Data and Business Understanding

Sebastián Castro
Omar Jazouli
Freddy Tapia
Ignacio Solórzano

14-April-2021

- Data understanding
- Clients Data Analysis
- Campaigns Data Analysis
- Recommendations

Data Understanding

Data Understanding

Clients Data

1. age: int
2. job: type of job (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')
3. marital: marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)
4. education:(categorical: 'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','university.degree','unknown')
5. default: has credit in default? (categorical: 'no','yes','unknown')
6. housing: has housing loan? (categorical: 'no','yes','unknown')
7. loan:has personal loan? (categorical: 'no','yes','unknown')

Data related with the last contact

8. contact: contact communication type (categorical: 'cellular','telephone')
9. day_of_week: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')
10. month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
11. duration: last contact duration, in seconds (numeric). Important note: 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 y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

Data Understanding

Data related with the last campaign

12. campaign:number of contacts performed during this campaign and for this client (numeric, includes last contact)
13. 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)
14. previous: number of contacts performed before this campaign and for this client (numeric)
15. poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

Data related with social and economic context¶

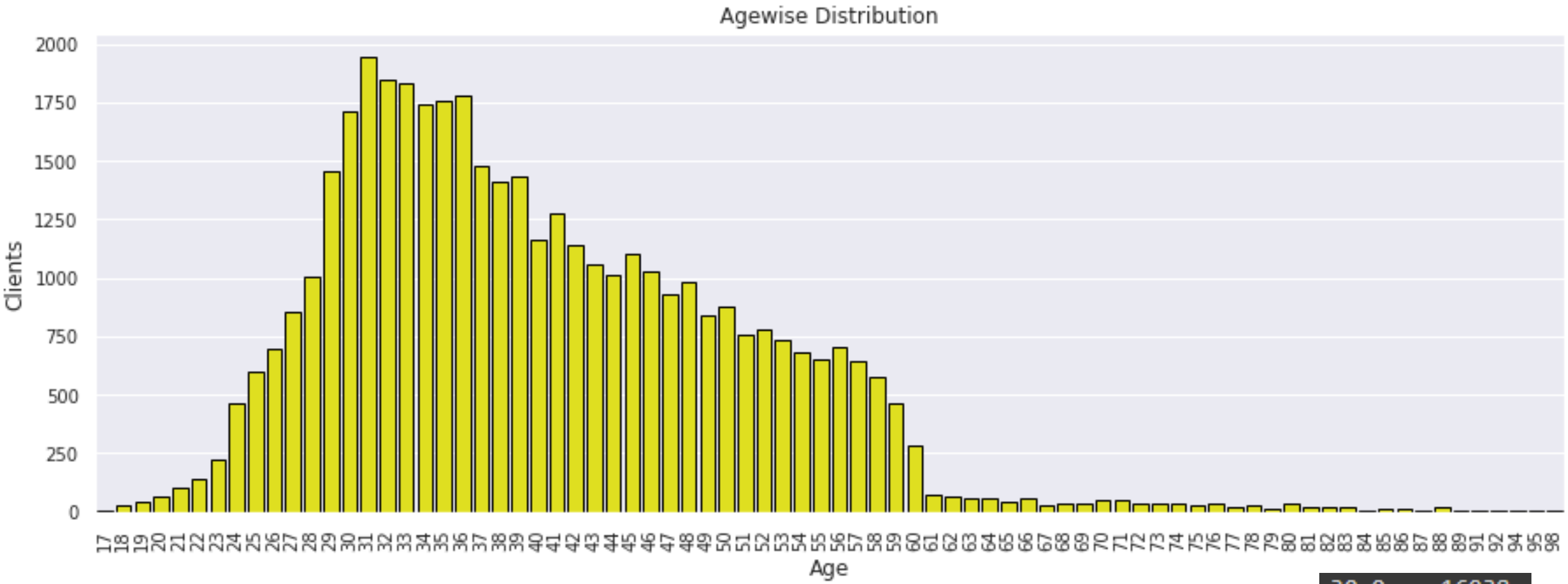
16. emp.var.rate: employment variation rate - quarterly indicator (numeric)
17. cons.price.idx: consumer price index - monthly indicator (numeric)
18. cons.conf.idx: consumer confidence index - monthly indicator (numeric)
19. euribor3m: euribor 3 month rate - daily indicator (numeric)
20. nr.employed: number of employees - quarterly indicator (numeric)

