Bank on this : Strategies for winning clients

Bank marketing dataset analysis

Ketaki Sahasrabudhe

Abstract

This report takes an in-depth look at the marketing campaigns of a Portuguese bank focused on promoting term deposits. By examining the data, we explore how client backgrounds, campaign strategies, and economic factors shape subscription rates. We discovered that clients in managerial or retired positions, especially those with higher education, are more inclined to subscribe to these offers. Interestingly, longer and more meaningful conversations during calls tend to yield better results than simply making numerous contact attempts.

Moreover, the success of these marketing efforts is closely linked to the economic environment, particularly during months when Euribor rates are low and employment rates fluctuate positively. We also identified that timing is key; targeting potential clients during peak subscription months like March, September, and Fridays can significantly boost success rates.

These findings offer valuable recommendations for banks, emphasizing the importance of focusing on quality interactions and aligning marketing strategies with economic trends, rather than relying on the quantity of contacts alone.

Introduction

The dataset contains data from the marketing campaigns conducted by a Portuguese banking institution, with the primary objective of promoting bank term deposits to clients. These marketing campaigns were carried out via phone calls, with each client receiving one or more calls. The goal of each contact was to assess whether the client would subscribe to the term deposit (classified as 'yes') or not (classified as 'no'). The dataset includes a mix of client-specific information (such as age, job type, marital status, and education level), along with data from the marketing campaign itself, including the duration of calls, number of contacts made, and days passed since the last contact. Furthermore, economic indicators like the employment variation rate, consumer price index, and Euribor rates are also incorporated, providing context for the broader economic environment during the campaigns. We will use this dataset for analyzing and predicting customer behavior in response to marketing efforts,and in turn help banks strategize more efficiently for future campaigns. Analysis of the background of the client base which subscribed will also help the bank to know which customer base to focus on and target more.

Data Description

The dataset consists of 41,188 observations (rows) and 21 variables (columns).

Structure of the data:

Observations: 41,188

Variables: 21 (demographic, financial, and campaign-related)

Age: Range from 17 to 98 years (mean = 40.02)

Categorical variables: Job, marital status, education, housing loan, personal loan, contact method, month, day of the week

Target variable: Subscribed (whether a client subscribed to a term deposit)

Economic indicators: Employment variation rate (-3.4 to 1.4), consumer price index (92.2 to 94.77), consumer confidence index (-50.8 to -26.9), euribor 3-month rate (0.63 to 5.04)

Call duration: Range from 0 to 4918 seconds (mean = 258.3)

Campaign contacts: Number of contacts made to a client (1 to 56)

Pdays: Most values set to 999, indicating no prior contact

Previous campaign outcomes: Recorded in poutcome, with majority labeled "nonexistent"

Variables:

age: Age of the client (numeric).

job: The type of job held by the client (categorical, e.g., admin., technician, blue-collar, etc.).

marital: Marital status of the client (categorical, e.g., married, single).

education: Education level of the client (categorical, e.g., university degree, high school).

default: Whether the client has credit in default (categorical: yes, no).

housing: Whether the client has a housing loan (categorical: yes, no).

loan: Whether the client has a personal loan (categorical: yes, no).

contact: Communication type (categorical: cellular, telephone).

month: Last contact month (categorical).

day_of_week: Day of the week of the last contact (categorical).

duration: Duration of the last contact (numeric).

campaign: Number of contacts during the current campaign (numeric).

pdays: Days since last contact in a previous campaign (numeric, with 999 meaning not previously contacted).

previous: Number of previous contacts before this campaign (numeric).

poutcome: Outcome of the previous marketing campaign (categorical: success, failure, nonexistent).

emp.var.rate: Employment variation rate (numeric, quarterly indicator).

cons.price.idx: Consumer price index (numeric, monthly indicator).

cons.conf.idx: Consumer confidence index (numeric, monthly indicator).

euribor3m: Euribor 3-month rate (numeric, daily indicator).

nr.employed: Number of employees (numeric, quarterly indicator).

subscribed: Whether the client subscribed to a term deposit (binary: yes, no).

