**REPORT ON BANKING DS PROJECT**

1. DATASET

There are 45,211 rows and 18 columns and the data is from the time period May 2008 to November 2010.

* Age: Client’s age. Numeric Data.
* Job: Type of job the client has. Categorical.
* Marital: Marital status of the client. Categorical.
* Education: Education Level of the client. Categorical.
* Default: Whether the client has credit in default or not. Categorical. Binary.
* Balance: Average Yearly Balance of the client. Numeric Data.
* Housing: Home Loan, Yes or No.
* Loan: Personal Loan, Yes or No.
* Contact: How the client was contacted.
* Day: Last contacted day
* Month: Last contacted month of the year
* Duration: How long the call lasted. Numeric Data.
* Campaign: No of contact performed during **this** campaign. Numeric Data.
* Pdays: No of days since the client was contacted in the last campaign. Numeric Data.
* Previous: No of contacts performed before this campaign. Numeric Data.
* Poutcome: Outcome of the precious campaign. Categorical Data.

1. METHODOLOGY

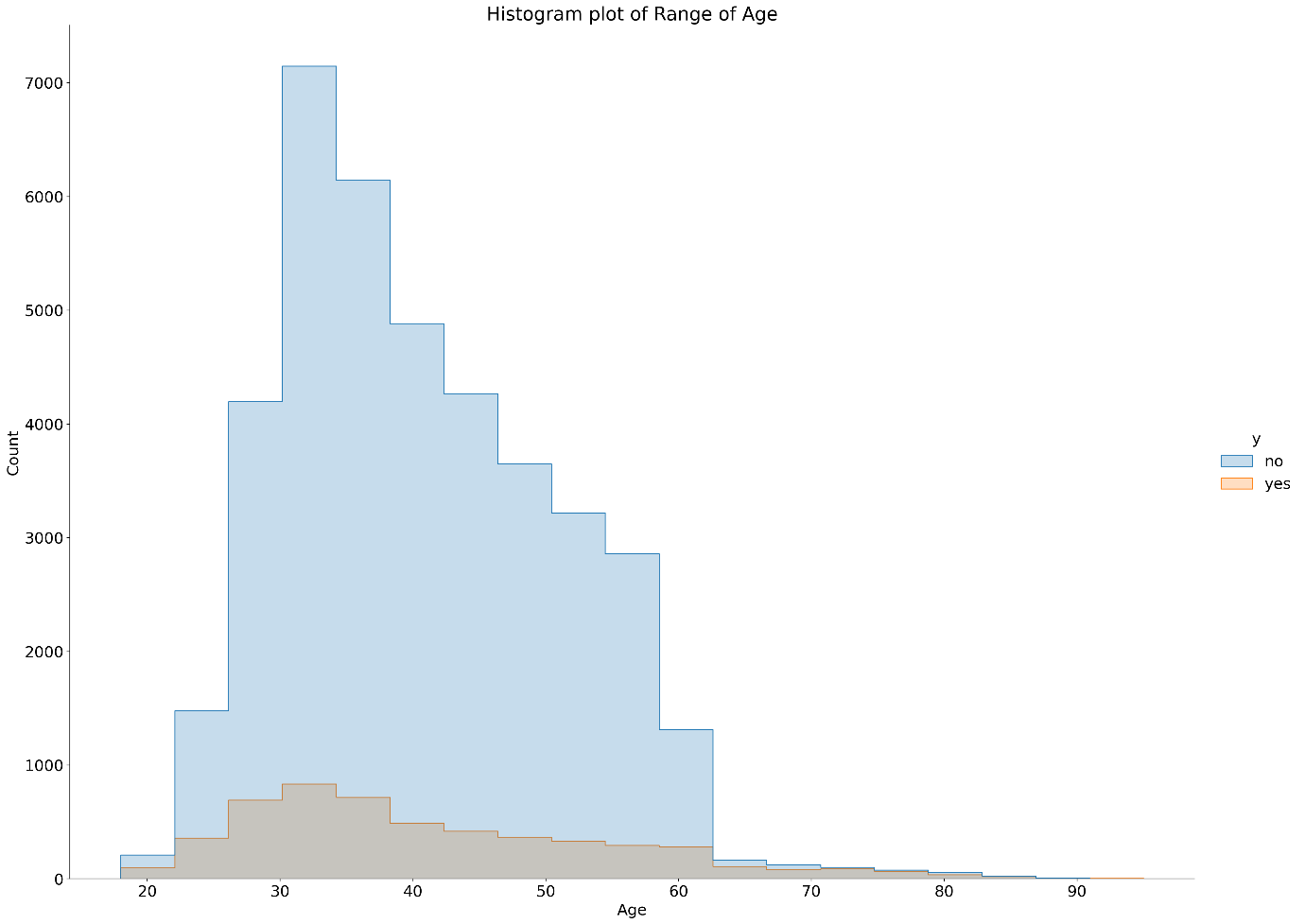
I have a labelled dataset with 17 Independent Variables and one target feature which is whether the person will subscribe to the term deposit or not.

**Step 1: Things of the top of my mind.**

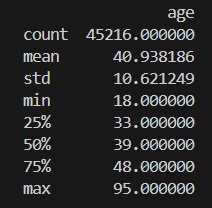
1. I will first find out the range of the Balance column.
2. I will convert all the Categorical Data into Numeric Value.
3. Scale only the originally numeric data and bring it together with the categorical data.

**Step 2: Following Instructions**

1. **What is the distribution of age among the clients?**

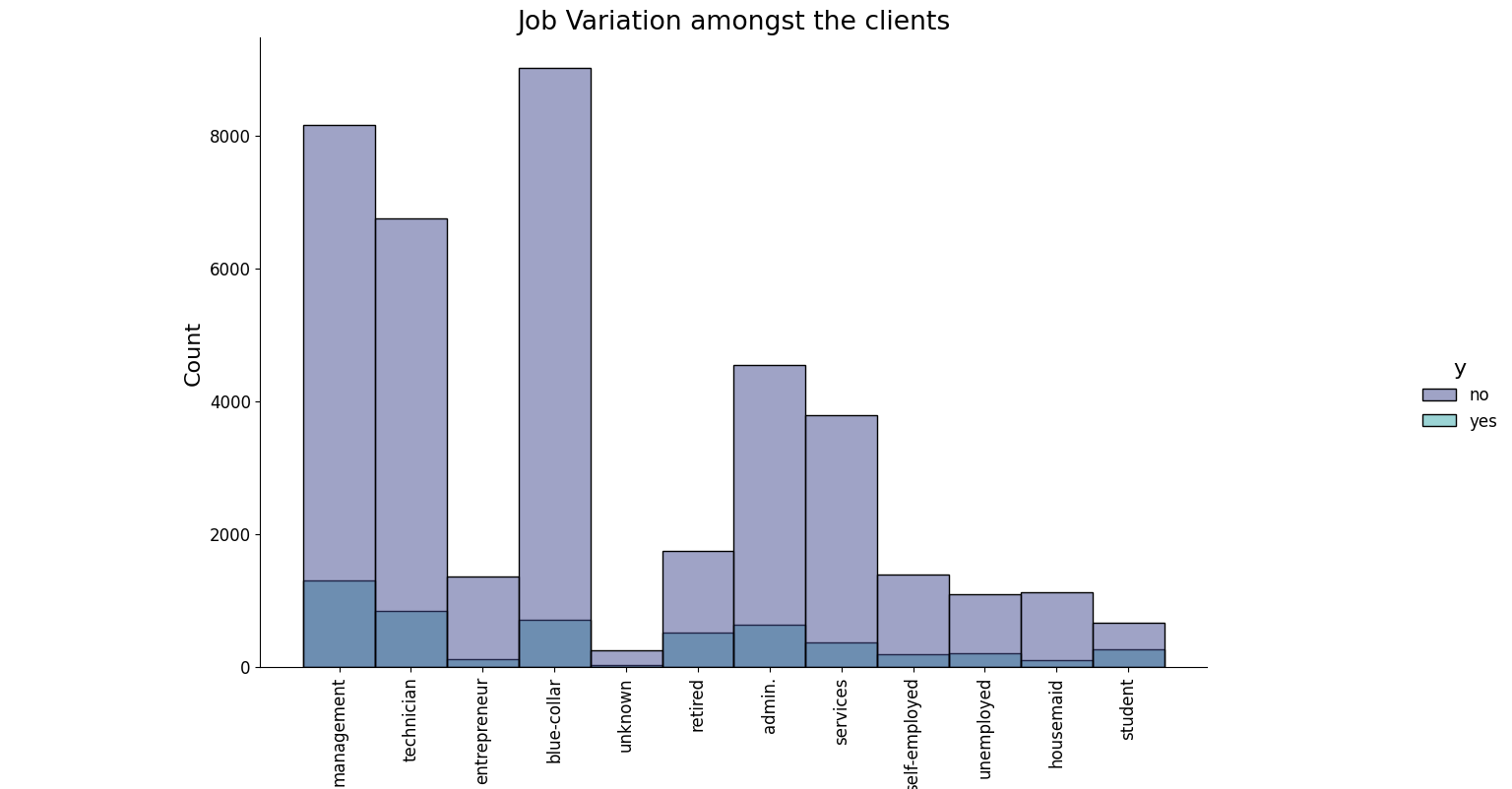


From the graph above, we can see that the age of client varies from 20 yrs to 90 yrs. The most row count is found in the range of 30yrs to 35yrs.

