

Cross selling recommendation

INTERNSHIP BATCH: [LISUM09](#)

DATA GLACIER VIRTUAL INTERNSHIP 2022

Outline

- Problem Statement
- Datasets Information
- EDA (Exploratory data analysis)

Problem description & Business understanding

XYZ Credit Union, located in Latin America, does well in selling banking products such as: credit cards, deposit accounts, retirement accounts, safe deposit boxes, etc. However, after statistics, they found that their existing customers basically only buy one product, which means that the bank does not perform well in cross-selling. So XYZ Credit Union wants analysts to build models such as marketing models through machine learning to solve their problems.

Datasets Information

Train.csv details:

Total number of observations	13,647,309
Total number of files	1
Total number of features	48
Base format of the file	csv
Size of the data	2.13 GB

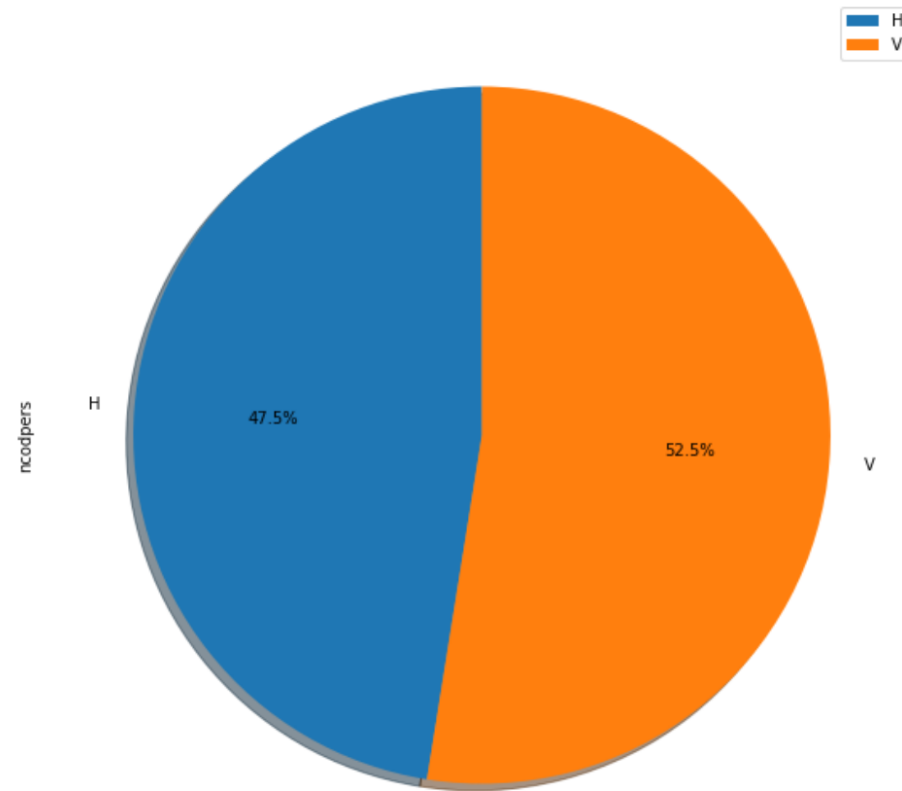
Test.csv details:

Total number of observations	929,615
Total number of files	1
Total number of features	24
Base format of the file	csv
Size of the data	105 MB

