Exploratory Data Analysis

Bank Marketing Campaign

16.12.2023

Data Glacier

Your Deep Learning Partner

Group Name	Tentacool
Name	Murat Kıran
Email	murattkiran@gmail.com
Country	Türkiye
Specialization	Data Science

Github Repository: https://github.com/murattkiran/Bank-Marketing-Campaign-DG-Final-Project/tree/main/Week_11

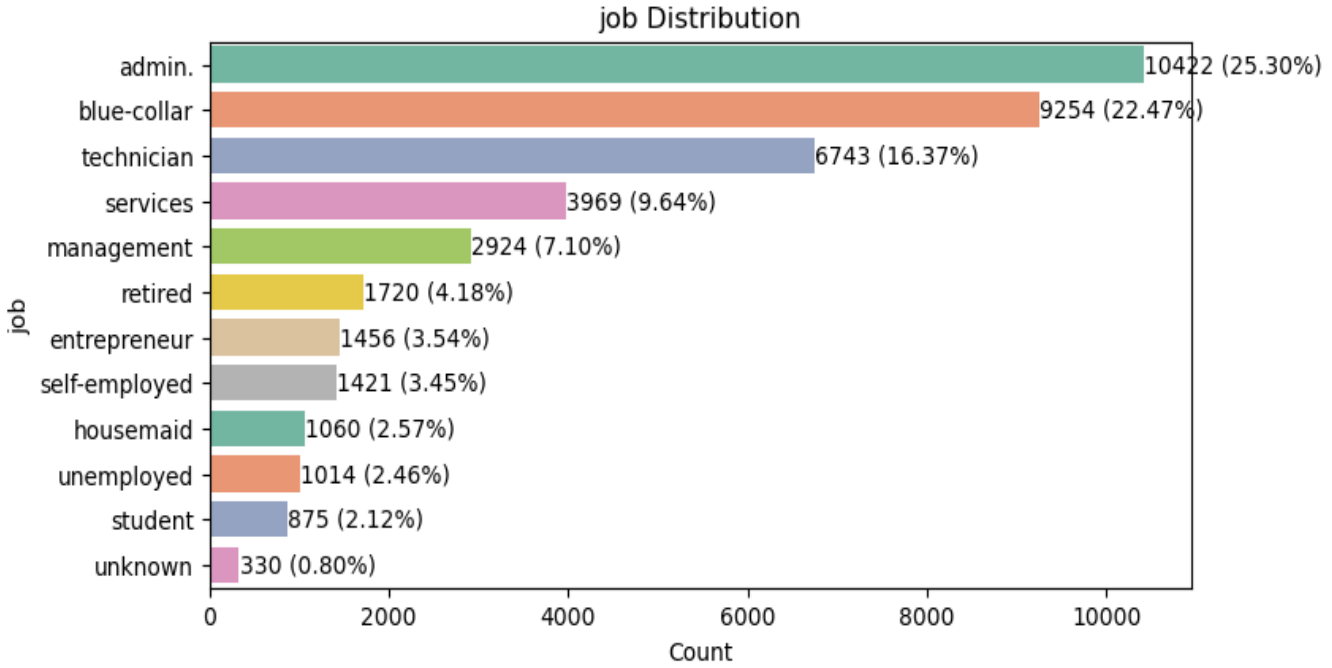
Problem Statement

- ABC Bank is on the verge of launching a new term deposit product and aims to boost its success by developing a predictive model.
- The goal is to identify whether a customer will subscribe to the term deposit ('yes') or not ('no') based on past interactions.
- The challenge lies in optimizing marketing efforts and tailoring strategies to maximize customer engagement.
- The dataset, derived from Portuguese banking campaigns, contains various client details and campaign outcomes.
- The objective is to create a robust predictive model that provides insights into factors influencing subscription decisions, empowering ABC Bank to refine its marketing approach for the impending product launch.