Exploratory Data Analysis:

Data cleaning

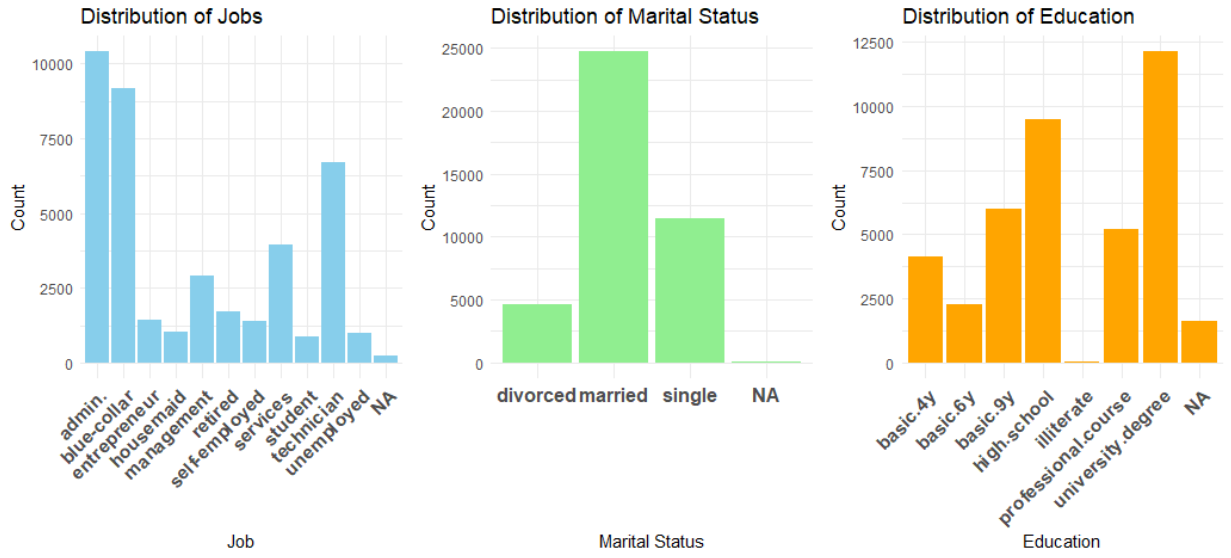
We can see that some variables contain some fields marked as "unknown" which means the customers haven't provided that information, so the rows which contain more that two "unknown" values are basically of no use for us, as it doesn't contain much useful data, so we remove those rows and convert all "unknown" values to NA, in order to be able to process it further.

```
#Identifying unknown values
sapply(D, function(x) sum(x == "unknown"))
#This counts how many "unknown" values are there in each column
#replacing the "unknown" values with NA
D[D == "unknown"] = NA
#Cleaned Dataset
D_cleaned <- D[apply(D, 1, function(x) sum(is.na(x)) <= 2), ]
```

We also convert the categorical variables to factors. After removing the rows, we remove about 1% of the rows, which is not a significant loss, so we can proceed with further analysis.

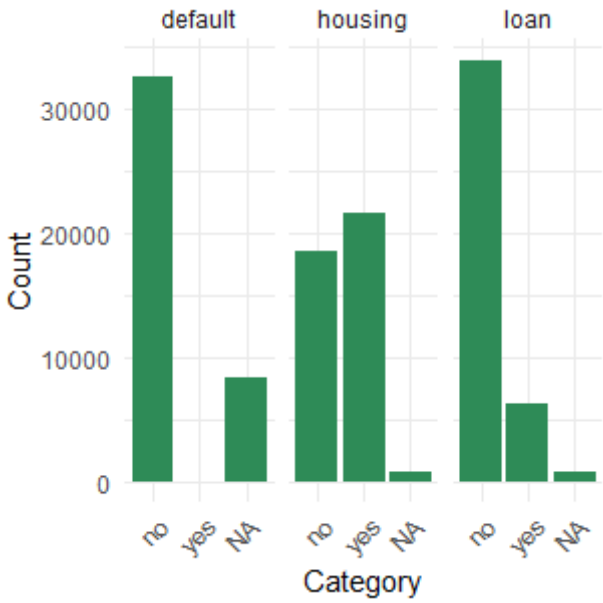
Univariate Analysis

We use bar plots to plot the categorical variables We plot Job, Marital status, and education in order to get an overview of the customers' background



