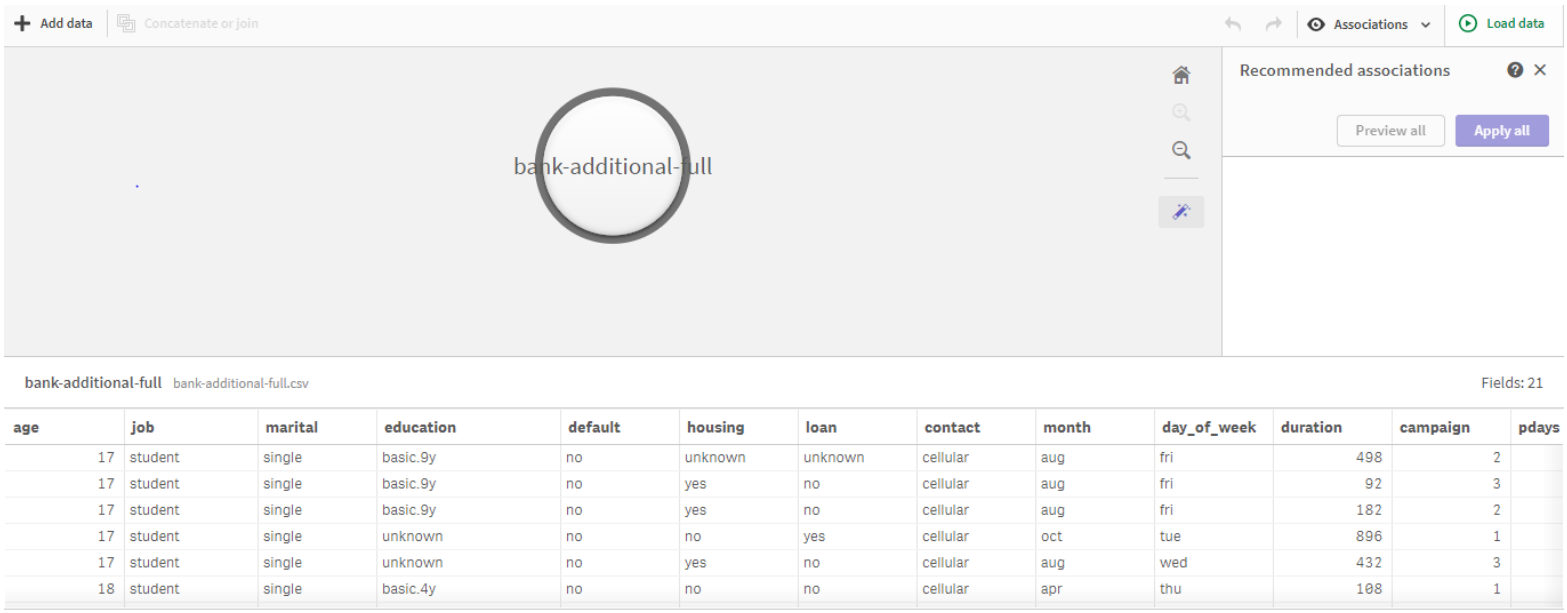
Business Intelligence – Qlik Assignment

Name: Asma Abid Karim

ERP ID; 19685

Step 1: Select and upload any ONE of the following datasets.

https://www.kaggle.com/datasets/volodymyrgavrysh/bank-marketing-campaigns-dataset



bank-additional-full bank-additional-full.csv Fields: 21

age	job	marital	education	default	housing	loan	contact	month	day_of_week	duration	campaign	pdays
17	student	single	basic.9y	no	unknown	unknown	cellular	aug	fri	498	2	
17	student	single	basic.9y	no	yes	no	cellular	aug	fri	92	3	
17	student	single	basic.9y	no	yes	no	cellular	aug	fri	182	2	
17	student	single	unknown	no	no	yes	cellular	oct	tue	896	1	
17	student	single	unknown	no	yes	no	cellular	aug	wed	432	3	
18	student	single	basic.4y	no	no	no	cellular	apr	thu	108	1	

Here the dataset is uploaded to Qlik cloud service accessed through the academic license. The only table we have here is bank-additional-full. This data is based on 21 columns and 41188 rows. The information is related to bank marketing campaigns for opening a term deposit.

Step 2: Acquire domain knowledge (if you don't have) and then write a problem statement which you will solve using BI (just one liner).

> The dataset describes the Portugal bank marketing campaigns results. Most of the efforts of offering the opening of new term deposit were conducted over the phone, with bank clients being offered the opportunity to put a term deposit. If the client consented to make a deposit after all the marking efforts, the goal variable is set to 'yes,' otherwise to 'no.'

Problem statement:

We analyze overall what was the total of a positive response. Then further link this positive response to the factors which are affecting these answers like duration, job, loan etc. The same insights process is mapped for the negative response and try to reach the major factors affecting that answer. Reading into the factors affecting the positive or negative answers will let us find out who should be targeted for the campaign.

So the problem here can be simply summarized to 'Who should be target of next campaign keeping in mind the result of the previous – Most offers should be made were the prediction of a answer is yes depending on other factors'.

Step 2: Understanding the data and renaming the columns

Client personal Data stored with bank:

1. age
    - Numeric column
    - Shows the age of the client
    - **Renaming age to Client_age**

    - *Description:* The age of the client may play an important role, since the term deposit, as our common sense say will be close to the mind of a young adult rather than a person who is his/her late sixties or a person younger than eighteen or twenty.
2. job
    - categorical column
    - Type of job the client does.
    - categories available
        - "admin."
        - "blue-collar"
        - "entrepreneur"
        - "housemaid"
        - "management"
        - "retired"
        - "self-employed"
        - "services"
        - "student"
        - "technician"
        - "unemployed"
        - "unknown"

    - **Renaming job to Client_job**
    - *Description:* This can also play a part as the source of income is the person's job mostly.

3. marital
    - categorical column
    - Describes the marital status of the client.

- categories available
    - "divorced" - means divorced or widowed
    - "married"
    - "single"
    - "unknown"
- **Renaming marital to Marital_status**

- *Description:* Mostly people make term deposit - to save money. These money saving is for the people attached to the client. So, if a person is married there are more chance of a term deposit. But there is a downside to this, is the person is married, means the most part of income is already being used which in return decrease the possible of the term deposit.

4. education
    - categorical column
    - illustrates the highest education or qualification of client
    - categories available
        - "basic.4y"
        - "basic.6y"
        - "basic.9y"
        - "high.school"
        - "illiterate"
        - "professional.course"
        - "university.degree"
        - "unknown"
    - **Renaming education to Education**

    - *Description:* The education affects the thought process of the person. It also has effect on the type of job and the amount of income of a client in general. People with different education lines will have a different in-build perspective of a term deposit.

5. default
    - categorical column
    - does client have credit in default?
    - categories available
        - "no"
        - "yes"
        - "unknown"

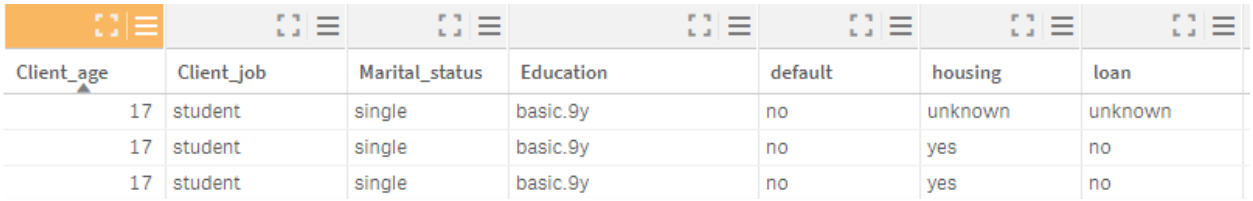
    - *Description:* The person in the default list means he/she is already juggling with finances and have no money in abundance for a term deposit.

6. housing
    - categorical column
    - Does client have a housing loan?
    - categories available
        - "no"
        - "yes"
        - "unknown"

- *Description:* If the person has a house loan means he/she is already juggling with finances and have no money in abundance for a term deposit.

7. loan
    - categorical column
    - Does client have a personal loan?
    - categories available
        - "no"
        - "yes"
        - "unknown"

    - *Description:* If the person has a personal loan means he/she is already juggling with finances and have no money in abundance for a term deposit.



Client_age	Client_job	Marital_status	Education	default	housing	loan
17	student	single	basic.9y	no	unknown	unknown
17	student	single	basic.9y	no	yes	no
17	student	single	basic.9y	no	yes	no

Information on the last contact for the current campaign with the client:

8. contact
    - categorical column
    - type of communication
    - categories available
        - "cellular"
        - "telephone"
    - **Renaming contact to Type_of_contact**

9. month
    - categorical column\
    - last contact month of year
    - categories available
        - ("jan", "feb", "mar", …, "nov", "dec")

    - **Renaming month to Last_month_of_contact**

10. day*of*week
    - categorical column
    - last contact day of the week

- categories available
  - "mon"
  - "tue"
  - "wed"
  - "thu"
  - "fri"

11. duration
    - numeric column
    - last contact duration, in seconds
    - this attribute highly affects the output target (e.g., if duration=0 then y="no").
    - Yet, the duration is not known before a call is performed. Also, after the end of the call Client_answer is obviously known.
    - **Renaming duration to Last_contact_duration**

Description: This a matter of importance, when was the campaign last time offered to the client. How the client was approached. Most people are very comfortable with a non-personal approach. If a telephone was used, then a person is not every time available on it. Cellular is a better option in that sense. So, means of contact is a matter of debate and should be dealt carefully. When, in which month and day of week the contact was last made. Another important factor is the duration of the contact. With the amount of duration, we can gauge a farther reaction. Like if the client listens till the end, then he/she was quite interested. Many more such assumption can be made based on the duration.

Other Attributes:

12. campaign
    - numeric column
    - number of contacts performed during this campaign and for this client.
    - **Renaming campaign to Client_contact_per_campaign**
13. pdays
    - numeric column
    - number of days that passed by after the client was last contacted from a previous campaign
    - 999 means client was not previously contacted

14. Previous
    - Numeric column
    - number of contacts performed before this campaign and for this client
    - **Renaming previous to Client_contact_before_campaign**

15. poutcome
    - Categorical column
    - outcome of the previous marketing campaign
    - categories available
      - "failure"
      - "nonexistent"
      - "success"

Attributes in context of Economy and society(social):

16 . emp.var.rate

- Numeric column
- employment variation rate - quarterly indicator.

17 cons.price.idx

- numeric column
- consumer price index - monthly indicator

18. cons.conf.idx

- numeric column
- consumer confidence index - monthly indicator

19. euribor3m

- numeric column
- euribor 3 month rate - daily indicator

20. nr.employed

- numeric column
- number of employees - quarterly indicator

21. y

- binary column
- has the client subscribed a term deposit?
- Options are: "yes" or "no"
- **Renaming y to Client_answer**

Last_month_...	day_of_week	Last_contact...	Client_conta...	pdays	Client_conta...	poutcome	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed	Client_answer
aug	fri	498	2	999	1	failure	-2.9	92.201	-31.4	0.869	5076.2	yes
aug	fri	92	3	4	2	success	-2.9	92.201	-31.4	0.869	5076.2	no
aug	fri	182	2	999	2	failure	-2.9	92.201	-31.4	0.869	5076.2	no
oct	tue	896	1	2	2	success	-3.4	92.431	-26.9	0.742	5017.5	yes

Description: Other attributes can also play a role. Like the number of contacts for the same offer or for another offer this can lead to irritation in the client. Have the client ever accepted the previous campaign's offers. In the same way the social and economic variable like the rate of employment, price indexes and confidence can have large impact on the results of the campaign.

Missing values:

There are several missing values in some categorical attributes, all coded with the "unknown" label. These missing values can be treated as a possible class label.

Step 4: List down dimensions (with different types as done in class) along with KPIs (to be extracted from transformed data)

Columns

- Client_age
- Client_job
- Marital_status
- Education
- default
- housing
- loan
- Type_of_contact
- Last_month_of_contact
- dayofweek
- Last_contact_duration
- Client_contact_per_campaign
- pdays
- Client_contact_before_campaign
- poutcome
- emp.var.rate
- cons.price.idx
- cons.conf.idx
- euribor3m
- nr.employed
- Client_answer

Dimensions:

- Normal
  - Client_age
  - Client_job
  - Marital_status
  - Education
  - default
  - housing
  - loan
  - Type_of_contact

- Time:
  - Last_month_of_contact
  - dayofweek
  - Last_contact_duration

- Economic factor
    - emp.var.rate
    - cons.price.idx
    - cons.conf.idx
    - euribor3m
    - nr.employed
- Previous Campaign and contact
    - Client_contact_per_campaign
    - pdays
    - Client_contact_before_campaign
    - Poutcome

Measure:

- Client_answer

Step 5: Some potential Analysis, Question to be answered

Questions:

Which dimensions affected the client answer negatively and which have a positive effect and why?

