Banking Dataset

Term deposits serve as a significant revenue stream for banks, representing cash investments held within financial institutions. These investments involve committing funds for a predetermined period, during which they accrue interest at an agreed-upon rate. To promote term deposits, banks employ various outreach strategies including email marketing, advertisements, telephonic marketing, and digital marketing.

Despite the advent of digital channels, telephonic marketing campaigns persist as one of the most effective means of engaging customers. However, they necessitate substantial investment due to the requirement of large call centers to execute these campaigns. Therefore, it becomes essential to pre-identify potential customers likely to convert, enabling targeted outreach efforts via phone calls.

The data is related to direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe to a term deposit (variable y).

Content

The data is related to the direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed by the customer or not. The data folder contains two datasets:-

Banking_data.csv: 45,211 rows and 18 columns ordered by date (from May 2008 to November 2010)

Detailed Column Descriptions:

age: This column represents the age of the bank client. It's a numeric variable indicating the age in years.

job: This column indicates the type of job the client has. It's a categorical variable with options such as "admin.", "unknown", "unemployed", "management", etc.

marital: This column represents the marital status of the client. It's a categorical variable with options such as "married", "divorced", or "single".

education: This column indicates the level of education of the client. It's a categorical variable with options such as "unknown", "secondary", "primary", or "tertiary".

default: This column indicates whether the client has credit in default. It's a binary variable with options "yes" or "no".

balance: This column represents the average yearly balance in euros for the client. It's a numeric variable.

housing: This column indicates whether the client has a housing loan. It's a binary variable with options "yes" or "no".

loan: This column indicates whether the client has a personal loan. It's a binary variable with options "yes" or "no".

contact: This column represents the type of communication used to contact the client. It's a categorical variable with options such as "unknown", "telephone", or "cellular".

day: This column represents the last contact day of the month. It's a numeric variable.

month: This column represents the last contact month of the year. It's a categorical variable with options such as "jan", "feb", "mar", etc.

duration: This column represents the duration of the last contact in seconds. It's a numeric variable.

campaign: This column represents the number of contacts performed during this campaign and for this client. It's a numeric variable.

pdays: This column represents the number of days that passed by after the client was last contacted from a previous campaign. It's a numeric variable where -1 means the client was not previously contacted.

previous: This column represents the number of contacts performed before this campaign and for this client. It's a numeric variable.

poutcome: This column represents the outcome of the previous marketing campaign. It's a categorical variable with options such as "unknown", "other", "failure", or "success".

```
import seaborn as sns

import matplotlib.pyplot as plt

import pandas as pd

df=pd.read_csv(r'project1.csv')

 pd.set_option('display.max_rows',None)

 pd.set_option("display.max_columns",None)

 print(df)
```

Q1 - What is the distribution of age among the clients?

```
age_1=df["age"].value_counts()

print(age_1)

plt.hist(age_1,color="blue",edgecolor="skyblue")

plt.xlabel("clients")

plt.ylabel("age")

plt.title("Age vs clients")

plt.savefig('graph1.png')

plt.show()
```

Q2 - How does the job type vary among the clients?

```
print(df.columns)

job_1=df["job"].unique()

print(job_1)

plt.hist(job_1,bins=10,color="pink",edgecolor="skyblue")
```

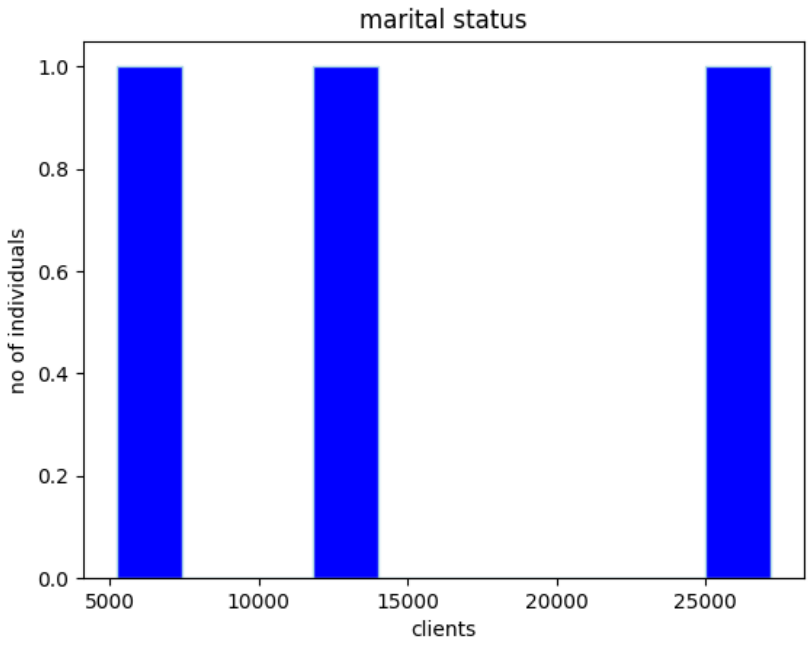
```
plt.xlabel("clients")

plt.ylabel("job")

plt.title("clients vs job")

plt.savefig("graph2.png")

plt.show()
```

