

Project Proposal

Review Phase #1

DAB103 – Section 4

November 5, 2021



ST. CLAIR
COLLEGE

ZEKELMAN
 **SCHOOL OF**
BUSINESS
& INFORMATION TECHNOLOGY



Project Proposal

Phase#1



Course

Analytic Tools and Decision Making

Course Code

DAB 103

Section No.

004

Academic Term

Fall 2021

Professor

Mr. Sathish Chandra Pichika

Team Introduction

Group Name

INSIGHT



S. ID: 0780765

Farzin Valiloo



S. ID: 0780763

Maryam Aliakbari



S. ID: 0783366

**Mohmad Faizal
Mohmad Faruk Kazi**



S. ID: 0795068

**Muhammad Thanazal
Kunduvalappil
Mottammal**

Dataset Description

- About Dataset
- Dataset Characteristic
- Other Works on this Dataset

About Dataset

Title

Bank Marketing Data

Source

Data.world

Dataset

<https://data.world/data-society/bank-marketing-data>

<http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

Data Link

<https://query.data.world/s/lqrsaugj7kwkyazkvvdowjuxjrybxx>

Description

The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution.

Main Article

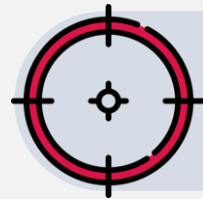
Moro et al., 2014, S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

About Dataset

There are four datasets extracted from Main Dataset used in the article:

- **bank-additional-full.csv** with all examples (41188) and 20 inputs, ordered by date (from May 2008 to November 2010), very close to the data analyzed in [Moro et al., 2014]
- **bank-additional.csv** with 10% of the examples (4119), randomly selected from 1), and 20 inputs.
- **bank-full.csv** with all examples and 17 inputs, ordered by date (older version of this dataset with less inputs).
- **bank.csv** with 10% of the examples and 17 inputs, randomly selected from 3 (older version of this dataset with less inputs).

The smallest datasets are provided to test more computationally demanding machine learning algorithms (e.g., SVM).



The classification goal was to predict if the client will subscribe (yes/no) a term deposit (variable y).

Dataset Characteristic

- The data is related with direct marketing campaigns of a **Portuguese banking institution**.
- The marketing campaigns were based on **phone calls**.
- Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be **('yes') or not ('no')** subscribed.
- The data is example of data very close to the data analyzed in the main article ordered by date from **May 2008 to November 2010**.

Other Works on this Dataset

Classification

:

<https://github.com/mcabinyaya/Bank-Marketing-Data-Analysis>

Predict
Subscription

:

<https://medium.com/swlh/using-machine-learning-to-predict-subscription-to-bank-term-deposits-for-clients-with-python-aec8a4690807>

Machine Learning

:

<https://rpubs.com/alfandash/lbb-classification-2>

Background & Motivation

- Industry Knowledge domain
- The Telemarketing Main Problems
- Background & Motivation

Industry Knowledge domain



Telemarketing is **the direct marketing of goods or services to potential customers over the telephone or the Internet.**

Four common kinds of telemarketing include outbound calls, inbound calls, lead generation, and sales calls.

Industry Knowledge domain

Financial telemarketers convince consumers to switch banks or their stock trading account by making those necessary telemarketing calls.



- Bank is one of the organization use telemarketing method for selling banking products or services.
- Telemarketing is a popular method used by bank to selling, because bank products and services sometimes too complicated for some users to understand.

The Telemarketing Main Problems

As much as we are crazy for telemarketing, we cannot ignore the persistent problems facing it today!



Rejections

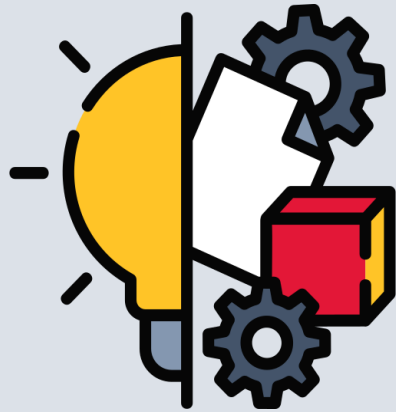
B2B marketers are getting rejected as more people have cynical whenever they receive a telemarketer



Increased costs

The Companies cannot initiate a cold calling campaign without churning out a hefty sum of money

Background & Motivation



Background & Motivation

- A good telemarketing campaign plan clearly outlines the details and the scope of the campaign, so everyone is clear about what it involves and how it'll work.
- Telemarketing campaigns can help you reach a group of targeted prospects or customers to communicate a message, gather feedback, and determine a next step for the relationship.

Problem Statement Project Proposal &

- **Problem Statement**
- **Project Proposal**
- **Project Audience**
- **Short Problem Statement**
- **Analysis Questions**

Problem Statement



Marketing strategy could evolve over time. As the bank learn more about what is and what is not working, it will build a deeper understanding of the marketplace. Building an effective campaign strategy can help bank and its marketing department to create the better positioning for understanding the customers requirements.

Benefits of effective Marketing Campaigns

- Attracts more sales
- Improve the reputation of the bank
- Improve Undertesting of the market & customers
- Better long-term marketing campaigns

Project Proposal

The product of this project will be a descriptive analytics that examines the characteristics of a successful phone call for the purpose of direct marketing in terms of targeted prospect customers and properties of the call session, based on the data set.

The output of this project may help marketing managers in similar industries (e.g. finance, other service industries) to improve the performance of their direct marketing efforts by increasing the efficiency and effectiveness of the phone calls by exploring some of the aspects of successful direct marketing calls recorded in the dataset.



Proposal Statement

Project Audience

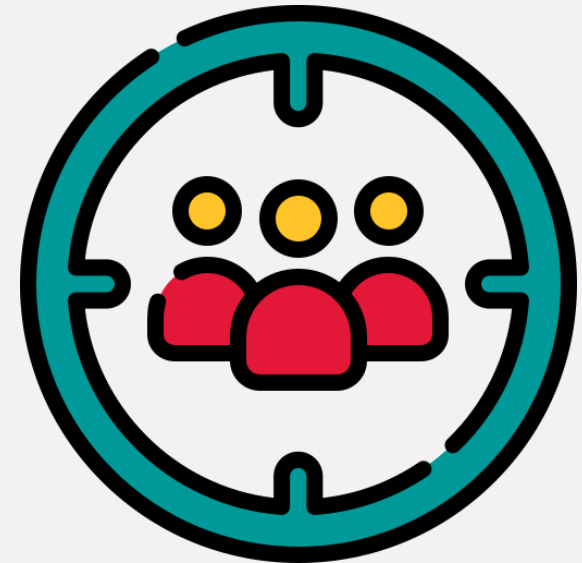
Audiences

Marketing
Department

Medium & High-
level Managers

Sales Department

Call Centers



Short Problem Statement



Problem Statement

How to increase effectiveness of direct marketing campaign based on the dataset

Analysis Questions



Question

Who are the top targeted prospects in terms of successful outcome?

