

EDA of Bank Subscription Campaign

The Assignment

The data is from 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 (or not) subscribed.

- Using the bank.csv file, perform the following tasks:
 - Import data into Excel and perform data cleaning
 - On a Word document, describe the changes you made and your observation
 - Perform filtering using table format
 - Group and summarize data – be creative
 - On the same word document, describe your observation
 - Use color scale or data bar to explore and highlight the data
 - On the same word document, describe your observation
 - Build a Pivot table to show relationship of column Y vs other factors (columns)
 - On the same word document, describe your observation and what hypotheses being observed?
 - Save the EXCEL file with the pivot tables with your name as the file name – submit it to Canvas.

The columns in bank.csv

- 1 age (numeric)
- 2 job : type of job (categorical: "admin.", "unknown", "unemployed", "management", "housemaid", "entrepreneur", "student", "blue-collar", "self-employed", "retired", "technician", "services")
- 3 marital : marital status (categorical: "married", "divorced", "single"; note: "divorced" means divorced or widowed)
- 4 education (categorical: "unknown", "secondary", "primary", "tertiary")
- 5 default: has credit in default? (binary: "yes", "no")
- 6 balance: average yearly balance, in euros (numeric)
- 7 housing: has housing loan? (binary: "yes", "no")
- 8 loan: has personal loan? (binary: "yes", "no")
-
- # related with the last contact of the current campaign:
 - 9 contact: contact communication type (categorical: "unknown", "telephone", "cellular")
 - 10 day: last contact day of the month (numeric)
 - 11 month: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")
 - 12 duration: last contact duration, in seconds (numeric)
-
- # other attributes:
 - 13 campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
 - 14 pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)
 - 15 previous: number of contacts performed before this campaign and for this client (numeric)
 - 16 poutcome: outcome of the previous marketing campaign (categorical: "unknown", "other", "failure", "success")
- 17 Output variable (desired target):
 - a. y - has the client subscribed a term deposit? (binary: "yes", "no")

Data Cleaning (Notes at the end of the file)

Revised the headers and order of some columns for clarity:

- 1 Age
- 2 Job
- 3 Status
- 4 Education
- 5 CreditDefault
- 6 AveYrlyBalance

- 7 HousingLoan
- 8 PersonalLoan
- 9 Contact : Communication Type
- 10 Day: Last Contact Day of the Month
- 11 Month: Last Contact Month of the Year
- 12 LastContactDuration: Duration of last contact, in seconds
- 13 FreqofContactPrev: Frequency of Contact in Previous Campaign
- 14 FreqofContactCur: Frequency of Contact in Current Campaign
- 15 DaysAfterLastContactPrev: Number of Days after Last Contact in Previous Campaign, -1 meaning no contact
- 16 OutcomePrev: Outcome of Previous Campaign
- 17 OutcomeCur: Outcome of Current Campaign, Subscribed to Term Deposit:Y/N, also known as current outcome.

FILTERING/GROUPING DATA

Filtered OutcomeCur (Outcome of Current Campaign) by Ascending Order
 YES = 521/4521, NO = 4000/4521

OBSERVATIONS

1. Of the 521 YES (in OutcomeCur), all had no default in credit
 But 3933/4000 NO (in OutcomeCur) also had no default in credit
 Therefore, default in credit may not be factor in the outcome

2. Of the YES, Frequency of Contact in Current Campaign was as follows:

No. of Yes	FreqContactCur
240	1
138	2
57	3
43	4
9	5
16	6

Looking at the No. of YES against the frequency of contact, you might say the more frequent the contact in the current campaign, the lower the YES.

However, the more frequent the contact, the lower the NO, too.

No. of No	FreqContactCur
1494	1
1126	2
501	3
282	4
158	5
139	6

Therefore, frequency of contact in the current campaign is also not a factor in the outcome.