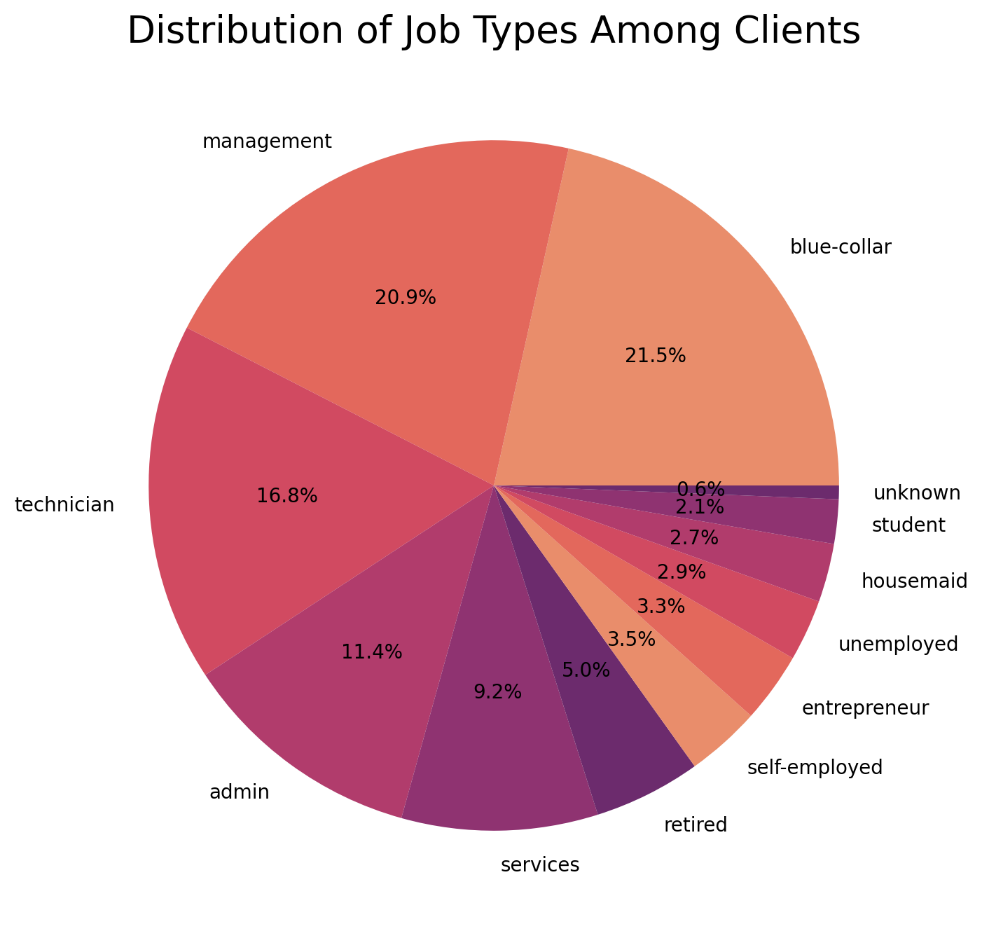
 The actual min value of age is 18 and the max value is 95.

1. **How does the job type vary among the clients?**

As clearly evident from the graph, most of the clients have a job in blue-collar domain, followed by management and technician.

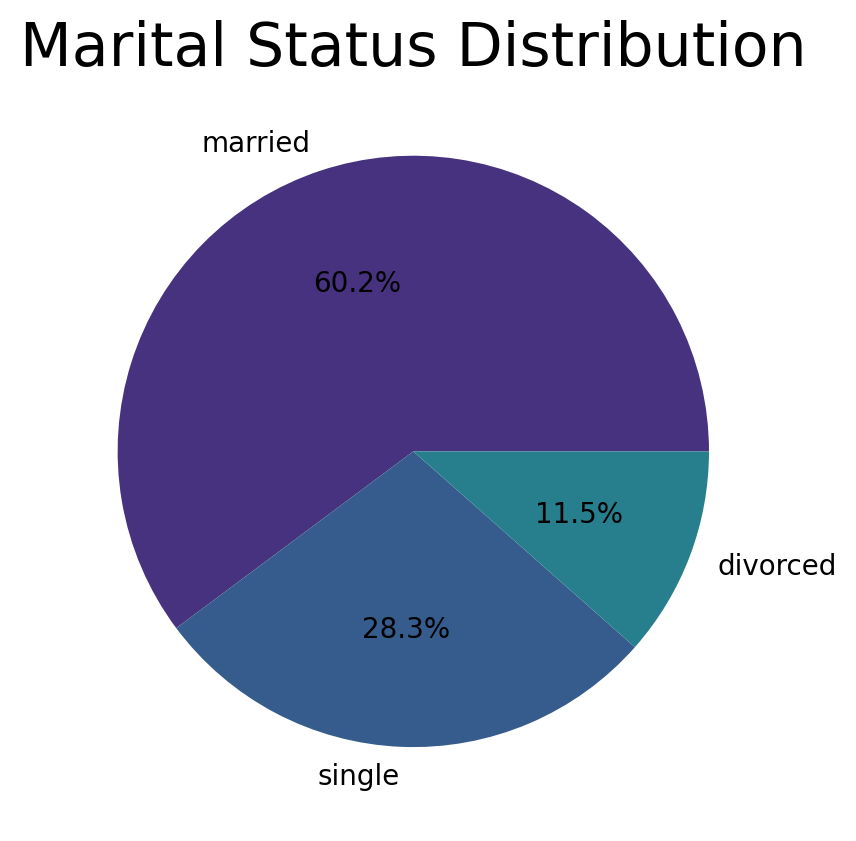
****

To get the exact distribution of various job, I plotted a pie chart that depicts the percentage of clients holding each kind of job.

****

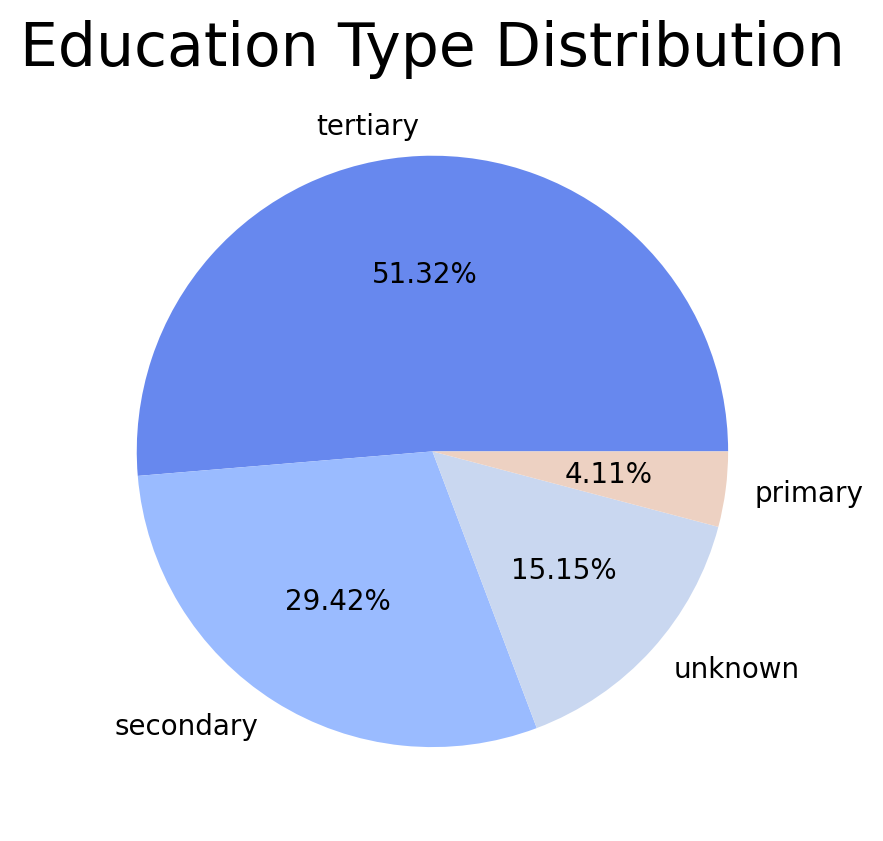
1. **What is the marital status distribution of the clients?**

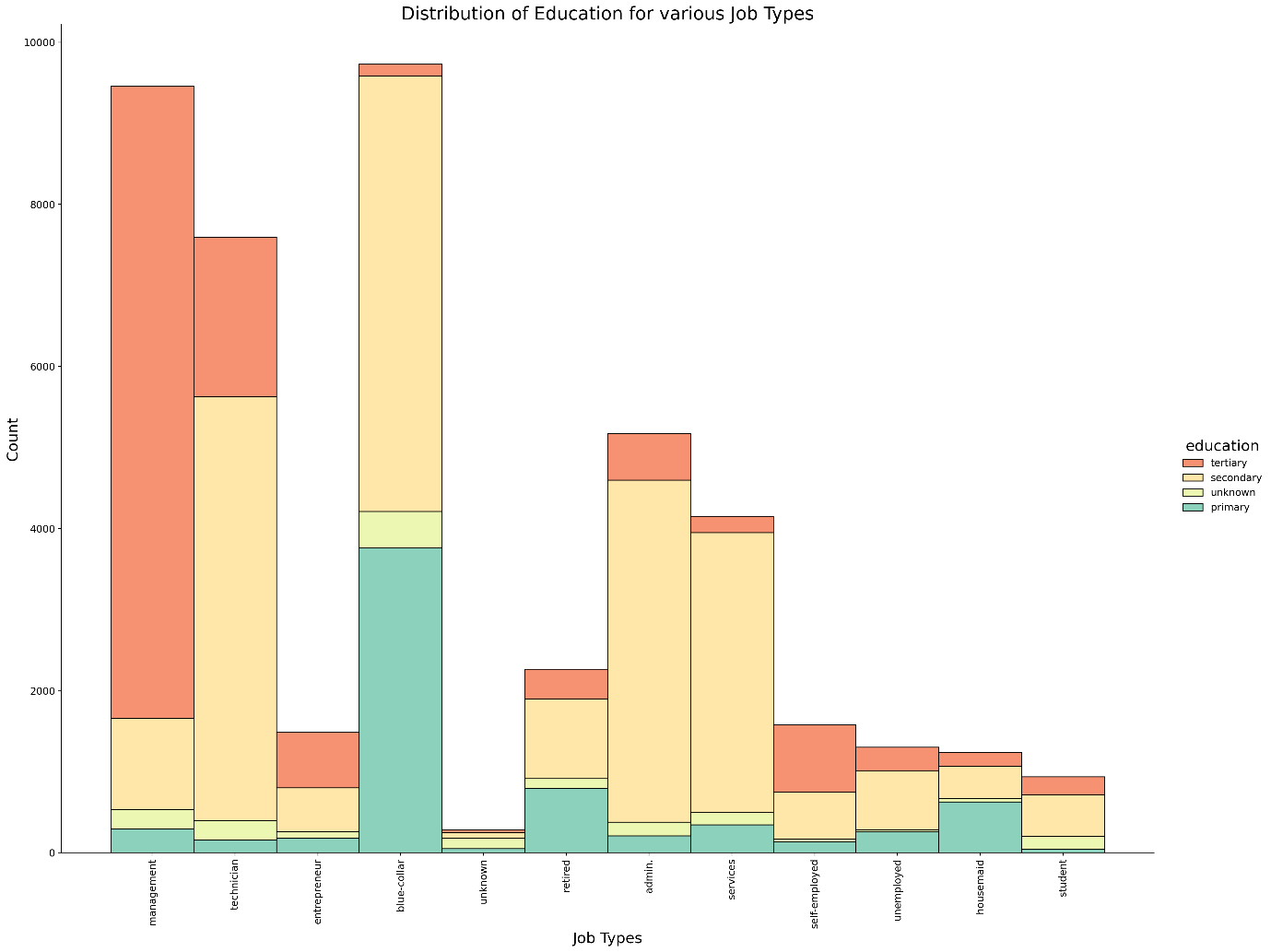
Amongst all the clients, 60% are married, 11.5% are divorced and 28.3% are single.

