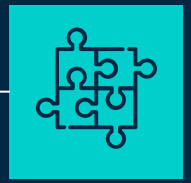
EDS Capstone 2022

[GROUP 3]
Suwaibatul Aslamiah binti Mohd Afandi
Sofiyah Binti Rahmat
Syafiqah Shahrul Aman
Nur Alyaa Nadhirah Mohd Roselee
Zarith Sofea Norezalee

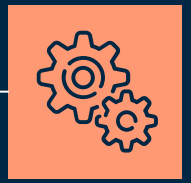
TABLE OF CONTENTS



01

INTRODUCTION

We will briefly go through the background of our analysis.



02

FINDINGS

We will present visuals that will aid in explaining our data and analysis process.

03

CONCLUSION

Based on the above findings, we will present a solution and suggest recommendations.

INTRODUCTION

01

Problem Statement
Objectives
Data Overview

UNDERSTANDING THE PROBLEM

PROBLEM

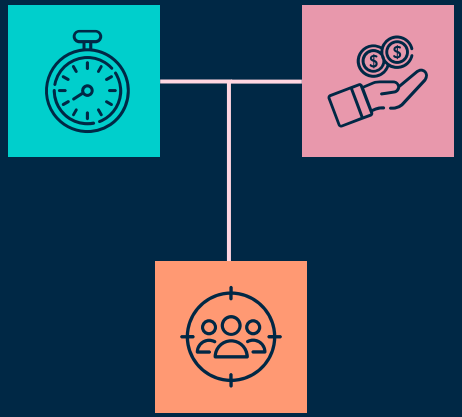
We're required to look into different factors that determine whether a stranger can be a potential customer based on their shopping behaviour.

SOLUTION

Through rigorous analysis and using a predictive model to investigate factors that influence customers' behaviour and make recommendations based on our findings.

OUR PROJECT OBJECTIVES

Understanding the problem and overall data.



Visualizing and explaining our findings from analysis.

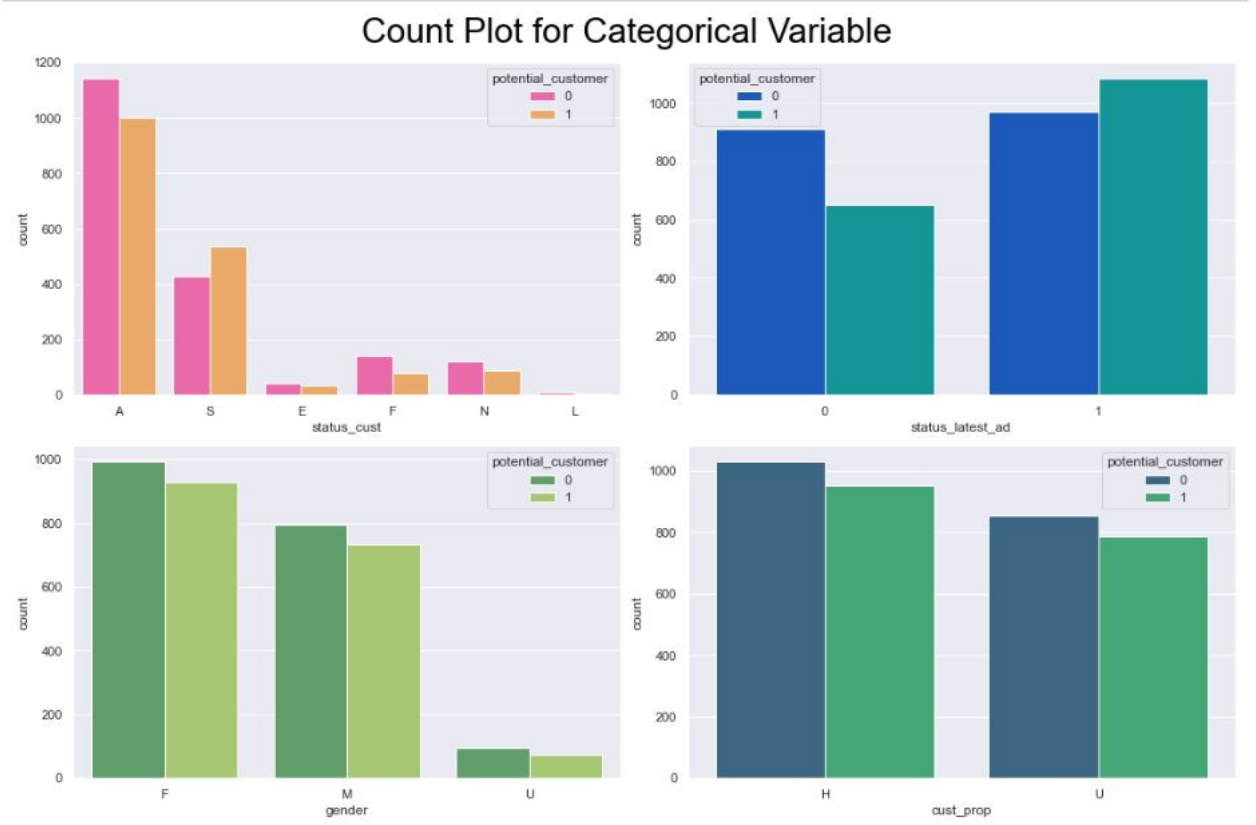
Extracting insights from our analysis and visualisations.