3. Of the 521 YES: 83 were a "Success" in OutcomePrev, 63 were a "Failure." However, 375 were "Other" or "Unknown."

Of the 4000 NO: 46 were a "Success" in OutcomePrev, 427 were a "Failure," 3527 were "Other" or "Unknown."

Therefore, no conclusion can be made regarding the relationship of the previous outcome with the current outcome.

4. Of the 4000 who said NO: 3368 were not previously contacted (-1 in column).

Of the 521 who said YES: 337 were not previously contacted. The rest of the YES were contacted within 1-804 days after (or an average of 196 days) after last campaign.

Therefore, the no. of days after last contact in the previous campaign is also not relevant.

5.

2559/4521 have a Housing Loan

692/4521 have a Personal Loan

407/4521 have both

Of the 4000 who said NO to the Term Deposit Subscription:

- 1661/4000 had no Housing Loan
- 2339/4000 had a Housing Loan. That's 58% of the total number of NOs
- 381/4000 had a housing and a personal loan

Of the 521 who said YES to the Term Deposit Subscription:

- 220/521 had a Housing Loan
- 301/521 had no Housing Loan. That's 57% of the total number of YES.
- 283/521 had no Personal Loan and no Housing Loan

Therefore, Housing Loan may have a relationship with the current outcome.

6.

723 out of 4521 customers have zero or negative balance

3798/4521 have a positive ave yearly balance, averaging 1725.23

461 of the 521 YES have a positive yearly balance, averaging 1802.33

3337 of the 4000 NO have a positive ave yearly balance, averaging 1714.58

663 of the 4000 NO have a zero or negative ave yearly balance, averaging -163.97

While majority of the customers have a positive ave yrly balance, those of the NO do have a lower average than those of the yes.

Therefore, Average Yrly balance may have a relationship with the current outcome.

7. Last Contact Duration vs. OutcomeCur

The 521 who said YES had a Last Contact Duration Average of 552 seconds or 9.2 minutes

The 4000 who said NO had an average 226 seconds or 3.7 minutes

Therefore, LastContactDuration may have a relationship with the outcome.

PIVOT TABLE 1

HOUSING LOAN VS. OUTCOME CURR (OUTCOME OF CURRENT CAMPAIGN, ie Subscription to Term Deposit)

Count of OutcomeCur	OutcomeCur		
HousingLoan	No	Yes	Grand Total
No	1661	301	1962
Yes	2339	220	2559
Grand Total	4000	521	4521

This table shows that the majority who said NO to the Term Deposit Subscription (as seen in "OutcomeCur", previously the "Y" Column) have a Housing Loan, while the Majority of those who said YES have no Housing Loan.

Therefore, those without a Housing Loan were/are more likely to say YES to the Term Deposit Subscription.

PIVOT TABLE 2

LASTCONTACTDURATION VS. OUTCOME CURR (OUTCOME OF CURRENT CAMPAIGN, ie Subscription to Term Deposit)

Average of LastContactDuration	
OutcomeCur	Total
No	226.3475
Yes	552.7428023
Grand Total	263.9612917

This table illustrates that the average duration of last contact (in seconds) that produced a YES is double that of the average duration of last contact that produced a NO.

An average of 552 seconds or 9.2 minutes was spent on last contact that yielded a YES vs. an average of 226 seconds or 3.7 minutes that yielded a NO.

Therefore, the longer the contact, the more likely it will get a YES.

By the way, GRAND TOTAL refers to the average contact duration of ALL calls, regardless of whether it yielded a YES or NO.

PIVOT TABLE 3

FREQUENCY OF CONTACT IN CURRENT CAMPAIGN VS. OUTCOME CURR (OUTCOME OF CURRENT CAMPAIGN, ie Subscription to Term Deposit)

Count of OutcomeCur	FreqContact															
OutcomeCur		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No		1494	1126	501	282	158	139	69	52	28	26	22	20	15	10	9
Yes		240	138	57	43	9	16	6	4	2	1		1	2		
Grand Total		1734	1264	558	325	167	155	75	56	30	27	22	21	17	10	9

This table shows that Frequency of Contact in the current campaign yields diminishing returns in terms of YES.

