REPORT ON BANKING DS PROJECT

Whether the client will agree on term deposit or not?

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.

METHODOLOGY

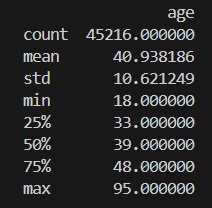
I have a labelled dataset with 17 Independent Variables and one target feature which is whether the person will subscribe to the term deposit or not.

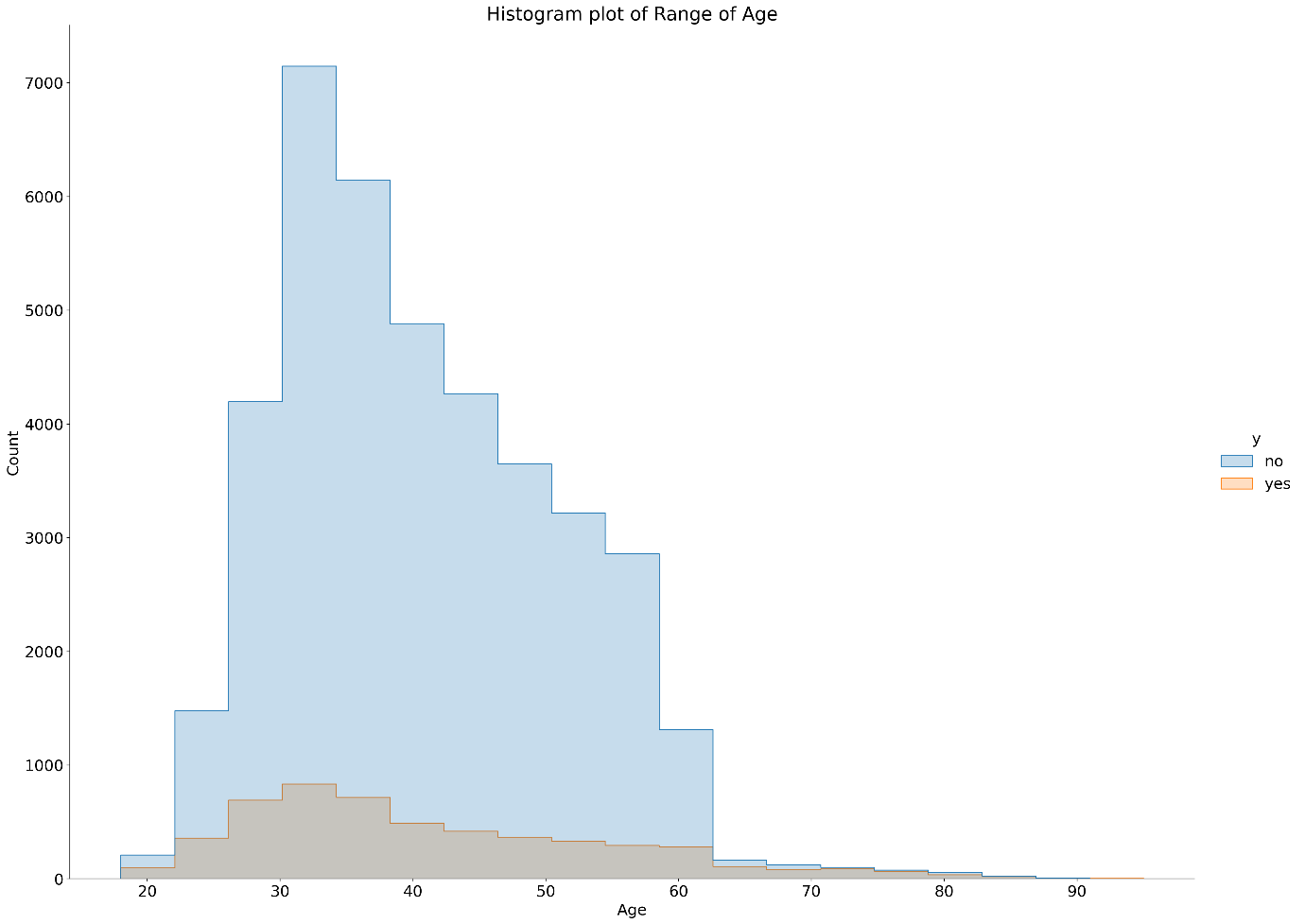
**Step 1: Things of the top of my mind.**

