

# Exploratory Data Analysis

Customer Segmentation

# Team Members

- Ray Ng
  - rayng2018@gmail.com
  - Canada
  - University of British Columbia
  - Specialization: Data Science
- Rita Uzoka
  - rita.uzoka@yahoo.com
  - United Kingdom
  - Sheffield Hallam University
  - Specialization: Data Science
- Fatemeh Bagheri
  - f.bagheri13@gmail.com
  - France
  - Université Jean Monnet St Etienne - Université de Lyon
  - Your Specialization: Data Science

# Problem

## Customer Segmentation

XYZ bank wants to roll out Christmas offers to their customers. But the bank does not want to roll out the same offer to all customers, instead they want to roll out personalized offers to particular sets of customers. If they manually start understanding the category of customer then this will not be efficient and also they will not be able to uncover the hidden pattern in the data ( pattern which groups certain kinds of customer in one category). Bank approached ABC analytics company to solve their problem. Bank also shared information with ABC analytics that they don't want more than 5 groups as this will be inefficient for their campaign.

# What kind of data does each column hold in the raw data?

Unnamed: 0: [ 0 1 2 ... 999997 999998 999999]

fecha\_dato: ['2015-01-28' '2015-02-28']

ncodpers: [1375586 1050611 1050612 ... 1149999 1150908 1183305]

ind\_empleado: ['N' nan 'A' 'B' 'F' 'S']

pais\_residencia: ['ES' nan 'CA' 'CH' 'CL' 'IE' 'AT' 'NL' 'FR' 'GB' 'DE' 'DO' 'BE' 'AR' 'VE'

'US' 'MX' 'BR' 'IT' 'EC' 'PE' 'CO' 'HN' 'FI' 'SE' 'AL' 'PT' 'MZ' 'CN'