The optimal number of times to contact the customer to achieve a YES is in the range of 1-3 or 1-4 times at the most. Contacting the customer more often than that may just be a waste of time and money.

NOTE: This table has many columns but I have highlighted the important ones.

PIVOT TABLE 4

OUTCOME OF PREVIOUS CAMPAIGN VS OUTCOME OF CURRENT CAMPAIGN

Count of OutcomeCur	OutcomeCur		
OutcomePrev	No	Yes	Grand Total
Failure	427	63	490
Success	46	83	129
Grand Total	473	146	619

The majority of customers who said NO to the current campaign (i.e. said NO to the Term Deposit Subscription) were also a "Failure" in the previous campaign.

The majority of customers who said YES to the current campaign were also a "Success" in the previous campaign.

We could recommend that the next campaign focus on those for whom the previous campaign/s was/were a "Success" to optimize outcomes.

However, there were too many "unknown" and "other" in the outcome of the previous campaign so this hypothesis cannot be verified, given the current data.

PIVOT TABLE 5

AVERAGE YEARLY BALANCE VS. OUTCOME CURR (OUTCOME OF CURRENT CAMPAIGN, ie Subscription to Term Deposit)

Average of AveYrlyBalance	
OutcomeCur	Total
No	1403.21175
Yes	1571.955854
Grand Total	1422.657819

The mean of the average yearly balance of all who said YES to the Term Deposit Subscription is higher than the mean of average yearly balance of all who said NO.

Therefore, the average yearly balance of a customer may have a relationship with the Outcome of the Current Campaign.

Possibly, the higher the average yearly balance of a customer, the more likely he will subscribe to a Term Deposit.

By the way, GRAND TOTAL refers to the average yearly balance of ALL customers, regardless of whether they said YES or NO.

PIVOT TABLE 6

CREDIT DEFAULT VS. OUTCOME CURR (OUTCOME OF CURRENT CAMPAIGN, ie Subscription to Term Deposit)

Count of OutcomeCur	CreditDefault	
OutcomeCur	No	Yes
No	3933	67
Yes	512	9
Grand Total	4445	76

While majority of those who said YES to the Term Deposit Subscription have no Credit Default, the majority who said NO likewise have no Credit Default.

Therefore, Credit Default has potentially no relationship with the Outcome of the Current Campaign.

Additional notes:

DATA CLEANING

- Capitalised some Headers (which had numeric values, e.g. Age, Balance, Day, Month, Duration) manually. Kept getting errors if I used '=Proper (text)' and pasted on to the header
- Used Trim (Proper) for Columns B, C, D, E, G, H, I, K, P, Q

- Removed Duplicates. 1044053 duplicate values found and removed. 4522 unique values remain. Though earlier similar attempt to “Remove duplicates” showed no duplicates. Not sure what happened there.
- Replaced “.” in “Admin.” with “” in Column B, also replaced “Blue-Collar” with “Blue Collar”
- Changed headers so they’re easier to understand at a glance:
 - Marital to Status
 - Default to CreditDefault
 - Balance to AveYrlyBalance
 - Housing to HousingLoan
 - Loan to PersonalLoan
 - Duration to LastContactDuration
 - Campaign to FreqContactCur
 - Pdays to DaysafterLastContact(Prev)
 - Previous to FreqContactPrev
 - Poutcome to OutcomePrev
 - Y to OutcomeCurrent
- I also moved “FreqContactPrev” column to just before “FreqContactCur” so that they’re beside each other for easier comparison (just in case it matters); it’s also more chronological

Other notes:

- Cleared Format. No change.
- HOME > FIND & SELECT > Go to Special > Blanks = always caused app to crash
- Also noted that CTRL-A in Windows = Command-A in Mac