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**

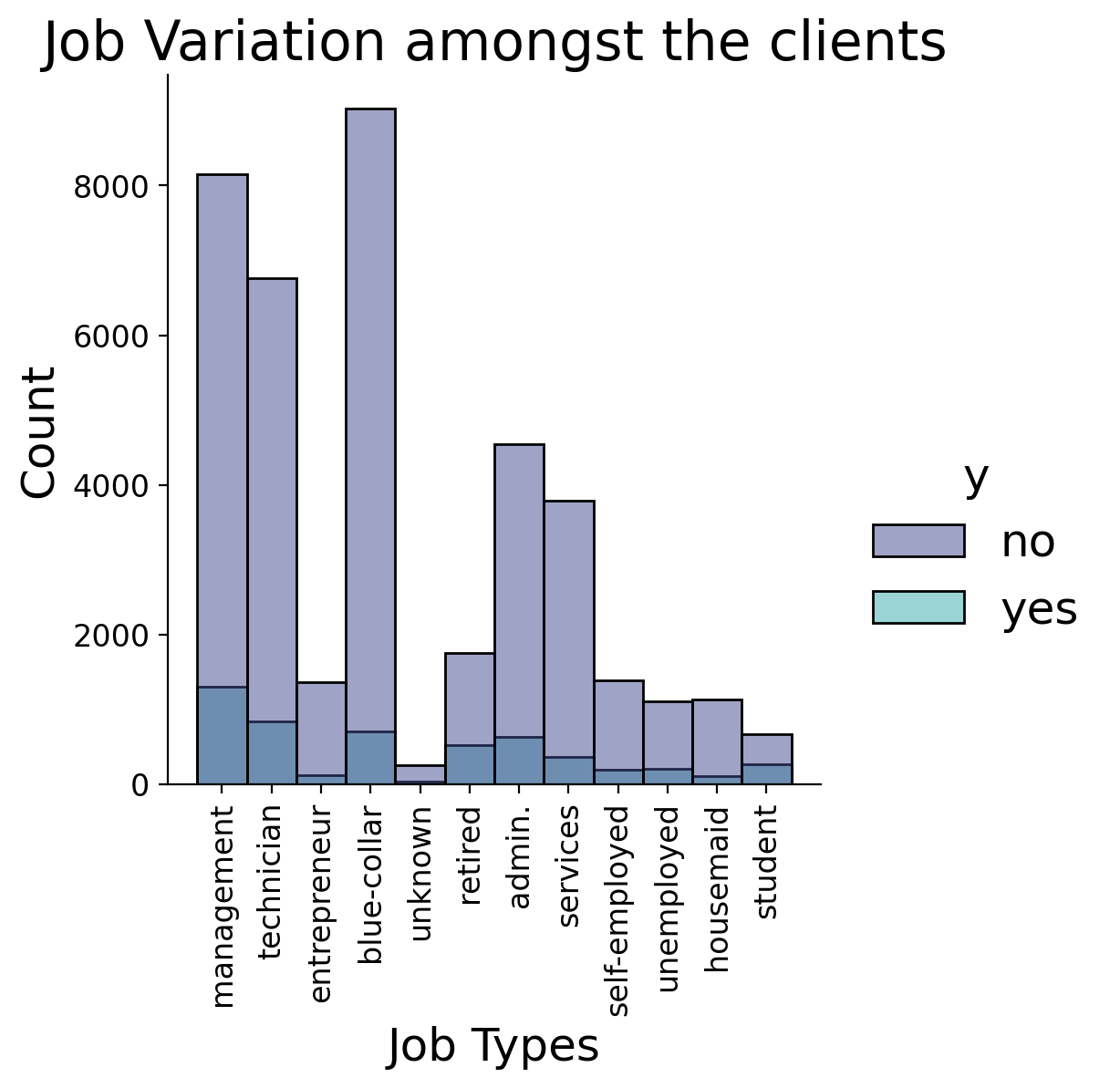
1. **What is the distribution of age among the clients?**

