Bank Marketing Campaign

Opening a Term Deposit

INTAN TJENDERA | M NAUFAL REFADI

A little bit of background...

WHAT IS A

Term Deposit?

Key Takeaways:

- A type of deposit account held at a financial institution where money is locked up for a set period of time.
- Offers higher interest rates than traditional liquid savings accounts, whereby customers can withdraw their money at any time.

source:

https://www.investopedia.com/terms/t/termdeposit.asp

In essence...

By allowing banks to hold onto a deposit for a specific amount of time, they can invest in higher gain financial products in order to make a profit.

They can also hold a better chance to persuade term deposit clients into buying other products such as funds or insurance to further increase their revenues.

Problem Statement & Goals

Problem Statement

Portugal Bank sees a potential decline in revenue as clients are not depositing into the bank as frequently as before

Objectives

- **Identify target customers with higher conversion rate**
- **Minimize marketing costs by mainly focusing on target customers**
- **Appropriate funds and labor allocations**

Supervised, Classification Problem

APPROACH

Create machine learning (ML) algorithm to achieve efficient marketing and also give an insight to banks how effective direct phone call campaigns are, as compared to other sales channels; and vice versa

Evaluation method: Accuracy, Precision, Recall, F1-score
Focus on Recall scores

Risks include overestimation of target customer behavior despite customer having all aspects of being a 'potential client'

Data Understanding & Data Cleaning

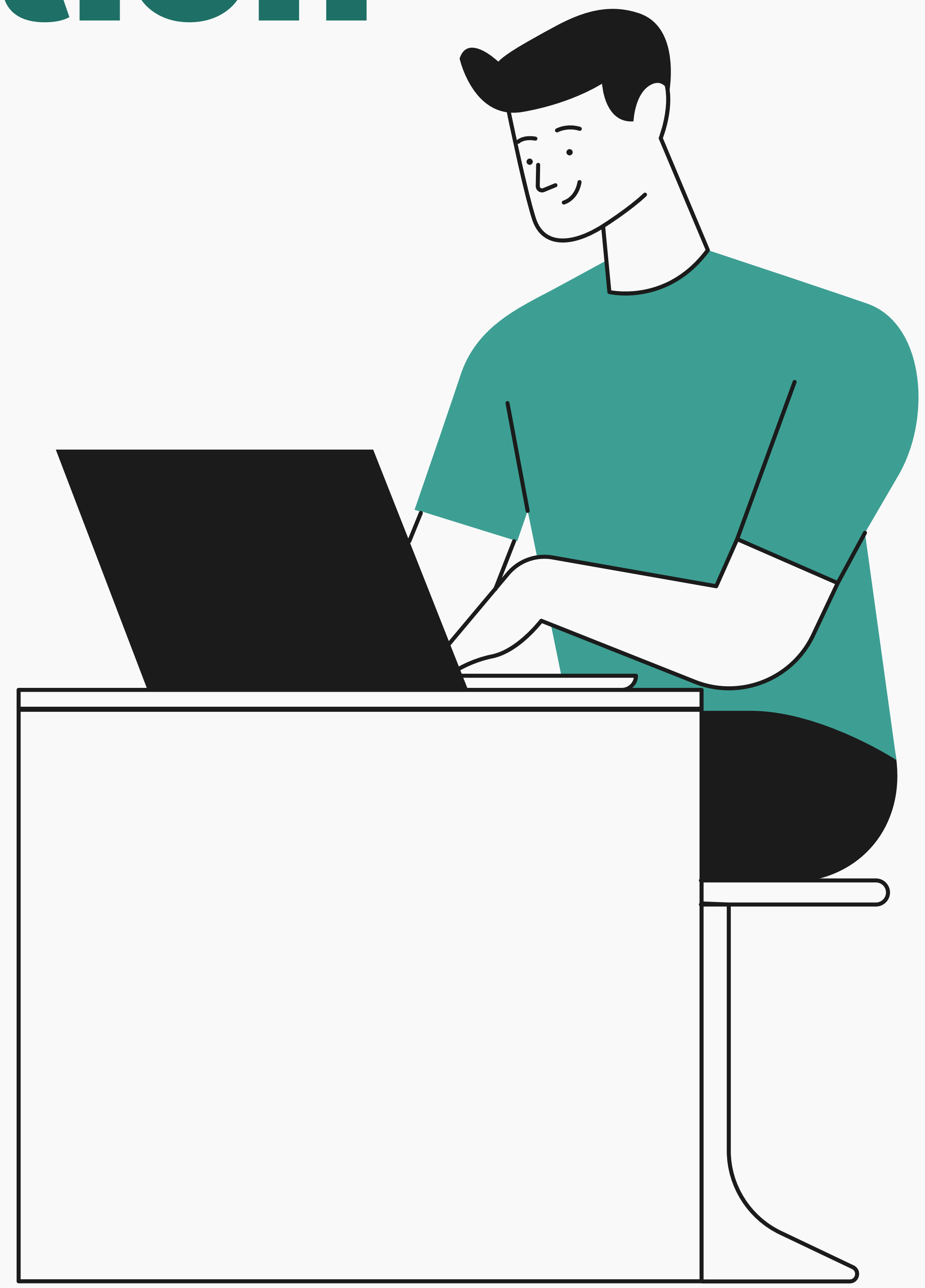
Data Collection

Data Source:
https://www.kaggle.com/volodymyrgavrysh/bank-marketing-campaigns-dataset

Data Description:
A dataset that describes the Portugal bank's marketing campaigns results. Conducted campaigns were based mostly on direct phone calls, offering bank clients to place a term deposit.

Number of Columns:
20 + output variable

Number of Rows:
41188



Data Understanding

- **age** : client's age
- **job** : type of job
- **marital** : marital status
- **education** : last education
- **default**: has credit in default?
- **housing**: has housing loan?
- **loan**: has personal loan?
- **contact**: contact communication type
- **month:** last contact month of year
- **day*of*week**: last contact day of the week
- **duration**: last contact duration, in seconds.
- **campaign**: number of contacts performed during this campaign and for this client
- **pdays:** number of days that passed by after the client was last contacted from a previous campaign
- **previous:** number of contacts performed before this campaign and for this client
- **poutcome**: outcome of the previous marketing campaign
- **emp.var.rate**: employment variation rate - quarterly indicator
- **cons.price.idx**: consumer price index - monthly indicator
- **cons.conf.idx**: consumer confidence index - monthly indicator
- **euribor3m**: euribor 3 month rate - daily indicator
- **nr.employed**:number of employees - quarterly indicator
- **y** : has the client subscribed a term deposit?

Check and drop columns with multiple null values

Check and drop columns with duplicated data (No null values)

DataFrame shape **BEFORE** null and duplicate check:
(41188 , 21)

DataFrame shape **AFTER** null and duplicate check:
(41176 , 21)

Data Grouping

'education' column

Categories 'basic.4y', 'basic.6y', 'basic.9y' are grouped together under 'Basic'

'age' column

Ages of the customers are grouped into 4 classes:

- 17-29 = Young
- 30-43 = Middle-age Adult
- 44-56 = Old-age Adult
- 56-100 = Elderly

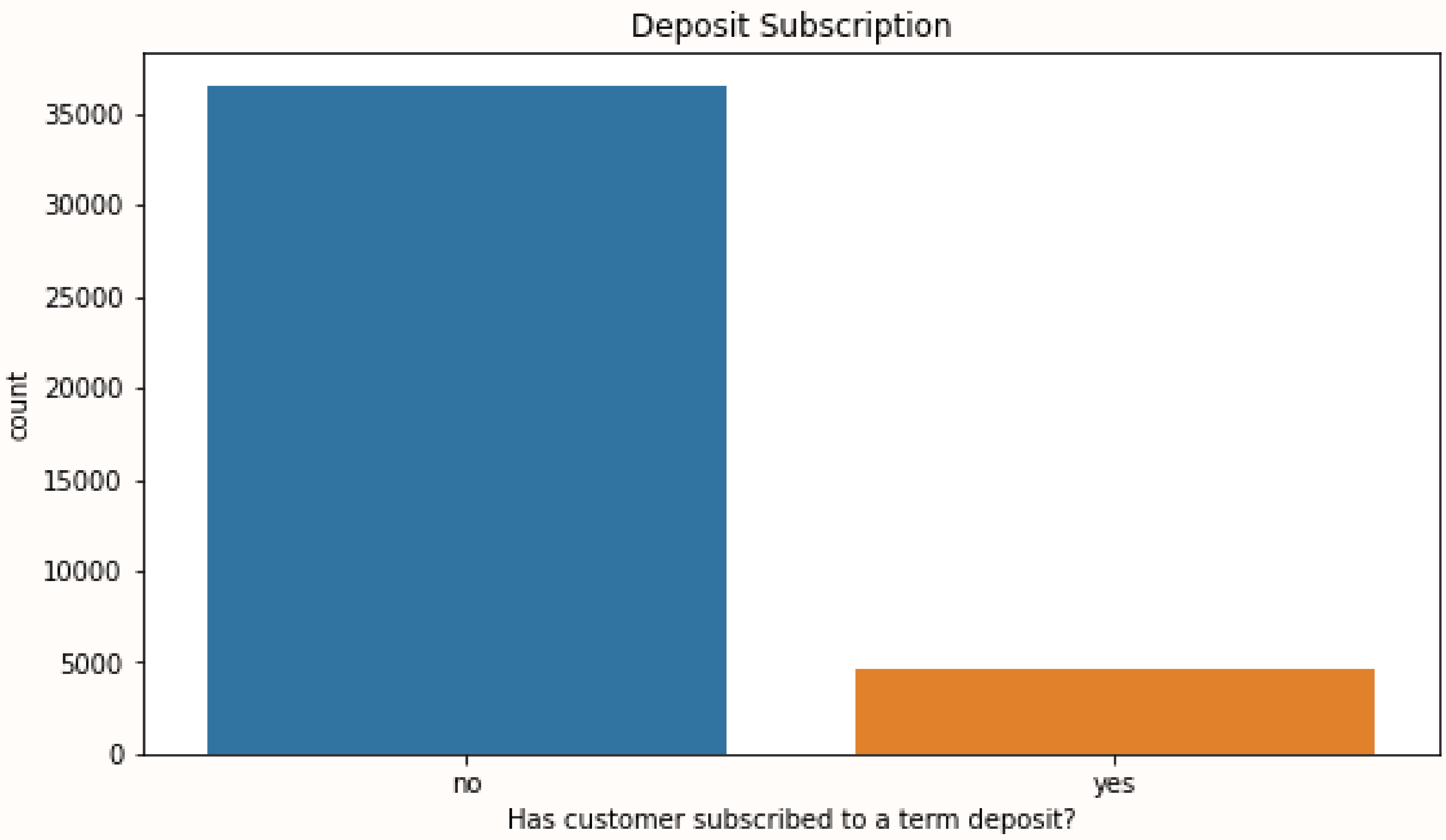
'pdays' column

Updated into binary numerical feature:

- '0' for '999' / no previous contact with clients
- '1' for previous contact made / other values aside '999'

