Predict Customer Clicked Ads Classification by Using Machine Learning

Table of Contents

Download Dataset Here

Customer Type and Behaviour Analysis on Advertisement

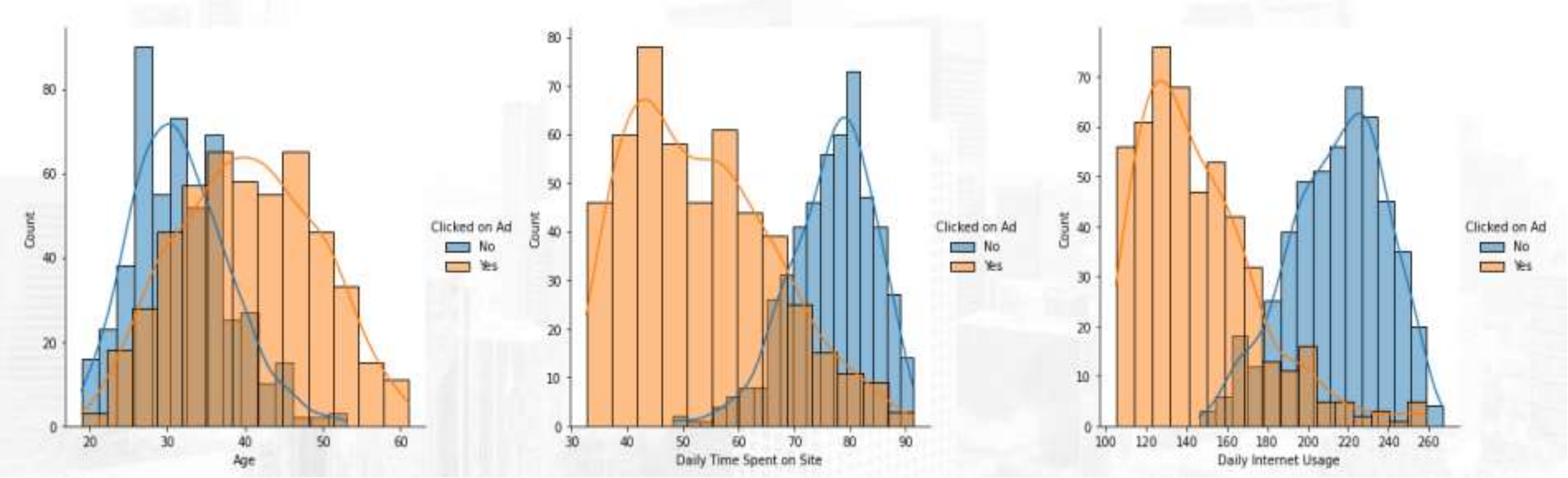

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 11 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Unnamed: 0                1000 non-null   int64
 1   Daily Time Spent on Site  987 non-null    float64
 2   Age                       1000 non-null   int64
 3   Area Income               987 non-null    float64
 4   Daily Internet Usage      989 non-null    float64
 5   Male                      997 non-null    object
 6   Timestamp                 1000 non-null   object
 7   Clicked on Ad             1000 non-null   object
 8   city                      1000 non-null   object
 9   province                  1000 non-null   object