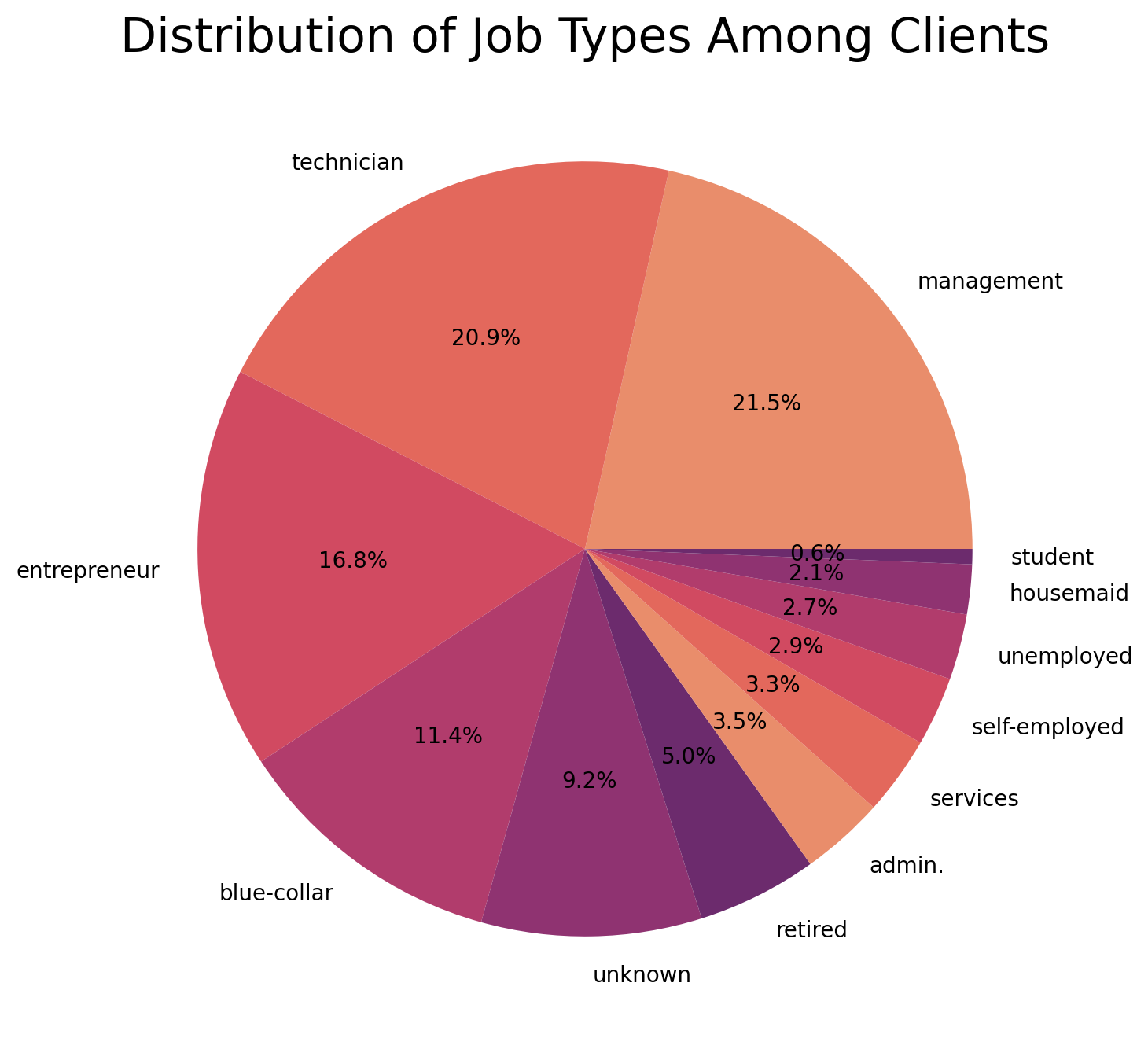




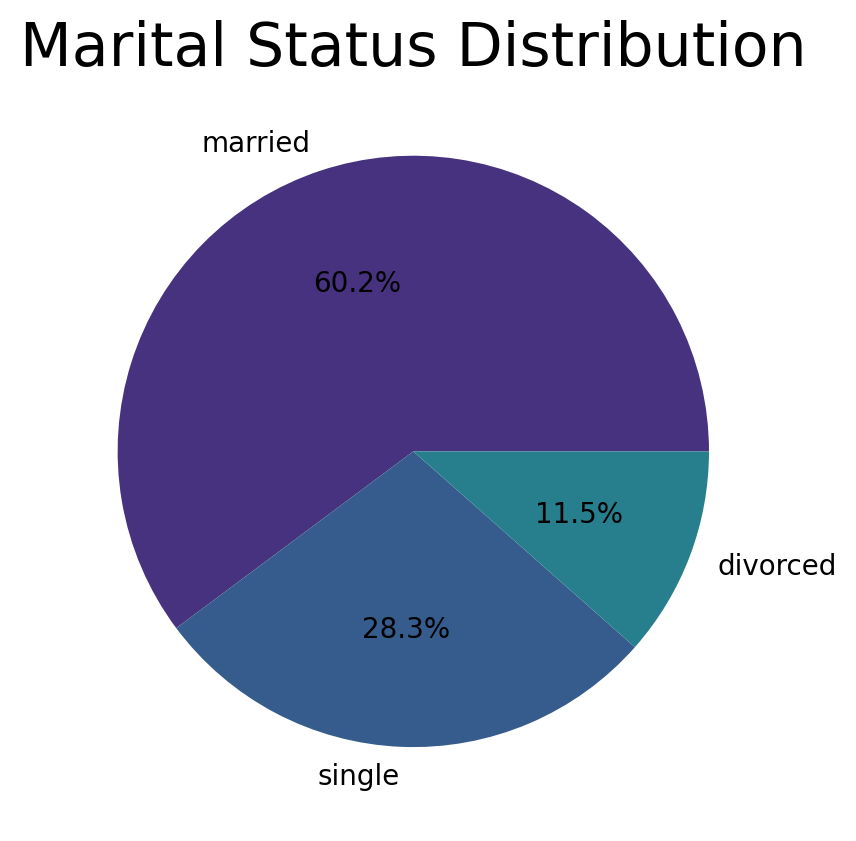
From the graph above, we can see that the age of client varies from 20 yrs to 95 yrs. The most row count is found in the range of 30yrs to 35yrs.