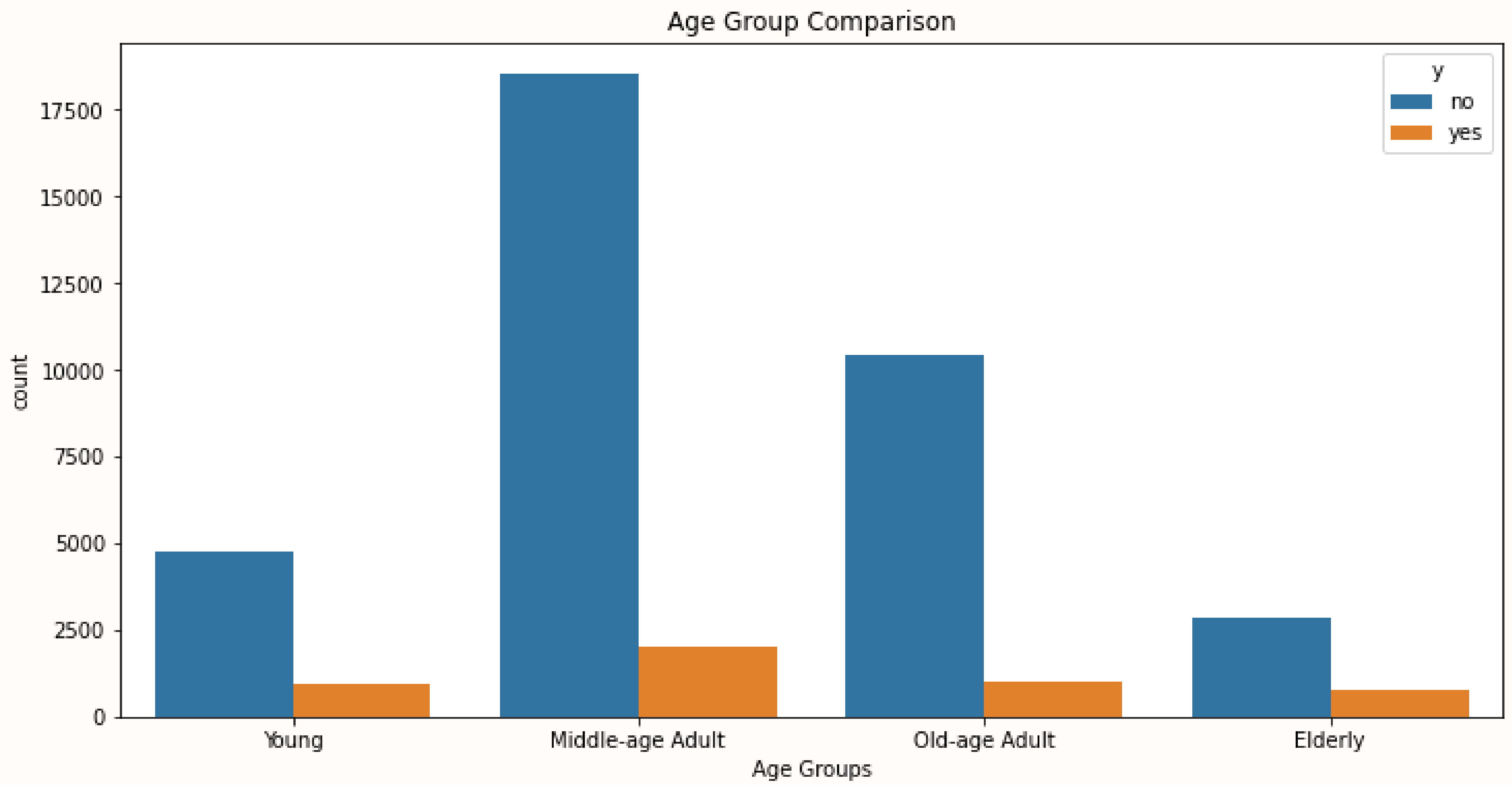
Exploratory Data Analysis

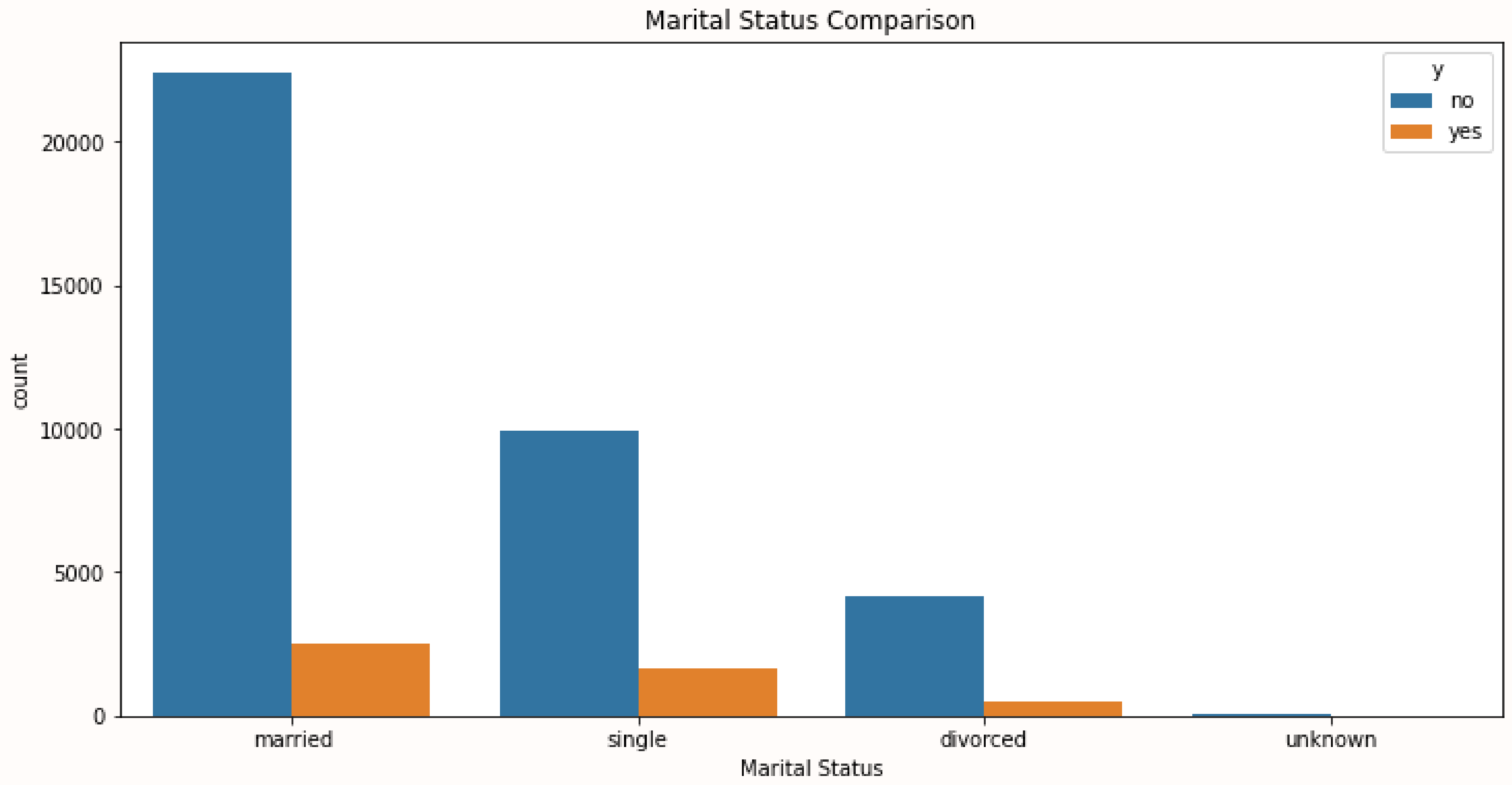
Distribution of Target Variable



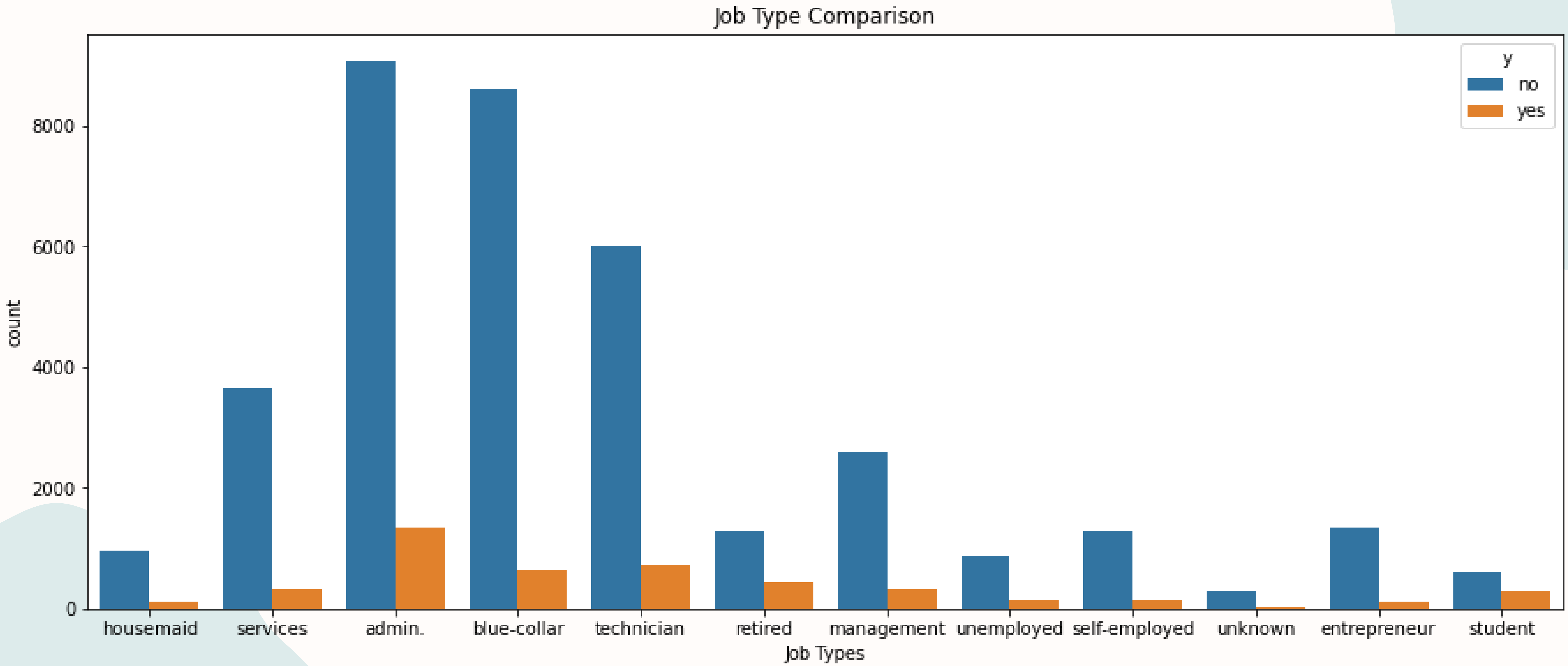
From the graph, the success/conversion rate for term depositors without a ML model is about 11 out of every 100 calls (~11%).

