



Bank Marketing Campaign



Domain

- Dataset contains customer information from a Portuguese banking institution during its marketing campaign through phone calls
- Our study mainly focus on
 - helping banks to campaign more efficiently by choosing the right target population
 - reducing banks' risk of having customers being default on credit¹

¹failing to pay an expected debt



Investigative Questions

1. **What attributes of the customers are associated with the outcome of the previous marketing campaign?**
 - a. **How does the proportion of successful outcome changes as each numerical variable increases?**
 - b. **for each associated variable, what is the group that has the highest successful proportion?**
2. Which kind of people has the highest proportion of successful outcome according to associated attributes from Q1?
 - a. What are the values of the personal information attributes for group of people having highest success proportion?
 - b. How about social and economic attributes?
3. Does the existence of default on credit vary by different marital, education and job groups?
 - a. For each attribute above from Q1, what value(group) results in the highest proportion of existence credit?

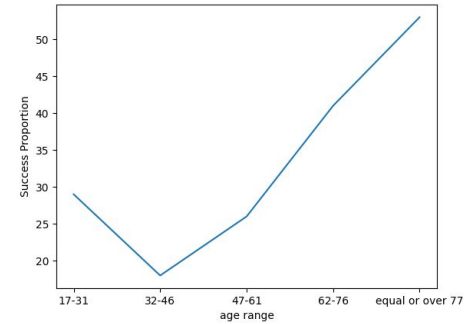
Results Q1

Associated attributes:

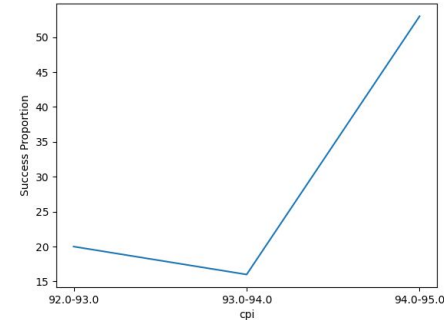
- **age:** 92 or over
- **job:** do not have job
- **education level:** high school or lower
- **the number of contacts this bank performed before the current marketing campaign for this person(pcontact):** multiple times
- **CPI¹:** 94.0 to 95.0
- **CCI²:** -33 to -24
- **whether this person subscribed a term deposit or not(subscribed):** yes

¹ an average change in prices over time that consumers pay for a basket of goods and services

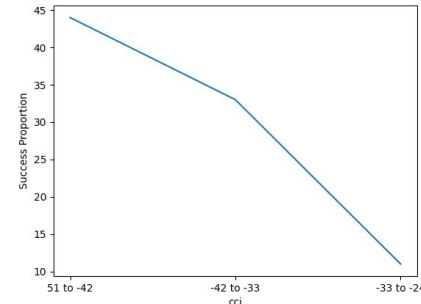
² how optimistic consumers are regarding the bank



age



cpi



cci



Investigative Questions

1. What attributes of the customers are associated with the outcome of the previous marketing campaign?
 - a. How does the proportion of successful outcome changes as each numerical variable increases?
 - b. for each associated variable, what is the group that has the highest successful proportion?
2. **Which kind of people has the highest proportion of successful outcome according to associated attributes from Q1?**
 - a. **What are the values of the personal information attributes for group of people having highest success proportion?**
 - b. **How about social and economic attributes?**
3. Does the existence of default on credit vary by different marital, education and job groups?
 - a. For each attribute above from Q1, what value(group) results in the highest proportion of existence credit?



Results Q2

- **Personal Attributes of People Have the highest proportion of successful outcome (0.81)**

agerange	educationlevel	havejobornot	pcontact	successproportion
equal or over 77	Lower than or eq to High School	HaveNoJob	multiple contact	63
32-46	Lower than or eq to High School	HaveNoJob	multiple contact	63
17-31	Higher than High School	HaveNoJob	multiple contact	67
equal or over 77	Higher than High School	HaveNoJob	multiple contact	71
32-46	Higher than High School	HaveNoJob	multiple contact	81

(last 5 rows)

Results Q2 (Continue..)