Dataset

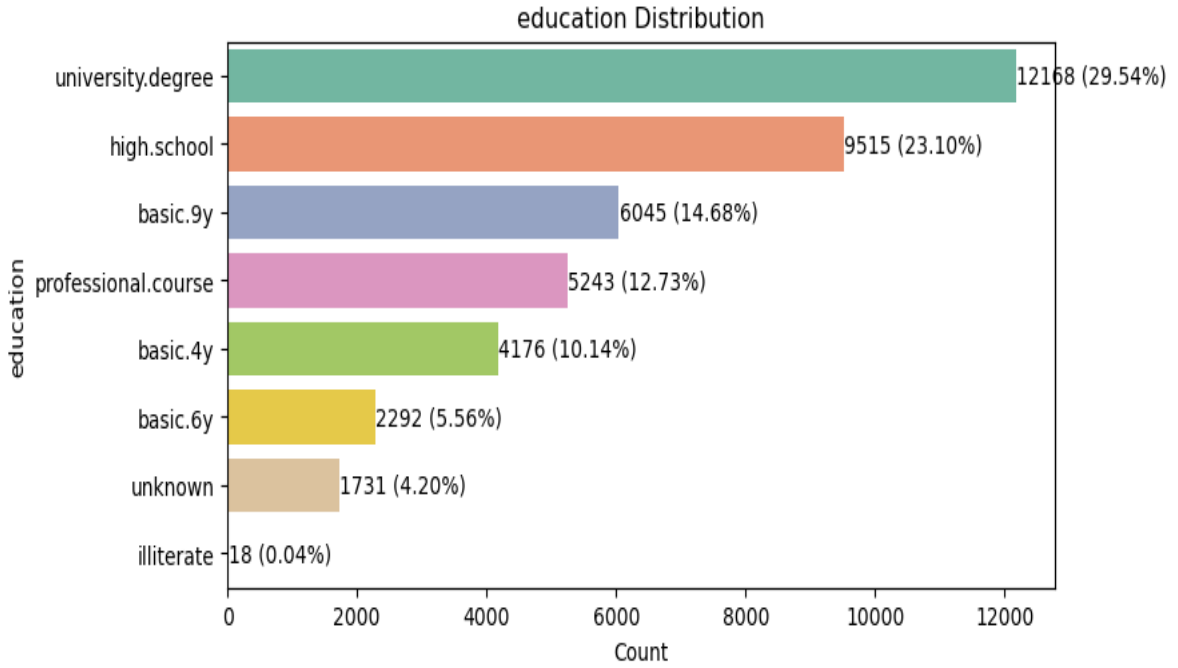
- Total number of observations 41188.
- There is no missing value in this dataset.
- **Categorical Columns**: 'job', 'marital', 'education', 'default', 'housing', 'loan', 'contact', 'month', 'day_of_week', 'poutcome‘
- **Numerical Columns:** 'age', 'campaign', 'pdays', 'previous', 'emp.var.rate', 'cons.price.idx', 'cons.conf.idx', 'euribor3m', 'nr.employed'
- The **duration** feature was removed from the dataset to avoid data leakage. This attribute highly affects the output target, and its value is known only after the call is performed, leading to unrealistic predictive models. The removal aligns with the intention to develop a realistic predictive model.
- Values labeled as "**unknown**" were not deleted. Sensible imputations were carried out, associating **education**, **housing**, and **loan** variables with the job. This approach aligns with the real-world scenario where job is correlated with education, housing, and loan status.

Job



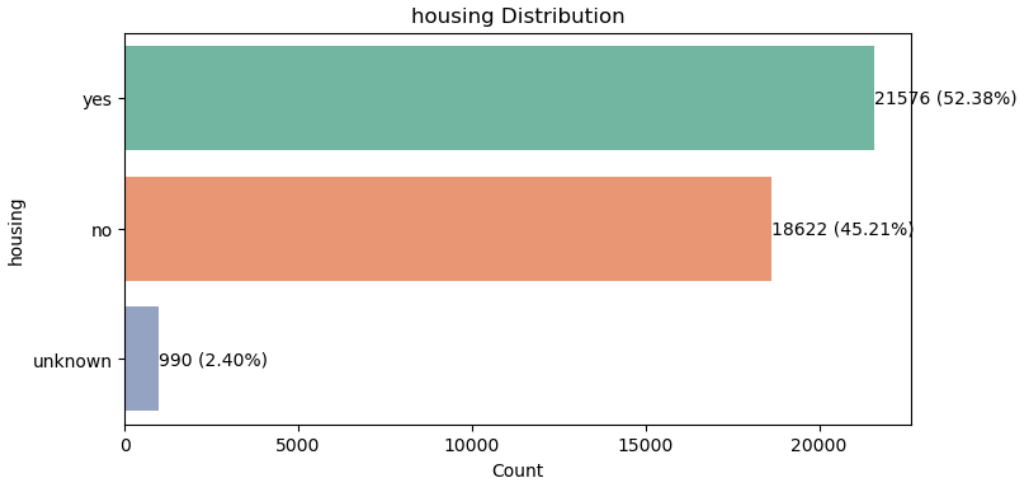
job	y_mean
admin.	0.129726
blue-collar	0.068943
entrepreneur	0.085165
housemaid	0.100000
management	0.112175
retired	0.252326
self-employed	0.104856
services	0.081381
student	0.314286
technician	0.108260
unemployed	0.142012
unknown	0.112121

