Data-Driven Insights for Customer Retention in Pizza Restaurants

Project Report

A guid by Husanbano Shamlik

INTRODUCTION

- Objectives:Improve customer retention and drive business growth through data-driven insights
- Agenda:

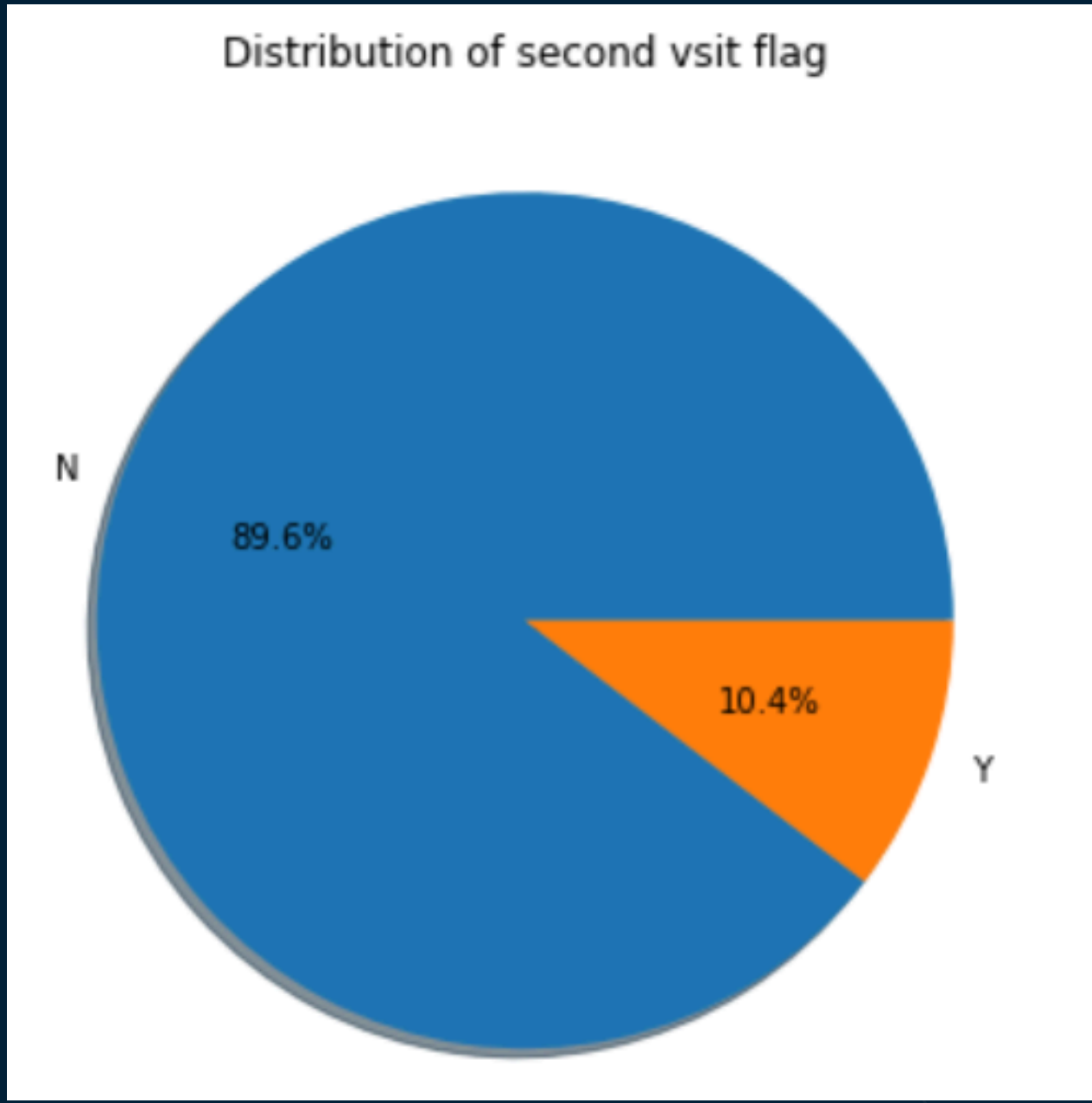
1. Enhance Repeat Visit Identification
2. Optimize Marketing and Promotions
3. Improve Customer Experience
4. Leverage Customer Feedback
5. Enhance Product and Sales Strategies
6. Address Class Imbalance
7. Implementation Plan

