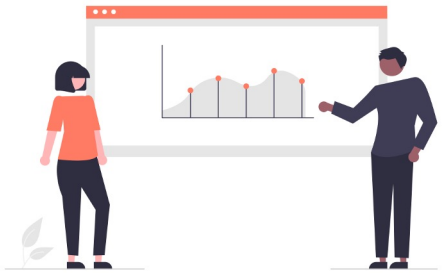
EDA Presentation and Modeling Technique Proposal

Data Science - Bank Marketing Campaign

Data Glacier - Team Datalux

Asmaa Alqurashi
Deepak Rawat
Huu Thien Nguyen

PROJECT SECTION



1. Team Introduction
2. Problem Description
3. GitHub Repository
4. EDA Presentation for Business Users
5. Machine Learning Model Recommendation

1. Team Introduction

Group Name: Datalux
Group Members: 3

Name	Email	Country	Uni/Company
Huu Thien Nguyen	nguyenhuuthien27296@gmail.com	Sweden	Skövde University
Asmaa Alqurashi	asmaa.idk@gmail.com	Saudi Arabia	Taif University
Deepak Rawat	deepakrawat68@gmail.com	Ireland	Dublin Business School

Specialisation: Data Science
Submitted to: Data Glacier canvas platform
Internship Batch: LISUM09

2. Problem Description

The ABC Bank wants to market its term deposit product to clients in this project.

A machine learning model that will assist them in determining whether a particular consumer would buy their product.

Goal: Save the time and resources and finally leads to optimised cost for this campaign.

3. GitHub repository

The link for GitHub: https://github.com/AndrewNguyen27296/DataGlacier

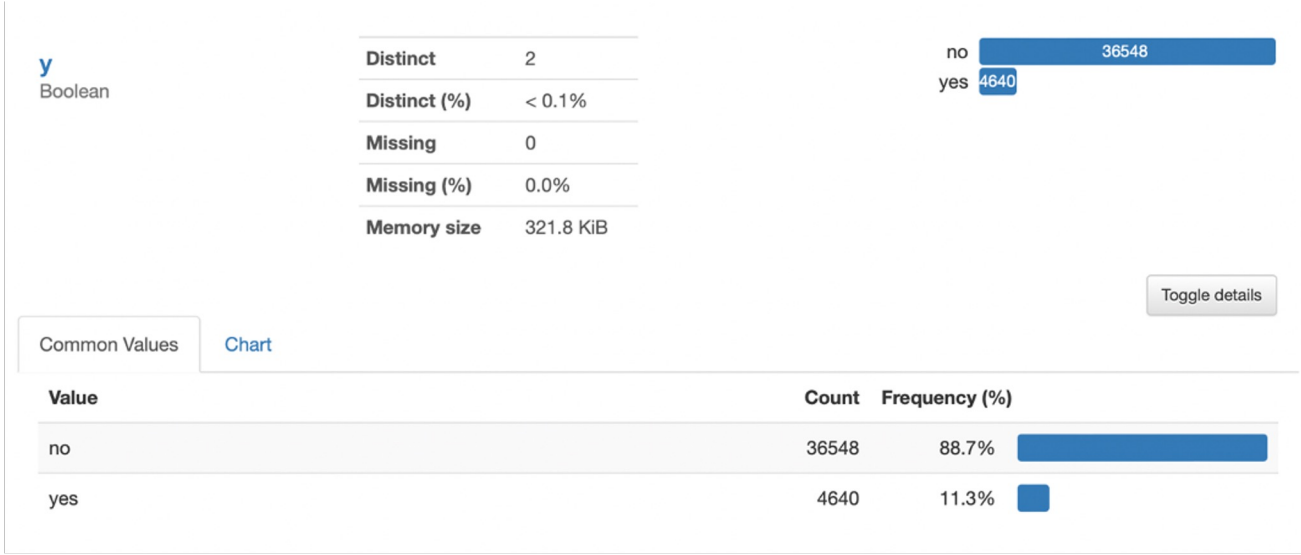
4. EDA Presentation for Business Users

Insight #1: Sampling techniques and ensemble methods are required when building classification due to an imbalanced dataset.