Target Variable Distribution Based On Other Features



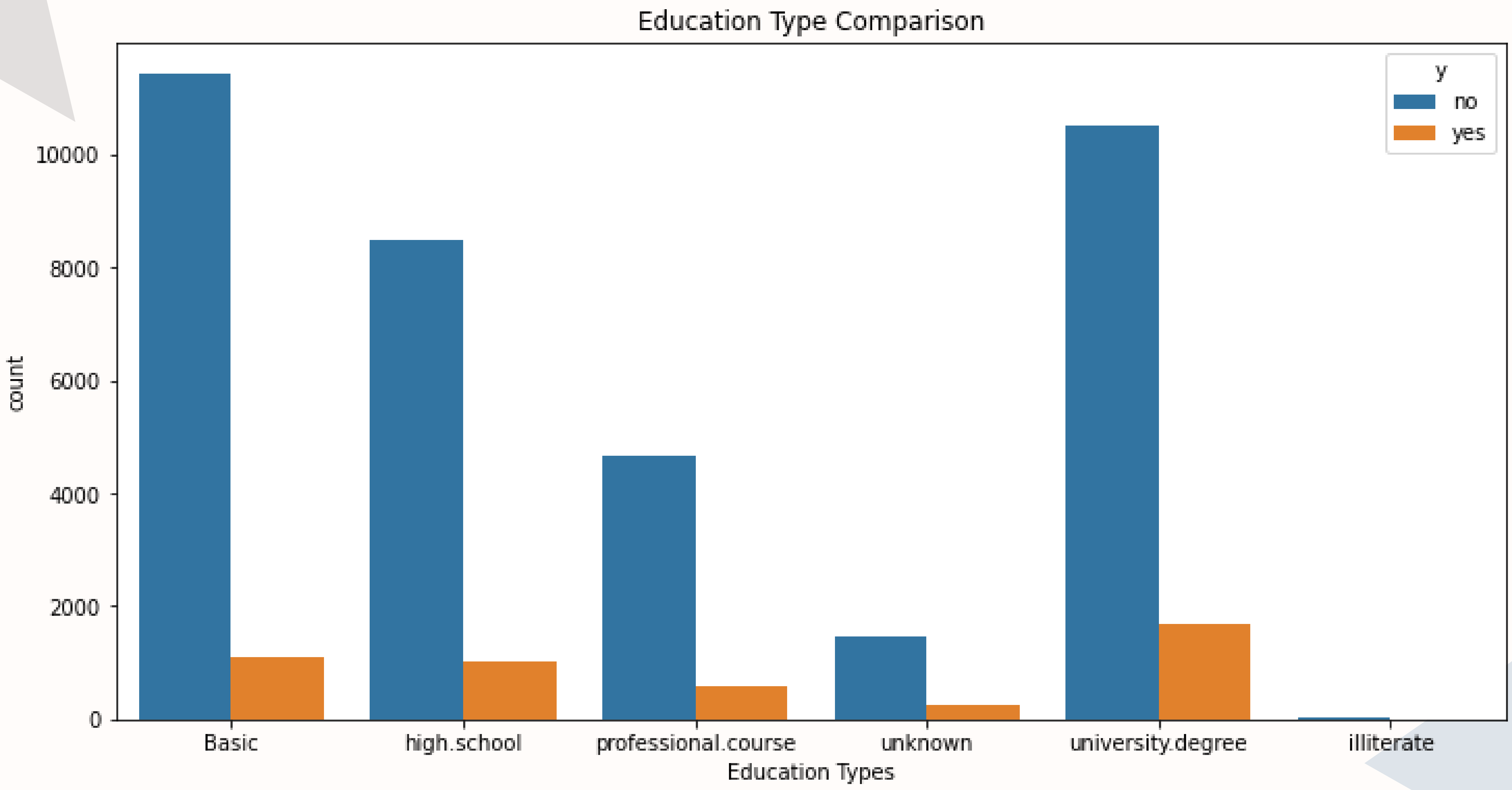


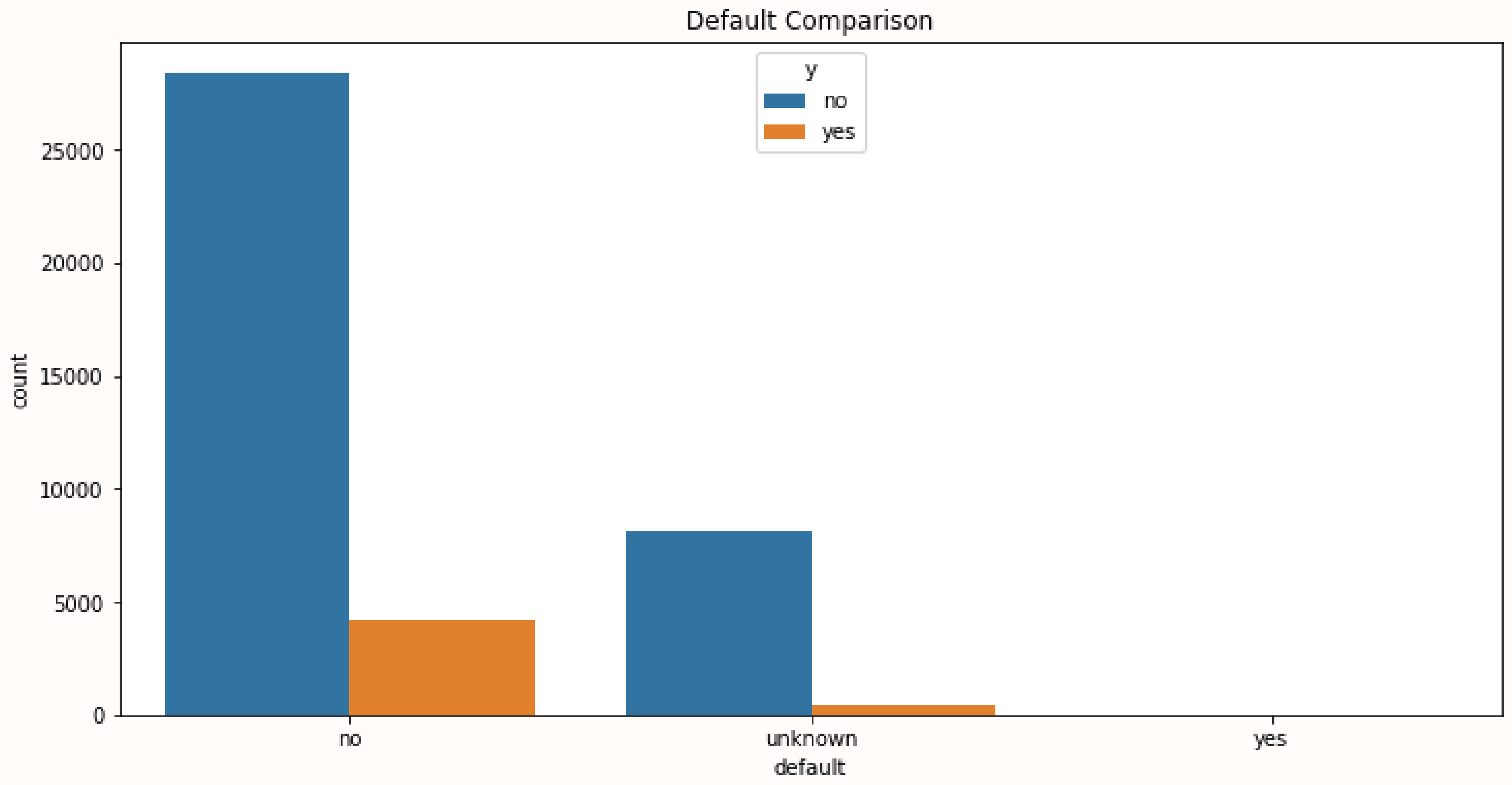
Success Rate
based on Age and Marital Status

