

BANKING DATASET CASE PROJECT

Analyze data from Portuguese bank phone campaigns to understand subscription to term deposits.

A case study carried out by Vansh Sinha

OUTLINE

- Introduction
- Data cleaning
- Key visualizations
- Analysis



INTRODUCTION

SUMMARY OF DATASET

The dataset details phone call marketing campaigns by a Portuguese bank for promoting term deposits. It aims to predict client subscription to these deposits, emphasizing the efficiency of telephonic outreach despite the costs. This focus on optimizing marketing efforts makes the dataset valuable for strategic planning in banking.

COLUMN DESCRIPTIONS

- **age:** Client's age.
- **job:** Job category.
- **marital:** Marital status.
- **education:** Education level.
- **default:** Credit default status.
- **balance:** Yearly balance.
- **housing:** Housing loan status.
- **loan:** Personal loan status.
- **contact:** Contact method.
- **day:** Last contact day.
- **month:** Last contact month.
- **duration:** Last contact duration.
- **campaign:** Contacts in this campaign.
- **pdays:** Days since last campaign.
- **previous:** Contacts before this campaign.
- **poutcome:** Outcome of previous campaign.
- **y:** Term deposit subscription (target).

DATA CLEANING

Resolving issues of messy and dirty data



ISSUES WITH DATASET

DIRTY DATA

- **Completion issues** like filling null values.
- **Validity issues** like wrong data-type assigned.
- **Accuracy issues** like duplicate entries.

MESSY DATA

- **Redundant columns**
- **Unnecessary columns**
- **Renaming columns**

KEY VISUALIZATIONS

Visualizations from the dataset

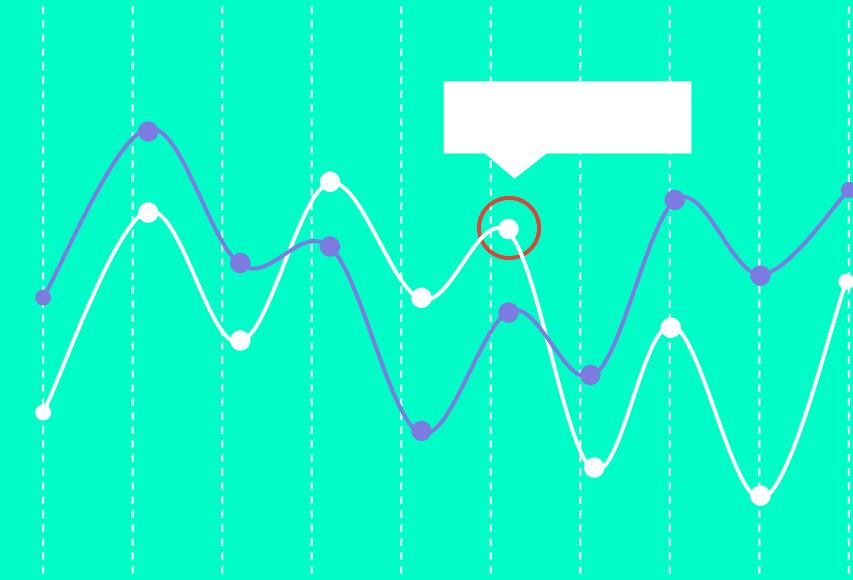


CHART 1: CLIENTS VS AGE

Client ages vary from 18 to 95, mostly late 30s to early 40s, with a mean age of 41 and a median of 39 years, indicating a slight skew towards younger clients.

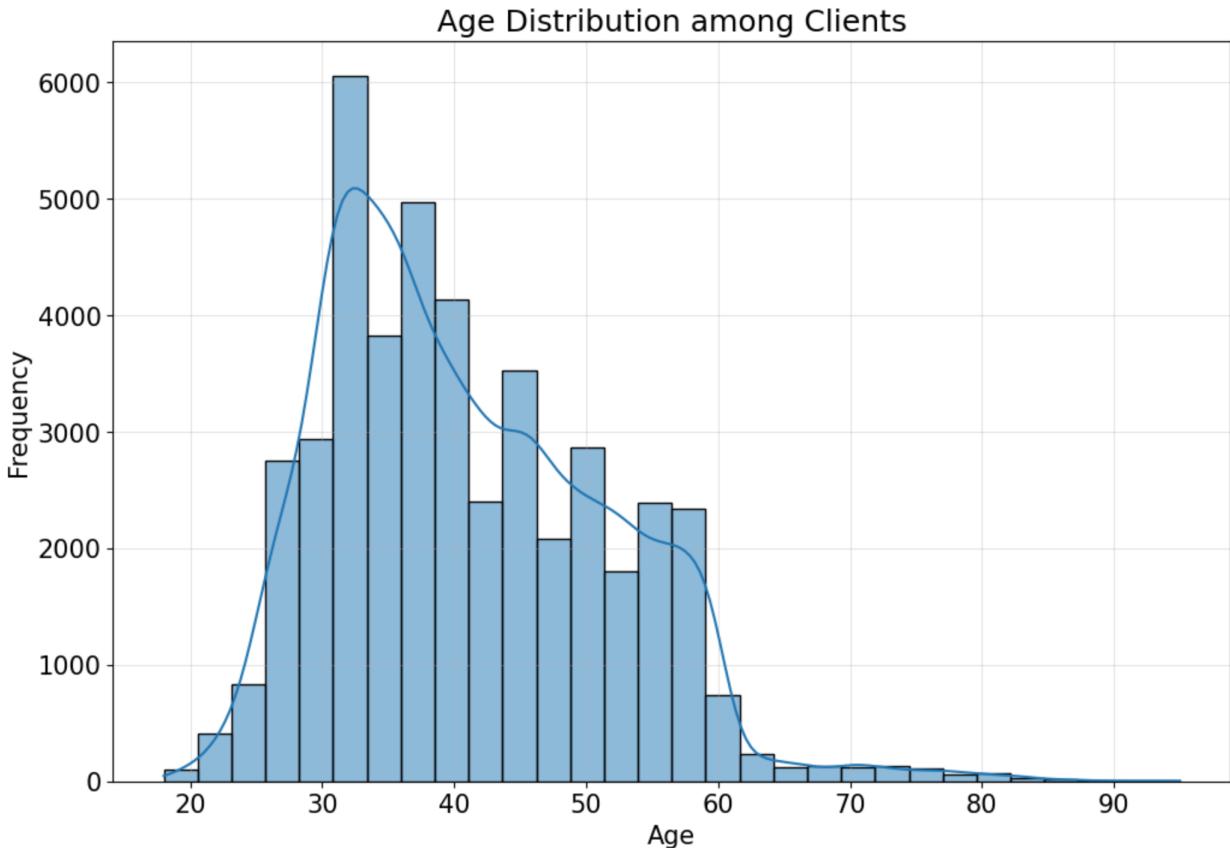
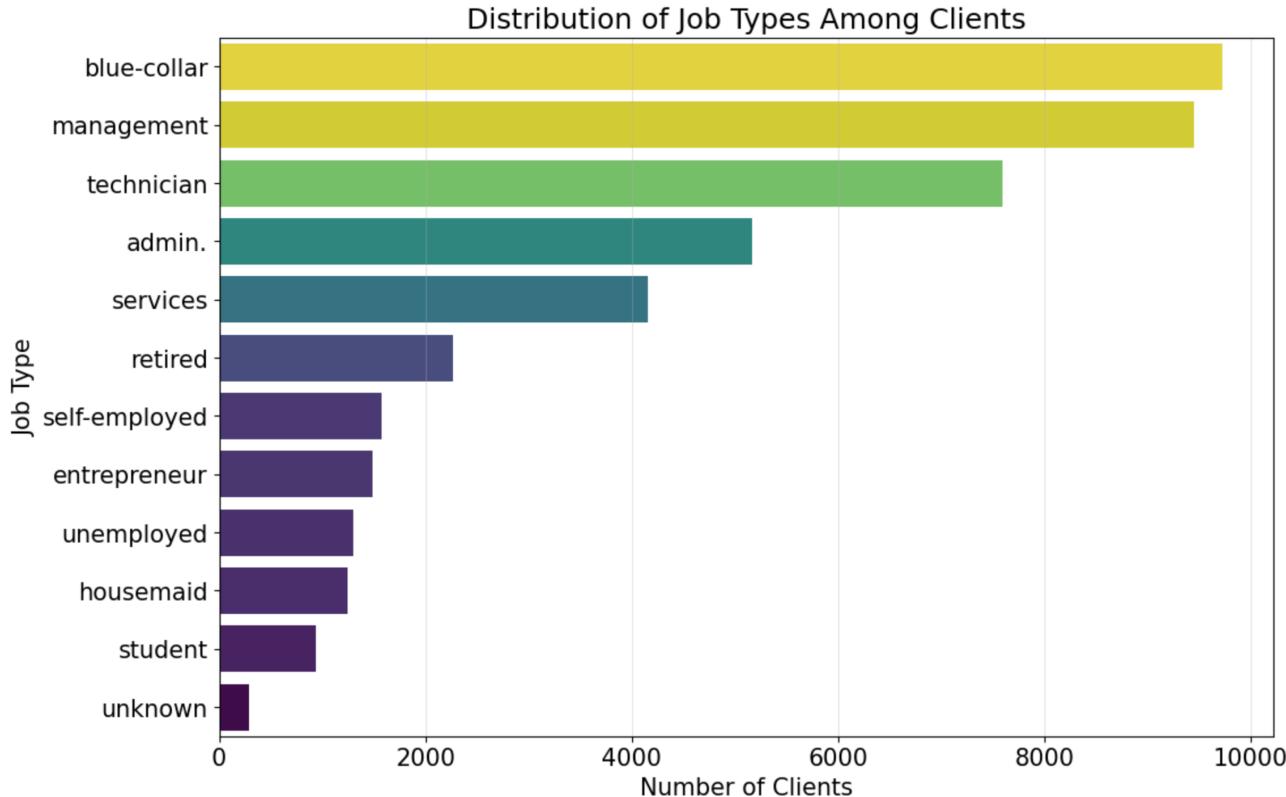