DATA DISCRIPTION

- Customer_ID: Unique identifier for each customer.
- TOTALQUANTITY: Total quantity of items purchased.
- Main_Plates_Sale: Revenue from sales of main plates.
- Salads_&_Powerbowls_Sale: Revenue from sales of salads and power bowls.
- Salads_Sale: Revenue from sales of salads.
- Beverages_Sale: Revenue from sales of beverages.
- Mixed_Drinks_Sale: Revenue from sales of mixed drinks.
- Pastas_Sale: Revenue from sales of pasta dishes.
- Soups_Sale: Revenue from sales of soups.
- Wine_Sale: Revenue from sales of wine.
- Mains_Sale: Revenue from sales of main dishes.
- Pizza_Sale: Revenue from sales of pizza.
- NABs_Sale: Revenue from sales of non-alcoholic beverages.
- Specialty_Sale: Revenue from sales of specialty items.
- Small_Plates_Sale: Revenue from sales of small plates.
- Lunch_Duos_Sale: Revenue from sales of lunch duo combinations.
- Combo_Boxes_Sale: Revenue from sales of combo boxes.
- Desserts_Sale: Revenue from sales of desserts.
- Kids_Sale: Revenue from sales of kids' meals.
- Sandwiches_Sale: Revenue from sales of sandwiches.
- TOTAL_DISCOUNT: Total discount applied to the order.
- %_Alcohol_of_Bill: Percentage of the total bill attributed to alcoholic beverages.
- GUEST_COUNT: Number of guests at the table.
- DISTANCE_TO_CLOSESTSTORE: Distance from the customer to the closest store.
- SECONDVISITFLAG: Flag indicating if this is a second visit by the customer.

ENHANCE REPEAT VISIT IDENTIFICATION

From the above pie chart, we conclude that the percentage of transactions during the second visit is less than 10.4%, compared to those who only visit for the first time, which is 89.6%



XGBoost : It is the best model for predicting the probability of a customer making a repeat visit, as it has the highest ROC-AUC score (0.91) and accuracy (0.90), making it the most reliable in distinguishing between customers who will and will not make repeat visits.

```
logistic Regression ROC-AUC Score: 0.69
confusion_matrix : [[42700     0]
 [ 5013     0]]

Decisoin Tree  Accuracy : 0.87
Decisoin Tree  ROC-AUC Score: 0.66
confusion_matrix : [[39491  3209]
 [ 3081  1932]]

Random Forest Accuracy : 0.90
Random Forest ROC-AUC Score: 0.89
confusion_matrix : [[41802   898]
 [ 4097   916]]

XGBoost Accuracy : 0.90
XGBoost ROC-AUC Score: 0.91
confusion_matrix : [[41713   987]
 [ 3820  1193]]

BernoulliNB Accuracy : 0.89
BernoulliNB ROC-AUC Score: 0.57
confusion_matrix : [[42700     0]
 [ 5013     0]]
```