Imbalance between the dependent variables is 88.7% for refusal (no) and 11.3% for customers who subscribe to a term deposit

Under-, over-, and random selection technique can be applied to solve this problem

The bagging, boosting, and stacking method can improve the ML's predictive score



The imbalance of dependent variables in the banking dataset

4. EDA Presentation for Business Users

Insight #2: Feature engineering with highly correlated numeric variables.

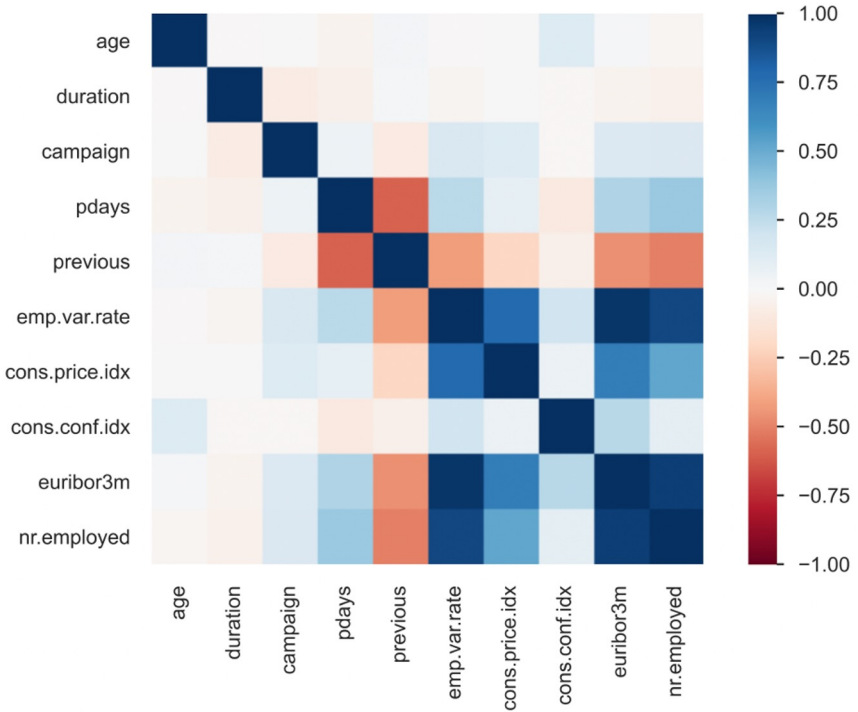
pdays feature should be excluded when modelling or re-examining

The distribution is imbalanced; 96,3% is a “999” value, which does not provide any valuable information

It correlated to the “previous” feature and will cause noises

Proposed methods: bin the pdays into two groups, “contacted” and “not contacted”.

“cons.conf.idx”, “euribor3m”, and “nr.employed” can be combined



Correlation matrix of numeric variables in the banking dataset

4. EDA Presentation for Business Users

Insight #3: Deploying marketing campaign on primary client segment (subscribed term deposit customers), which are married/single, non-existent poutcome, and do not have loans.

The segment of customers based on work profession, marriage status, poutcome, and loan history

The filter of subscribers (Y=yes) and count of subscribed cases> 110 was applied

If a client is single, the bank-firm needs to target admin and technicians to maximise the profit. Vice versa, admin, blue-collar, managers, retired, and technicians are the leading group of customers to focus on