What is the best time to contact prospect customers?

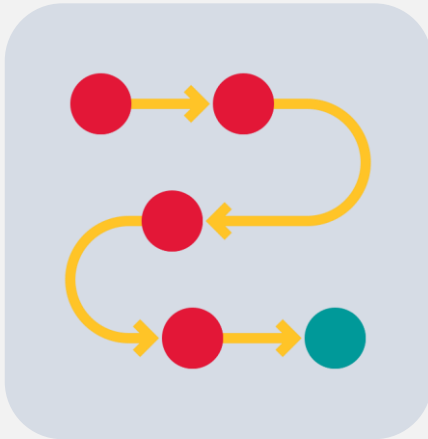
What are the top characteristics of a successful call (duration and time)?

Preliminary Exploratory Data Analysis (EDA)

- Data Description
- Univariate Preliminary Visualization: Numeric Variables
- Univariate Preliminary Visualization: Categorical Variables
- Future Exploration & Expected Transformation Steps

Preliminary EDA Steps

In this chapter, for the initial EDA:



- The data types of variables will be explored
- Each variable will be introduced
- A plot and short description of each variable will be discussed.
- A general exploration of correlation among numerical variables will be provided

Data Description

To describe and analyze the data, we need to understand the nature of data.

The type of statistical analysis is influenced by the type of data and can be performed on it.

DataFrame Structure

Primary Dataset Structure

No. Variables

21

No. Rows

41188



New-Structured Dataset

No. Variables

16

No. Rows

41188



Social and Economic context Attributes that have been Eliminated based on Problem statement

Name	Description	Data Type
emp.var.rate	employment variation rate - quarterly indicator	Numeric
cons.price.idx	consumer price index - monthly indicator	Numeric
cons.conf.idx	consumer confidence index - monthly indicator	Numeric
euribor3m	euribor 3 month rate - daily indicator	Numeric
nr.employed	number of employees - quarterly indicator	Numeric

Variables: Type & Description

Old Name	New Name	Data Type	Description
age	Age	Numeric	Client Age in Year
job	Job	Categorical	Client's Job type (admin, bluecollar, housemaid, management, retired, self-employed, services, student, technician, unemployed, unknown)
marital	Marital	Categorical	Marital status (divorced (divorced or widowed), married , single, unknown)
education	Education	Categorical	Level of education (basic.4y, basic.6y, basic.9y, high.school, illiterate, professional.course, university.degree, unknown)
default	Credit	Categorical	Has the client credit in default? (no , yes, unknown)
housing	Housing_loan	Categorical	Has the client housing loan? (no , yes, unknown)
loan	Personal_loan	Categorical	Has the client personal loan? (no , yes, unknown)
contact	Call_type	Categorical	Type of contact communication (cellular, telephone)
month	Last_month	Categorical	Last contact month of year (jan, feb, mar, apr, .., nov, dec)
day_of_week	Last_weekday	Categorical	Last contact day of the week (mon, tue, wed, thu, fri)
duration	LastCall_Dur	Numeric	Last contact duration, in seconds. The duration is not known before a call is performed. (if LastCall_Dur=0 then y='no')
campaign	NewCampaign_CallNo	Numeric	Number of contacts performed during this campaign and for this client
pdays	Campaign_Intervals_Day	Numeric	Number of days passed by after the client was last contacted from a previous campaign
previous	PrevCampaign_CallNo	Numeric	Number of contacts performed before this campaign and for this client
poutcome	PrevCampaign_Result	Categorical	Outcome of the previous marketing campaign (failure, nonexistent, success)
y	Campaign_Success	Categorical	Has the client subscribed a term deposit? (binary: yes , no)

Data Type

Categorical

Job

Education

Marital

Credit

Personal_loan

Housing_loan

Last_month

Last_weekday

PrevCampaign_Result

Call_type

Campaign_Success



Numerical

Age

LastCall_Dur

NewCampaign_CallNo

PrevCampaign_CallNo

Campaign_Intervals_Day

Missing & Unknown Values

"Unknown" Values

Missing Values

0

Job

330

% 0.8

Marital

80

% 0.19

Education

1731

% 4.2

Credit

8597

% 20.9

Housing_Loan

990

% 2.4

Personal_Loan

990

% 2.4

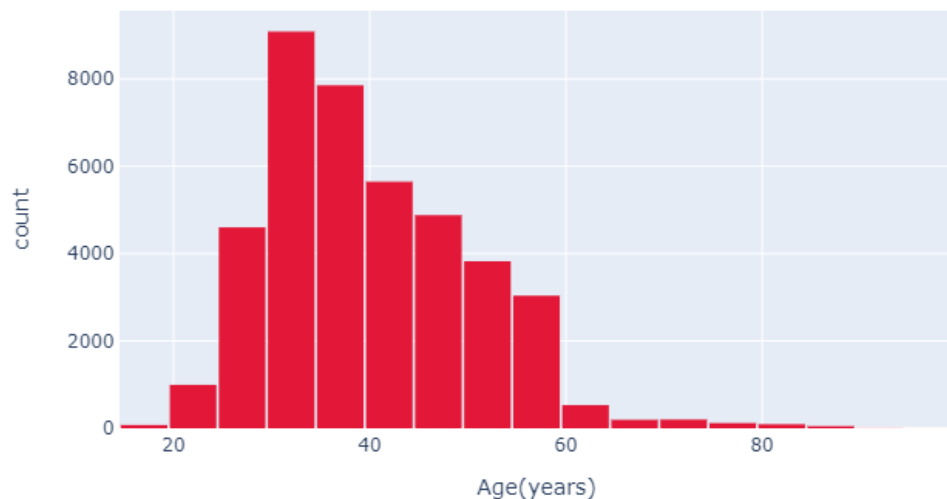
Univariate Preliminary Visualization

Numeric Variables

Preliminary Visualization: Numeric Variables

Age

Age Distribution



The data for age variable is between 15 and 98 and more than 50% of the clients are younger than 40 years old.

