

Bank Marketing

This dataset consists of direct marketing campaigns by a Portuguese banking institution using phone calls. The campaigns aimed to sell subscriptions to a bank term deposit (see variable `y`).

Not sure where to begin? Scroll to the bottom to find challenges!

[1] 41188

...	↑↓	...	↑↓	job	...	↑↓	...	↑↓	education	...	↑↓	...	↑↓	...	↑↓	...	↑↓	c.	...	↑↓	...	↑↓	da..	...	↑↓
1	56	housemaid		married	basic.4y				no		no		no		telephone		may		mon					▲	
2	57	services		married	high.school				unkno...		no		no		telephone		may		mon					▼	
3	37	services		married	high.school				no		yes		no		telephone		may		mon						
4	40	admin.		married	basic.6y				no		no		no		telephone		may		mon						
5	56	services		married	high.school				no		no		yes		telephone		may		mon						
6	45	services		married	basic.9y				unkno...		no		no		telephone		may		mon						
7	59	admin.		married	professional.course				no		no		no		telephone		may		mon						
8	41	blue-collar		married	unknown				unkno...		no		no		telephone		may		mon						
9	24	technician		single	professional.course				no		yes		no		telephone		may		mon						
10	25	services		single	high.school				no		yes		no		telephone		may		mon						
11	41	blue-collar		married	unknown				unkno...		no		no		telephone		may		mon						
12	25	services		single	high.school				no		yes		no		telephone		may		mon						
13	29	blue-collar		single	high.school				no		no		yes		telephone		may		mon						
14	57	housemaid		divorced	basic.4y				no		yes		no		telephone		may		mon						
15	35	blue-collar		married	basic.6y				no		yes		no		telephone		may		mon						
16	54	retired		married	basic.9y				unkno...		yes		yes		telephone		may		mon					▼	

Rows: 100

Data Dictionary

Column	Variable	Class
age	age of customer	
job	type of job	categorical: "admin.", "blue-collar", "entrepreneur", "housemaid", "management", "retired", "self-employed", "services", "student", "technician", "unemployed", "unknown"
marital	marital status	categorical: "divorced", "married", "single", "unknown"; note: "divorced" means divorced or widowed
education	highest degree of customer	categorical: "basic.4y", "basic.6y", "basic.9y", "high.school", "illiterate", "professional.course", "university.degree", "unknown"
default	has credit in default?	categorical: "no", "yes", "unknown"
housing	has housing loan?	categorical: "no", "yes", "unknown"
loan	has personal loan?	categorical: "no", "yes", "unknown"
contact	contact communication type	categorical: "cellular", "telephone"
month	last contact month of year	categorical: "jan", "feb", "mar", ..., "nov", "dec"
day_of_week	last contact day of the week	categorical: "mon", "tue", "wed", "thu", "fri"
campaign	number of contacts performed during this campaign and for this client	numeric, includes last contact
pdays	number of days that passed by after the client was last contacted from a previous campaign	numeric; 999 means client was not previously contacted
previous	number of contacts performed before this campaign and for this client	numeric
poutcome	outcome of the previous marketing campaign	categorical: "failure", "nonexistent", "success"
emp.var.rate	employment variation rate - quarterly indicator	numeric
cons.price.idx	consumer price index - monthly indicator	numeric
cons.conf.idx	consumer confidence index - monthly indicator	numeric
euribor3m	euribor 3 month rate - daily indicator	numeric
nr.employed	number of employees - quarterly indicator	numeric
y	has the client subscribed a term deposit?	binary: "yes", "no"

[Source](#)  of dataset.

Citations:

- S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014
- S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimaraes, Portugal, October, 2011. EUROSIS.

Don't know where to start?

Challenges are brief tasks designed to help you practice specific skills:

-  **Explore:** What are the jobs of the people most likely to subscribe to a term deposit?
-  **Visualize:** Create a plot to visualize the number of people subscribing to a term deposit by month.
-  **Analyze:** What impact does the number of contacts performed during the last campaign have on the likelihood that a customer subscribes to a term deposit?

Scenarios are broader questions to help you develop an end-to-end project for your portfolio:

You work for a financial services firm. The past few campaigns have not gone as well as the firm would have hoped, and they are looking for ways to optimize their marketing efforts.

