CASE PROJECT : BANKING

```
# importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Load the Dataset
df = pd.read_csv('banking_data.csv')
df.shape
```

```
(45216, 19)
```

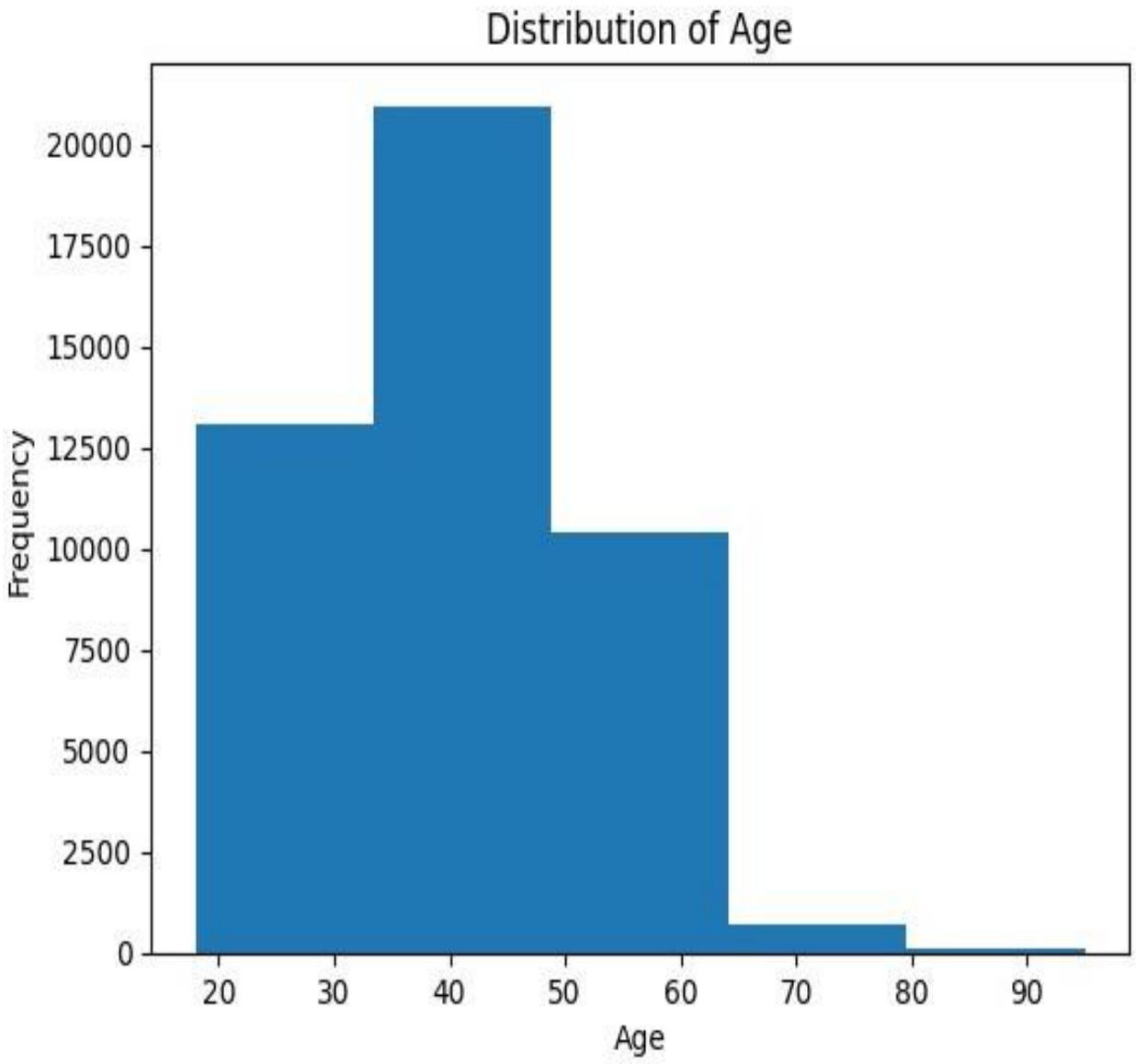
```
#unique counts in each column
df.nunique()
```

```
age                 77
job                 12
marital              3
marital_status       3
education            4
default              2
balance           7168
housing              2
loan                 2
contact              3
day                 31
month               12
day_month          318
duration          1573
campaign            48
pdays              559
previous            41
poutcome             4
y                    2
dtype: int64
```

Answers:

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

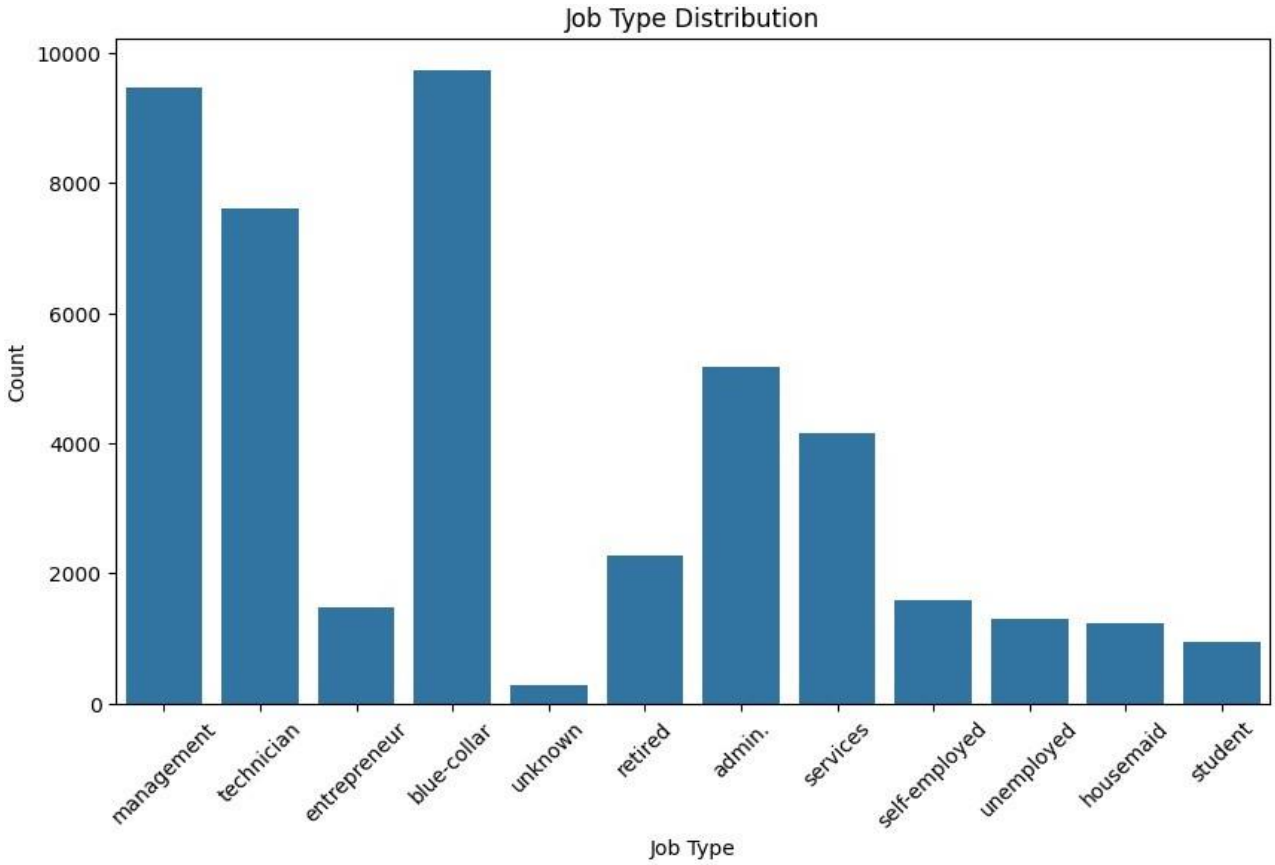
```
#plot histogram

plt.hist(df['age'], bins=5)
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.title('Distribution of Age')
plt.show()
```



2.- How does the job type vary among the clients?

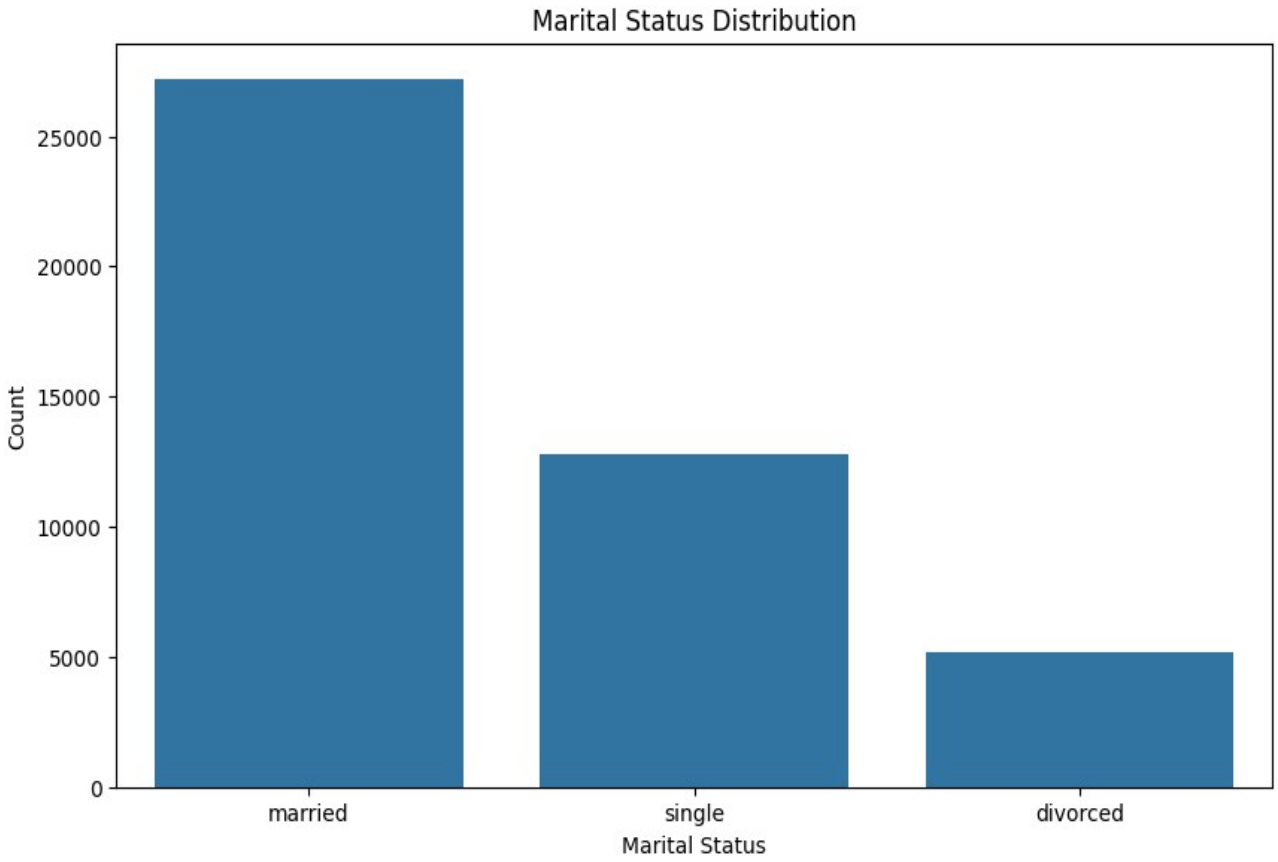
```
plt.figure(figsize=(10, 6))
sns.countplot(x='job', data=df)
plt.xlabel('Job Type')
plt.ylabel('Count')
plt.title('Job Type Distribution')
plt.xticks(rotation=45)
```

```
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11],
 [Text(0, 0, 'management'),
  Text(1, 0, 'technician'),
  Text(2, 0, 'entrepreneur'),
  Text(3, 0, 'blue-collar'),
  Text(4, 0, 'unknown'),
  Text(5, 0, 'retired'),
  Text(6, 0, 'admin.'),
  Text(7, 0, 'services'),
  Text(8, 0, 'self-employed'),
  Text(9, 0, 'unemployed'),
  Text(10, 0, 'housemaid'),
  Text(11, 0, 'student')])
```



