BANK MARKETING

PREDICTION

GHAZY RAFIF AGUSTIAN

Who Am I?

Ghazy Rafif Agustian

I am a Data Science Enthusiast who is deepening Data Science in a Bootcamp. My goal is to become a reliable Data Science.

Education

- Binus University
- Dibimbing.id

Working Experience

Desa Jongkon Cafe & Resto

Digital Marketing & Business Development

Aug 2020 - Mar 2023

PT. BUENA PERSADA MINING SERVICE

Strategic Management Intern

Feb 2020 - Aug 2020

Sartre Coffee

Financial Officer

Jan 2018 - Dec 2018

Table of Content

Notebook

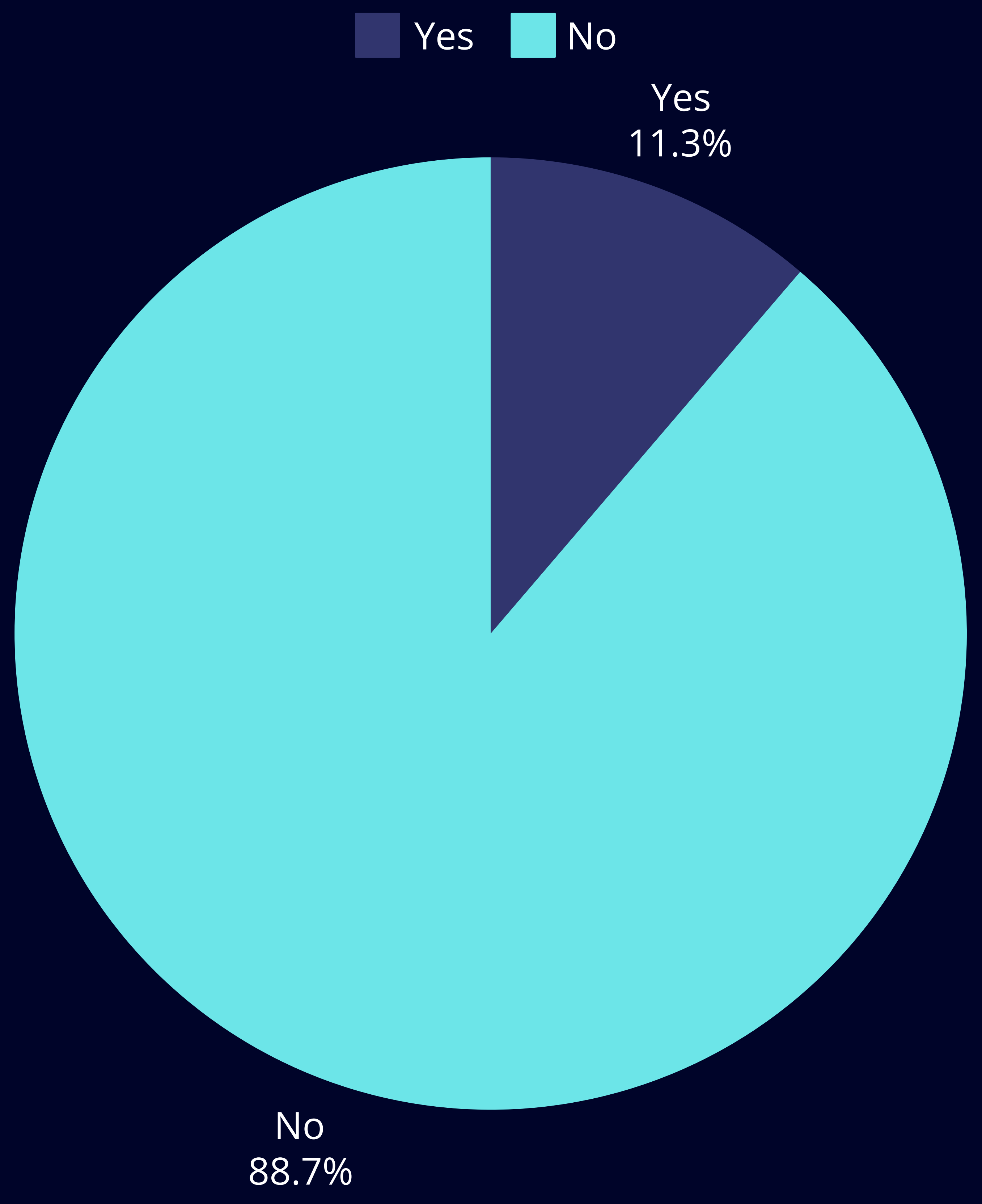
DATA BACKGROUND

Background

The case study used data from the bank's marketing campaign aimed at promoting its time deposit products. A prediction model was developed based on this data to improve customer conversion rates.

Case Study

The bank wanted to improve the effectiveness of its time deposit product marketing campaign by identifying customers who had a higher potential to subscribe. Previous campaigns had high costs but low conversion rates, thus requiring a data-driven solution.



Problems

Low Conversion Rate:
Many customers are contacted but do not subscribe.

High Campaign Cost:
Campaigns require large resource allocation with sub-optimal results.

Lack of Customer Segmentation:
No clear strategy to prioritize high-potential customers.

Goals

- Increase the success rate of marketing campaigns by using data-driven predictions.
- Reduce campaign costs by focusing efforts on customers with the most potential to subscribe.

Objectives

- Create predictive models to identify customers with a high likelihood to subscribe.
- Analyze critical features that influence customer decisions.
- Provide strategic recommendations to maximize the effectiveness of marketing campaigns.

DATA PREPROCESSING

Datasets Information

Rows

41.118

Features

20

Client
- Age
- Job
- Marital
- Education
- Default
- Housing
- Loan

Related
- Contact
- Month
- Day of Week
- Duration

Other
- Campaign
- Pdays
- Previous
- Poutcome

Social & Economic
- Emp.var.rate
- Cons.price.idx
- Cons.conf.idx
- Euribor3m
- Nr.employed