Q3 -What is the marital status distribution of the clients?

```
m_status=df["marital_status"].value_counts()

print(m_status)

plt.hist(m_status,bins=10,color="blue",edgecolor="lightblue")
```

```
plt.xlabel("clients")

plt.ylabel("no of individuals")

plt.title("marital status")

plt.savefig("graph3.png")

plt.show()
```

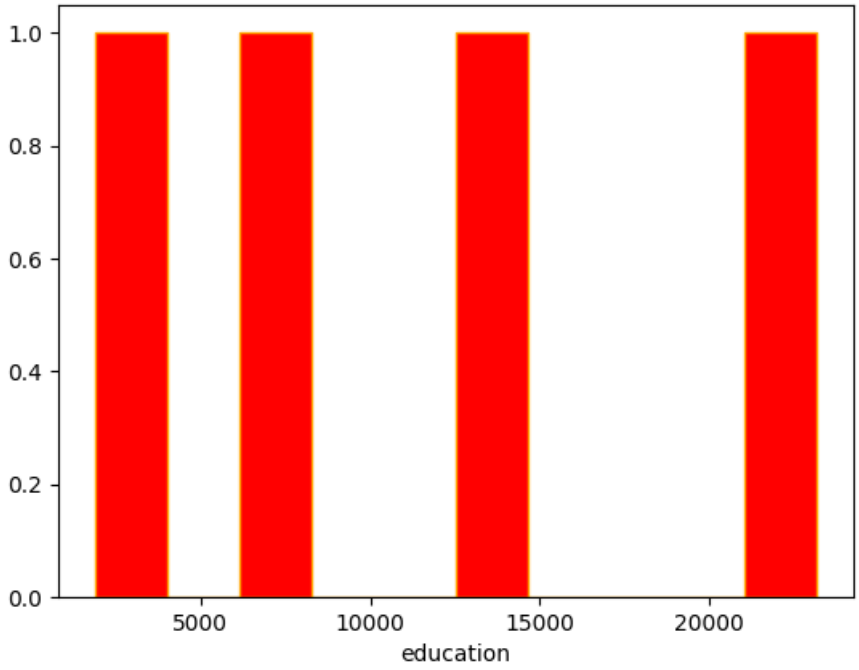
```
marital_status

married     27216

single      12790

divorced     5207
```



Q4 - What is the level of education among the clients?

```
edu=df["education"].value_counts()

print(edu)

plt.hist(edu,bins=10,color="red",edgecolor="orange")

plt.xlabel("education")

plt.savefig("graph4.png")

plt.show()
```

```
education

secondary    23204

tertiary     13301

primary       6851

unknown       1857

Name: count, dtype: int64
```