CHART 2: CLIENTS VS JOB

The top three job categories among clients are blue-collar workers, management, and technicians, while students form the smallest client group.



Marital Status Distribution Among Clients

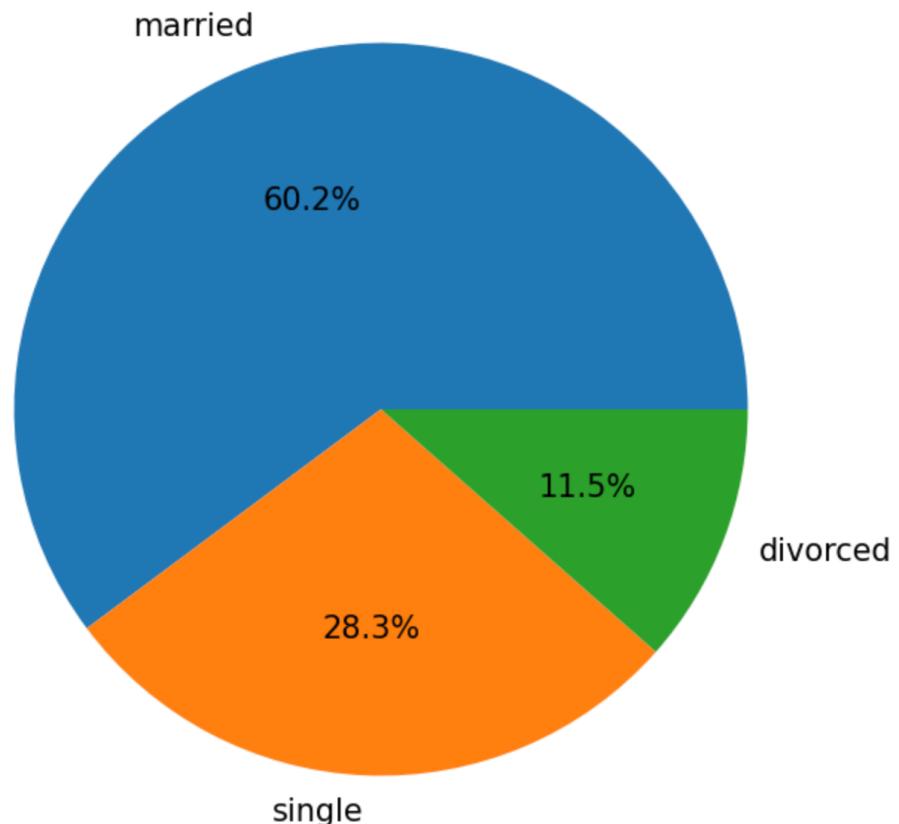


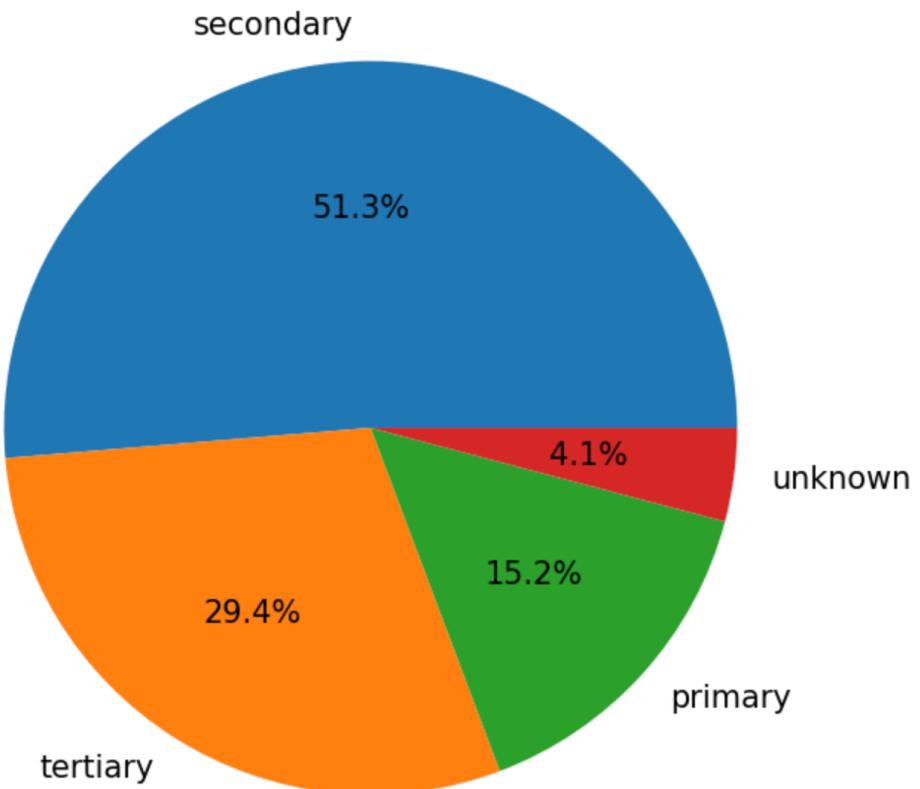
CHART 3: CLIENTS VS MARITAL STATUS

Around 60% clients
are married, 10%
are divorced, 30%
are single.

Education Level Distribution Among Clients

CHART 4: CLIENTS VS EDUCATION

Over 50% of clients have secondary education, around 30% have tertiary, and 15% have primary education, with a small percentage having an unknown education level.