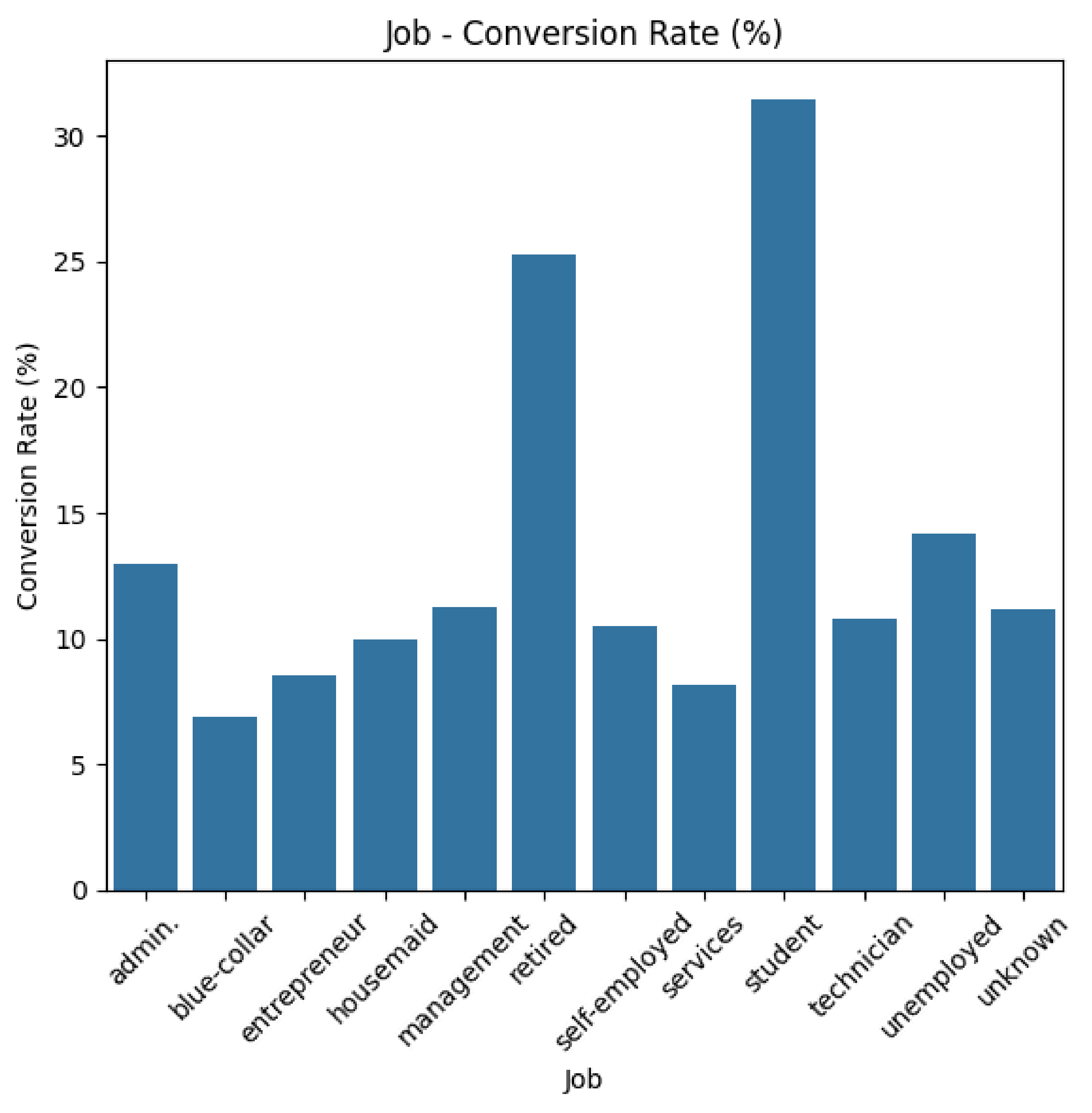
Target

1

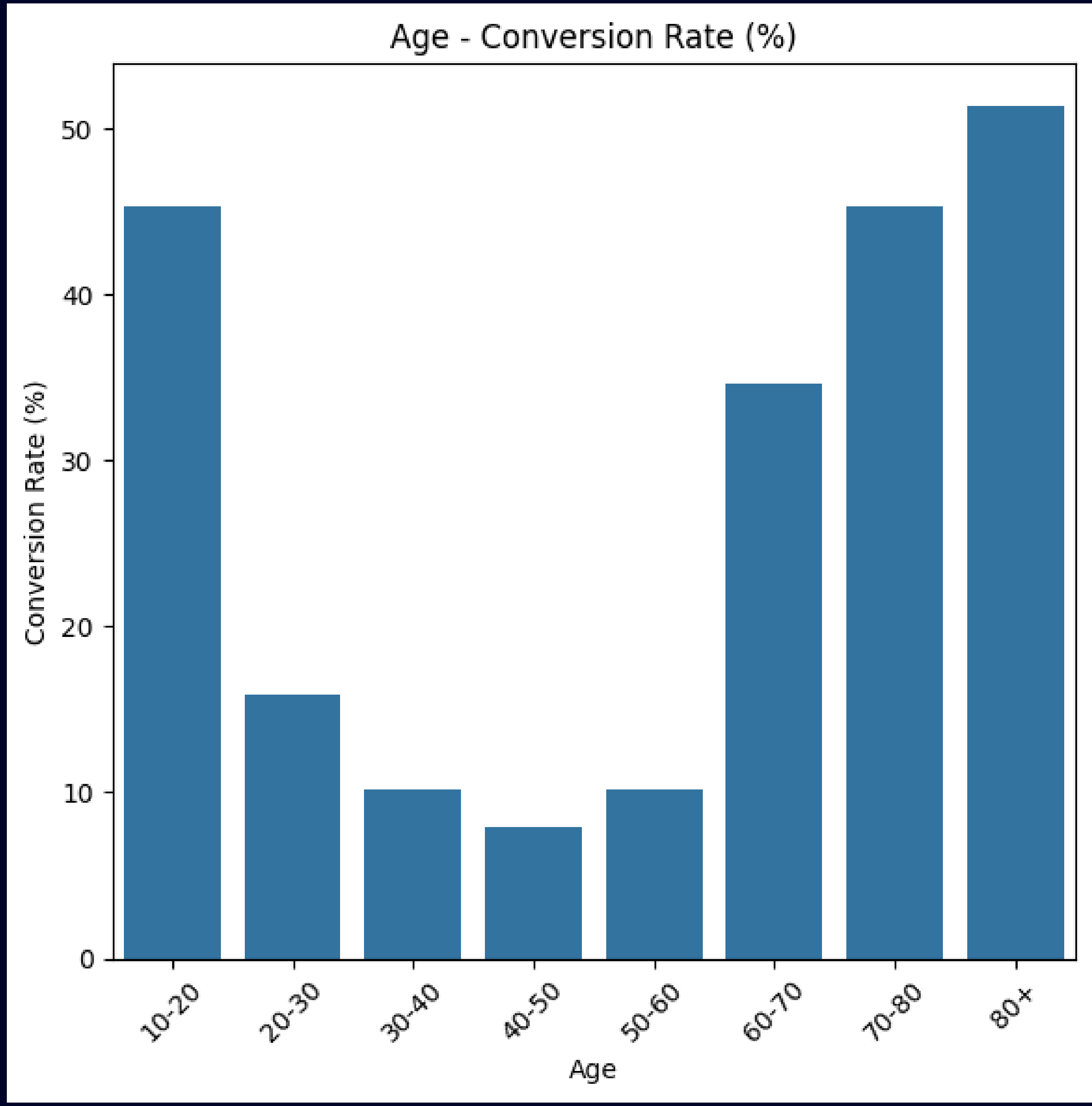
Output
- y

Client Information

The highest conversion rate was found in the "student" category, followed by "retired".



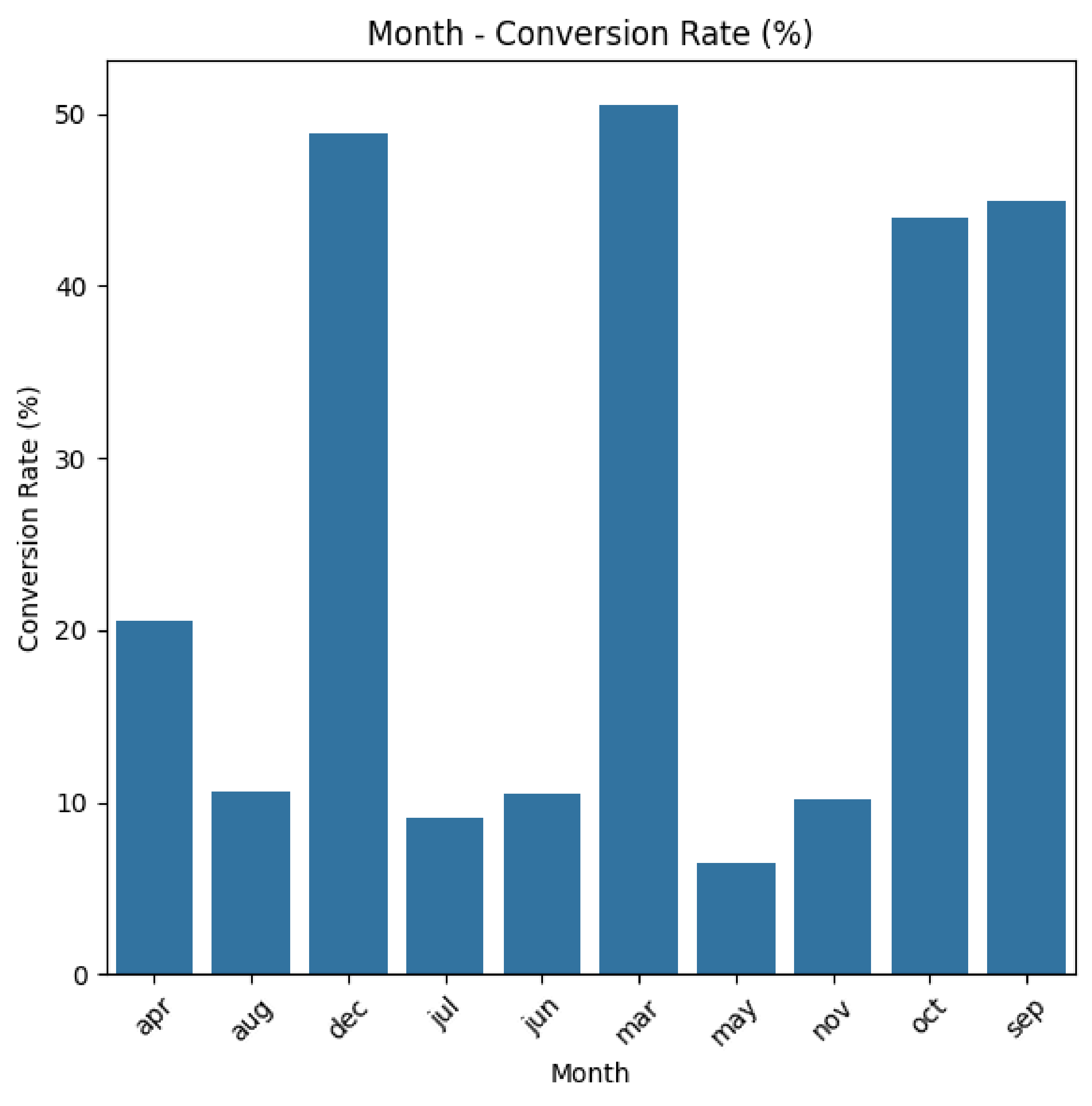
The highest conversion rates are in the 10-20 years and 80+ years age groups.