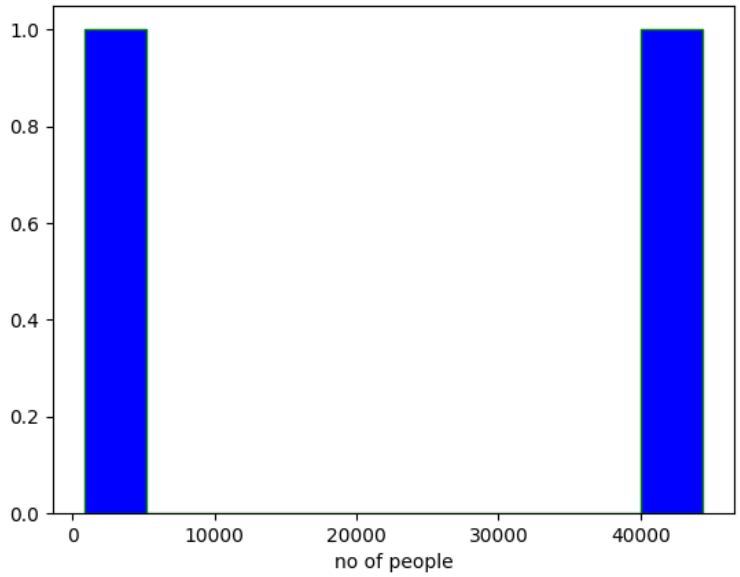


Q5 - What proportion of clients have credit in default?

```
credit_customer=df["default"].value_counts()
print(credit_customer)
plt.hist(credit_customer,bins=10,color="blue",edgecolor="green")
plt.xlabel("no of people")
plt.savefig("graph5.png")
plt.show()
```

```
default
no     44401
yes      815
```

Name: count, dtype: int64



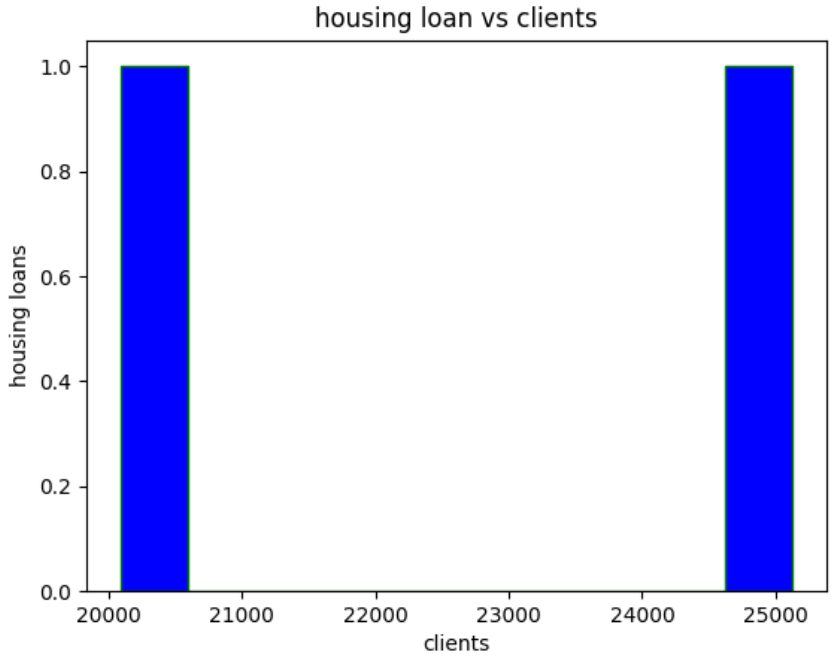
Q6 - -What is the distribution of average yearly balance among the clients

```
yearly_balance=df["balance"].value_counts()
print(yearly_balance)
plt.hist(yearly_balance,bins=10,color="yellow")
plt.xlabel("yearly balance")
plt.ylabel("clients")
plt.title("yearly balance")
plt.savefig("graph6.png")
plt.show()
```

Q7 - - How many clients have housing loans?

```
print(df.columns)

loan=df["housing"].value_counts()

print(loan)

plt.hist(loan,bins=10,color="blue",edgecolor="green")

plt.xlabel("clients")

plt.ylabel("housing loans")

plt.title("housing loan vs clients")

plt.savefig("graph7.png")

plt.show()
```

```
yes    25130

no     20086

Name: count, dtype: int64
```



Q8 How many clients have personal loans?