Result of the current campaign

21. y - has the client subscribed a term deposit? (binary: 'yes','no')

Clients Data Analysis

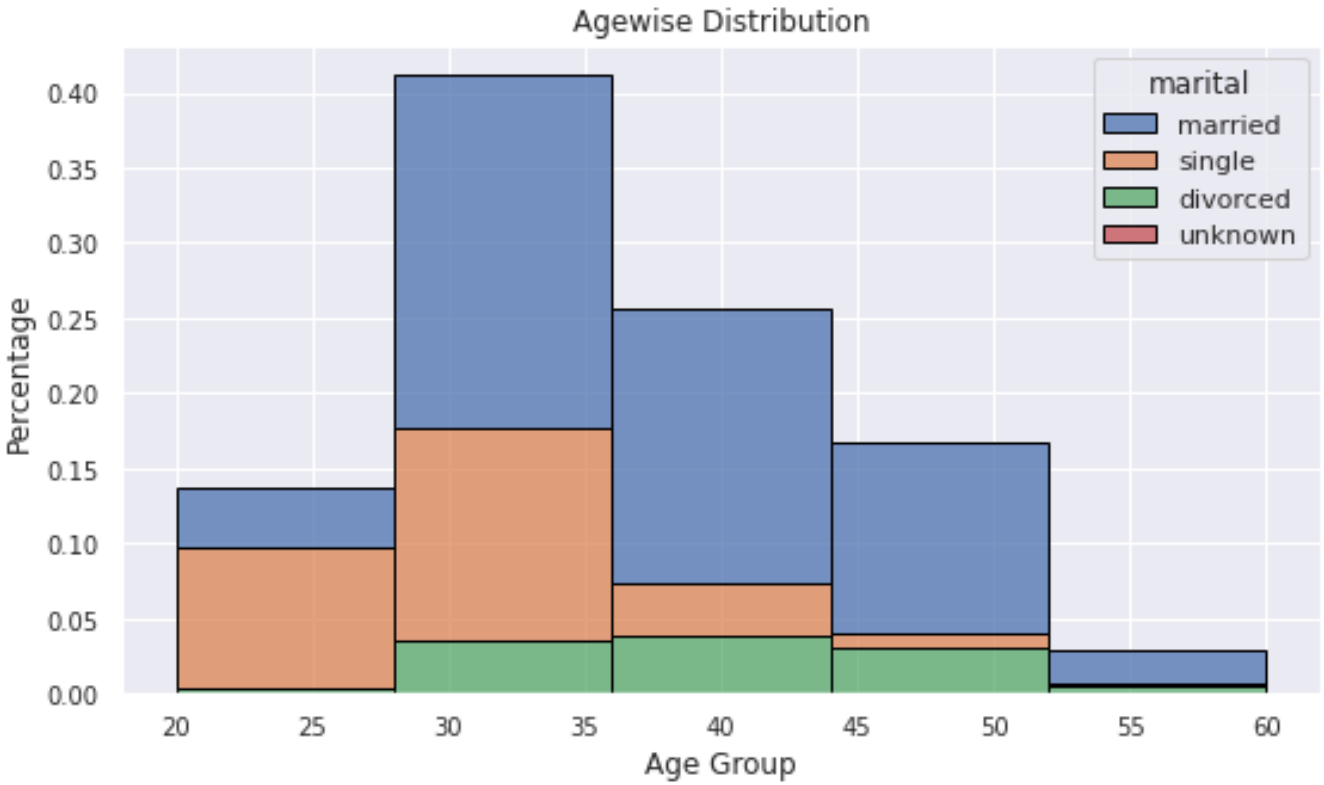
Clients Data Analysis

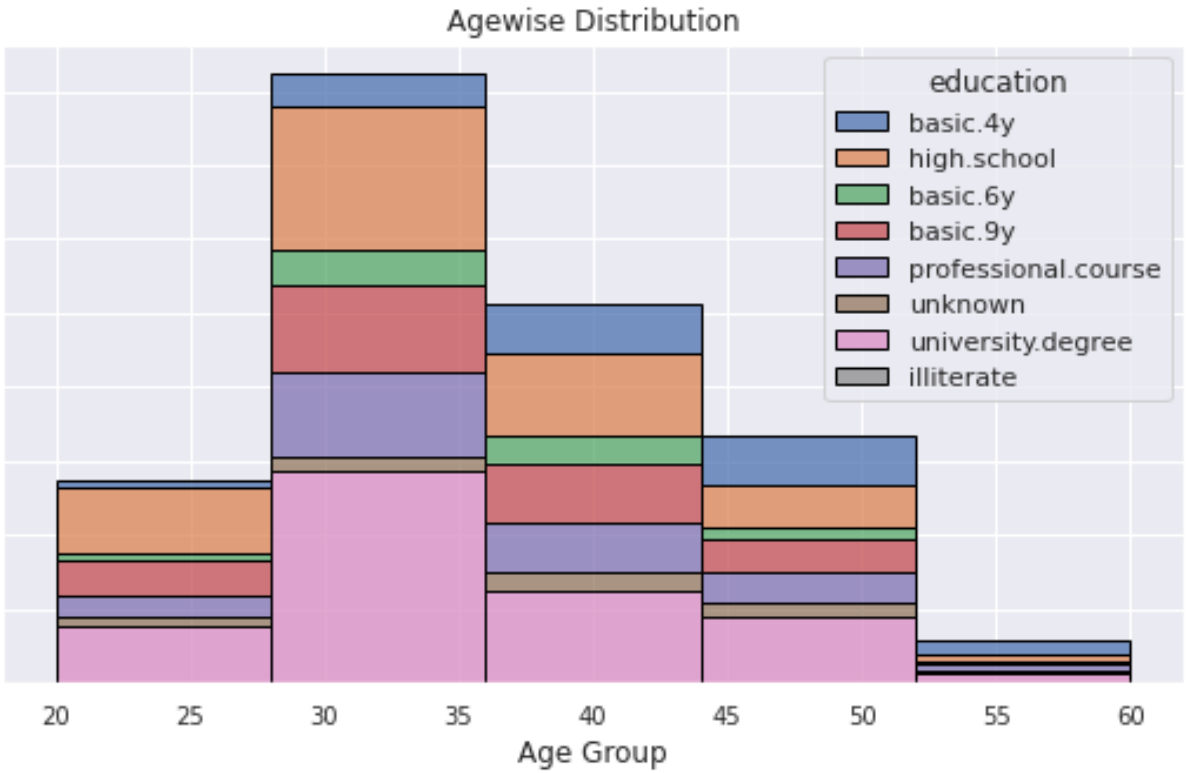


In order to avoid skewed data, we group the ages into different age ranges

30.0	16938
40.0	10526
50.0	6862
20.0	5669