Unknown in marital is imputed as married
Unknown values for job and education will be kept as it is

Related Information

The highest conversion rates occur in March and December, followed by October and September.

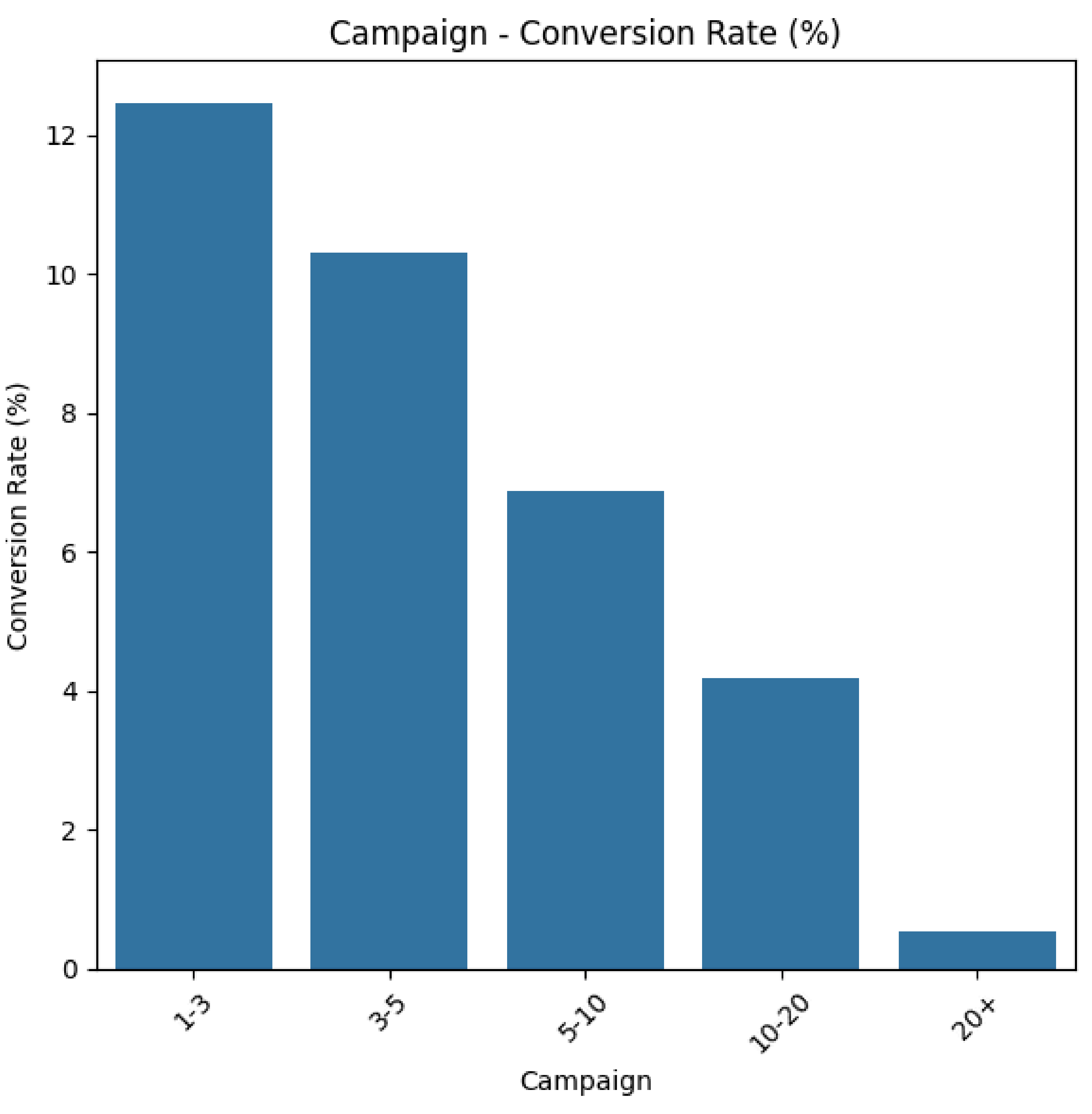


Contact is dropped from the dataset, as well as duration

Other Information

The highest conversion rates were found for customers contacted 1-3 times, with a gradual decline as the number of contacts increased.

Customers contacted more than 10 times had a much lower conversion rate, with a conversion rate of almost zero for the more than 20 times category.

