

## Prueba 2

### I Description

Teniendo en cuenta los datos de pedidos en el tiempo:

El objetivo del estudio es conocer mejor a los clientes, poder predecir correctamente la demanda basandonos en la recurrencia y realizar las actuaciones

#### Se pide:

- 1 Clasifica a tu juicio los clientes en funcion de la calidad en 4 grupos.
- 2 Realiza un análisis de cohortes como el descrito en este artículo
- 3 ¿Cómo lo podrías mejorar?
- 4 ¿Qué conclusiones de utilidad para el negocio puedes sacar de estos datos? Cita al menos 6.

**Nota: el CAC son 30€.**

En caso de duda, emplea tus propias hipótesis.

### II Results

#### 1. Data cleaning

1.a) Bad ids , see appendix

1.b) 0 Gastos , see appendix

1.c ) **Age : 1505 has age 2017 out of 4748 wrong !**

1 d) Geography according provided zip codes

top 10, in appendix they are all provided

Madrid 3548

Barcelona 645

Valencia 116

A Coruna 40

Alicante 37

Malaga 33

Seville 25

Asturias(formerly Oviedo) 23

Bizkaia 21

Zaragoza 20

1. e) Gender

Mujer vs Hombre

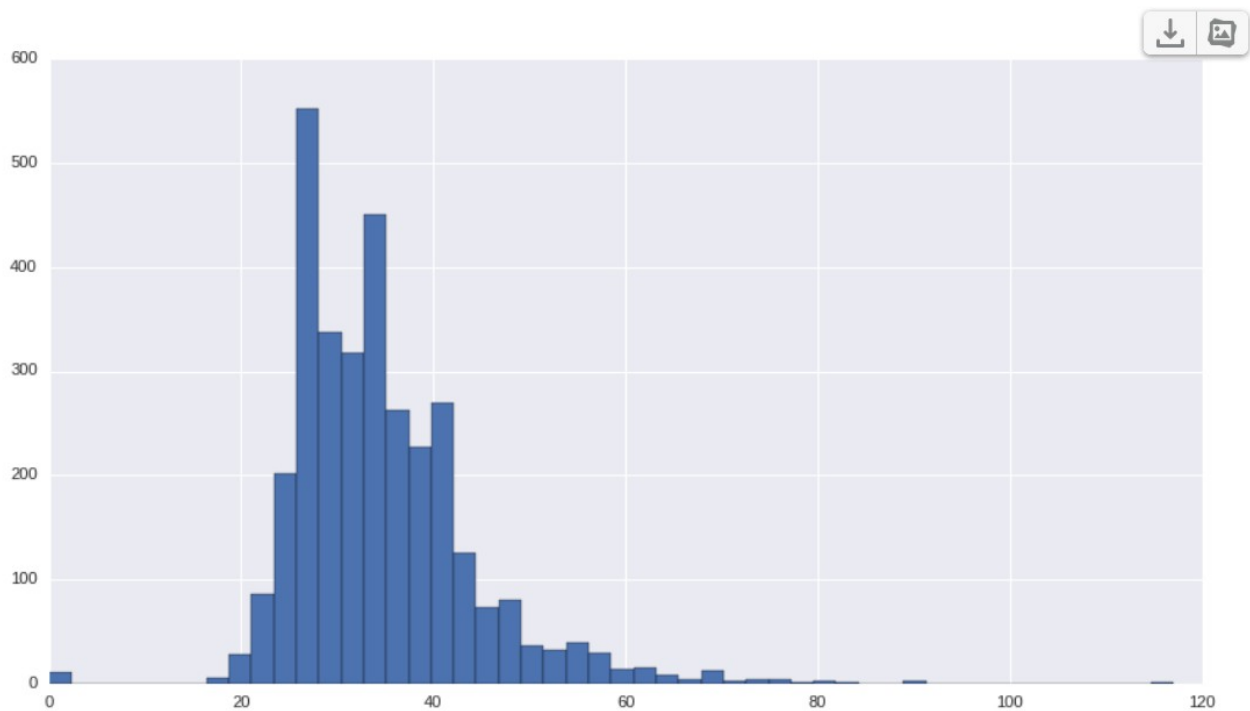
Female : 4457 Male : 291 this is obviously wrong or not ?