Clients with Credit in Default

CHART 5: CLIENTS VS CREDIT DEFAULT

Approximately 98.2% of clients do not have credit in default, while about 1.8% of clients do have credit in default.

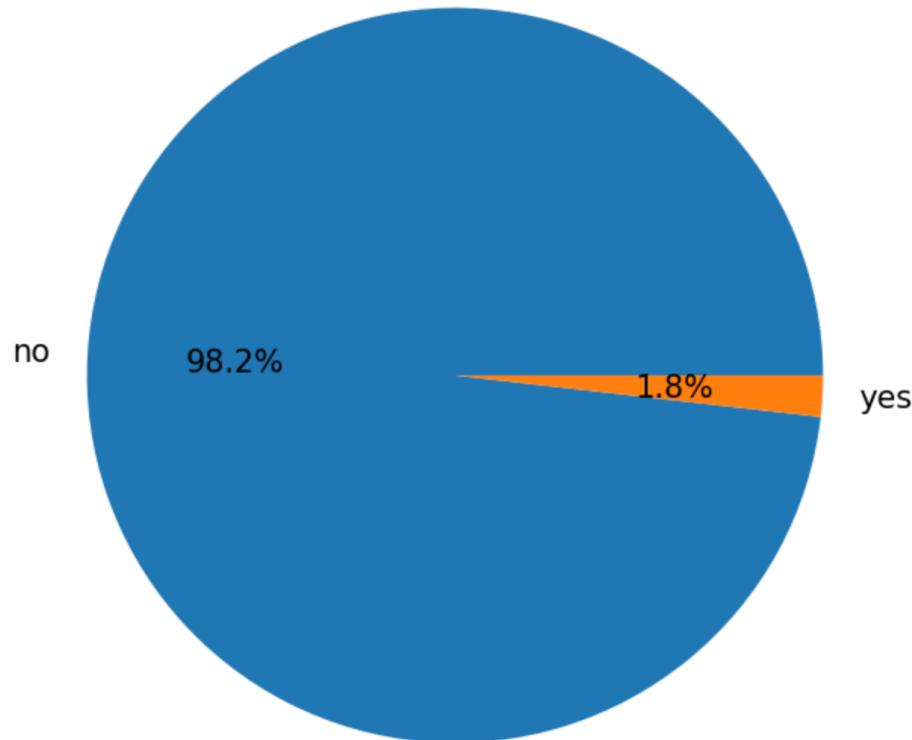


CHART 6: CLIENTS VS AVG. BALANCE

The average balance is €1,362, with a median of €448.5, indicating a skew towards lower balances among clients, ranging from -€8,019 to €102,127, highlighting both overdrawn accounts and significantly high balances.

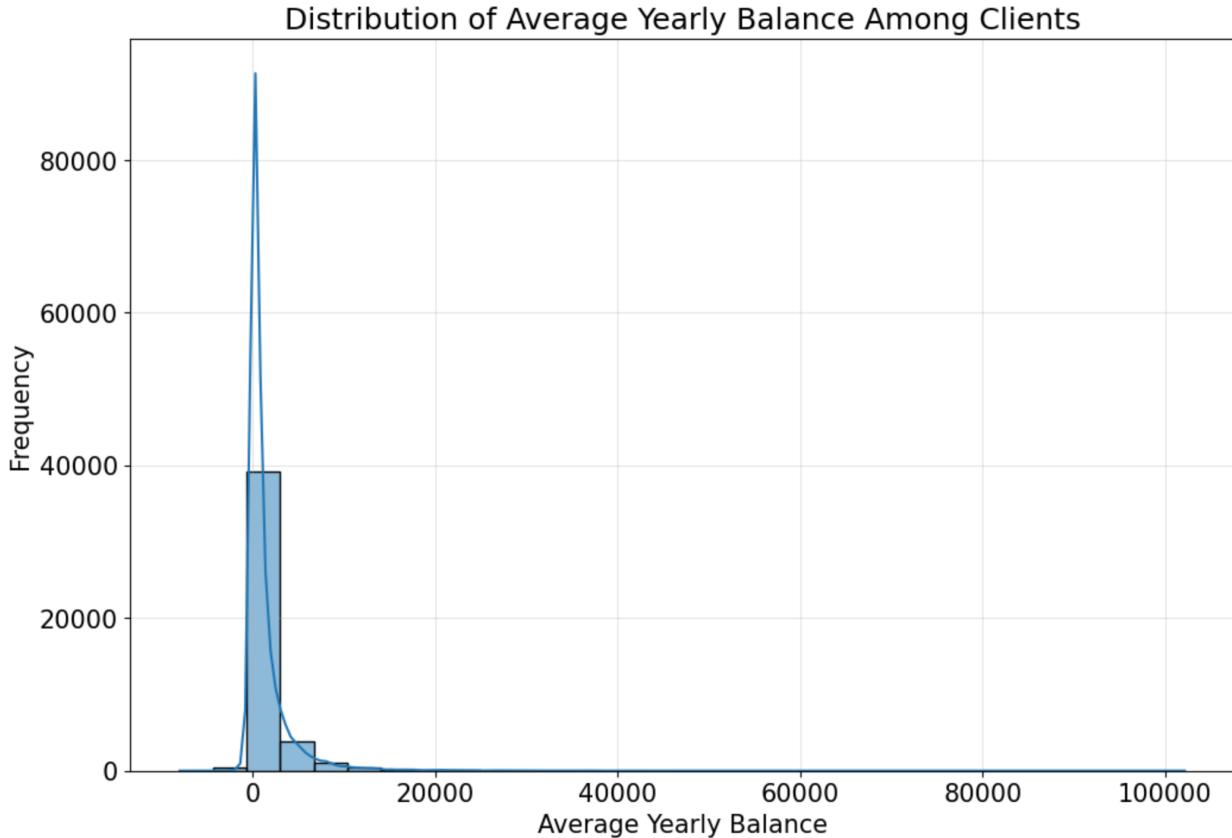


CHART 7: CLIENTS VS HOUSING LOAN

Over half of the clients, about 55.6%, possess housing loans, in contrast to the 44.4% who do not have such loans.