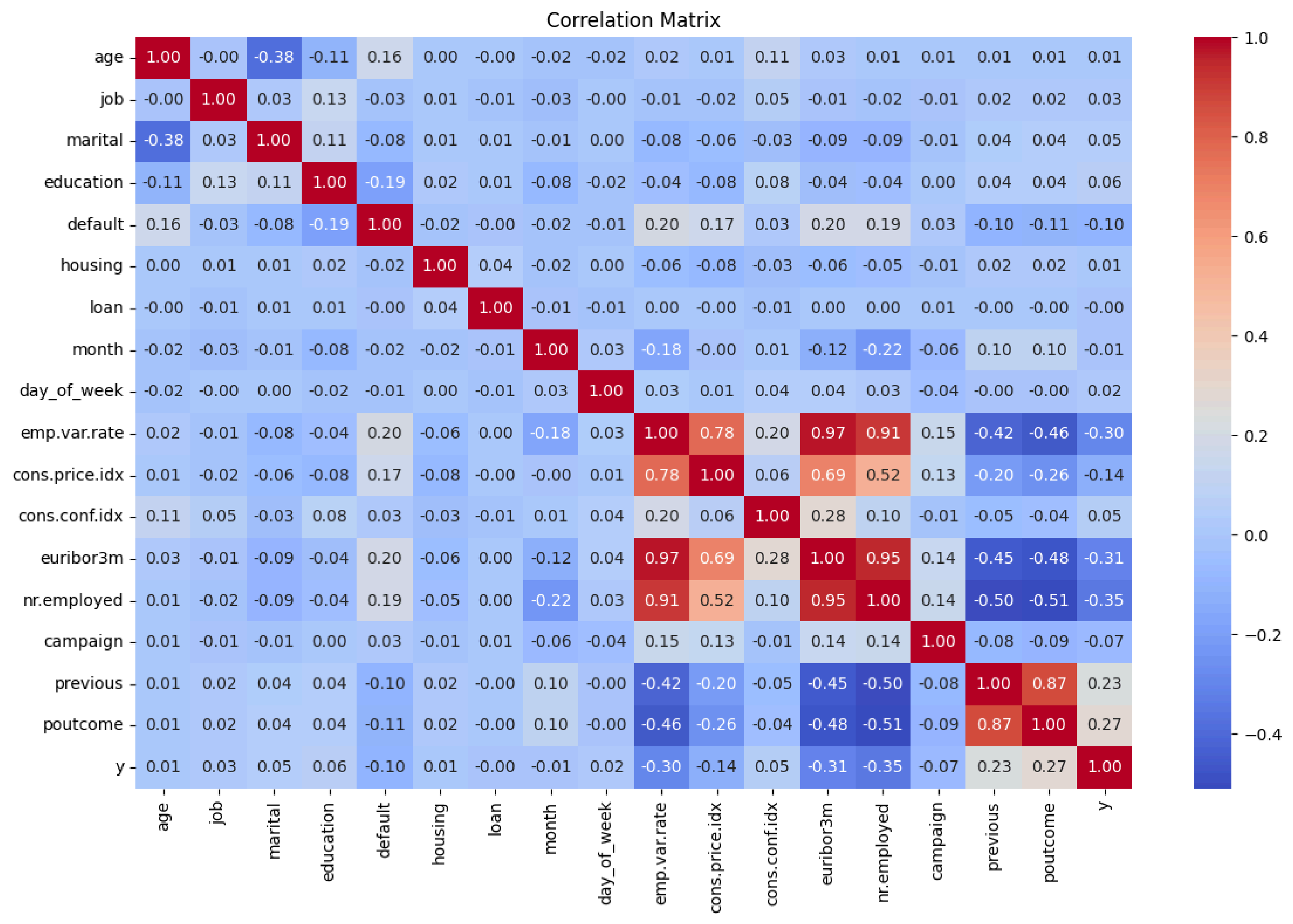


Pdays are dropped from the dataset, since Poutcome, Pday and previous are highly correlated to each other

Correlation Checking

cons.price.idx, nr.employed and pdays are dropped because they have high VIF values

	Feature	VIF
0	age	10.066567
1	job	2.118804
2	marital	5.647149
3	education	4.424382
4	default	1.395648
5	housing	2.201486
6	loan	1.207827
7	month	6.496976
8	day_of_week	3.078460
9	emp.var.rate	39.788742
10	cons.price.idx	29567.630549
11	cons.conf.idx	129.463993
12	euribor3m	313.245995
13	nr.employed	35271.441346
14	campaign	1.916175
15	previous	4.757227
16	poutcome	32.345130



MODELING

Base Model Classifier Comparison

The dataset is splitted into train data (80%) and test data (20%)

Model	Accuracy	ROC-AUC	Recall	Precision	F1-Score
Random Forest	0.8847	0.7569	0.27	0.52	0.35
Gradient Boosting	0.8967	0.7950	0.21	0.71	0.32
LightGBM	0.8963	0.7958	0.21	0.70	0.32
XGBoost	0.8966	0.7952	0.20	0.72	0.32

The Gradient Boosting, LightGBM and XGBoost models will be subjected to Hyperparameter Tuning, as they have high Precision values that are not too different.

Legend :

Accuracy : Percentage of correct predictions to total data.

ROC-AUC : Measures the ability of the model to distinguish between positive and negative classes.

Recall : Percentage of positive class data successfully recognized by the model.

Precision : The percentage of positive predictions that are actually positive.

F1-Score : Harmonious average between Precision and Recall.

