**BANKING DATASET**

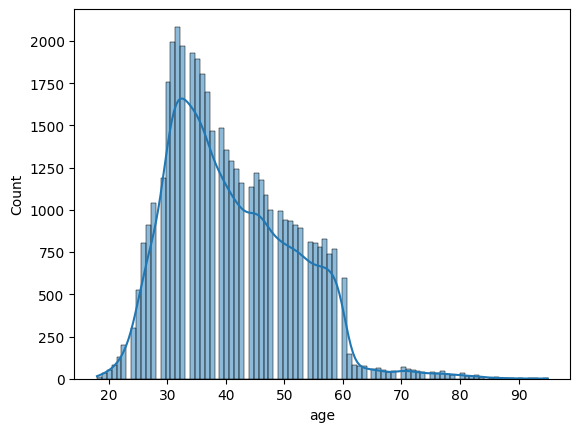
* **Sharon Blessy J V**

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

**Ans.:** Majority of the clients are within the age 30 to 40. The following histogram plot is plotted using the following code snippet.

# Histogram depicting the distribution of the age of the clients

*sns.histplot(bd['age'], kde = True)*

****

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

**Ans.:** The following bar graph and pie chart describe the job type of the clients. The code snippet used to obtain the same is also mentioned below.

*# Bar graph for counts*

*sns.catplot(y="job", kind="count", data=bd, palette='viridis')*

*plt.title('Distribution of Job Types Among Clients')*

*plt.xlabel('No. of Clients')*

*plt.ylabel('Job Type')*

*plt.show()*

*# Pie chart for the percentages*

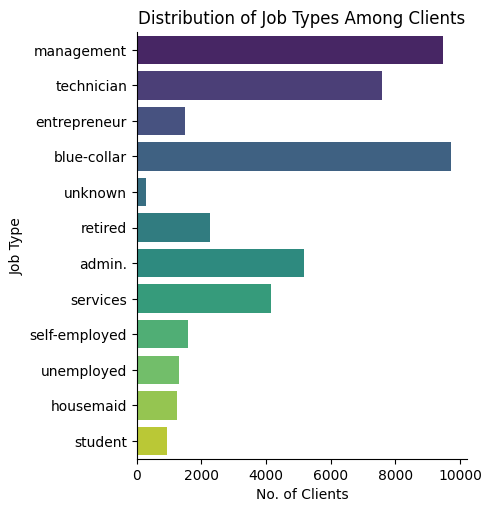
*job\_percent = bd['job'].value\_counts(normalize=True) \* 100*

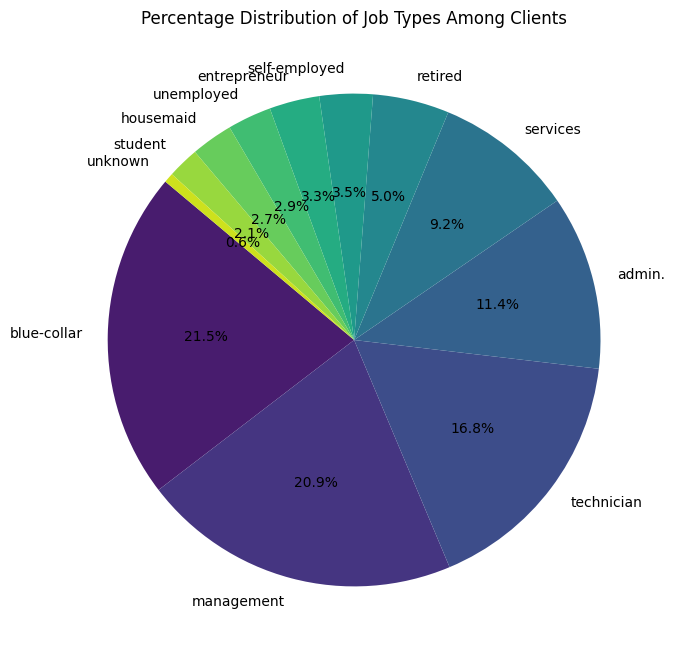
*plt.figure(figsize=(8, 8))*

*plt.pie(job\_percent, labels=job\_percent.index, autopct='%1.1f%%', startangle=140, colors=sns.color\_palette('viridis', len(job\_percent)))*

*plt.title('Percentage Distribution of Job Types Among Clients')*

*plt.show()*



****

Majority of the clients are having blue-collar jobs.