Mean

40.02

Median

38

Var

108.6

Distinct No

78

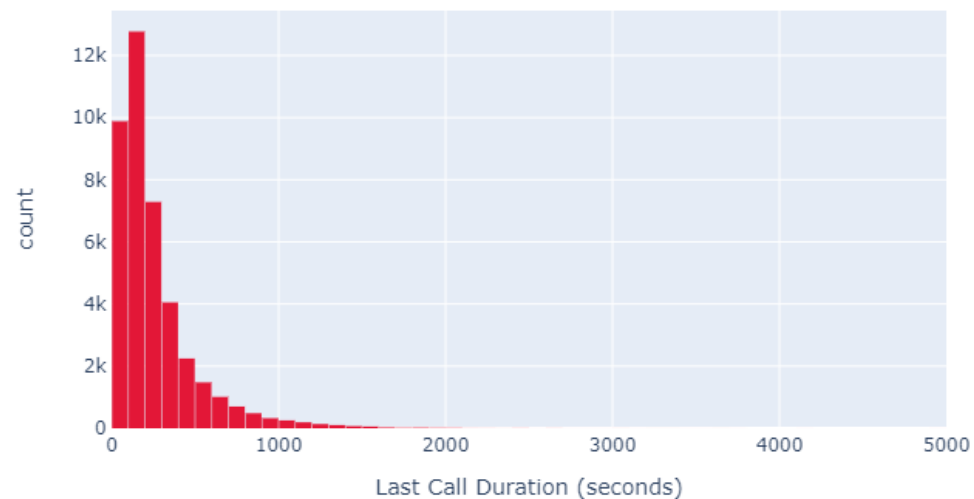
Skew

Right

November 5, 2021

LastCall_Dur

Last Call Duration Distribution



More than 30% of the call durations fall between 100 and 200 seconds. There maybe some outlier data in this variable.

Mean

258.3

Median

180

Var

259.3

Distinct No

1544

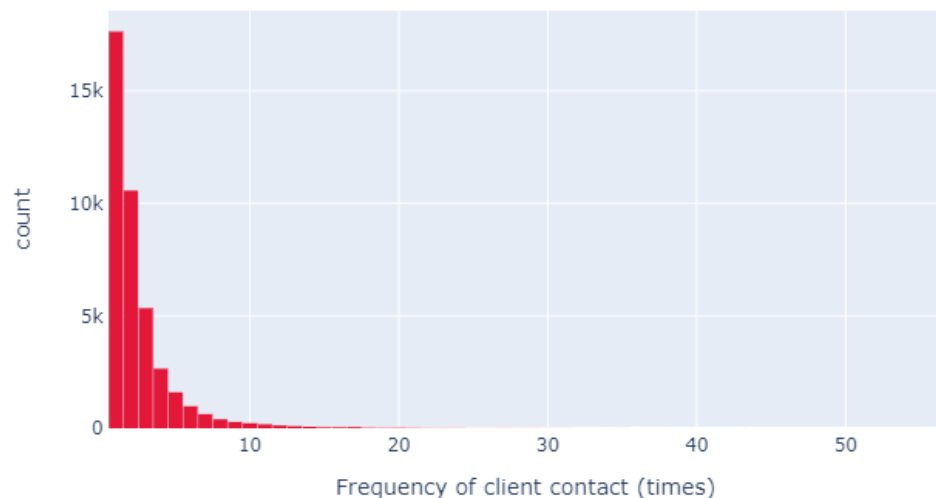
Skew

Right

Preliminary Visualization: Numeric Variables

NewCampaign_CallNo

Distribution of the times that a client has been contacted



This variable is highly skewed and almost 70% of the observations were under 2. It has a long tail, and it has possible a few outlier data.

Mean

2.56

Median

2

Var

7.67

Distinct No

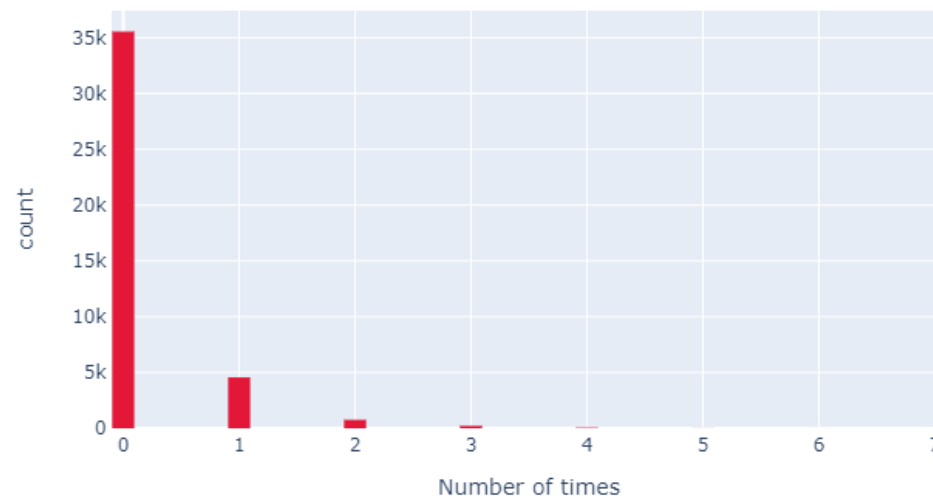
42

Skew

Right

PrevCampaign_CallNo

Distribution of the number of the client calls during the last campaign



This plot shows that most of the clients has not been included in the previous campaign (about 85%).