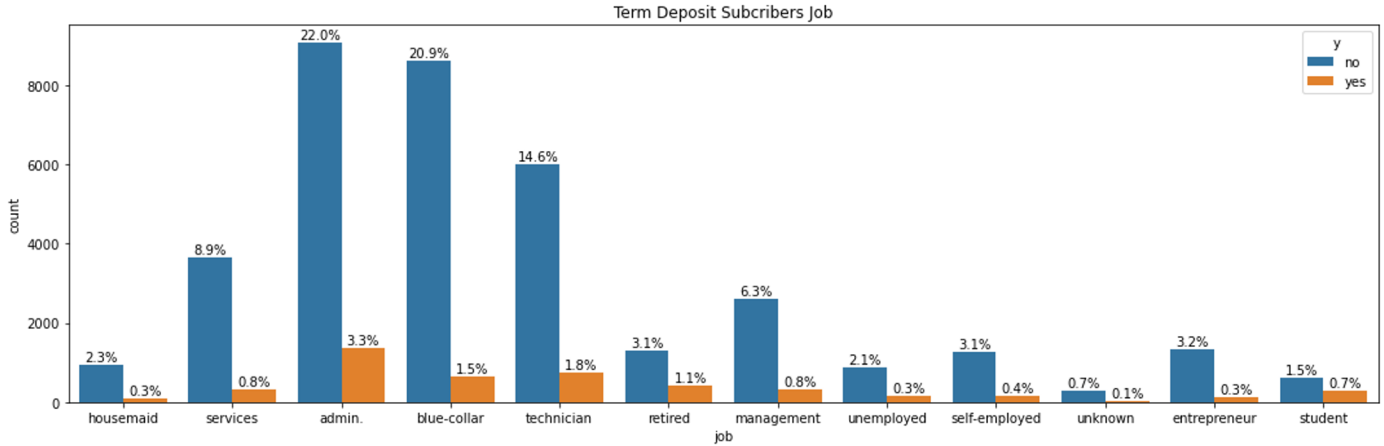
Subscribe a term deposits (Primary customer segment)

			Job				
Loan	Poutcome	Marital	admin.	blue-coll..	manage..	retired	technician
no	nonexistent	married	348	289	123	160	204
		single	324				153

Primary customer segment, which needed to focus on the marketing/customer service

4. EDA Presentation for Business Users

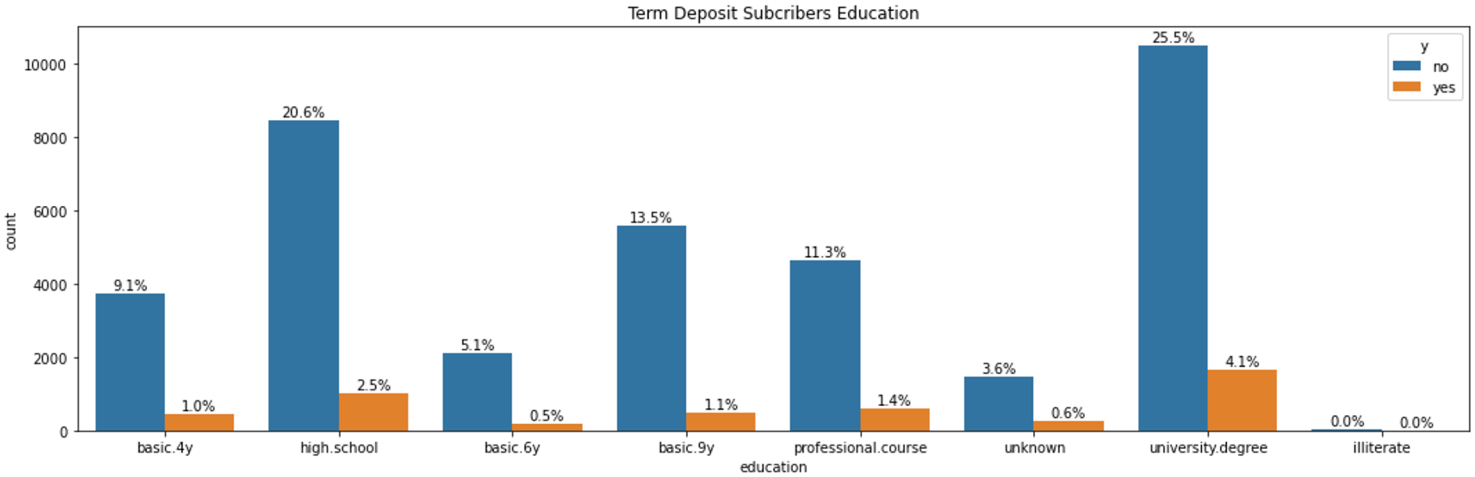
Insight #4: Profile the clients to target the right group for the campaign.