Education

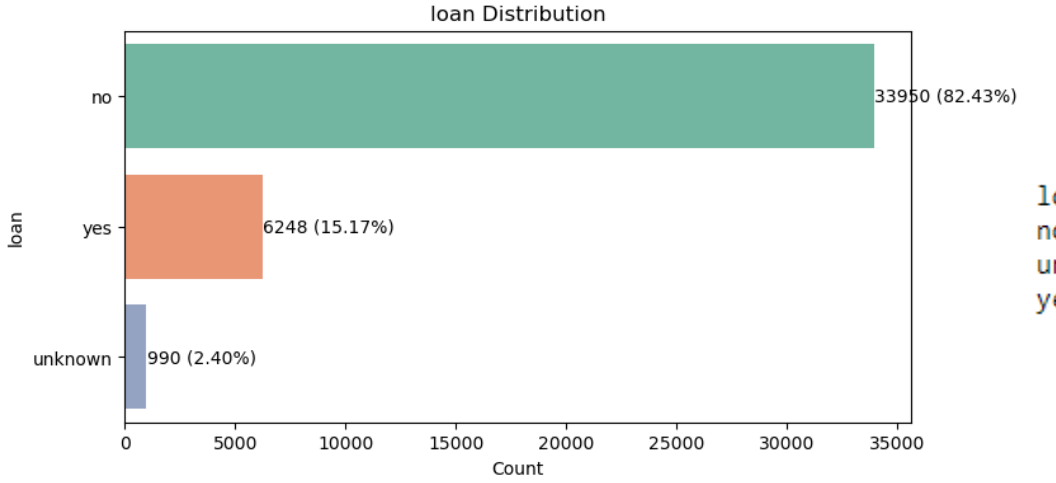


education	y_mean
basic.4y	0.102490
basic.6y	0.082024
basic.9y	0.078246
high.school	0.108355
illiterate	0.222222
professional.course	0.113485
university.degree	0.137245
unknown	0.145003