Clients with Housing Loans

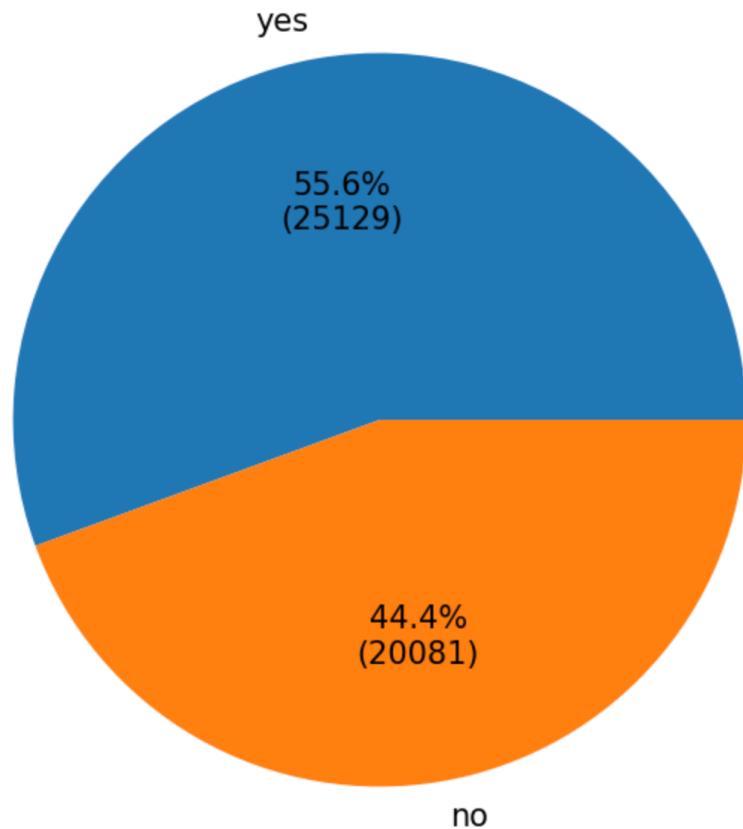
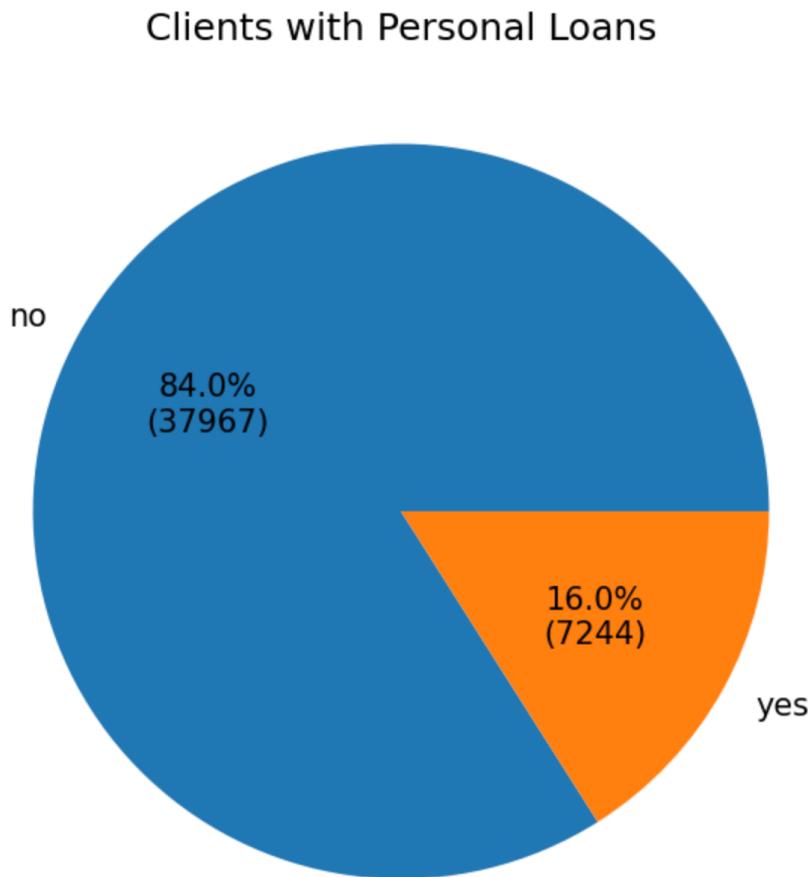


CHART 8: CLIENTS VS PERSONAL LOAN

Approximately 16% of clients have taken out personal loans, while the vast majority, 84%, do not have personal loans.



Communication Types Used for Contacting Clients

CHART 9: CLIENTS VS COMMUNICATION TYPE

Cellular phones were the most frequently used communication method, followed by instances where the communication type was not specified ('unknown'), and telephone calls were the least used method.