OUR FINDINGS

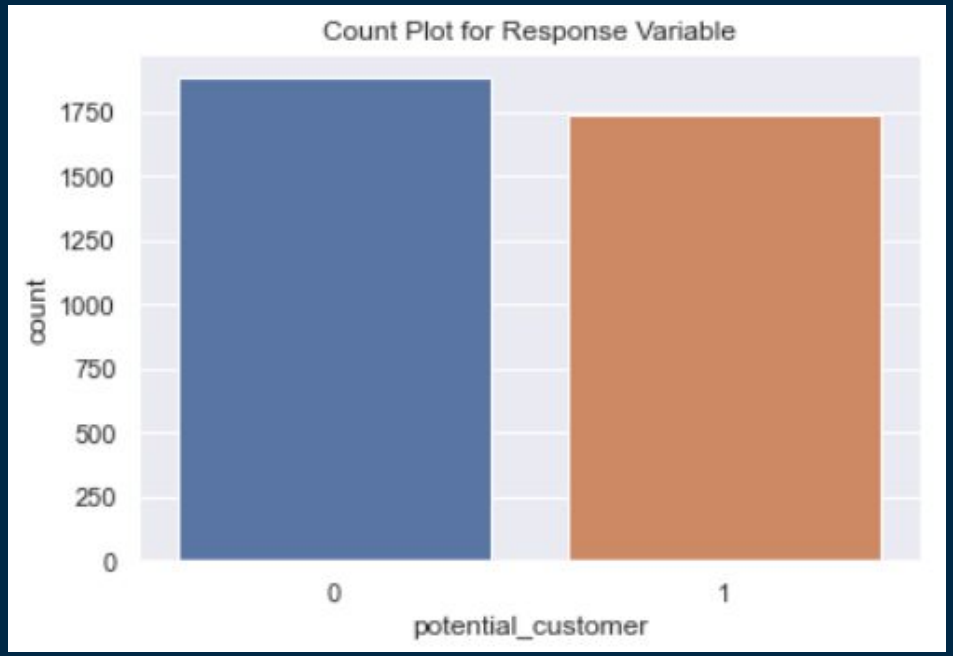
02

Visualisations of analysis
Explaining the model

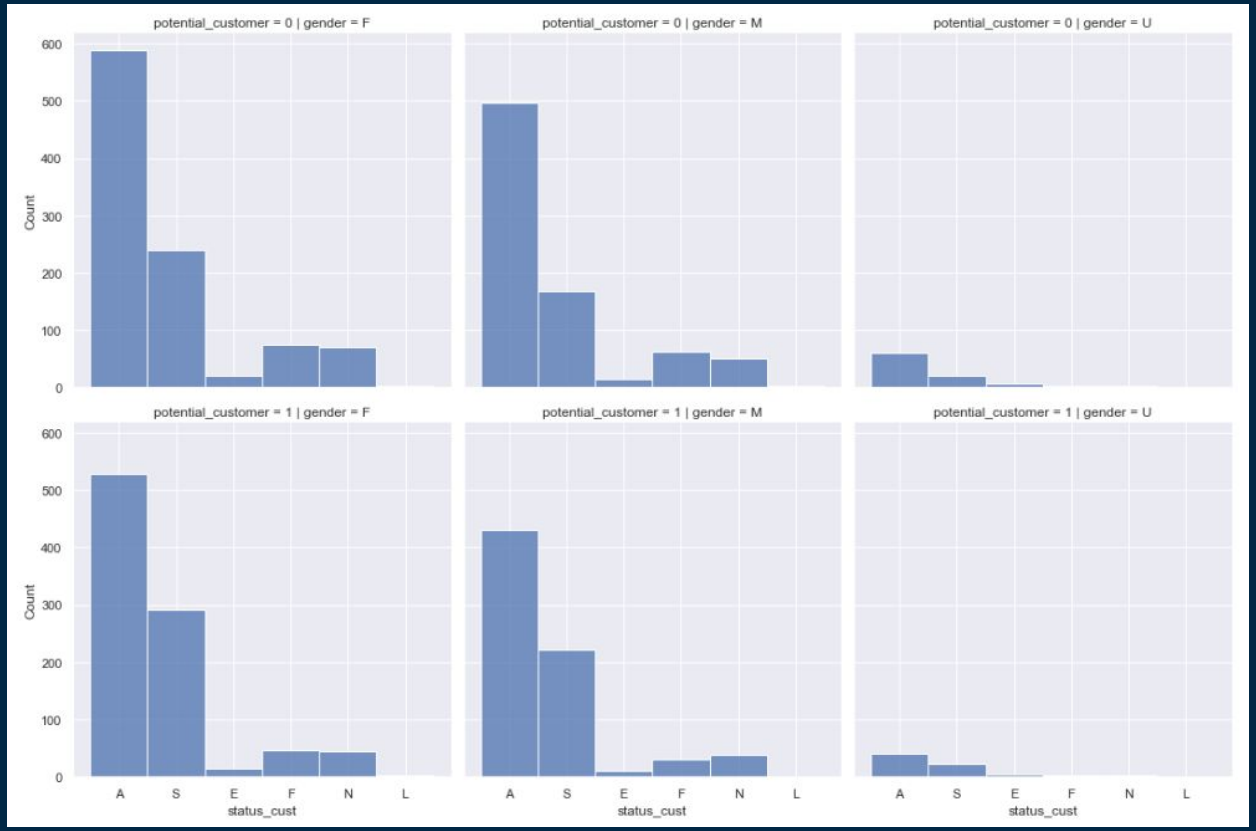
1. UNDERSTANDING THE CUSTOMERS



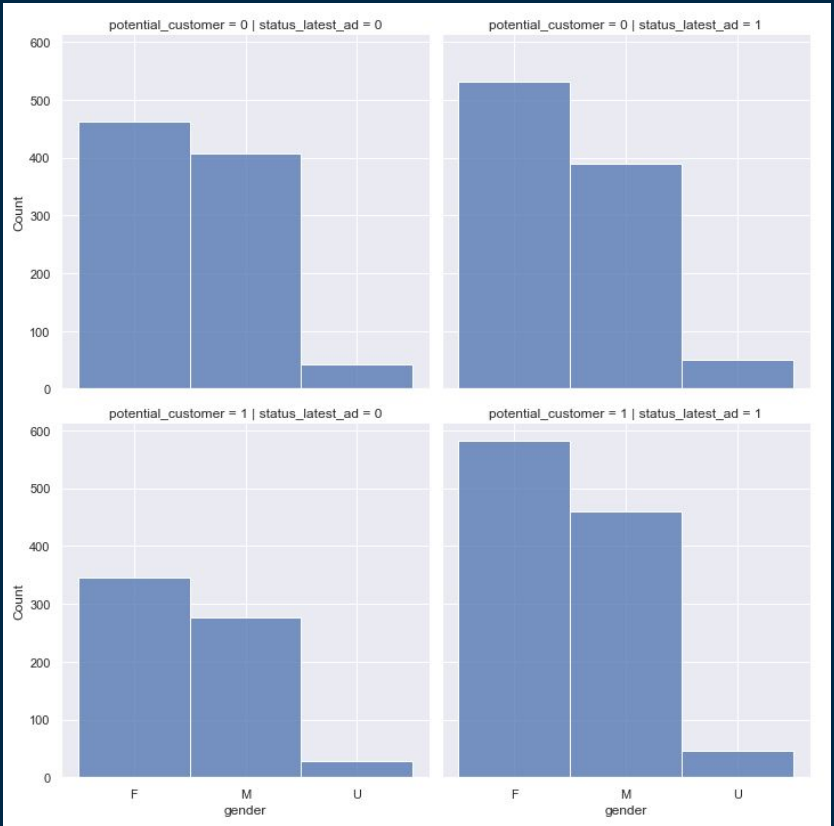
2. PROPORTION OF POTENTIAL CUSTOMERS



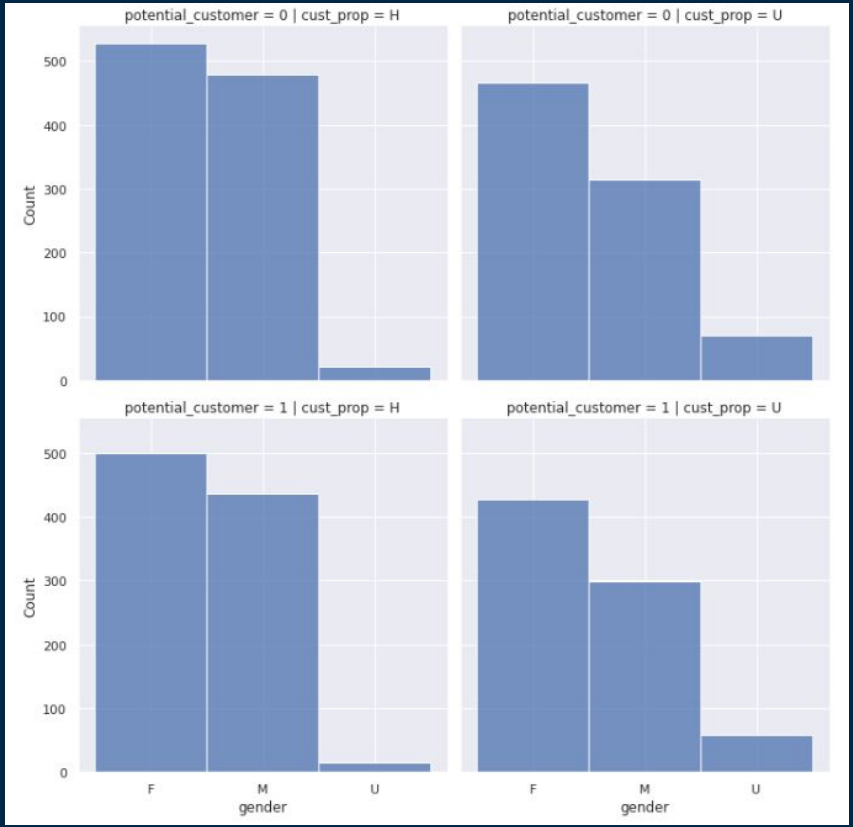
3. TYPE OF CUSTOMERS BASED ON PURCHASE



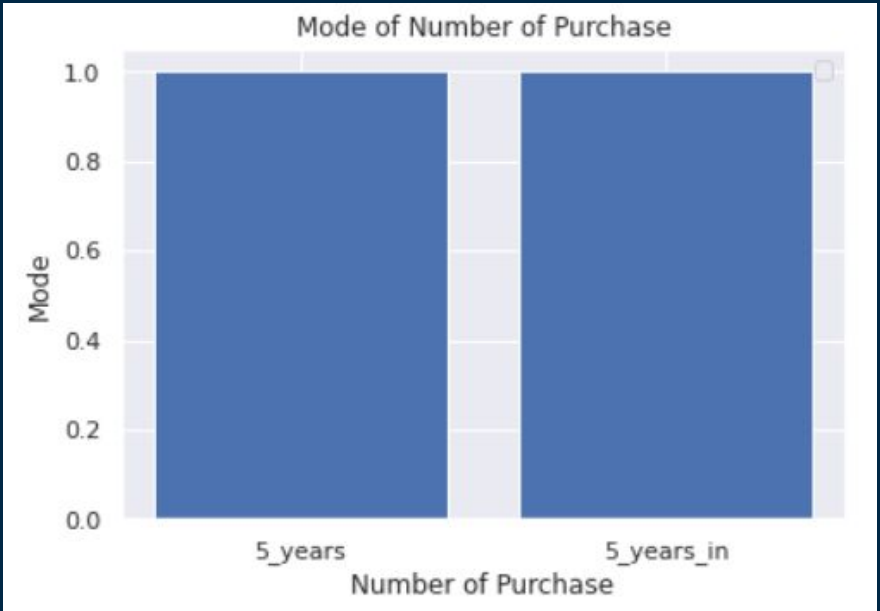
4. TYPE OF CUSTOMERS BASED ON PURCHASE