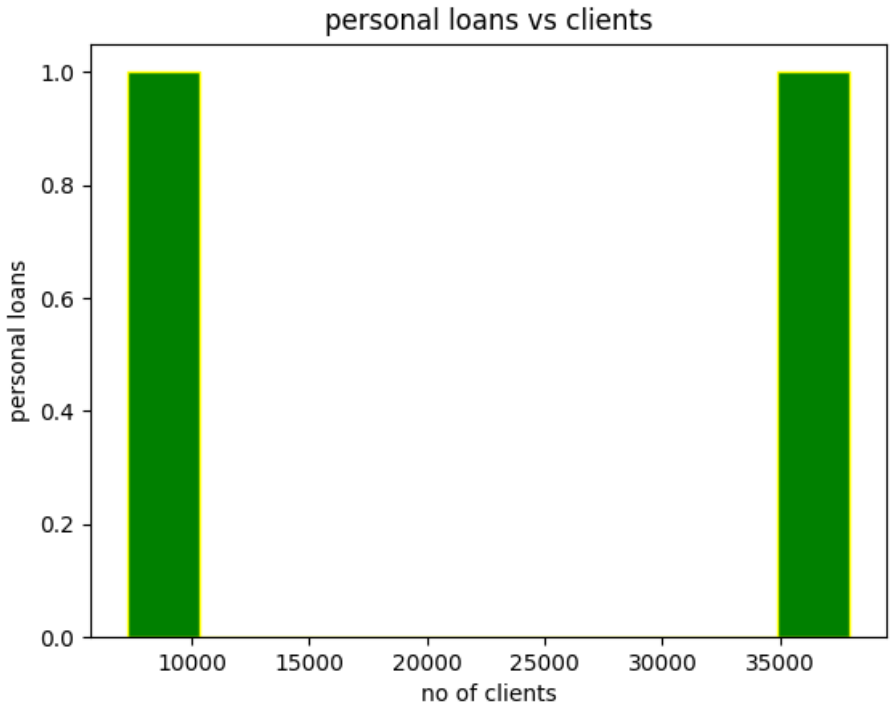
```
loan_1=df["loan"].value_counts()

print(loan_1)

plt.hist(loan_1,bins=10,color="green",edgecolor="yellow")

plt.xlabel("no of clients")

plt.ylabel("personal loans")

plt.title("personal loans vs clients")

plt.savefig("graph8.png")

plt.show()
```

loan

no 37972

yes 7244

Name: count, dtype: int64

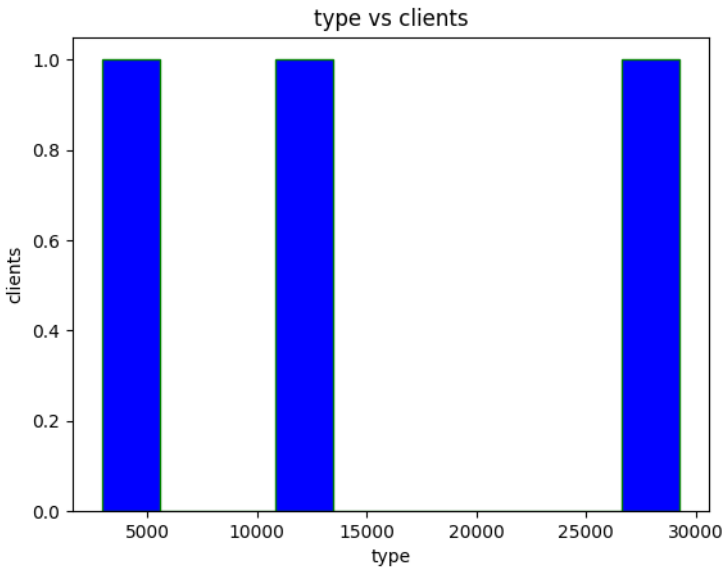


Q9 What are the communication types used for contacting clients during the campaign?

```
type=df["contact"].value_counts()

print(type)

plt.hist(type,bins=10,color="blue",edgecolor="green")
```

```
plt.xlabel("type")

plt.ylabel("clients")

plt.title("type vs clients")

plt.savefig("graph9.png")

plt.show()
```

```
contact

cellular     29290

unknown      13020

telephone     2906

Name: count, dtype: int64
```



Q10 What is the distribution of the last contact day of the month?

```
day_1=df["day"].value_counts()

print(day_1)

plt.hist(day_1,bins=10,color="blue",edgecolor="green")

plt.xlabel("clients")

plt.ylabel("day")

plt.title("day vs clients")

plt.savefig("graph10.png")

plt.show()
```

```
day
20    2752
18    2308
21    2026
17    1942
6     1932
5     1910
14    1848
8     1842
28    1830
7     1817
```