- **Social and Economic Attributes of People Have the highest proportion of successful outcome (0.77)**

cpirange	ccirange	subscribed	successproportion	totalsuccess	totaloutcome
93.0-94.0	-51 to -42	no	5	30	573
92.0-93.0	-51 to -42	no	6	104	1727
93.0-94.0	-42 to -33	no	8	74	866
92.0-93.0	-42 to -33	no	21	30	141
92.0-93.0	-33 to -24	no	29	134	459
94.0-95.0	-42 to -33	no	29	75	252
94.0-95.0	-51 to -42	no	29	32	108
92.0-93.0	-51 to -42	yes	31	53	169
93.0-94.0	-51 to -42	yes	36	34	93
92.0-93.0	-42 to -33	yes	54	50	92
92.0-93.0	-33 to -24	yes	60	255	423
93.0-94.0	-42 to -33	yes	64	162	252
94.0-95.0	-42 to -33	yes	70	264	372
94.0-95.0	-51 to -42	yes	77	76	98
(14 rows)					



Overall Results about Right Target of Campaign

When CPI between 94.0 and 95.0, the bank should pay more effort to contact frequently with people that have attributes:

- age between 32 and 46
- no job
- education level higher than high school
- CCI between -51 to -42
- subscribed to a term deposit



Investigative Questions

1. What attributes of the customers are associated with the outcome of the previous marketing campaign?
 - a. How does the proportion of successful outcome changes as each numerical variable increases?
 - b. for each associated variable, what is the group that has the highest successful proportion?
2. Which kind of people has the highest proportion of successful outcome according to associated attributes from Q1?
 - a. What are the values of the personal information attributes for group of people having highest success proportion?
 - b. How about social and economic attributes?
3. **Does the existence of default on credit vary by different marital, education and job groups?**
 - a. **For each attribute above from Q1, what value(group) results in the highest proportion of existence credit?**

Results Q3

- group of people with highest proportion of having existent credit in default:
 - marital status: married
 - education level: professional course
 - job: unemployed
- we do not have sufficient evidence to conclude that people's marital, education and job attributes have a relation with their default on credit.

marital	creditproportion	havecredit	totalcredit
single	0	0	10083
divorced	0	0	3781
married	0.000160788937721085	3	18658
(3 rows)			

Marital Status

Challenges

When GROUP BY too many attributes that each have many distinct values(eg. Job has 11 values), there would be only 1 or 2 observations in some groups which is meaningless to do analysis.

```
job
-----
retired
housemaid
admin.
management
entrepreneur
blue-collar
services
technician
self-employed
student
unemployed
(11 rows)
```

age	education	job	pcontact	successproportion	totalsuccess	totaloutcome
37	high. school	blue-collar	2	0	0	1
24	basic. 9y	admin.	1	100	1	1
49	professional. course	admin.	1	0	0	1
31	basic. 9y	self-employed	1	0	0	2
31	university. degree	admin.	5	100	1	1
38	basic. 9y	self-employed	1	0	0	3
40	basic. 9y	blue-collar	2	50	1	2
51	basic. 4y	retired	1	100	1	1
44	basic. 9y	technician	1	100	1	1

(first 9 rows)

Solution

- we divide the values of attributes into useful groups, such as
 - splitting 11 values of jobs into 'HaveJob' and 'HaveNoJob'
 - splitting education into 'Lower than or eq to High School' and 'Higher than High School'

agerange	educationlevel	havejobornot	pcontact	successproportion	totalsuccess	totaloutcome
32-46	Higher than High School	HaveJob	one contact	18	172	931
32-46	Lower than or eq to High School	HaveNoJob	one contact	18	15	79
47-61	Higher than High School	HaveJob	one contact	21	80	370
17-31	Higher than High School	HaveJob	one contact	26	120	452
32-46	Higher than High School	HaveNoJob	one contact	27	12	43
47-61	Lower than or eq to High School	HaveNoJob	one contact	28	29	101
62-76	Higher than High School	HaveNoJob	one contact	29	13	44
17-31	Higher than High School	HaveNoJob	one contact	31	11	35
47-61	Lower than or eq to High School	HaveNoJob	multiple contact	33	10	30

Lessons

- Use SQL to solve real world question, in the future, can transform in R to make more specific and advanced analysis
- As mentioned in Challenges, sometimes after GROUP BY a list of attributes, each resulting group only has a few observations. We learned that we can divide the values of attributes into useful groups such as splitting jobs into 'HaveJob' and 'HaveNoJob'.

agerange	educationlevel	havejobornot	pcontact	successproportion	totalsuccess	totaloutcome
32-46	Higher than High School	HaveJob	one contact	18	172	931
32-46	Lower than or eq to High School	HaveNoJob	one contact	18	15	79
47-61	Higher than High School	HaveJob	one contact	21	80	370
17-31	Higher than High School	HaveJob	one contact	26	120	452
32-46	Higher than High School	HaveNoJob	one contact	27	12	43
47-61	Lower than or eq to High School	HaveNoJob	one contact	28	29	101
62-76	Higher than High School	HaveNoJob	one contact	29	13	44
17-31	Higher than High School	HaveNoJob	one contact	31	11	35
47-61	Lower than or eq to High School	HaveNoJob	multiple contact	33	10	30

(first 9 rows)



Questions?