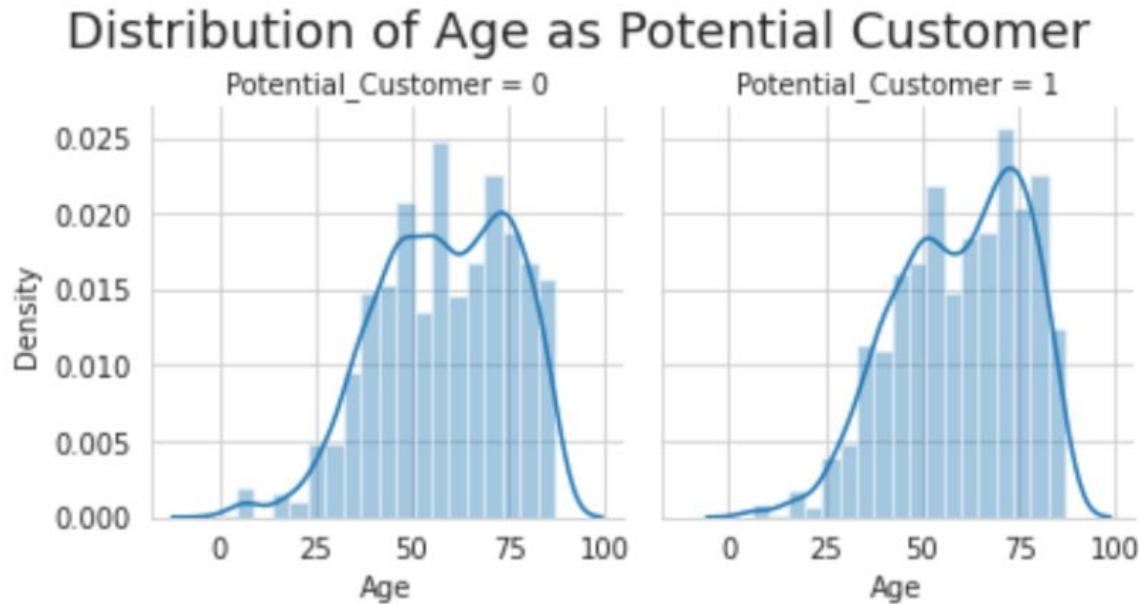
Non-Potential Customer appears to be **more active** than Potential



Female customer has higher active and star customer status compared to Male Customer. However, they also mostly prone to be a non-potential customer.

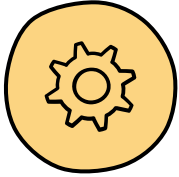


Young People show **interest** to be **Potential** Customer



Left skew of the graph distributions shows the potential customers of purchasing property attracts young-adult demographic customers.

Findings



(+) **Corr** between Purchasing and Total Ads Responded

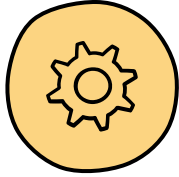
```
[154] cust_data_new[["Total_Ad_Res_3_year", "Total_Pur_3_years", "Pur_3_years_Avg"]].corr()
```

	Total_Ad_Res_3_year	Total_Pur_3_years	Pur_3_years_Avg
Total_Ad_Res_3_year	1.000000	0.504730	-0.024606
Total_Pur_3_years	0.504730	1.000000	-0.342615
Pur_3_years_Avg	-0.024606	-0.342615	1.000000

From our correlation study, we found that the **Total Amount of Purchase** from the customers have a **moderately positive correlation** to the **Total Ads Responded within 3 years** (0.504)

This shows that the **marketing campaign have a positive effect on the purchase of customers**, but should also work more to increase stronger positive correlation on average purchase.

Findings



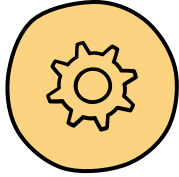
Narrowing down, we can also see that customers' purchase amount are more impacted from Indirect Advertisements from the company compared to Direct Advertisements

This can be shown from the higher correlation between the Customers' Response to Indirect Advertisement against the purchase amount made in 3 years as compared to Customers' response to Direct Advertisement

```
[230] cust_data[["Ad_Res_3_Year", "Ad_Res_Ind_3_Year", "Pur_3_years"]].corr()
```

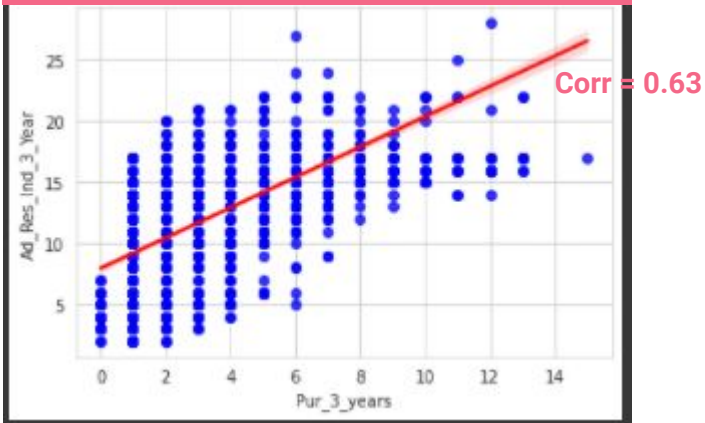
	Ad_Res_3_Year	Ad_Res_Ind_3_Year	Pur_3_years
Ad_Res_3_Year	1.000000	0.624917	0.384991
Ad_Res_Ind_3_Year	0.624917	1.000000	0.593459
Pur_3_years	0.384991	0.593459	1.000000