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

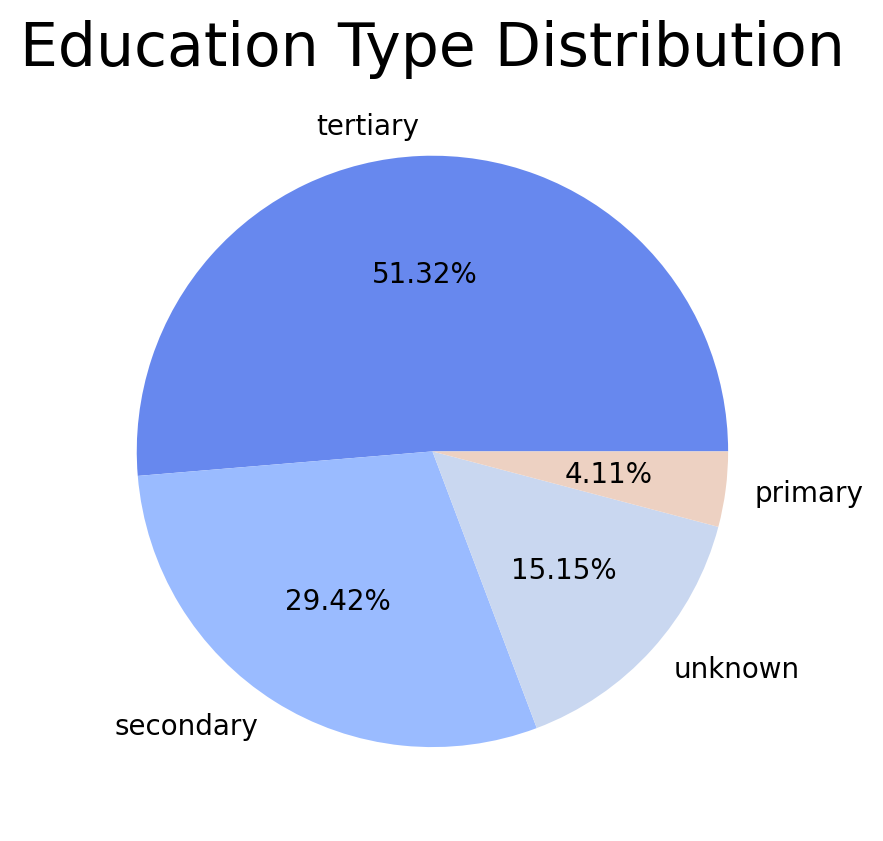
****

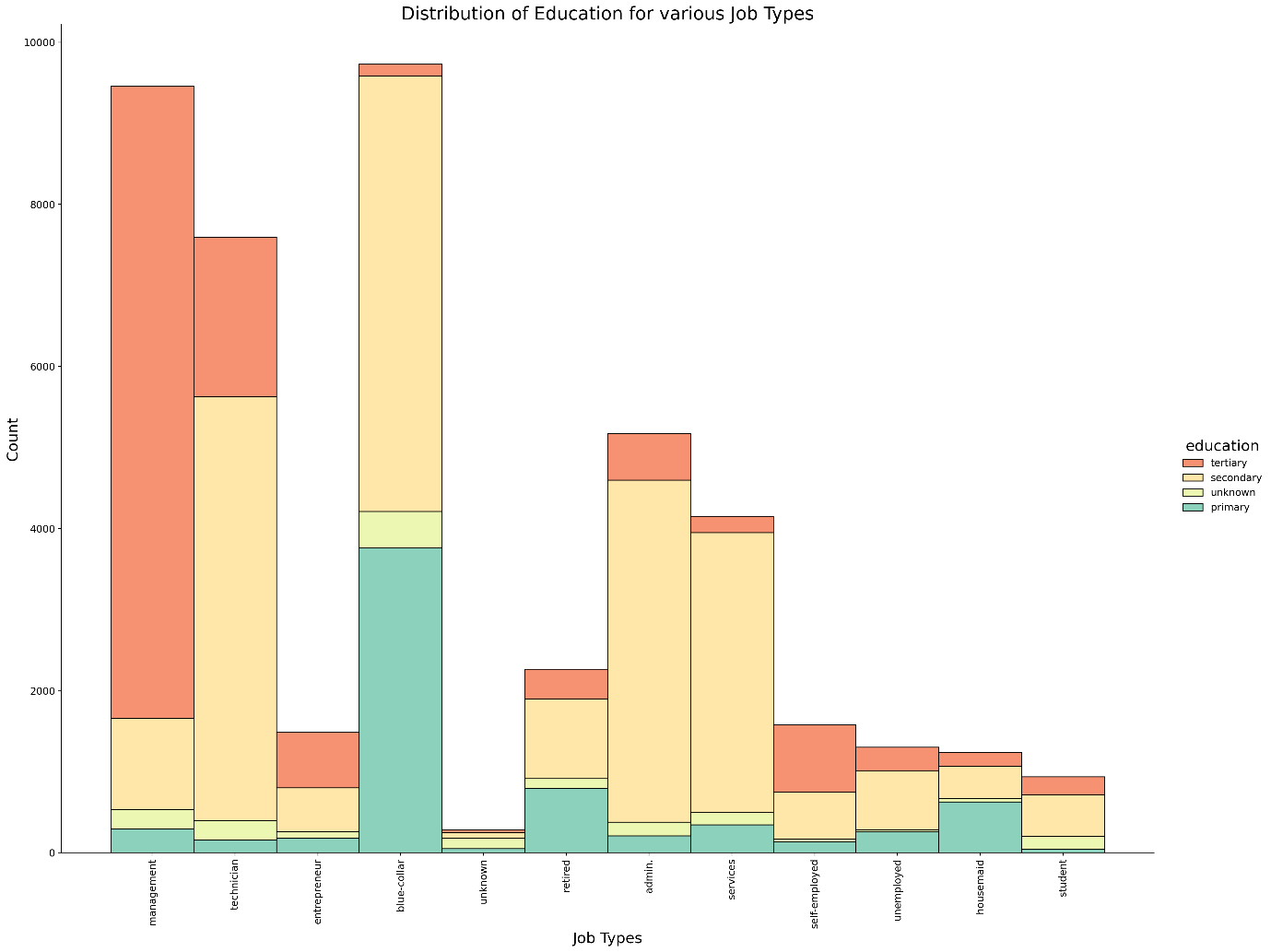
****

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