 10  category                  1000 non-null   object
dtypes: float64(3), int64(2), object(6)
memory usage: 86.1+ KB
```

The data consists of 1000 lines

There are several missing value in the Daily Time column Spend on Site, Income Area , Daily Internet Usage, and Gender (Male/Female)

For details, see jupyter notebook here

Univariate Analysis



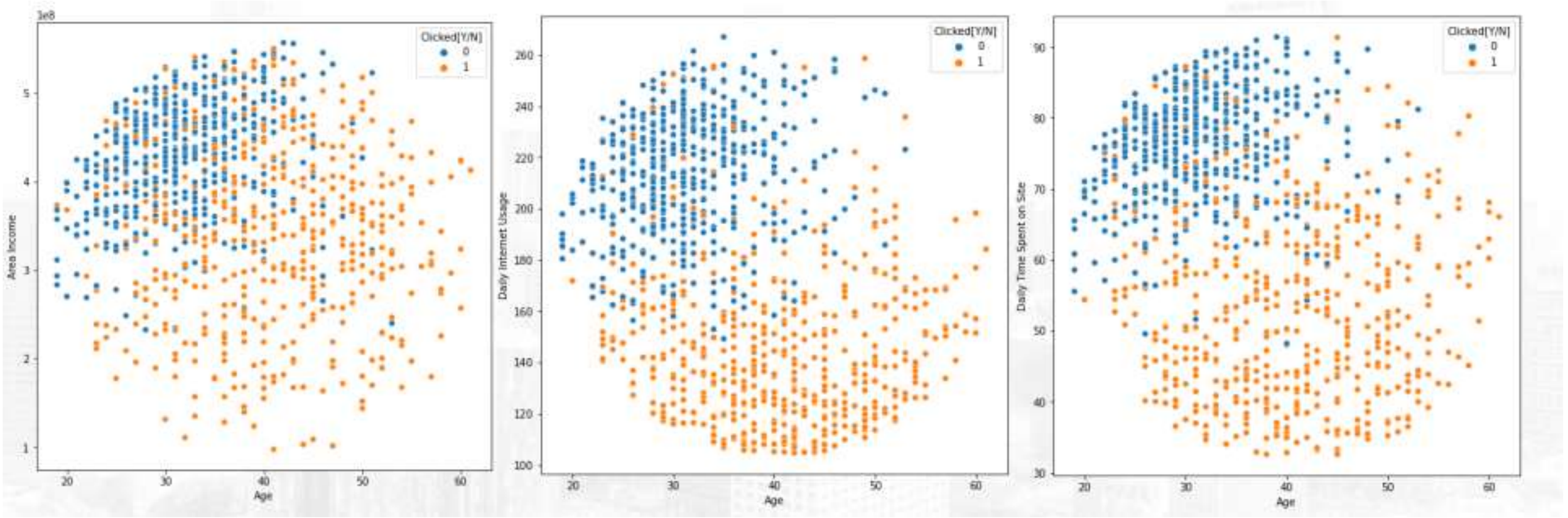
The age range of internet users is 20 - 60 years old, and people aged 40 and over tend to click on advertisements served compared to the age below

Internet users' usage time of less than 70 minutes per day tends to click on advertisements. Person -it is possible for people like this to find the information needed according to the advertisement needed.

People with relatively small internet usage tend to be more effective at clicking on ads than those with make great use of

For details, see jupyter

Bivariate