First the major analysis for the yes part of the data and the attributes affecting it. Then we come to the no part.

Who should be offered the term deposit? What are their properties. What these people might have in common.

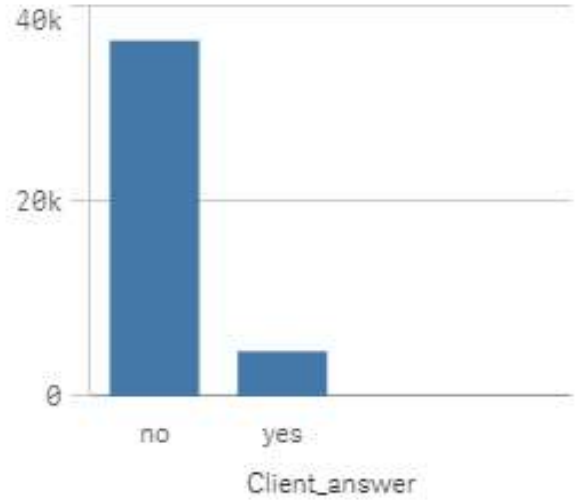
Mostly who say no? Offering to these types of people should be minimum.


Chart And Analysis

Client Answer

Click sheet to make selections | Reset selections | Go to sheet

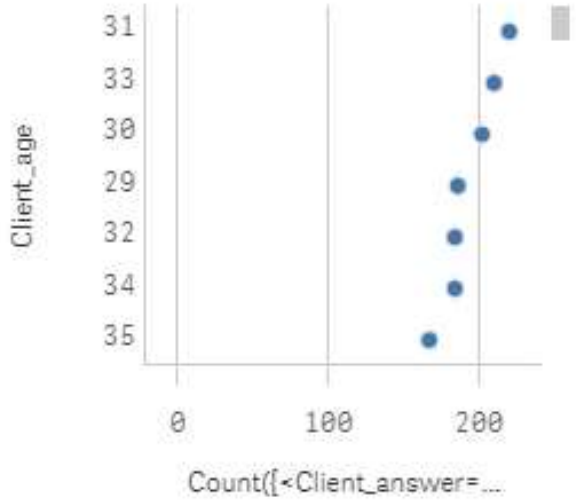
Client Answer count



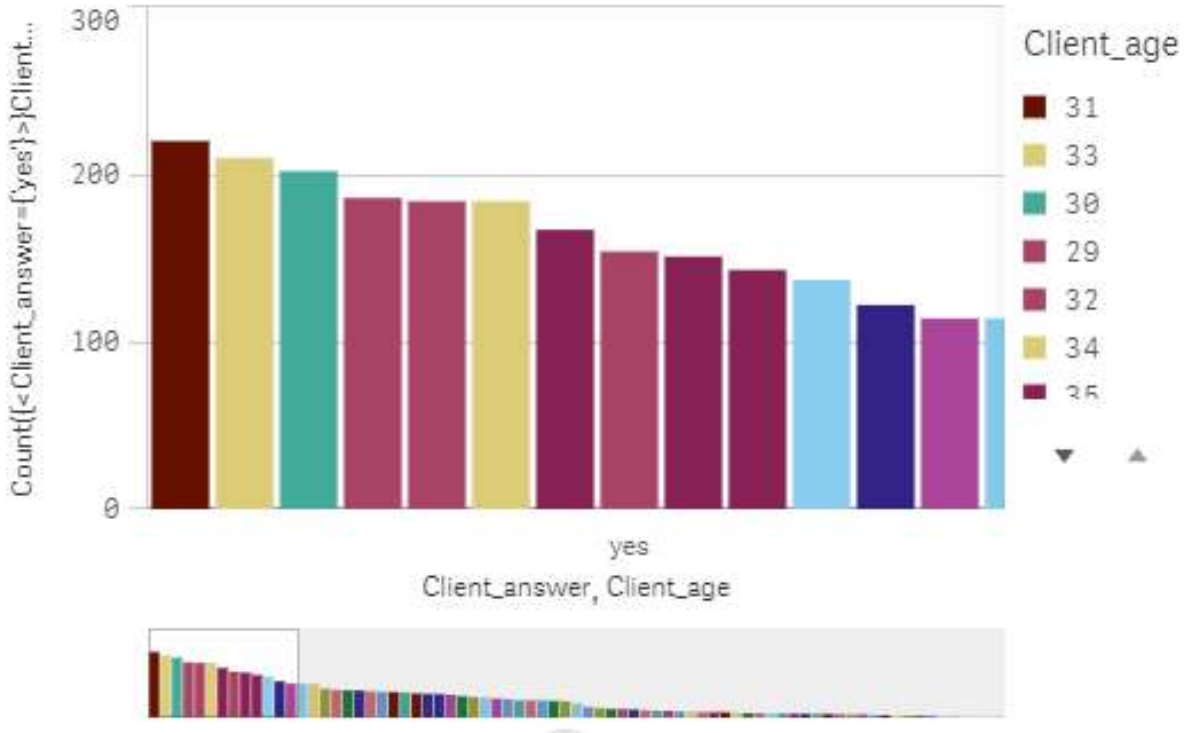
Client Answer count absolute numbers

Client_answer	Count(Client_answer)
Totals	**41188**
no	36548
yes	4640

Client age with alient answer



Client age when answer yes



In here we see how many positive and negative answers we got. From here we try to focus on factors on which we got positive answers. In this sheet we build a connection between client's age and if he/she wants a term deposit. Most people to get the term deposit are from 30 and above. Which is a good representation of whom should be offered.

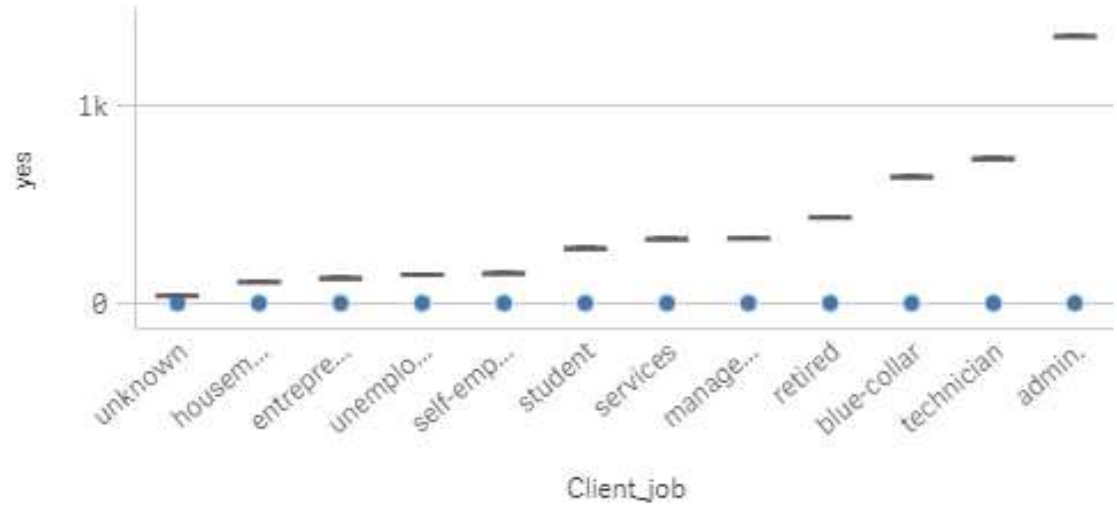
Client Job

Click sheet to make selections | Reset selections | Go to sheet

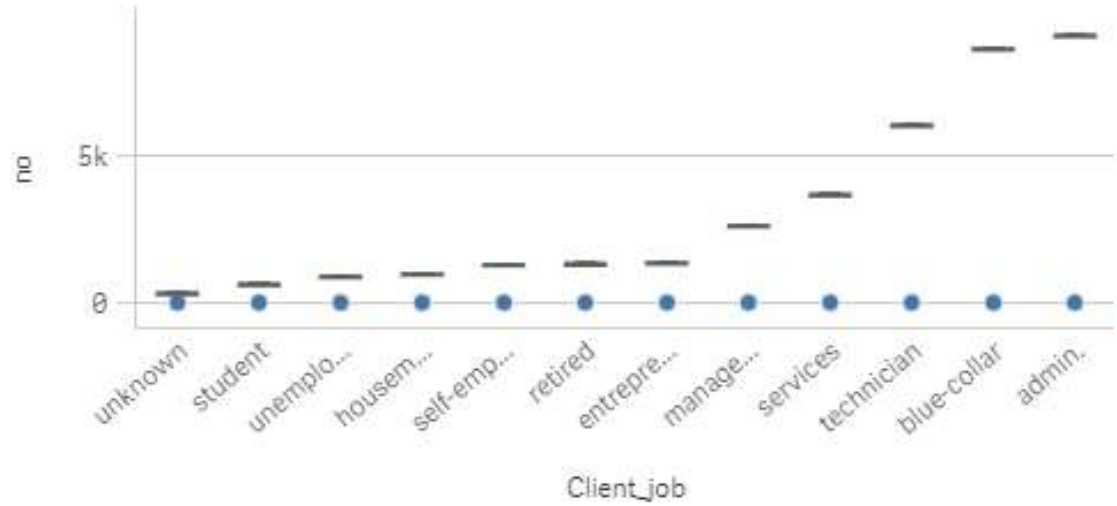
Client answer and there job

Client_...	[Client_job]	Count(...
Totals		**41188**
no	admin.	9070
no	blue-collar	8616
no	entrepreneur	1332
no	housemaid	954
no	management	2596
no	retired	1286
no	self-employed	1272
no	services	3646
no	student	600
no	technician	6013
no	unemployed	870
no	unknown	293
yes	admin.	1352
yes	blue-collar	638
yes	entrepreneur	124
yes	housemaid	106
yes	management	328
yes	retired	434
yes	self-employed	149