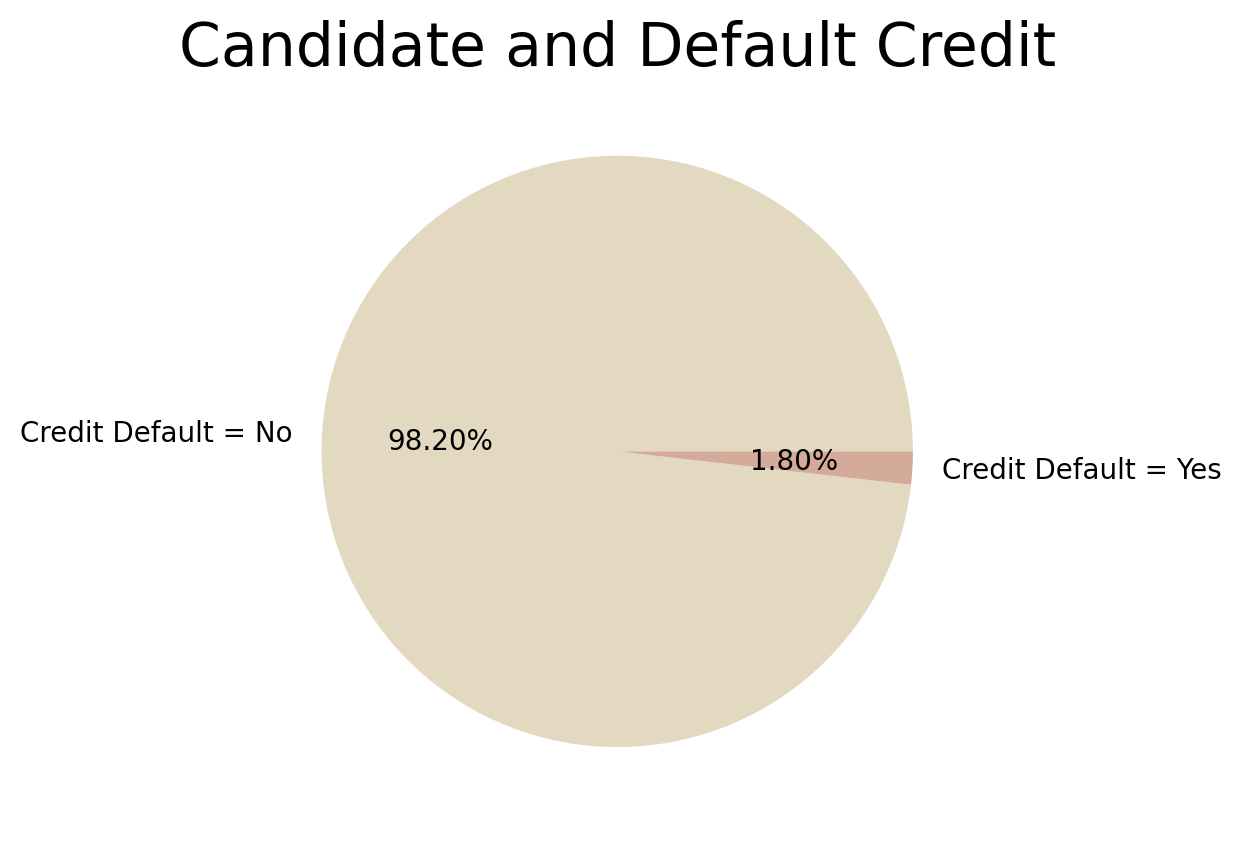
****

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

****

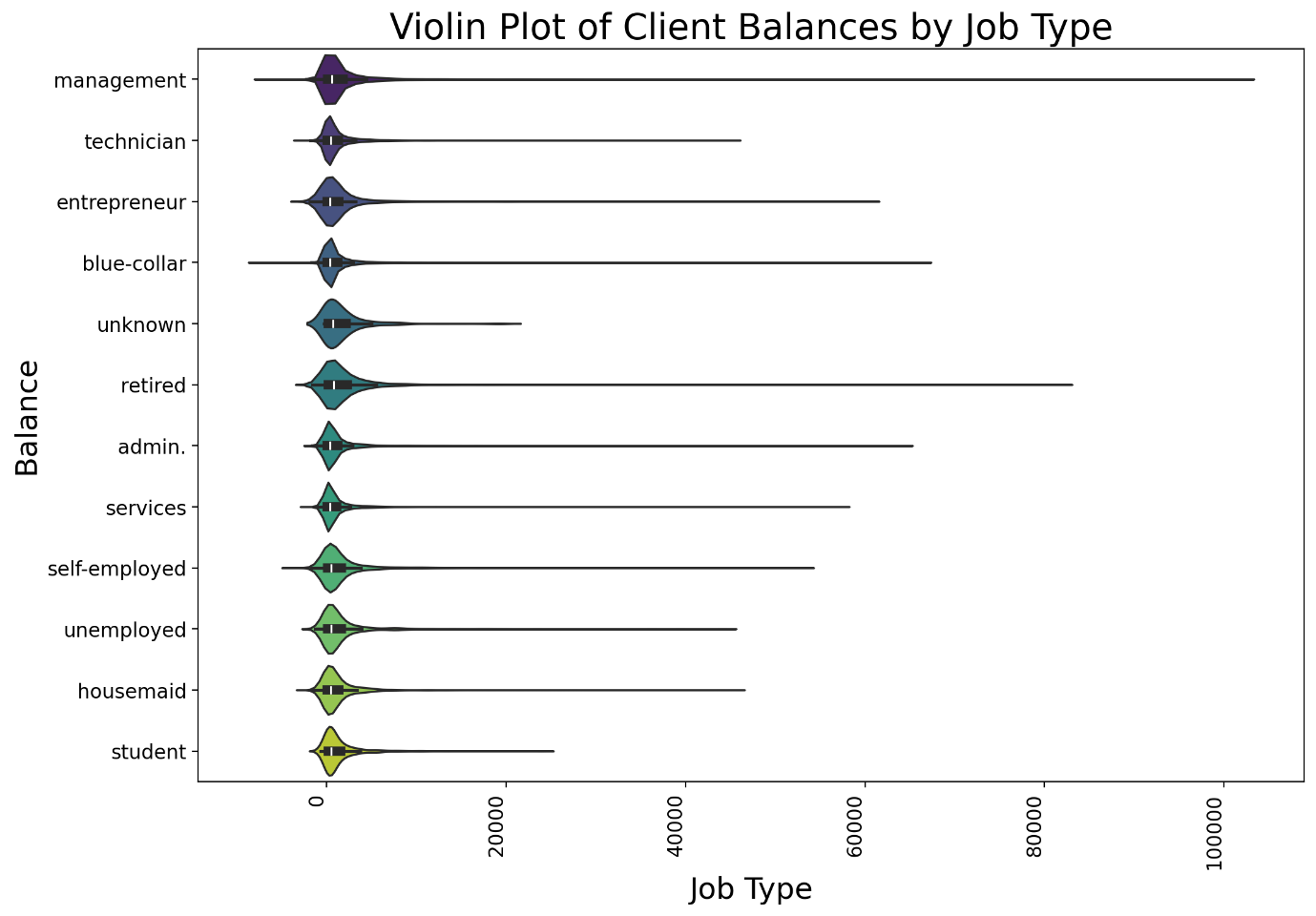
****

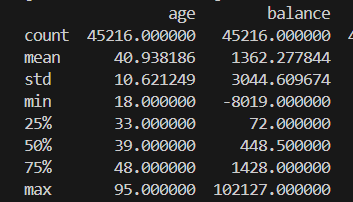