Jobs: Targeting administer (3.3%) and technician (1.8%) clients. Clints with these jobs are most subscribers to the term deposit.

4. EDA Presentation for Business Users

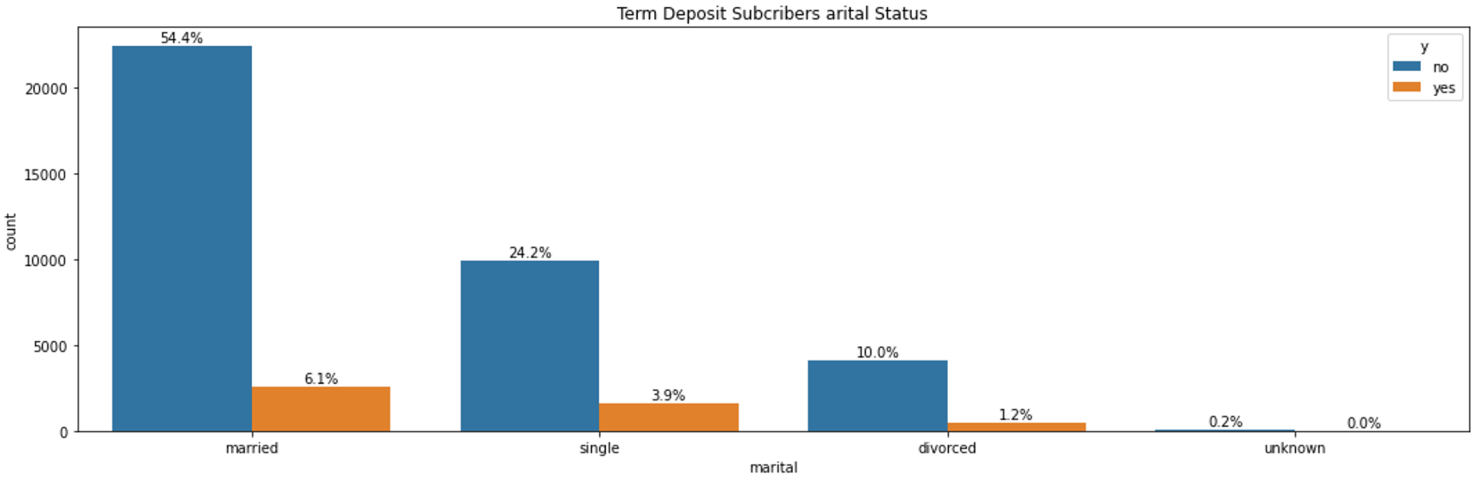
Insight #4: Profile the clients to target the right group for the campaign.



Education: More subscriber to the term deposit with a university degree (4.1%). We recommend targeting these clints.