Mean

0.17

Median

0

Var

0.25

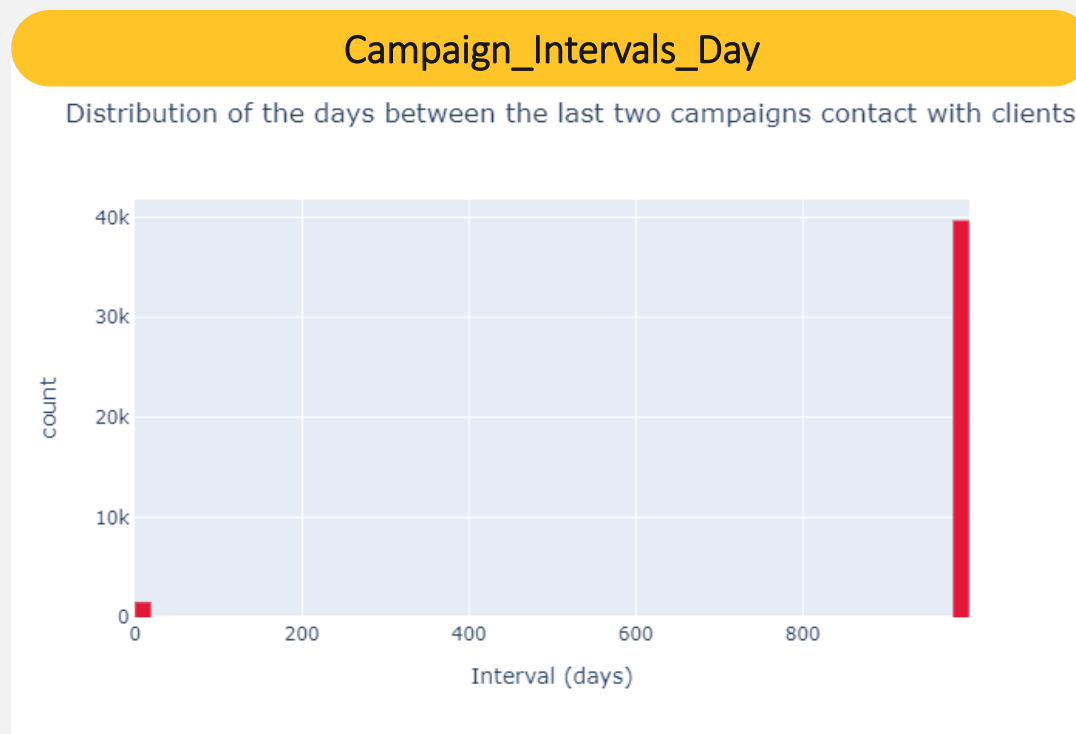
Distinct No

8

Skew

Right

Preliminary Visualization: Numeric Variables



This plot is consistent with the previous plot and shows that most of the clients have not been contacted during the previous campaign.

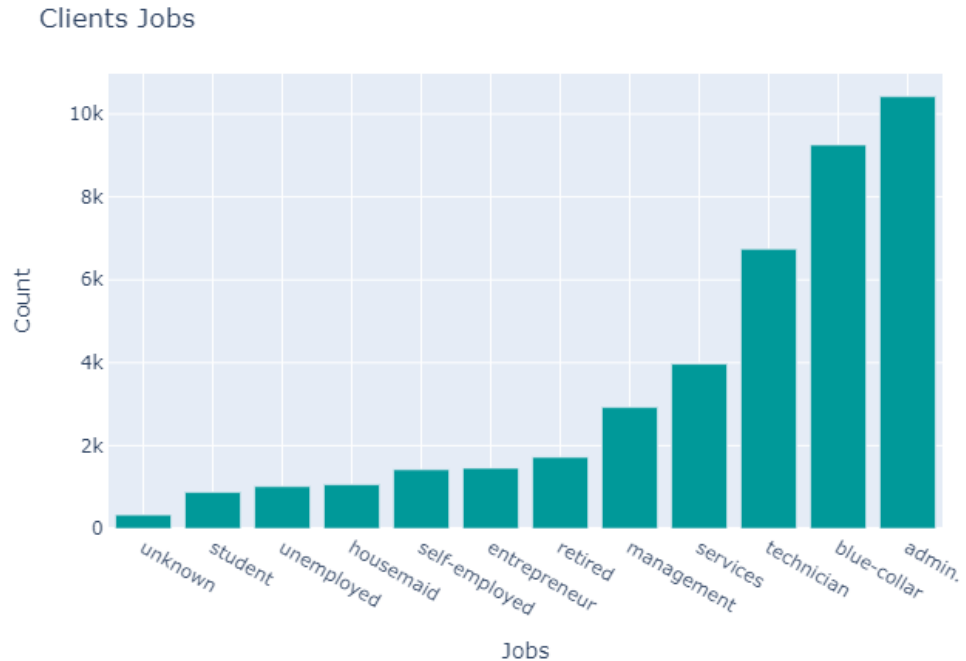
Mean	Median	Var	Distinct No	Skew
962.4	999	186.911	27	Left

Univariate Preliminary Visualization

Categorical Variables

Preliminary Visualization: Categorical Variables

Job

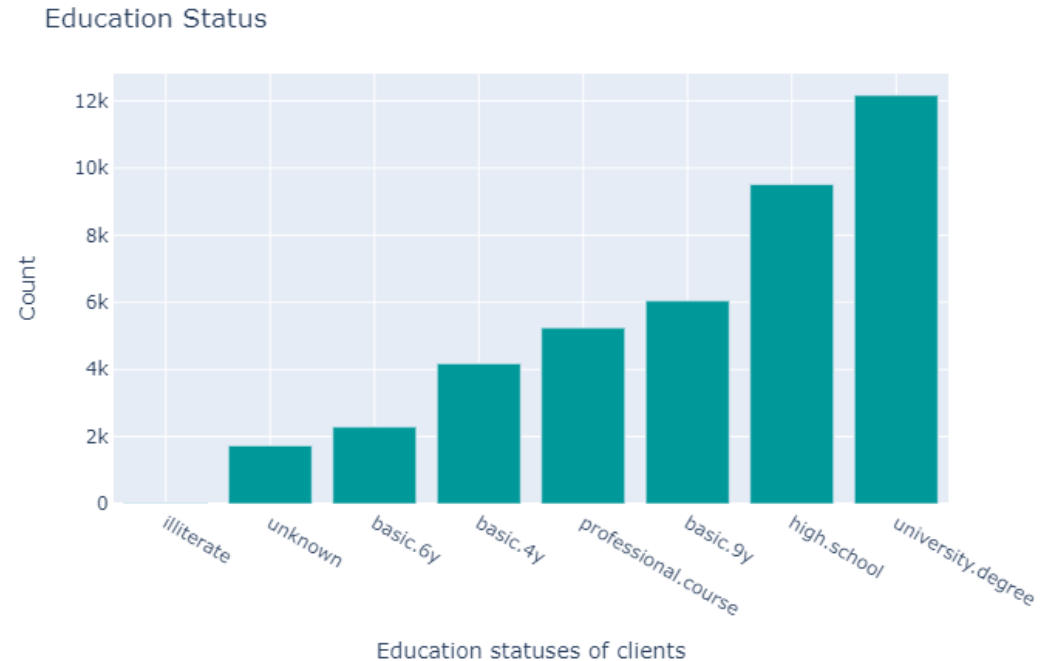


About half of the phone calls are made with clients that either have administrative roles or are blue-collar workers.