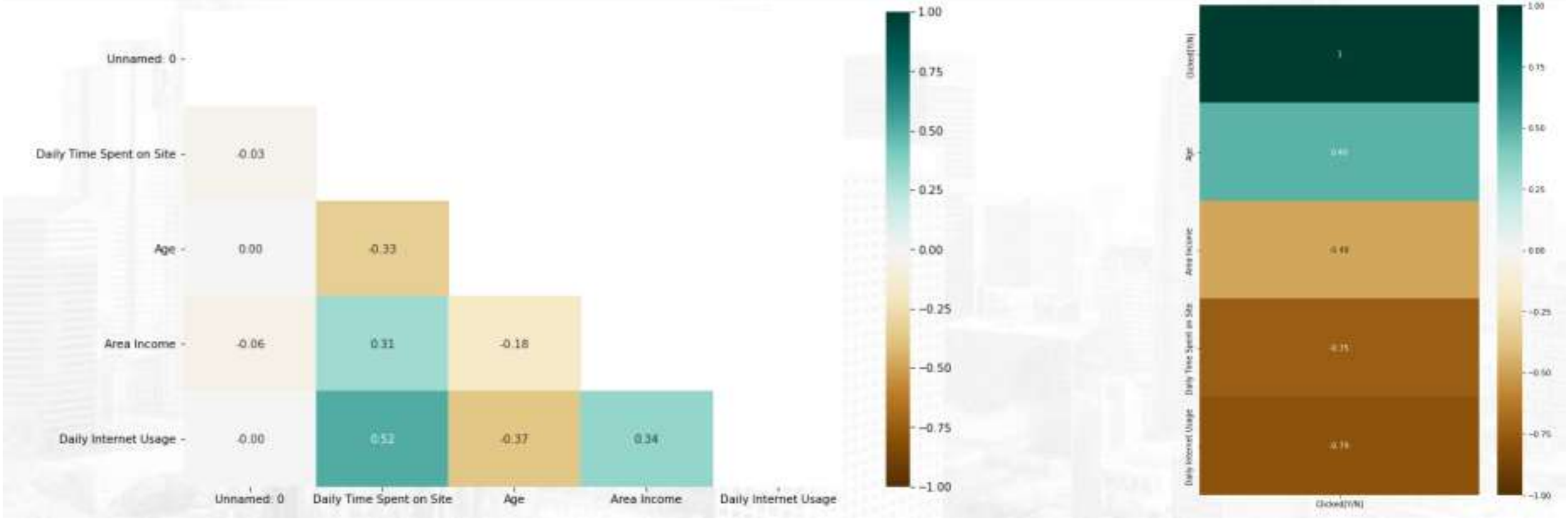


The distribution of the target market's age group in general leads to the age of 20 – 60 years with us age internet and time is short, it is likely the intended target is productive workers who only have take a minute to access the internet (open a website) because they work.

Ad clickers are likely to be people with low incomes aged 20 – 40 years40 years and over have the potential to click on ads at the age of 40 and over.

For details, see jupyter

Heatmap



The correlation between age and daily spend time shows a moderate correlation, namely 0.33, so the age is higher the higher the income

Daily internet usage and time used show a fairly high relationship, namely 0.52, these 2 features considered highly related

On the Heatmap (right side), shows the correlation between features against the target (Clicked [Y/N]) where Time and Internet usage has a strong correlation