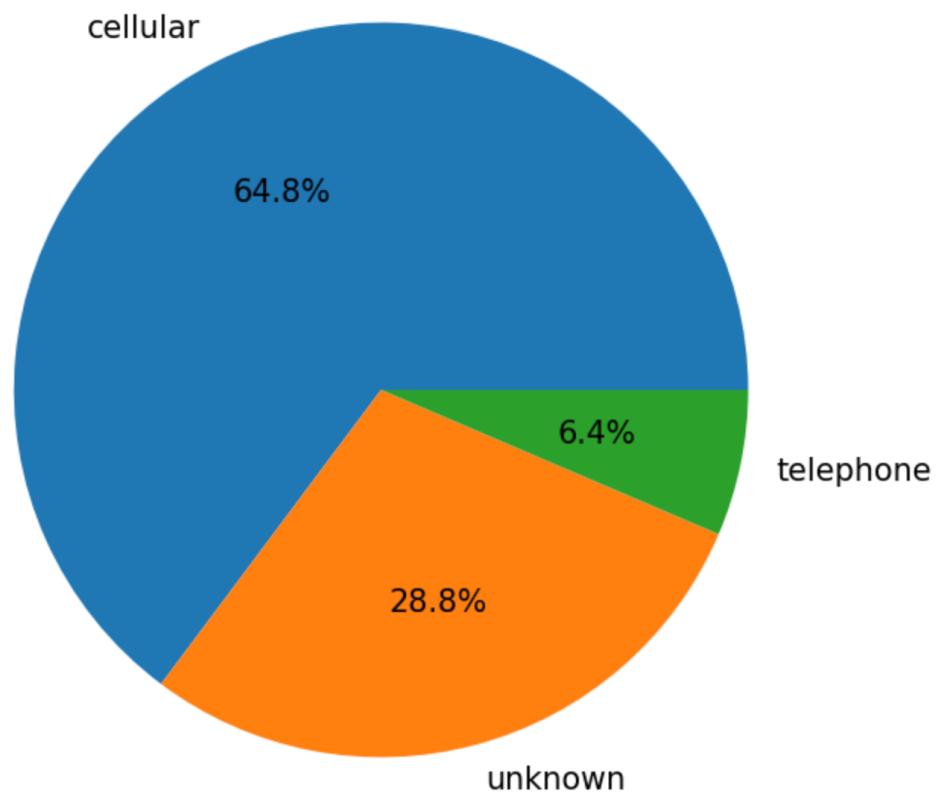


CHART 10: CONTACTS VS CONTACT DAY

The 20th of the month sees the peak in contact numbers, with the 18th and 21st also experiencing high activity. Contact frequency dips at the start and end of the month.

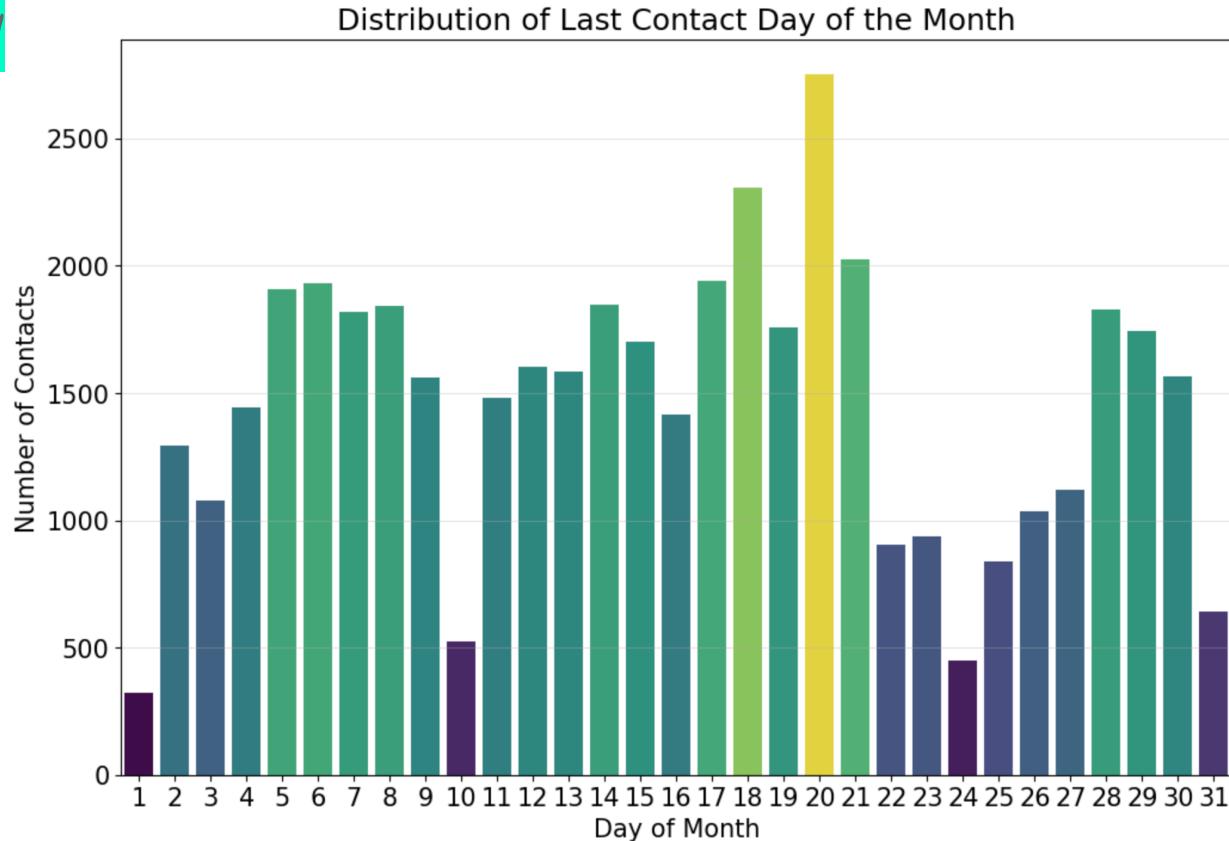


CHART 11: CONTACTS VS CONTACT MONTH

May witnessed the most client contacts in the campaign, followed by July and August, while December had the fewest.