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

**Ans.:** The following graph shows the distribution of marital status of the clients obtained using the code snippet below.

*# Distribution of marital status of clients*

*Marital\_Status = bd['marital'].tolist()*

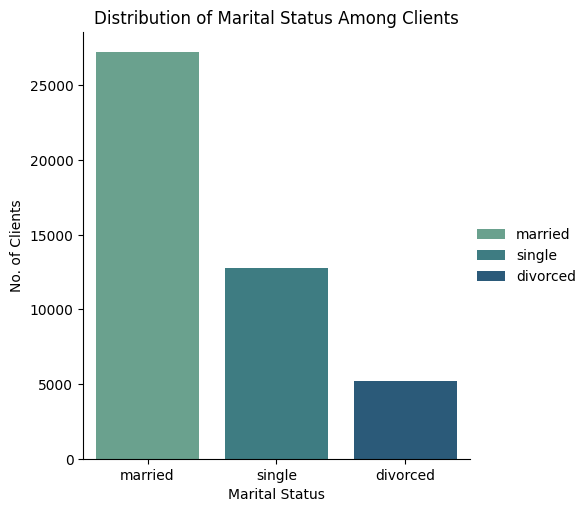
*sns.catplot(x = 'marital', hue=Marital\_Status, data = bd, kind = "count" , palette = 'crest')*

*plt.title('Distribution of Marital Status Among Clients')*

*plt.xlabel('Marital Status')*

*plt.ylabel('No. of Clients')*

*plt.show())*

****

Most clients are married and some are unmarried while others are divorced.

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

**Ans.:** The following graph shows the distribution of level of education among the clients obtained using the code snippet below.

#Distribution of education level among clients

plt.figure(figsize=(10, 6))

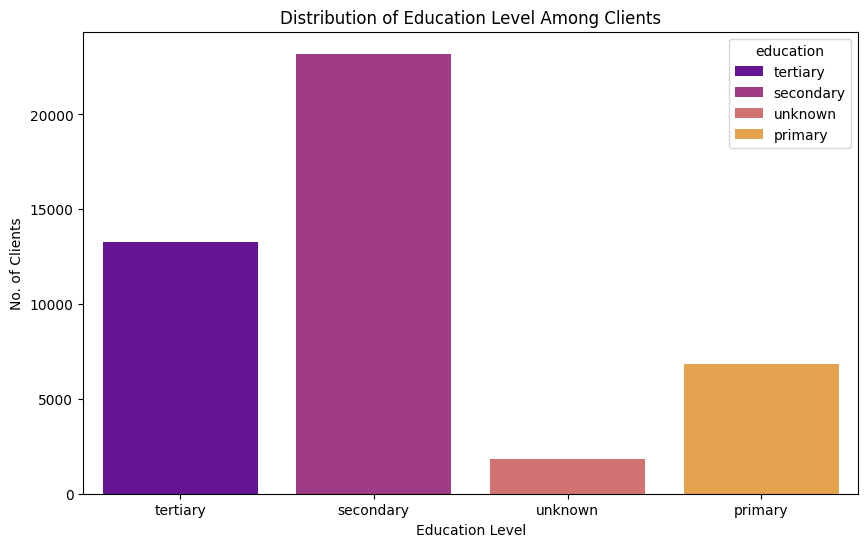
sns.countplot(x='education', hue='education', data=bd, palette='plasma')

plt.title('Distribution of Education Level Among Clients')

plt.xlabel('Education Level')

plt.ylabel('No. of Clients')

plt.show()

****

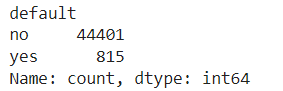
Many clients have secondary level of education.

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

**Ans.:** The count of the no. of clients who have and don’t have credit in default can be found out by the following code snippet.

*print(bd['default'].value\_counts())*

**Output:**

****

Therefore, the ratio of clients who have credit in default to those who don’t is 815:44401

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

**Ans.:**

The distribution of average yearly balance can be analysed by plotting a graph.