If I see first 566 then (array([u'Female', u'Male'], dtype=object), array([277, 290]))

2. Data Understanding , discreptive statistic

2.1 Age without 2017

Age



Age < 18 : 13

id	sexo	zip	age	w1	fo1	w2	fo2	w3	fo3	\
184	185.0	Female	28007.0	17.0	0.00	0.0	29.80	1.0	0.00	0.0
433	434.0	Male	28006.0	1.0	29.14	1.0	28.34	0.0	32.29	0.0
565	566.0	Female	28005.0	0.0	47.29	1.0	53.08	0.0	63.18	0.0
610	611.0	Female	41013.0	1.0	0.00	0.0	74.99	1.0	123.66	0.0
611	612.0	Female	28012.0	1.0	0.00	0.0	26.24	1.0	35.98	0.0
738	739.0	Female	28049.0	0.0	0.00	0.0	0.00	0.0	0.00	0.0
747	748.0	Female	28014.0	17.0	0.00	0.0	0.00	0.0	0.00	0.0
2865	2866.0	Female	28045.0	0.0	0.00	0.0	0.00	0.0	0.00	0.0
3489	3490.0	Female	28039.0	1.0	0.00	0.0	0.00	0.0	0.00	0.0
4236	4237.0	Female	46022.0	0.0	0.00	0.0	0.00	0.0	0.00	0.0
4278	4279.0	Female	28014.0	0.0	0.00	0.0	0.00	0.0	0.00	0.0
4628	4629.0	Female	28005.0	0.0	0.00	0.0	0.00	0.0	0.00	0.0
4685	4686.0	Female	28028.0	0.0	0.00	0.0	0.00	0.0	0.00	0.0