3.- What is the marital status distribution of the clients?

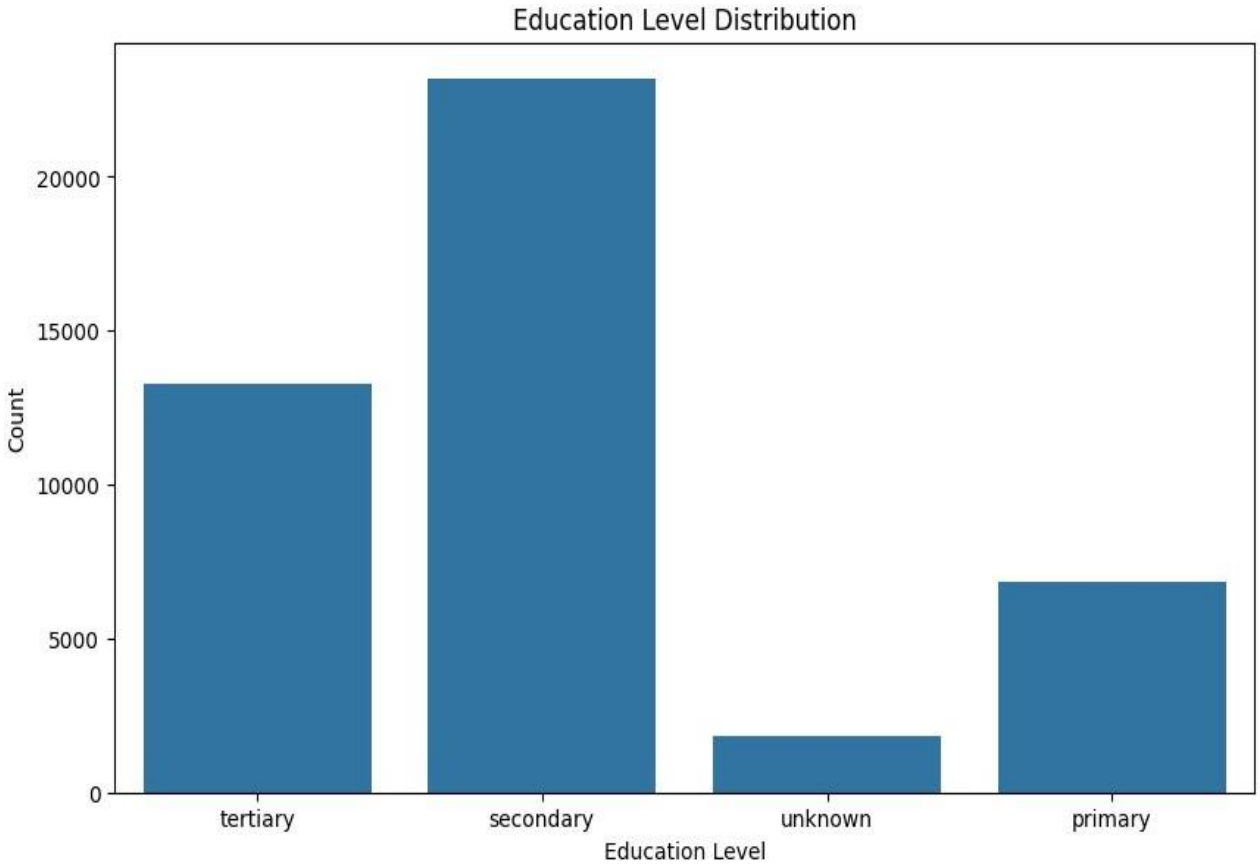
```
plt.figure(figsize=(10,6))

sns.countplot(x='marital', data=df)
plt.xlabel('Marital Status')
plt.ylabel('Count')
plt.title('Marital Status Distribution')
plt.show()
```