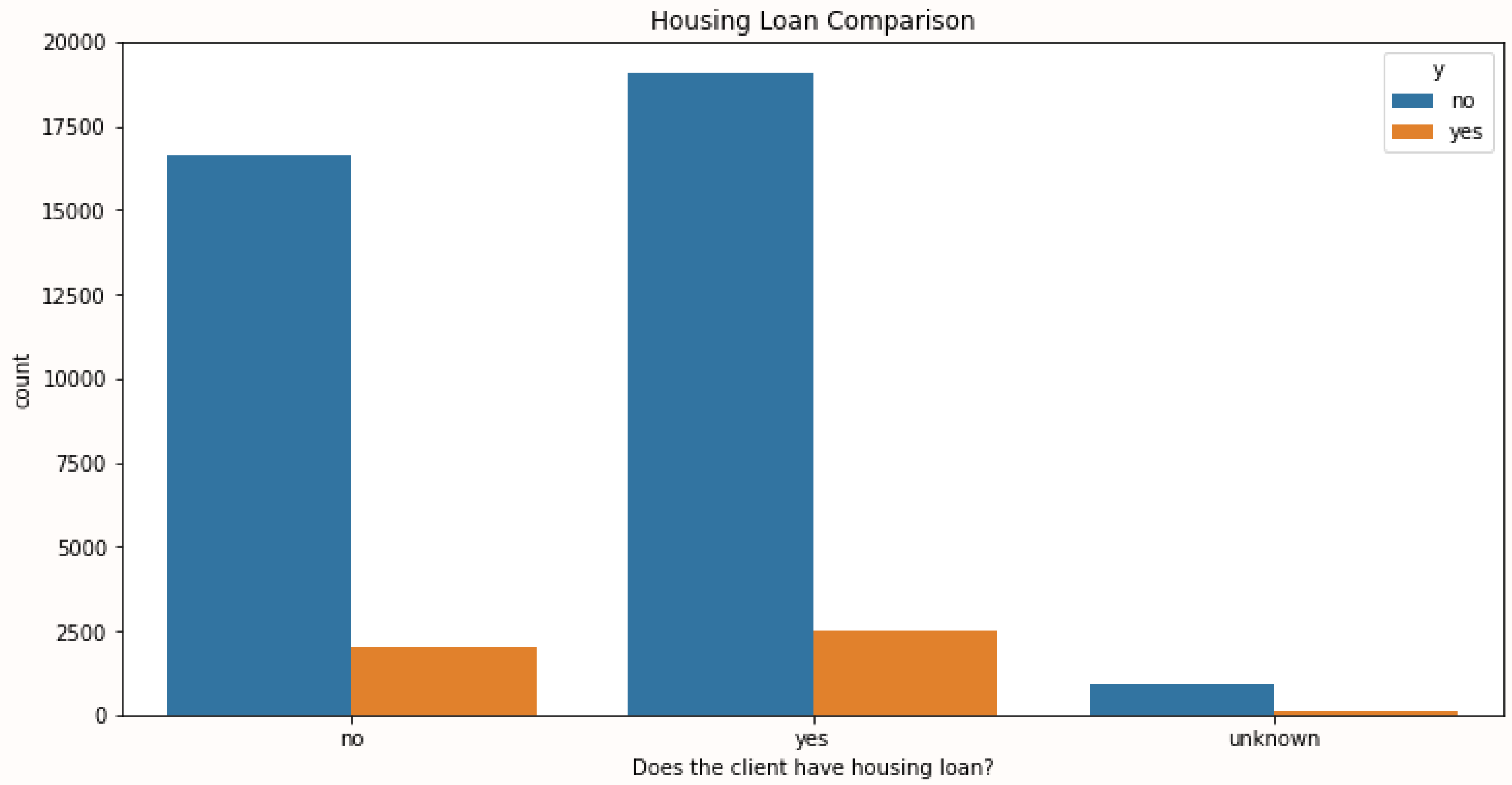


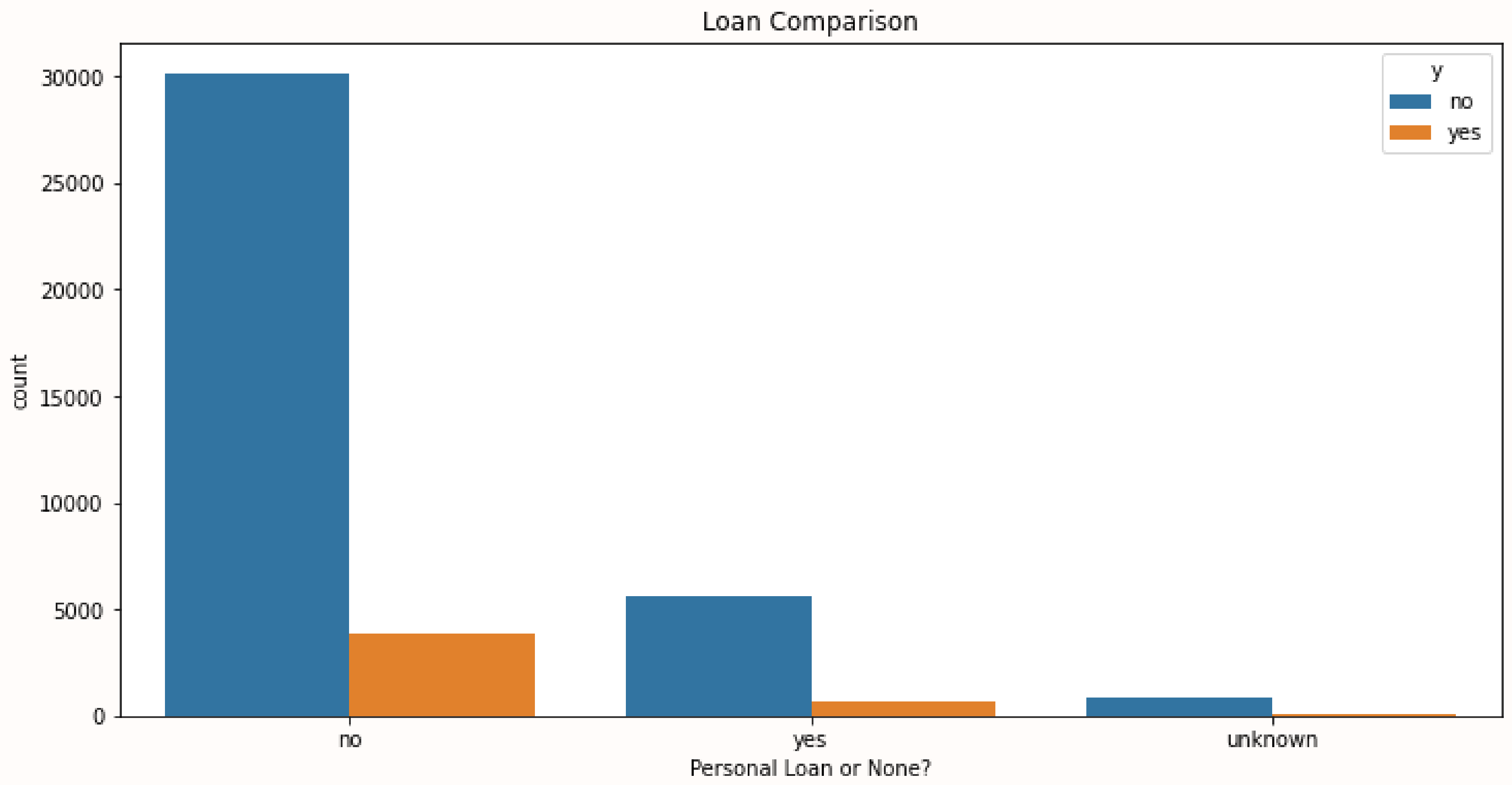
'student' (31.43%) or 'retired' (25.23%) are more likely to make a term deposit. Least likely to make a term deposit are 'blue-collared' clients (6.89%).

Aside from clients who are 'illiterate' or whose education background is 'unknown', those with 'university degree' have the next highest conversion rate