For details, see jupyter

Data Cleaning & Preprocessing

Missing Value and Duplicate Data

Missing value. there is a missing value in the Daily time spent on site column of 13, Area income of 13, Daily internet usage of 11. Gender of 3 Duplicate data.

There are no duplicate data.

To handle missing values for the daily time spent in site, income area, and daily internet usage columns, use the median. Meanwhile, gender uses fashion.

For details, see jupyter

Feature Encoding and Engineering

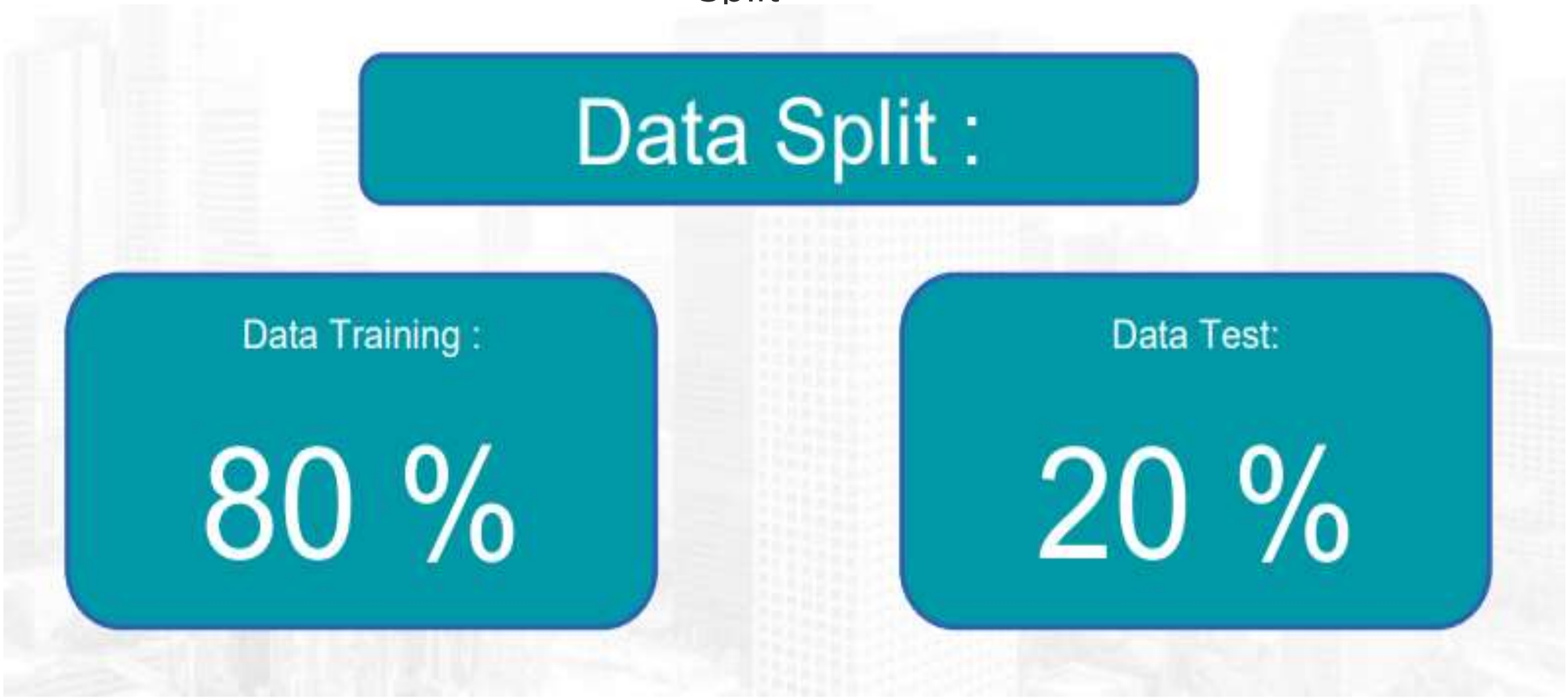
One hot encoding. Not doing one hot encoding

Encoding label. Change male gender to 1 and female to 0. And province column

Extract timestamps. From timestamp to year, month, day and hour.

For details, see jupyter

Train Test Split



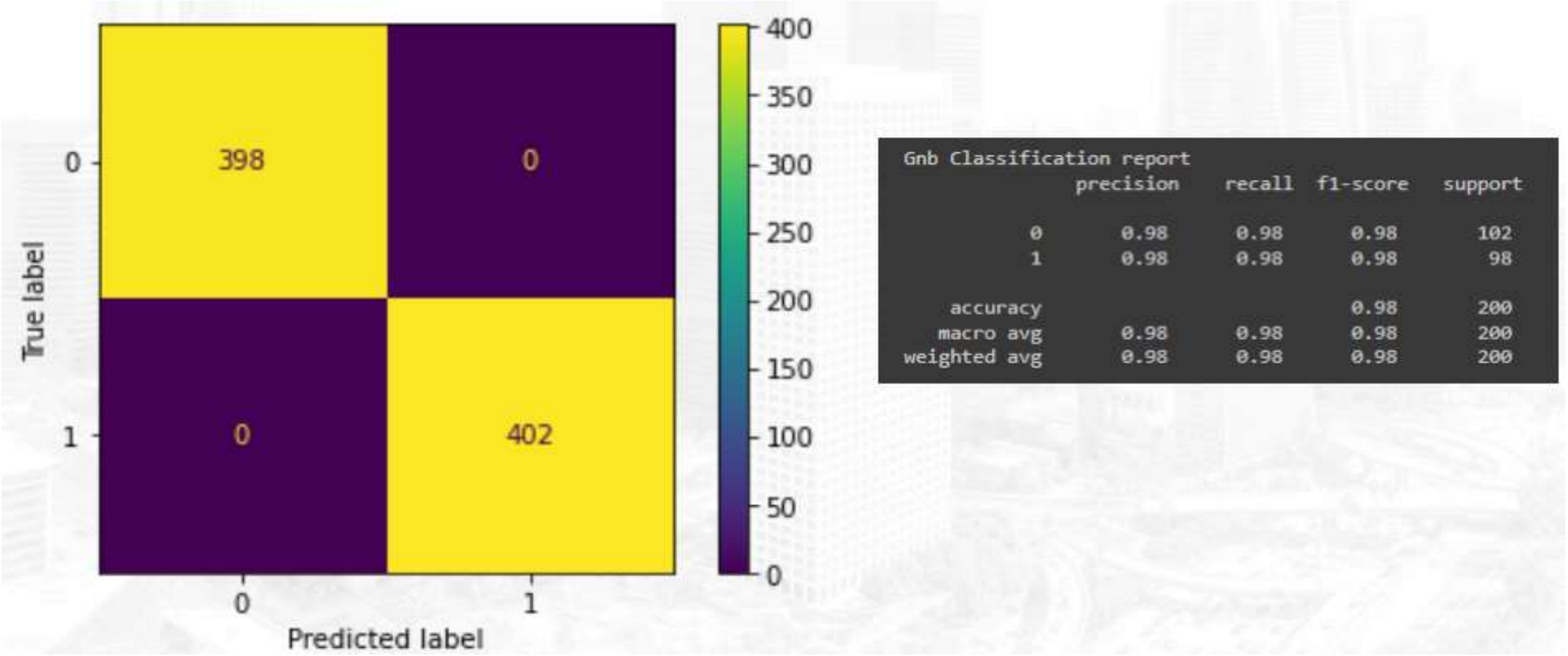
For details, see jupyter