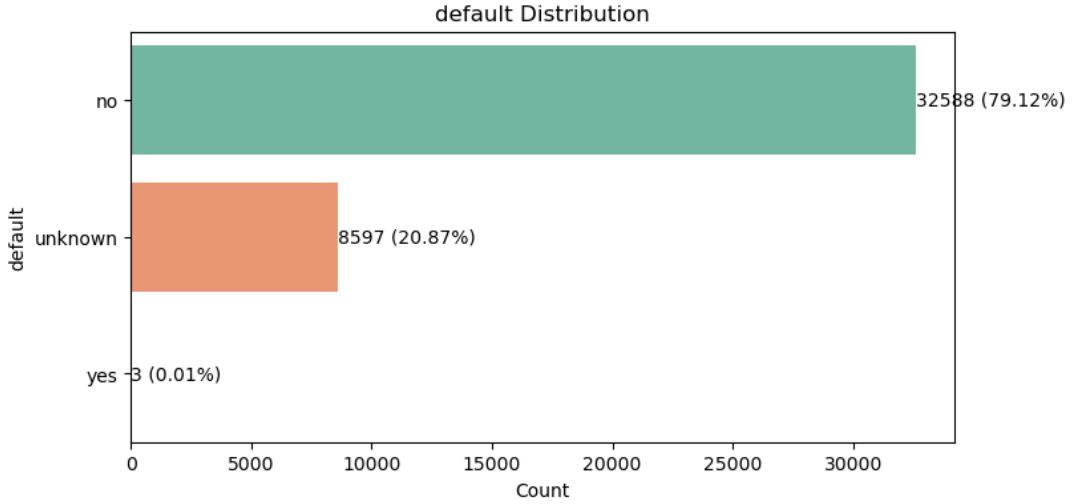
Default, housing and loan



housing	y_mean
no	0.108796
unknown	0.108081
yes	0.116194

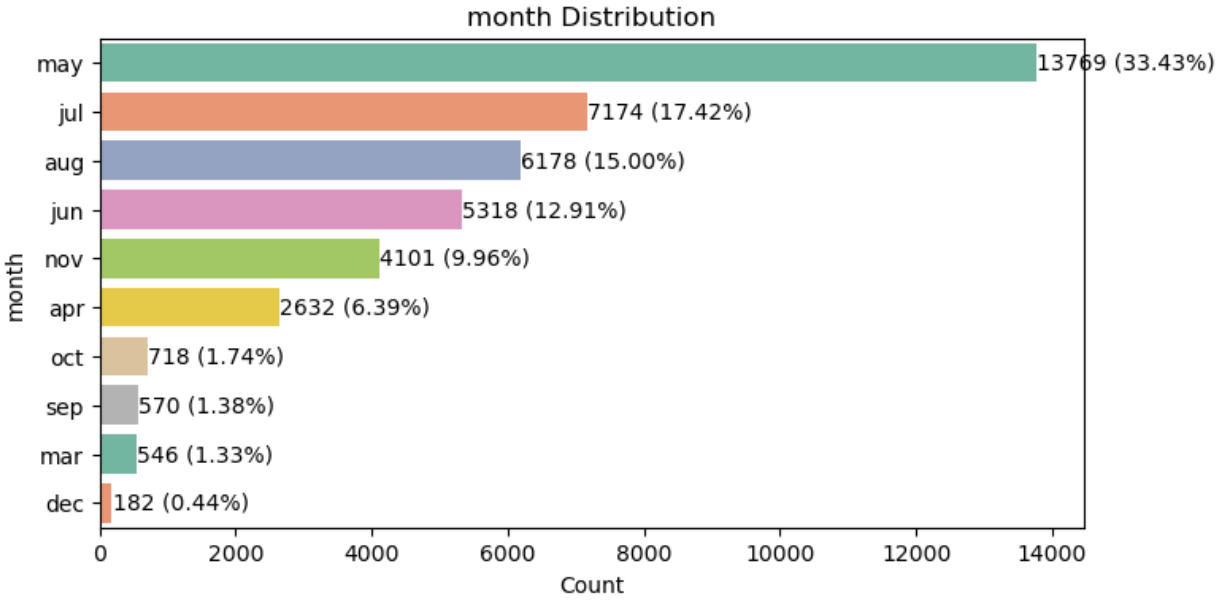


loan	y_mean
no	0.113402
unknown	0.108081
yes	0.109315



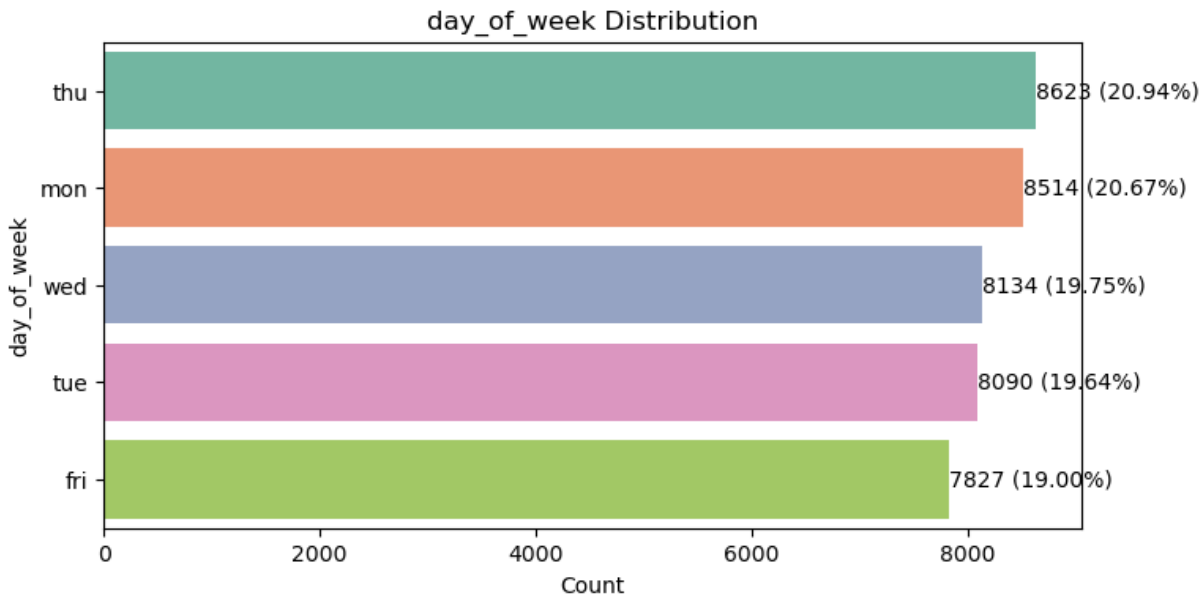
default	y_mean
no	0.12879
unknown	0.05153
yes	0.00000