Distinct No

12

Education



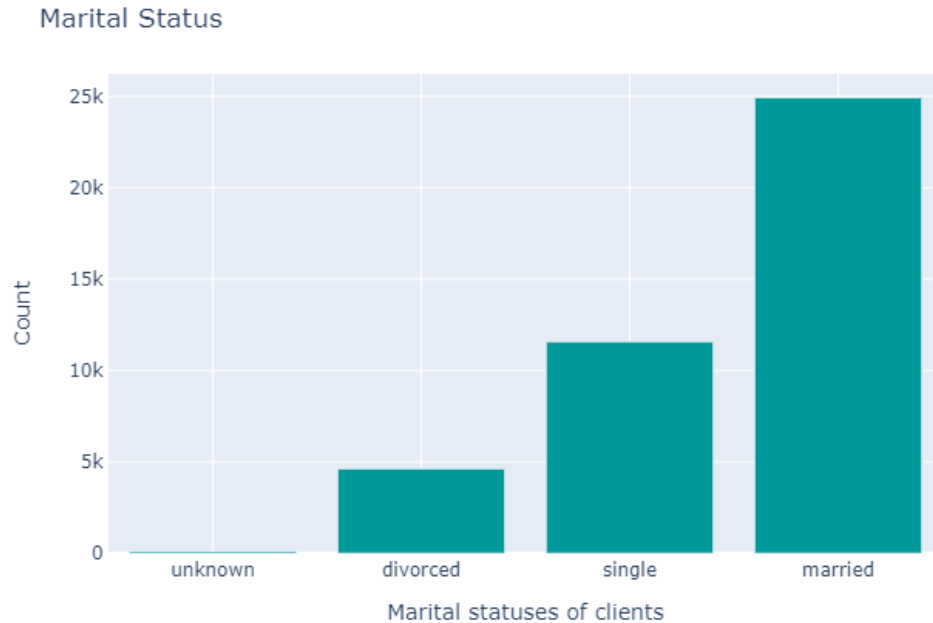
About 50% of the clients in this campaign have at least high-school level education. Education level of about 4% is unknown.

Distinct No

8

Preliminary Visualization: Categorical Variables

Marital

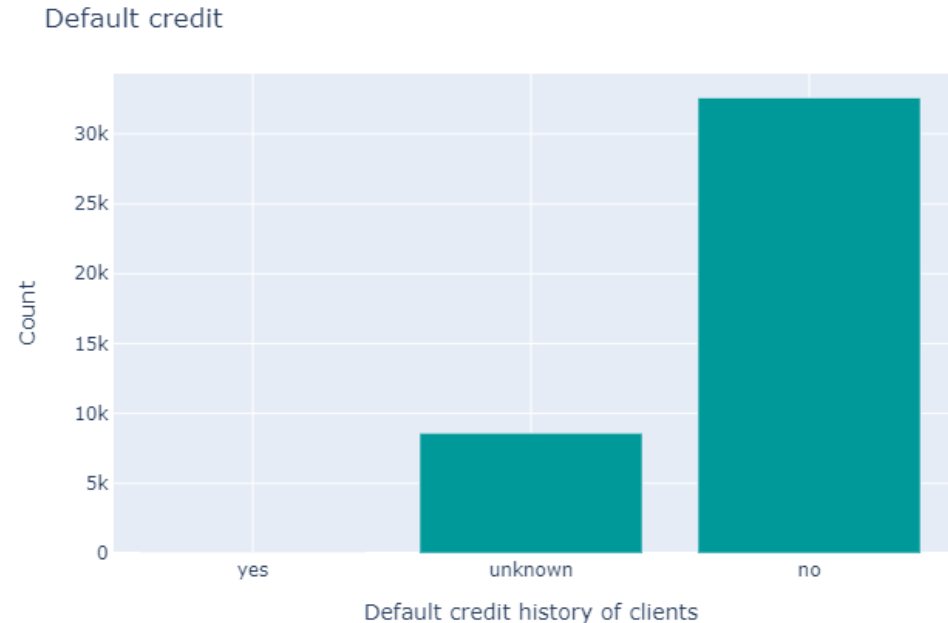


About 60% of the correspondents are married. Comparison of this data with the age group may have interesting outcomes.

Distinct No

4

Credit



Only 3 correspondents have default credits. The status of about 20% of the clients is unknown.

Distinct No

3

Preliminary Visualization: Categorical Variables

Personal_loan

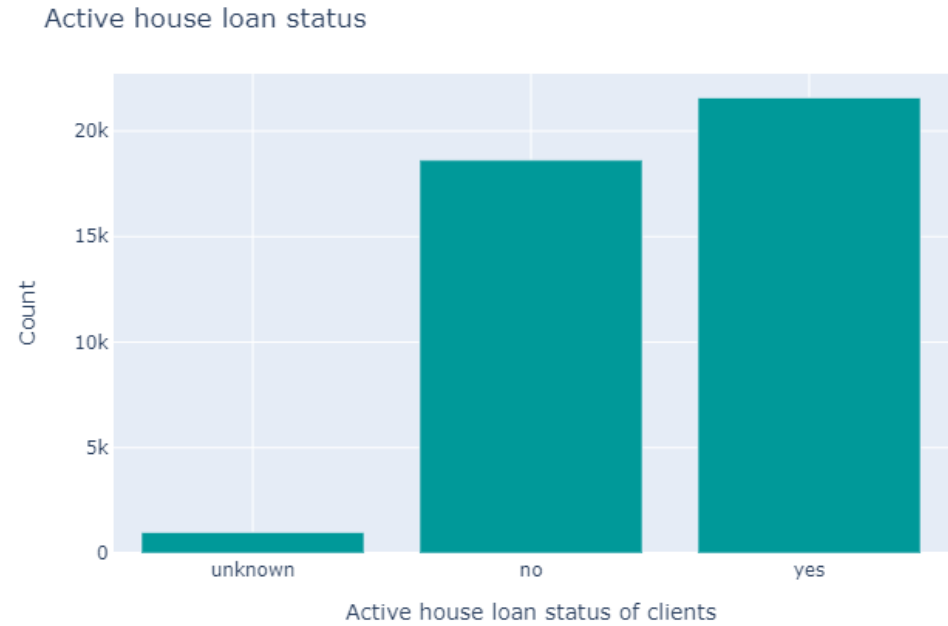


More than 80% percent of the correspondents have no personal loans.