****

1. **What is the level of education among the clients?**

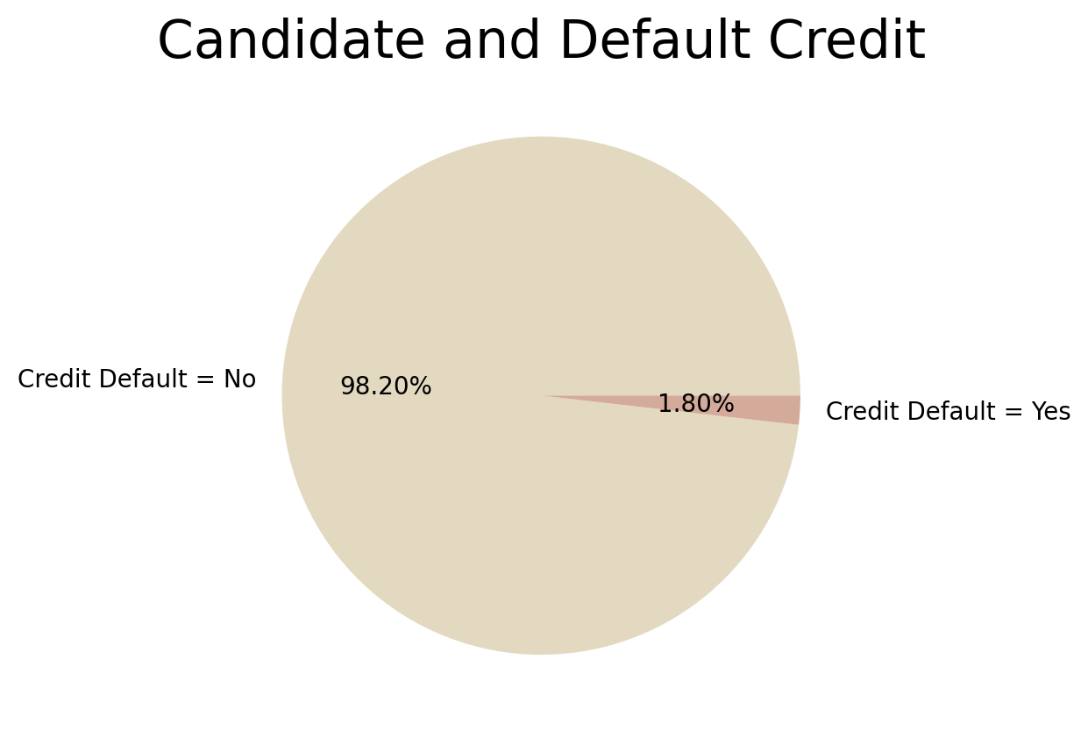
Amongst all the clients, 4.11% clients have primary education, 30% approx. have secondary education, 51.32% have tertiary education and the rest we have no knowledge.

****

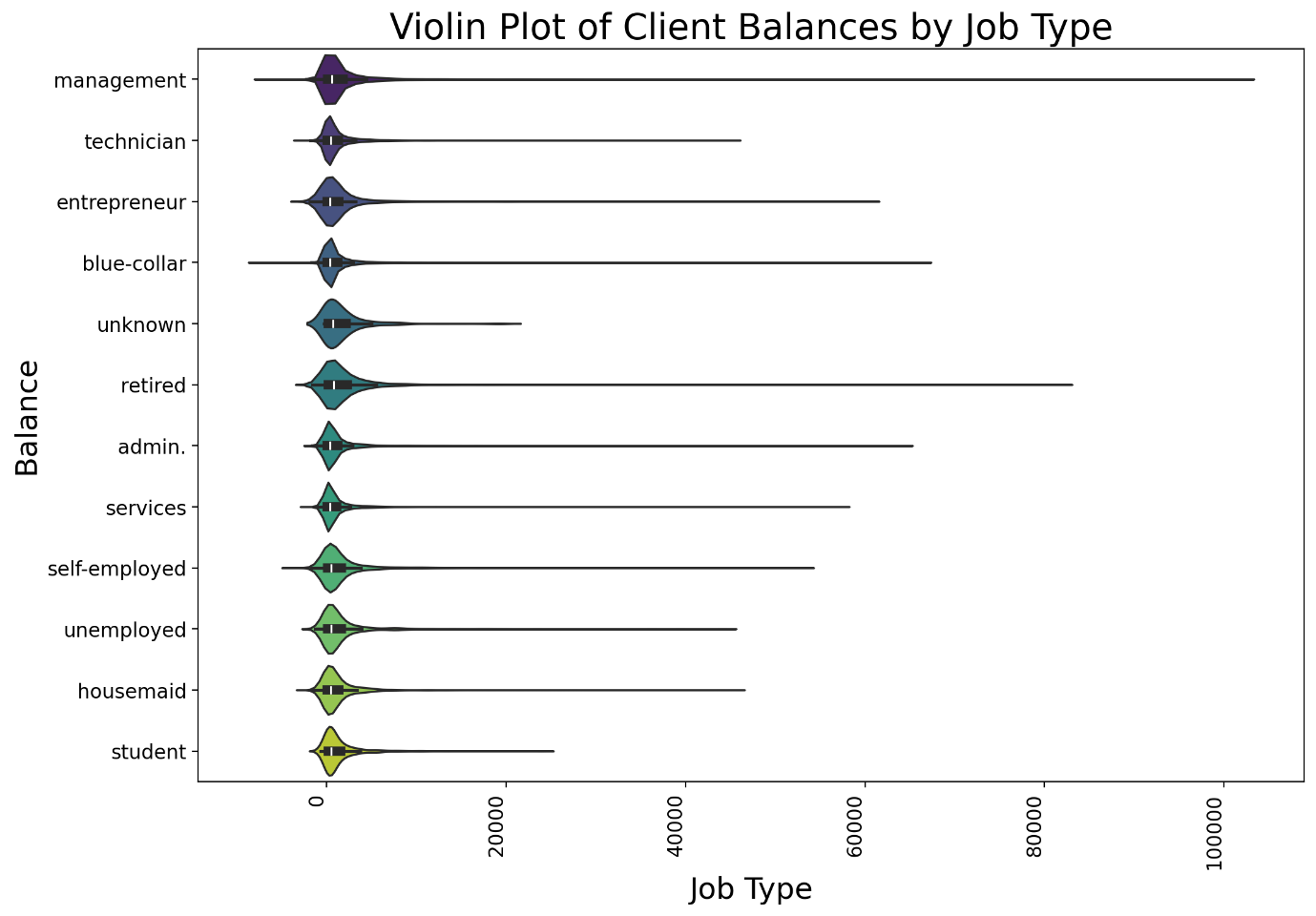
****

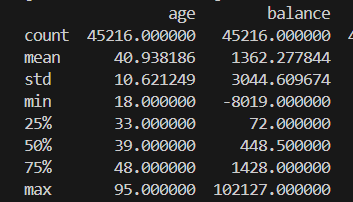
1. **What proportion of clients have credit in default?**

We can clearly infer from the pie chart that 1.8% of clients have credit by default.