4. EDA Presentation for Business Users

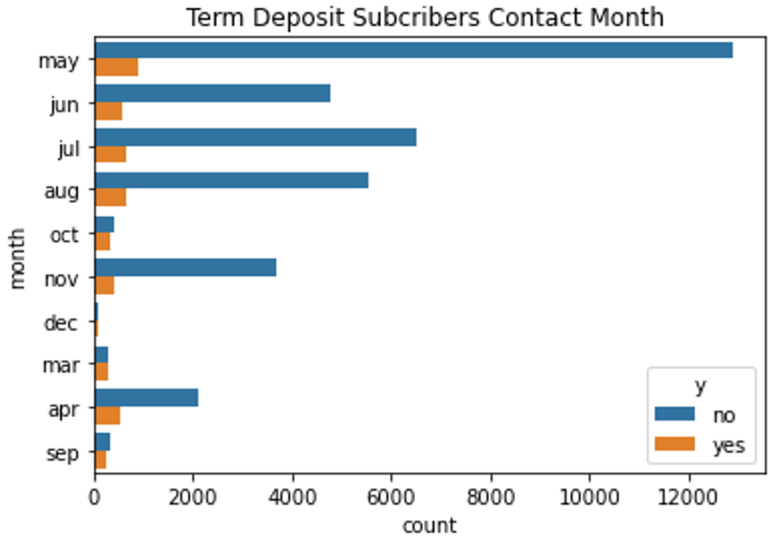
Insight #4: Profile the clients to target the right group for the campaign.



Marital Status: Most clients are married so the number of the married subscribers (6.1%) is higher but relatively singles (24.2%) are less but subscribed (3.9%) more to the term deposit so it's a good idea to target both.

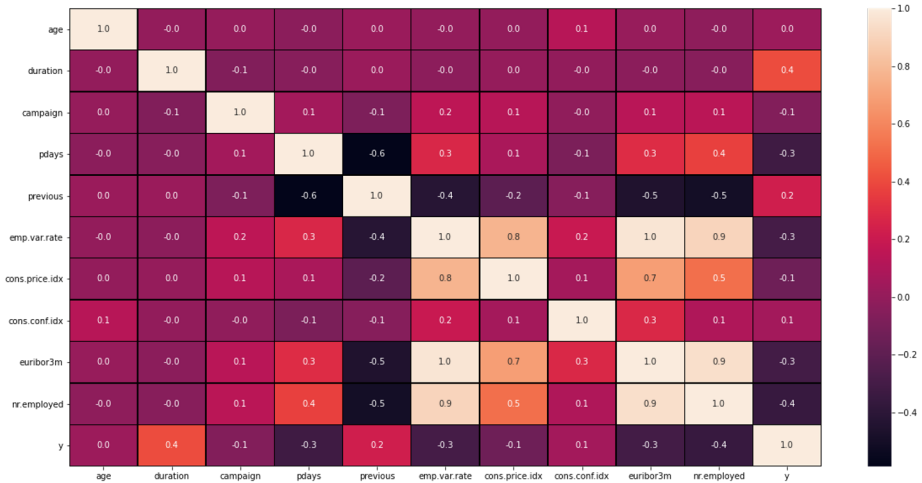
4. EDA Presentation for Business Users

Insight #5: The month of the contact impacts the response of the clients (most clients subscribed in May), and more calls were made in May. The calls to target the clints shouldn't be focused only on May but also on October, September and March.