60.0	1193

Agewise Distributions – Marital & Education

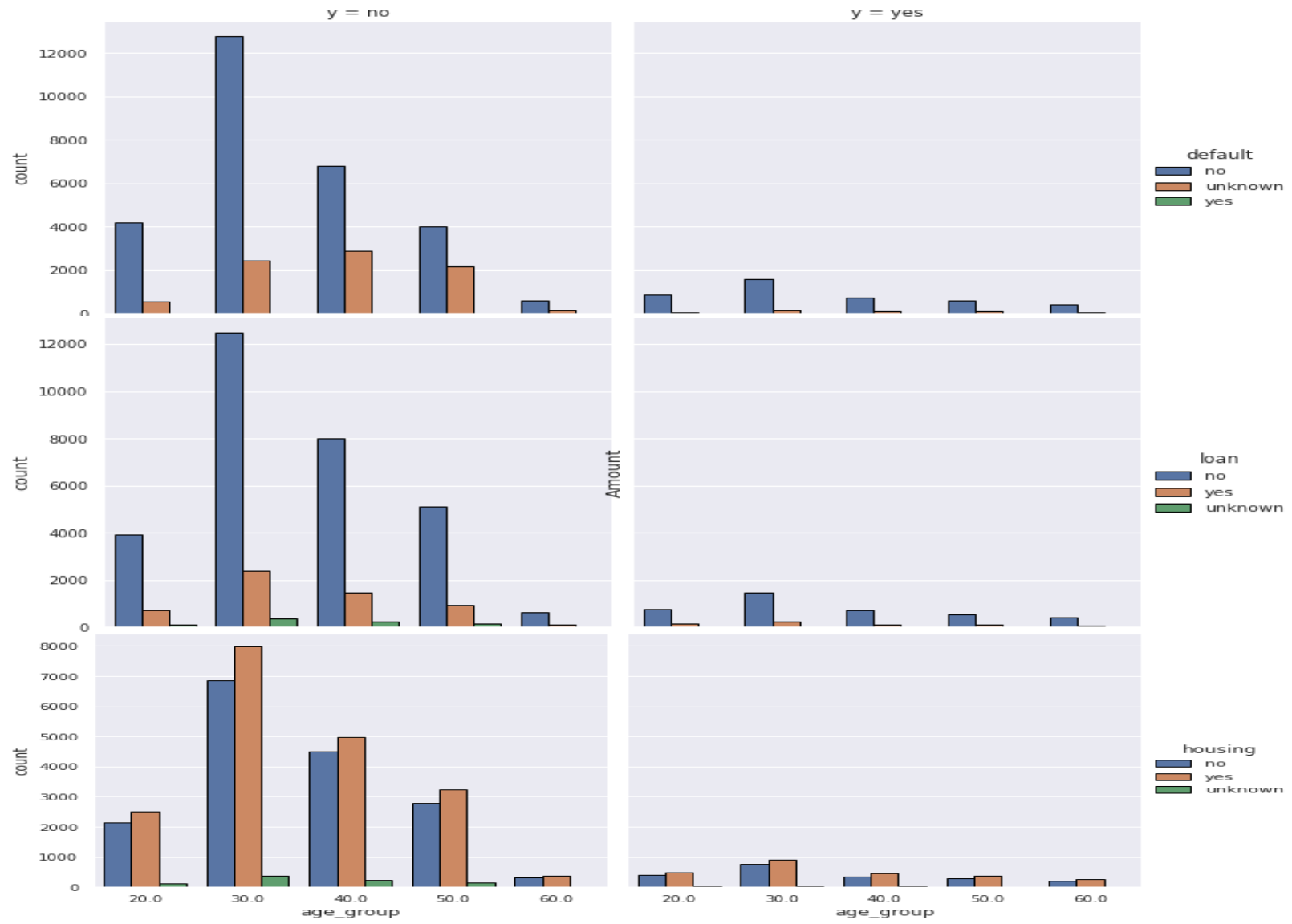




Marital status distribution through age ranges

Client's education distribution through age ranges

Default – Loan – Housing



Default distribution agewise.
Left: Customers with/out default who did not buy the term deposit
Right: Customers with/out default who bought the term deposit

Loan distribution agewise.
Left: Customers with/out loan who did not buy the term deposit
Right: Customers with/out loan who bought the term deposit

Housing distribution agewise.
Left: Customers with/out house who did not buy the term deposit
Right: Customers with/out housing who bought the term deposit