****

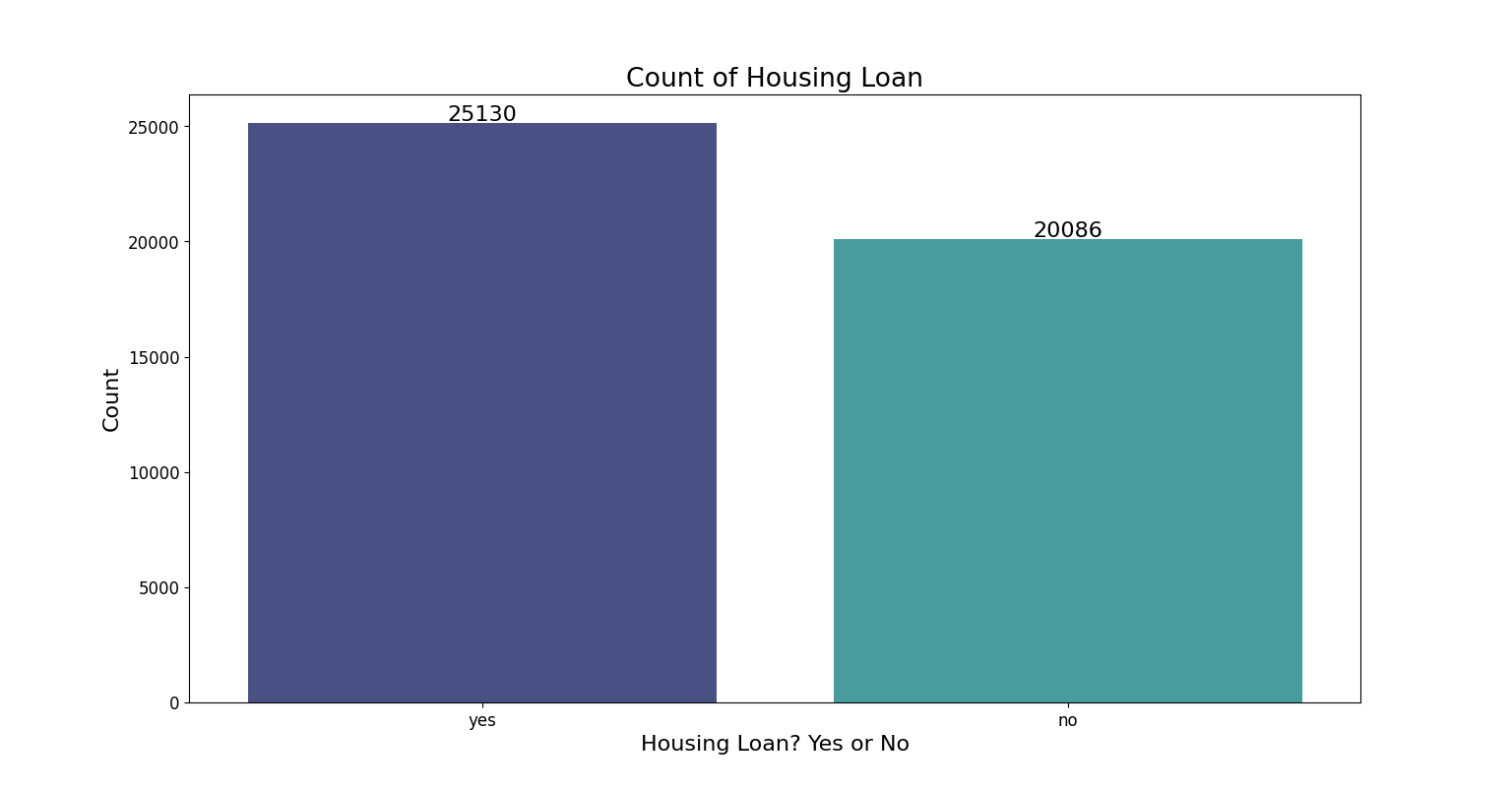
1. **What is the distribution of average yearly balance among the clients?**

****

****

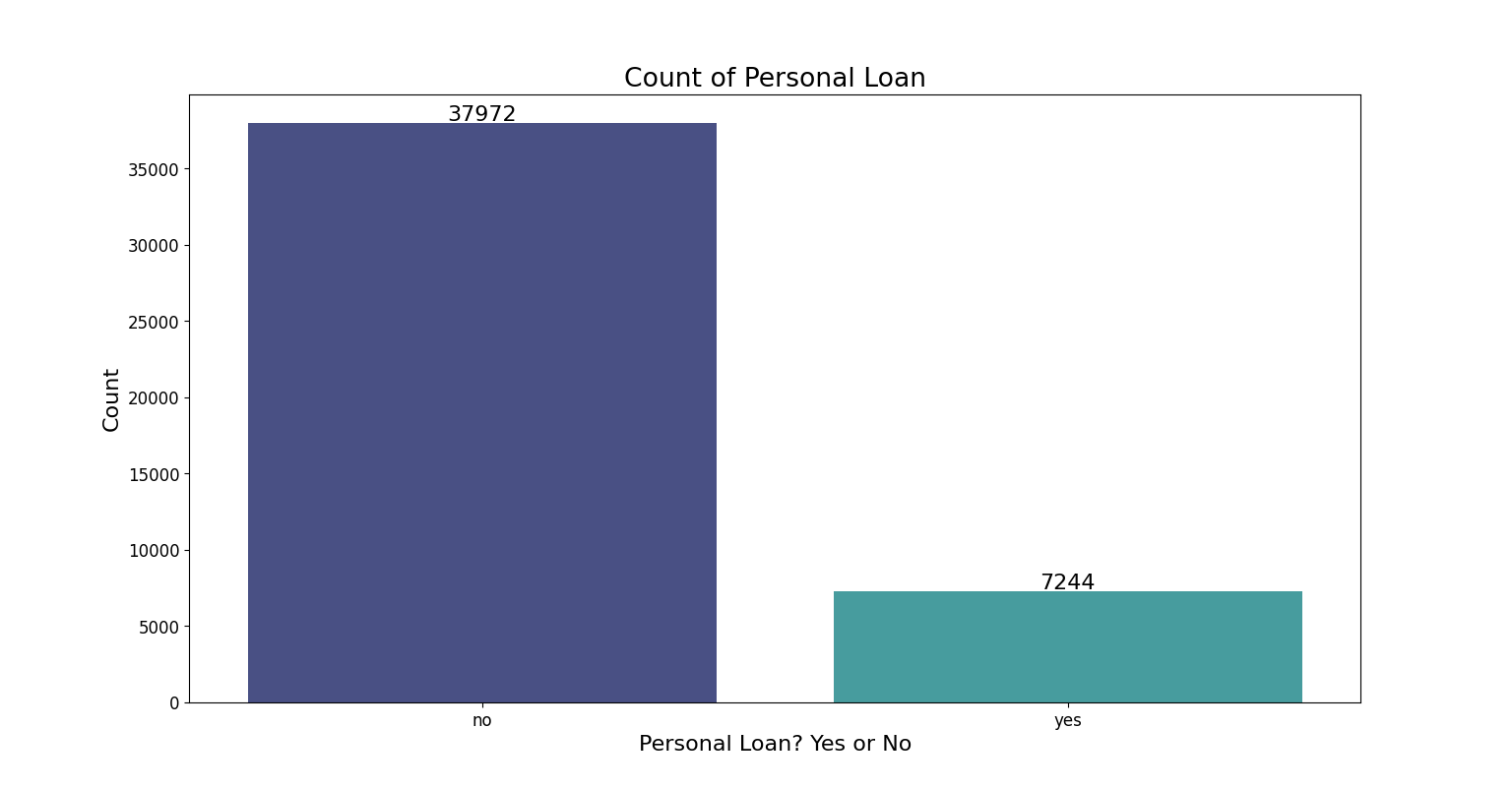
As you can gather from the graph and the statistics, the average yearly balance ranges from -8019 to 102127. To understand the distribution and trend more I also grouped the yearly balance using job type.

1. **How many clients have housing loans?**

****

As clear from the graph, 25130 clients have housing loans

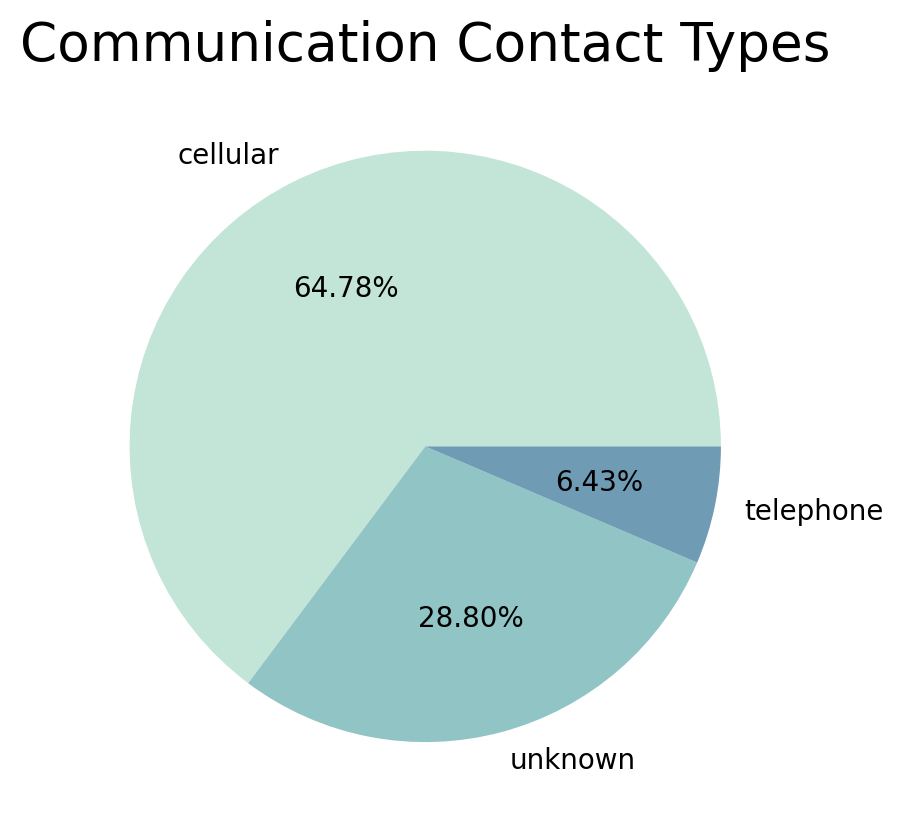
1. **How many clients have personal loans?**