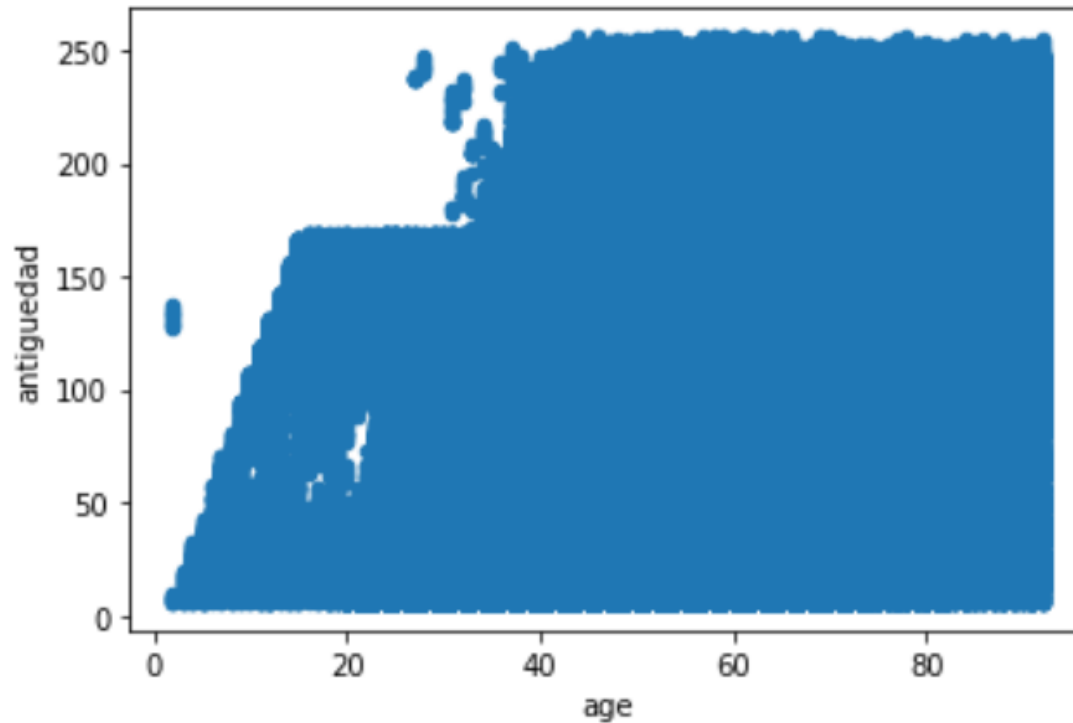
EDA (Exploratory data analysis)

Customer's sex - Column Name : sexo

ncodpers	
sexo	
H	3421063
V	3784510



Customer's Age vs. Customer seniority (in months)