% Subscription vs. % Agewise Range

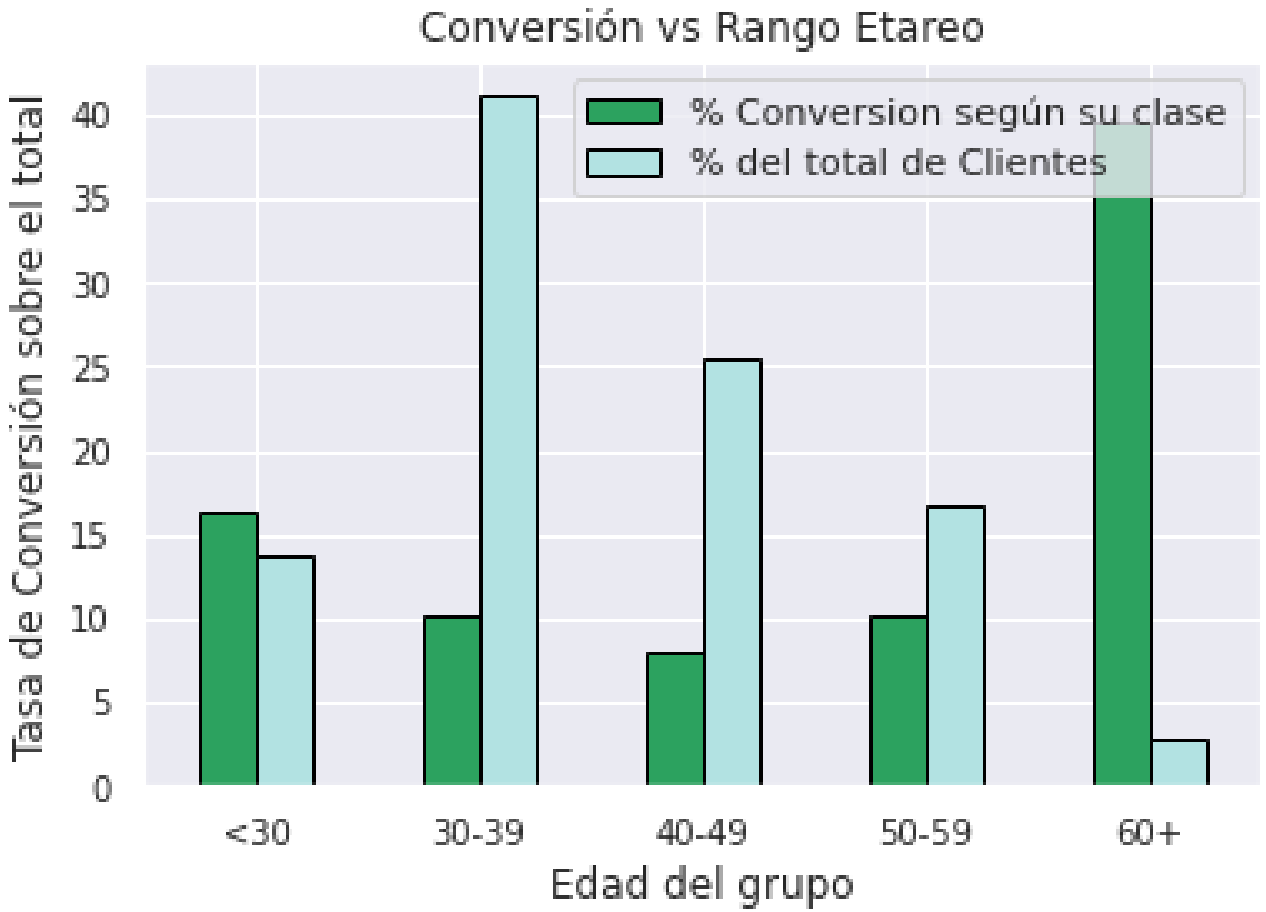
Conclusions:

• The most efficient conversion is in the group +60. This class converts 39.6%. [**4 out of 10 people '+60' who connect with the bank, sign up**]. The point against, is that this class represents 2.89% of total customers

• Of all the conversions, the 30's group is the highest .In other words, **conversions in this group represent almost 40% of the bank's customers.** The downside is that it is not very efficient, since the conversion rate of its class is 10.12%.