Data Modeling

Model	Accuracy	
	Non Normalisasi	Normalisasi
Logistic Regression	49%	96%
Random Forest	94%	95%
Naïve Baiyes	71%	98%

For details, see jupyter

Confusion Matrix

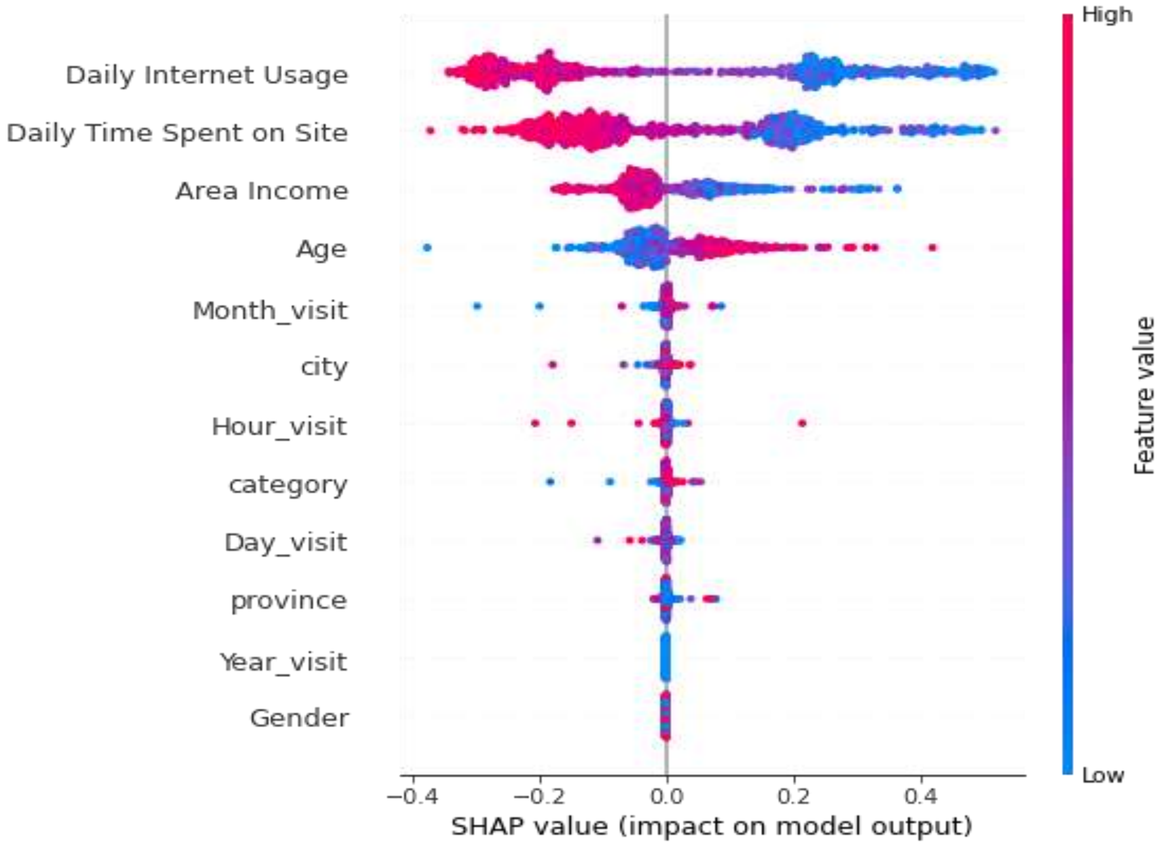


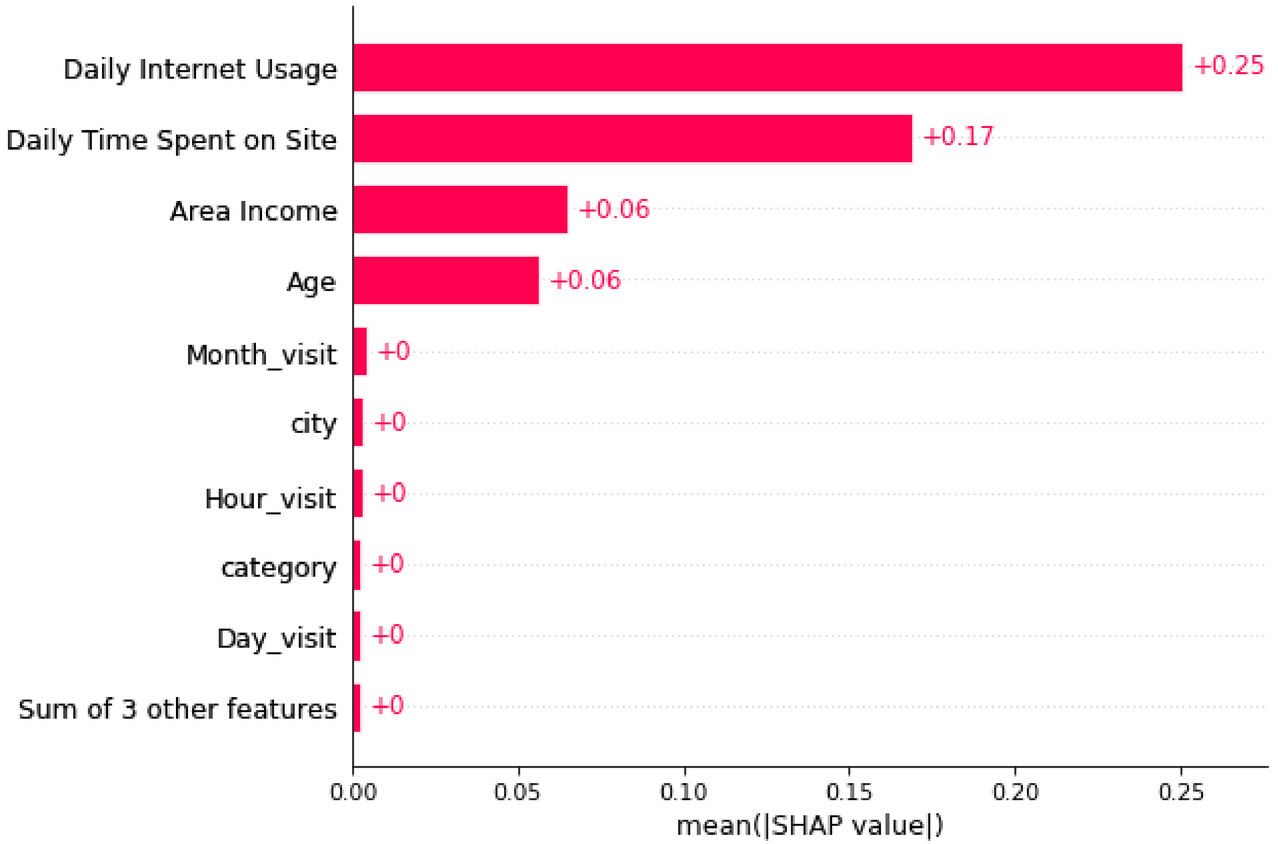
Gnb Classification report

	precision	recall	f1-score	support
0	0.98	0.98	0.98	102
1	0.98	0.98	0.98	98
accuracy			0.98	200
macro avg	0.98	0.98	0.98	200
weighted avg	0.98	0.98	0.98	200

For details, see jupyter

Feature Importance Shap Value





Interpretation :

There are people who access the internet for not too long but have impact on the target and actually click on the ads

Older people tend to click (targeted market)

People with high incomes, old age, and little internet use tend to click on ads (targeted market)

For details, see jupyter

Business Insight