client saying yes per job type



client saying no per job type

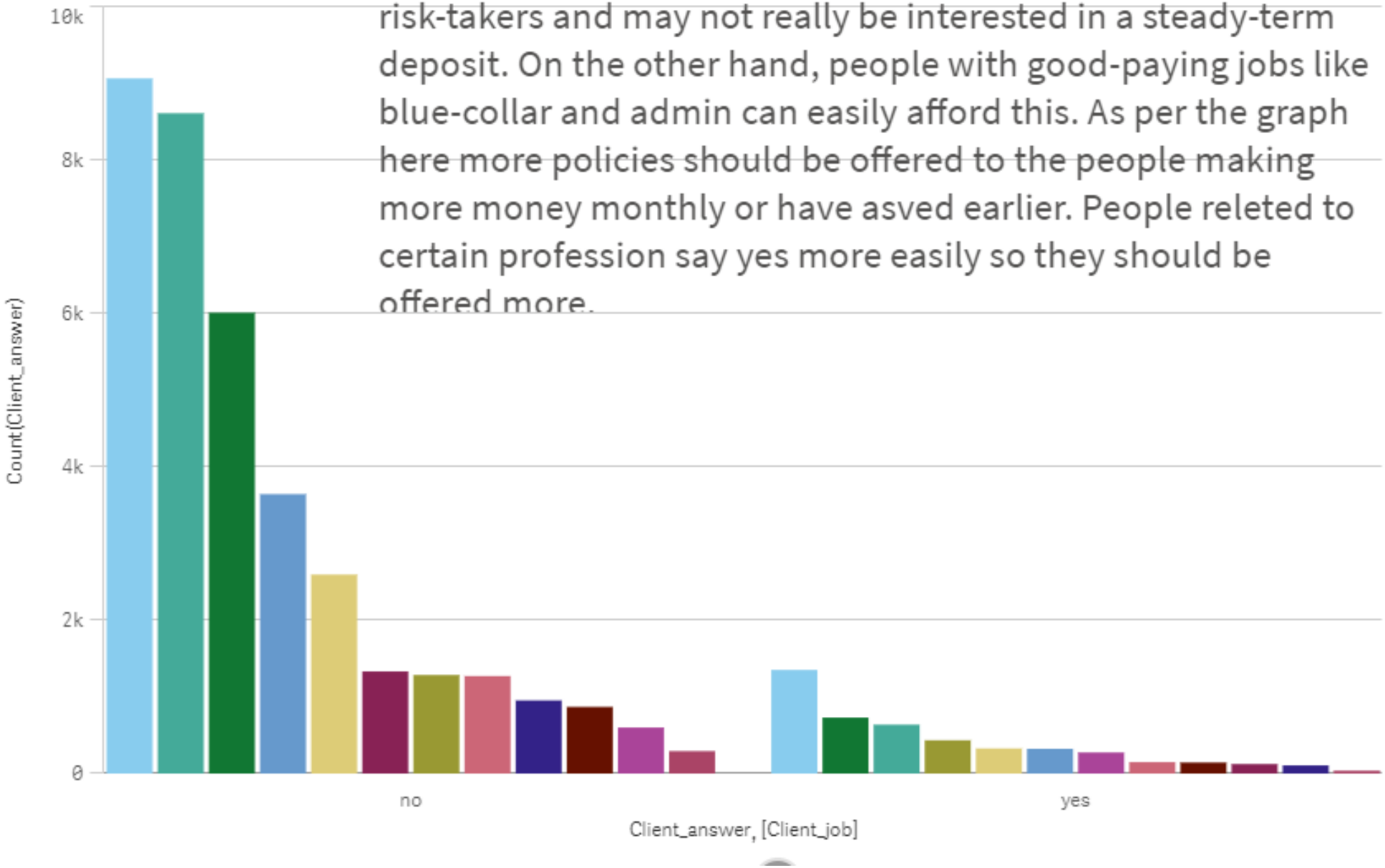


The profession or job of a client highly affect the answer to a term deposit offer. Now as the graph shows housemaids, self-employed, and even entrepreneurs have a low yes rate. while blue-collar and admin people have a high yes level.
We can assume a lot of common thinking related to these answers. Like housemaid may not have soo much every month to save. For self-employed same amount is majorly not possible.

Client Job Contd..

Click sheet to make selections | Reset selections | Go to sheet

An entrepreneur may have a low yes rate because they are risk-takers and may not really be interested in a steady-term deposit. On the other hand, people with good-paying jobs like blue-collar and admin can easily afford this. As per the graph here more policies should be offered to the people making more money monthly or have asved earlier. People releted to certain profession say yes more easily so they should be offered more.



we see the same demonstration for both yes and no with their professions.
This slide here also shows the people with yes and no. The colors show the jobs they do. Like it shows services people also have good contribution in yes.

Client Education

Click sheet to make selections | Reset selections | Go to sheet

Client answer there education

Client_answ...	Education	Count(Client_an...
Totals		**41188**
no	basic.4y	3748
no	basic.6y	2104
no	basic.9y	5572
no	high.school	8484
no	illiterate	14
no	professional.course	4648
no	university.degree	10498
no	unknown	1480
yes	basic.4y	428
yes	basic.6y	188
yes	basic.9y	473
yes	high.school	1031
yes	illiterate	4
yes	professional.course	595
yes	university.degree	1670
yes	unknown	251

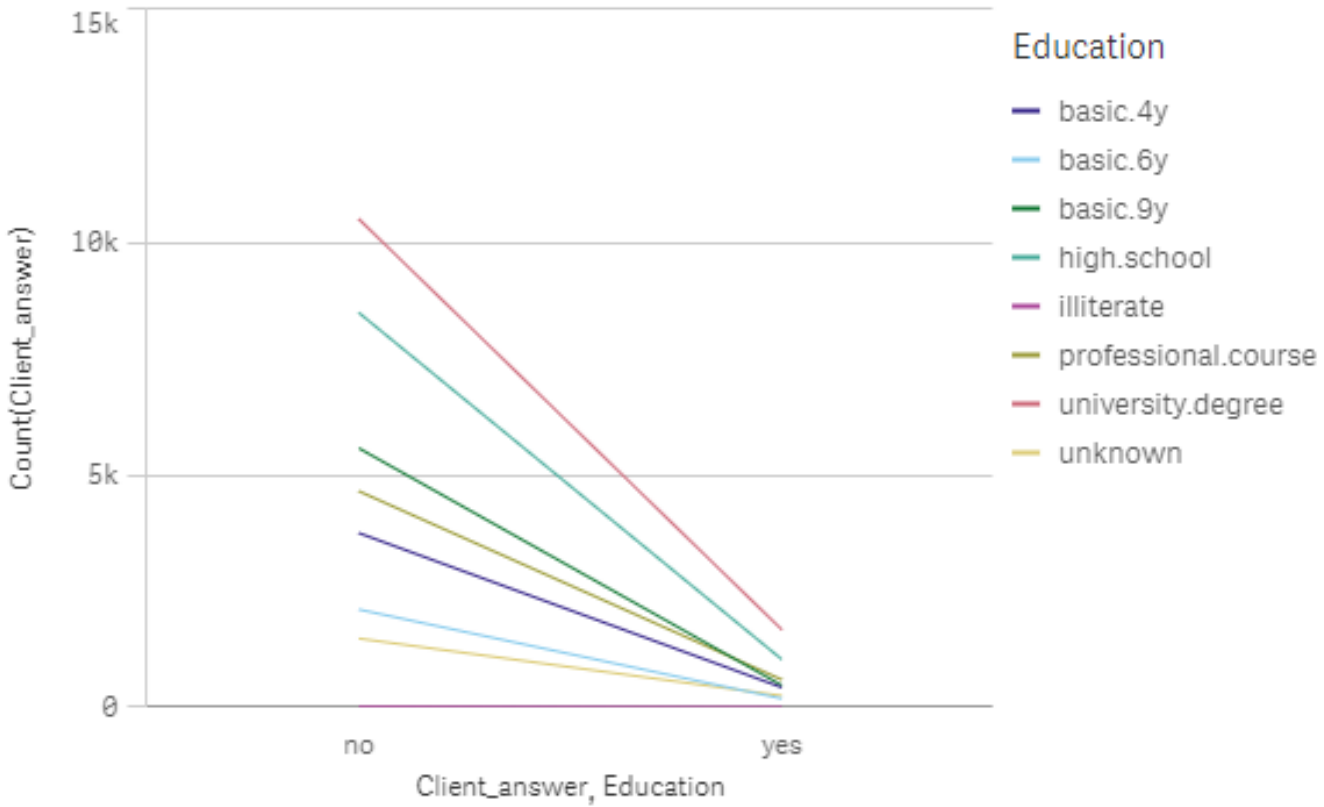
client saying yes per education

client saying no per education

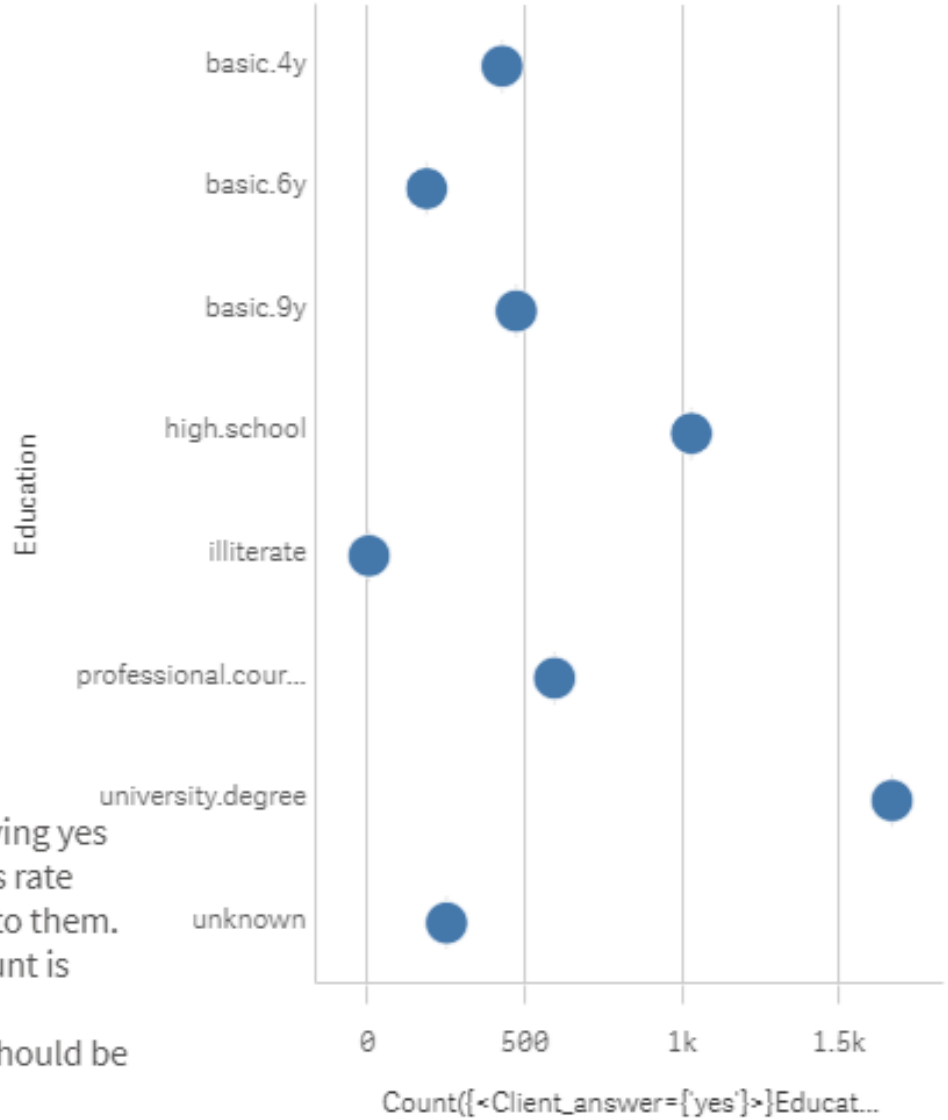
Client Education Contd..

Click sheet to make selections | Reset selections | Go to sheet

Client count saying yes and no per education



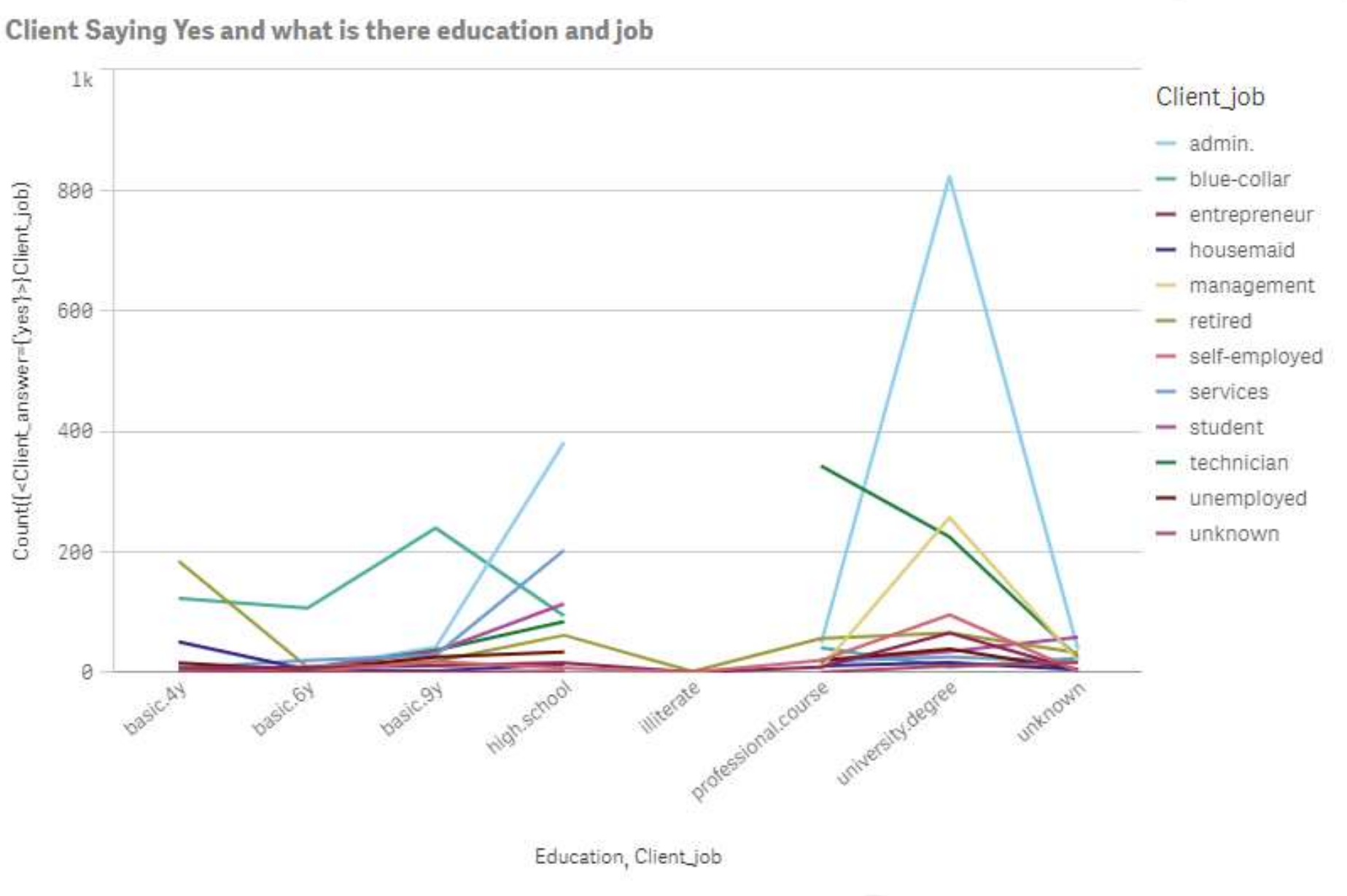
Client saying yes as per education - distribution plot



The education of a client also shapes his/her answer to the term deposit. The people saying yes more often have a university degree, or have done high school on the other hand low yes rate comes from basic 4years of study. This is definitely not because fewer offers were made to them. The yes and no have also the same treemap and line graph pattern because the total count is dominated by a no in the campaign. But if we compare yes coming from high educated individuals and less educated individuals then we see the people with more education should be made more offers.