Months



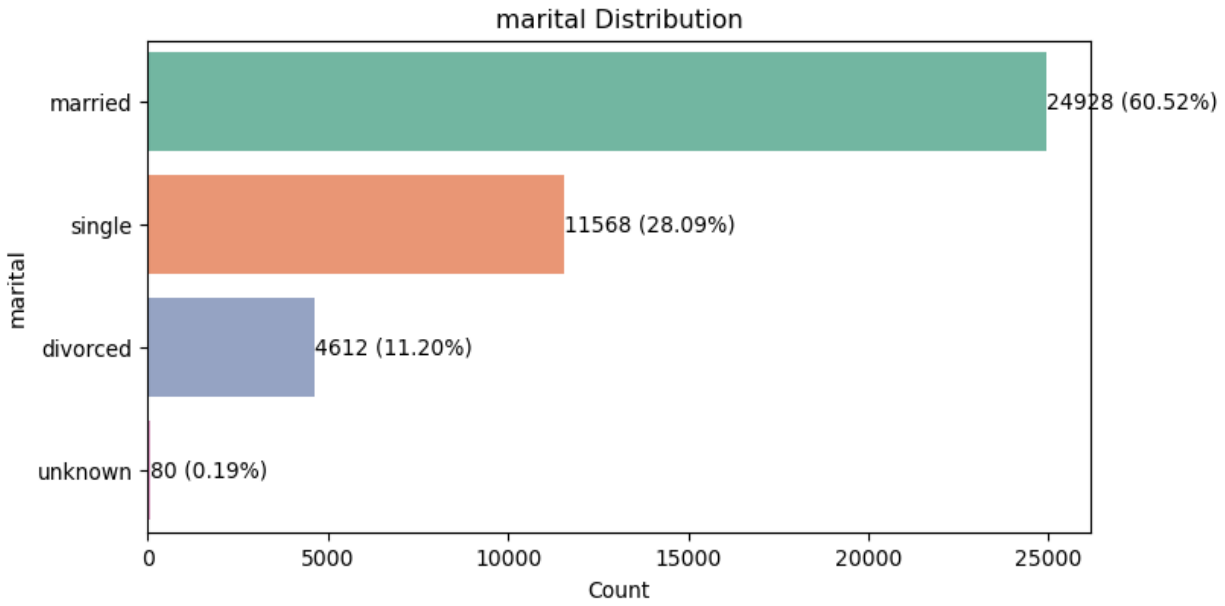
month	y_mean
apr	0.204787
aug	0.106021
dec	0.489011
jul	0.090466
jun	0.105115
mar	0.505495
may	0.064347
nov	0.101439
oct	0.438719
sep	0.449123

Days



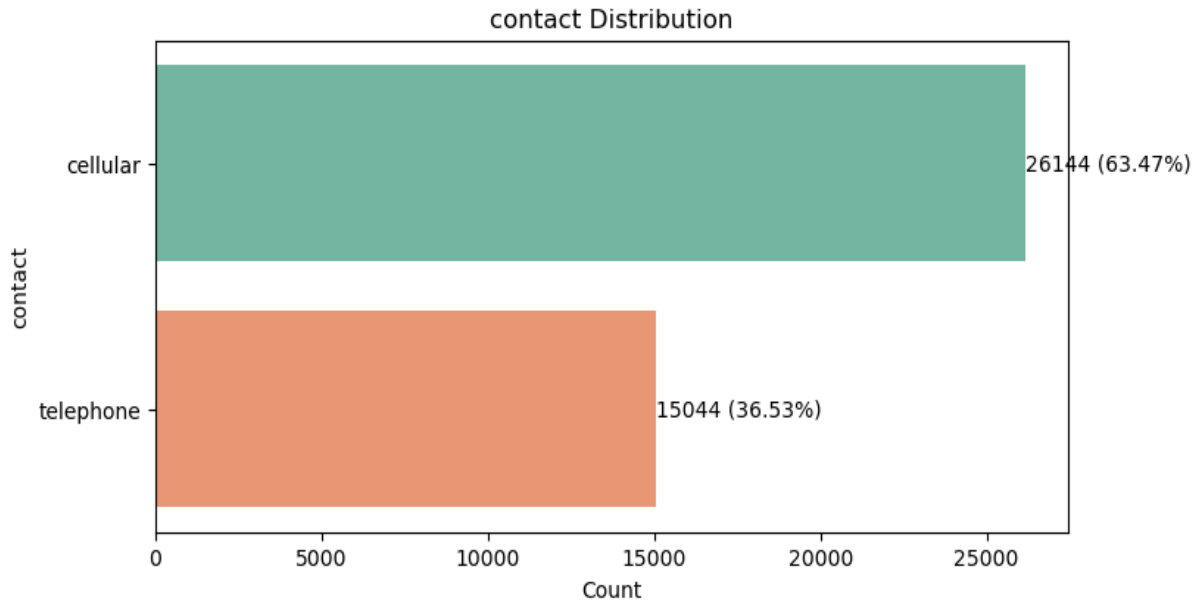
day_of_week	y_mean
fri	0.108087
mon	0.099483
thu	0.121188
tue	0.117800
wed	0.116671