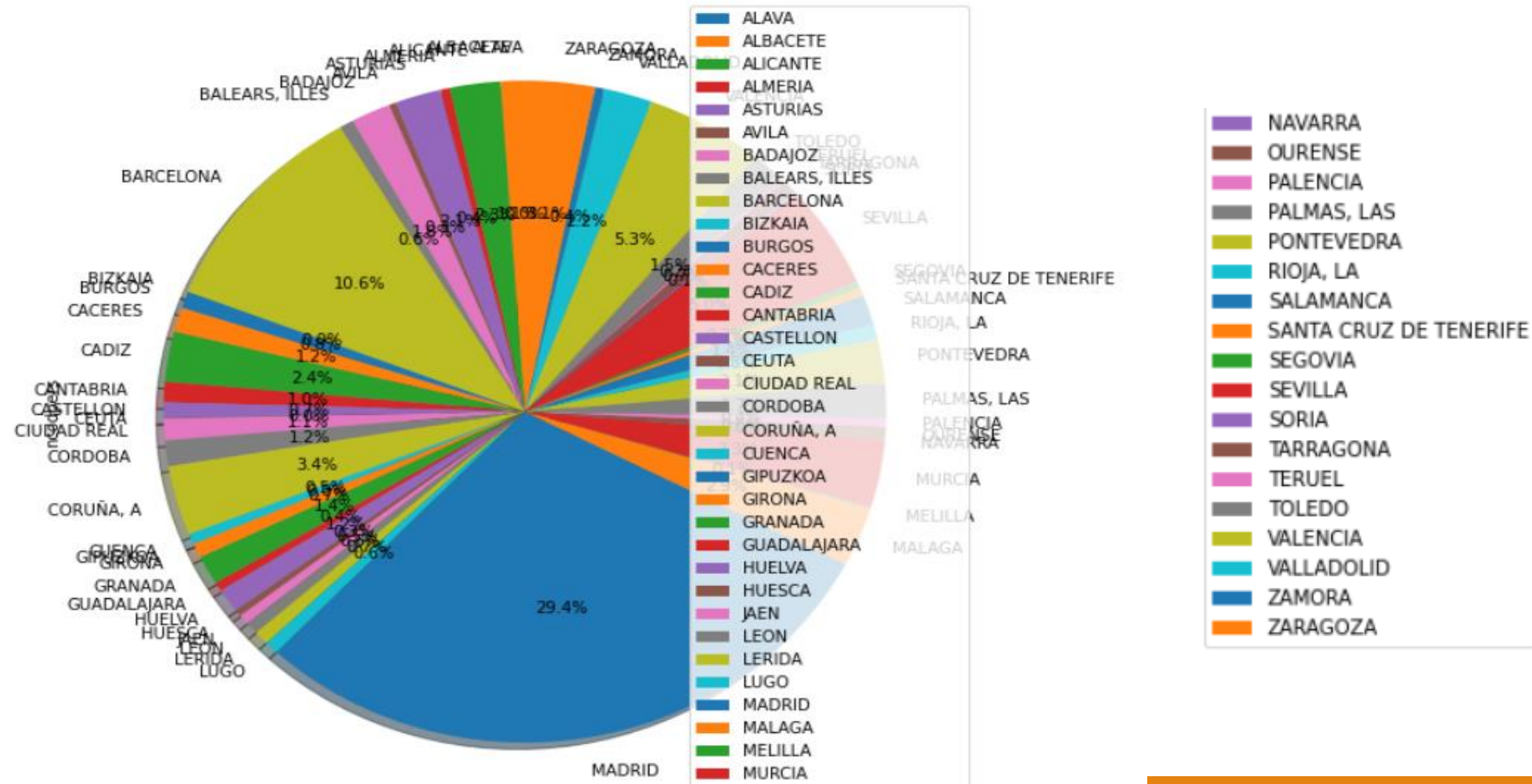


Customer's Province name (Count)

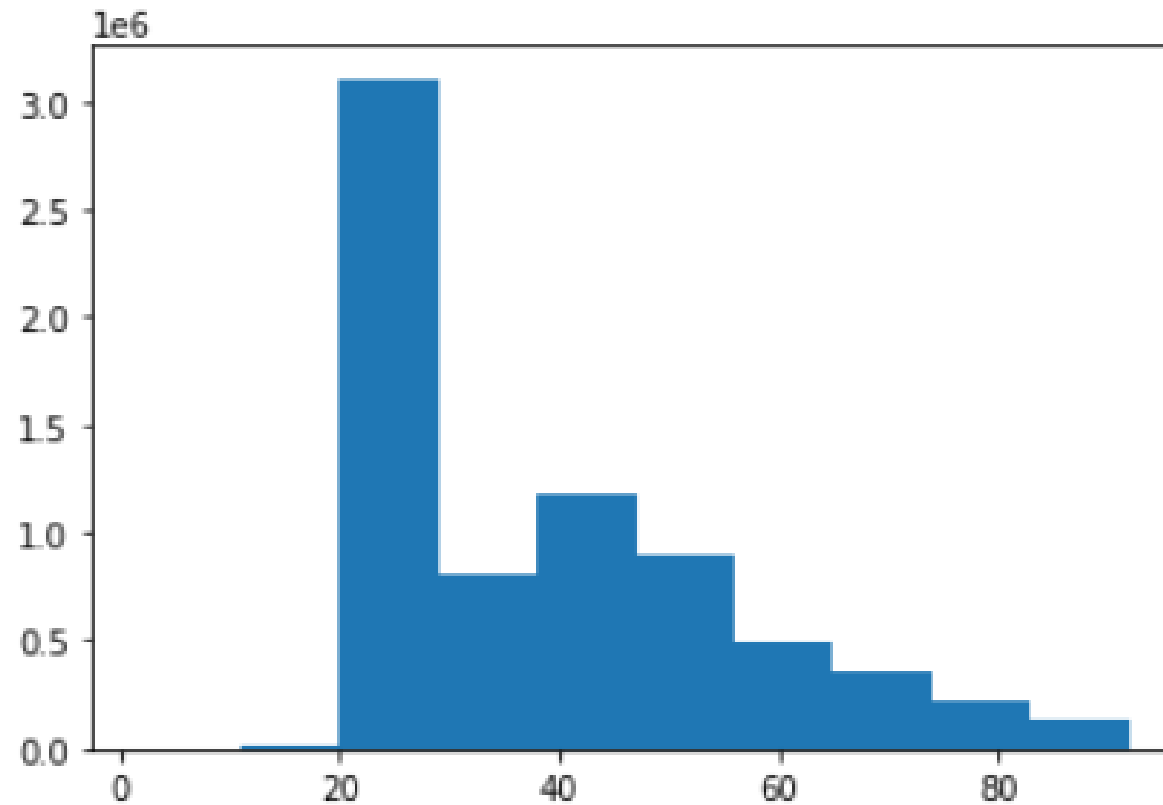
ALAVA	5
ALBACETE	77945
ALICANTE	162647
ALMERIA	30581
ASTURIAS	148461
AVILA	22670
BADAJOS	128064
BALEARS, ILLES	45551
BARCELONA	760799
BIZKAIA	50
BURGOS	58629
CACERES	85989
CADIZ	175633
CANTABRIA	72042
CASTELLON	53935
CEUTA	2927
CIUDAD REAL	76075
CORDOBA	89996
CORUÑA, A	244916
CUENCA	36099
GIPUZKOA	22
GIRONA	50819
GRANADA	100885
GUADALAJARA	32090
HUELVA	83041
HUESCA	23183