5. TYPE OF CUSTOMERS BASED ON PURCHASE



6. NUMBER OF PURCHASE (DIRECT AND INDIRECT)




```
1 data.pur_5_years.max()
```

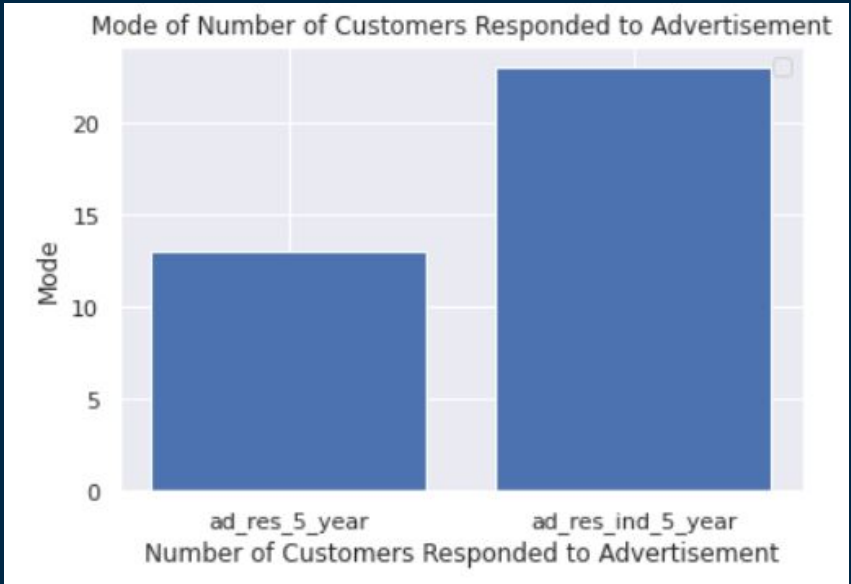
```
91
```

```
1 data.pur_5_years_indirect.max()
```

```
41
```

7. CUSTOMERS THAT RESPOND TO ADS

0.076 %



0.39 %

```
1 data.ad_res_5_year.max()
```

```
157
```

```
1 data.ad_res_ind_5_year.max()
```

```
56
```

THE MODEL

What model(s) did we create?