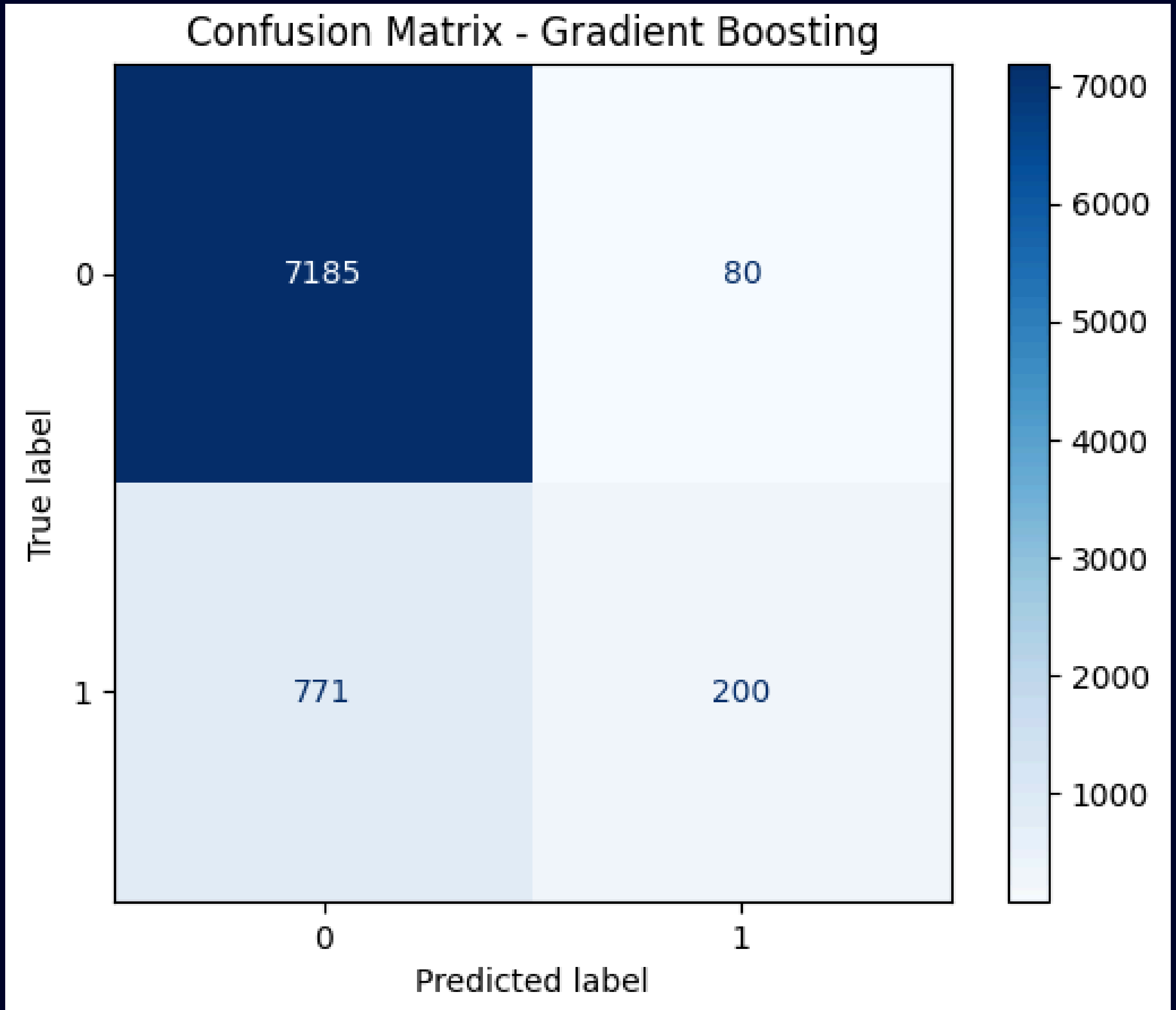
Business Insight by Base Model

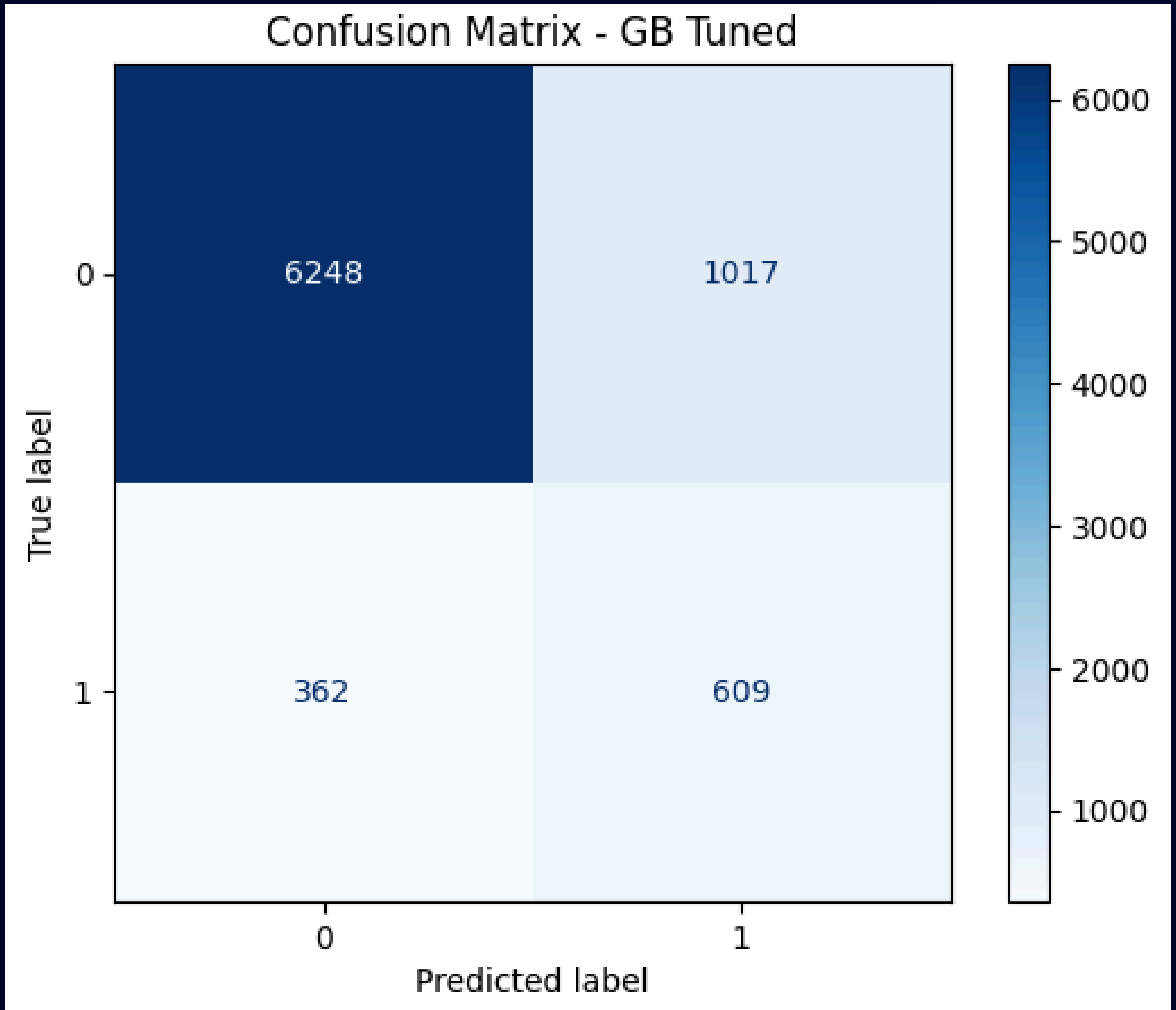
Model	True Positives	Predicted Positive	Profit	Conversion Rate %
Random Forest	261	501	21.090	52.09%
Gradient Boosting	200	280	17.200	71.42%
LightGBM	203	289	17.410	70.24%
XGBoost	196	273	16.870	71.79%

The Gradient Boosting, LightGBM and XGBoost models will be subjected to Hyperparameter Tuning, as they have high Conversion Rate values that are not too different.

Confusion Matrix - Gradient Boosting

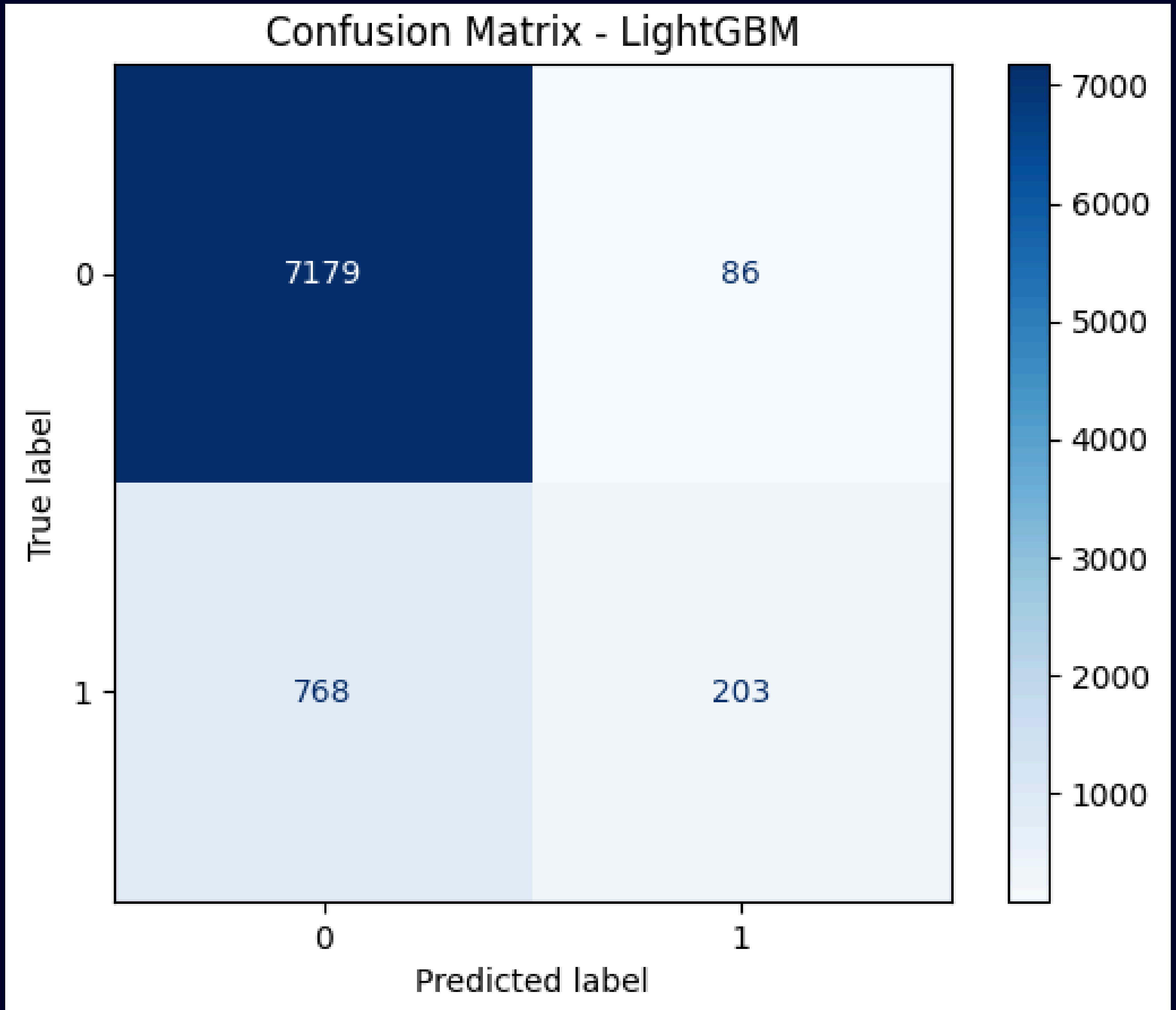
Since one of the campaign objectives is to target as many subscribed customers as possible, Gradient Boosting Tuning is better