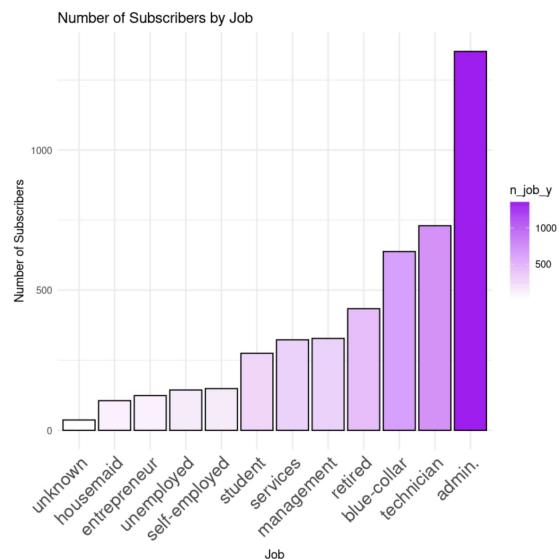
They have supplied you with data from a previous campaign and some additional metrics such as the consumer price index and consumer confidence index. They want to know whether you can predict the likelihood of subscribing to a term deposit. The manager would also like to know what factors are most likely to increase a customer's probability of subscribing.

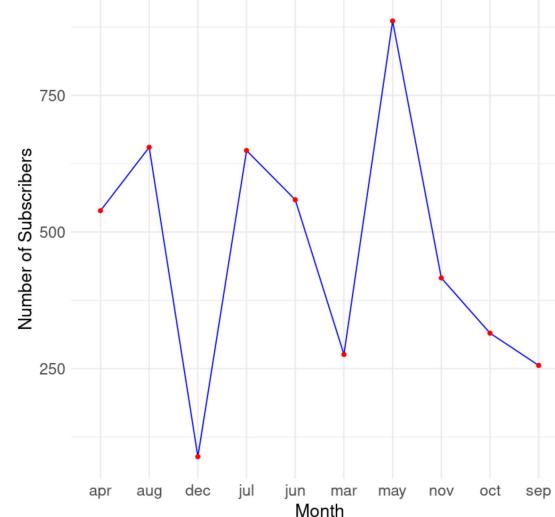
You will need to prepare a report that is accessible to a broad audience. It should outline your motivation, steps, findings, and conclusions.

41188

```
'age' · 'job' · 'marital' · 'education' · 'default' · 'housing' · 'loan' · 'contact' · 'month' · 'day_of_week' · 'duration' · 'campaign' · 'pdays' · 'previous' · 'poutcome' · 'emp.var.rate' · 'cons.price.idx' · 'cons.conf.idx' · 'euribor3m' · 'nr.employed' · 'y'
```

```
[1] "The jobs of the people most likely to subscribe to a term deposit is: "
# A tibble: 12 × 2
  job      n_job_y
  <chr>    <int>
1 admin.    1352
2 blue-collar  638
3 entrepreneur 124
4 housemaid   106
5 management   328
6 retired     434
7 self-employed 149
8 services     323
9 student      275
10 technician   730
11 unemployed   144
12 unknown      37
```



A number of subscriber to a deposit by month.

```
tibble [41,188 x 21] (S3: tbl_df/tbl/data.frame)
$ age      : num [1:41188] 56 57 37 40 56 45 59 41 24 25 ...
$ job      : chr [1:41188] "housemaid" "services" "services" "admin." ...
$ marital   : chr [1:41188] "married" "married" "married" "married" ...
$ education : chr [1:41188] "basic.4y" "high.school" "high.school" "basic.6y" ...
$ default   : chr [1:41188] "no" "unknown" "no" "no" ...
$ housing   : chr [1:41188] "no" "no" "yes" "no" ...
$ loan      : chr [1:41188] "no" "no" "no" "no" ...
$ contact   : chr [1:41188] "telephone" "telephone" "telephone" "telephone" ...
$ month     : chr [1:41188] "may" "may" "may" "may" ...
$ day_of_week: chr [1:41188] "mon" "mon" "mon" "mon" ...
$ duration  : num [1:41188] 261 149 226 151 307 198 139 217 380 50 ...
$ campaign  : num [1:41188] 1 1 1 1 1 1 1 1 1 1 ...
$ pdays     : num [1:41188] 999 999 999 999 999 999 999 999 999 999 ...
$ previous  : num [1:41188] 0 0 0 0 0 0 0 0 0 0 ...
$ poutcome  : chr [1:41188] "nonexistent" "nonexistent" "nonexistent" "nonexistent" ...
$ emp.var.rate: num [1:41188] 1.1 1.1 1.1 1.1 1.1 1.1 1.1 1.1 1.1 1.1 ...
$ cons.price.idx: num [1:41188] 94 94 94 94 94 94 ...
$ cons.conf.idx: num [1:41188] -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 ...
$ euribor3m   : num [1:41188] 4.86 4.86 4.86 4.86 4.86 ...
$ nr.employed: num [1:41188] 5191 5191 5191 5191 5191 ...
$ y          : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	1.000	2.000	2.568	3.000	56.000

Wilcoxon rank sum test with continuity correction