# How many NA are in each column?

Unnamed: 0: 0

fecha\_dato: 0

ncodpers: 0

ind\_empleado: 10782

pais\_residencia: 10782

sexo: 10786

age: 10782

fecha\_alta: 10782

ind\_nuevo: 10782

antiguedad: 10782

indrel: 10782

ult\_fec\_cli\_1t: 998899

indrel\_1mes: 10782

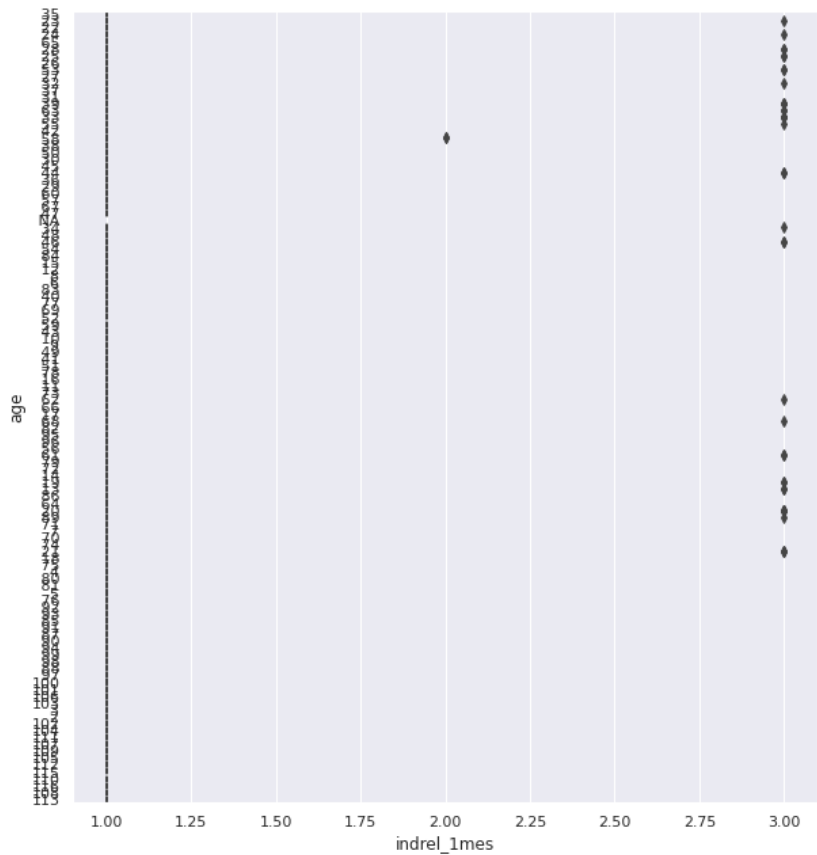
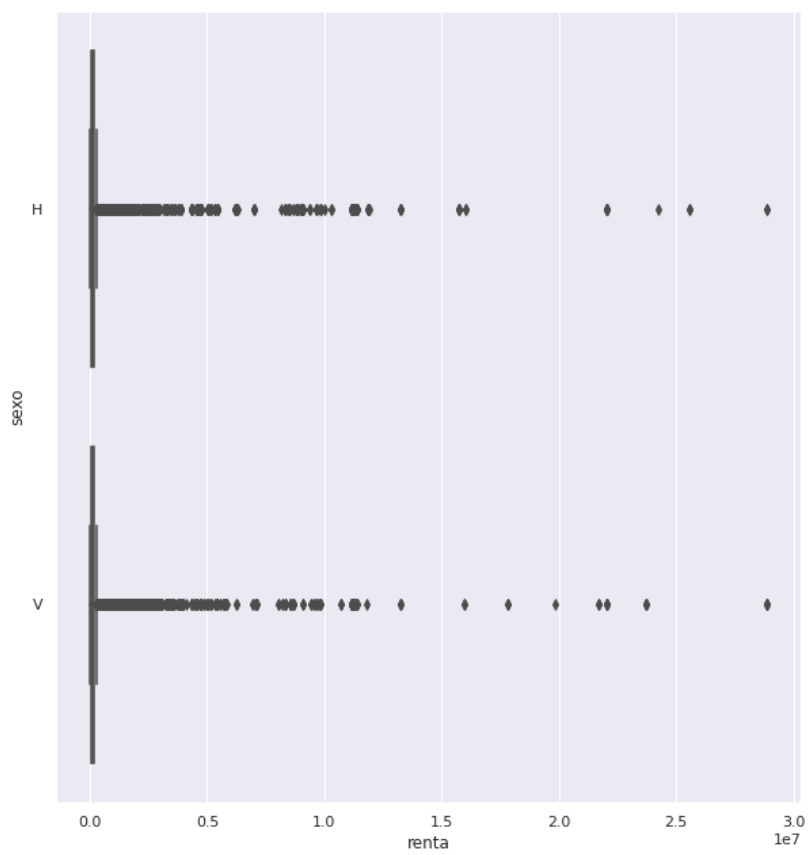
# Table 1

[illegible]