*plt.figure(figsize=(10, 6))*

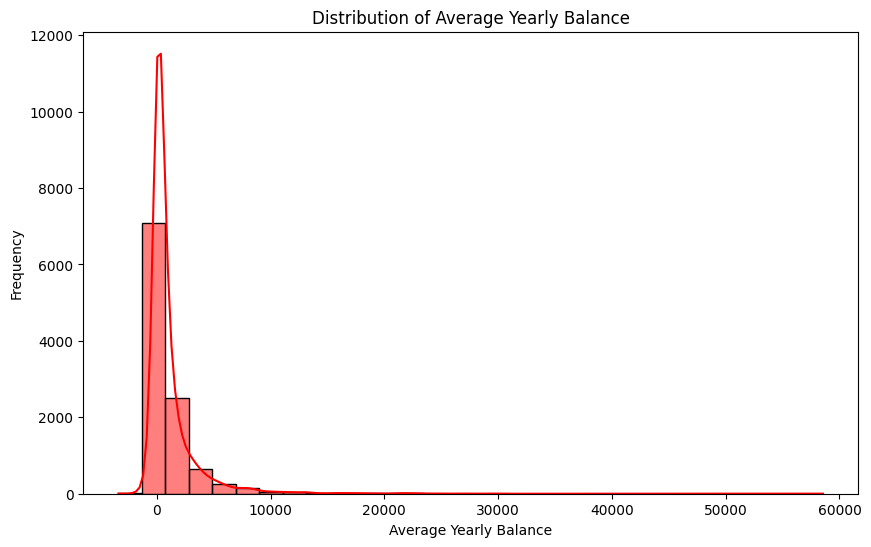
*sns.histplot(bd['balance'], bins=30, kde=True, color='red')*

*plt.title('Distribution of Average Yearly Balance')*

*plt.xlabel('Average Yearly Balance')*

*plt.ylabel('Frequency')*

*plt.show()*

****

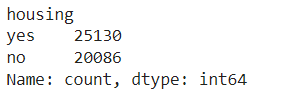
Majority of the clients have unpaid amounts in the banks i.e they owe money to the bank.

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

**Ans.:** The no. of clients who have a housing loan can be found out using the following code.

*print(bd['housing'].value\_counts())*

**Output:**

****

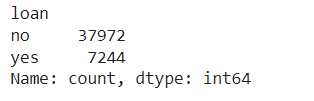
No. of clients who have housing loan is 25130 while the remaining do not have housing loan.

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

**Ans.:** The no. of clients who have personal loan can be found out using the following code.

*print(bd['loan'].value\_counts())*

**Output:**

****

No. of clients who have personal loan is 37972 while the remaining do not have personal loan.

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

**Ans.:**

The different types of communication modes used for contacting clients during campaign can be found out by the code below.

*print(bd['contact'].unique())*

**Output:**

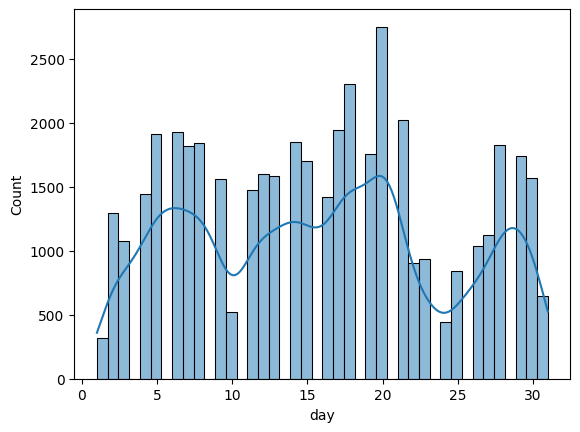
****

The different modes include cellular, telephone and other unknown methods.