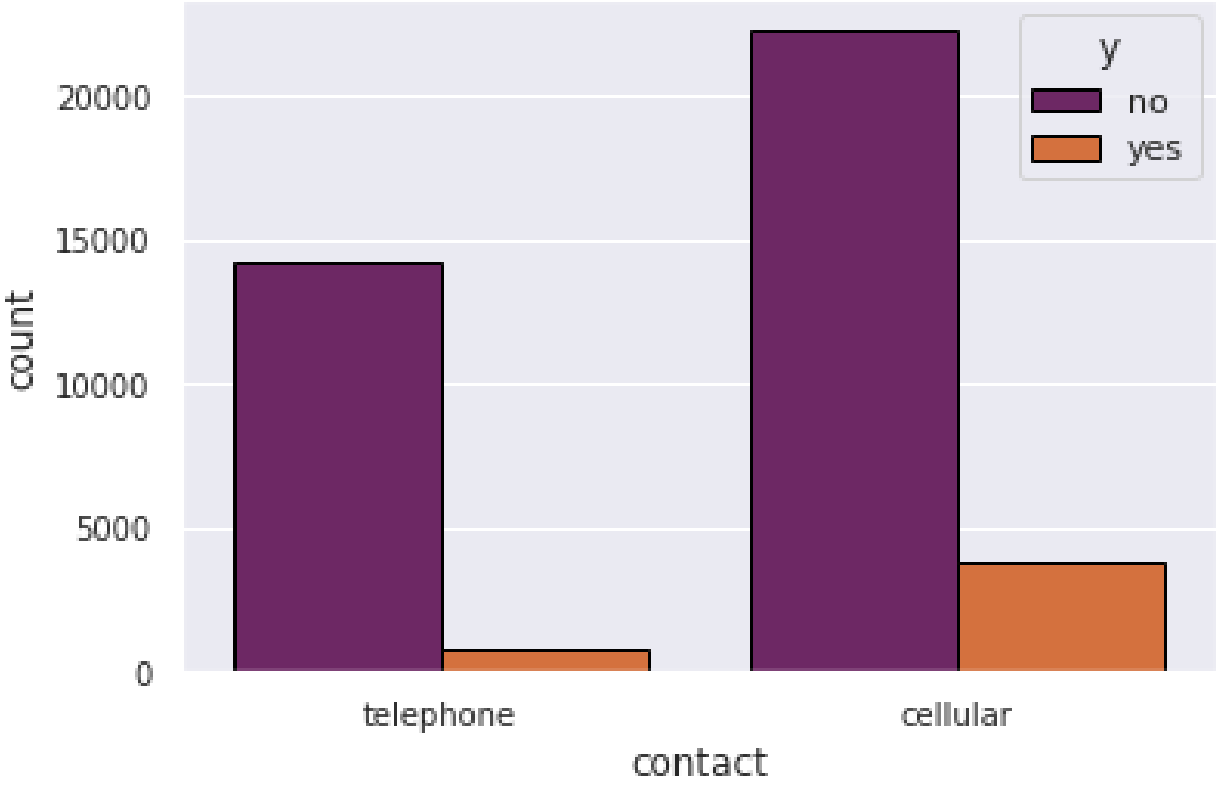
•The least efficient conversion is in the group of 40 with 7.9%

	% del total de Clientes	% Conversion según su clase	age
20.0	13.763718	16.263891	20
30.0	41.123628	10.125162	30
40.0	25.555987	7.923238	40
50.0	16.660192	10.157389	50
60.0	2.896475	39.564124	60