Client Education Contd ...

Click sheet to make selections | Reset selections | Go to sheet



This shows the relation between the client's job, client education, and their answers. So the count of a maximum number of people saying yes is university degree holders having a position of admins. So basically people like these should be targeted more.
Then at every distinct education level, we have a distinct job and their count of yes matches previous findings too.
Technicians and management people are also a good target.

Client answer with default loan, housing loan

Click sheet to make selections | Reset selections | Go to sheet

Client answer count -default

Client_answer	default	Count(Client_answer)
Totals		**41188**
no	no	28391
no	unknown	8154
no	yes	3
yes	no	4197
yes	unknown	443

Client answer count -housing loan

Client_answer	housing	Count(Client_answer)
Totals		**41188**
no	no	16596
no	unknown	883
no	yes	19069
yes	no	2026
yes	unknown	107
yes	yes	2507

Client answer count -loan

Client_answer	loan	Count(Client_answer)
Totals		**41188**
no	no	30100
no	unknown	883
no	yes	5565
yes	no	3850
yes	unknown	107
yes	yes	683

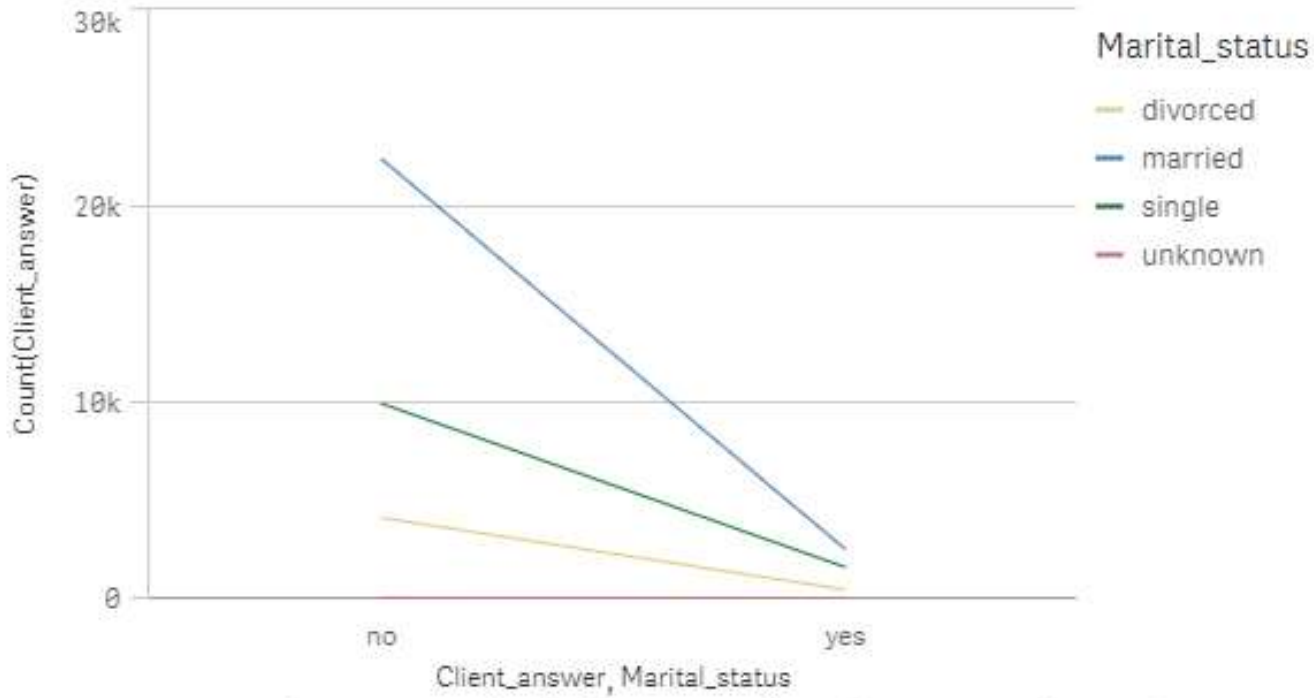
Here we see factors like default, loans, and housing loans. The first table represents the default and client answers. we see people having defaulted and not saying yes to a term deposit. This definitely makes sense. and gives us the understanding that the offer should not be paid to the people having defaults. For people with a loan, we see the same pattern meaning a very less amount of people have a loan except for the offer. So our conclusion here can be that people with some kind of loan and default are not the best shots for a term deposit offer.

Data last loaded: *May 10, 2022, 3:47 PM*

Client Marital Status

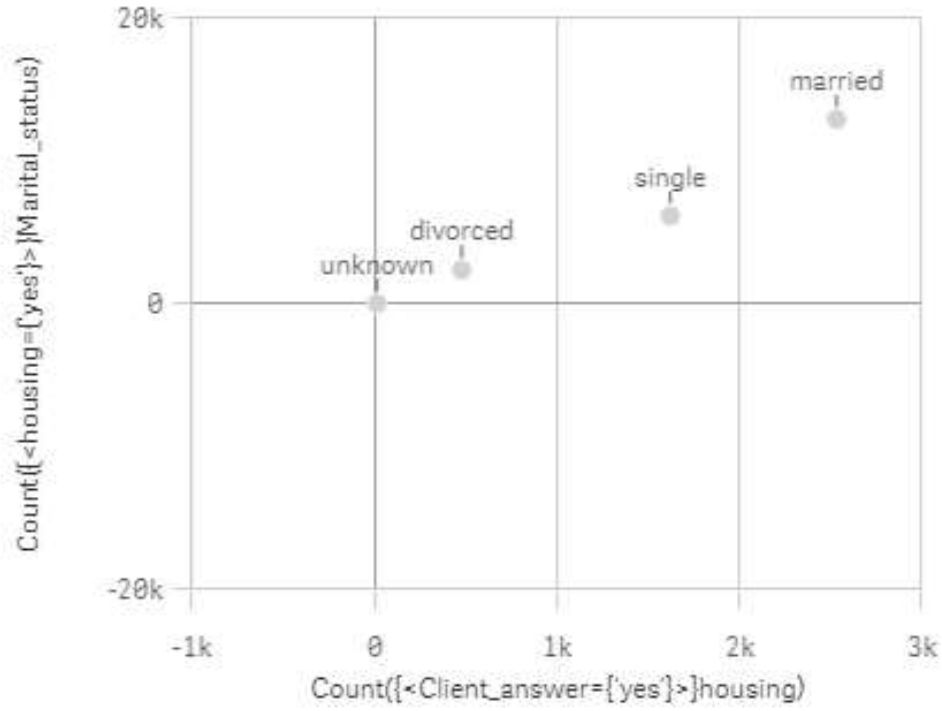
Click sheet to make selections | Reset selections | Go to sheet

Marital Status and client answer count



Here we see that married clients take the tern deposit policy. This totally makes sense if they are saving> there are less amount of people saying yes if they are single and almost negligible who are divorced. Simple offer more to married people. Make single people a second priority and decrease the number of offers made to divorced clients.

Client Saying yes, married or not, have a housing loan?



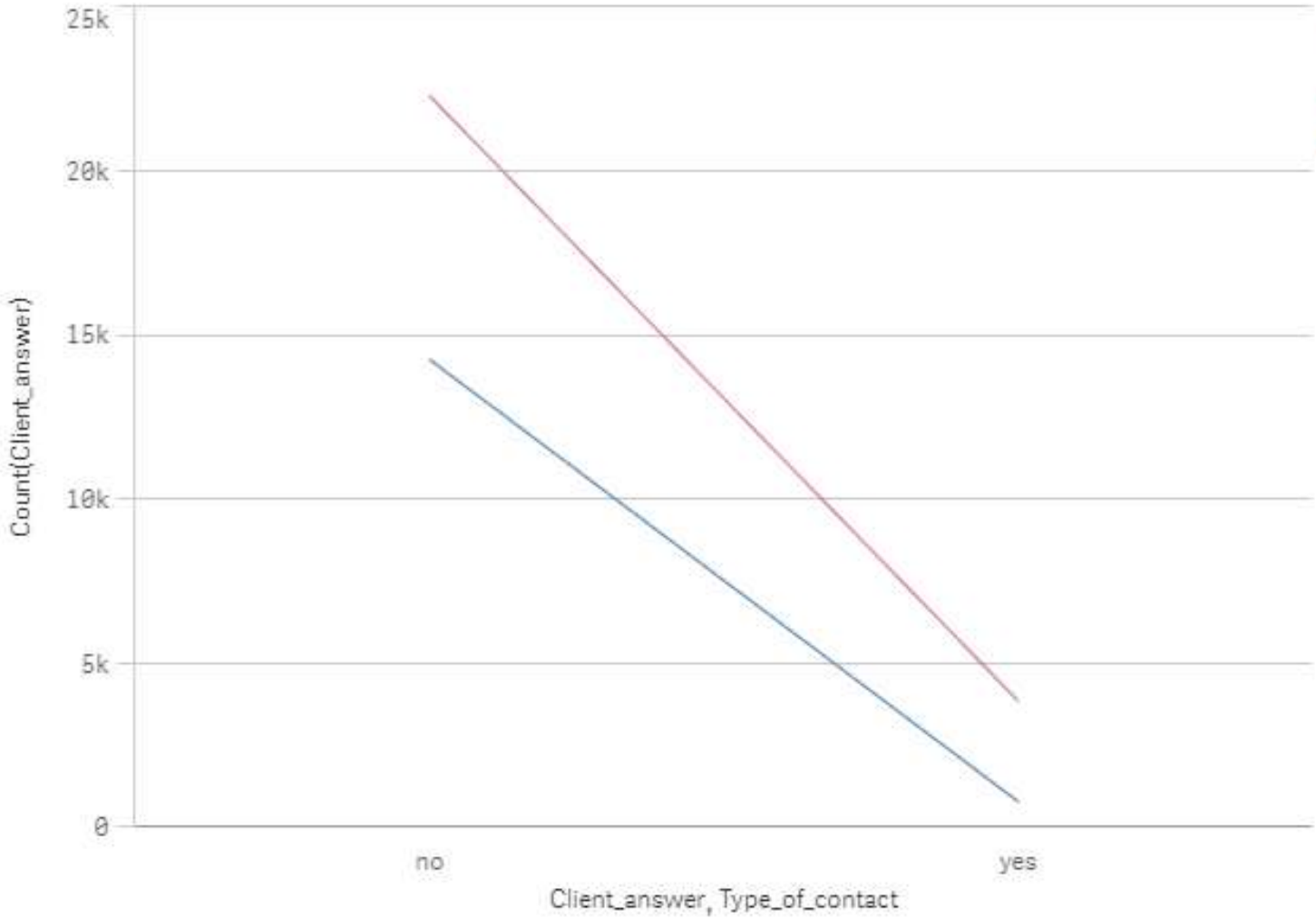
In the second chart, we see that people who are married and have a housing loan accept the term deposit offered more often. So potential clients saying yes are who are married and may or may have a housing loan. Housing can make sense if they are trying to save for the loan taken every month.