Distinct No

3

Housing_loan



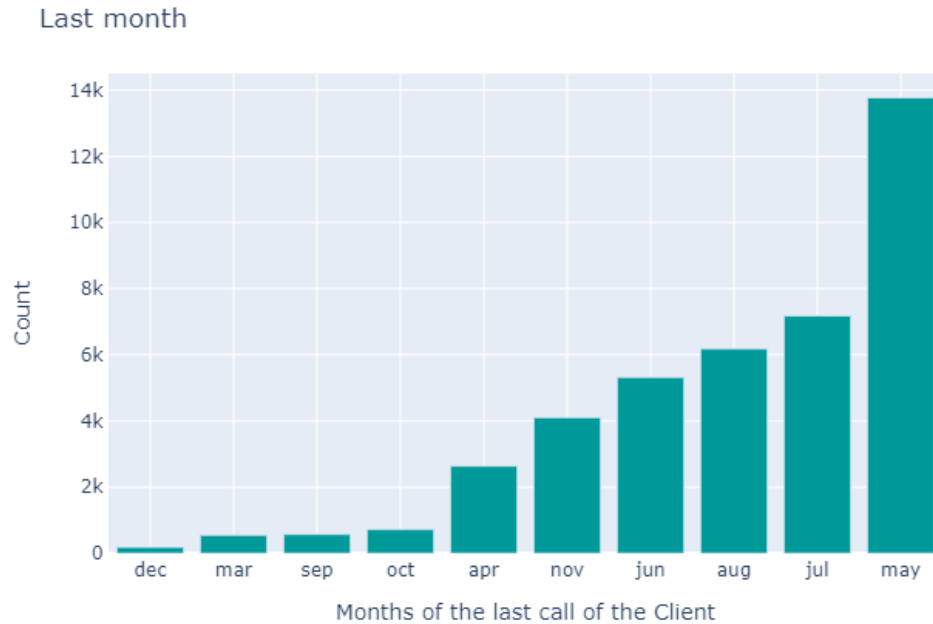
More than 50% of the correspondents have some kind of housing loans.

Distinct No

3

Preliminary Visualization: Categorical Variables

Last_month

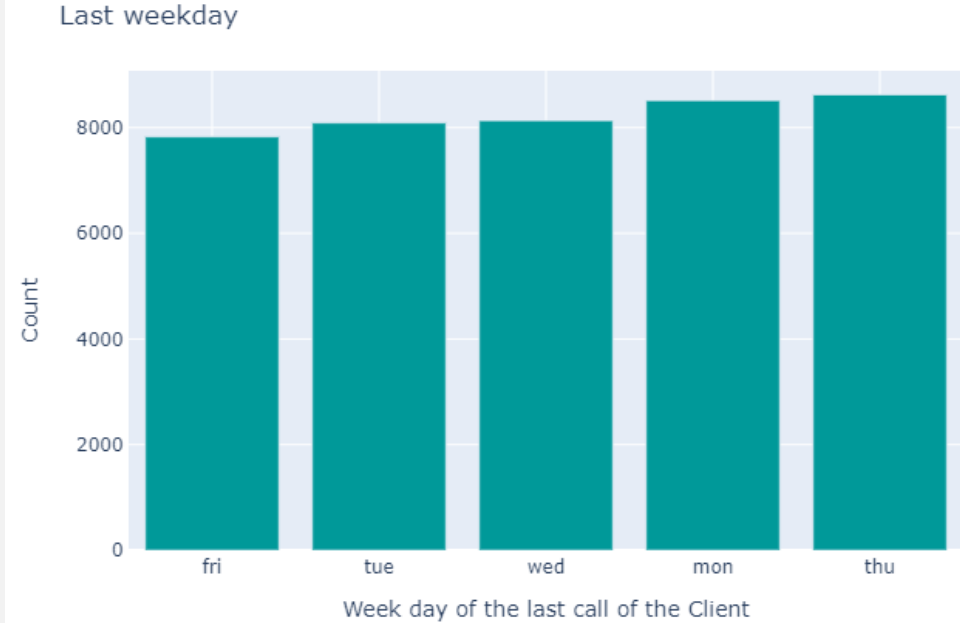


No calls were made during January and February. Most of the calls were made during May (approx. 34%).

Distinct No

10

Last_weekday



It seems that the calls were made in balance regarding the day of the week that calls were made.