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

**Ans.:** The distribution of the last contact day of the month is obtained by plotting a histogram.

*sns.histplot(bd['day'], kde = True )*

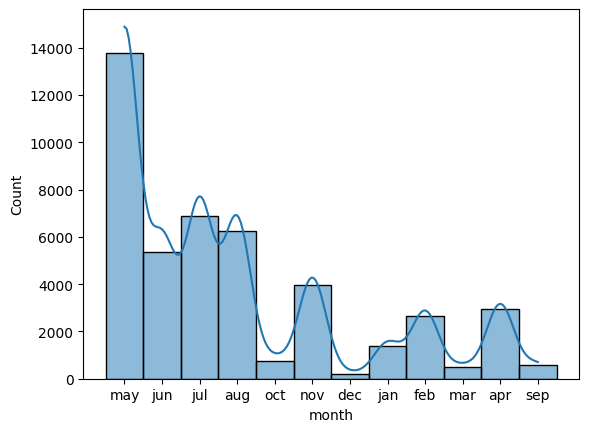
****

Most clients were last contacted on 17th to 22nd of the month.

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

**Ans.:** The variation of the last contact month among the clients can be seen by plotting a histogram using the below code.

*sns.histplot(bd['month'], kde = True )*

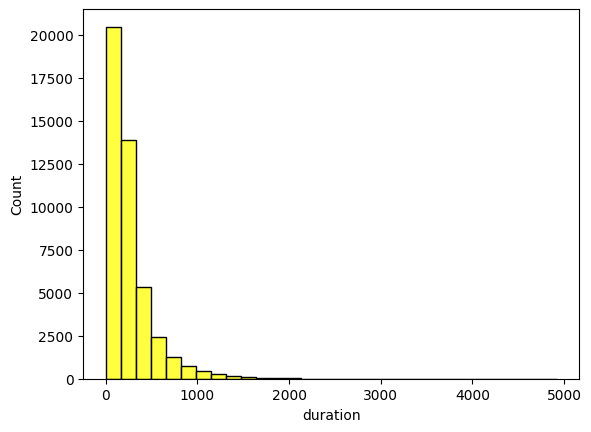


Most clients were last contacted in the month of May and the least no. of clients were contacted in December.

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

**Ans.:** Distribution of the duration of the last contact can be analysed by the histogram obtained by the code below.

*sns.histplot(bd['duration'], bins = 30 ,kde = True )*

****

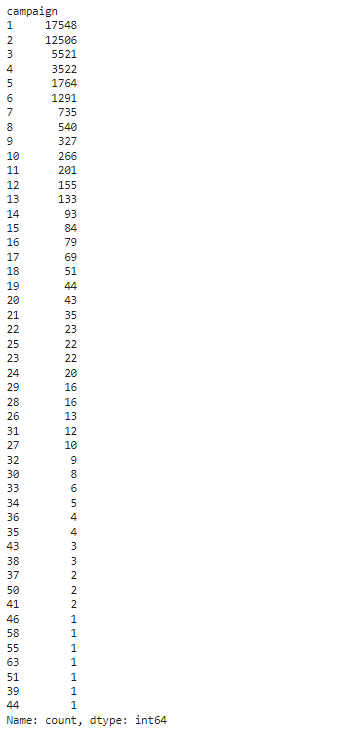
Most clients were contacted for duration from 0-180 seconds.

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

**Ans.:** The list can be obtained by the following code.

*print(bd['campaign'].value\_counts())*

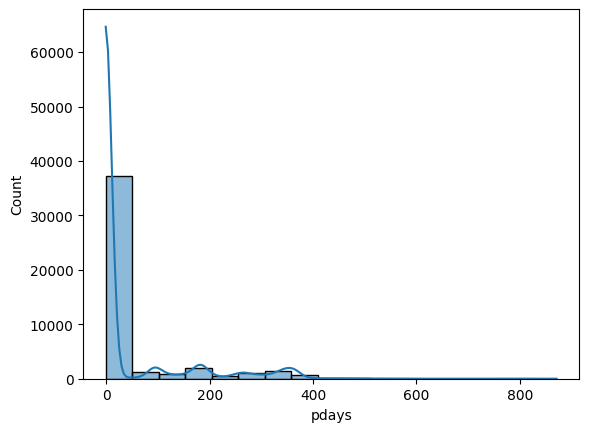
**Output:**

****

1. **What is the distribution of the number of days passed since the client was last contacted from a previous campaign?**