19 1757

29 1745

15 1703

12 1603

13 1585

30 1566

9 1561

11 1479

4 1445

16 1417

2 1293

27 1121

3 1079

26 1035

23 939

22 905

25 840

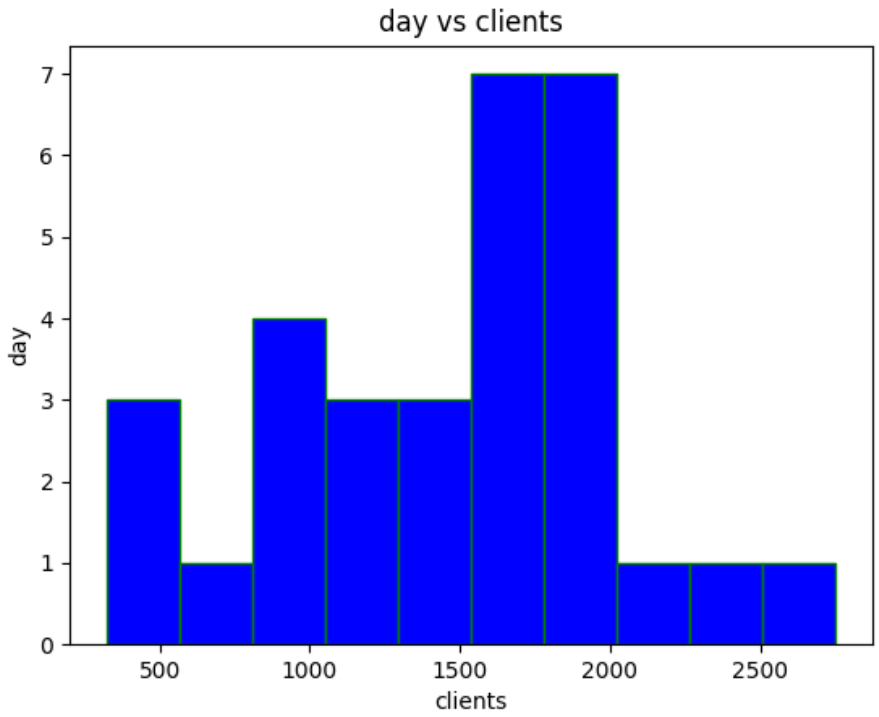
31 643

10 524

24 447

1 322

Name: count, dtype: int64



Q11 How does the last contact month vary among the clients?

```
month1=df["month"].value_counts()
print(month1)
plt.hist(month1,bins=10,color="orange",edgecolor="green")
plt.xlabel("clients")
plt.ylabel("month")
plt.title("month vs clients")
plt.savefig("graph11.png")
plt.show()
```