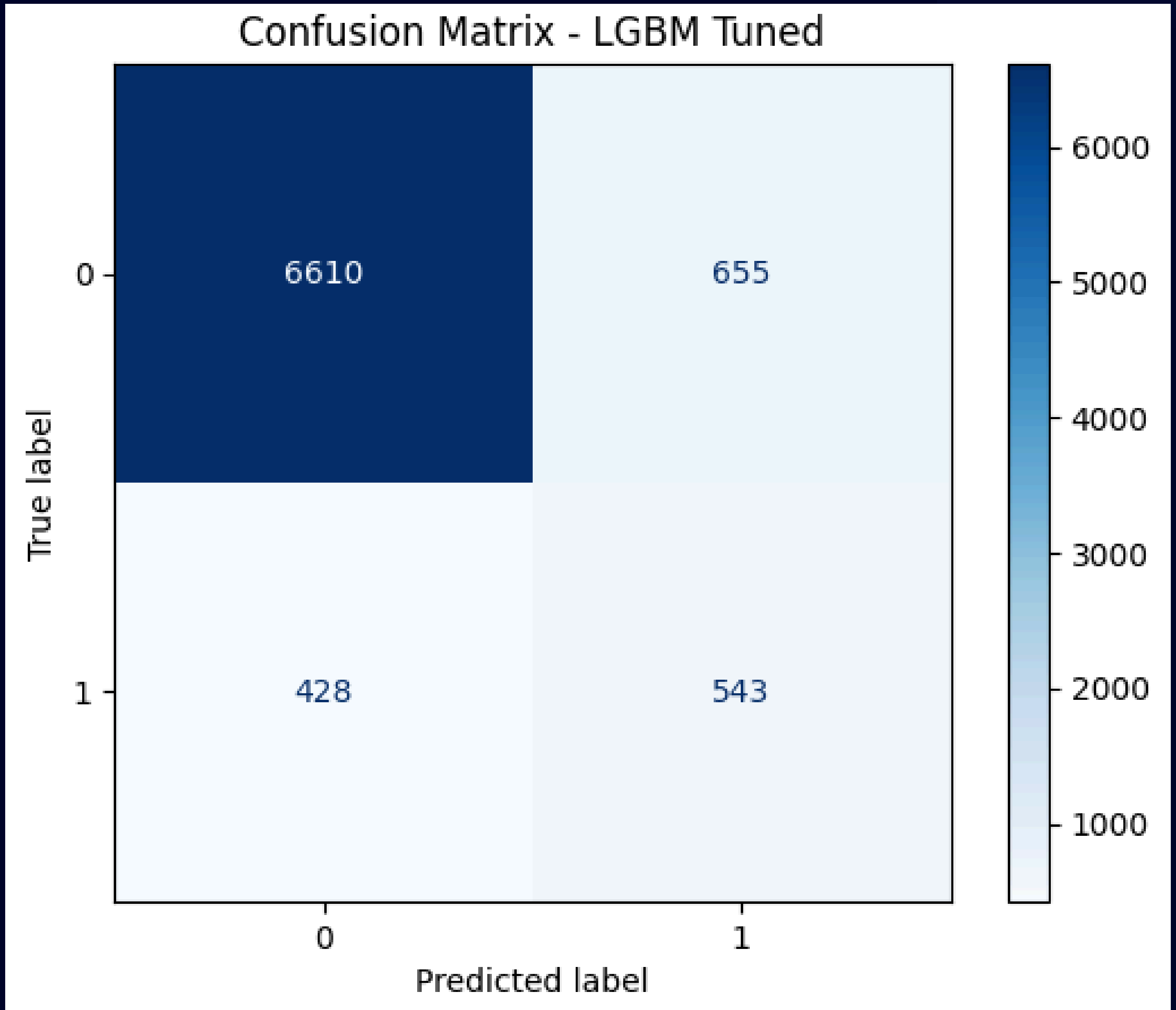




Confusion Matrix - LightGBM

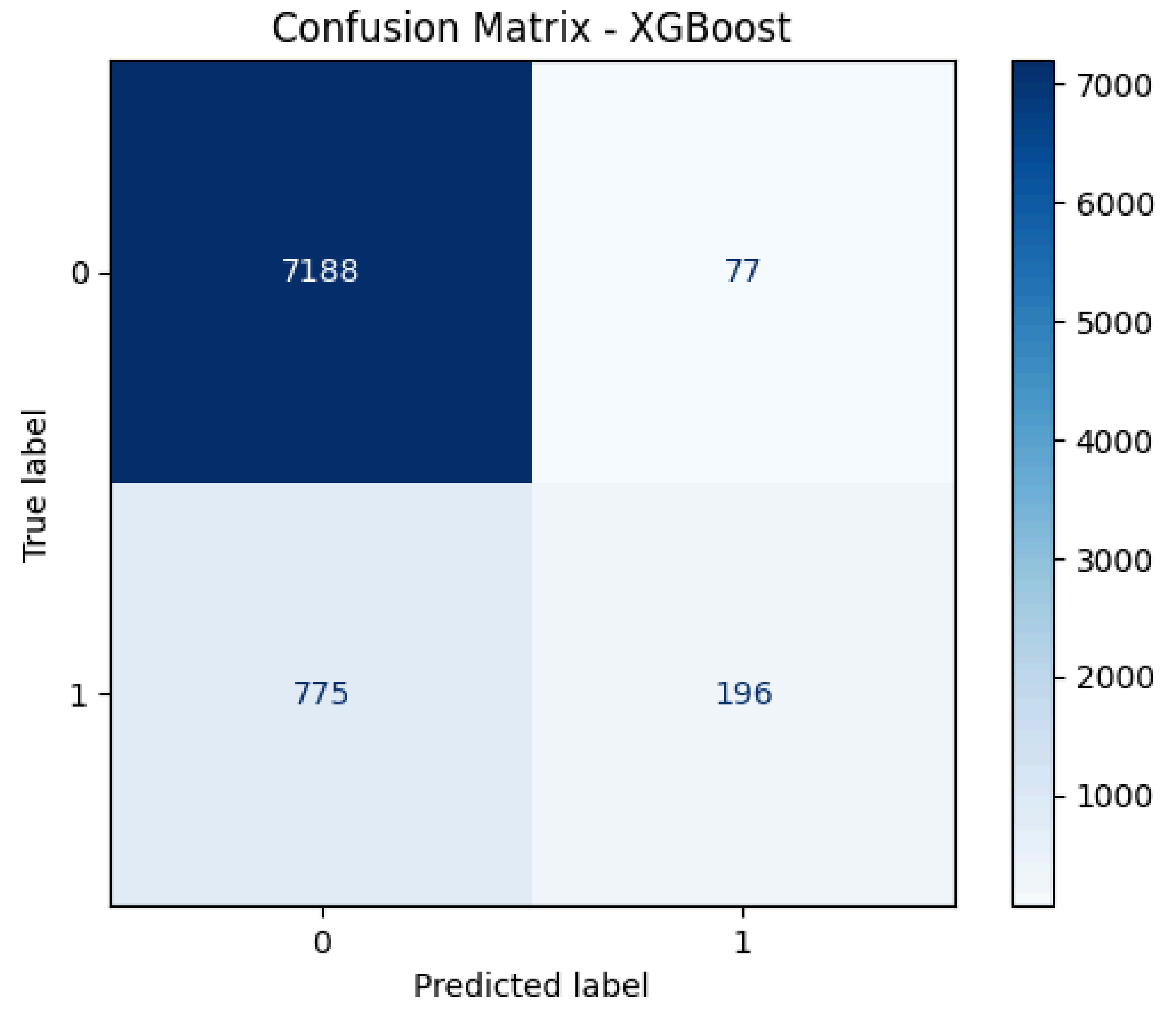
Since one of the campaign objectives is to target as many subscribed customers as possible, Gradient Boosting Tuning is better