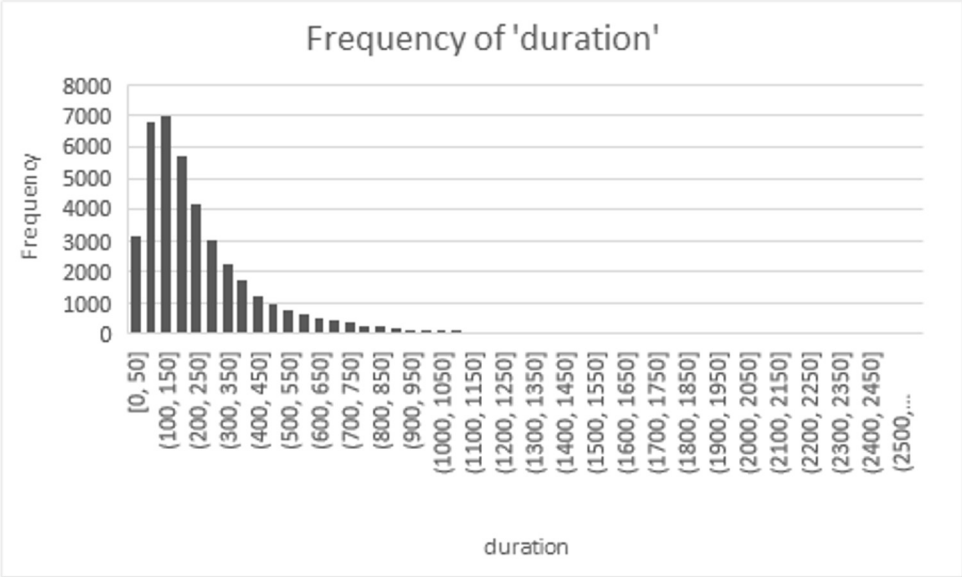
4. EDA Presentation for Business Users

Insight #6: Increasing the call duration impacts the response to the campaign.



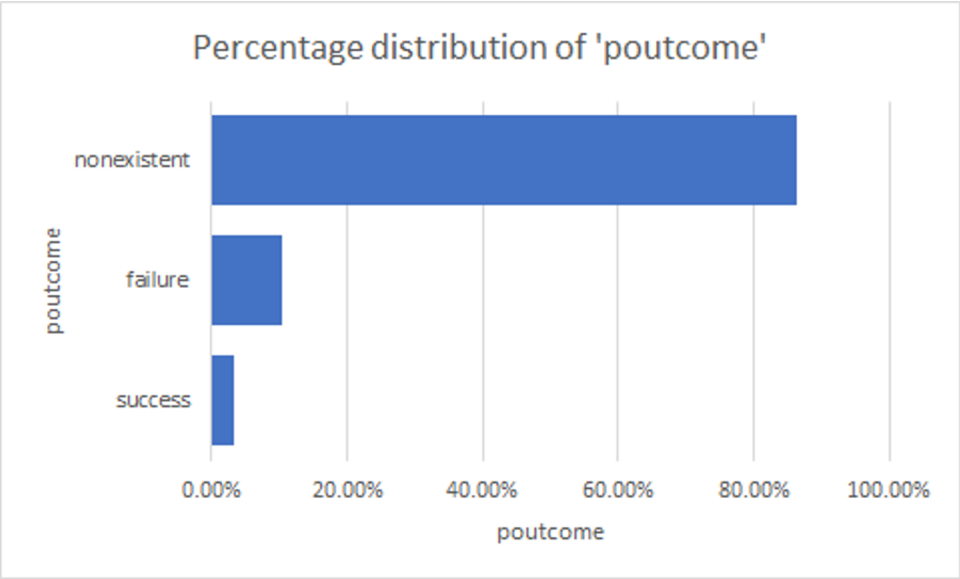
Duration of the call has high correlation with the target which means the longer the call with clients the higher the chance of the client subscribing to the term deposit.