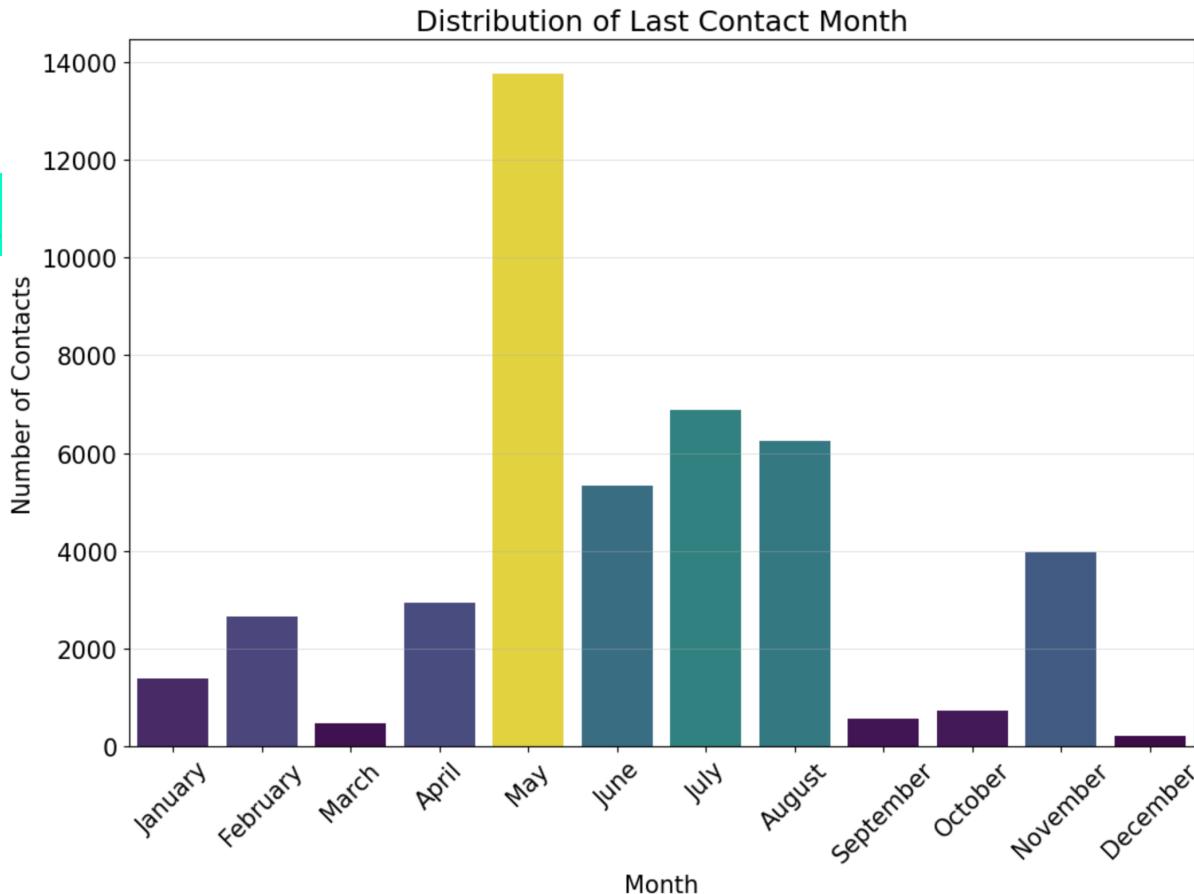


CHART 12: CONTACTS VS LAST CONTACT DURATION

The average last contact duration is 258 seconds, with a median of 180 seconds, ranging from 0 to 4918 seconds. The data shows a positive skew, meaning most contacts are short, with a few exceptionally long ones.

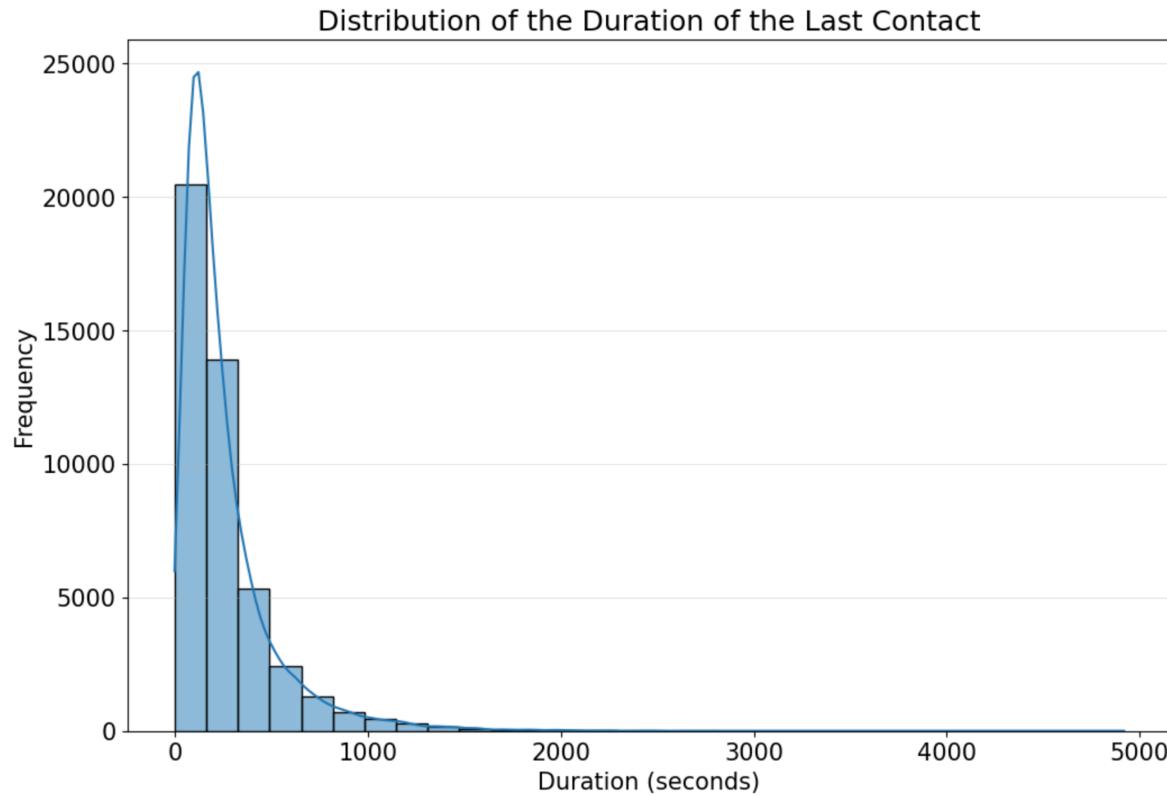


CHART 13: CLIENTS VS NO. OF CONTACTS

On average, clients received 2.8 contacts, with a median of 2, and contacts ranging from 1 to 63. The distribution is positively skewed, showing most clients had few contacts, while a small group experienced many.

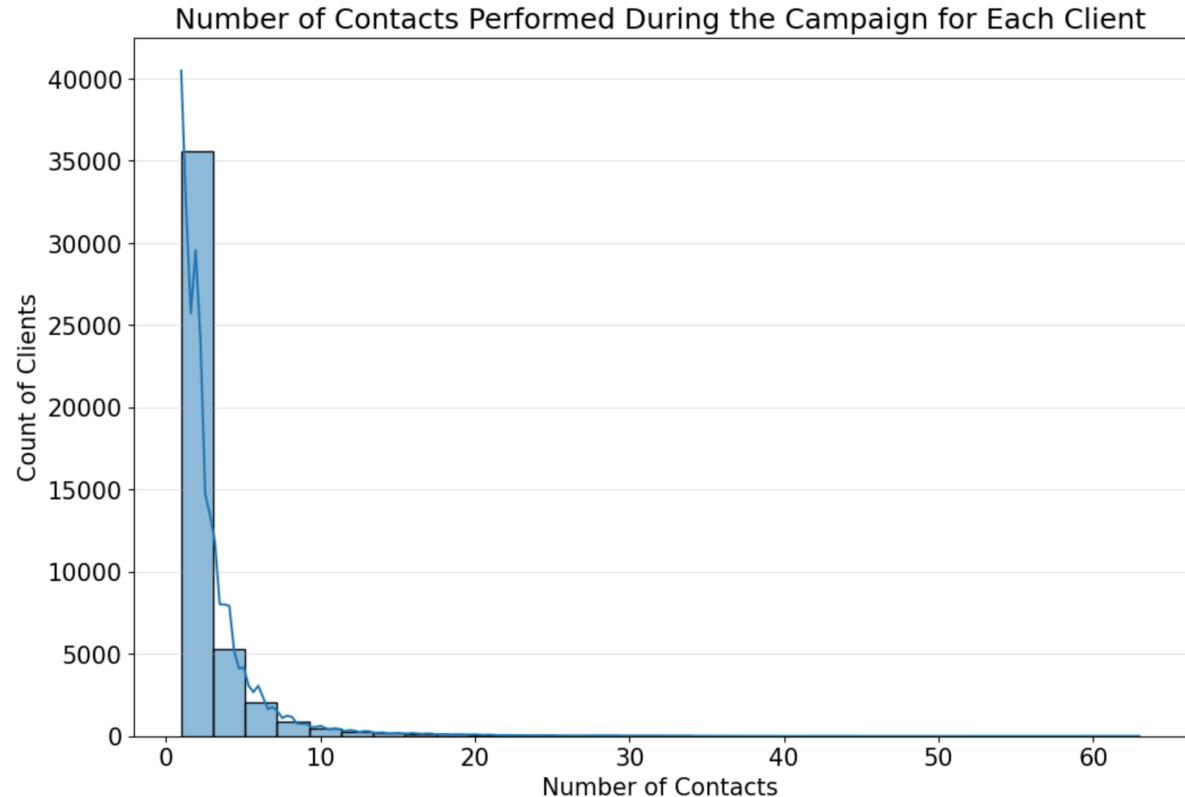


CHART 14: LAST CONTACT DURATION FREQ.