1. **What proportion of clients have credit in default?**

****

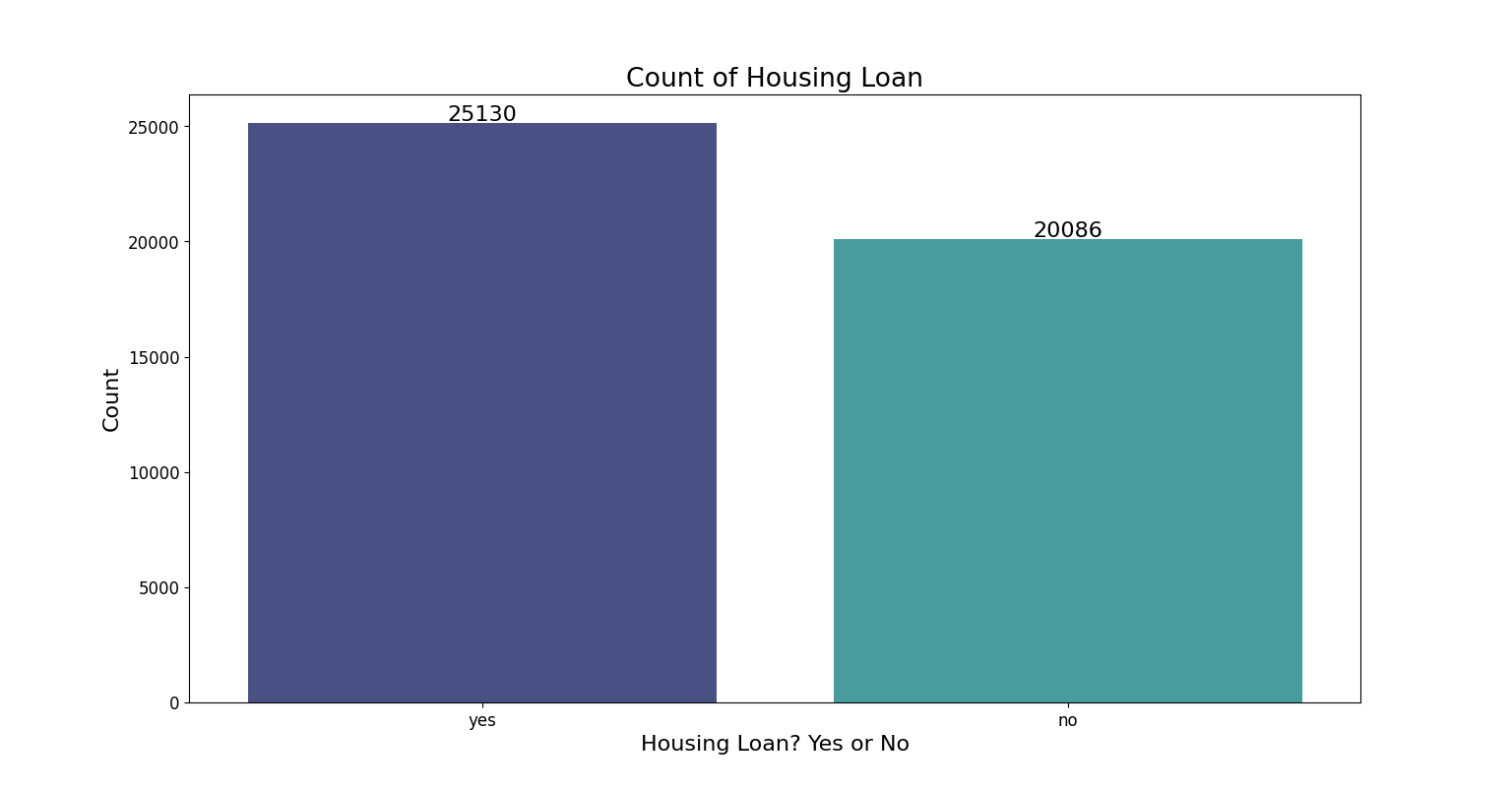
We can clearly infer from the pie chart that 1.8% of clients have credit by default.