4.- What is the level of education among the clients?

```
plt.figure(figsize=(10, 6))
sns.countplot(x='education', data=df)
plt.xlabel('Education Level')
plt.ylabel('Count')
plt.title('Education Level Distribution')
```

```
Text(0.5, 1.0, 'Education Level Distribution')
```



5. What proportion of clients have credit in default?

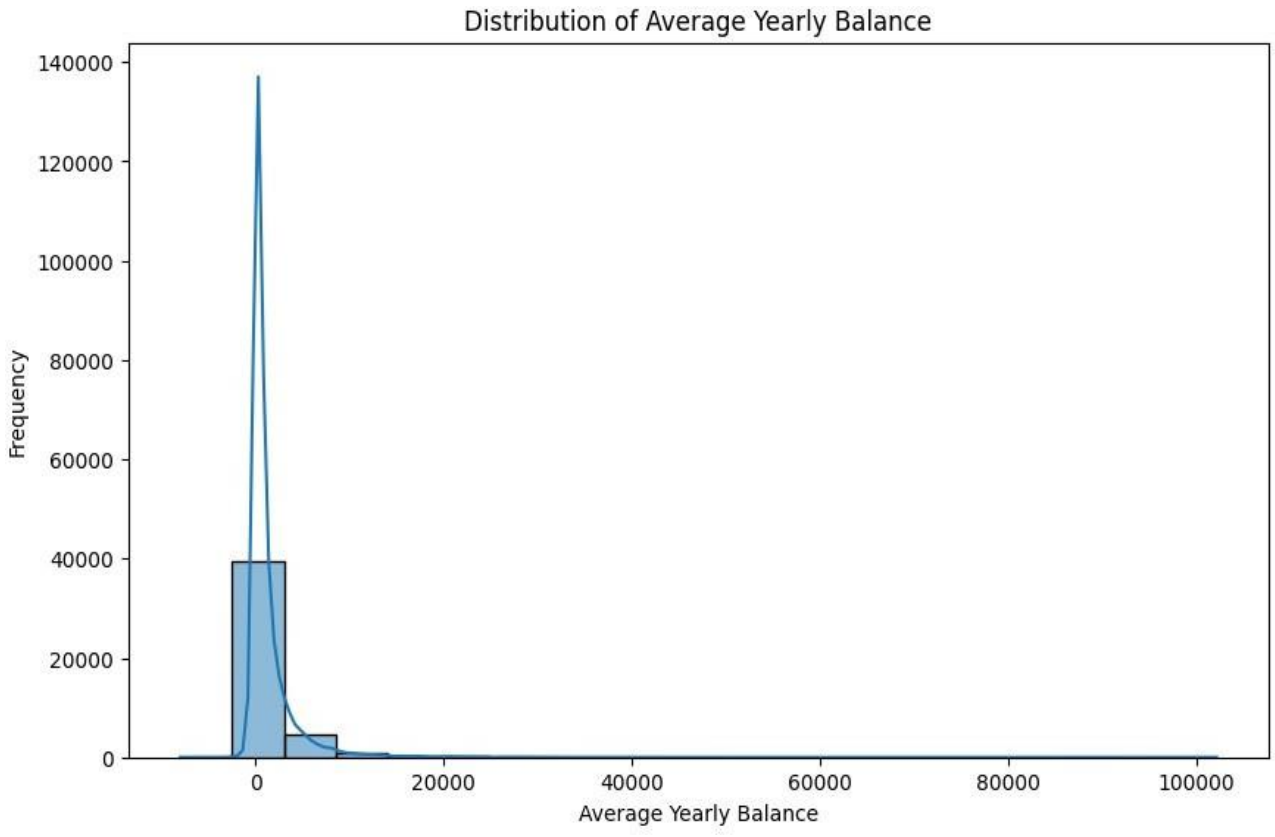
```
#proportion of clients have credit
default_proportion = df['default'].value_counts(normalize=True)
print(default_proportion)
```

```
default
no     0.981975
yes    0.018025
Name: proportion, dtype: float64
```

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

```
#distribution of average yearly balance
plt.figure(figsize=(10, 6))
sns.histplot(df['balance'], bins=20, kde=True)
plt.xlabel('Average Yearly Balance')
plt.ylabel('Frequency')
plt.title('Distribution of Average Yearly Balance')
```

```
Text(0.5, 1.0, 'Distribution of Average Yearly Balance')
```



7. How many clients have housing loans?

```
housing_loan_count = df['housing'].value_counts()
print(housing_loan_count)
```

```
housing
yes    25130
no     20086
Name: count, dtype: int64
```

8. How many clients have personal loans?

```
personal_loan_count=df['loan'].value_counts()
print(personal_loan_count)
```

```
loan
no     37972
yes     7244
Name: count, dtype: int64
```