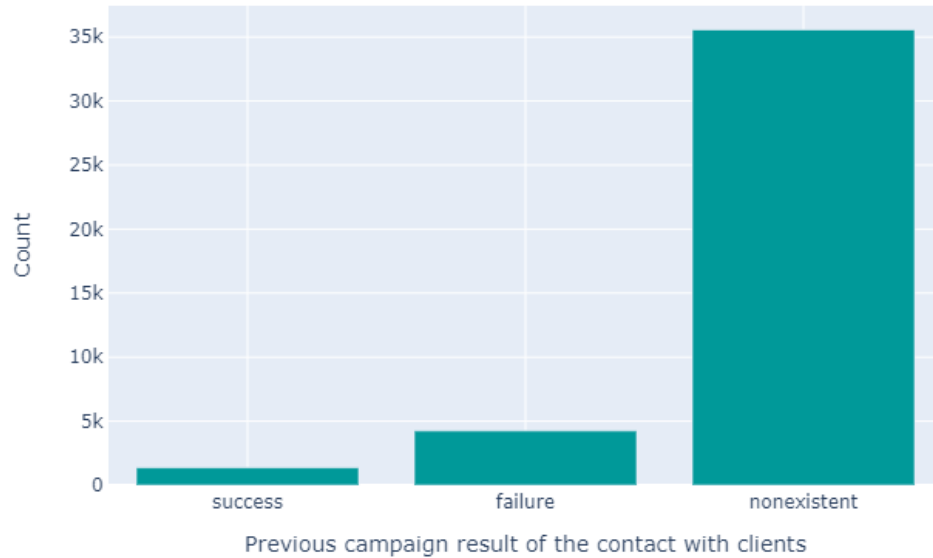
Distinct No

5

Preliminary Visualization: Categorical Variables

PrevCampaign_Result

The result of the previous campaign



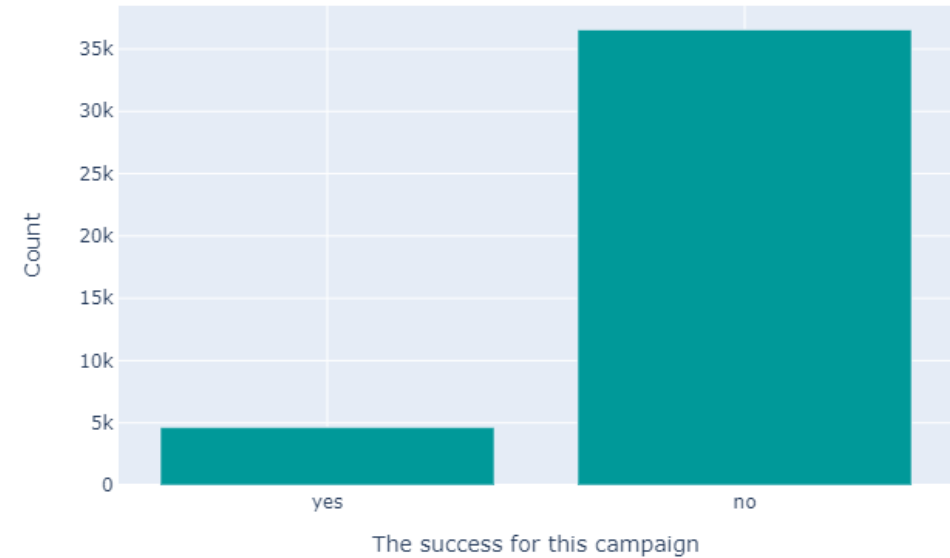
It seems majority of the correspondents are selected from clients that have not been targeted during previous campaigns (87%).

Distinct No

3

Campaign_Success

The success status for the campaign



The rate of success (signing up to deposit into a special account) in this campaign was 11.2%

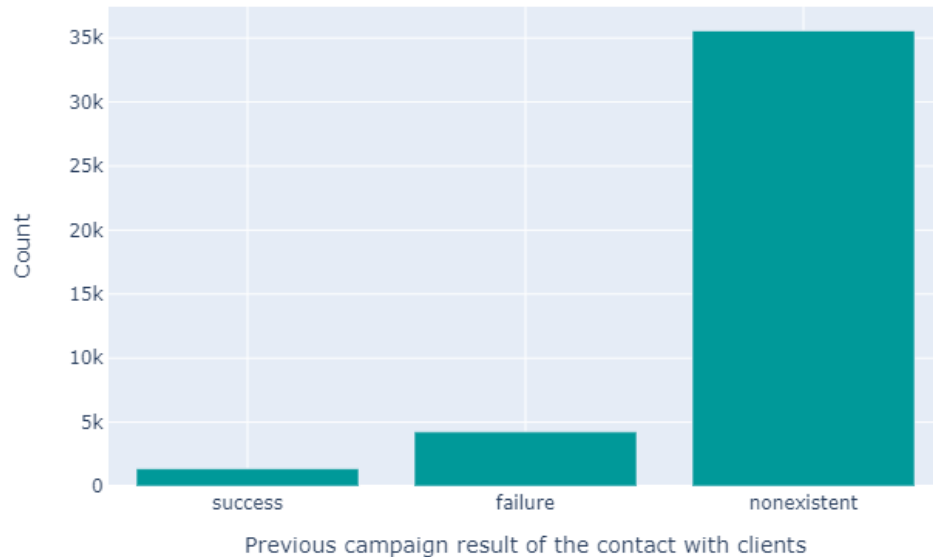
Distinct No

2

Preliminary Visualization: Categorical Variables

PrevCampaign_Result

The result of the previous campaign



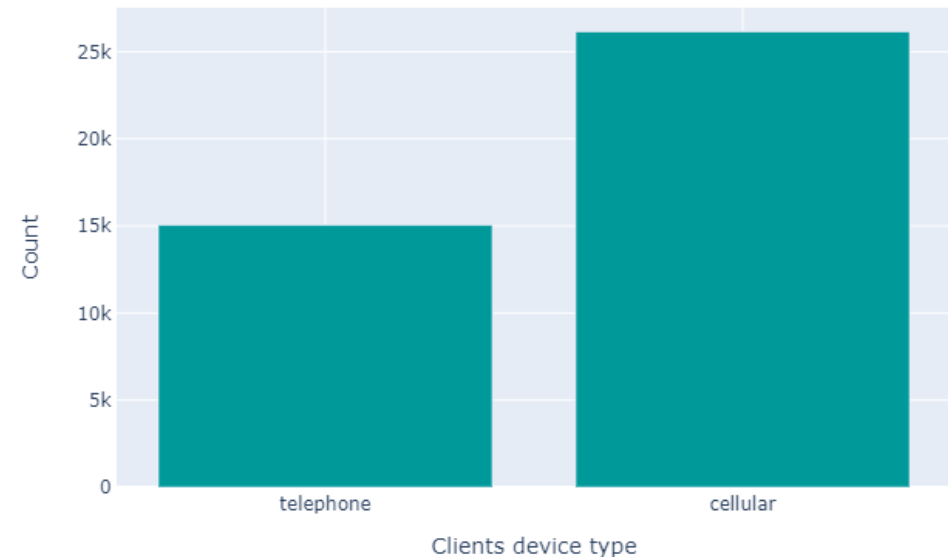
It seems majority of the correspondents are selected from clients that have not been targeted during previous campaigns (87%).