```
data: campaign by y
W = 94153912, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

```
Call:
glm(formula = y ~ campaign, family = binomial, data = bank)
```

```
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.776717  0.024716 -71.89  <2e-16 ***
campaign    -0.125440  0.009192 -13.65  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

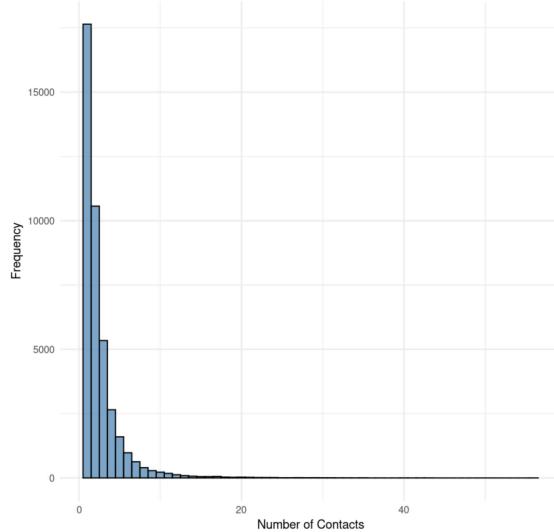
```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 28999  on 41187  degrees of freedom
Residual deviance: 28748  on 41186  degrees of freedom
AIC: 28752
```

```
Number of Fisher Scoring iterations: 5
```

```
# A tibble: 42 x 4
  campaign count_y total_customers sucess_rate
  <dbl>    <int>        <int>       <dbl>
1 1        2300        17642      0.130
2 2        1211        10570      0.115
3 3         574        5341       0.107
4 4         249        2651       0.0939
5 5         120        1599       0.0750
6 6          75         979       0.0766
7 7          38         629       0.0604
8 8          17         400       0.0425
9 9          17         283       0.0601
10 10         12         225       0.0533
# i 32 more rows
```

Histogram of Campaign Contacts



...	↑↓	...	↑↓	job	...	↑↓	...	↑↓	education	...	↑↓	...	↑↓	...	↑↓	...	↑↓	...	↑↓	c.	...	↑↓	...	↑↓	da...	...	↑↓
1		56		housemaid			married		basic.4y			no		no		no		no		telephone		may		mon			
2		57		services			married		high.school			unkno...		no		no		no		telephone		may		mon			
3		37		services			married		high.school			no		yes		no		no		telephone		may		mon			
4		40		admin.			married		basic.6y			no		no		no		no		telephone		may		mon			
5		56		services			married		high.school			no		no		yes		yes		telephone		may		mon			
6		45		services			married		basic.9y			unkno...		no		no		no		telephone		may		mon			
7		59		admin.			married		professional.course			no		no		no		no		telephone		may		mon			
8		41		blue-collar			married		unknown			unkno...		no		no		no		telephone		may		mon			
9		24		technician			single		professional.course			no		yes		no		no		telephone		may		mon			
10		25		services			single		high.school			no		yes		no		no		telephone		may		mon			

Rows: 4,761 △ truncated from 41,188 rows

From the result Summary of Results:

- Intercept (-1.7767, p < 0.001): This represents the log-odds of success ($y=1$) when the campaign variable is 0
- Campaign (-0.1254, p < 0.001): The coefficient is negative, meaning that as the number of contact attempts increases, the likelihood of a successful outcome decreases.
- Statistical Significance: Both the intercept and campaign variable are highly significant ($p<0.001$).

Business insights

1. The results show that as the number of contact attempts increases, customers are more likely to reject the campaign. This is likely because they become annoyed when they receive too many calls.
2. The optimal number of contact attempts is one. Therefore, employees should maximize their efforts during the first call and use a well-structured script to engage customers effectively.
3. The company should explore alternative ways to reach customers instead of relying solely on phone calls. Options include SMS, online advertisements, and other digital marketing strategies.