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

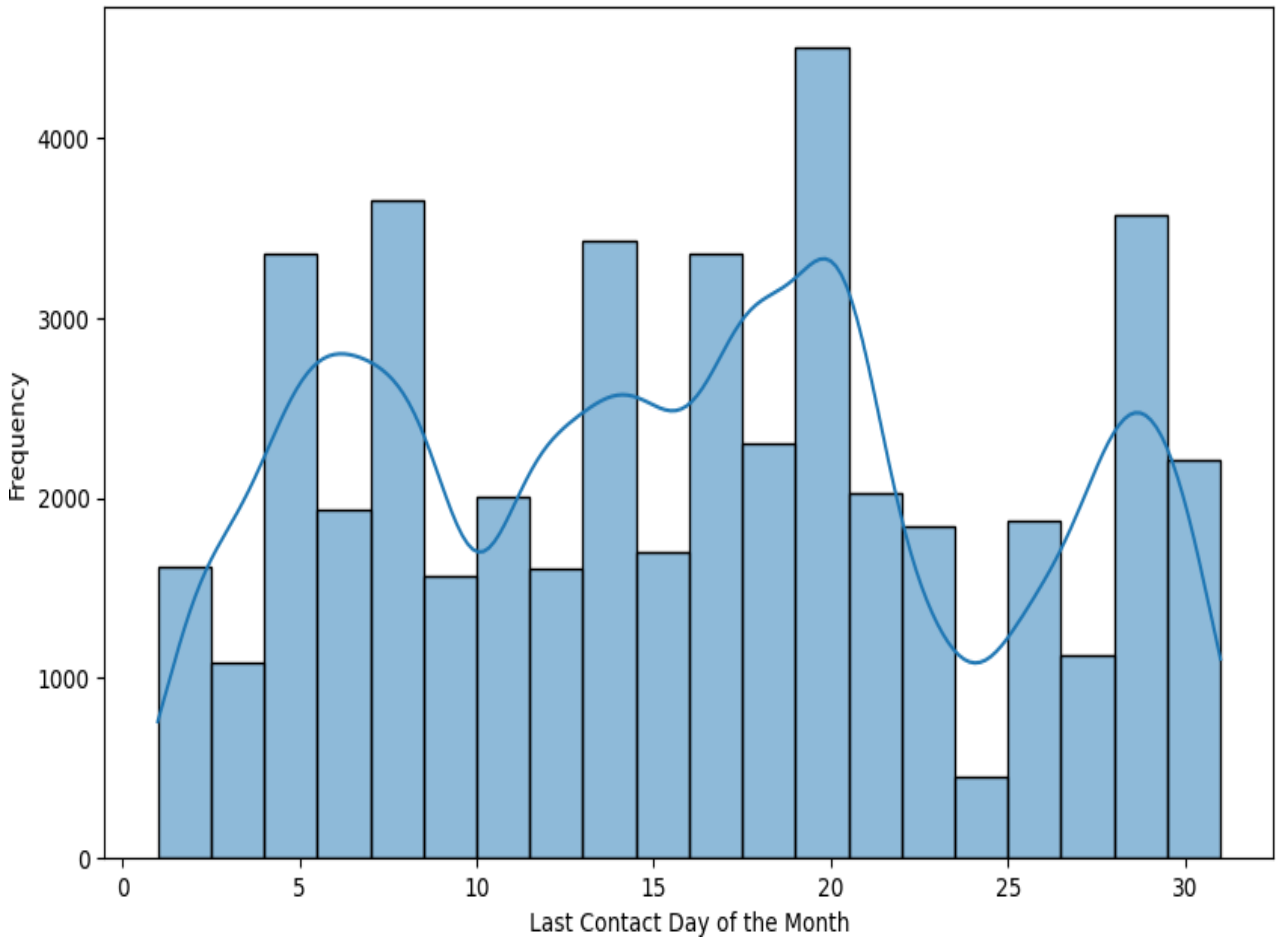
```
communication_types=df['contact'].unique()
print(communication_types)
```

```
['unknown' 'cellular' 'telephone']
```

10. What is the distribution of the last contact day of the month?

```
plt.figure(figsize=(10, 6))
sns.histplot(df['day'], bins=20, kde=True)
plt.xlabel('Last Contact Day of the Month')
plt.ylabel('Frequency')
```

```
Text(0, 0.5, 'Frequency')
```