**Ans.:** The distribution is obtained by plotting the following histogram.

*sns.histplot(bd['pdays'], kde = True )*

****

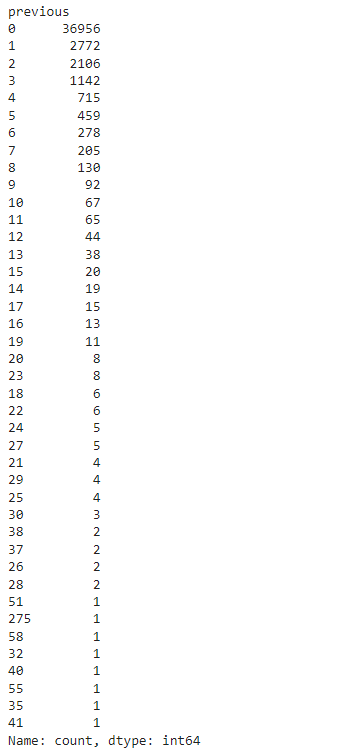
Most clients were last contacted before 50 days.

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

**Ans.:**

The number of contacts performed before the current can be listed out by the following code.

*print(bd['previous'].value\_counts())*

****

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

**Ans.:**

The outcomes of the previous marketing are identified along with their distribution by the following plot.

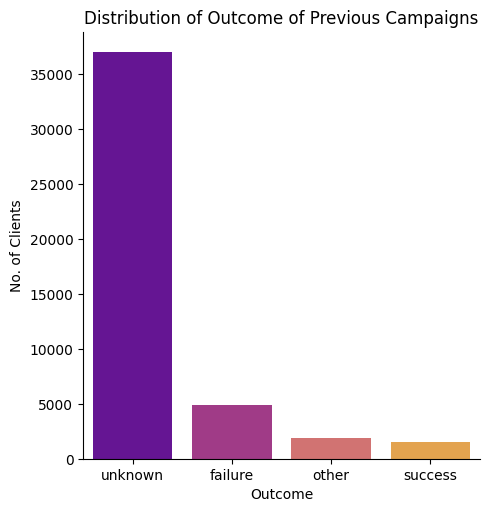
*sns.catplot(x="poutcome", kind="count", data=bd, palette='plasma')*

*plt.title('Distribution of Outcome of Previous Campaigns')*

*plt.xlabel('Outcome')*

*plt.ylabel('No. of Clients')*

*plt.show()*

****

Outcomes of majority of the clients remain unknown.

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

**Ans.: The distribution of clients who subscribed or did not subscribe to a term deposit can be seen by the following code.**

*print(bd['y'].value\_counts())*

*sns.countplot(x='y', data=bd, palette='plasma')*

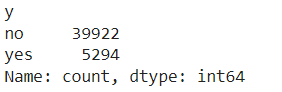
*plt.title('Distribution of Subscription')*

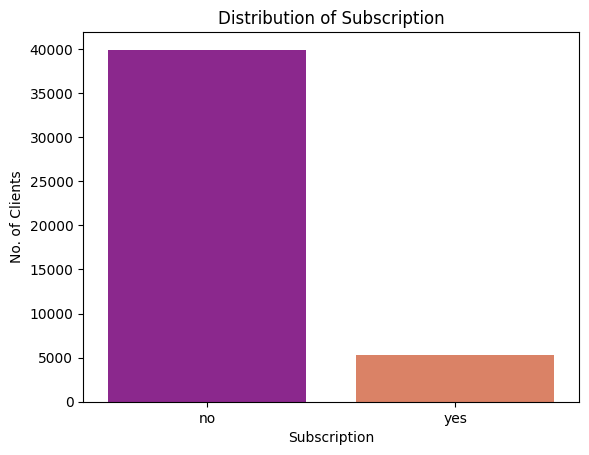
*plt.xlabel('Subscription')*

*plt.ylabel('No. of Clients')*

*plt.show()*

**Output:**

****

****

Majority of the clients haven’t subscribed for a term deposit.

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

**Ans.:**

There are no clear correlations between other attributes and the likelihood of subscribing to a term deposit.

The following is the comparison of likelihood of subscribing to a term deposit by job can be seen below.

*subs\_job = bd.groupby('y')['job'].value\_counts()*

*print(subs\_job)*

*sns.countplot(x='job', hue='y', data=bd, palette='plasma')*

*plt.xticks(rotation=45)*