GRANADA	100885
GUADALAJARA	32090
HUELVA	83041
HUESCA	23183
JAEN	34301
LEON	42448
LERIDA	49431
LUGO	45581
MADRID	2119228
MALAGA	210317
MELILLA	4824
MURCIA	236278
NAVARRA	41
OURENSE	46517
PALENCIA	30866
PALMAS, LAS	121606
PONTEVEDRA	153945
RIOJA, LA	54366
SALAMANCA	101863
SANTA CRUZ DE TENERIFE	29166
SEGOVIA	23062
SEVILLA	363568
SORIA	9418
TARRAGONA	49443
TERUEL	13735
TOLEDO	108268
VALENCIA	384133
VALLADOLID	155422
ZAMORA	29474
ZARAGOZA	225218

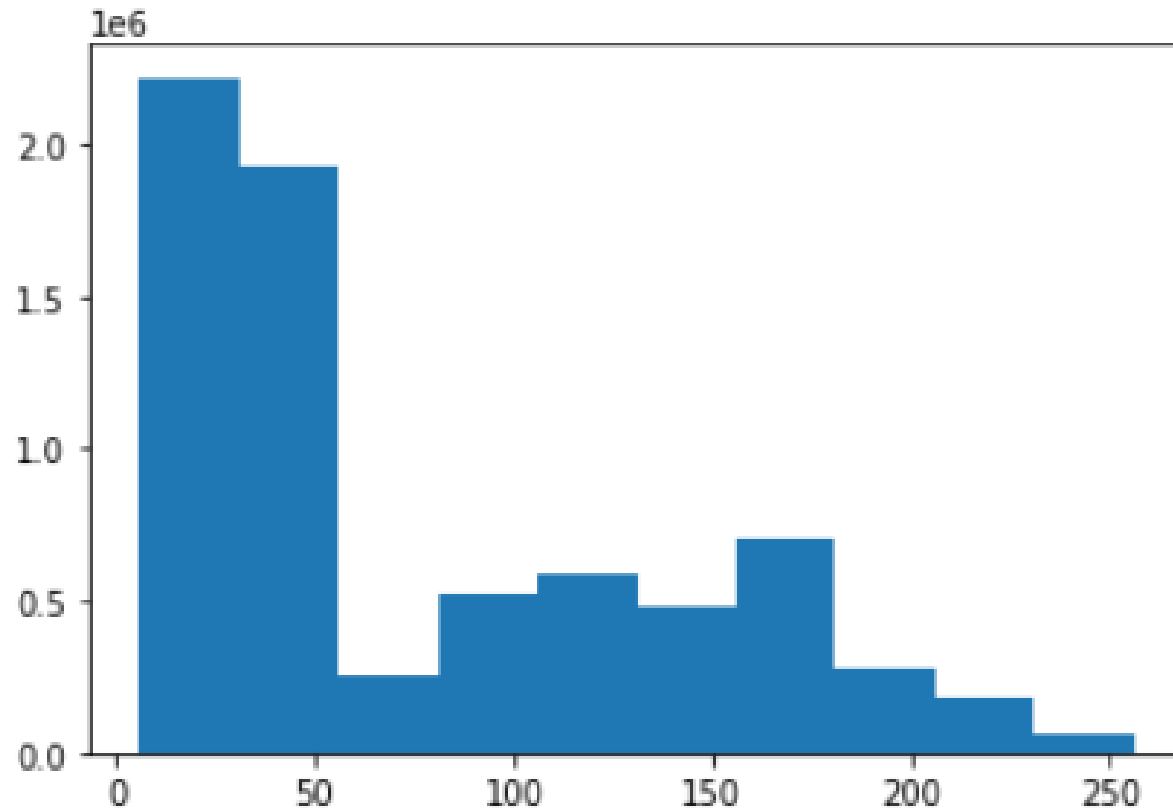
Customer's Province name (Pie chart)