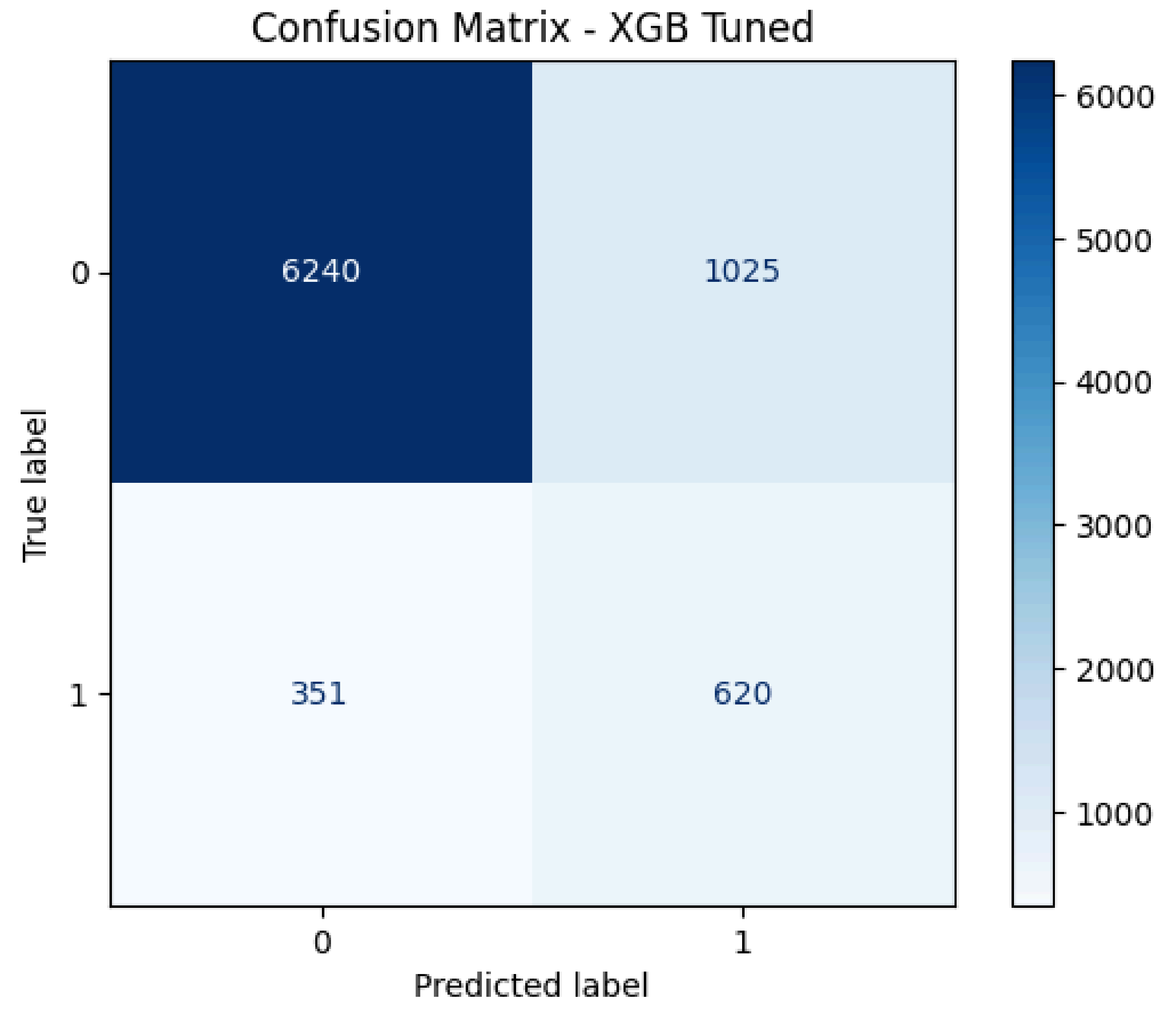




Confusion Matrix - XGBoost

Since one of the campaign objectives is to target as many subscribed customers as possible, Gradient Boosting Tuning is better



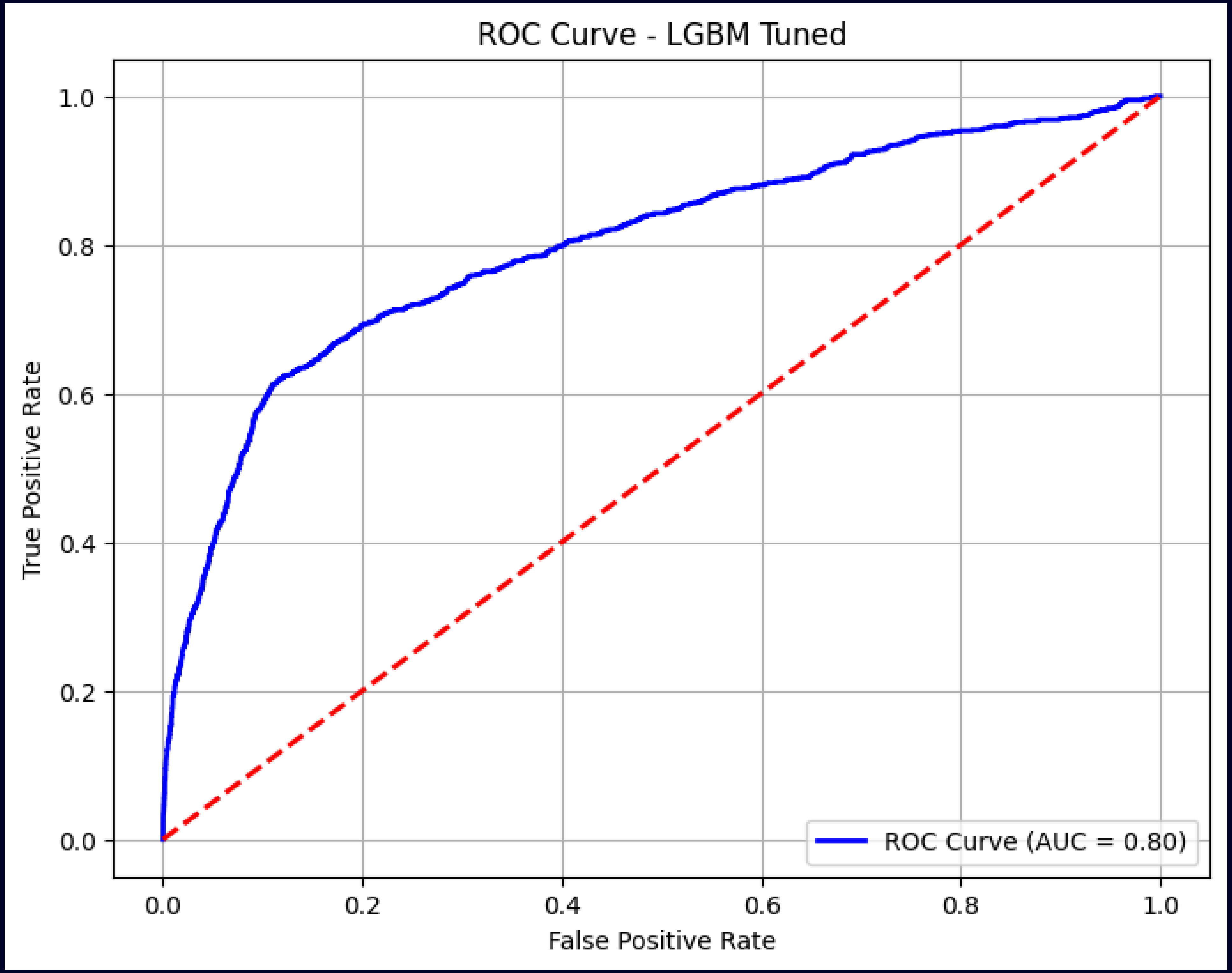


Base Model Classifier Comparison

Since campaign efficiency is a priority, LGBM Tuned is the best model.

Model	True Negatives	False Positives	False Negatives	True Positives
GB Base Model	7185	80	771	200
GB Tuned	6248	1017	362	609
LGBM Base Model	7179	86	768	203
LGBM Tuned	6610	655	428	543
XGB Base Model	7188	77	775	196
XGB Tuned	6240	1025	351	620