Marital Status



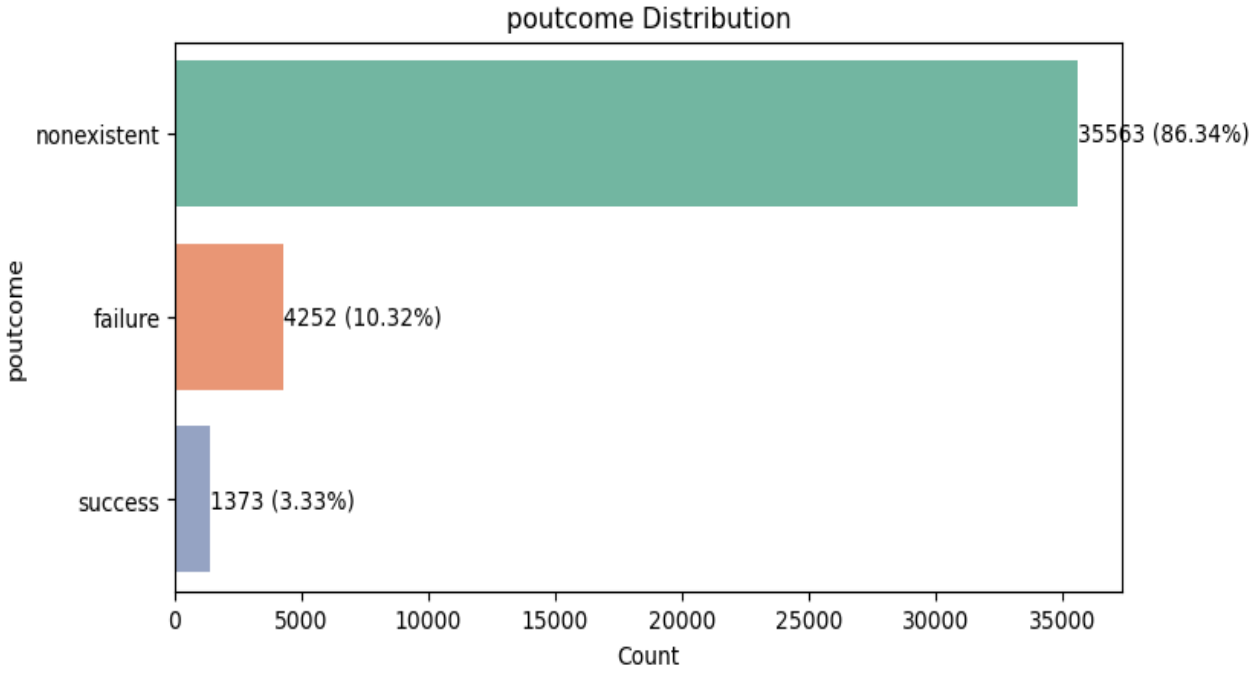
marital	y_mean
divorced	0.103209
married	0.101573
single	0.140041
unknown	0.150000