# Correlation of variables

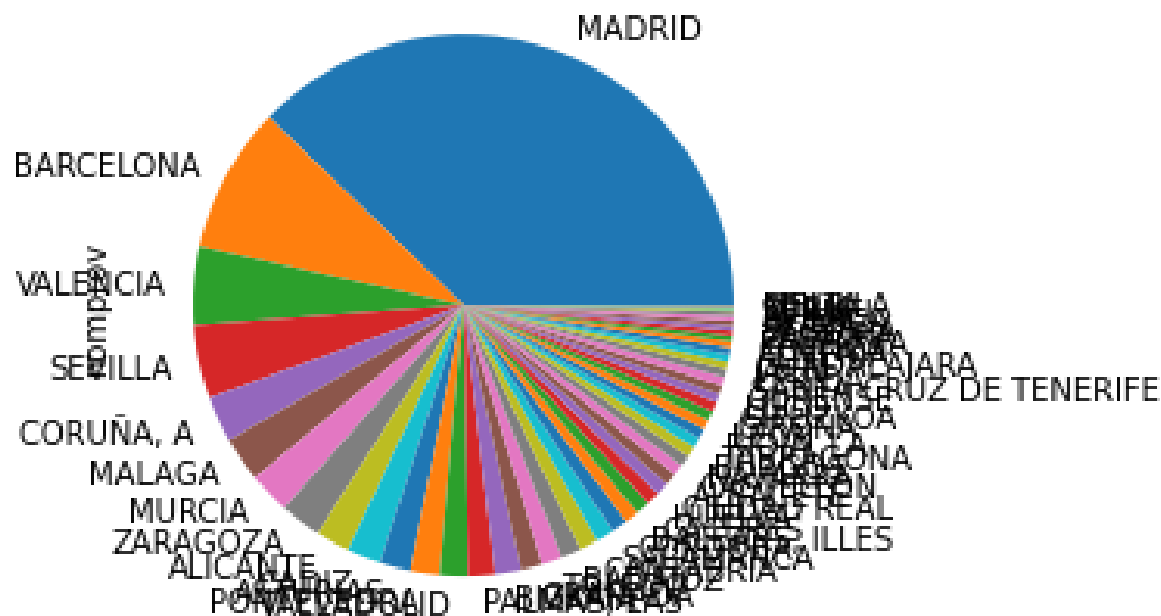
	Unnamed: 0	ncodpers	ind_nuevo	indrel	indrel_1mes	tipodom	codprov	ind_actividad_cliente	renta	ind_ahor_fin_ult1	...	ind_hip_fin_ult1	ind_pla_fin_ult1	ind_pre_fin_ult1	ind_reca_fin_ult1	ind_tjcr_fin_ult1	ind_valo_fin_ult1	ind_viv_fin_ult1	ind_nomina_ult1	ind_nom_pens_ult1
Unnamed: 0	1.000000	-0.447119	0.008312	-0.007374	-0.000538	NaN	0.035614	0.116265	0.044528	0.005893	...	0.043927	0.048669	0.029035	0.092628	0.082870	0.077629	0.035114	0.069204	0.074190
ncodpers	-0.447119	1.000000	0.002898	0.011554	0.001253	NaN	0.040761	-0.187022	-0.088417	-0.013469	...	-0.100045	-0.125874	-0.027066	-0.189129	-0.190208	-0.177782	-0.073552	-0.138254	-0.152622
ind_nuevo	0.008312	0.002898	1.000000	0.026681	0.268051	NaN	0.000279	0.008165	-0.000986	-0.000296	...	-0.002224	-0.002701	-0.001432	-0.003034	-0.004998	-0.004265	-0.001789	-0.004081	-0.004497
indrel	-0.007374	0.011554	0.026681	1.000000	0.004462	NaN	0.001728	-0.030518	-0.000745	-0.000447	...	-0.002745	-0.003067	-0.001219	-0.005014	-0.005641	-0.003062	-0.002699	-0.005771	-0.005919
indrel_1mes	-0.000538	0.001253	0.268051	0.026681	1.000000	NaN	0.000779	-0.000859	0.000615	-0.000088	...	-0.000659	-0.000800	-0.000424	-0.001801	-0.001127	-0.001333	-0.000530	-0.000162	-0.000319