Hyperparameter Tuning - ROC Curve - LightGBM Tuning

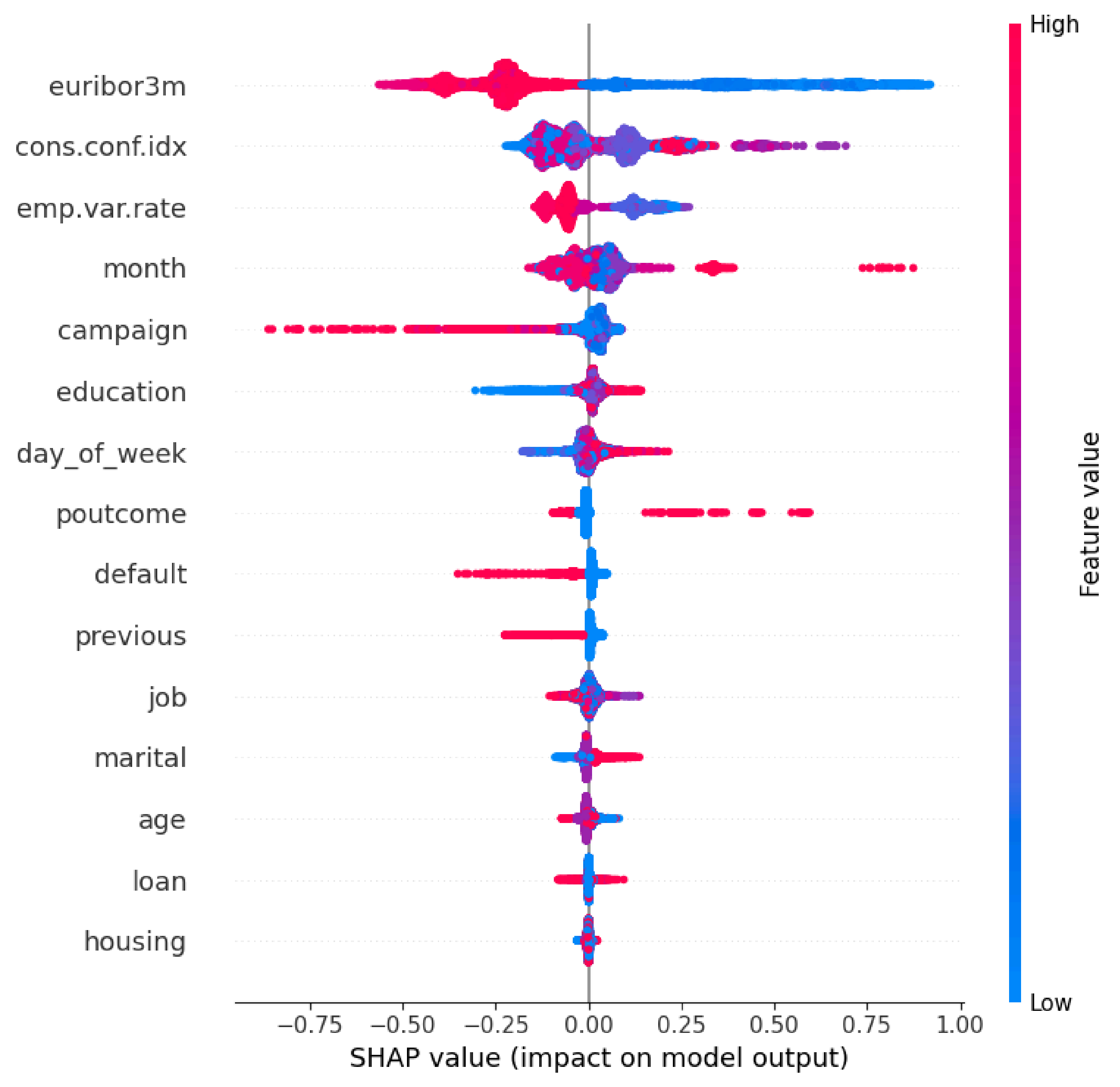


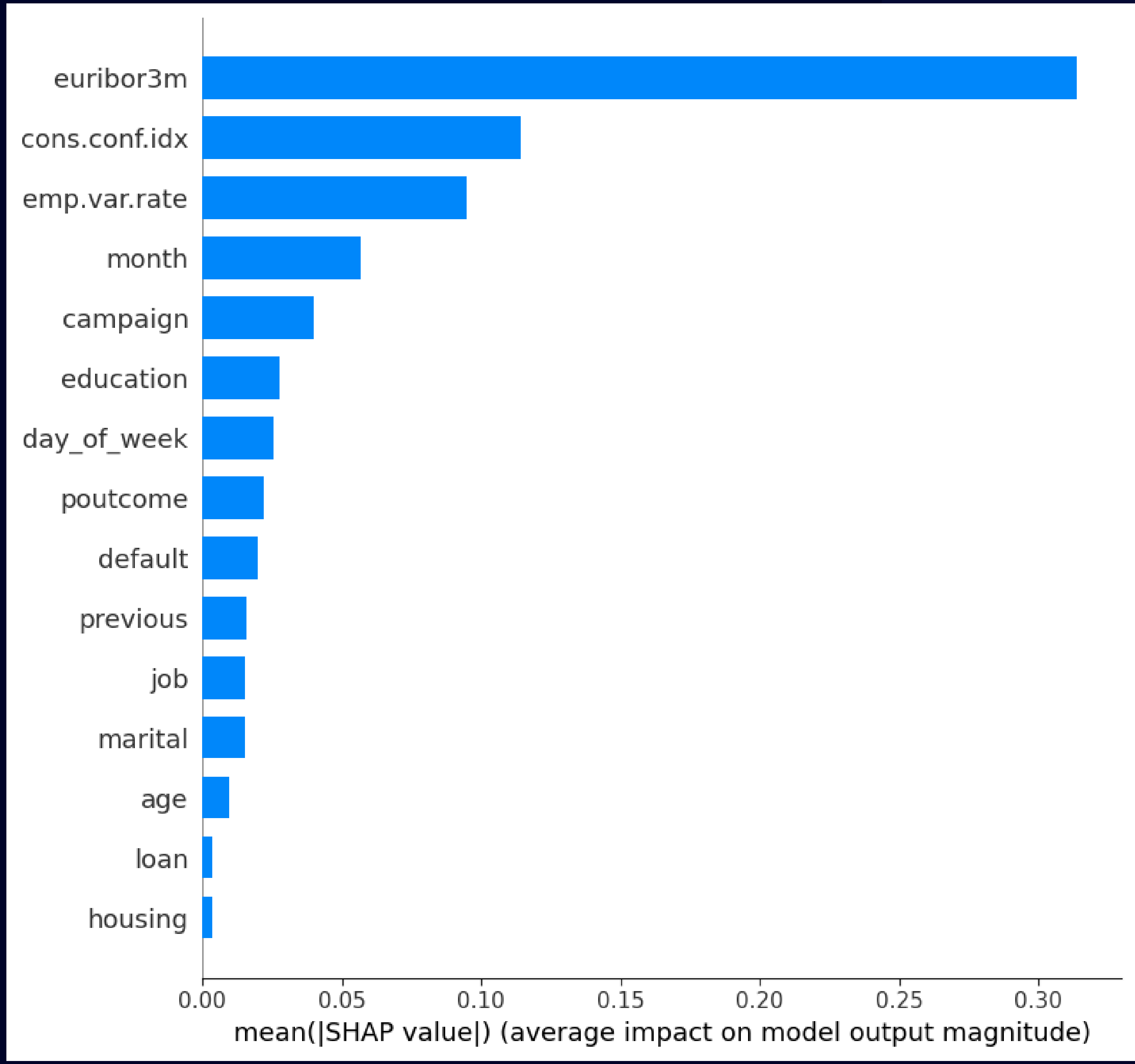
The model does not experience overfitting. The model is good enough

BUSINESS INSIGHT

Best Model Interpretation - LightGBM Tuning

Socio-economic factors, month and campaign are the features with the best contribution.





Business Insight

*143%

35.47%

Conversion Rate

$42.320

Profit

Assumption :

Cost per Customer = $10, Revenue per Customer = $100.

Formula

Profit = (True Positve X Revenue per Customer) - (Predicted Positive X Cost per Customer)

***the percentage increase in profit is from the base model**

RECOMMENDATION

Action Recommendation

- Target Based on Market Conditions (Euribor3m):
  - Focus campaigns on times when Euribor3m is low, as stable financial market conditions are more likely to increase customer conversion rates.
  - Monitor financial conditions regularly to determine the best time to run campaigns.

- Optimize Consumer Confidence (Cons.conf.idx):
  - Campaigns should be intensified when the Consumer Confidence Index is high.
  - Review consumer sentiment data regularly to adjust campaign strategies.

- Analyze Economic Stability (Emp.var.rate):
  - Ensure that campaigns are run when job variations are stable, as good labor market conditions are more likely to increase customer interest.

- Seasonal Strategy (Month):
  - Further analysis is required to determine which months are most effective based on customer behavior patterns.
  - Adjust the campaign frequency to the months that show a positive impact.

- Campaign Frequency (Campaign):
  - Avoid contacting customers too frequently, as overly aggressive campaigns can have a negative impact.
  - Use customer segmentation to optimize the number of contacts.

- Segment Customers Based on Historical Data (Poutcome, Previous):
  - Use previous campaign outcome data to target customers who have a history of positive responses.
  - Avoid customers who have repeatedly not responded positively.

Strategic Recommendation

- **Focus campaigns on customer segments that are in favorable market conditions (low Euribor3m, high cons.conf.idx).**
- **Use seasonal approach and segmentation based on historical data to increase effectiveness.**
- **Reduce campaign frequency on certain segments to avoid customer saturation.**

THANKS!