The age that has the potential to click on ads is age 40 and over. Matter it is possible that the products/services suit this age group.(Recommendation: Ads are set up for ages 40 and over to be effective budget and more targeted advertising)

Target market are those who only spend less than70 minutes per day is likely the only productive worker had time to access less than that duration of time. (Recommendation : Create an ad campaign with keywords that are adjusted to the duration time (certain promos to make transactions with a time limit of 60minutes to do the conversion)

For details, see jupyter

Business Insight

Machine Learning is able to predict accuracy up to 98%, so with the model it can be seen the behavior of people who click , so that it can focus on determining the effectiveness of the appropriate advertising budget with the majority of visitor behavior/character.

Recommendation : Because the model already has high accuracy, people's behavior on True Negative there is no need to be targeted and focus on optimizing advertisements based on the main character of the True Negative results, namely: 'People with high income, old age, and less internet use click on ads (targeted market)'

For details, see jupyter

Simulation

Before Using Machine Learning

- Ad Budget 3,000,000
- CPM (Cost per million) : 20,000 (Target Impression 150,000)
- Cost CPC (Cost per click): 3,000
- Before using ML from 1000 people there are 500 Yes and 500 No (50% Yes)
- 1000 visitors with 500 clicks .

After using machine learning

- CPC is lowered and the target market is specified in the age group
- After using ML, ads are optimized and focused based on people's character on True Positive(Moving amount on True Negative to True Positive) up to 80%
- 1000 visitors with 800 clicks.