Most clients had no prior contact before the current campaign, shown by a peak at -1. The average days since last contact is 40, with a range from -1 to 871 days, and systematic follow-ups occur every 50 to 100 days.

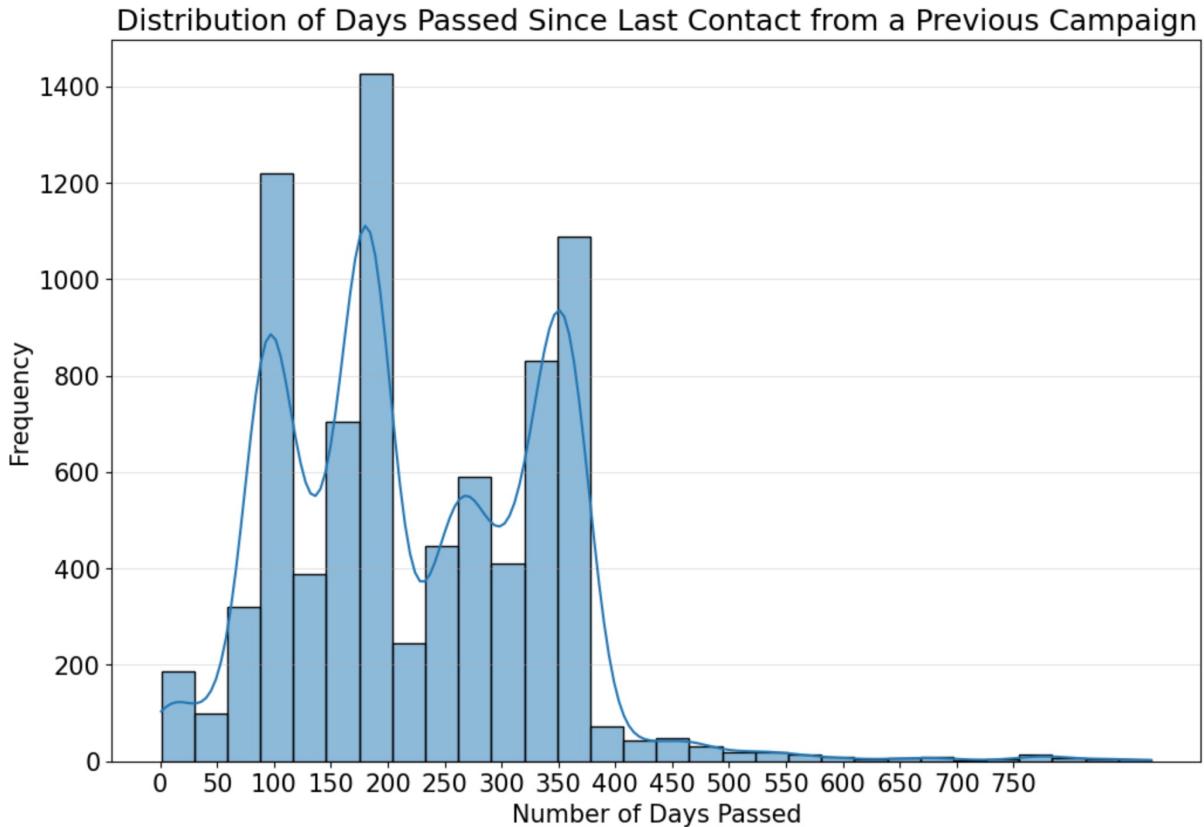
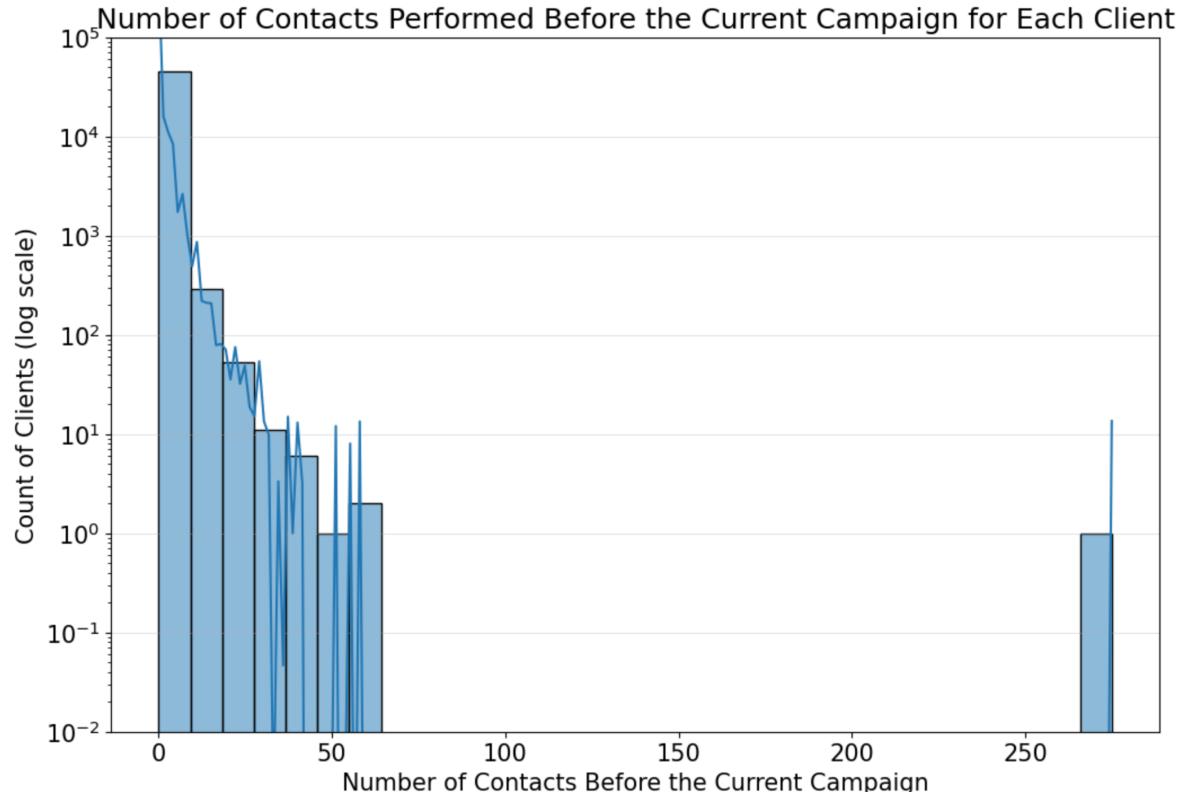