# CHECKING FOR OUTLIERS IN (2)NUMERICAL DATA



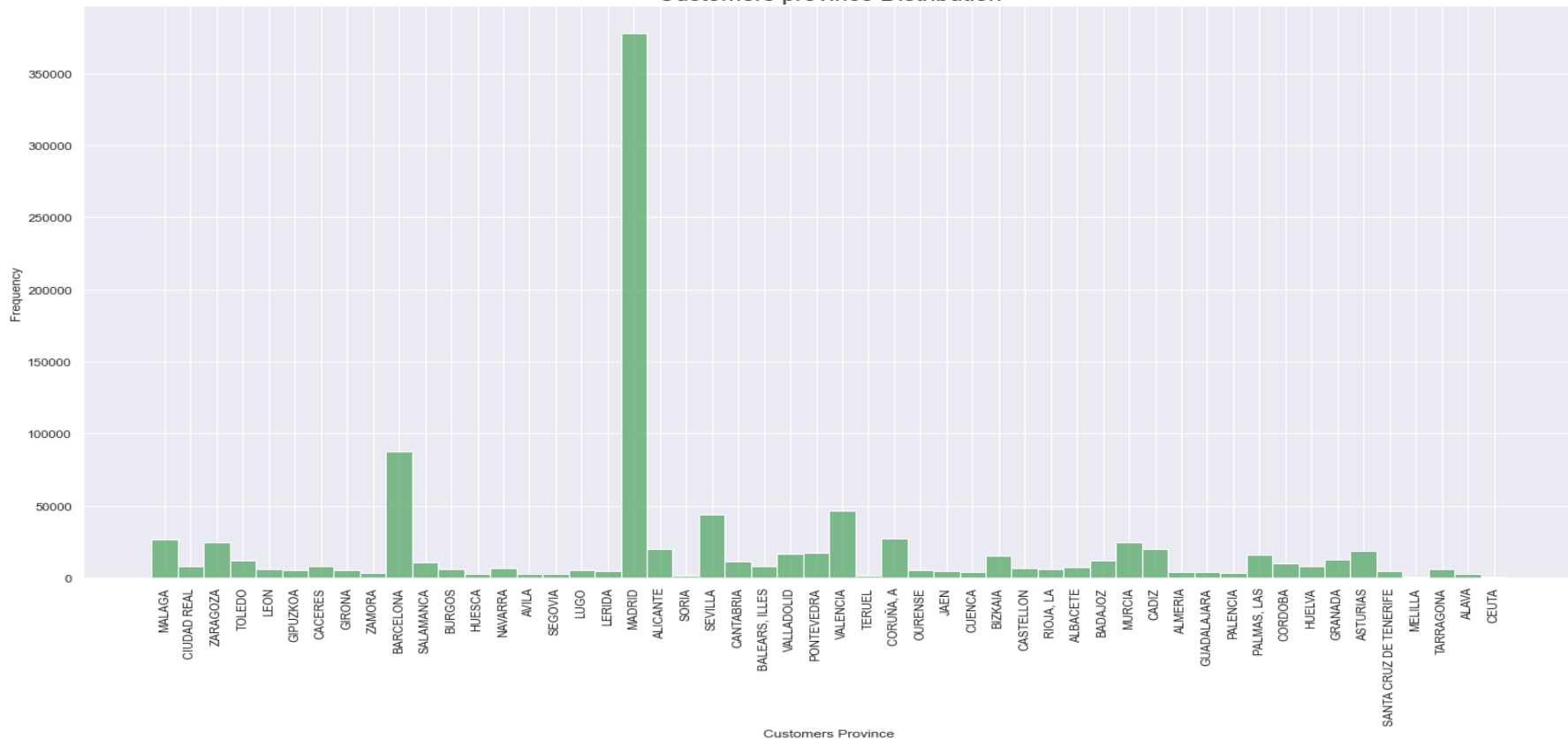


# Share of customers by province

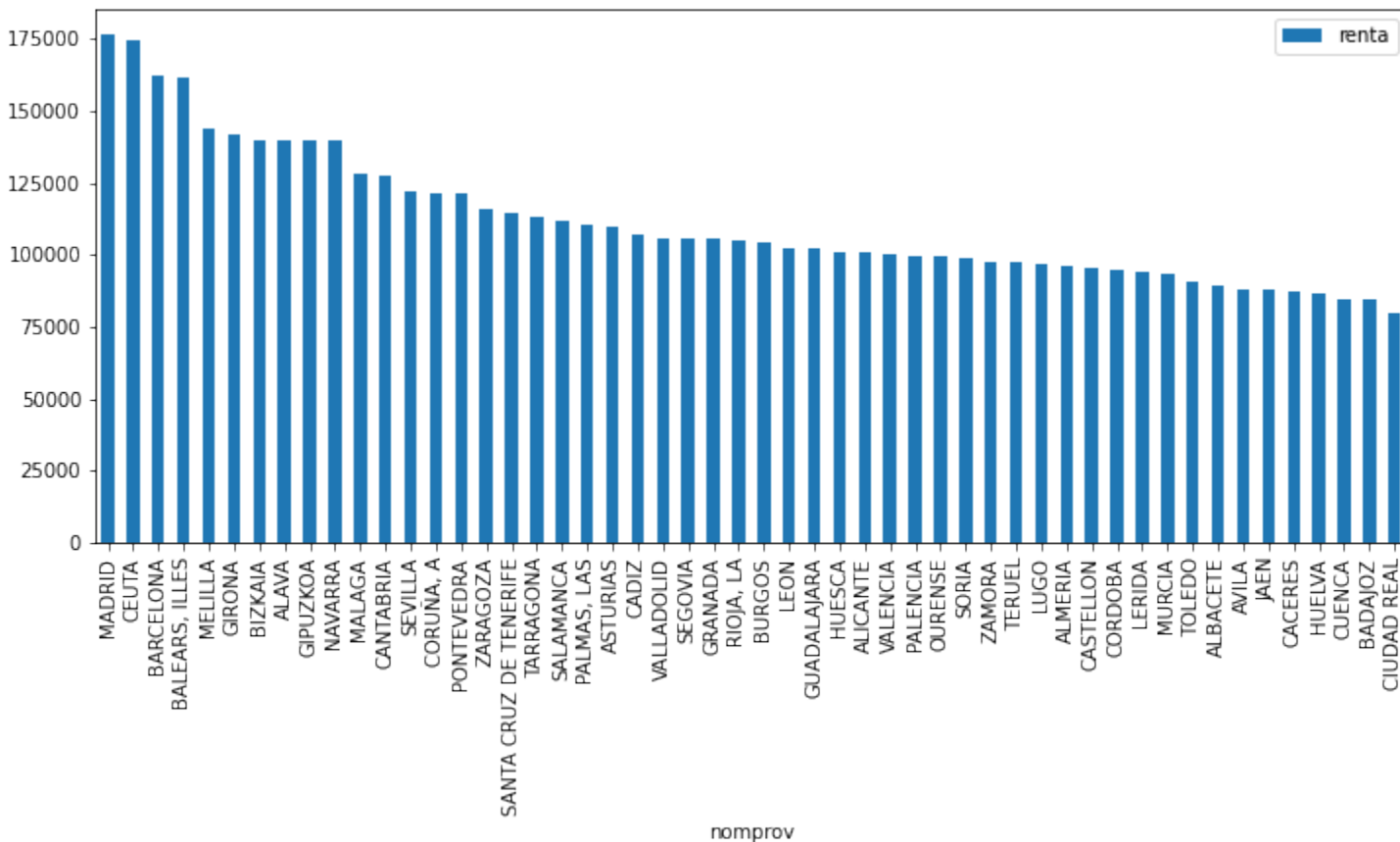