month

may 13766

jul 6895

aug 6247

jun 5341

nov 3975

apr 2932

feb 2649

jan 1403

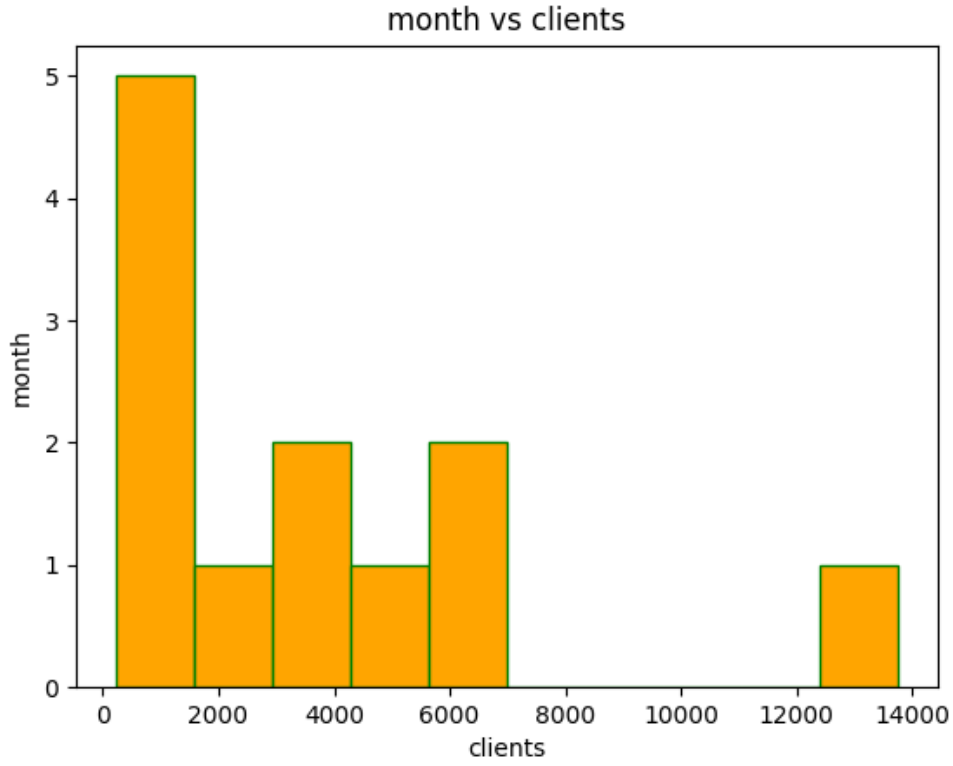
oct 738

sep 579

mar 477

dec 214

Name: count, dtype: int64



Q12 What is the distribution of the duration of the last contact?

```
# dur=df["duration"].value_counts()

# print(dur)

# plt.hist(dur,bins=10,color="violet",edgecolor="green")

# plt.xlabel("clients")

# plt.ylabel("duration")

# plt.title("clients vs duration ")

# plt.savefig("graph12.png")

# plt.show()
```

```
duration

124    188

90     184
```

89 177

104 175

122 175

...

1833 1

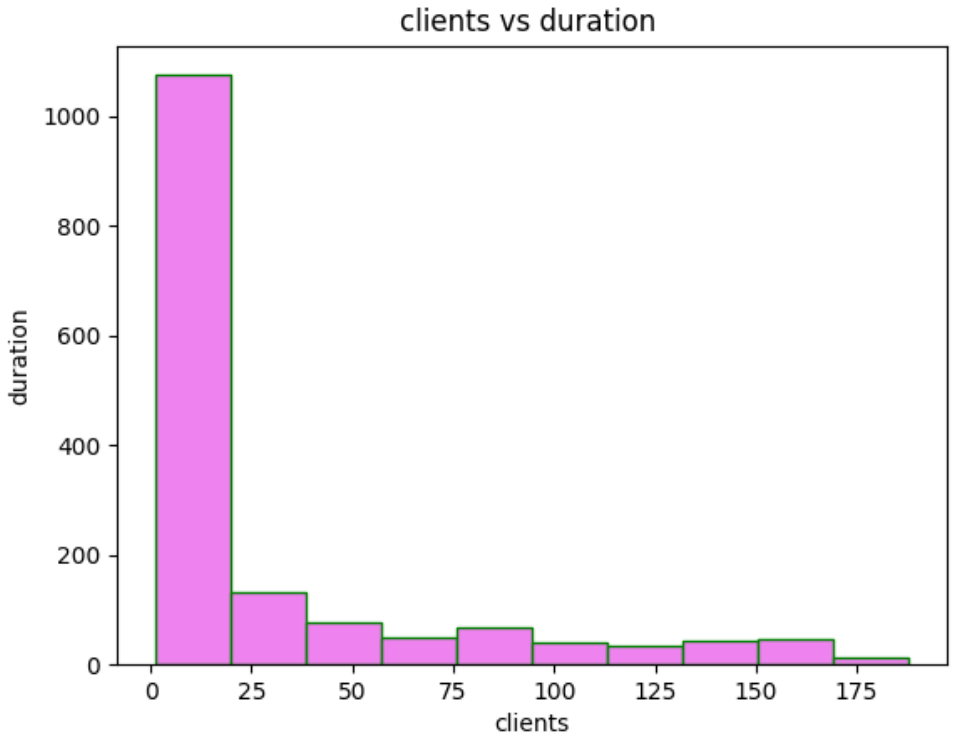
1545 1

1352 1

1342 1

1556 1

Name: count, Length: 1573, dtype: int64



Q13 How many contacts were performed during the campaign for each client?

```
camp=df["campaign"].value_counts()
print(camp)
plt.hist(camp,bins=10,color="orange",edgecolor="green")
plt.xlabel("clients")
plt.ylabel("campaign")
plt.title("campaign vs clients")
plt.savefig("graph13.png")
plt.show()
```

```
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

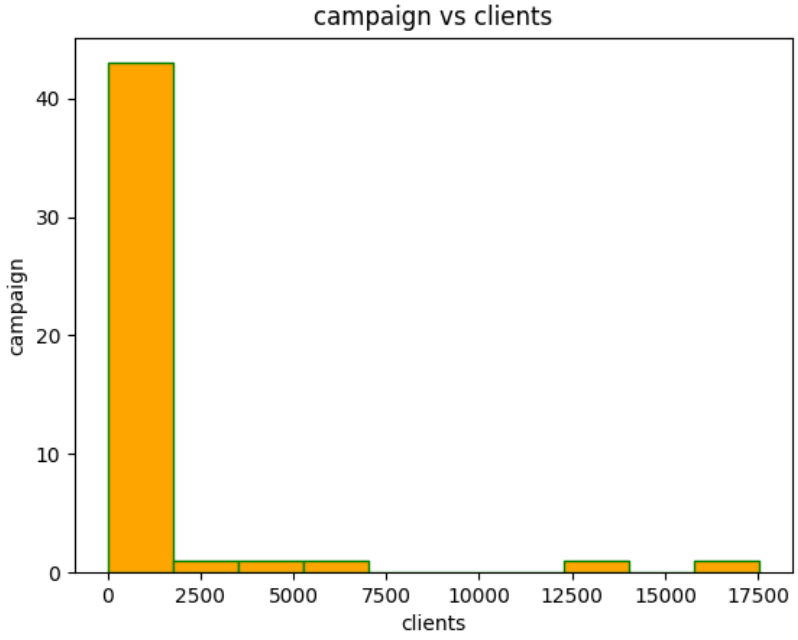
63 1

51 1

39 1

44 1

Name: count, dtype: int64



Q14 What is the distribution of the number of days passed since the client was last contacted from a previous campaign?