Contact



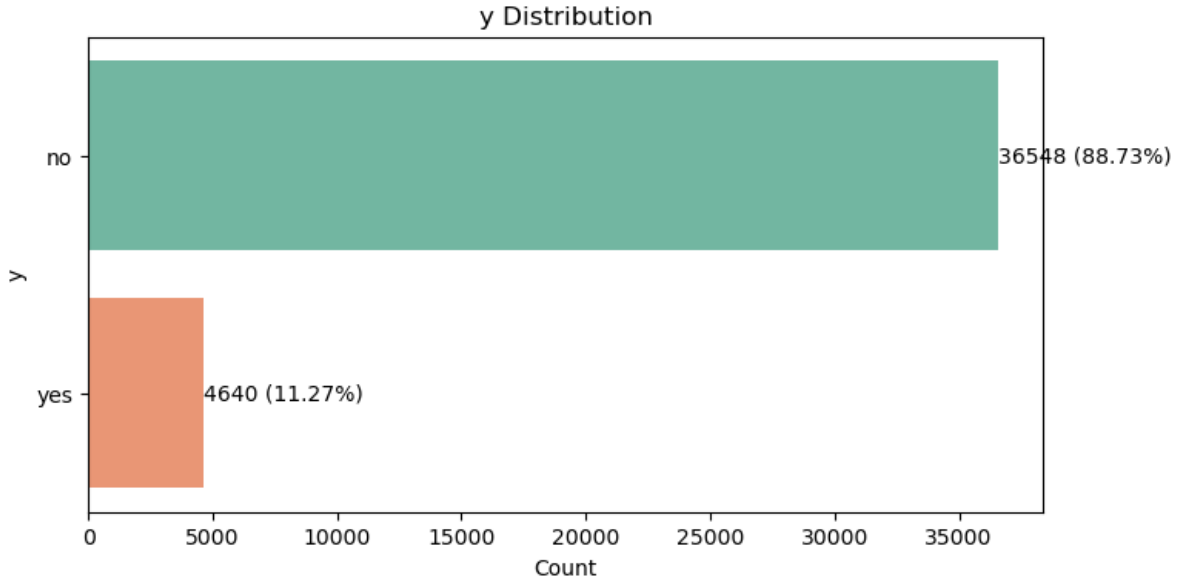
contact	y_mean
cellular	0.147376
telephone	0.052313

Poutcome



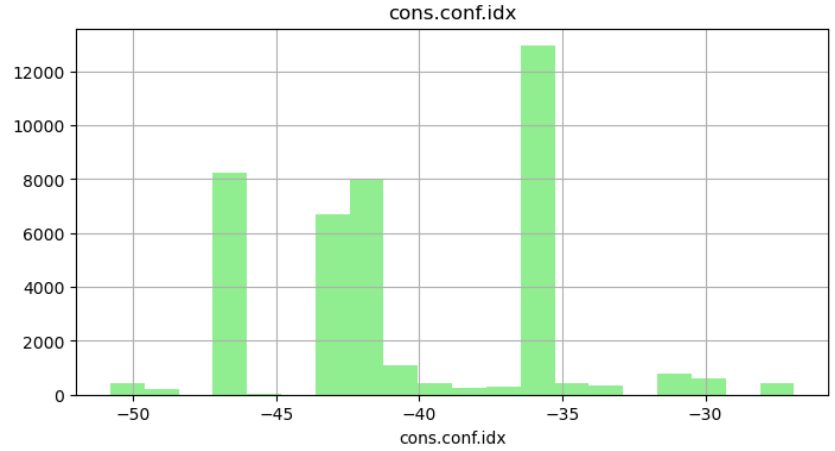
poutcome	y_mean
failure	0.142286
nonexistent	0.088322
success	0.651129