Customer's age

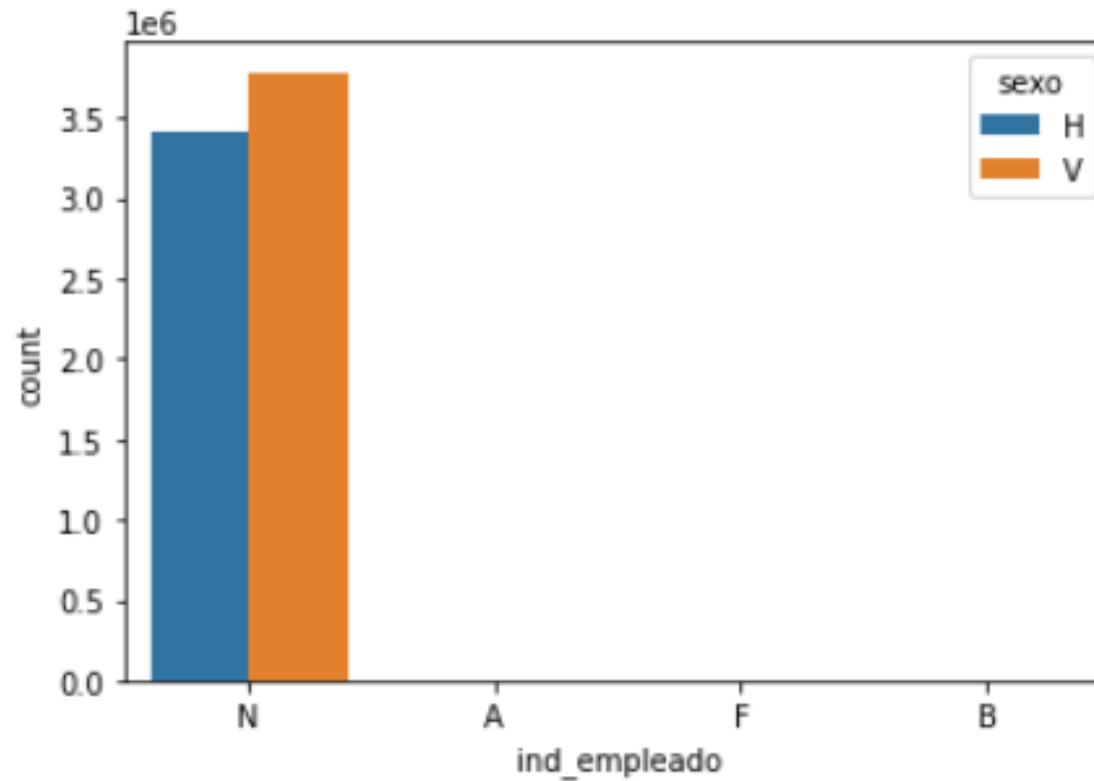


Customer seniority (in months)