4. EDA Presentation for Business Users



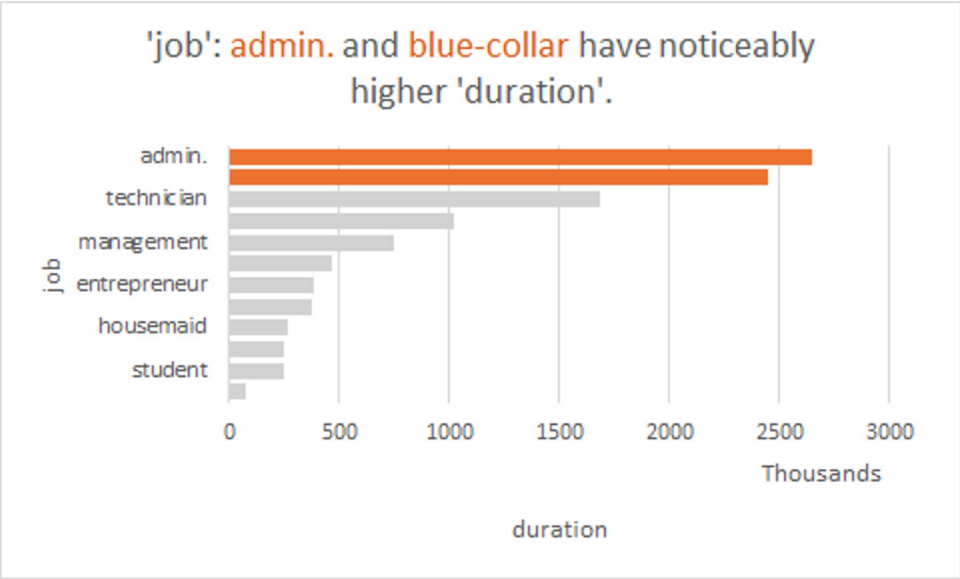
Insight #7 : The duration of most calls is under 600 seconds (10 minutes). The average call duration is 4 minutes, it is recommended to target the customers to have a call duration of approx 5 minutes.

4. EDA Presentation for Business Users



Insight #8:The outcome of the previous marketing campaign is nonexistent for more than 85%, so we can target these customers for this campaign along with successful customers.

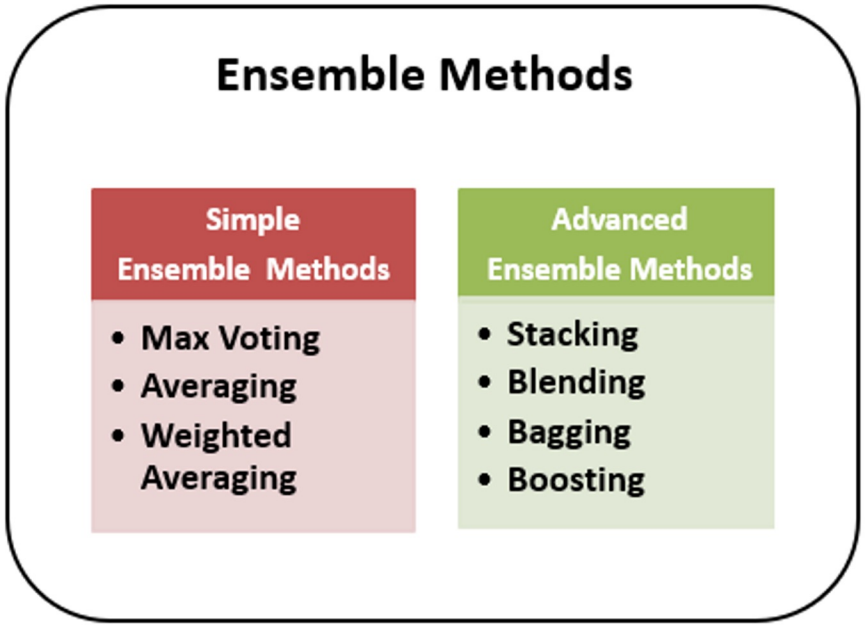
4. EDA Presentation for Business Users



Insight #9: The total duration of call for admin and blue collar jobs is high as compared to other professions, so they may take more time of callers in the campaign

5. Machine Learning Model Recommendation

Based on Insight #1: Sampling techniques and ensemble methods are required when building classification due to an imbalanced dataset.



5. Machine Learning Model Recommendation

After Modeling there's many evaluation methods to choose the best classification model for our problem such as:

- Classification accuracy: shows how many of the predictions are correct.
- Confusion matrix: it provides insight into the predictions and show the correct and incorrect (i.e. true or false) predictions.
- Precision and recall: Precision measures how good our model is when the prediction is positive. Recall measures how good our model is at correctly predicting positive classes.
- F1 score: the weighted average of precision and recall