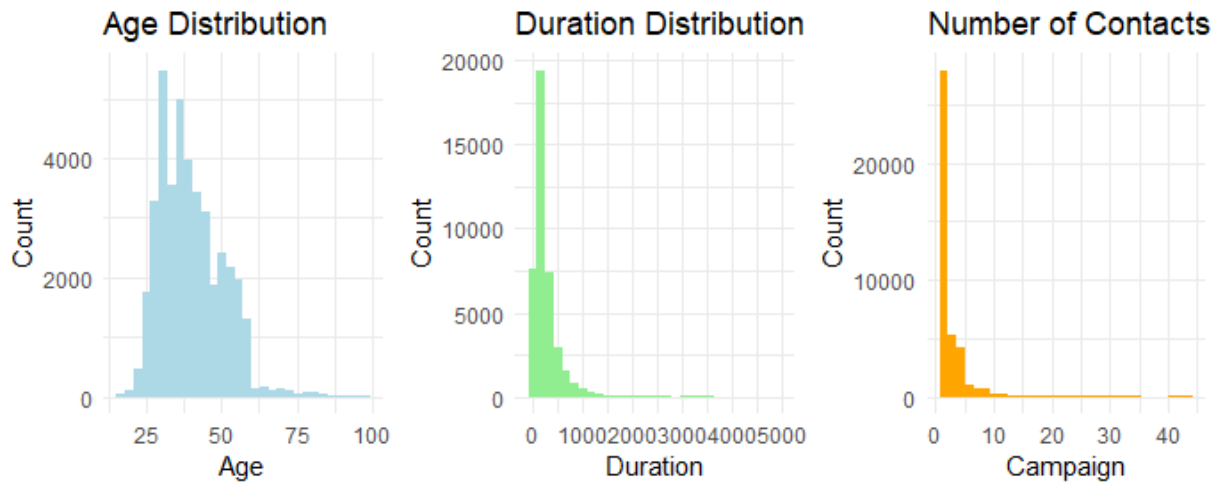
The plots show that the majority of clients work in "admin." or "blue-collar" jobs, are mostly married, and tend to have university or high school education. There are some missing values across all categories.

We plot Housing loan, Personal loan and Default records in order to get an idea about the financial stability of the customers.



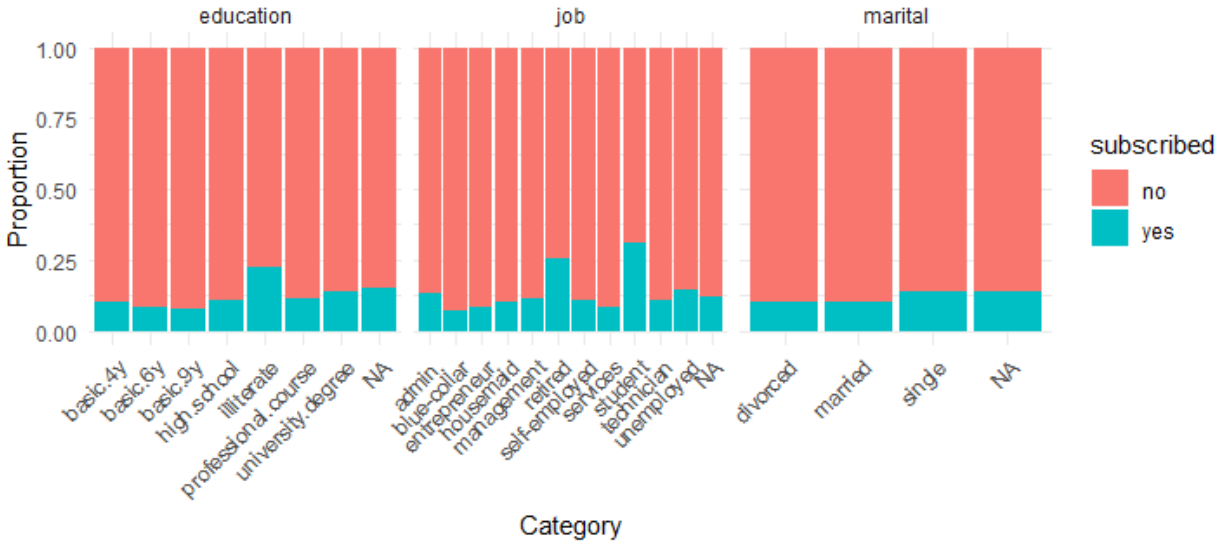
The majority of clients have no defaults, many hold housing loans, and relatively fewer have personal loans.