Data last loaded: *May 10, 2022, 3:47 PM*

Offer made through which type of contact

Click sheet to make selections | Reset selections | Go to sheet

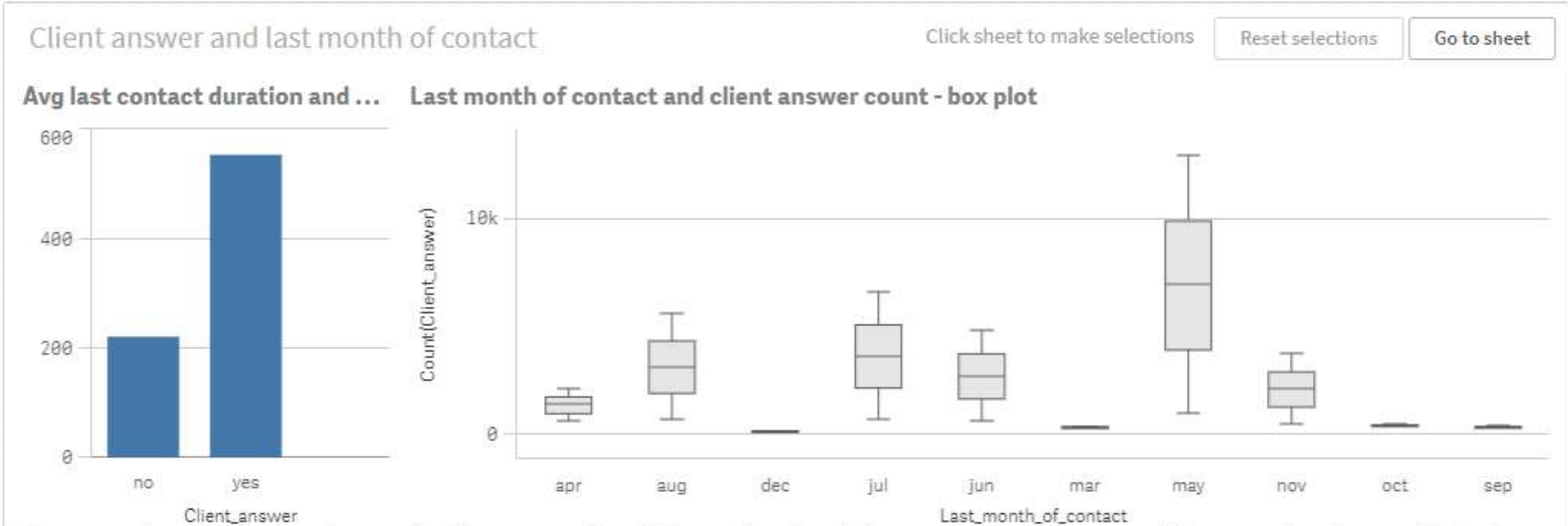
Type of contact and count of client answer



As the line chart shows that the contact made through cellular type of contact have more counts of yes.

This can simply have a reason of availability, the client is having his/her cells at all times, especially for the better part of the day. Another explanation can be that the telephone is mostly available in places where clients can mostly be busy, like home and office. If the bank is lucky to contact the client in his/her leisure hour the chances of being heard over cells are more bright.
So this simply means at least a major part of contact should be made over cell phones.

Data last loaded: *May 10, 2022, 3:47 PM*



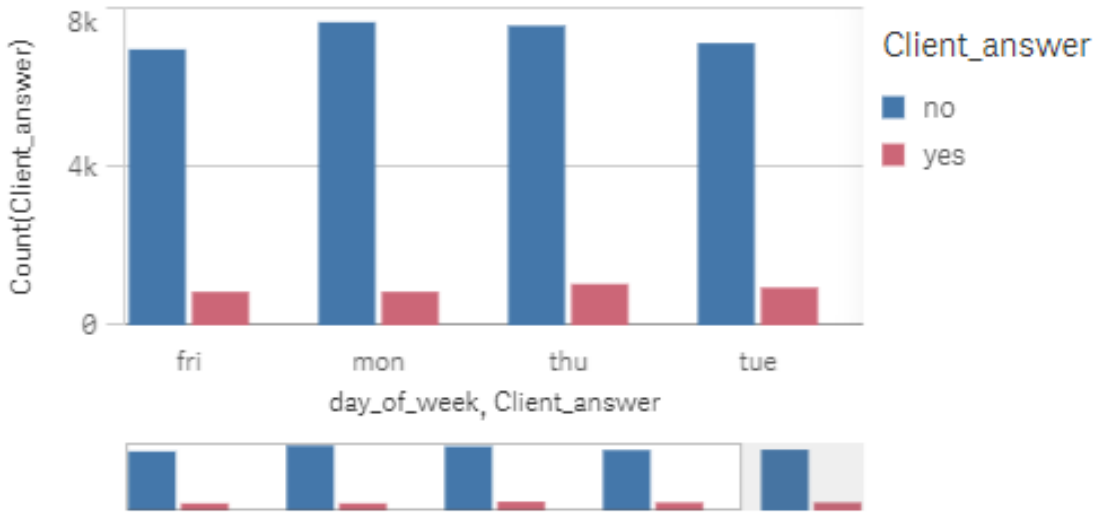
The most activity is shown in the month of May. Maybe this could be an overall important month in terms of some holiday or some kind of bonus. So major efforts should be made this month. No outliers are shown since we are only counting client answers overall.

Another bar chart shows the average time duration of contact for both the answers, yes and no. For a positive answer, the average duration is almost 600 seconds. This is also an average of about 10 to 12 minutes. Meaning a client needs to get persuaded which definitely needs time. So our earlier explanation that the client should have a little time to give to you is important. Secondly, another impactful thing will be the starting conversation. The opening pitch have the most value, this will persude client to atleast listen more so

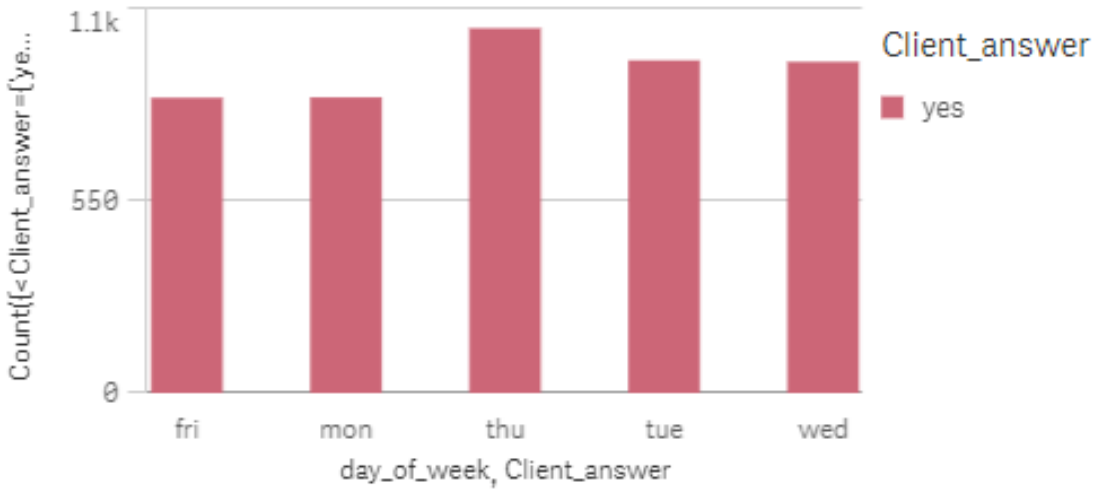
Client answer and day of week of contact

Click sheet to make selections | Reset selections | Go to sheet

dayofweek and client answer count



dayofweek and client answer count

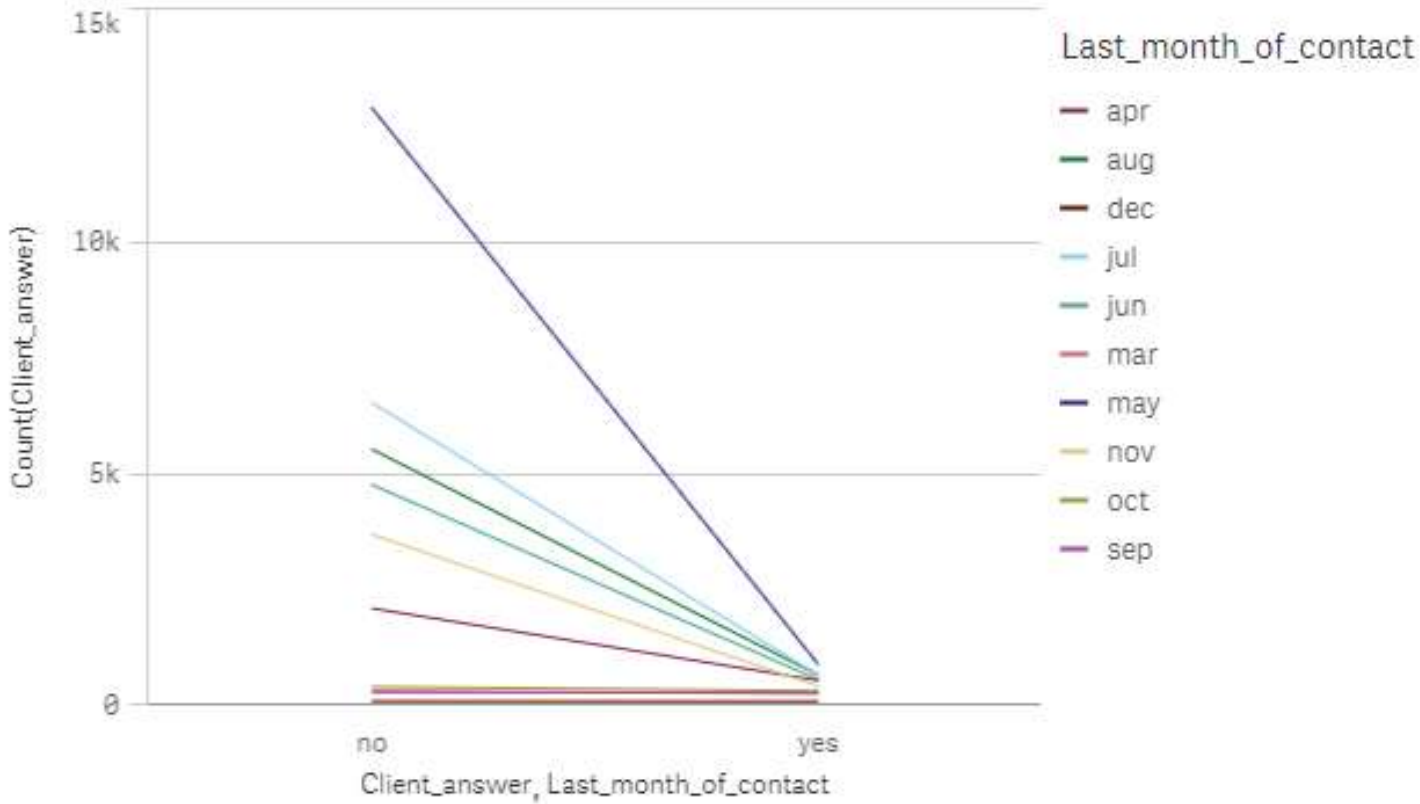


The most lest are on Thursday. There is not much difference in count of days but we are only offering five days a week which can a problem. I think bank can try to offer term-deposit on weekends-this can be a hetic and effort-consuming task but maybe this can have a positive impact in the answers of the client.