```
days=df["pdays"].value_counts()

print(days)

plt.hist(days,bins=10,color="pink",edgecolor="green")

plt.xlabel("clients")

plt.ylabel("duration")

plt.title("clients vs duration")

plt.savefig("graph14.png")

plt.show()
```

pdays

-1 36956

182 167

92 147

91 126

183 126

...

32 1

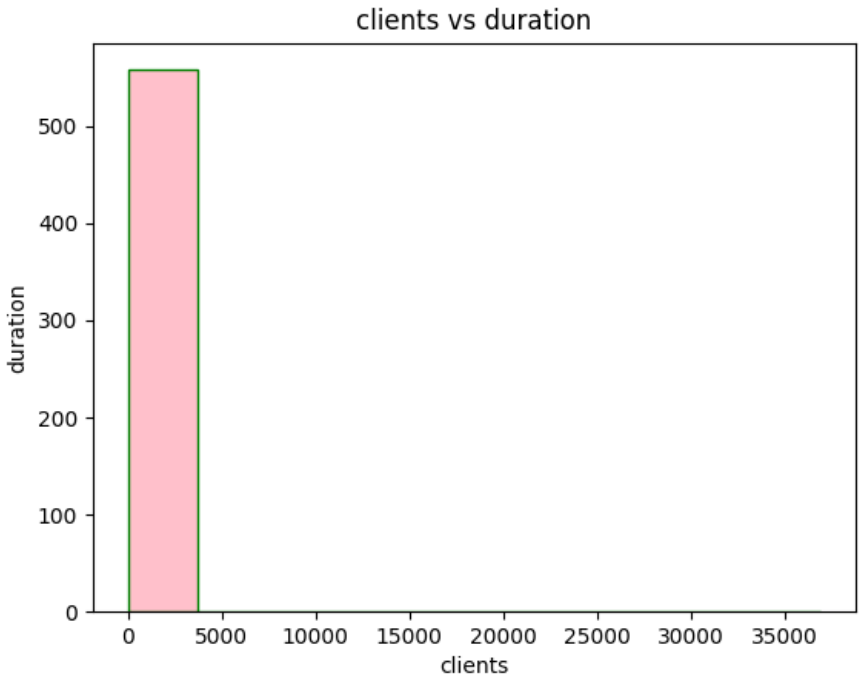
551 1

670 1

65 1

530 1

Name: count, Length: 559, dtype: int64



Q15 How many contacts were performed before the current campaign for each client?

```
contact=df["previous"].value_counts()

print(contact)

plt.hist(contact,bins=10,color="darkgreen",edgecolor="green")

plt.xlabel("clients")

plt.ylabel("previous")

plt.title("previous vs clients")

plt.savefig("graph16.png")