- If calculated with the same budget, namely 3,000,000, the number of clicks that can be obtained:○ Without ML : 10,000 impressions (maximum) + 933 clicks○ Using ML : 10,000 impressions (maximum) + 1166 clicks

For details, see jupyter

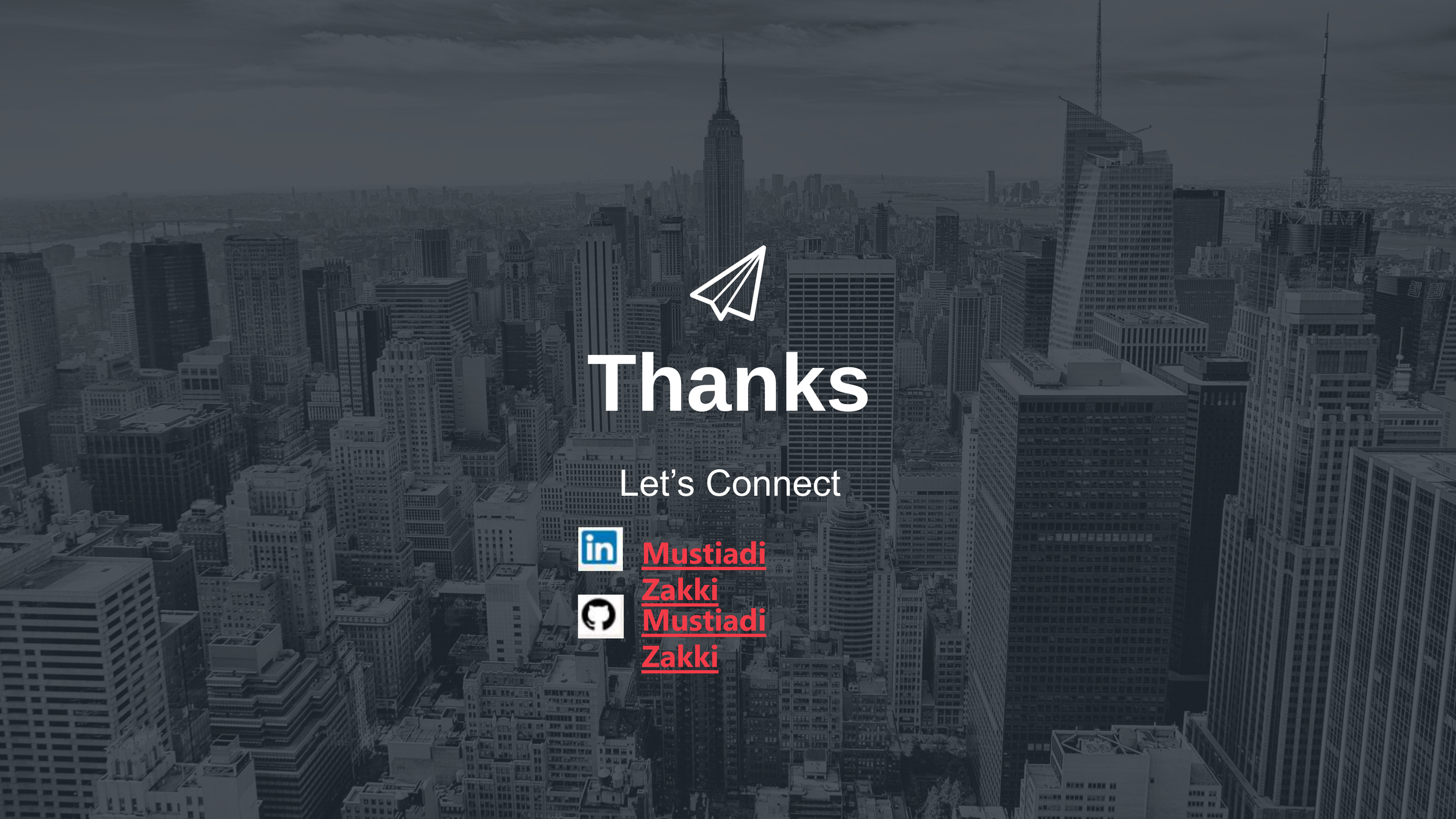
Simulation

If it is assumed per performance ads there is a 30% conversion from the comparison of the two schemes , conversion gets Gross Revenue worth 300 thousand and profit per product of 50,000No ml

- GROSS = 0.3 * 933 * 300,000 = 83,970,000
- PROFIT = 0.3 * 933 * 50,000 = 13,995,000Using ML
- GROSS = 0.3 *1166* 300,000 = 104,940,000
- PROFIT = 0.3 *1166* 50,000 = 17,490,000

Profit Difference Profit = 3,495,000(If you want to increase profit, you need to increase your budgeting)

For details, see jupyter



Thanks

Let’s Connect

Mustiadi Zakki

Mustiadi Zakki