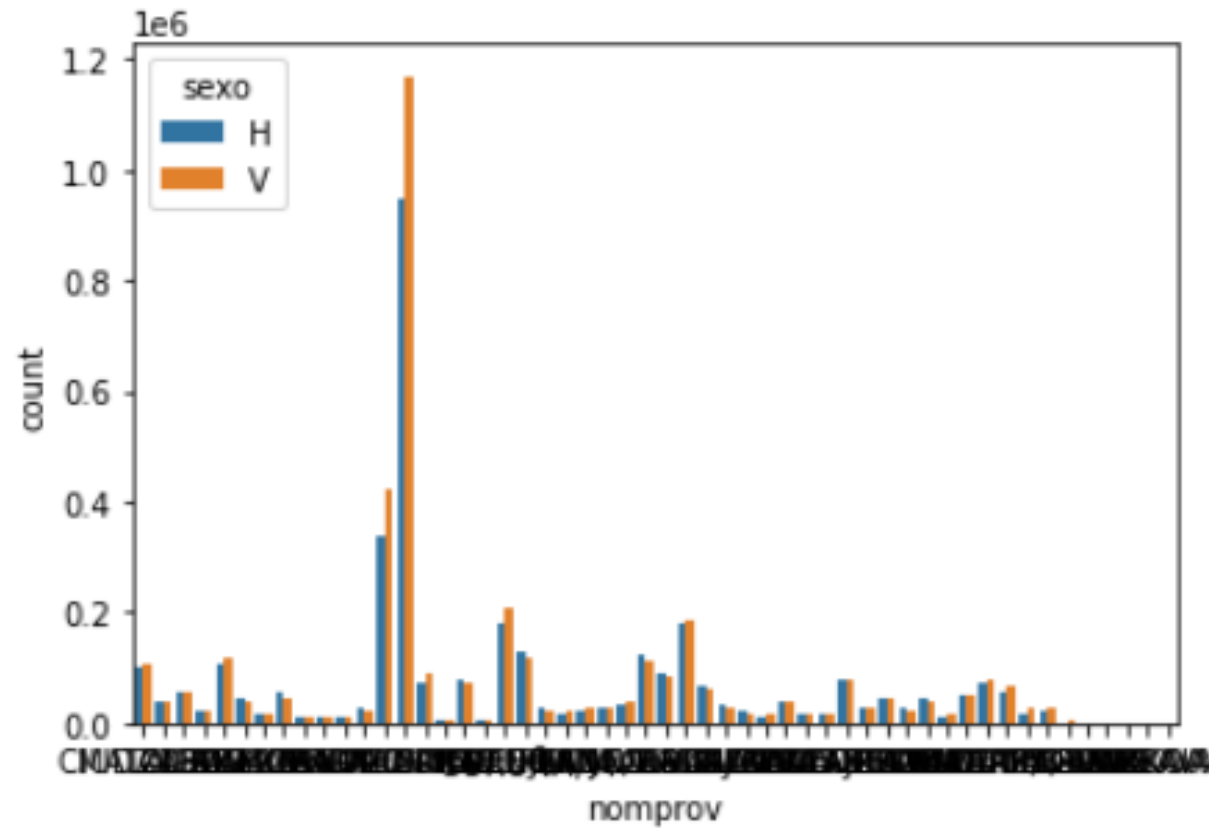
Employee index vs. Customer's sex

Employee index: A active, B ex employed, F filial, N not employee, P pasive



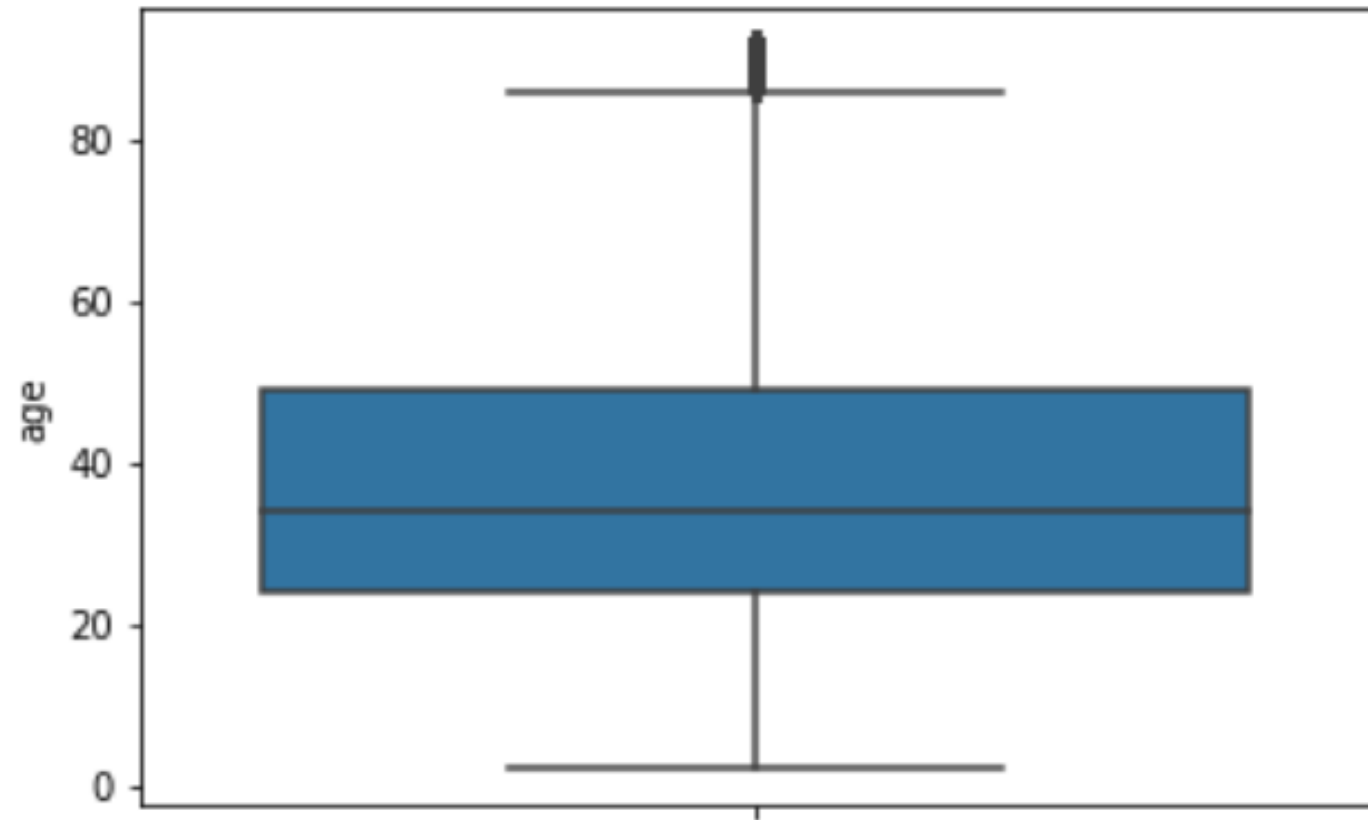
Customer's Province name vs. Customer's sex