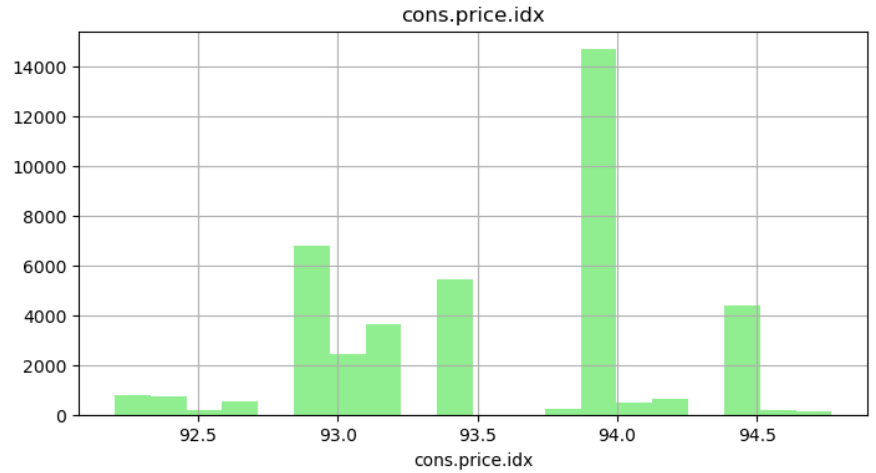
Target variable

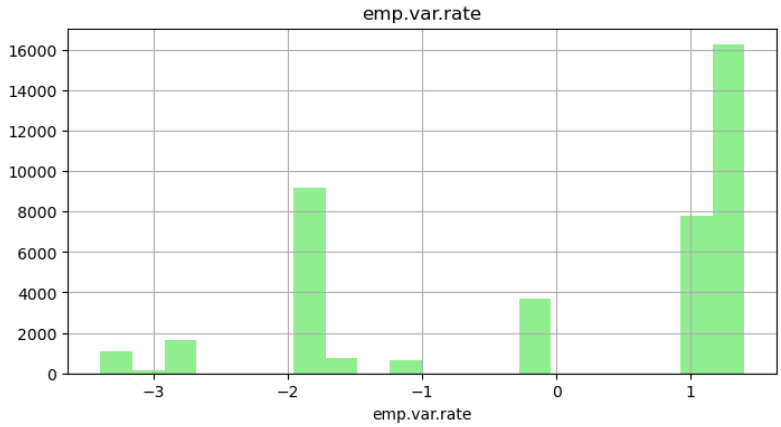


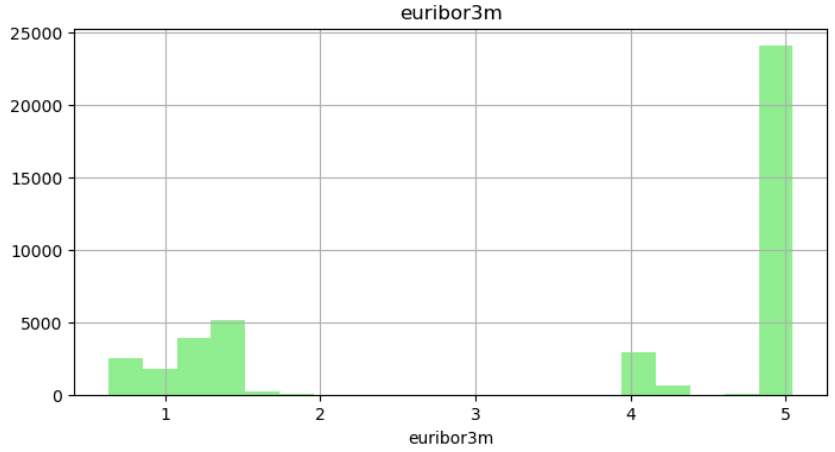
- When dealing with imbalanced datasets, there are various techniques to make accurate observations and achieve balance. One of these is the selection of the right metric.
- The ROC AUC score metric will be employed for model evaluations.

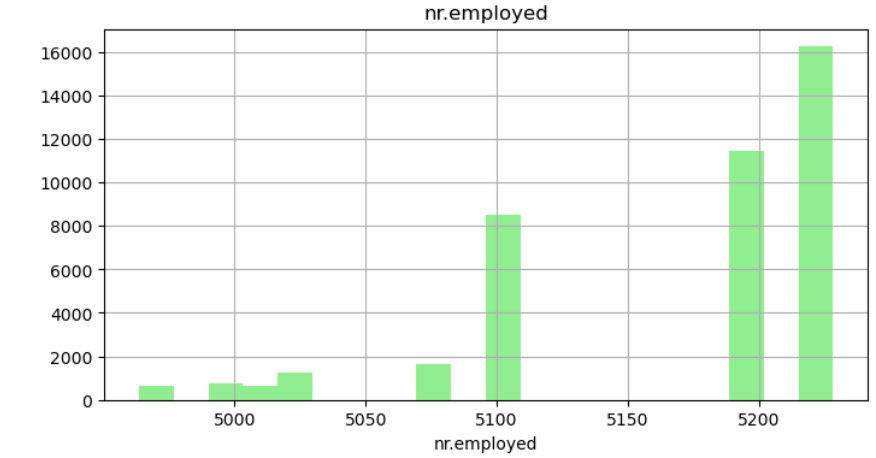
Social and Economic Context Attributes












```
##########################################
     emp.var.rate
y
no        0.248875
yes      -1.233448

##########################################
     cons.price.idx
y
no         93.603757
yes        93.354386

##########################################
     cons.conf.idx
y
no       -40.593097
yes      -39.789784

##########################################
     euribor3m
y
no     3.811491
yes    2.123135

##########################################
     nr.employed
y
no    5176.166600
yes   5095.115991
```

Recommended Models

- In this study, logistic regression, random forest, and XGBoost models will be employed to perform predictions.
- In these models, necessary hyperparameter optimizations will be conducted, and improvements in the ROC-AUC score will be examined to identify the most successful model.

Thank You