# Share of customers by province

Customers province Distribution

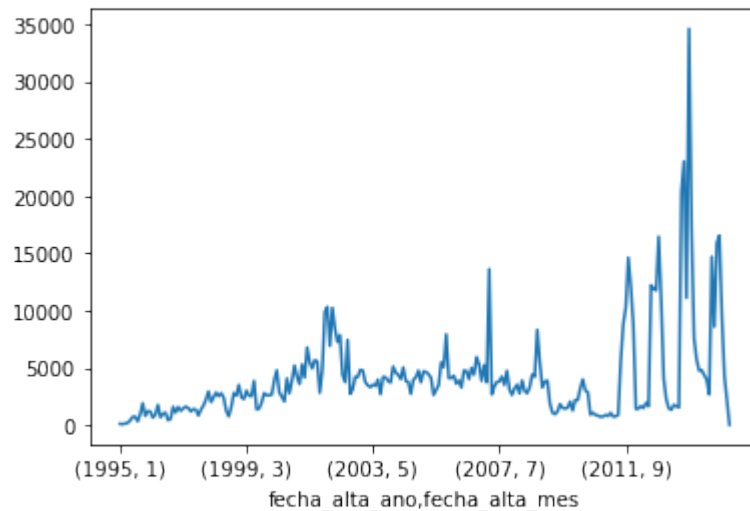
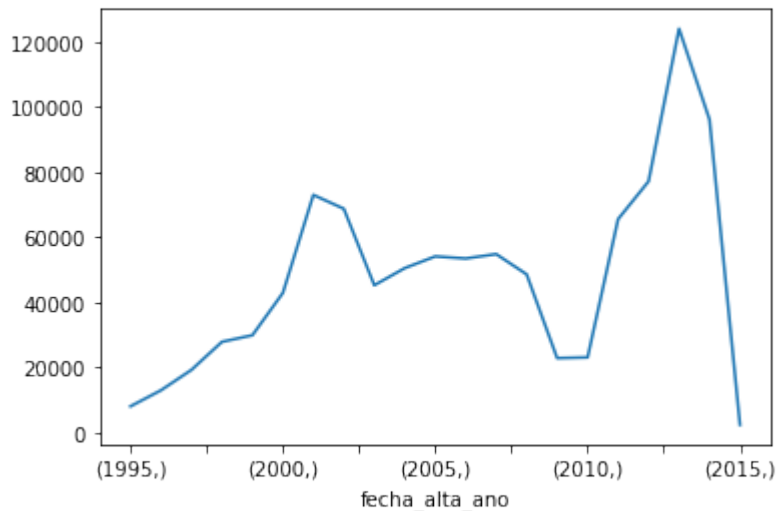