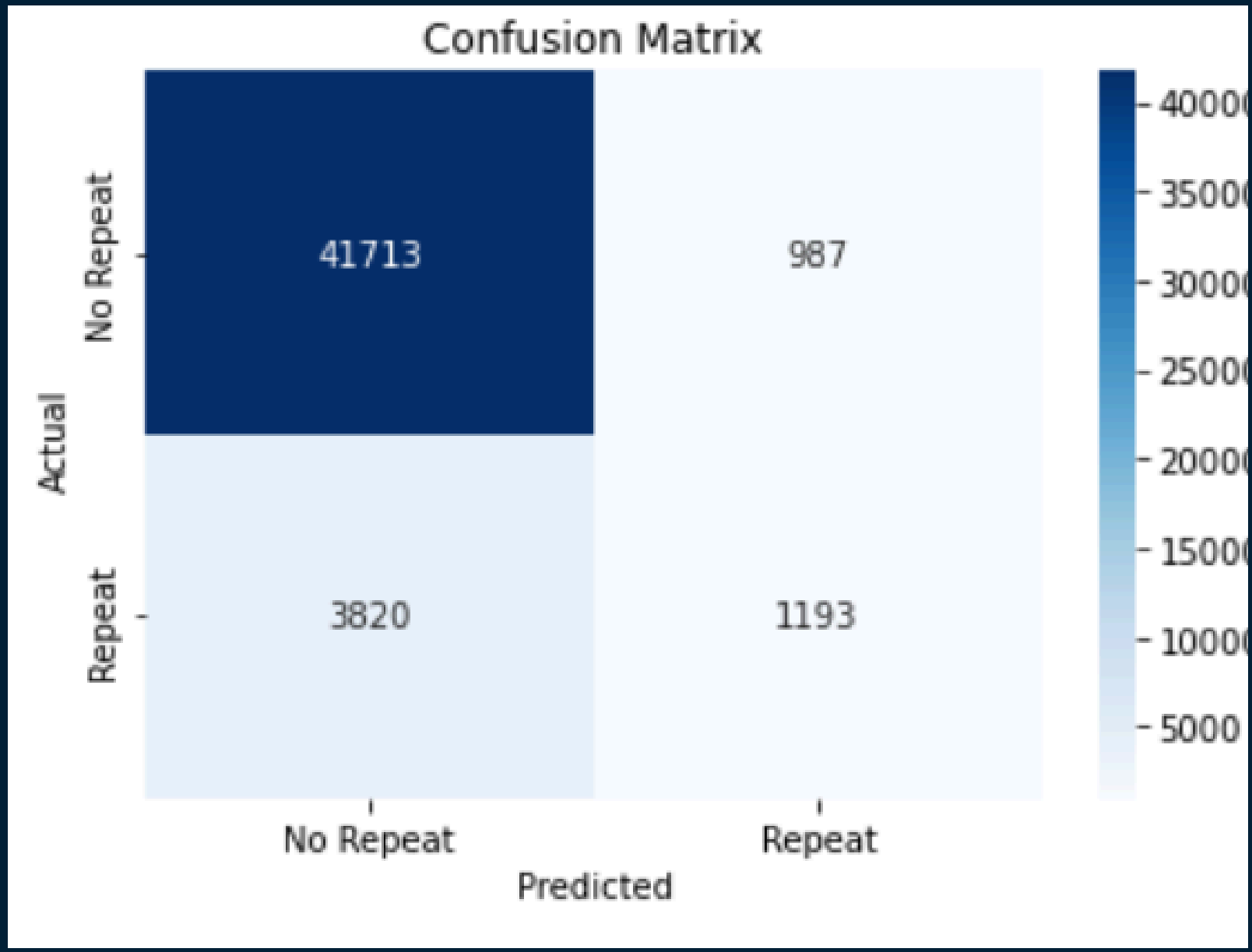
The most significant factors influencing repeat visits are Customer_ID and TOTAL_DISCOUNT. Other important factors include Pizza_Sale, GUEST_COUNT, and DISTANCE_TO_CLOSESTSTORE. Features related to specific sales categories and quantities contribute less significantly but can still provide valuable insights for tailoring marketing strategies and improving customer retention.

```
d features: Index(['Customer_ID', 'Salads_Sale', 'Pizza_Sale', 'NABs_
'Appetizers_Sale', 'Lunch_Duos_Sale', 'TOTAL_DISCOUNT', 'GUEST_COUNT'
'DISTANCE_TO_CLOSESTSTORE', 'Total_Sales'],
type='object')
```

Accuracy: 0.8992517762454677
Classification Report:

	precision	recall	f1-score	support
0	0.92	0.98	0.95	42700
1	0.55	0.24	0.33	5013
accuracy			0.90	47713
macro avg	0.73	0.61	0.64	47713
weighted avg	0.88	0.90	0.88	47713

ROC-AUC Score: 0.9075619618500097


Confusion Matrix
41713
987
3820
1193
No Repeat
Repeat
Actual
No Repeat
Repeat
Predicted
40000
35000
30000
25000
20000
15000
10000
5000

Thanks for

PLAYING!