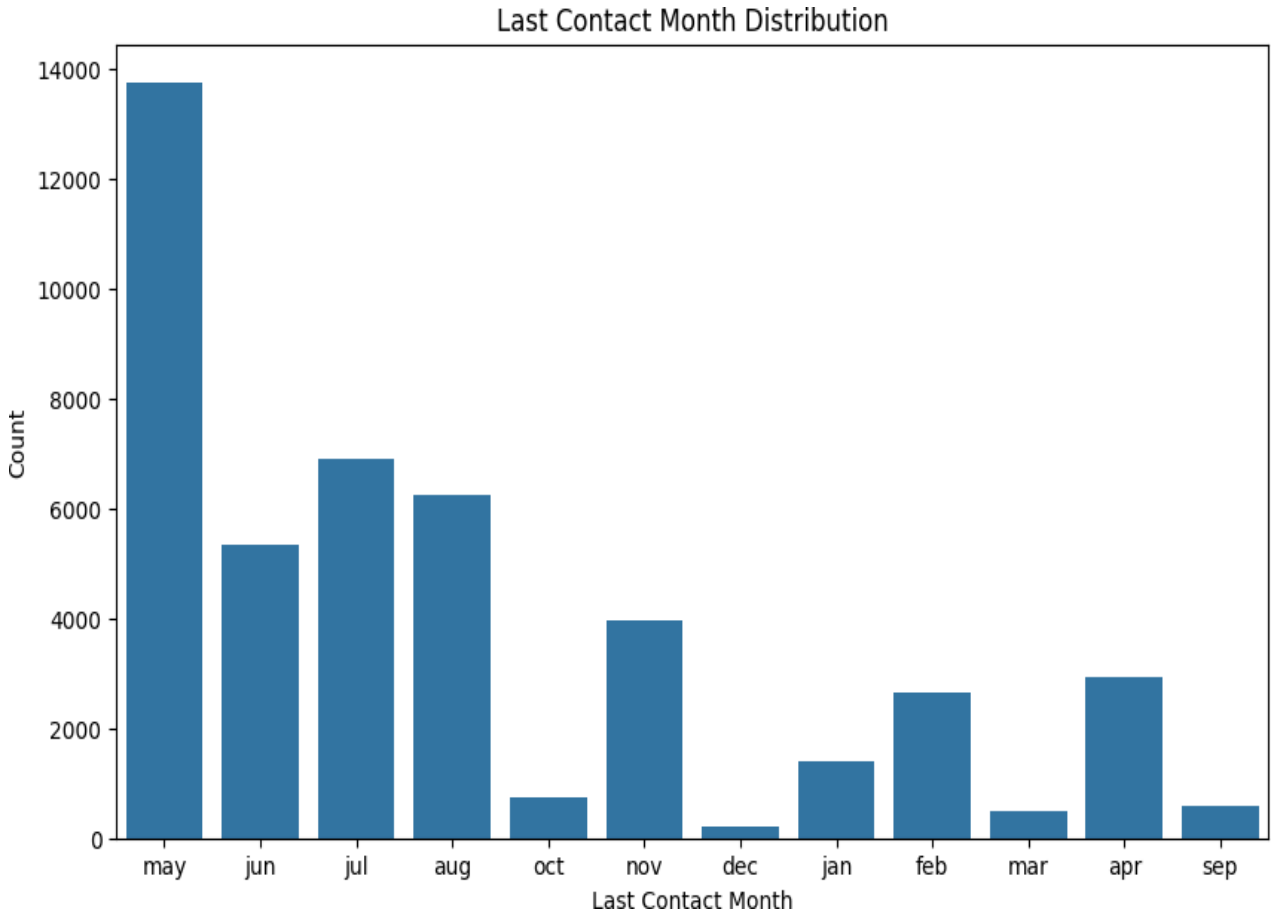


11. How does the last contact month vary among the clients?

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

```
sns.countplot(x='month', data=df)
plt.xlabel('Last Contact Month')
plt.ylabel('Count')
plt.title('Last Contact Month Distribution')
```

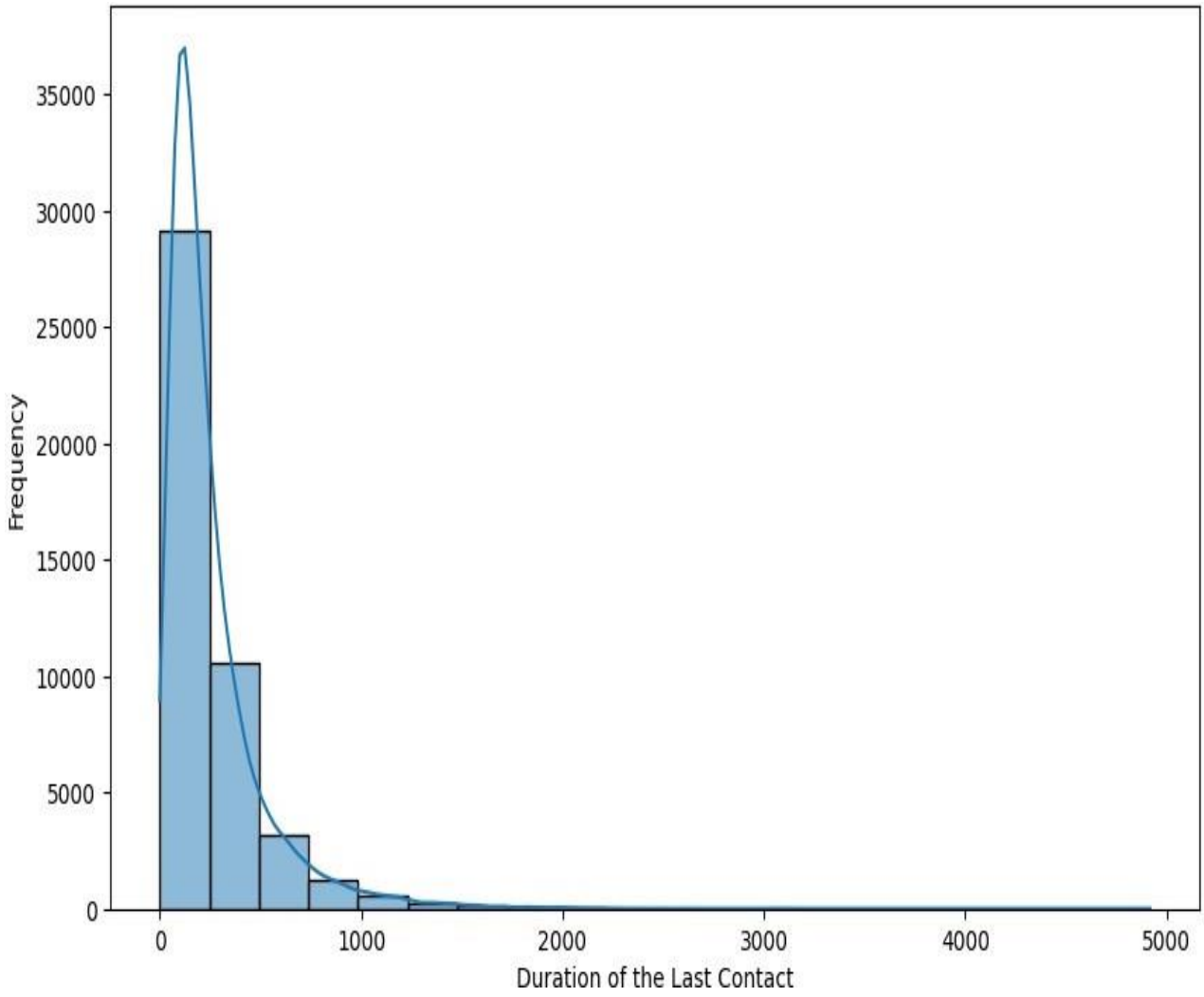
```
Text(0.5, 1.0, 'Last Contact Month Distribution')
```