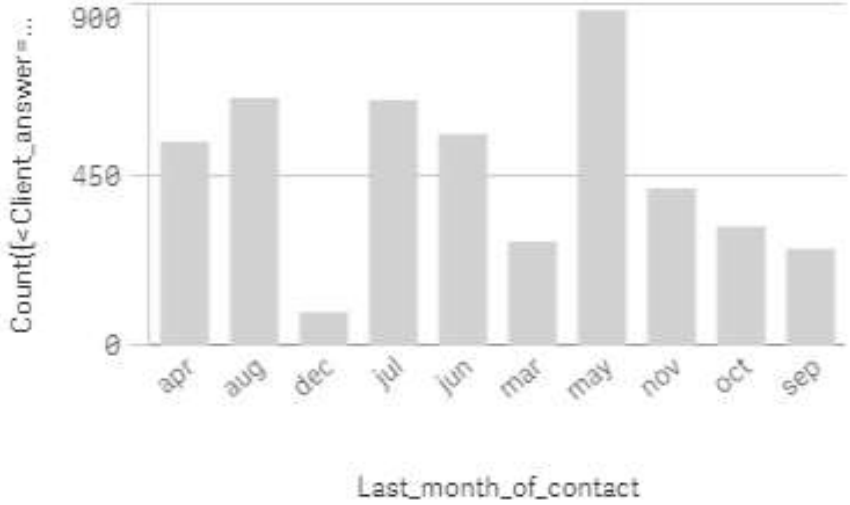
Client answer and last month of contact

Click sheet to make selections | Reset selections | Go to sheet

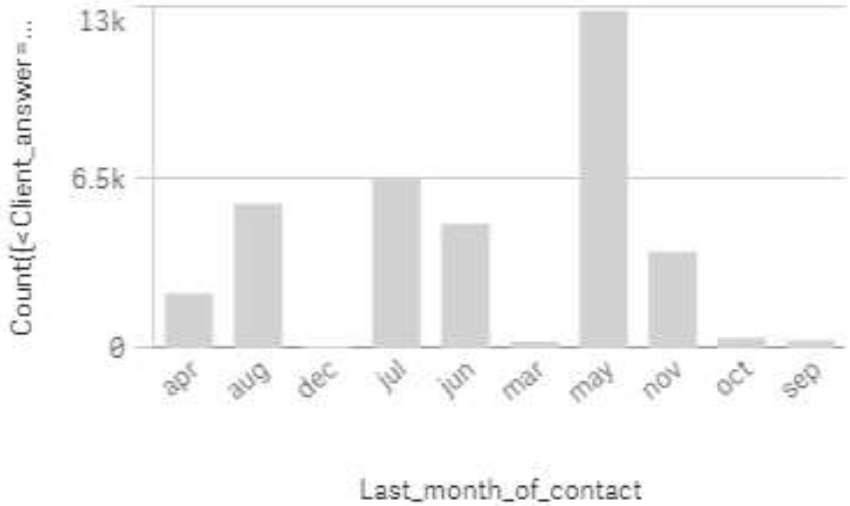
Client answer and last month of contact - line graph



Count of saying yes and month of contact



Count of saying no and month of contact



The most number of outputs are noted in the month of may as shown by the line chart.

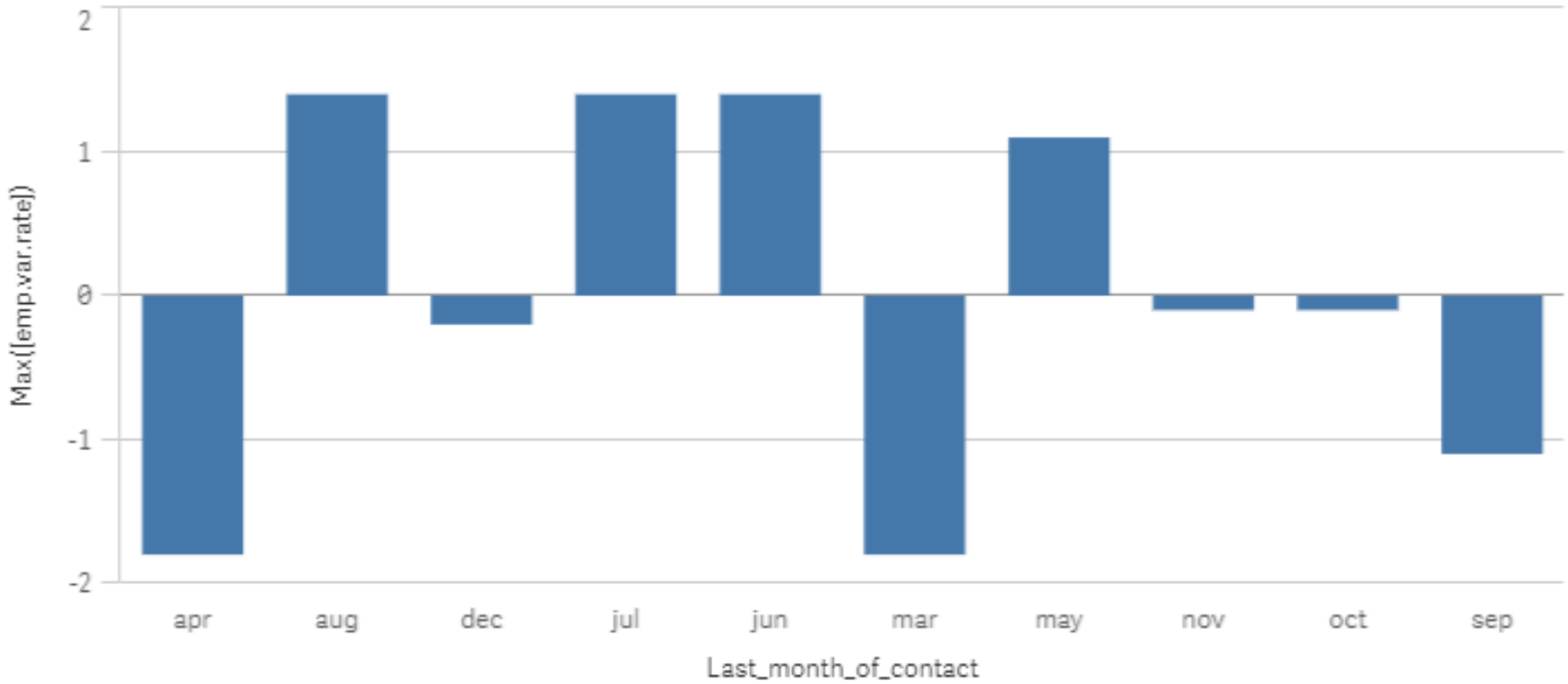
We have almost 900 yes output in that month meaning in this month more offers were made or simply this month have some important factor affecting this output.

Employee Variation rate and consumer confidence index

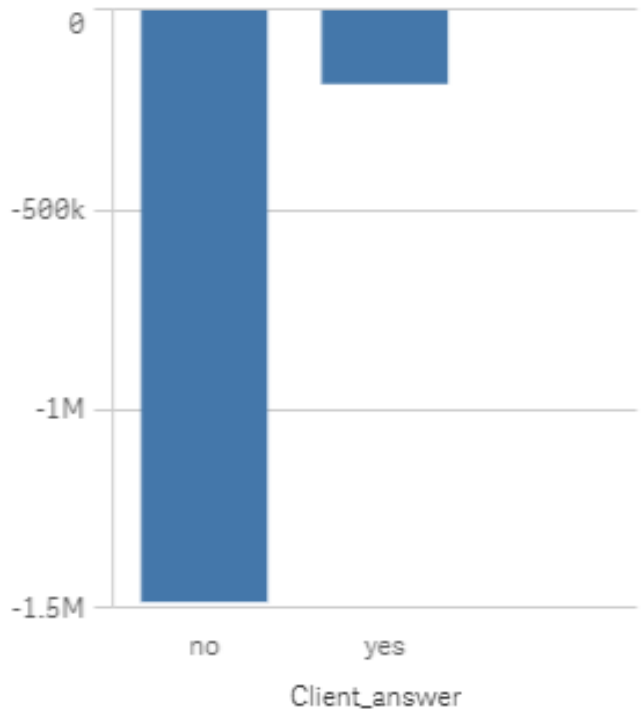
Click sheet to make selections | Reset selections | Go to sheet

Avg Emp.var.rate

Maximum Employee variation rate in the last month of contact



Consumer.confidence and client an...



Avg consumer confidence

Avg([cons.conf.idx])

-40.5

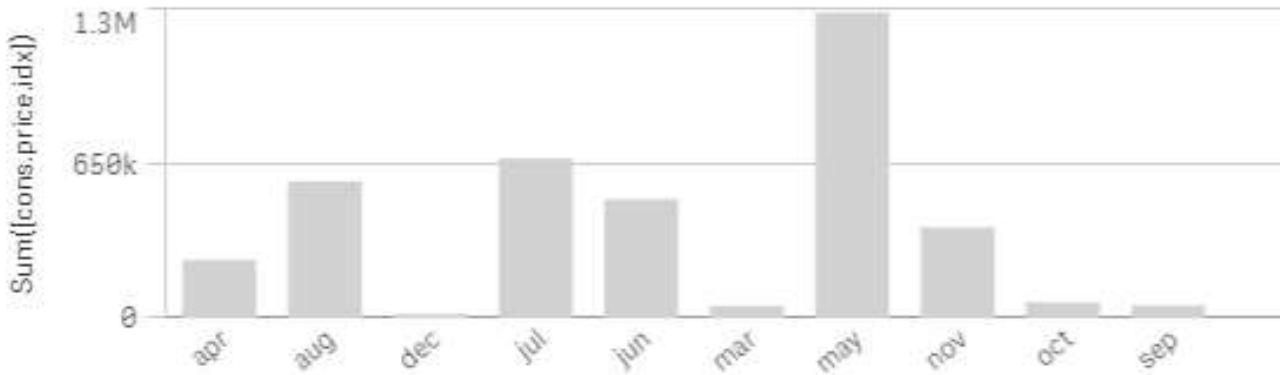
The average employee variation rate is 0.08. We plot it per month and get that most month has a negative employee variation rate, meaning variation is very low but still, we have a positive variation rate for the month of may which have the highest yes output.

For consumer confidence, we see that to get a yes if the consumer confidence should e more(in a negative scale sense) as shown in the chart.

Consumer price index

Click sheet to make selections | Reset selections | Go to sheet

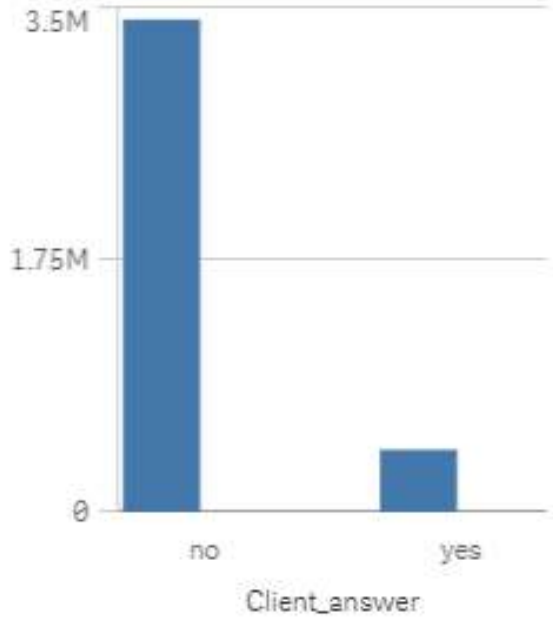
Consumer price index in last contact of month



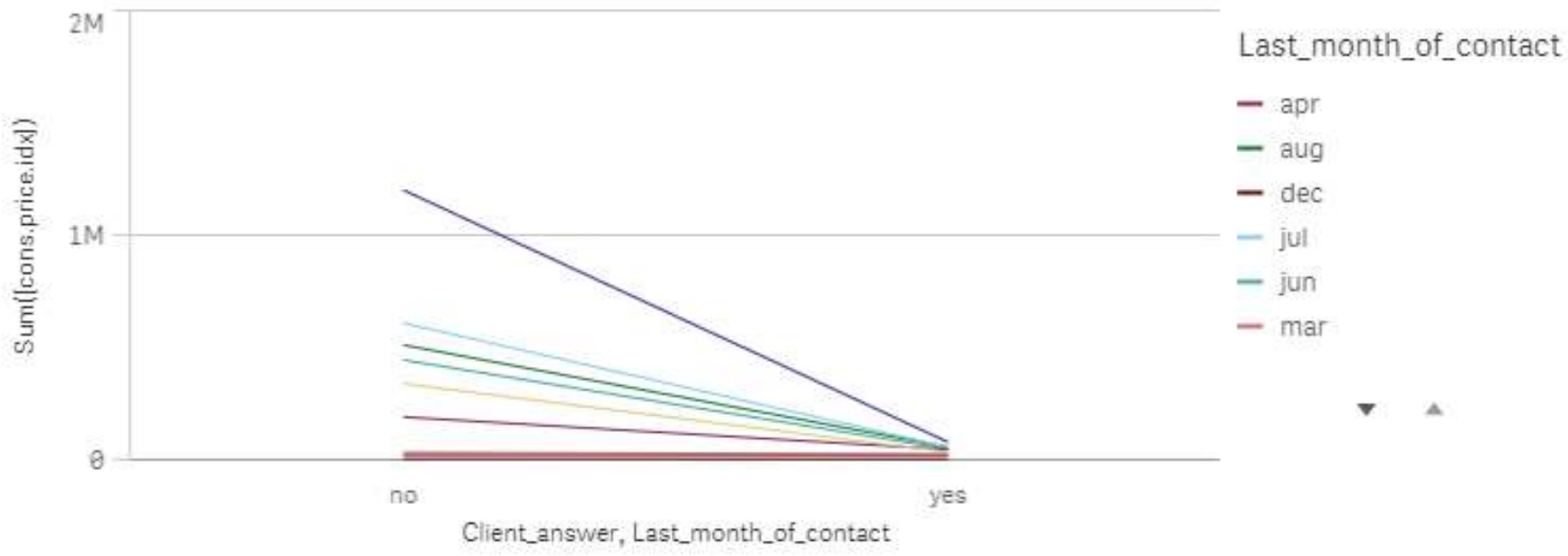
The consumer price index is the highest in the month of May - this matches our previous pattern. the most sales are done in this month so the sum is highest for this.