# Average household income by province



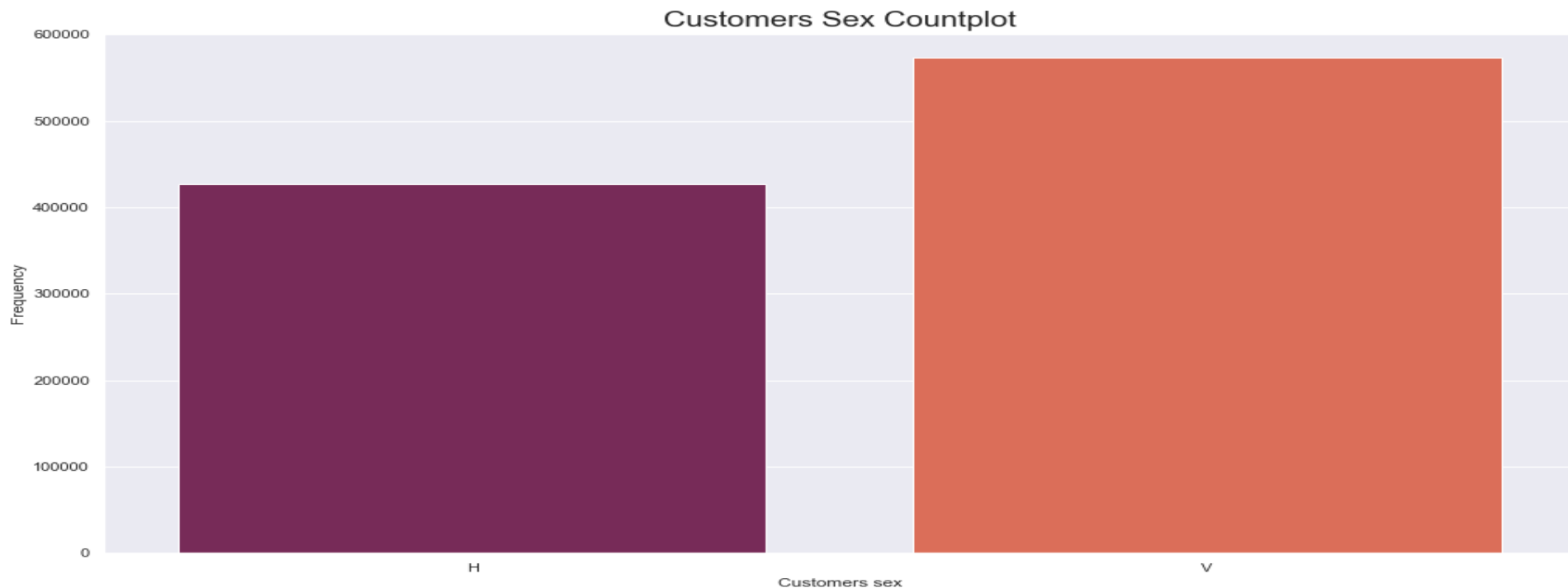
# Monthly and Annual Trends in Customer Registration

Number of customer registrations by year and by month. Note that the period after 2011 has more registrations compared to before.