****

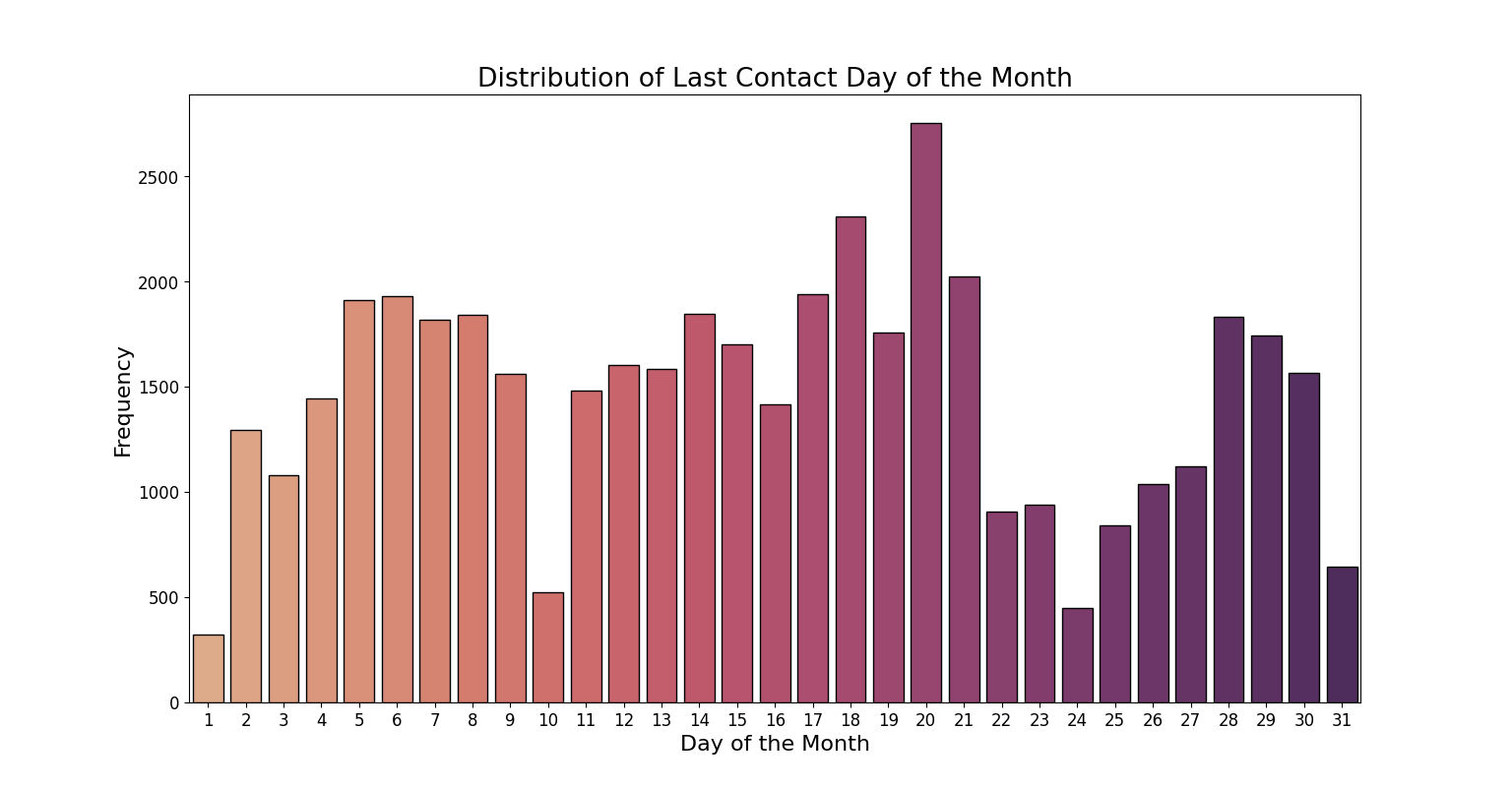
As clear from the graph, 7244 clients have personal loan

1. **What are the communication types used for contacting clients during the campaign?**

There are broadly two main types of communication, cellular and telephone.

****

1. **What is the distribution of the last contact day of the month?**

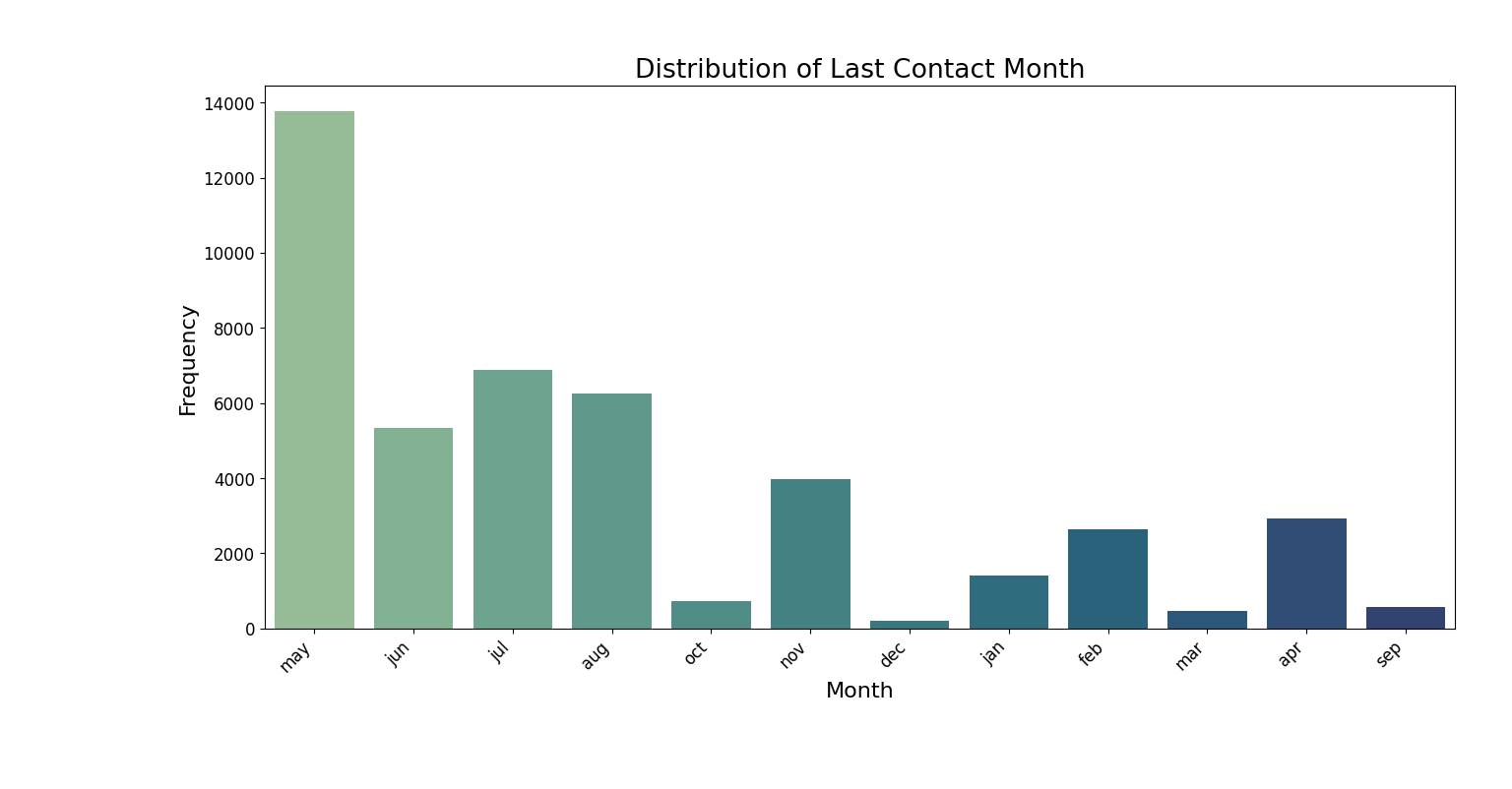
****

The clients were contacted the most on the 20th of the month.

The approximate range of dates is from 14th to 20th.

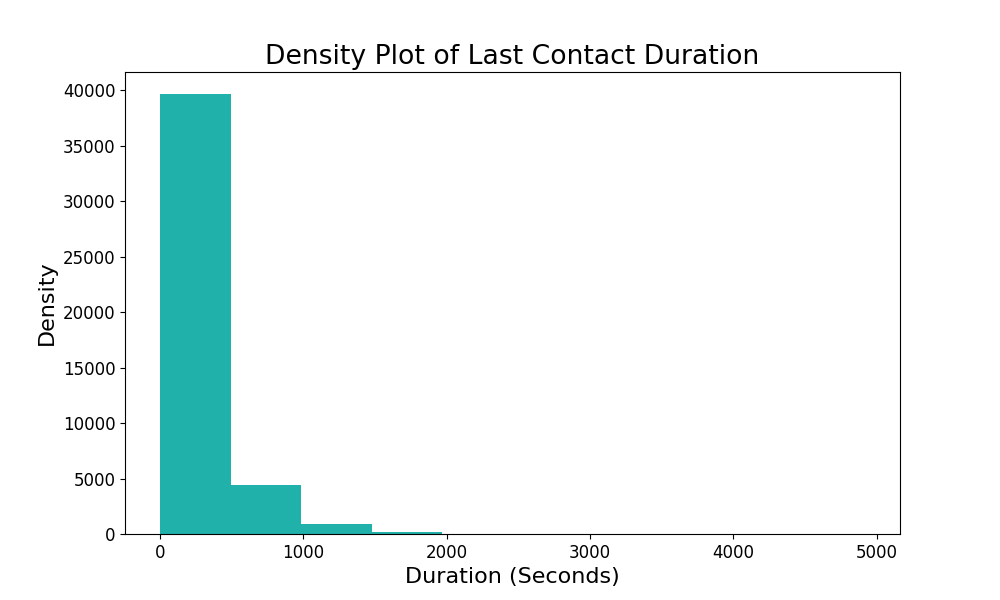
1. **How does the last contact month vary among the clients?**

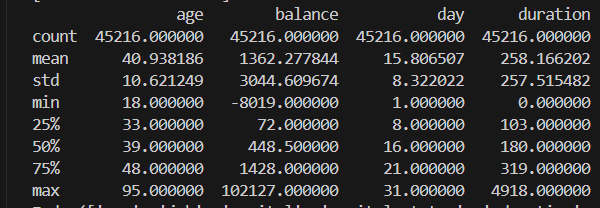
Most of the clients were contacted in the month of May.

****

1. **What is the distribution of the duration of the last contact?**

Majority of clients had a duration of call ranging between 0 to 500secs.