The bar chart shows the sum of the consumer price index where the answers are yes and no.

Client ans and sum.con.price.ind



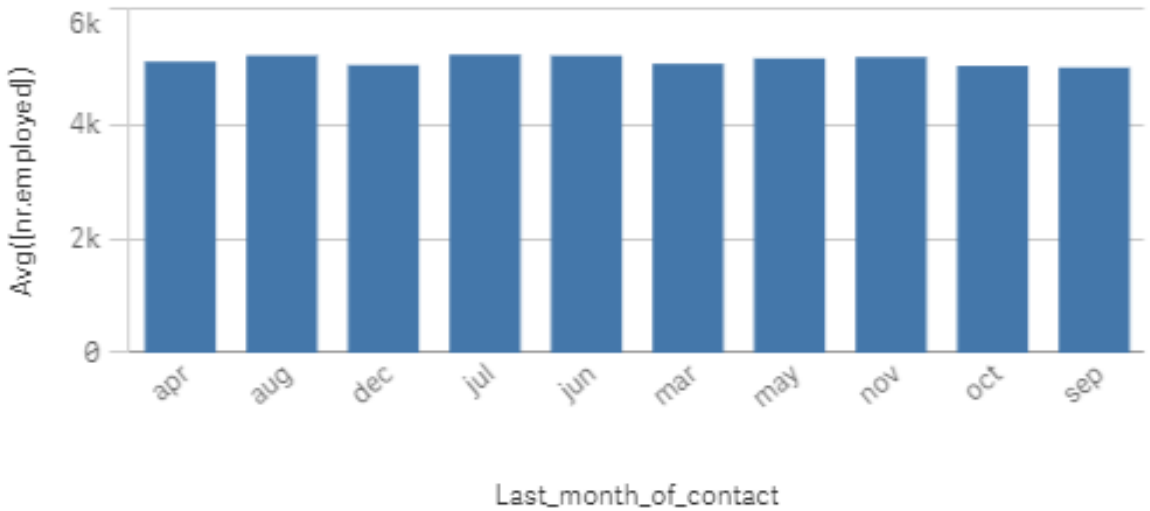
Con.price.ind with client answer in a month



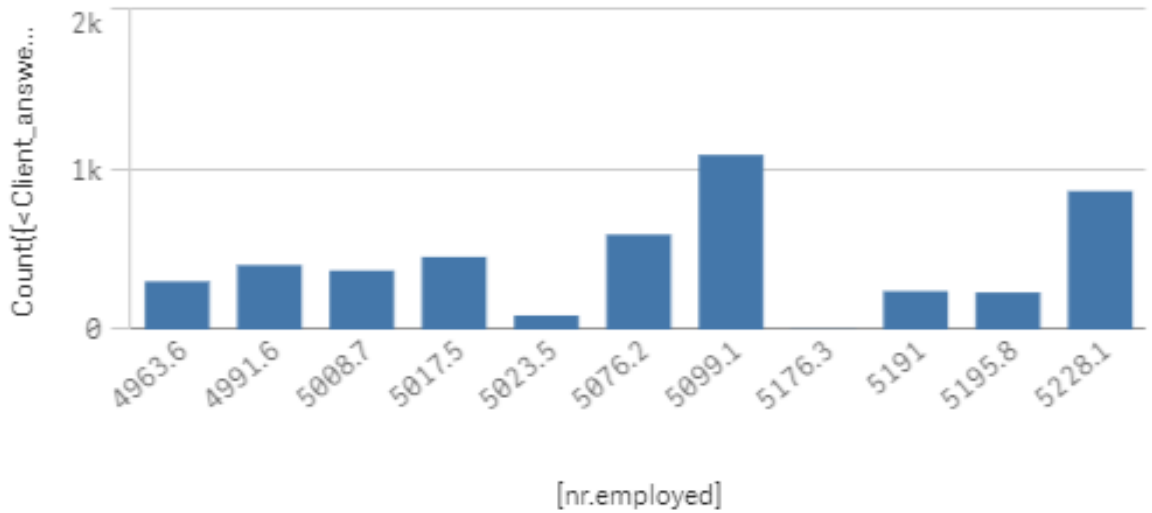
Number of employees and euribor

Click sheet to make selections | Reset selections | Go to sheet

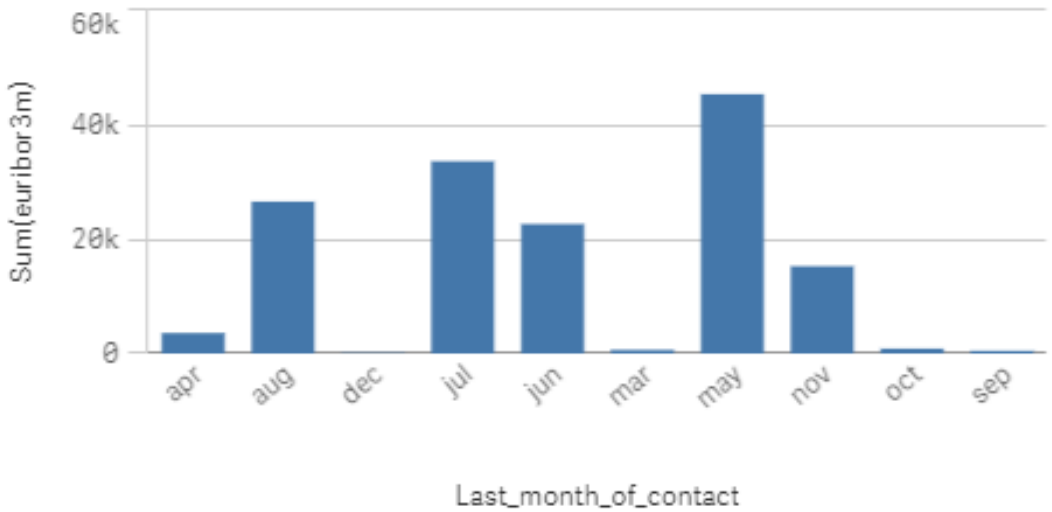
Number of employee average in a month



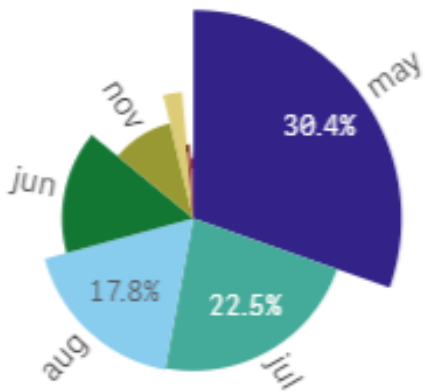
Count of client yes and number of employees



Sum of euribor in a month



Month_euribor and yes



In a month the average number of employees is almost 5500. For the values possible of employees where we have the highest number of employees.

The 3 month Euribor interest rate is the interest rate at which a selection of European banks lend one another funds denominated in euros whereby the loans have a maturity of 3 months.

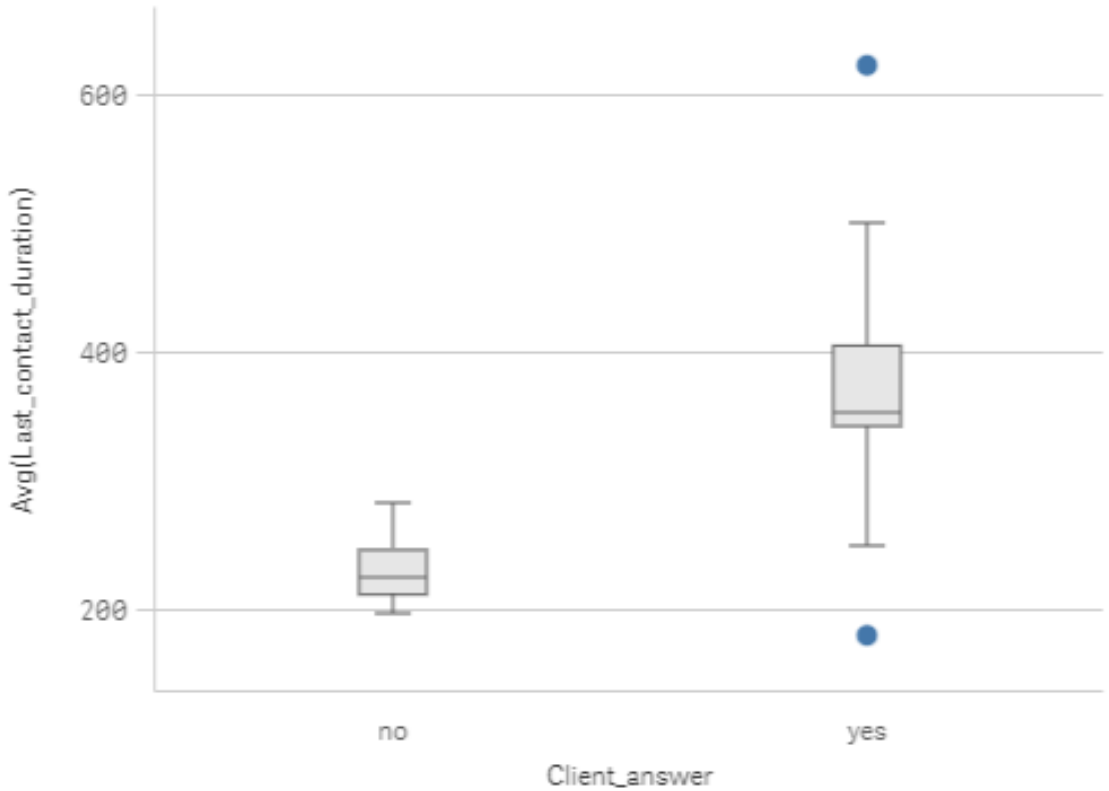
Radius shows the number of yes, the angle show the sum of Euribor per a month(shown by color). The highest contribution comes from may.