Campaign Data Analysis

Contact Data Analysis

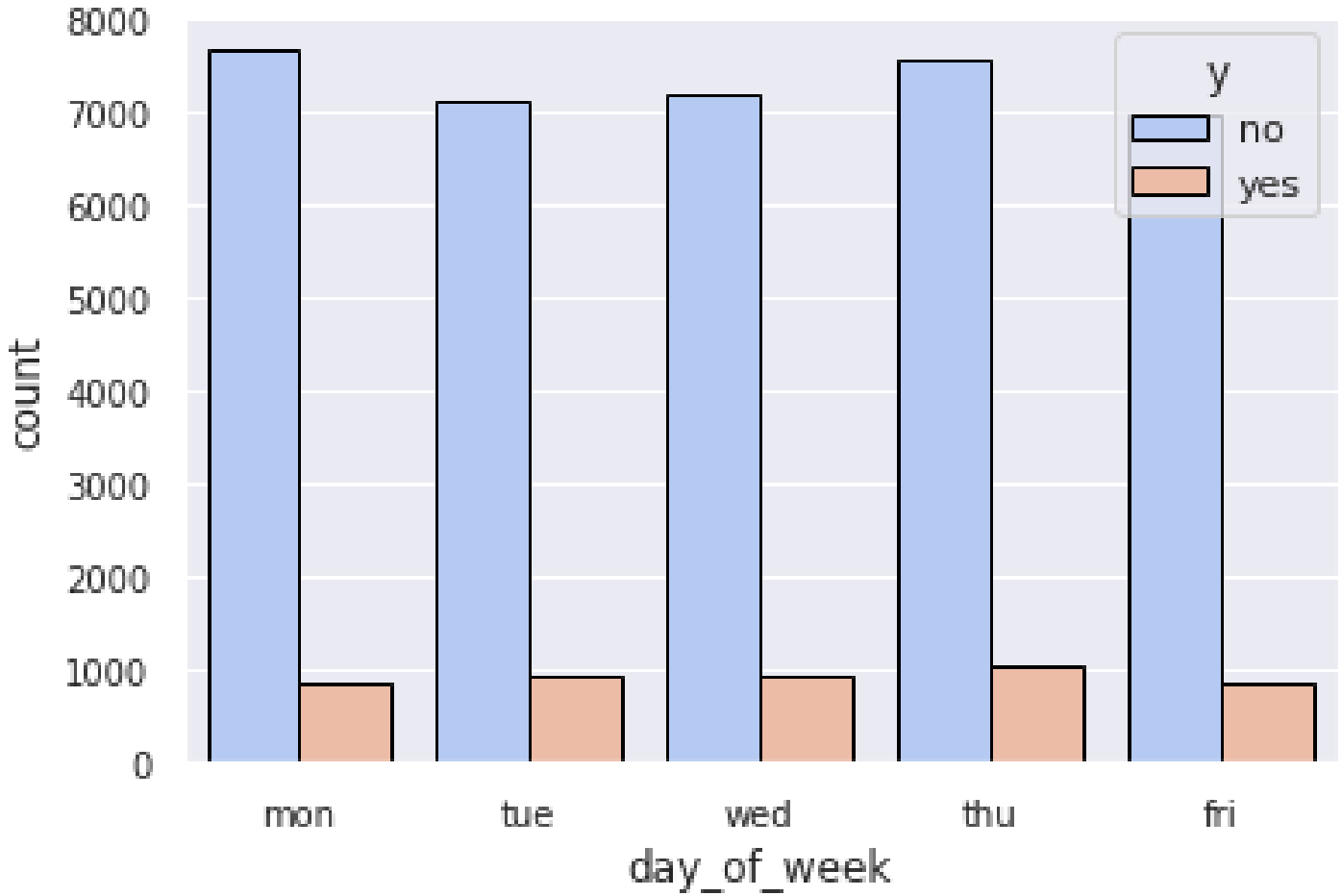


Telephone's conversions are 5% of all telephone's calls

Cellular's conversions are 15% of all Cellular's calls

Daywise distribution

Due to it’s constant distribution we could not say that there is any unusual day to make the calls in the meanwhile…



Recommendations

First sight recommendations

- Aim the calls to the ethereal extremes.

It makes sense that age extremes are risk averse, the young because they can save for education and the elderly because they need to ensure their retirement.

- Young people have more loans and are also more in default

Thank You

Data Glacier

Your Deep Learning Partner