We use histograms for numerical variables, as they better represent continuous data



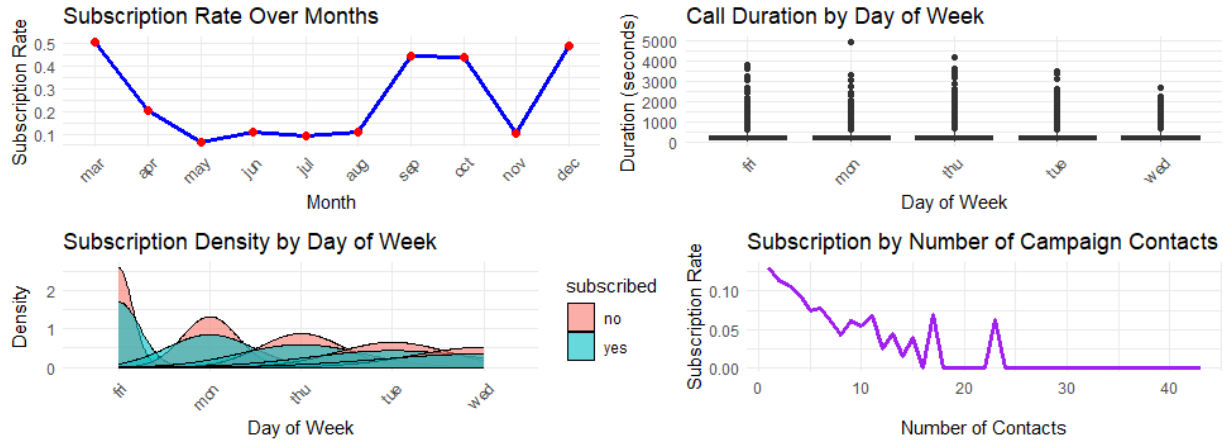
Bivariate Analysis

Here our taget variable is "subscribed" which is the ultimate goal of the campaign, so we analyse each variable with respect to the target variable, which will give us an idea about people from which kind of background tend to subscribe most of the times.



Clients with university degrees and management/retired jobs have higher subscription rates compared to other categories. Marital status shows little variation, with all categories having a similar subscription proportion.

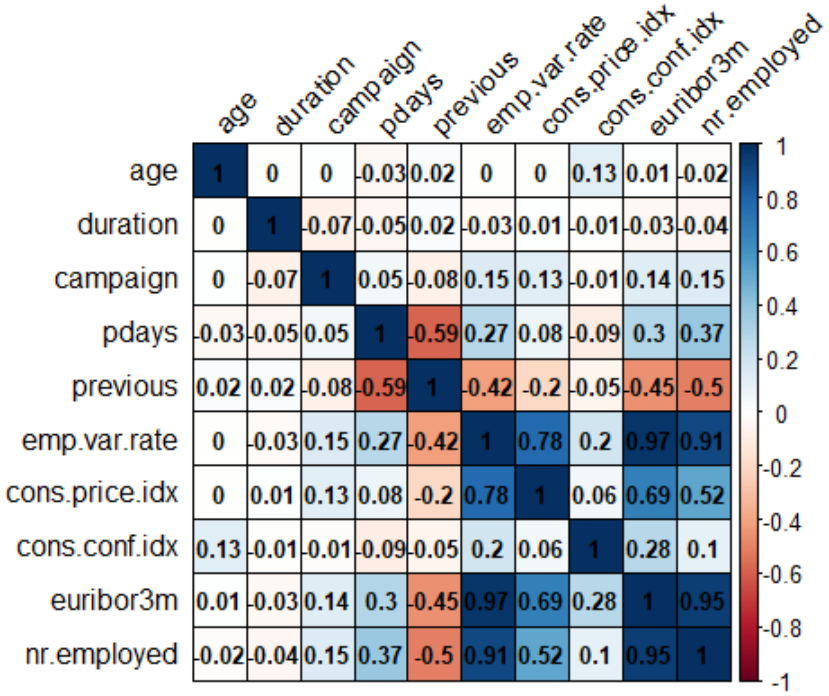
Time analysis:



Subscription trends peak during March, September, November, and December. Call durations are quite consistent across weekdays, with Tuesday showing more variation. Subscription density is higher on Fridays compared to other weekdays. Number of contacts negatively impacts the subscription rate, with diminishing returns after a few attempts.