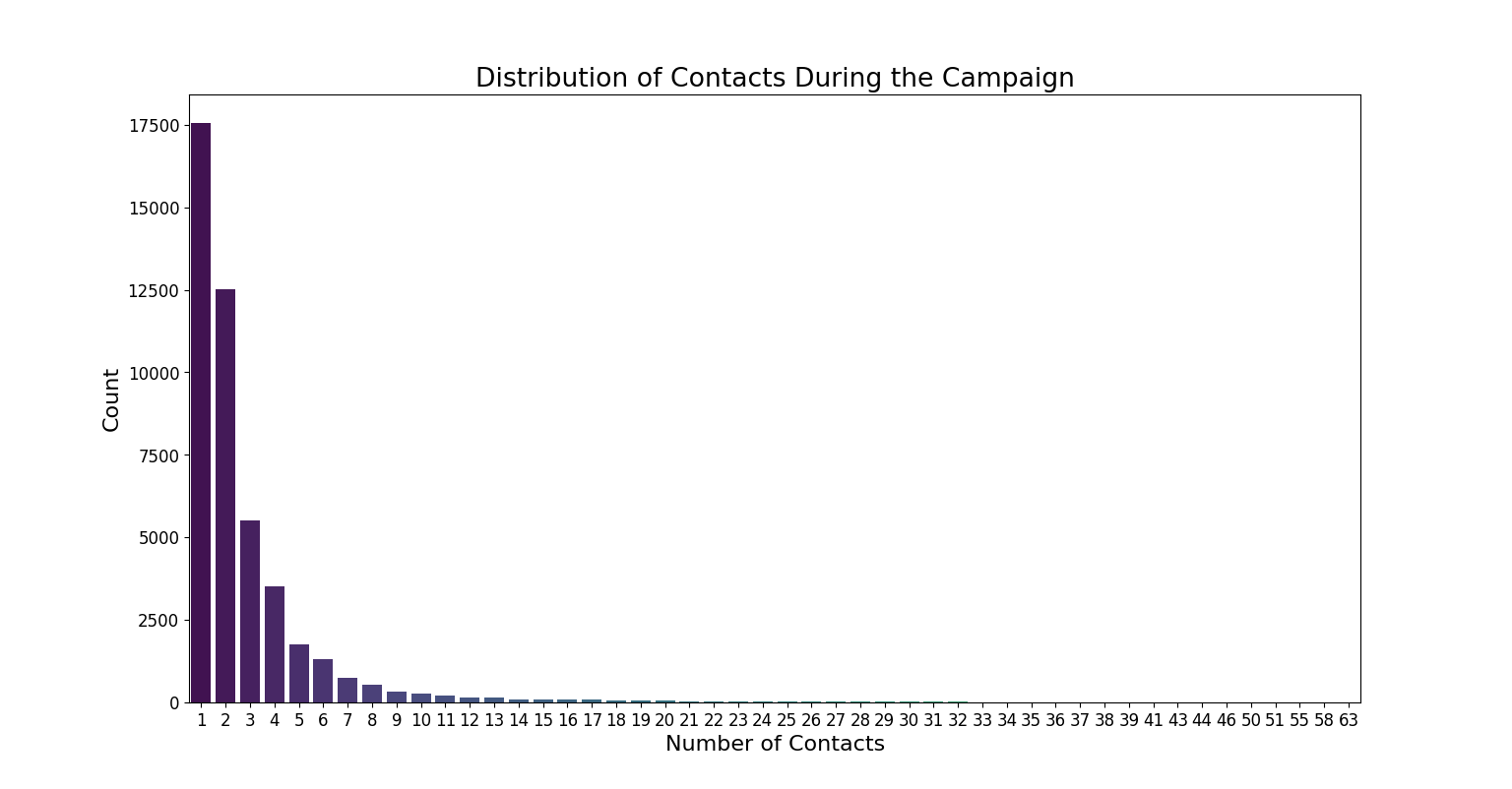
****

****

The maximum duration of call is 4918sec and the average call duration is 258sec.

1. **How many contacts were performed during the campaign for each client?**

Most of the clients were contacted only once during the campaign. The exact number is 17548 clients. They were contacted only once.

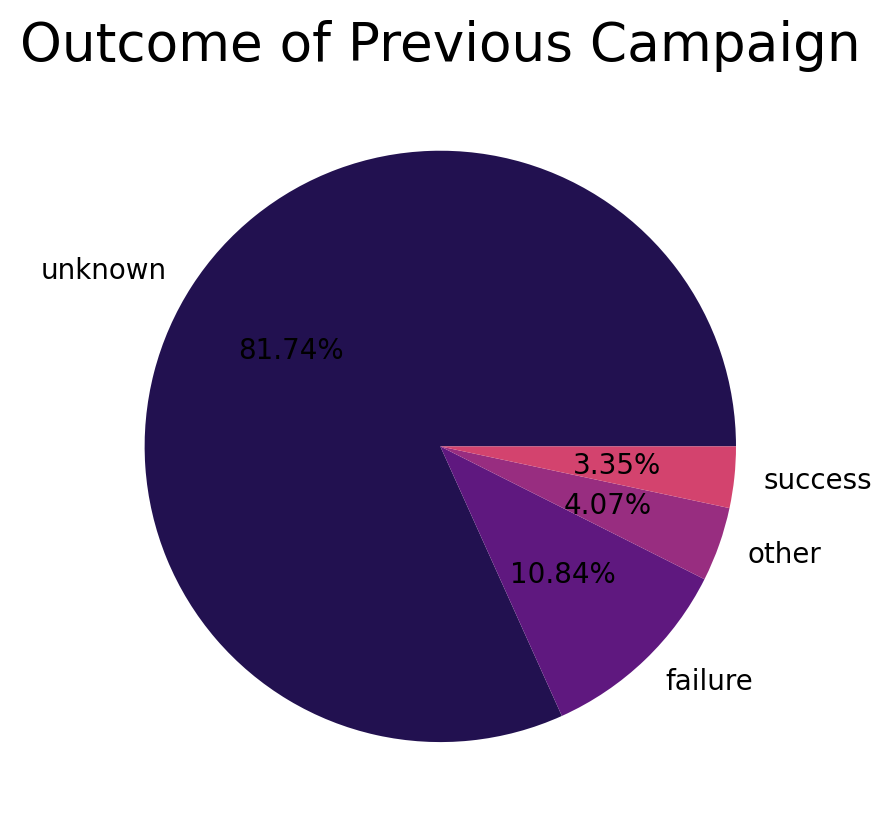
****

In the table given below, the number on the left side depicts the number of contacts and the number on right depicts the no of clients. These are the exact no of contacts that happened for how many clients.

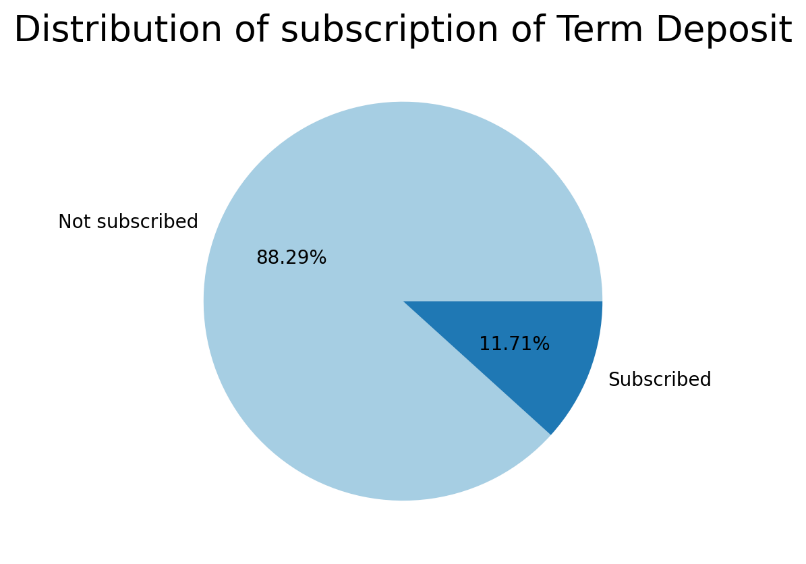
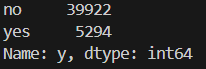
|  |  |  |  |
| --- | --- | --- | --- |
| 1 17548 | 2 12506 | 3 5521 | 4 3522 |
| 5 1764 | 6 1291 | 7 735 | 8 540 |
| 9 327 | 10 266 | 11 201 | 12 155 |
| 13 133 | 14 93 | 15 84 | 16 79 |
| 17 69 | 18 51 | 19 44 | 20 43 |
| 21 35 | 22 23 | 25 22 | 23 22 |
| 24 20 | 29 16 | 28 16 | 26 13 |
| 31 12 | 27 10 | 32 9 | 30 8 |
| 33 6 | 34 5 | 36 4 | 35 4 |
| 43 3 | 38 3 | 37 2 | 50 2 |
| 41 2 | 46 1 | 58 1 | 55 1 |
| 63 1 | 51 1 | 39 1 | 1. 1 |

1. **What were the outcomes of the previous marketing campaigns?**

Most of the contacts in the previous campaign gave no clear results with the percentage being 81.74%. The success rate was as low as 3.35% and the failure rate was significantly high at almost 11%

****

1. **What is the distribution of clients who subscribed to a term deposit vs. those who did not?**



Amongst 45216, only 5294 subscribed to the term deposit, which is approximately 12% of the total clientele.