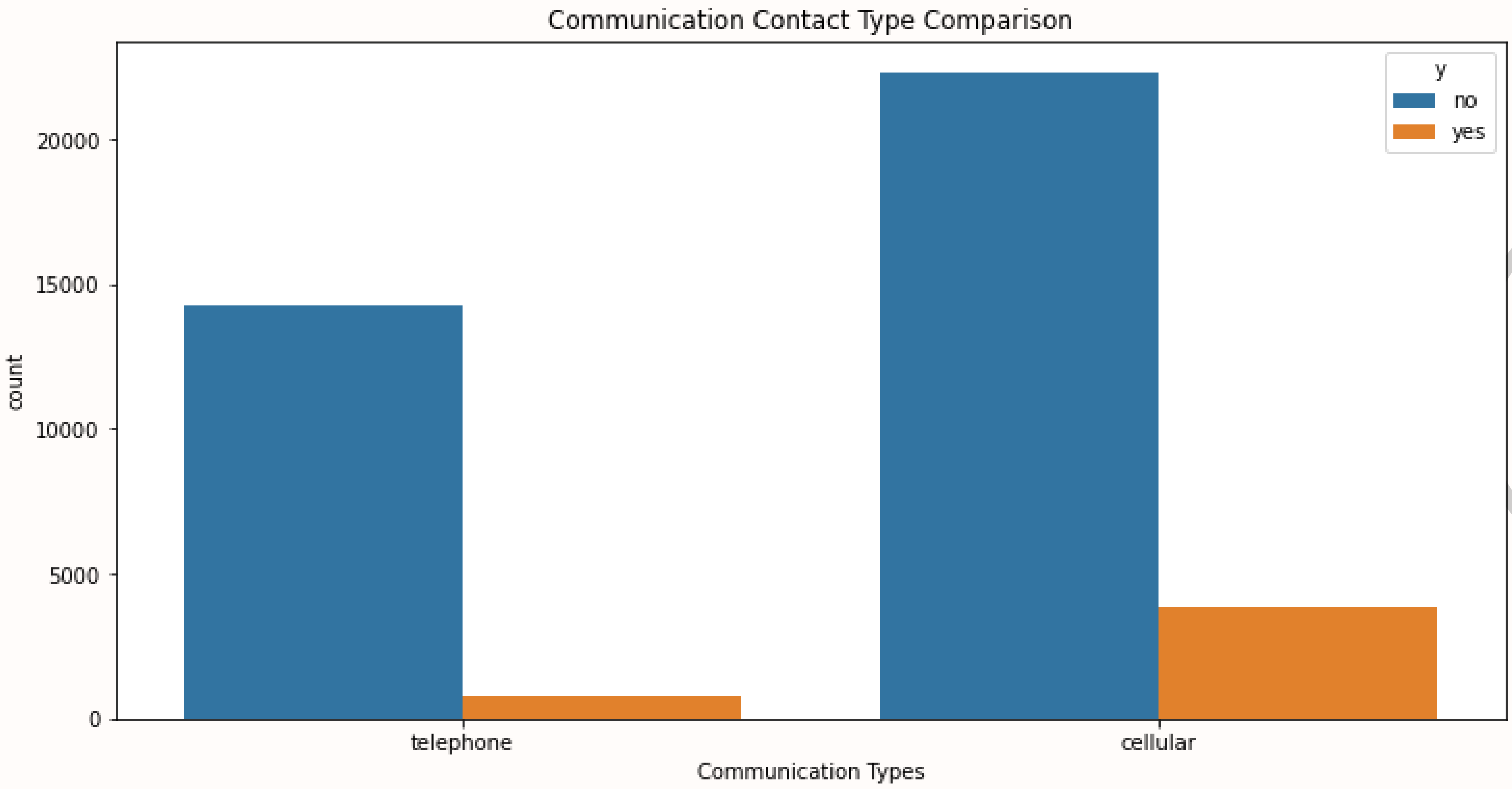




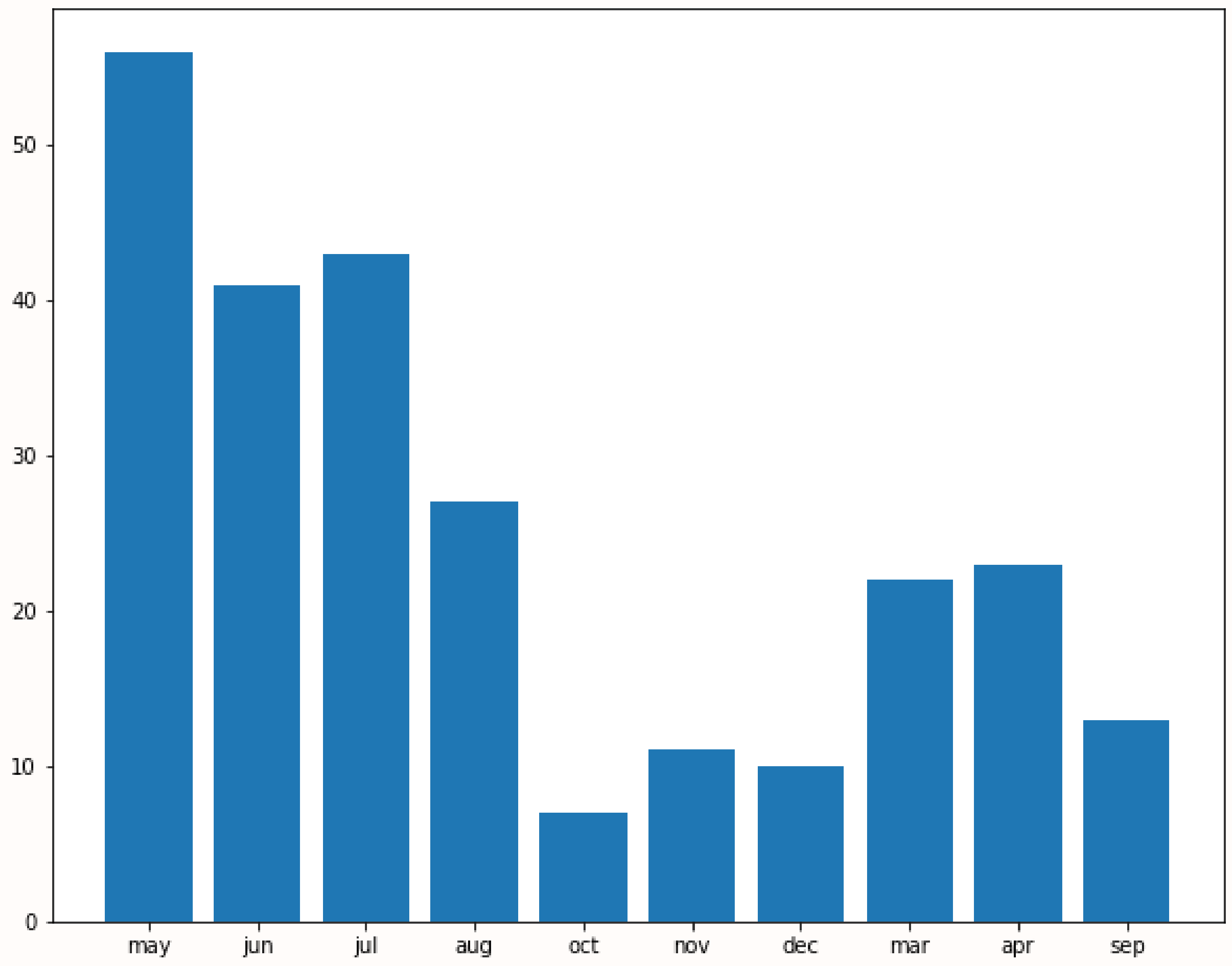


Success Rate based on Credit, Housing or Personal Loan Possession

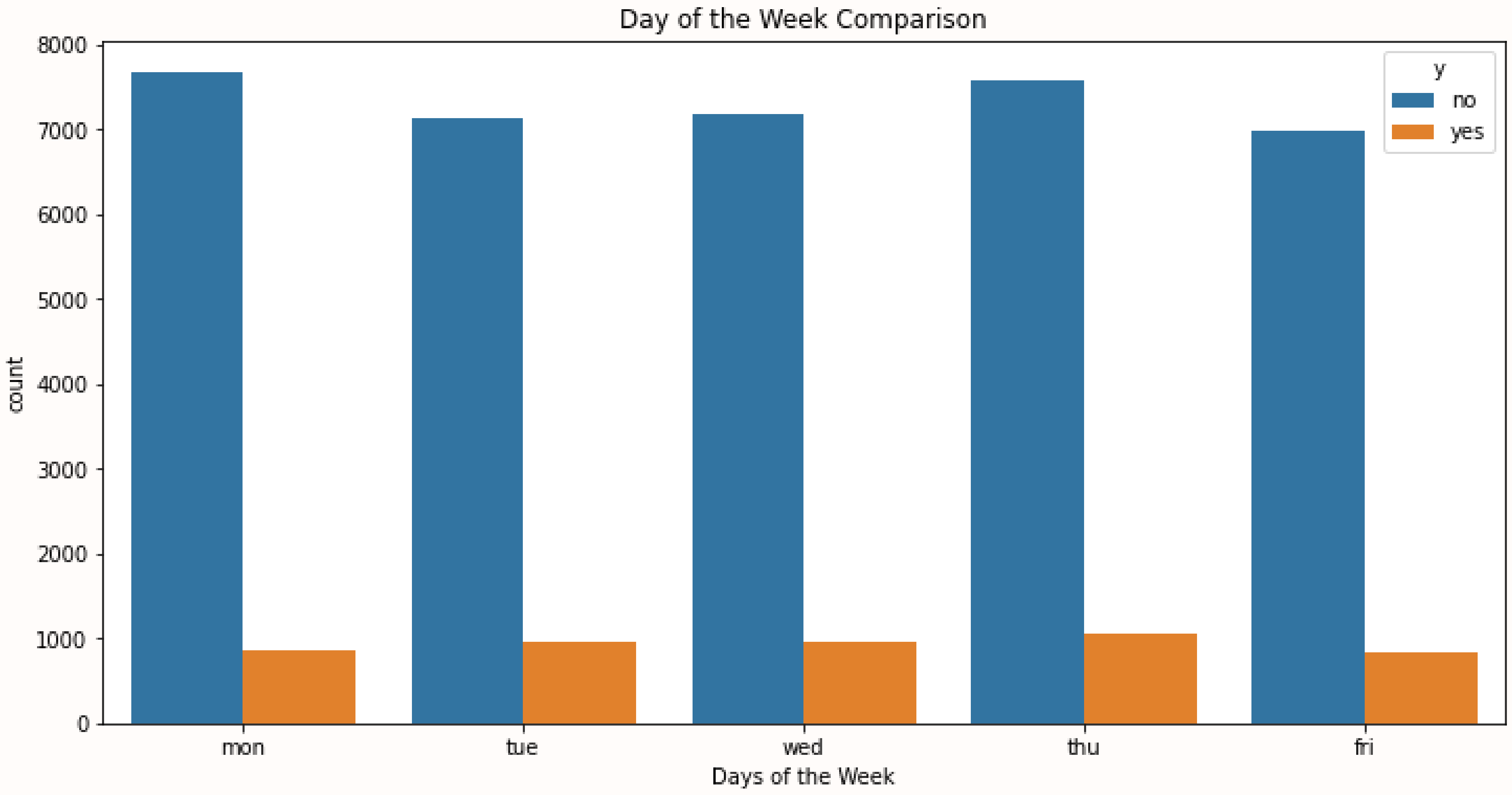
Success rate relatively higher for marketing via cellular means



- Clients are mostly contacted during May - July period; Jan and Feb details are missing
- Conversion rate highest for Sept, Oct and Dec
- Equal success rate regardless of the day of the week
- Campaign only carried out during weekdays