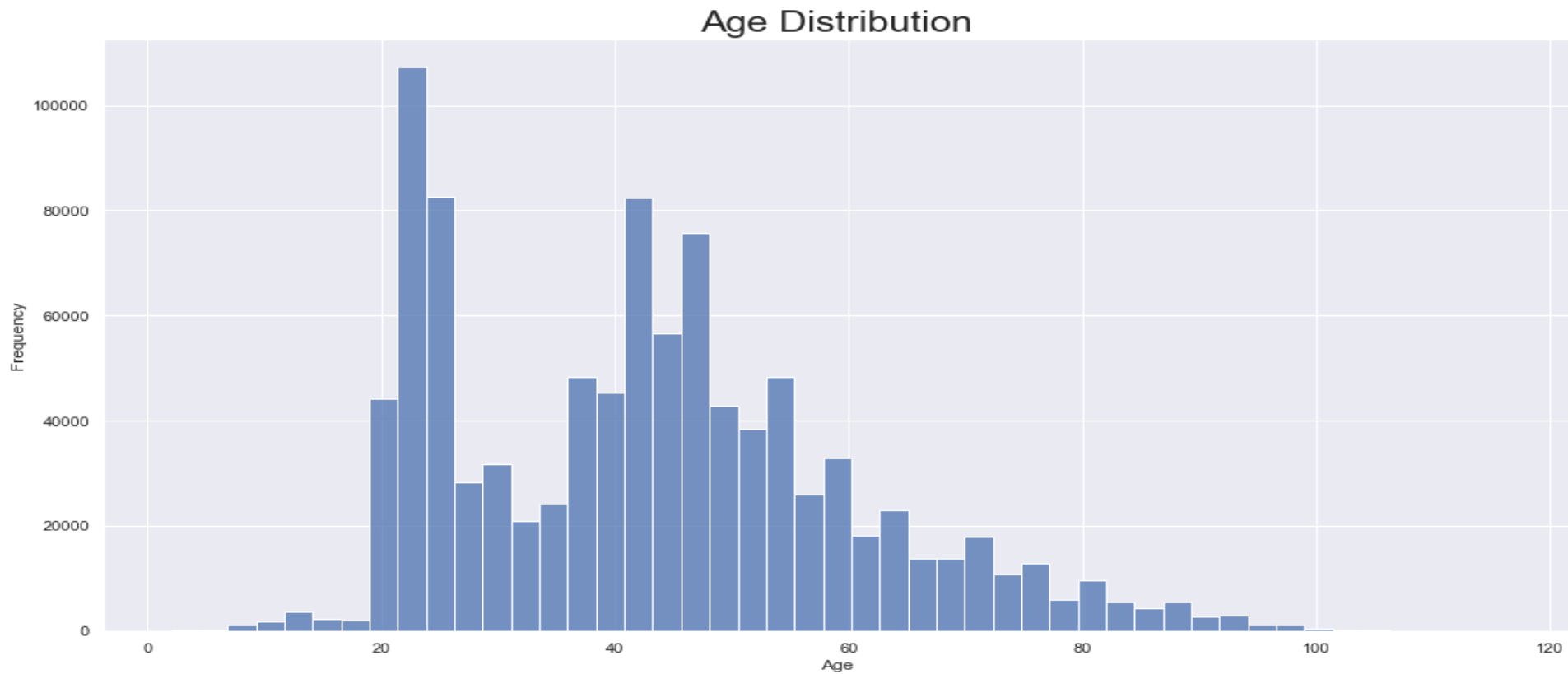


# Customers Sex Category Distribution

Most of the customers belong to the sex category “V”.



# Customers Age Distribution



# Recommendation

- We recommend to create a customer segment specially for the Madrid province residents and separately for residents outside Madrid, considering that they make up the largest group of customers and have a higher household income than other provinces.
- Another segment would be for customers who have registered from 2011, as the customers from those years represent a relatively large portion of customers compared to those from previous years.
- A sex category segment for customers will be appropriate as well. (Not sure if this is very appropriate - is this discrimination? I don't think it's discrimination because we have a separate column that specifies the sex of all the customers, this is necessary because one wouldn't want a Christmas package designed for sex "v" to be going to Sex "h" remember it is a personalized Christmas package, I think the company should have a lot of varieties for future purposes as well.)
- Customer Demographic Segment: a segment for customers and their respective province.