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

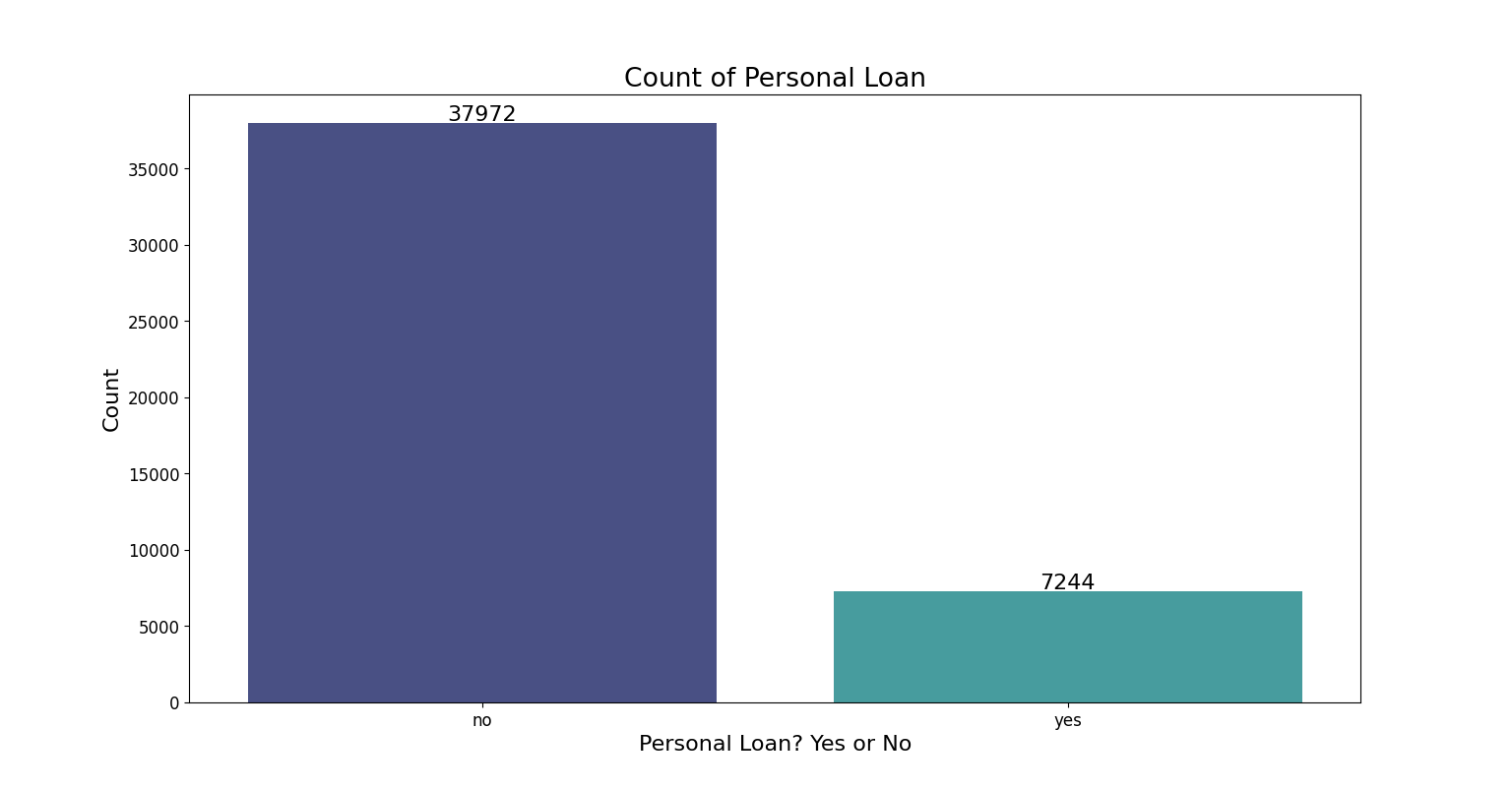
****

****

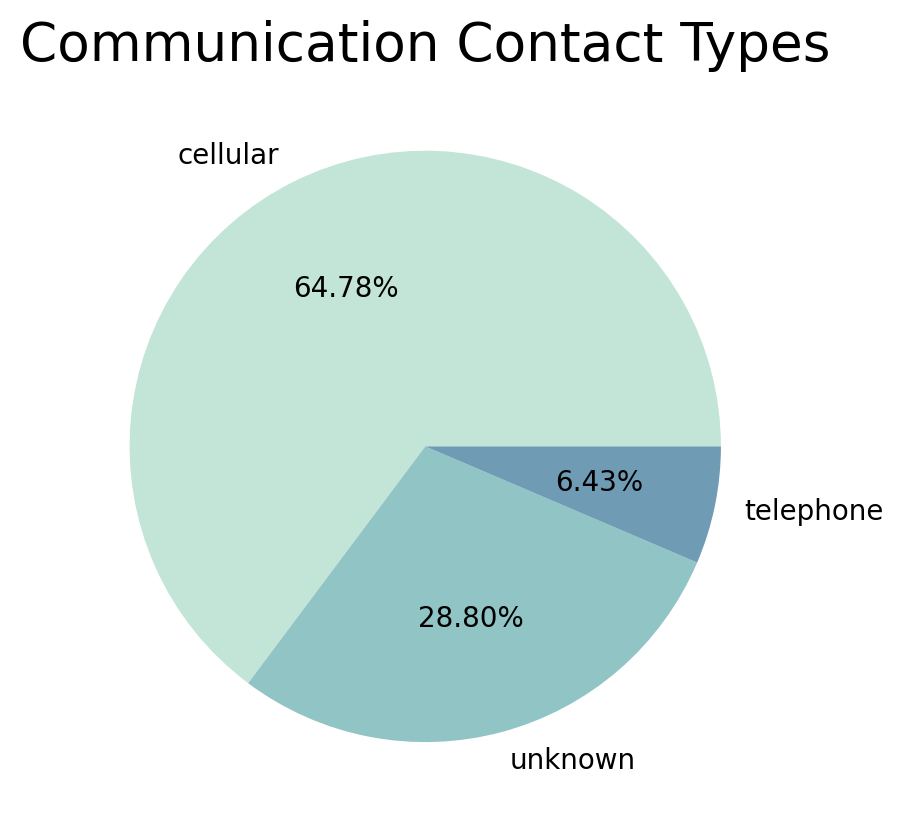
1. **How many clients have housing loans?**

****

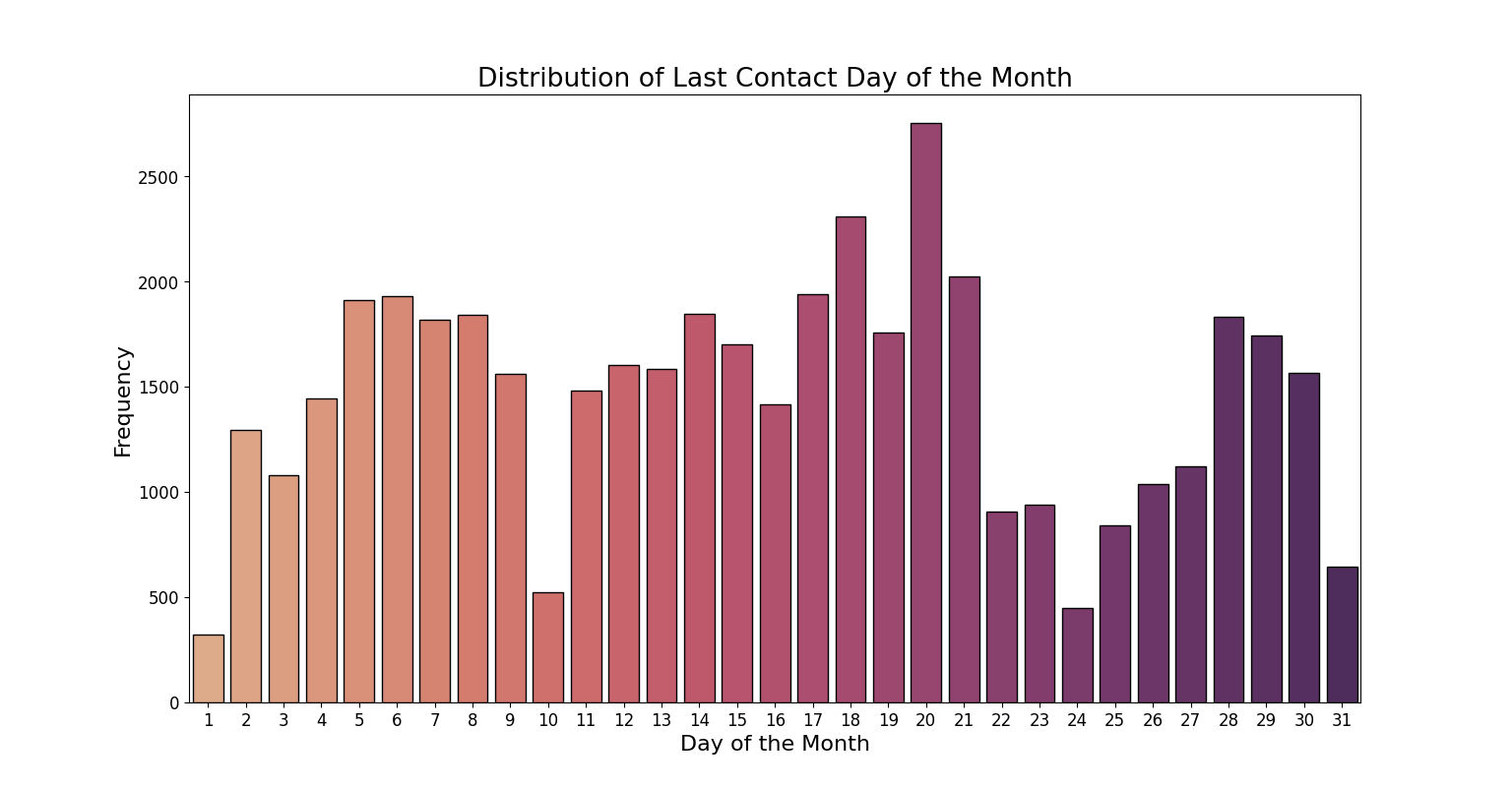
1. **How many clients have personal loans?**