Multivariate analysis:

A correlation matrix was created to identify relationships between numerical variables, helping us understand which factors are interdependent. This analysis provides insights into how economic indicators, campaign features, and client attributes interact, allowing us to pinpoint variables that play a crucial role in subscription success.



	age	duration	campaign	pdays	previous	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed
age	1	0	0	-0.03	0.02	0	0	0.13	0.01	-0.02
duration	0	1	-0.07	-0.05	0.02	-0.03	0.01	-0.01	-0.03	-0.04
campaign	0	-0.07	1	0.05	-0.08	0.15	0.13	-0.01	0.14	0.15
pdays	-0.03	-0.05	0.05	1	-0.59	0.27	0.08	-0.09	0.3	0.37
previous	0.02	0.02	-0.08	-0.59	1	-0.42	-0.2	-0.05	-0.45	-0.5
emp.var.rate	0	-0.03	0.15	0.27	-0.42	1	0.78	0.2	0.97	0.91
cons.price.idx	0	0.01	0.13	0.08	-0.2	0.78	1	0.06	0.69	0.52
cons.conf.idx	0.13	-0.01	-0.01	-0.09	-0.05	0.2	0.06	1	0.28	0.1
euribor3m	0.01	-0.03	0.14	0.3	-0.45	0.97	0.69	0.28	1	0.95
nr.employed	-0.02	-0.04	0.15	0.37	-0.5	0.91	0.52	0.1	0.95	1

From the correlation matrix, **duration of contact** shows no strong correlation with other variables but plays a significant role in predicting subscription success, suggesting that longer conversations are crucial. **Campaign contacts** also have minimal correlation with other variables, indicating that repeated contacts don't necessarily improve outcomes. Strategizing around **quality interactions** rather than quantity is recommended.

The negative correlation between **pdays** and **previous** suggests that re-engaging clients with prior successful contacts after a gap could be a smart move. Additionally, economic indicators like **euribor3m**, **employment rate**, and **number of employees** are closely related, so aligning campaigns with favorable economic trends could enhance success.

Results:

Demographics: Clients in managerial, retired, or administrative jobs with higher education levels have the highest subscription rates. Marital status has minimal impact on subscriptions.

Economic Indicators: The Euribor 3-month rate, employment variation rate, and number of employees show strong interdependencies, meaning campaign success is influenced by broader economic conditions.

Contact Strategy: The number of campaign contacts shows diminishing returns after a few attempts. Longer call durations are significantly associated with higher subscription rates, highlighting the importance of quality over quantity in client interactions.

Time Factors: Subscriptions peak in March, September, November, and December. Fridays and specific months see higher subscription rates.

Conclusion:

Client Demographics: Targeting clients in management or retired roles and those with higher education yields better subscription rates. Banks should focus marketing efforts on this segment.

Campaign Approach: Effective client engagement comes from longer, meaningful conversations rather than repeated contact attempts. Banks should emphasize personalized interactions.

Economic Timing: Aligning campaigns with favorable economic conditions, especially during periods of low Euribor rates and higher employment variations, can enhance subscription success.

Timing of Campaigns: Conducting campaigns during peak months and targeting Fridays could boost subscription rates.

Discussion:

This report shows that focusing on clients in managerial or retired roles with higher education can lead to more subscriptions for term deposits. It's clear that having quality conversations with clients is more important than just making a lot of calls; meaningful interactions make a real difference.

Additionally, timing marketing efforts to coincide with favorable economic conditions, especially when Euribor rates are low, can help improve campaign success. In conclusion, the success of a marketing campaign relies on targeting the right people, aligning efforts with economic trends, and building better relationships with clients. Looking ahead, it would be helpful to study how changing economic factors affect client decisions.

References:

Kaggle Bank Marketing Dataset