plt.show()
```

```
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

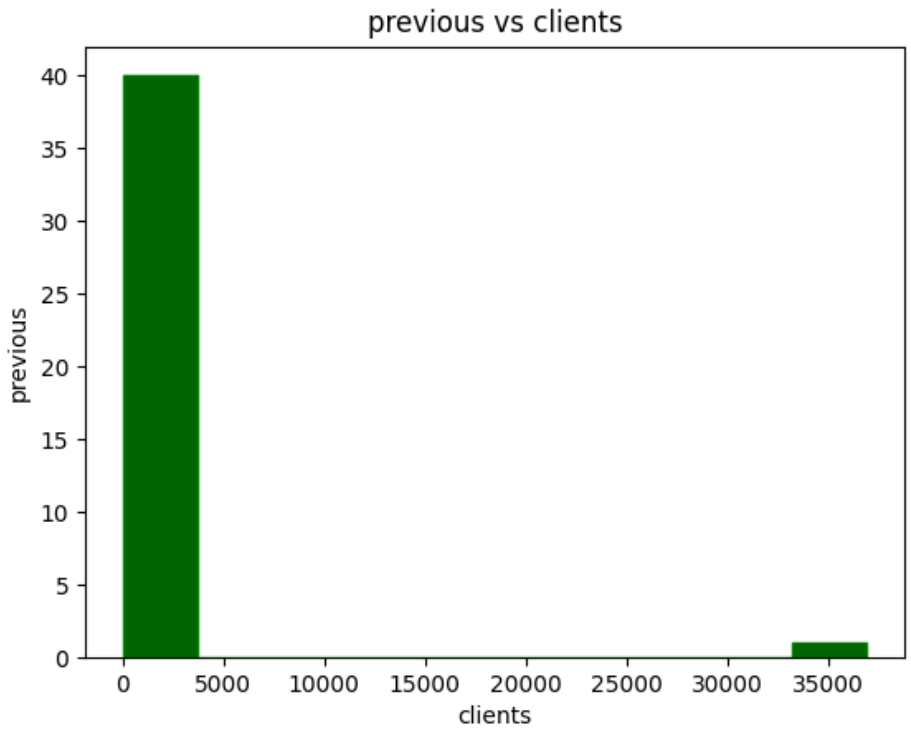
40 1

55 1

35 1

41 1

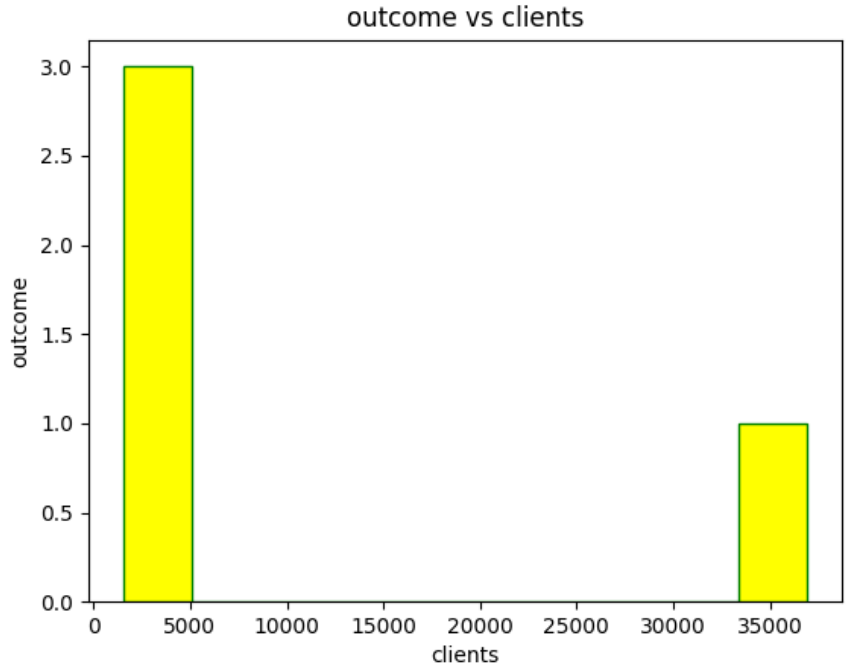
Name: count, dtype: int64



Q16 What were the outcomes of the previous marketing campaigns?

```
outcome=df["poutcome"].value_counts()
print(outcome)
plt.hist(outcome,bins=10,color="yellow",edgecolor="green")
plt.xlabel("clients")
plt.ylabel("outcome")
plt.title("outcome vs clients")
plt.savefig("graph17.png")
plt.show()
```

```
poutcome
unknown    36961
failure     4902
other       1840
success     1513
Name: count, dtype: int64
```



Q17 What is the distribution of clients who subscribed to a term deposit vs. those who did not?

```
term_deposit=df["y"].value_counts()
print(term_deposit)
plt.hist(term_deposit,bins=10,color="blue",edgecolor="green")
plt.xlabel("clients")
plt.ylabel("term deposit")
plt.title("term deposit vs clients")
plt.savefig("graph18.png")
plt.show()
```

y

no 39922

yes 5294

Name: count, dtype: int64