Max count: MADRID



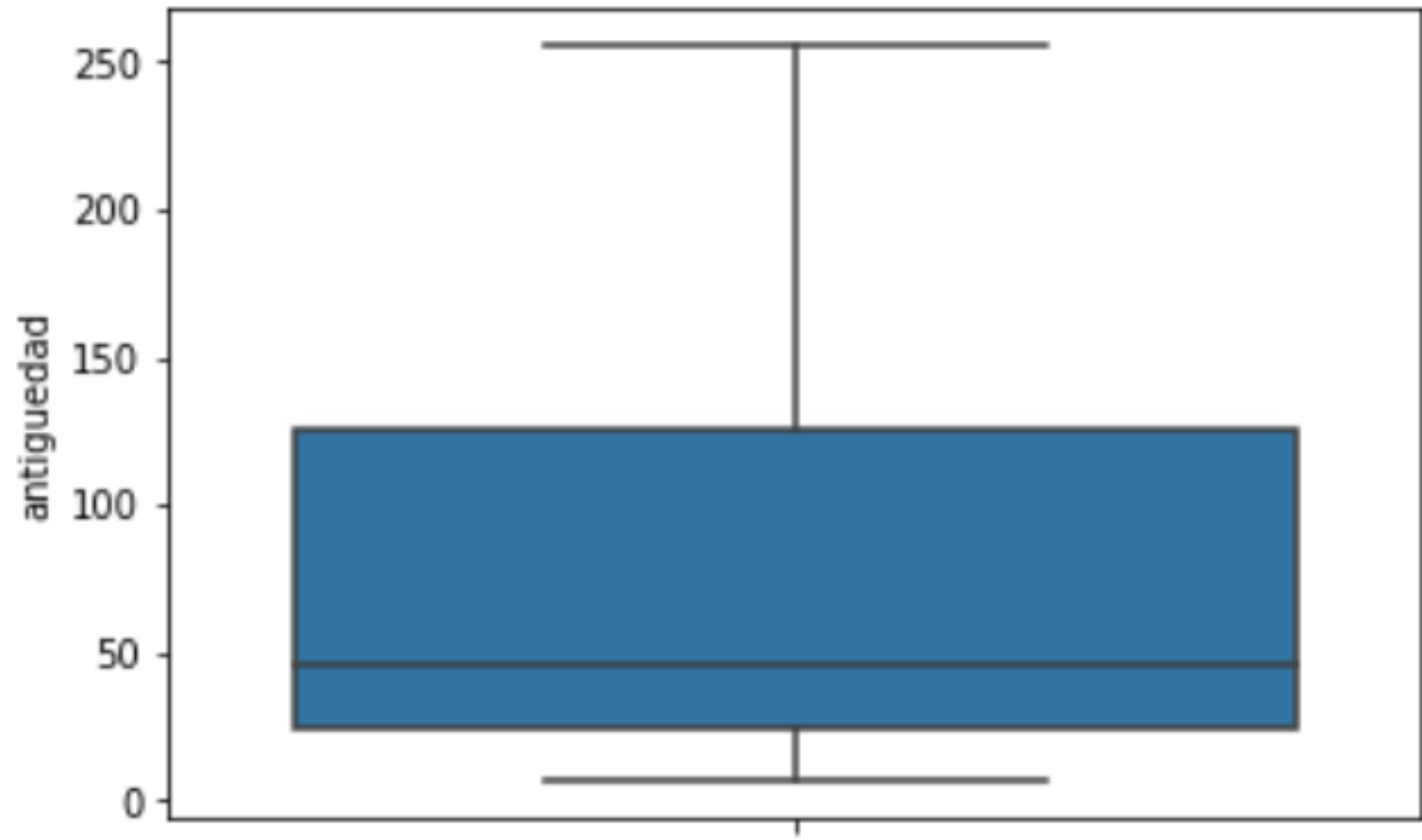
Boxplot - Age

Mean:
About 35

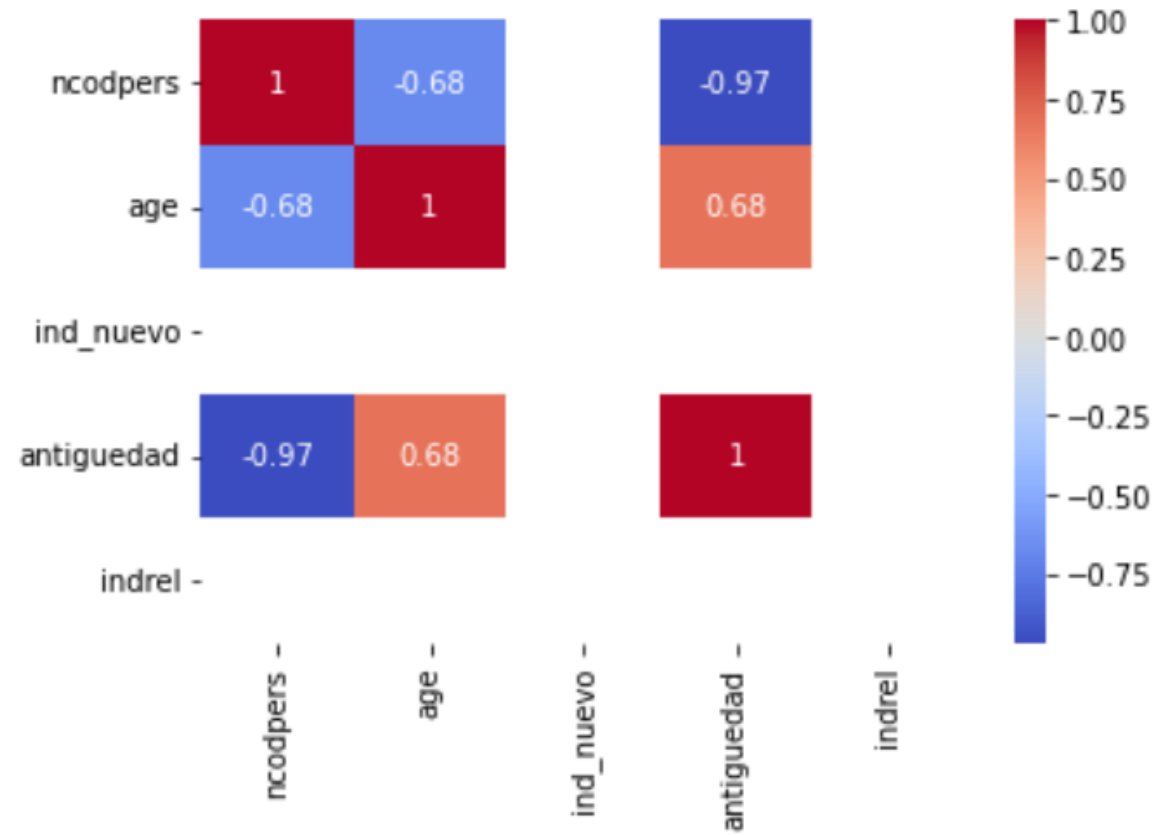


Boxplot - Customer seniority (in months)

Mean:
About 50



Heatmap



Thank you