Findings

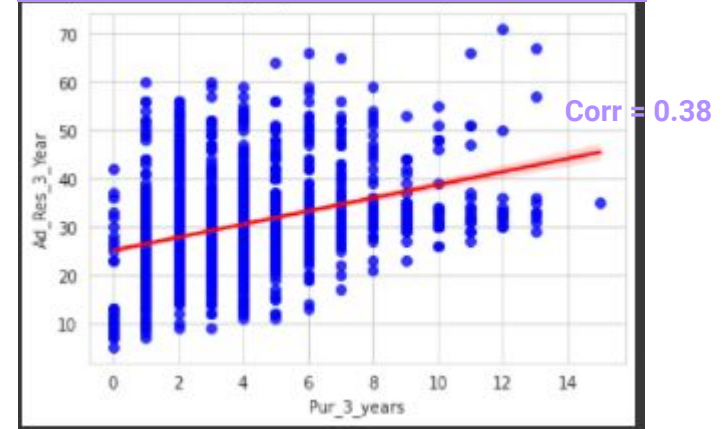


Indirect Advertisements have a stronger positive correlation to the customers amount of purchase compared to **Direct Advertisements**

Response to Indirect Ads



Response to Direct Ads



Solution



Hire a strategic analyst

This role is focused on developing ROI measurements for marketing. Based on our findings, customers high likely engage with indirect advertisement compared to direct advertisement



Focus on top marketing

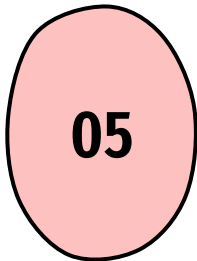
Invest more on indirect advertisement because that are the most source of revenue that the company can retrieve



Strategise on a better targeting of customers

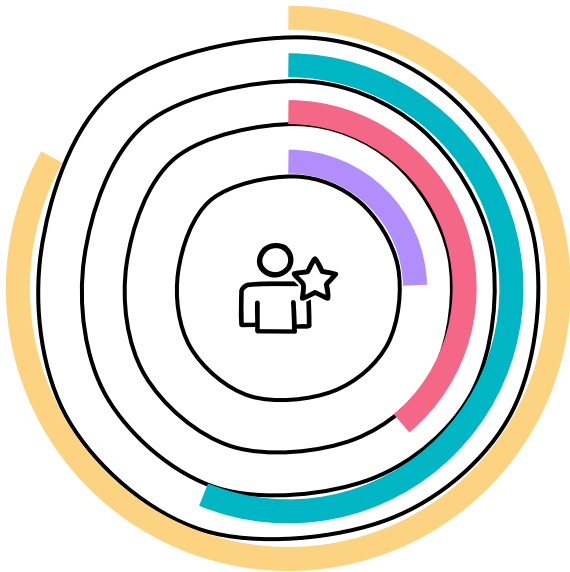
Identifying a target market helps the company develop effective marketing communication strategies. Eg. Star Customer Status, Female Customers.

PROJECT CONTENT NOTIFIER

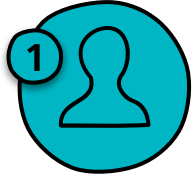


Upswing

Recommendations, Q&A,
Improvements



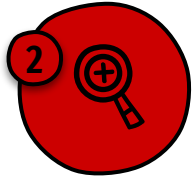
Recommendation



Define Target Audience

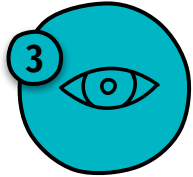
Age, Gender, Income, Location

Do some research to truly **understand the personas** of existing **customers**. Avoid selection bias.



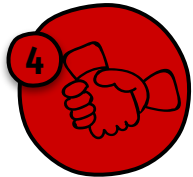
Research Influencers

Find **Influencers** that are famous at the current time so the he/she can **attract more** of their **followers**.



Identifying Social Media Platforms

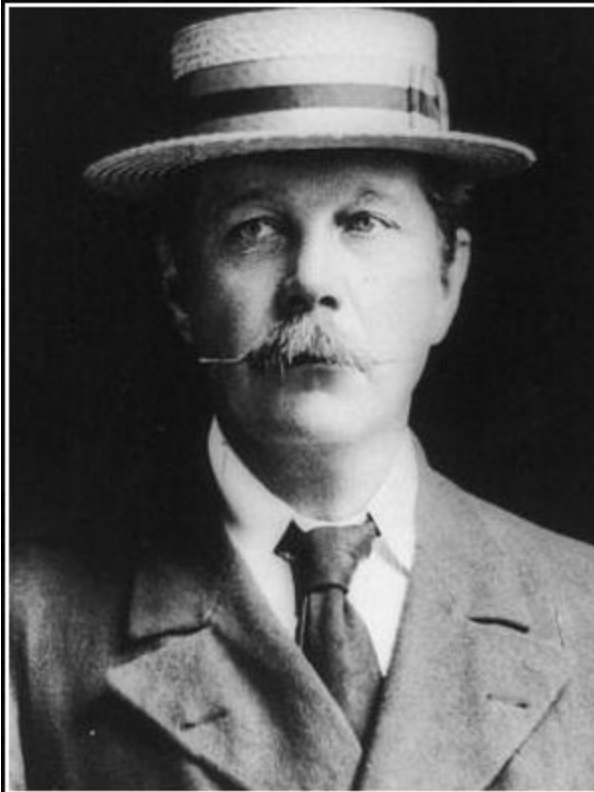
Keep **creating and posting guidance type of content** in social media such as TikTok, etc to educate customers on product. Be **customer experience**.



Collaboration & Human Engagement

Find some local product to do collaboration so that it will be more diversify.

Be more **meaningful in interactions with customers** by acknowledging the challenge faced by customers in purchasing.



[Sherlock Holmes:] The temptation
to form premature theories upon
insufficient data is the bane of our
profession.

— *Arthur Conan Doyle* —

AZ QUOTES

Q&A Time!