12. What is the distribution of the duration of the last contact?

```
plt.figure(figsize=(10, 6))
sns.histplot(df['duration'], bins=20, kde=True)
plt.xlabel('Duration of the Last Contact')
plt.ylabel('Frequency')
```

```
Text(0, 0.5, 'Frequency')
```



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

```
#How many contacts were performed during the campaign
contact_counts=df['campaign'].value_counts()
print(contact_counts)
```

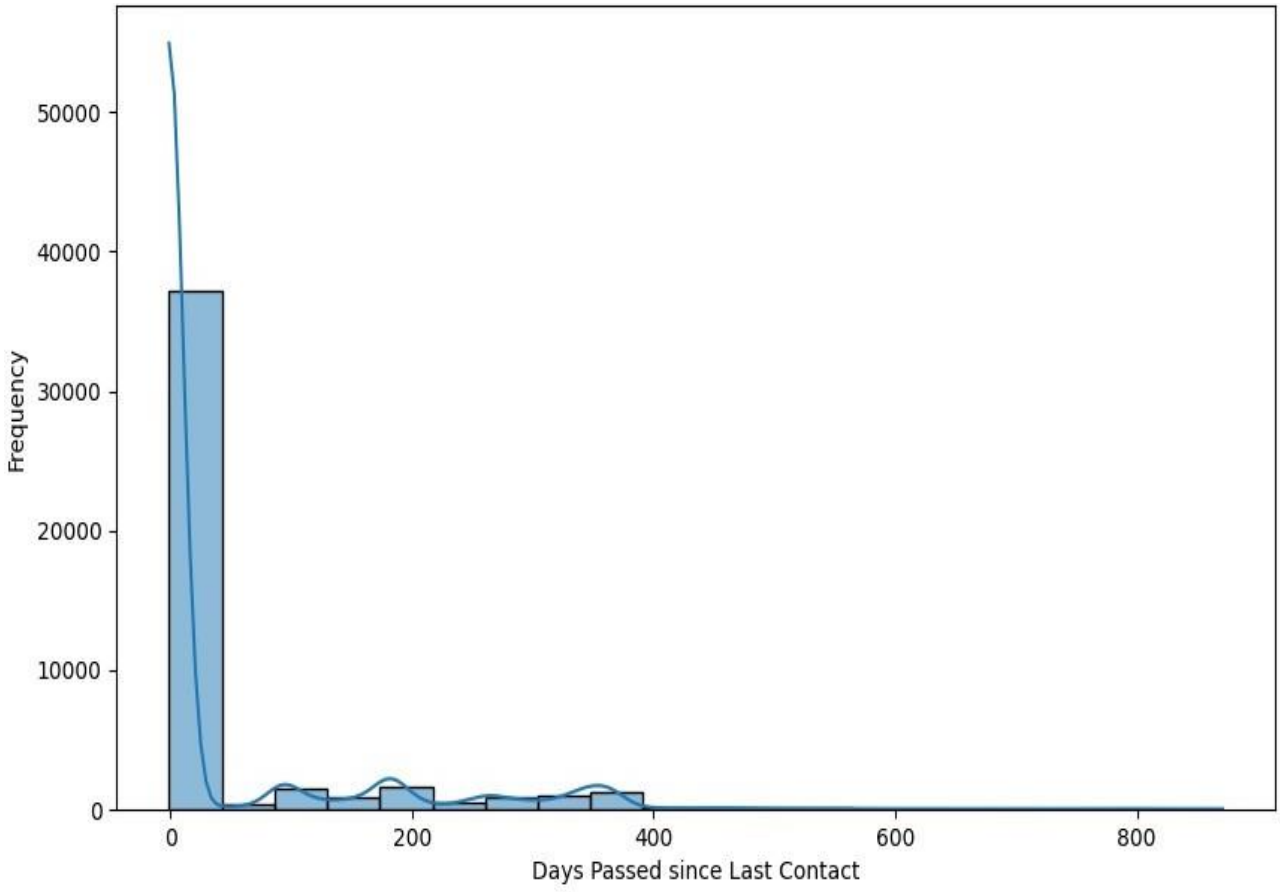
```
campaign
1     17548
2     12506
3      5521
4      3522
5      1764
6      1291
7       735
8       540
9       327
10      266
11      201
12      155
13      133
14       93
15       84
16       79
17       69
18       51
19       44
20       43
21       35
22       23
25       22
23       22
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
44          1
Name: count, dtype: int64
```