CHART 15: NUM. OF PREVIOUS CONTACTS

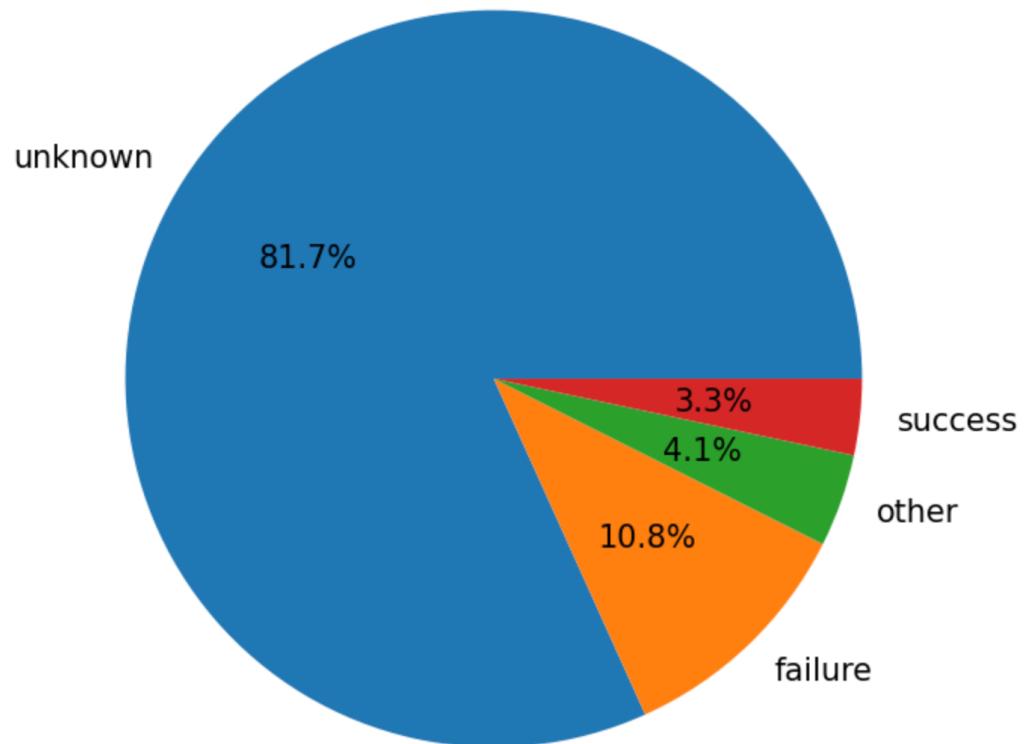
A significant peak at zero indicates many clients had no contacts before the current campaign. The average number of prior contacts is roughly 0.58, with most clients having none, and a range from 0 to 275 contacts for some clients.



Outcomes of Previous Marketing Campaigns

CHART 16: PREV. CAMPAIGN OUTCOME

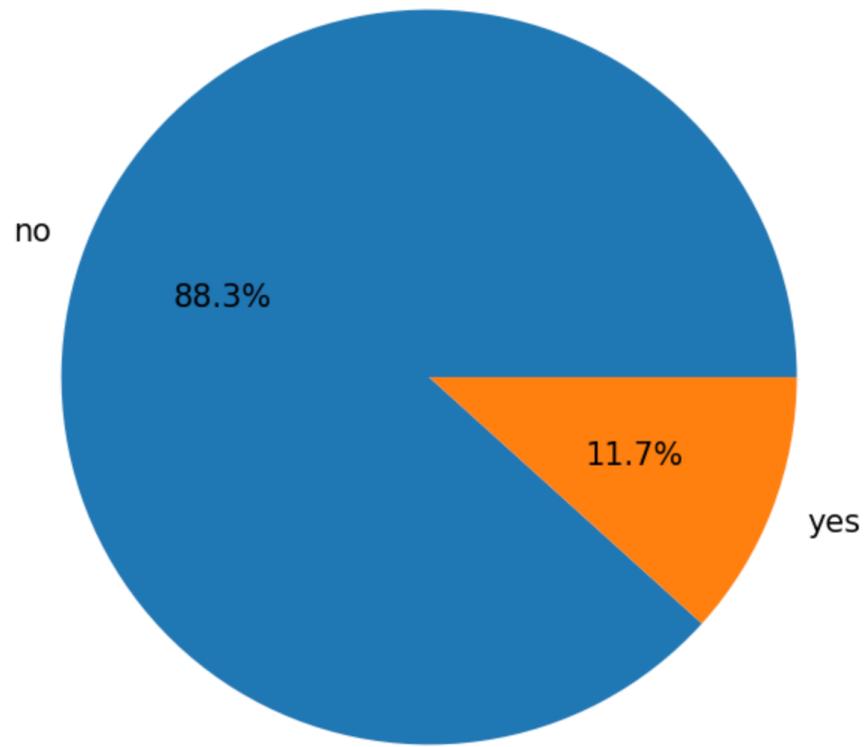
The majority of previous campaign outcomes are unknown, accounting for 81.7% of cases. Failures make up 10.8%, successes are at 3.3%, and other outcomes represent 4.1%.



Distribution of Clients Subscribed to a Term Deposit

CHART 17: CLIENTS VS TERM DEPOSIT

Among the clients approached, 88.3% did not subscribe to a term deposit, while the remaining 11.7% did, indicating a clear divide in the subscription preferences within the dataset.



ANALYSIS



KEY FINDINGS

1. Subscription rates are **highest** among those under 30 and over 60.
2. Students and retirees are **most likely** to subscribe, while blue-collar workers and entrepreneurs are **least likely**.
3. Singles have the **highest** subscription rates, followed by divorced, with married clients having the **lowest**.
4. Individuals with tertiary education are **more inclined** to subscribe compared to those with primary education.
5. Clients without credit defaults show a **higher** subscription likelihood.
6. Higher average account balances correlate with a **greater** likelihood of subscription.

KEY FINDINGS CONTD...

- 7.Those without housing loans are **more likely** to subscribe than those with them.
- 8.Clients without personal loans have a **higher** chance of subscribing compared to those who have them.
- 9.Cellular phone contact has been linked to a **higher** chance of subscription.
- 10.Contacts in March indicate a **preference for** subscriptions, possibly due to the start of a new financial year.
- 11.Longer call durations are associated with a **higher** likelihood of securing a subscription.

THANK YOU!