Distinct No

3

Call_type

Clients device type

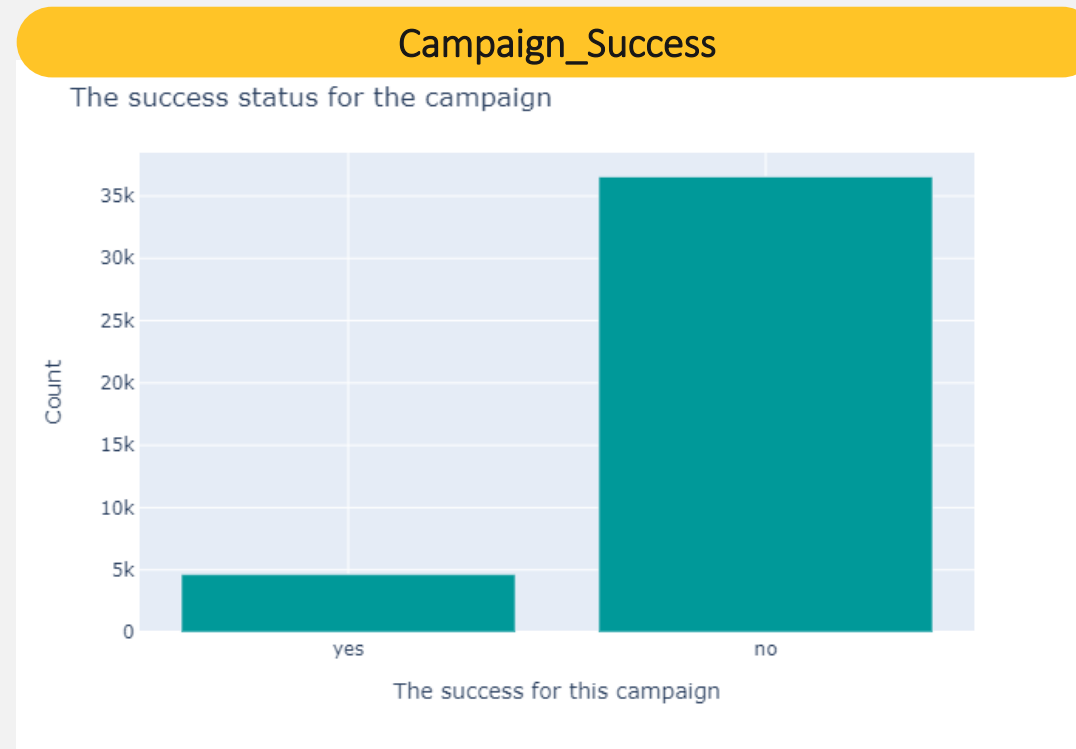


Most of the correspondents were contacted via their cellphones rather than their telephones (63.5% vs. 36.5%).

Distinct No

2

Preliminary Visualization: Categorical Variables



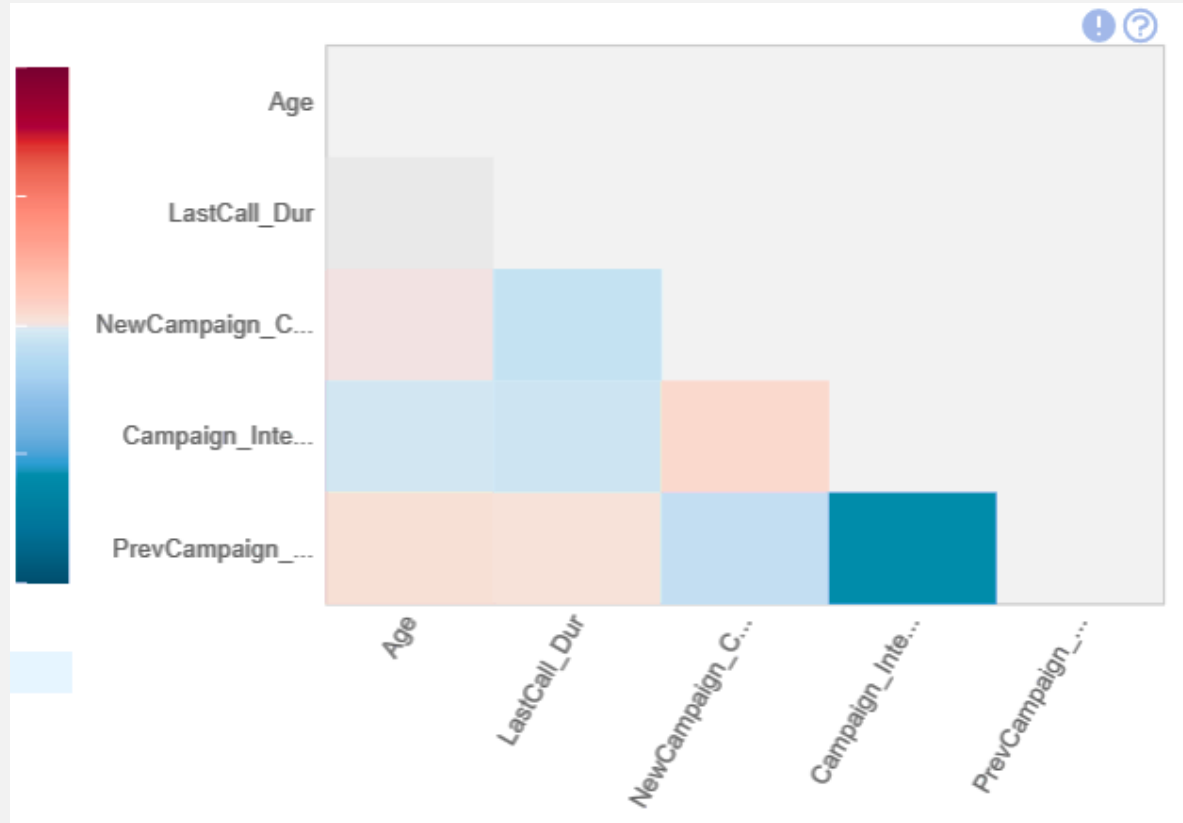
The rate of success (signing up to deposit into a special account) in this campaign was 11.2%

Distinct No
2

Correlations

- Pearson correlation matrix

Pearson correlation matrix (by dataprep.eda)



Cor. Method	Pearson	Spearman	KendallTau
Highest Positive Correlation	0.053	0.056	0.05
Highest Negative Correlation	-0.588	-0.51	-0.5
Lowest Correlation	0.001	0.001	0.001
Mean Correlation	-0.058	-0.054	-0.05

In a general view of the correlation matrix of a subset of the dataset (only the numerical variables are included in this subset), it seems only “PrevCampaign_CallNo” and “Campaign_Intervals_Day” have a considerable negative correlation. This correlation is somewhat expected due to the nature of these variables. More thorough analysis is needed to determine othe possible correlations.

Future Exploration & Transformation steps

Future Exploration & Transformation steps

Based on the initial exploration done in this phase these future analysis should be done in this project:



- Determination of outlier data
- Cleaning the data set
- Normalizing variables using appropriate methods
- Additional exploration of correlations
 - Exploring pair correlations
 - Including categorical variables in correlations

Resources

- <https://www.investopedia.com/terms/t/telemarketing.asp#:~:text=Key%20Takeaways-,Telemarketing%20is%20the%20direct%20marketing%20of%20goods%20or%20services%20to,lead%20generation%2C%20and%20sales%20calls>
- <https://data.world/data-society/bank-marketing-data>
- <https://glutch.com/telemarketing/create-a-plan-for-a-telemarketing-campaign>
- <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>



THANK YOU