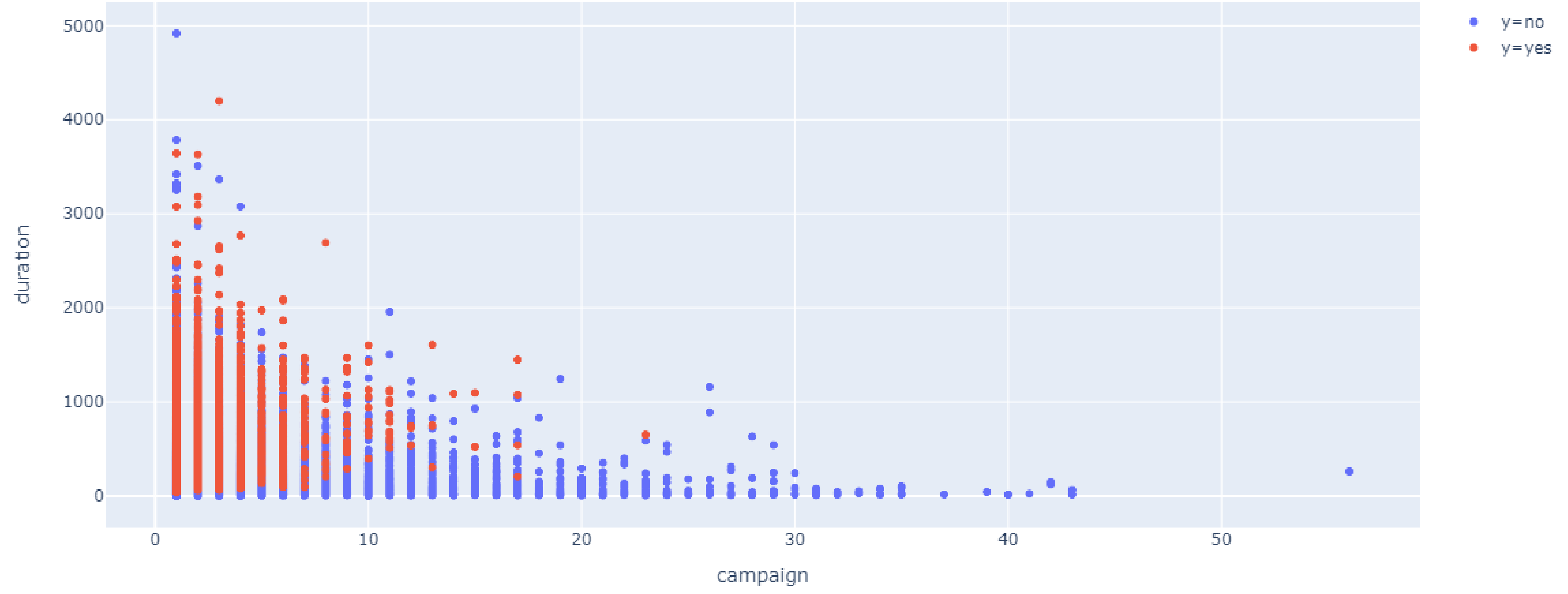


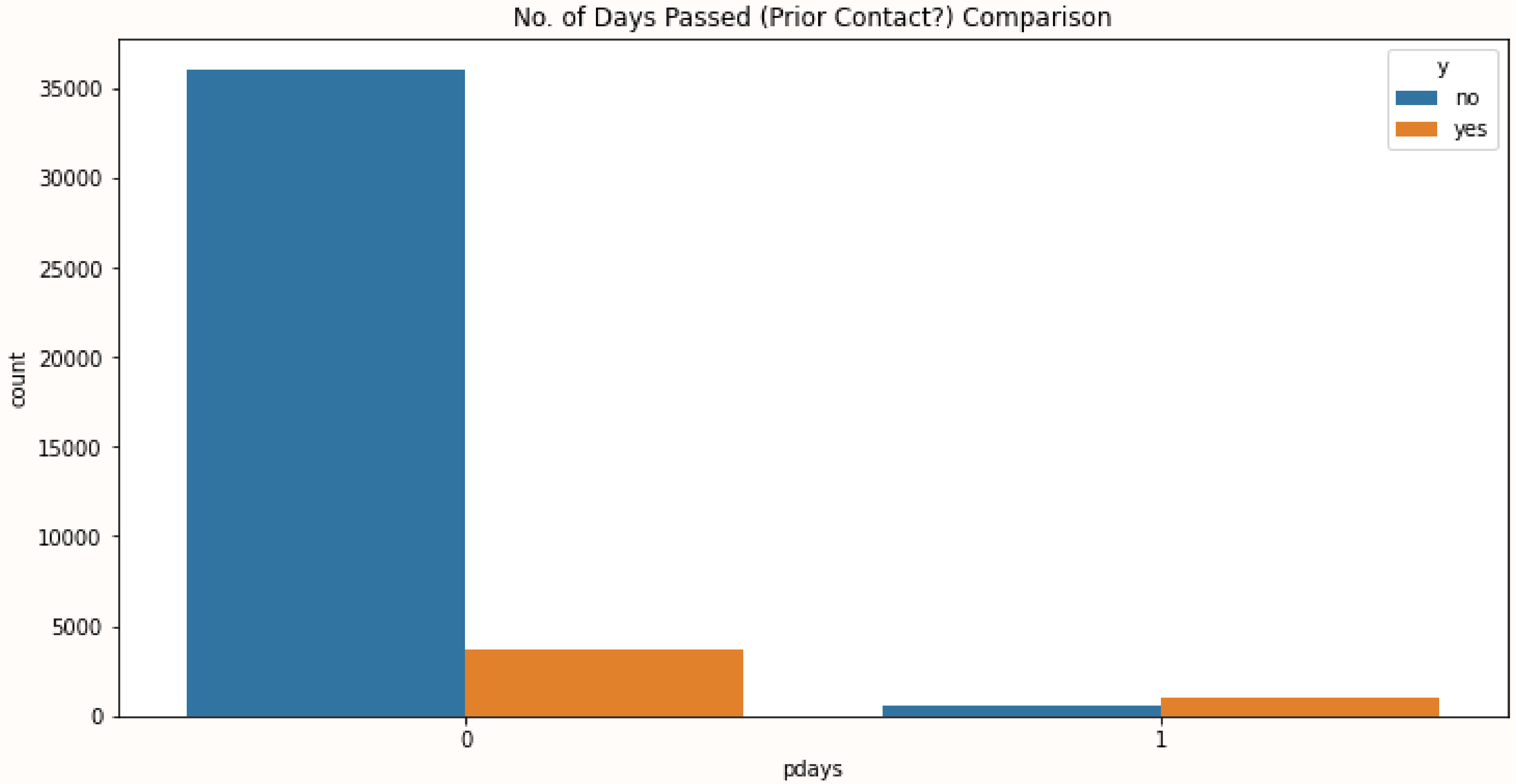
Month vs Campaign



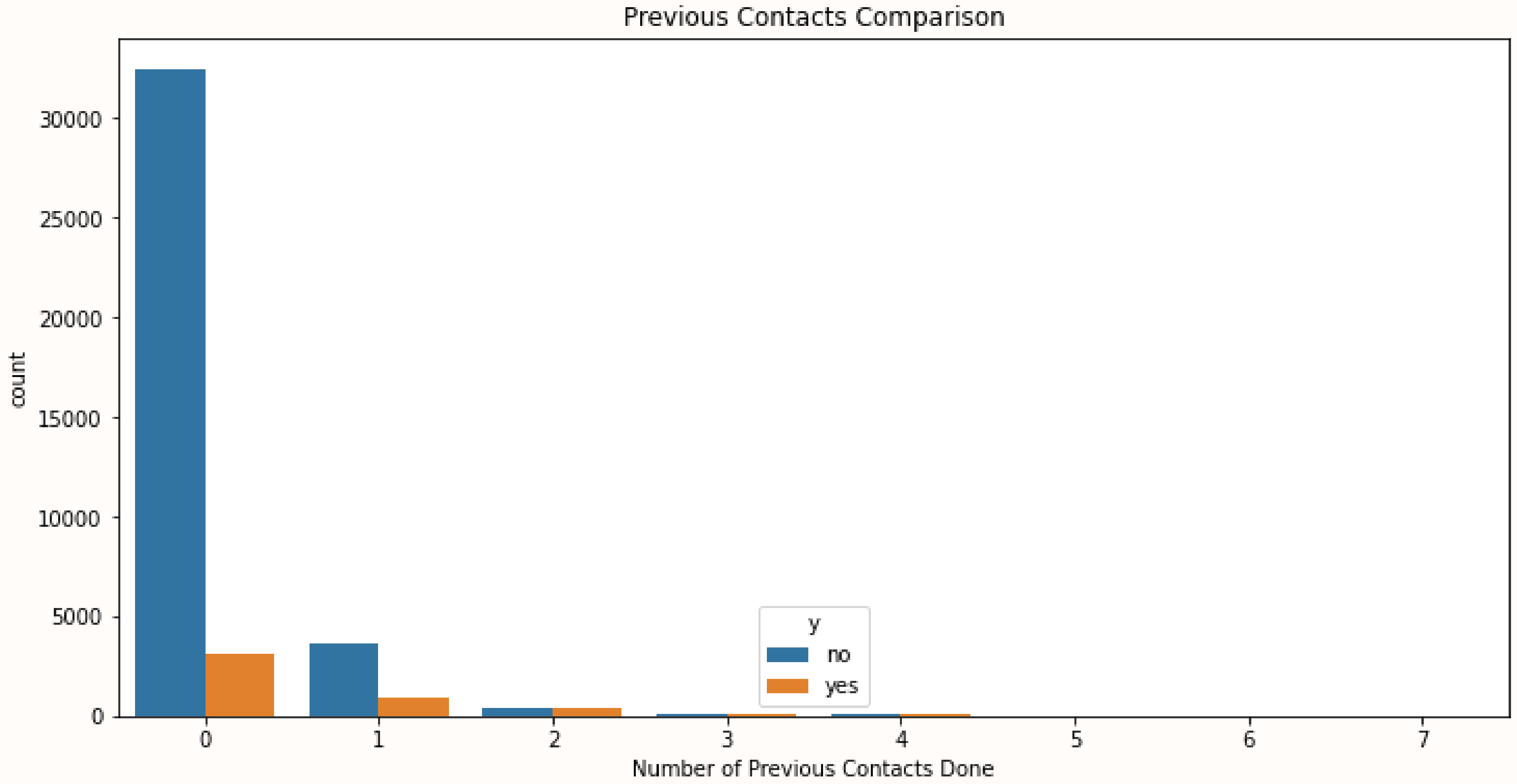
Success Rate based on Day of the Week

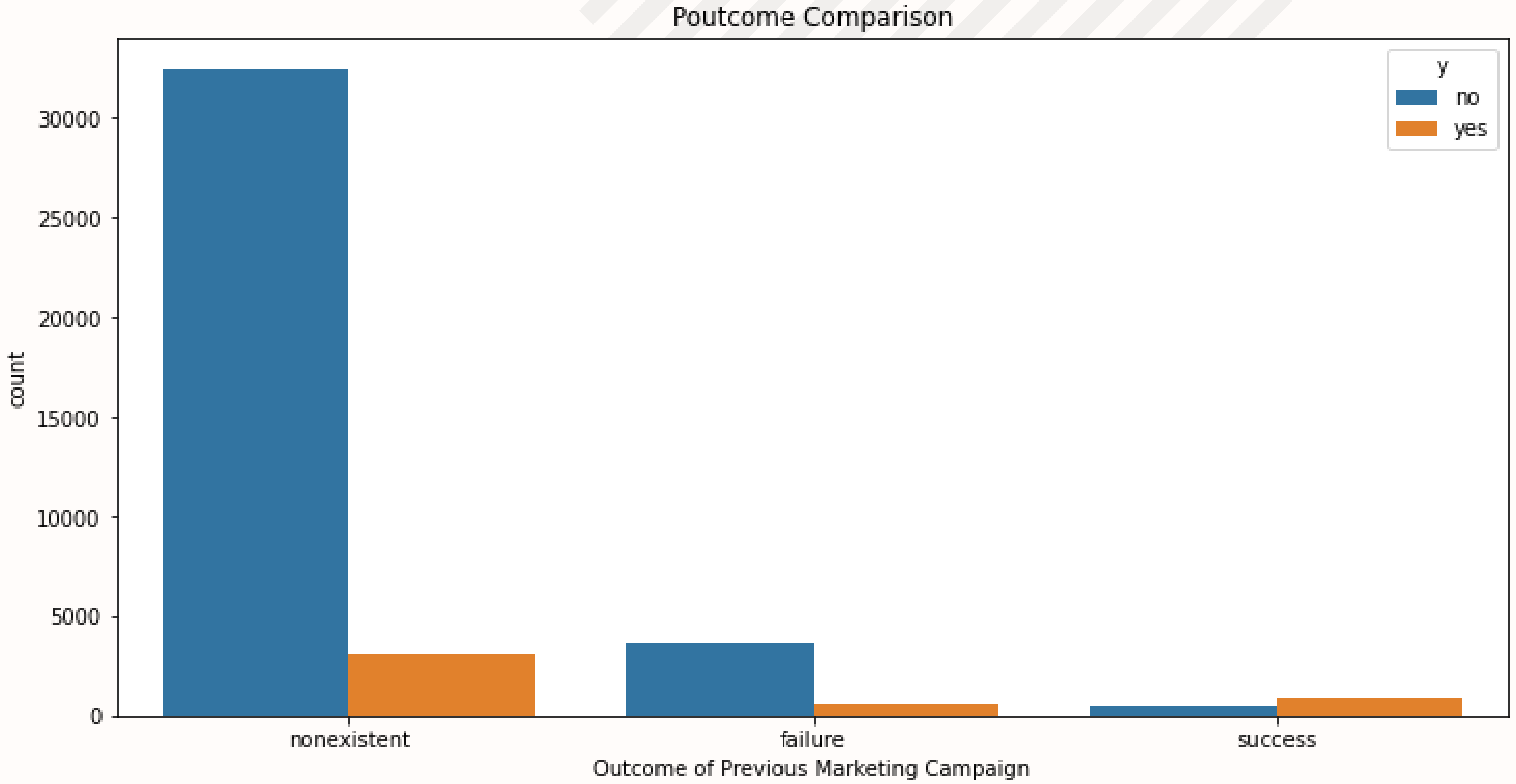
Campaign vs Duration



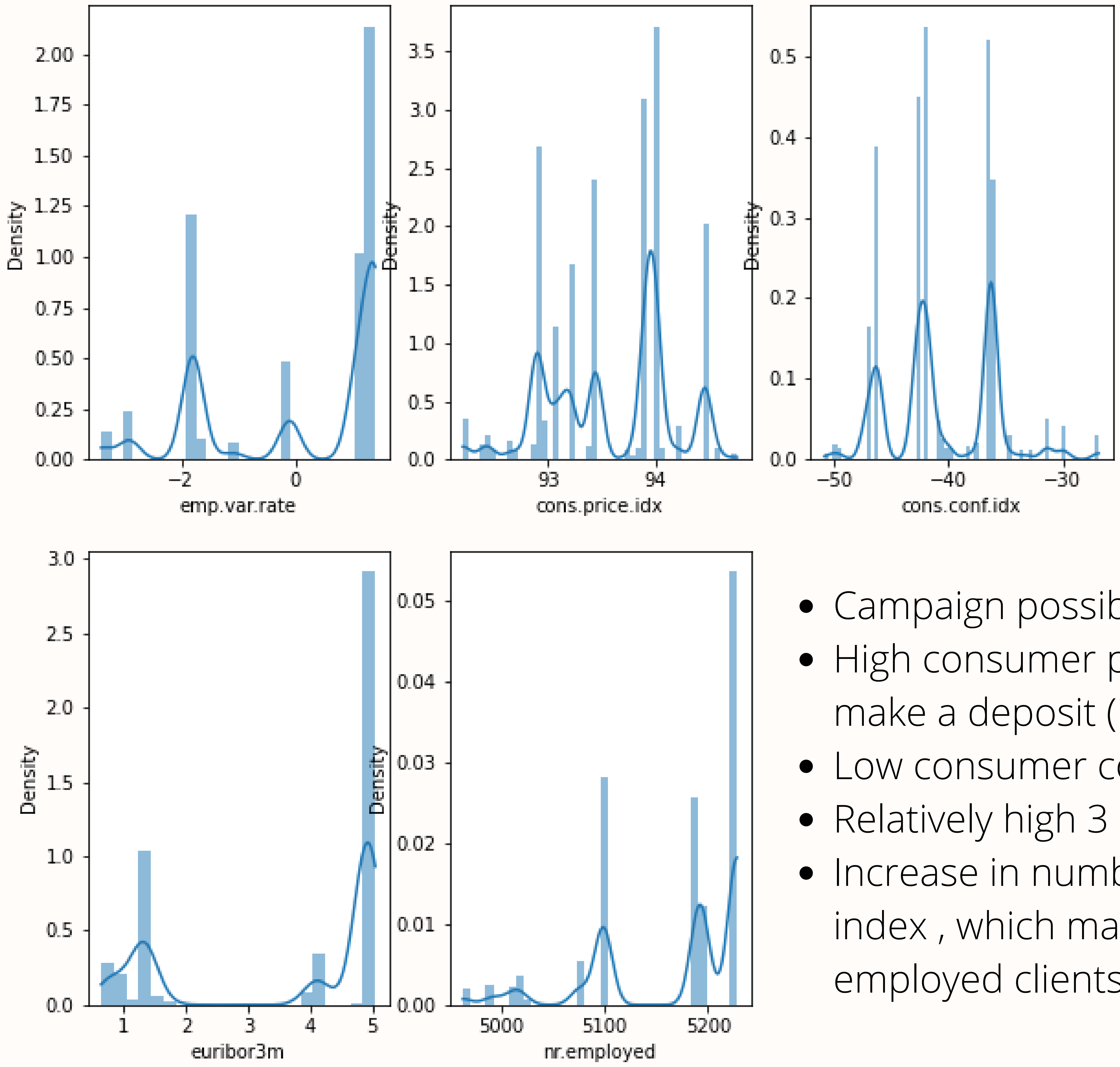


Customer conversion rate higher for those that had previous contact (~5x max) and were a success in a previous marketing campaign