1. **Are there any correlations between different attributes and the likelihood of subscribing to a term deposit?**

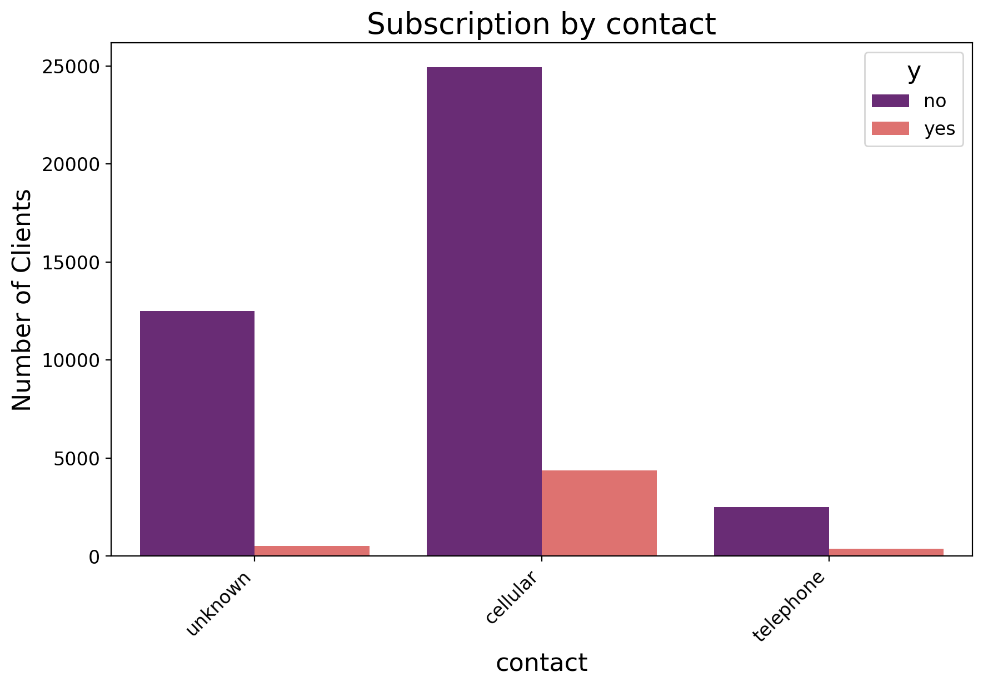
Since the features here are all categorical by nature to understand and analyze the relation between different attributes and the likelihood of subscribing to a term deposit, we will use **visualizing techniques.**

I will plot count plot for different features and group them by using the target variable using the ‘hue’ parameter to assess the likelihood.

1. Contact Type

This count plot is for various contact methods utilized to contact the client and convince them to subscribe to the term deposit.

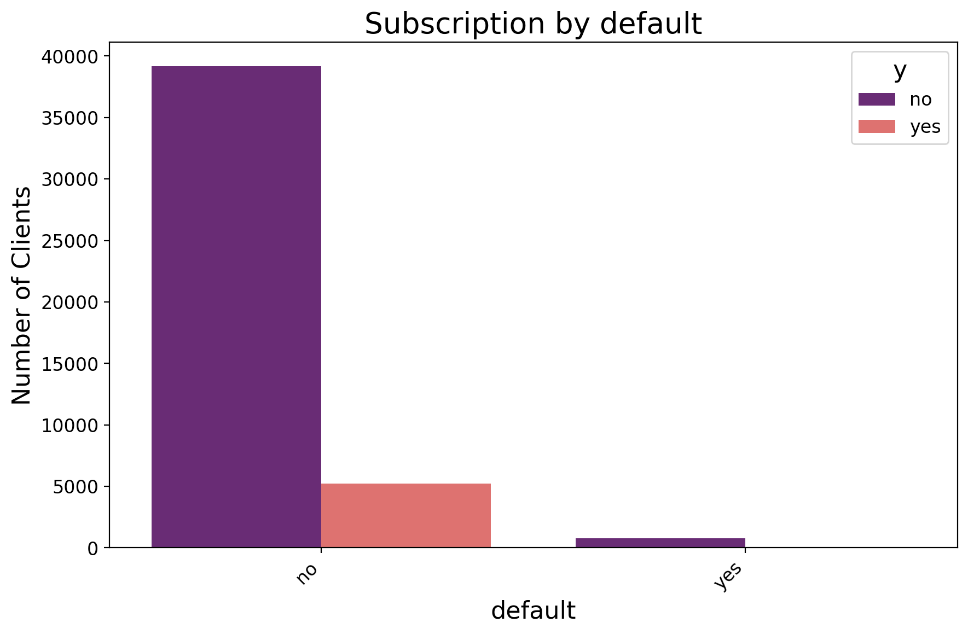
Clients that were contacted directly through their cell phones were more likely to subscribe to the term deposit than any other means.



1. Credit in default

This count plot is plotted for people who had credit in default and those who didn’t.

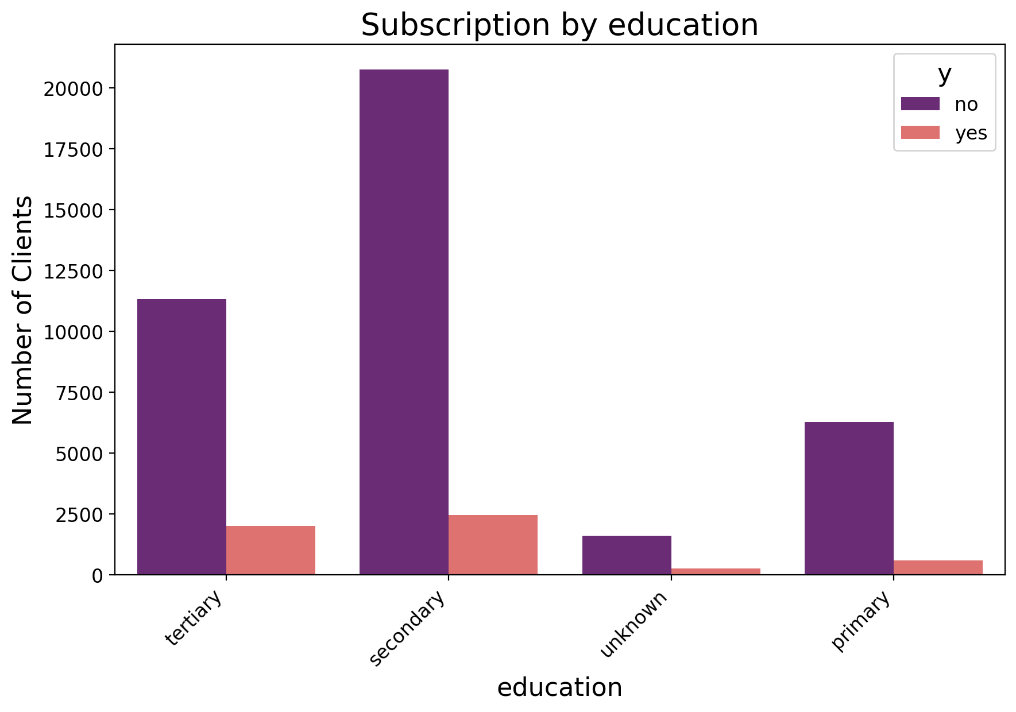
Clients that who had no credit by default were more likely to subscribe to the term deposit.



1. Education

This count plot represents the highest education level of clients and how that affected the subscription of term deposit.

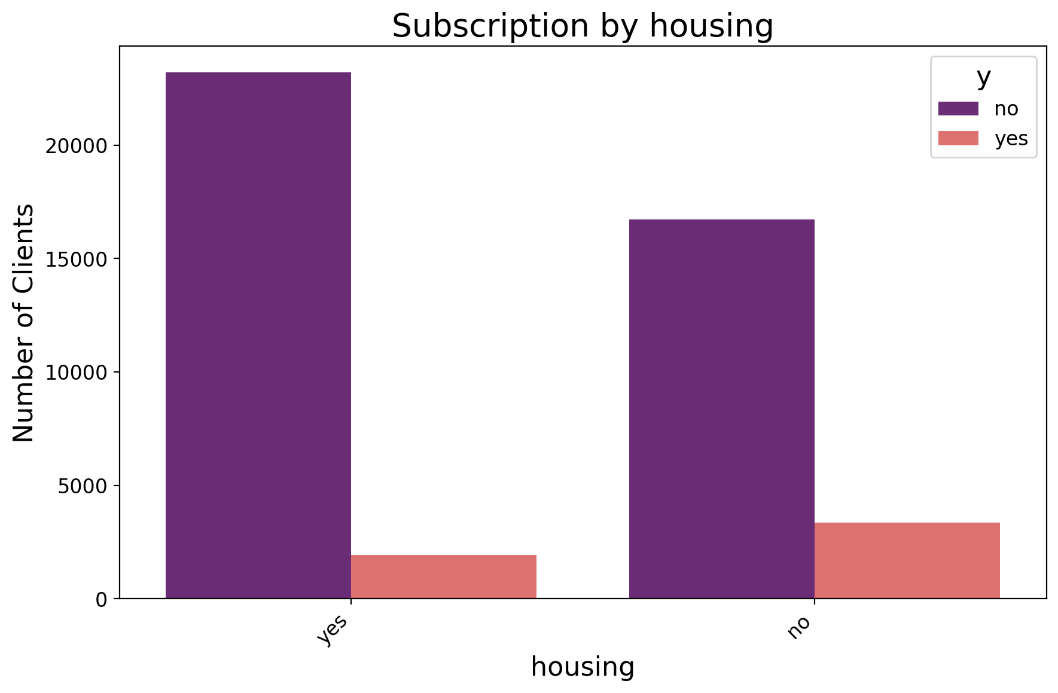
Clients whose highest education level was secondary were more likely to subscribe to the term deposit followed by tertiary education. Primary education clients were reluctant.



1. Housing Loan

This count plot captures the term deposit holders who had housing loans or not.