KNN Classifier, Decision Tree, Logistic Regression, Polynomial Logistic Regression, Naive Bayes

Which is the best model?

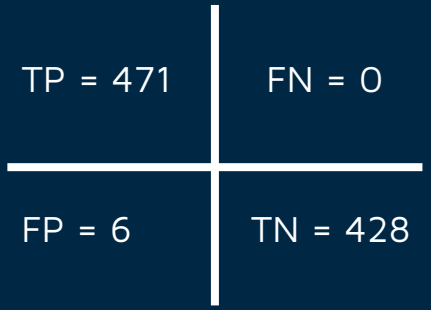
Decision Tree

Why this model is the best?

Decision Tree Best Train Scores: 0.9931079487636684

DT Best model Precision score on test data = 1.00
DT Best model Recall score on test data = 0.99
DT Best model F1 score on test data = 0.99
DT Best model Accuracy score on test data = 0.99

Confusion Matrix



CONCLUSION

Summary of Inferences
Actionables

03

INSIGHTS FROM OUR ANALYSIS

- Huge number of active female and male customers have left which we believe can contribute to great profit loss.
- “how reliance the model is when it says this data point is a potential customer”

  For decision tree model we built, the precision score turned to be 1.00, this would enable the business to predict future buyers confidently.

THANK YOU