Socio-Economic Attributes Distribution



- Campaign possibly made during high employee variation rate
- High consumer price index may be a stimulant for clients to make a deposit (idea of savings)
- Low consumer confidence index due to fluctuating economy?
- Relatively high 3 months Euribor interest rate
- Increase in number of employees result in higher income index , which may be a target factor for campaign to get employed clients to make a term deposit

ML Modelling

Term Depositors Predictor Modelling Steps

Feature Selection

Decide whether or not to drop speicifc features

Encoding

Scaling numerical columns
OneHot and Binary encode categorical features

Model Training & Selection

Cross Validation
Select best model

Model Tuning & Evaluation

Confusion Matrix
ROC - AUC Score
Feature Importance

Feature Selection

Two Approaches

Dropped:
'age'
'marital'
'education'
'housing'
'duration'
'dayofweek'
'loan'

No features dropped aside from 'age' and 'duration'

- 'duration' is dropped for both approaches in order to achieve a more realistic model
- 'age' dropped since new categorical column 'agegroup' made

Encoding

1

NUMERICAL ENCODING

StandardScaler

2

CATEGORICAL ENCODING

OneHotEncoder

3

BINARY ENCODING

BinaryEncoder

The following encoding are applied to both approaches.

Encoding

Split Train & Test Sets

Technique used to alleviate imbalanced data

Random OverSampling

Model Training

1. Logistic Regression
2. K-Nearest Neighbor
3. Random Forest
4. Gradient Boosting

Model Training

with dropped feature

	Train Accuracy	Test Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.797990	0.797950	0.742661	0.448324	0.558977
K-Nearest Neighbor	0.869314	0.821157	0.676585	0.715644	0.695290
Random Forest	0.925812	0.879209	0.786622	0.792008	0.789071
Gradient Boosting	0.816492	0.815154	0.744136	0.538175	0.624532

Model Training

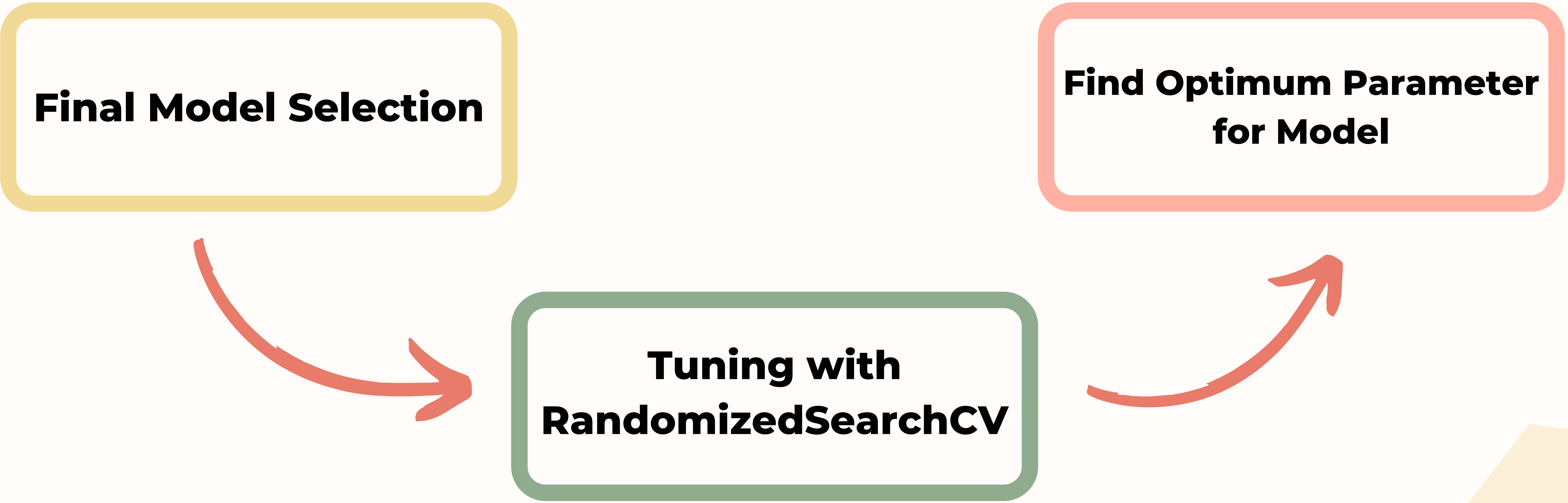
without dropped feature

	Train Accuracy	Test Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.800167	0.799561	0.740782	0.459260	0.566880
K-Nearest Neighbor	0.891895	0.824622	0.666257	0.772188	0.715096
Random Forest	0.984986	0.939507	0.854782	0.949591	0.899427
Gradient Boosting	0.815949	0.813543	0.740438	0.534934	0.621024

	Train Accuracy	Test Accuracy	Precision	Recall	F1-Score
Random Forest Without Drop	0.984986	0.939507	0.854782	0.949591	0.899427
Random Forest With Drop	0.925812	0.879209	0.786622	0.792008	0.789071

- Best model from the two approaches obtained = Random Forest
- Recall scores compared
- Final model selected = Random forest without any dropped features

Hyperparameter Tuning & Evaluation



- Decrease the execution time needed to compile
- Chances to find an optimized parameter comparatively higher as it tests all patterns randomly under the assumption that not all hyperparameters are equally important
- Reduces chance for overfitting to occur, despite potentially high variance due to being random

Confusion Matrix Plot

For better understanding, assume:

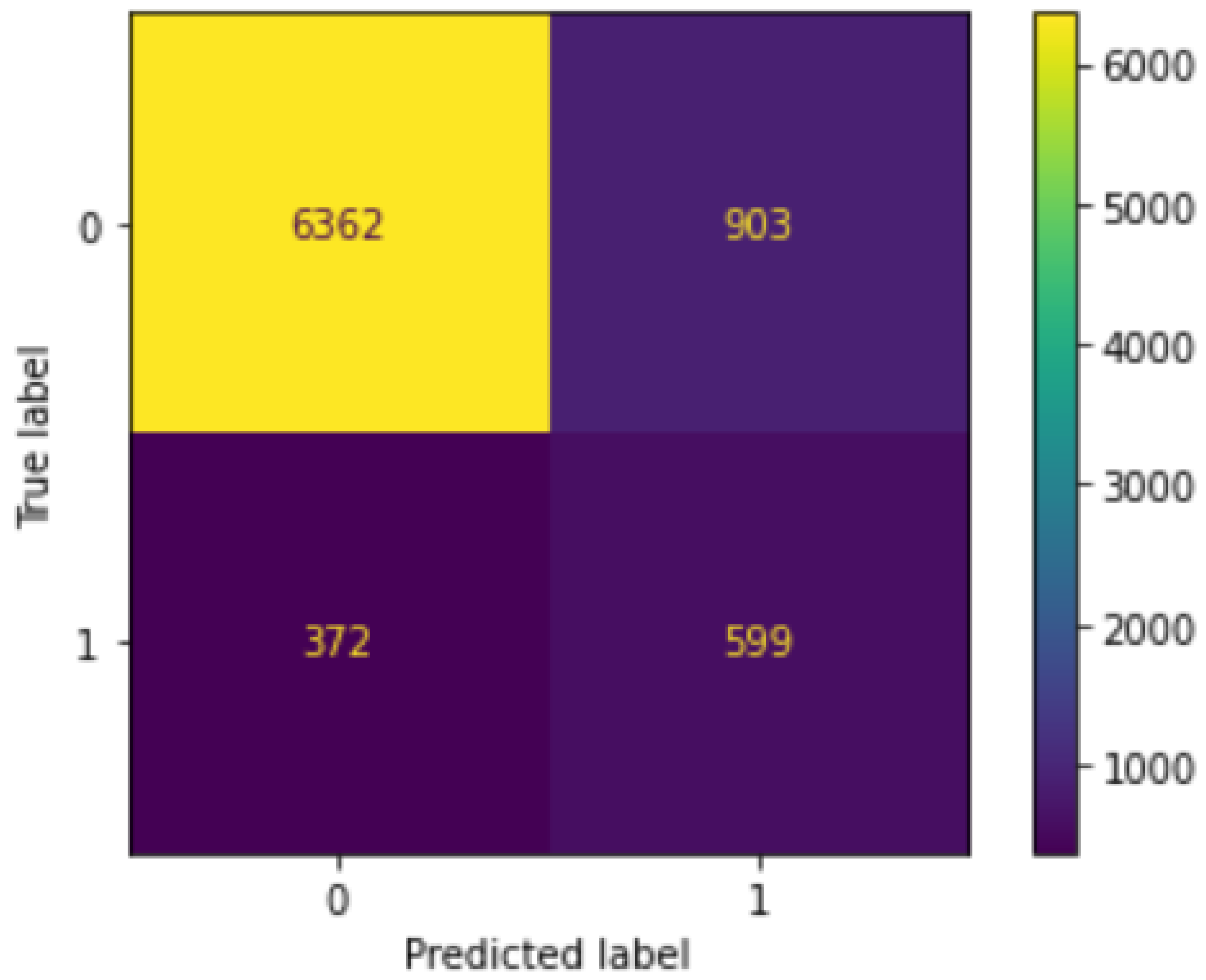
- Call cost = $1
- Profit from conversion = $10
- Revenue = $9