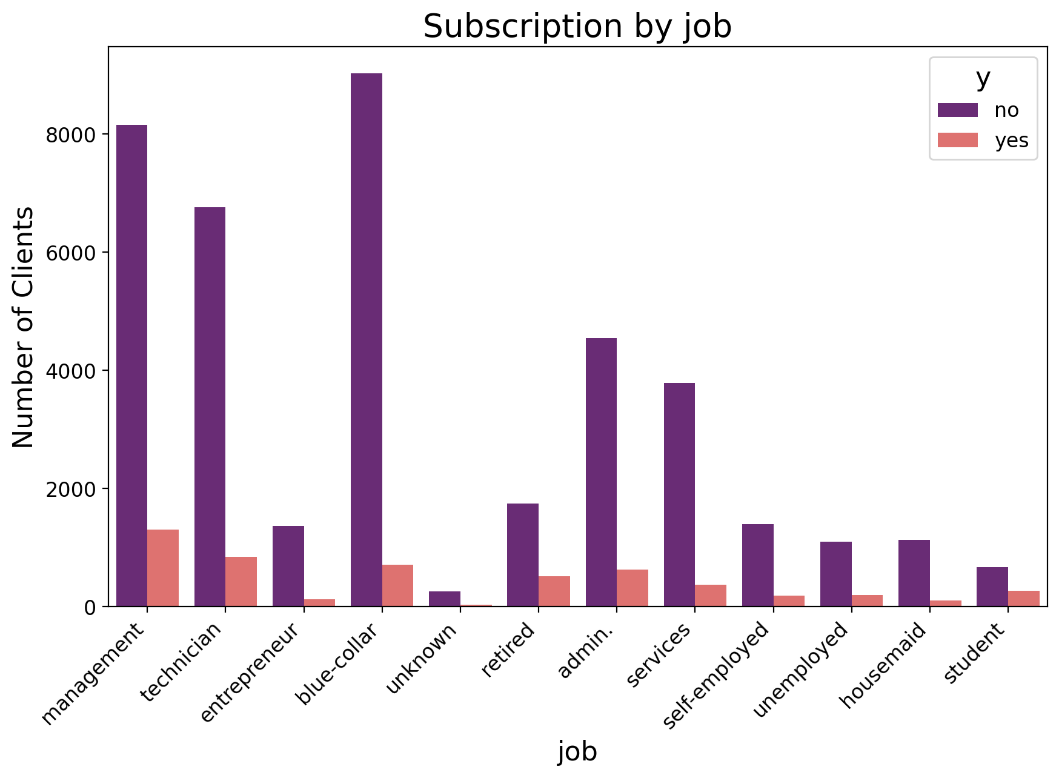
Clients who had no housing loan at the time of call were more likely to subscribe to the term deposit.



1. Job Type

This count plot gives the term deposit holders across various job types.

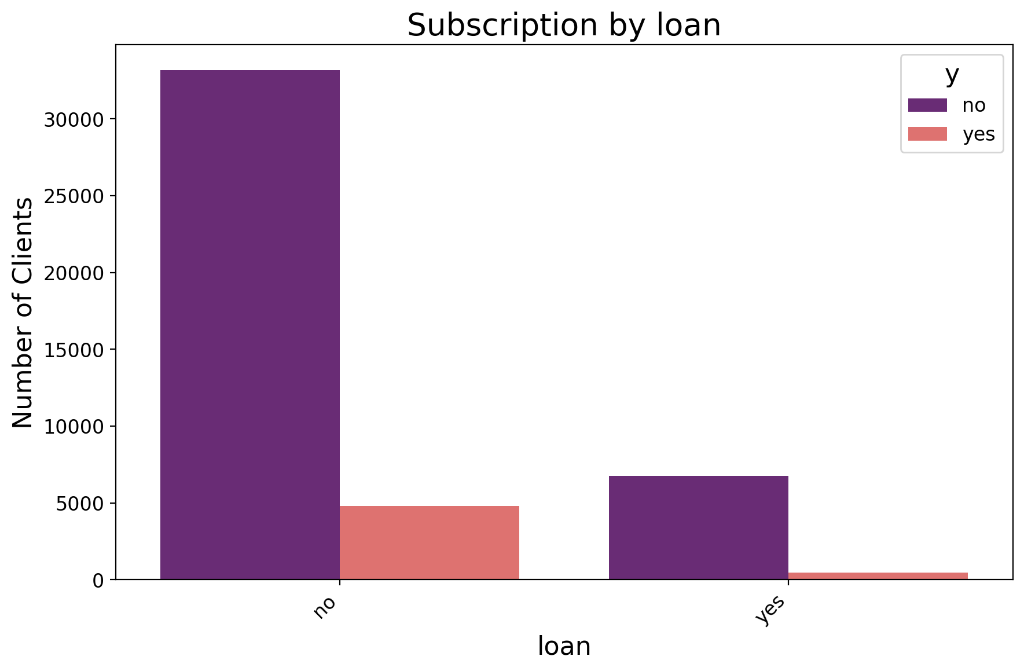
From the graph it is clear that the clients that had a job in management sector was more likely to subscribe to a term deposit followed by technician and then admin.



1. Personal Loan Holder

This count plot represents the data for clients that had a personal loan and clients that didn’t.

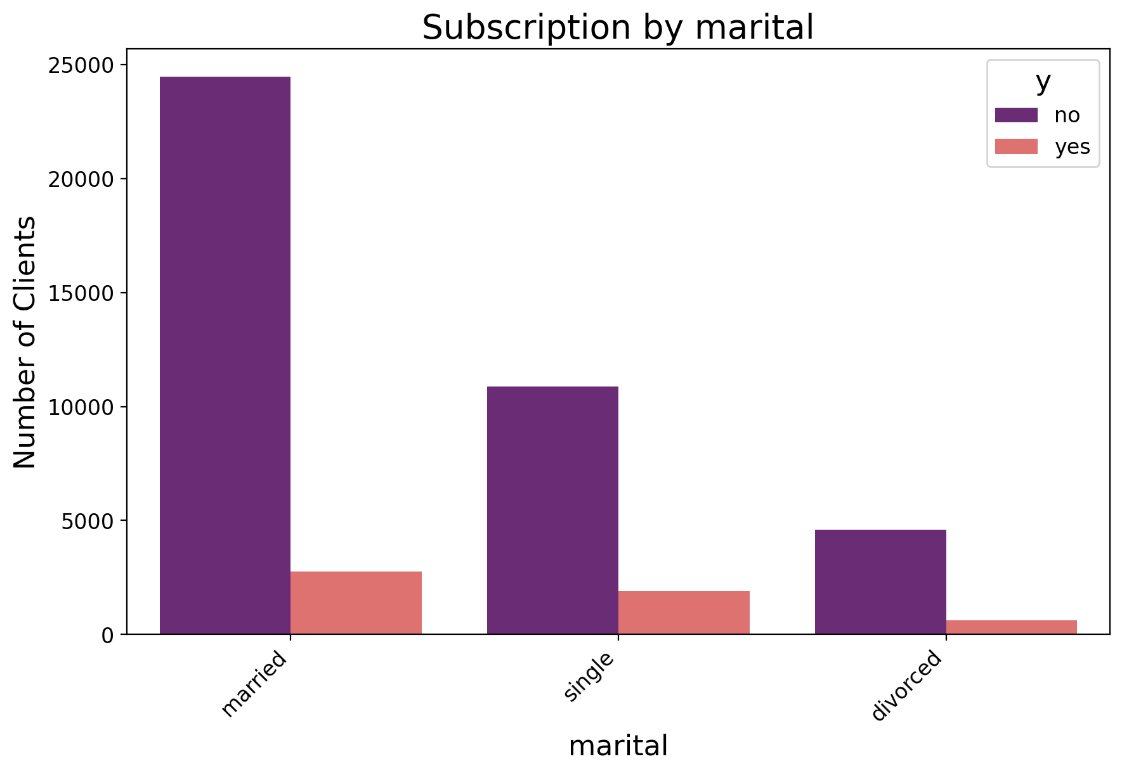
From the graph it is evident that clients who had no personal loan were more likely to subscribe to a term deposit as compared to the ones who had a personal loan to their name.



1. Marital Status

This particular count plot shows the stats for term deposit holders grouped by their marital status.

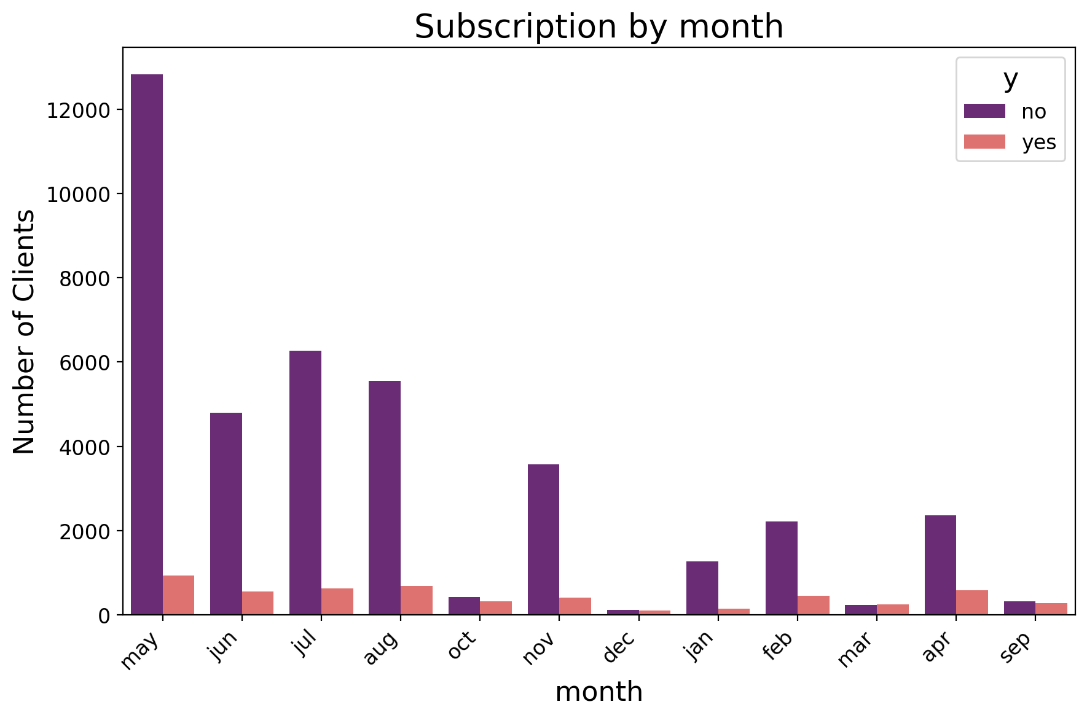
Clients who were married were more likely to subscribe to the term deposit.



1. Month in which the call was made

This count plot has the client’s subscription divided by the month in which the client was last contacted.

Clients who were contacted in may were more likely to subscribe to the term deposit than any other month.



1. Previous campaign’s outcome

This count plot counts the subscription count based on the outcome of previous campaign, i.e. whether the client subscribed to the term deposit when contacted in last campaign or not.

Clients that had no outcome in the last campaign were more likely to subscribe to term deposit in this campaign.