****

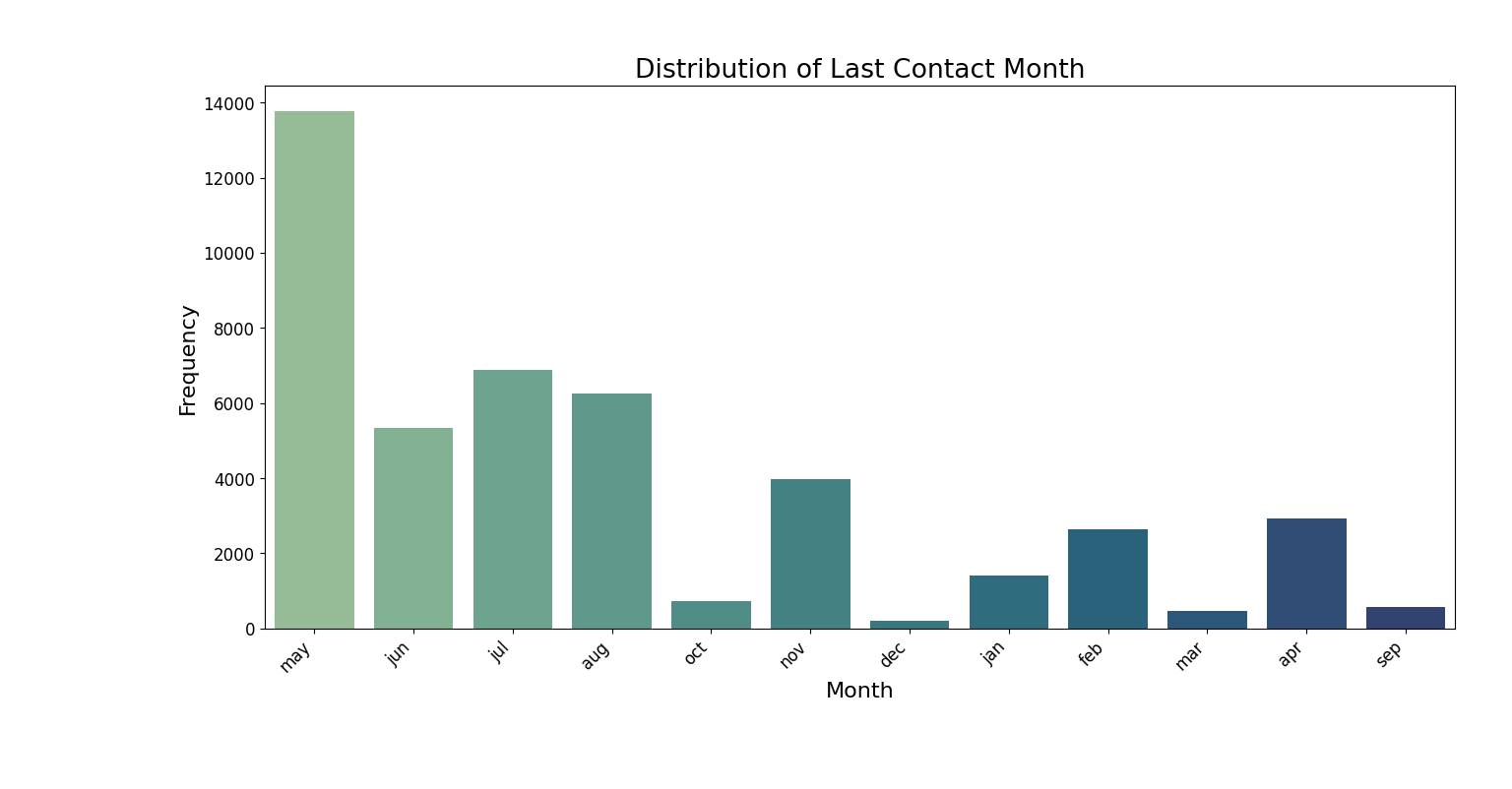
1. **What are the communication types used for contacting clients during the campaign?**

****

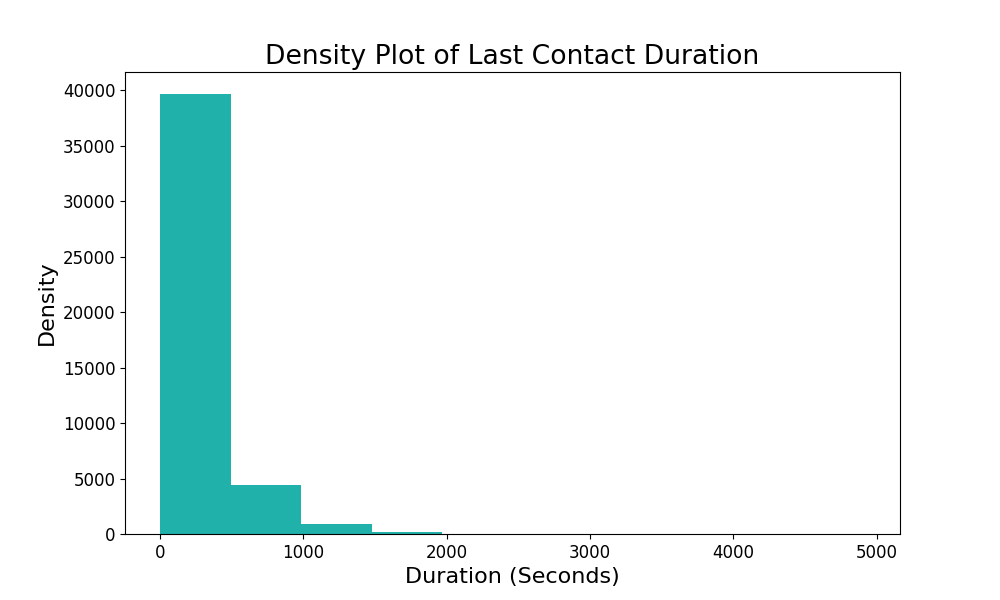
1. **What is the distribution of the last contact day of the month?**

****

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

****

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

****

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

* What is the distribution of the number of days passed since the client was last contacted from a previous campaign?
* How many contacts were performed before the current campaign for each client?
* What were the outcomes of the previous marketing campaigns?
* What is the distribution of clients who subscribed to a term deposit vs. those who did not?
* Are there any correlations between different attributes and the likelihood of subscribing to a term deposit?