Total Revenue w/o Model Use = $1,474

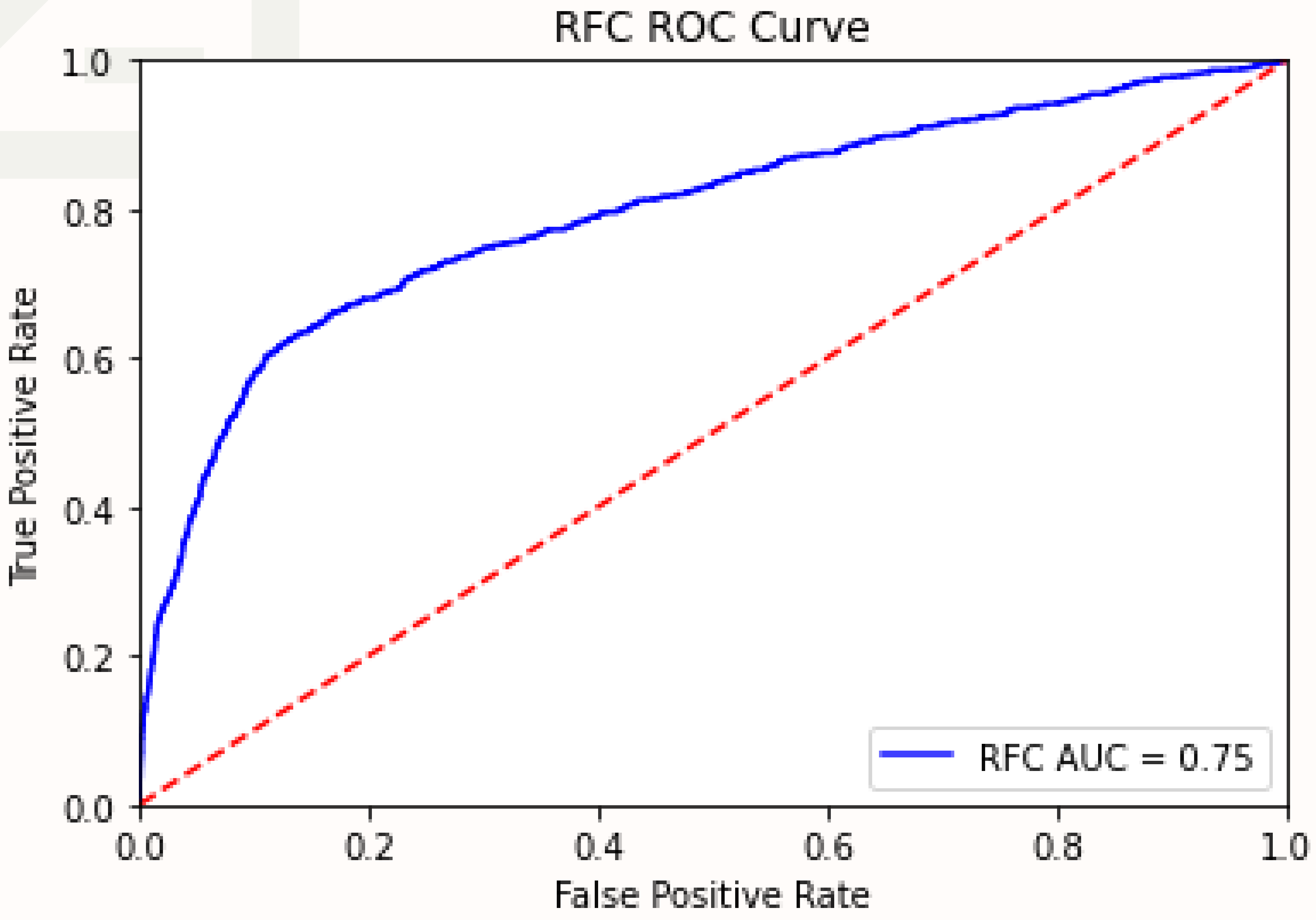
With Model Usage:

- Additional profit of $3,720 from 372 FN clients
- Additional revenue of ~$3,300 = ~2x initial revenue
- Avoid additional loss of $1,491 from FP and TN clients

	precision	recall	f1-score	support
0	0.94	0.88	0.91	7265
1	0.40	0.62	0.48	971
accuracy			0.85	8236
macro avg	0.67	0.75	0.70	8236
weighted avg	0.88	0.85	0.86	8236



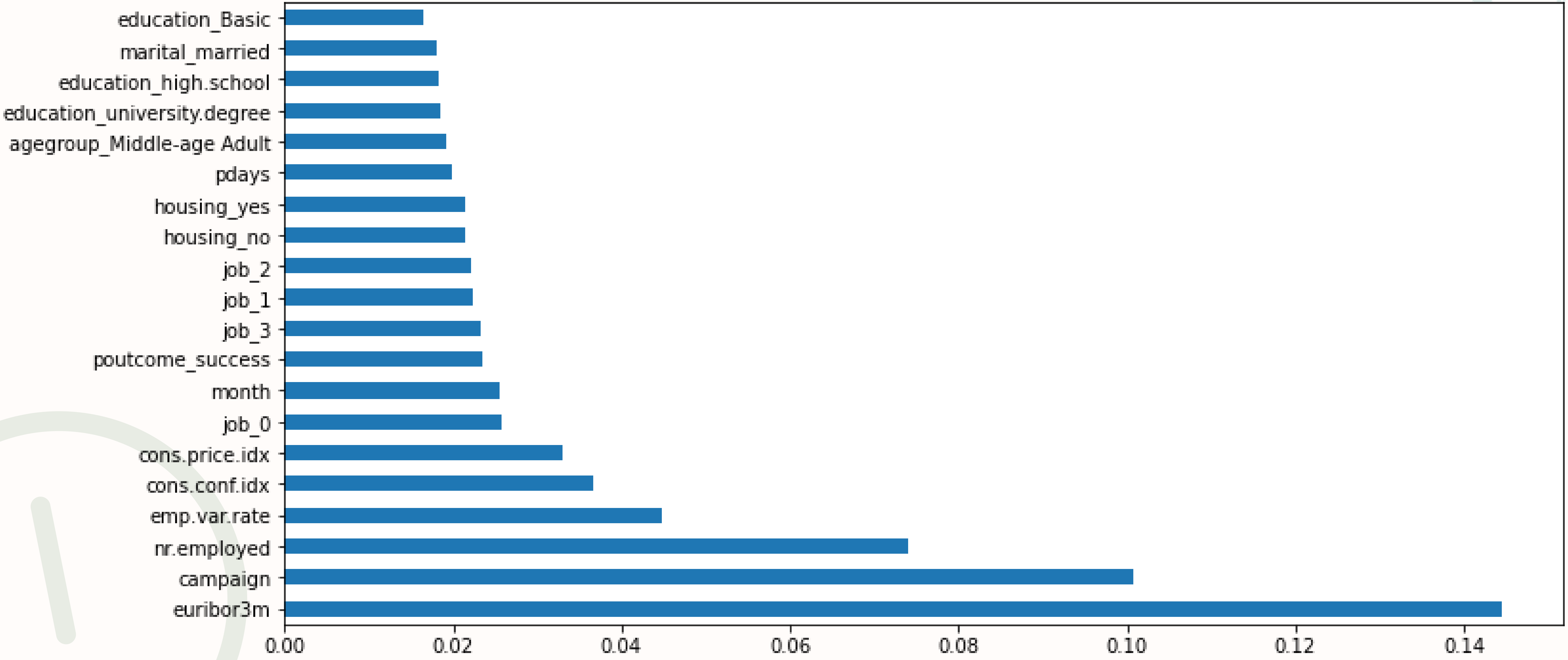
ROC - AUC Score



AUC = 0.75

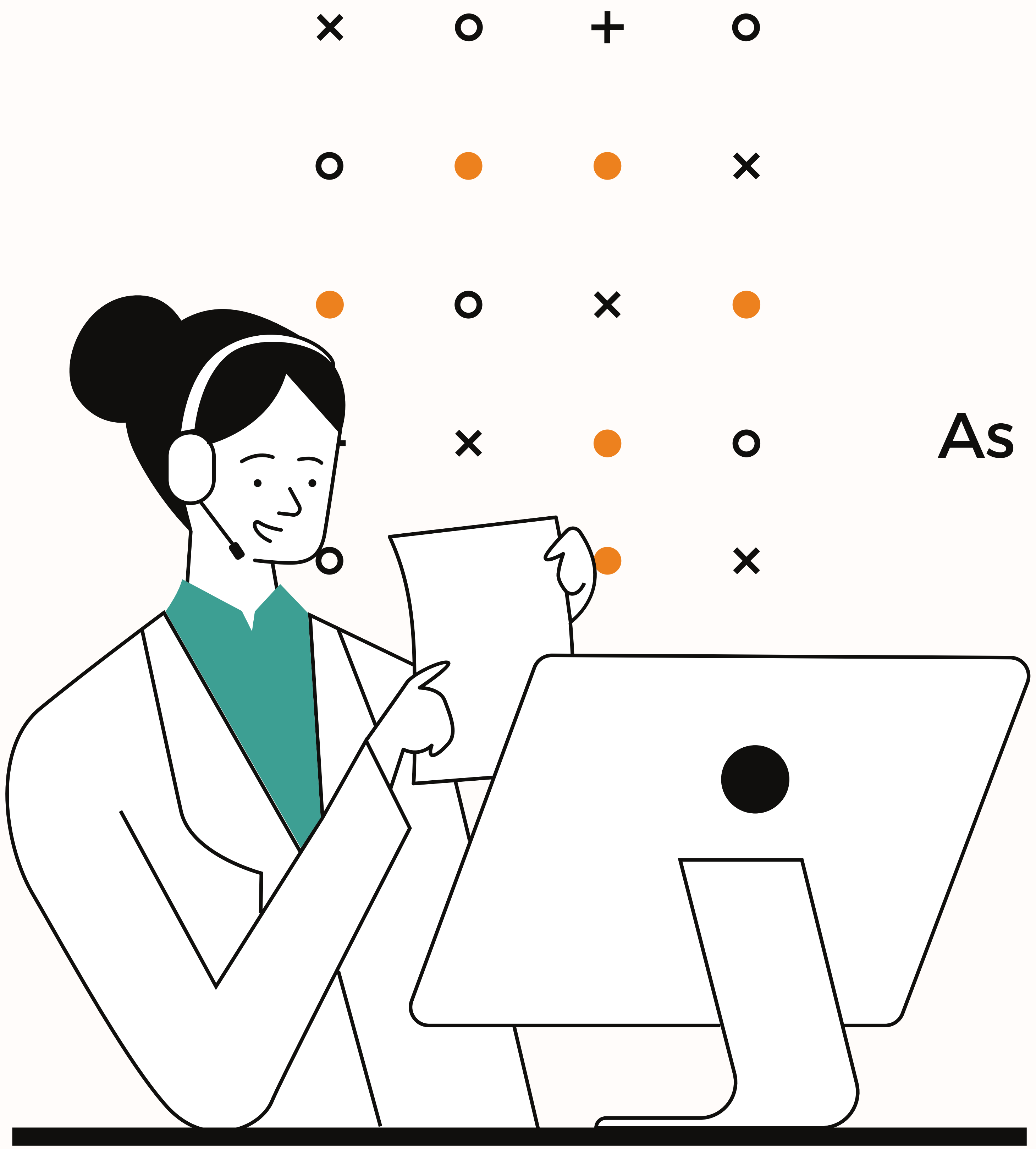
There is a 75% chance that the model will be able to distinguish between positive class and negative class

Feature Importance



Conclusion & Recommendation

To conclude...



Random Forest is the best go-to prediction model

Model can be utilised to help the bank's call center to prioritize the call for potential clients and maximize the conversion rate - all while managing minimum cost on marketing.

As mentioned in evaluation...

Using the model can:

- Acquire previously falsely identified potential clients
- Bring additional revenue of approximately ~$3,300 while maintaining conversion rate quality
- Avoid spending an additional marketing cost of $1,491 for falsely predicted clients
- Generate about 2x more revenue to the bank

Improvements

- Addressing the low precision and f1-scores
- Implementing some more advanced techniques such as dimensionality reduction, ensemble learning and deep learning may contribute to the model's performance
- Leveling the imbalanced data (e.g. there were less data on clients who agreed to make a term deposit) to improve its overall performance further

Recommendations

- Making 'Job' specification a mandatory column for clients to fill in
- Tune the campaign according to the national econometrics
- Always contact clients by cellphone when possible
- Creating a loyalty program for existing clients by giving them some bonuses and unique offers. Data shows that loyal clients, or clients with frequent past contacts are more likely to buy products/have higher conversion rate in making a term deposit

thank you