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

```
# distribution number of days passed from a previous campaign?

plt.figure(figsize=(10, 6))
sns.histplot(df['pdays'], bins=20, kde=True)
plt.xlabel('Days Passed since Last Contact')
plt.ylabel('Frequency')
```

```
Text(0, 0.5, 'Frequency')
```



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

```
df['previous'].value_counts()
```

```
previous
0      36956
1       2772
2       2106
3       1142
4        715
5        459
6        278
7        205
8        130
9         92
10        67
11        65
12        44
13        38
15        20
14        19
17        15
16        13
19        11
20         8
```

```
23         8
18         6
22         6
24         5
27         5
21         4
29         4
25         4
30         3
38         2
37         2
26         2
28         2
51         1
275        1
58         1
32         1
40         1
55         1
35         1
41         1
Name: count, dtype: int64
```

16. What were the outcomes of the previous marketing campaigns?

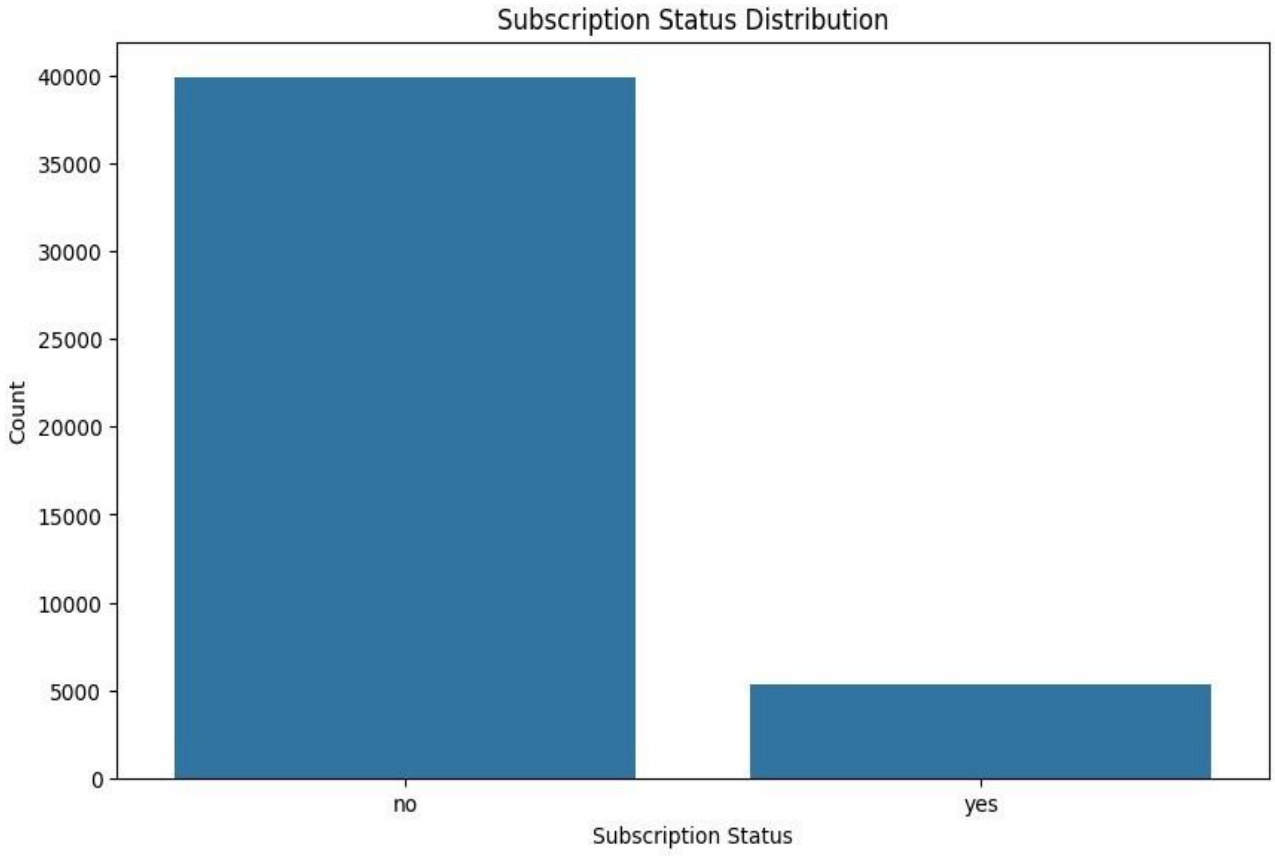
```
df['poutcome'].unique()
```

```
array(['unknown', 'failure', 'other', 'success'], dtype=object)
```

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

```
plt.figure(figsize=(10, 6))
sns.countplot(x='y', data=df)
plt.xlabel('Subscription Status')
plt.ylabel('Count')
plt.title('Subscription Status Distribution')
```

```
Text(0.5, 1.0, 'Subscription Status Distribution')
```



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

```
numeric_columns = df.select_dtypes(include=[np.number])
correlation_matrix = numeric_columns.corr()
plt.figure(figsize=(10, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
```

```
Text(0.5, 1.0, 'Correlation Matrix')
```