Data last loaded: *May 10, 2022, 3:47 PM*

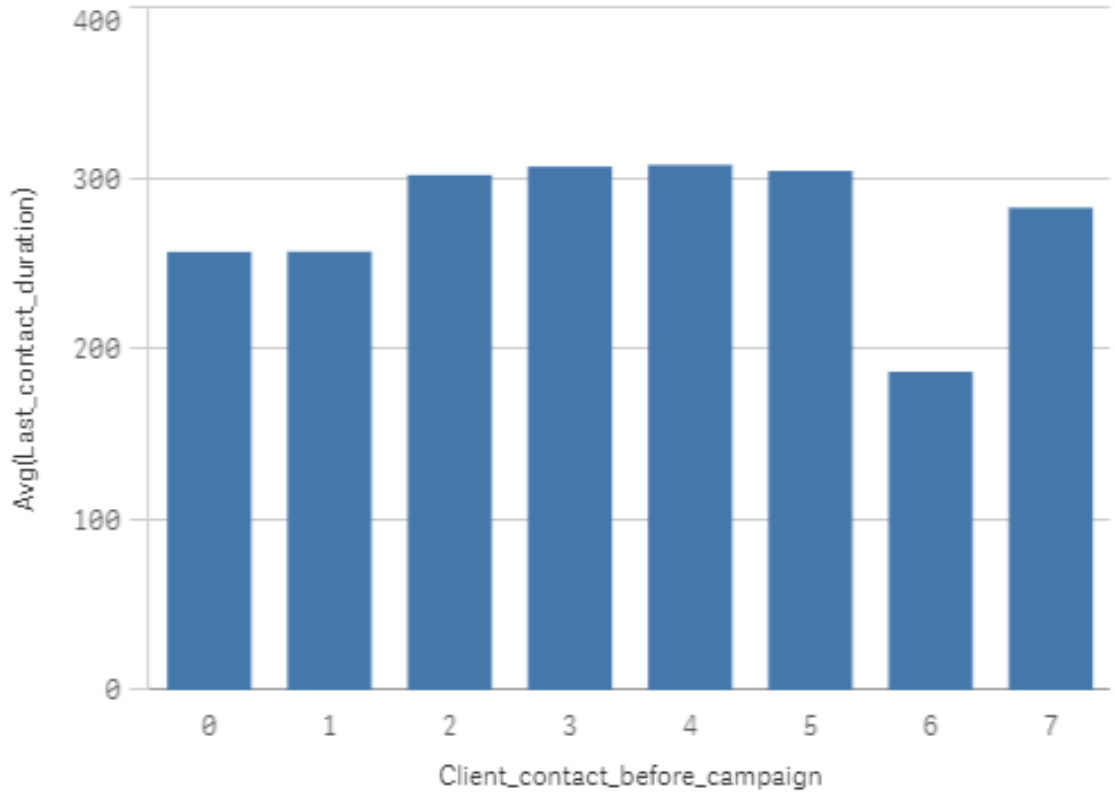
Number of Client_contact_before_campaign

Click sheet to make selections | Reset selections | Go to sheet

box-plot, client ans and duration with number contact before cam...



number contact before campaign and during of contact

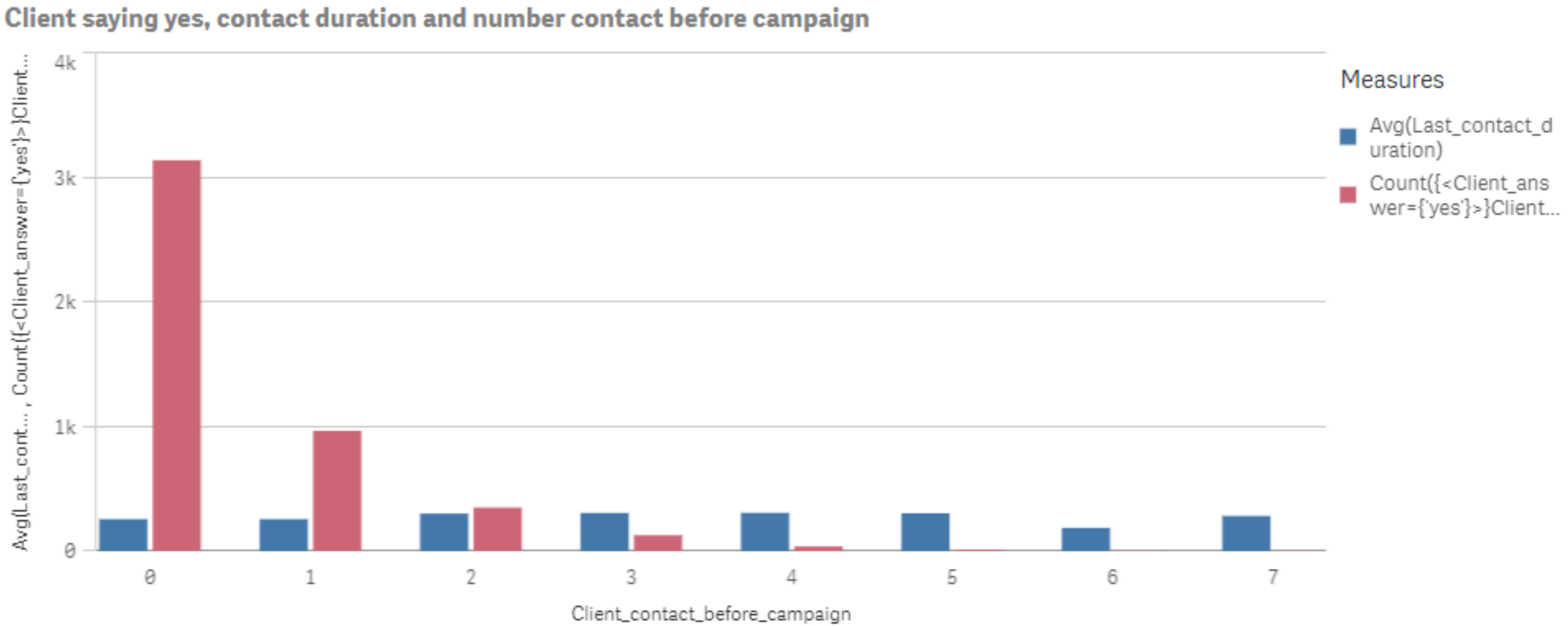


As the graphs show that there is more probability of saying yes if the client was contacted before the campaign.As the box plot shows the more the number of contact before the campaign, the more positive response we have. The same is shown in the bar chart. The duration is also important since we need to talk more time with the client to make him/her say yes. The conclusion can also be that clients are frequent clients so they know these offers well also. the permanent client tends to say yes more often. All major permanent clients should be offered all the time.

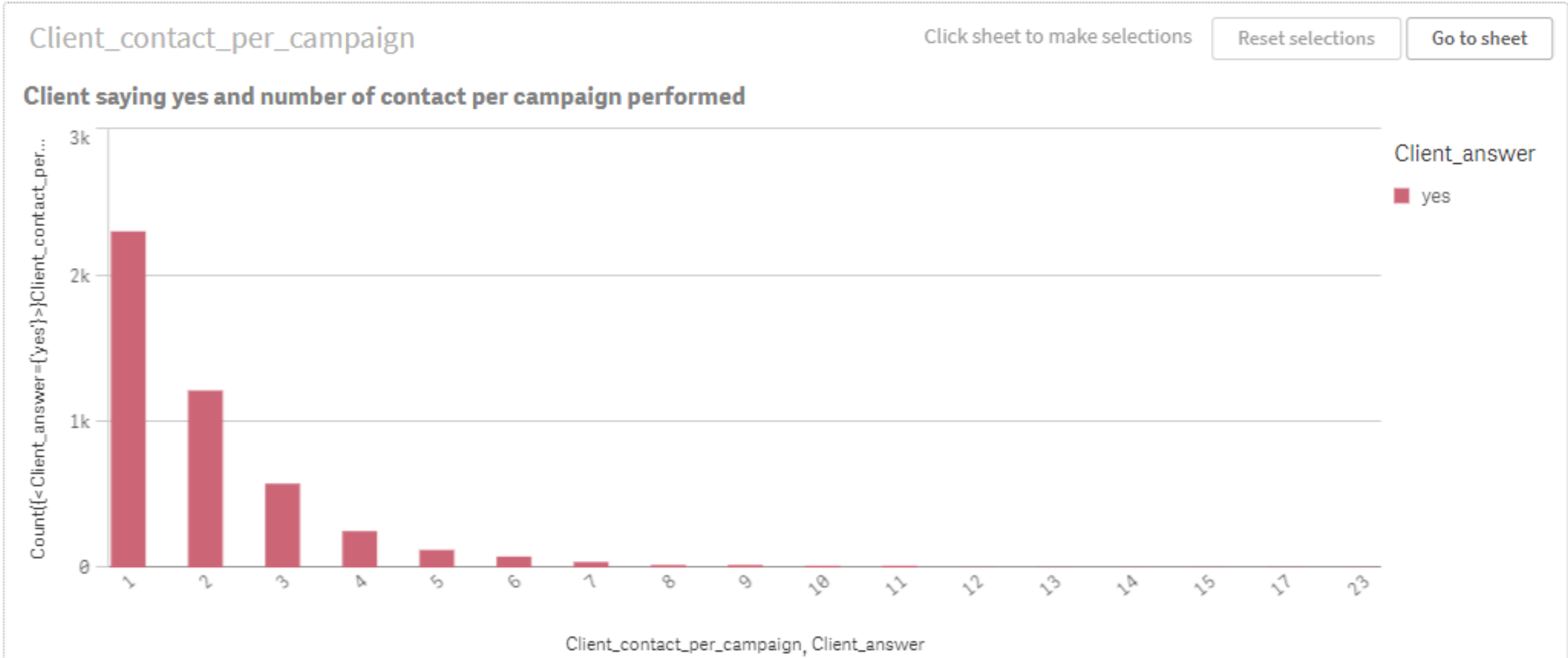
Data last loaded: *May 10, 2022, 3:47 PM*

Number of Client_contact_before_campaign

Click sheet to make selections | Reset selections | Go to sheet



But in this chart, as we analyze further we see that most client that said yes were contacted zero number of times before contact, meaning old clients are part of it but the new client also makes a major some in the count of yes.



This shows how many times a client was contacted before the campaign when he/she said yes to the current campaign. Majorly who were contacted one time only say yes to the campaign.

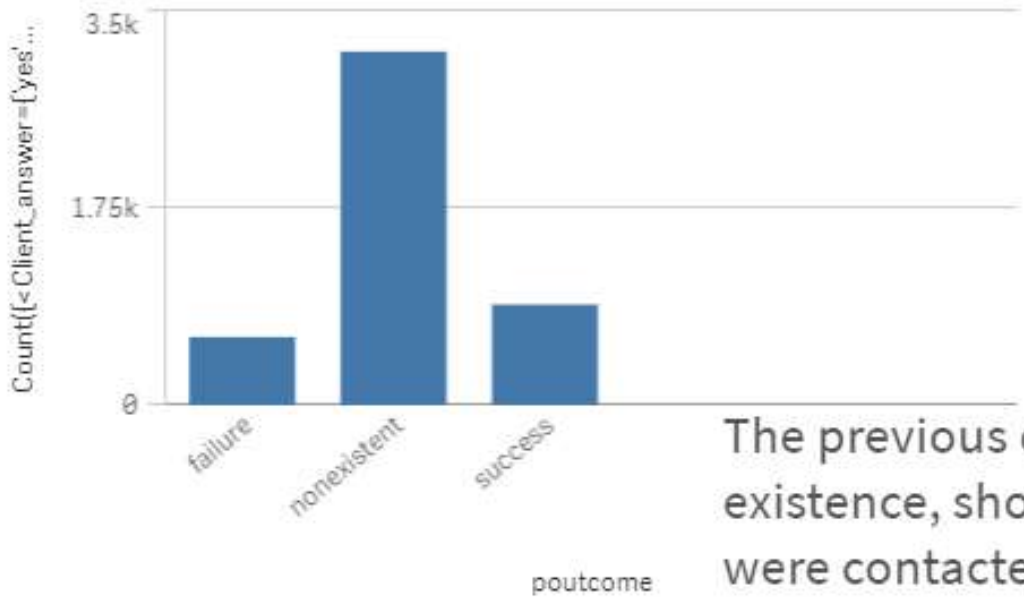
The more the number of contactless is the chances of saying yes. Only contact clients when pretty sure they will say yes. Or there is at least a good possibility.

Previous outcomes and p-days

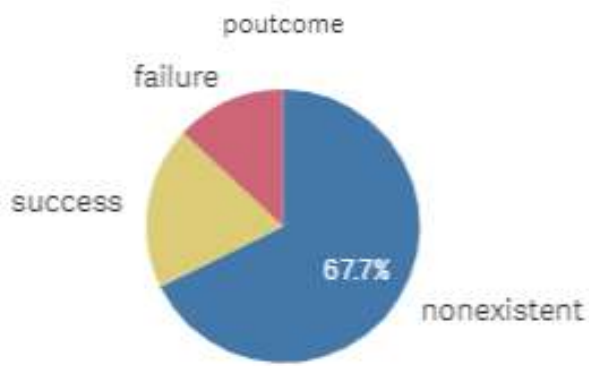
Click sheet to make selections | Reset selections | Go to sheet

P-days and client answers

Previous results and yes in this campaign



Pie chart with same results



The previous data for most client is non-existence, showing the same thing that they were contacted for the first time At least success rate is higher that the failure rate.

The count of yes are more for the people who are contacted for the first time. We have a good count in 3,4,6 days too, this can lead to the idea that they were approached first and then when they were approached again they accepted the term-deposit.

Data last loaded: *May 10, 2022, 3:47 PM*