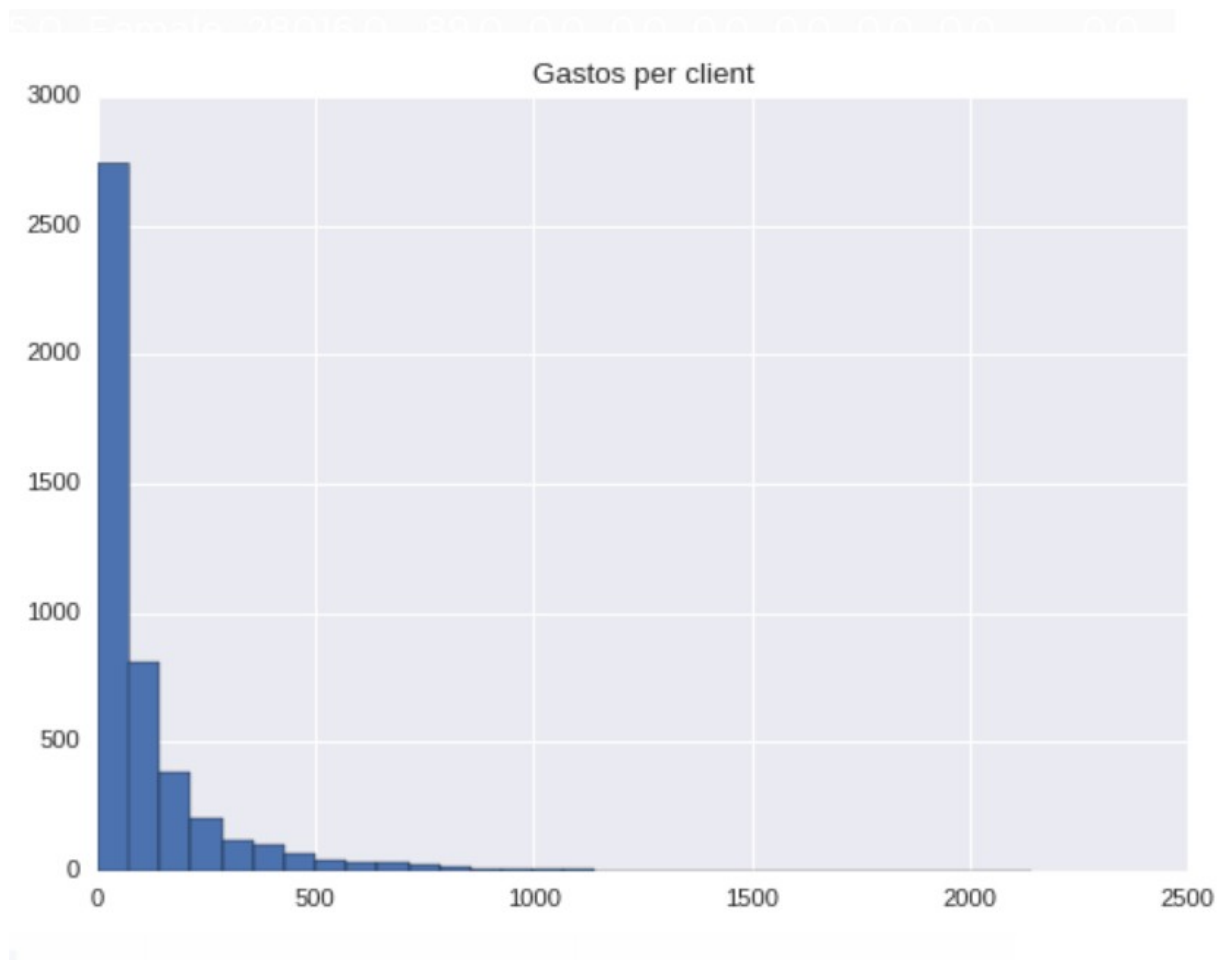
Age > 90 : 1

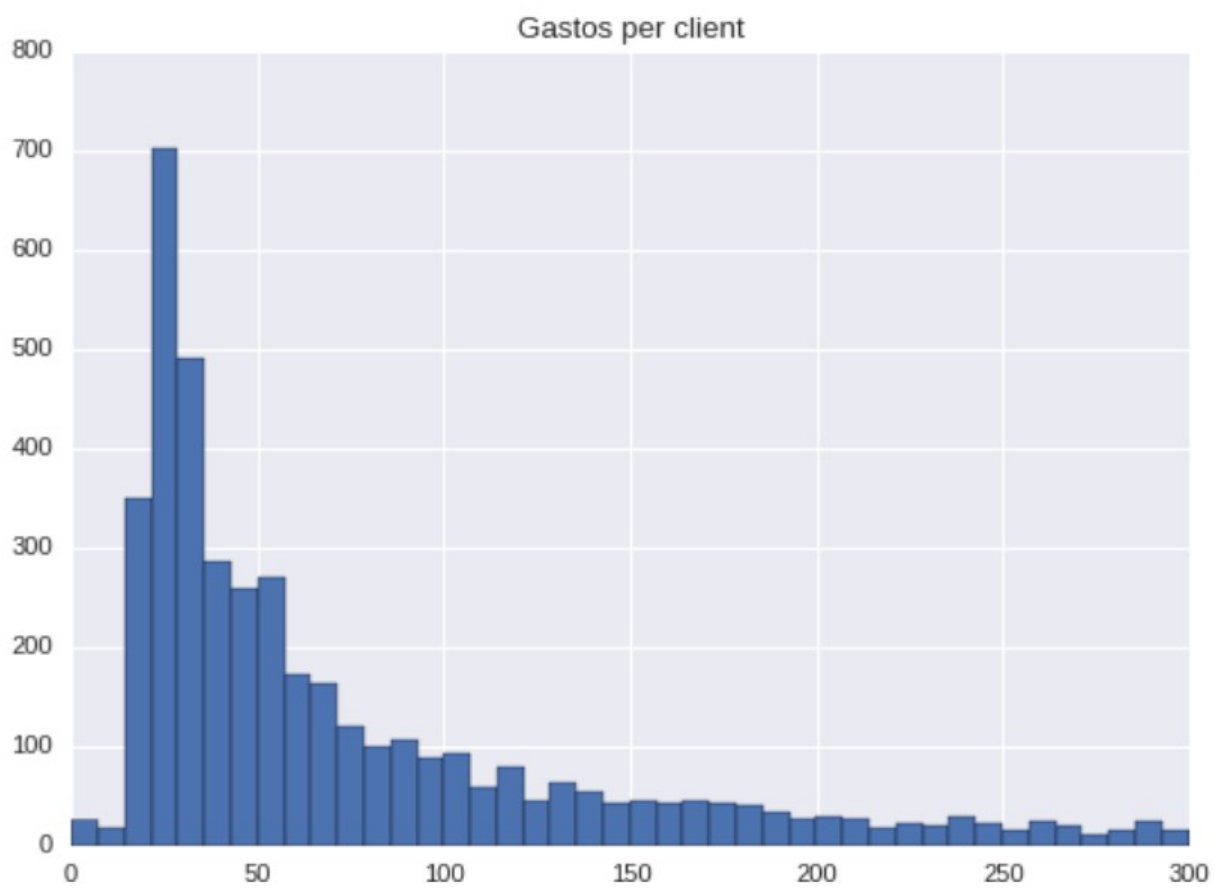
	id	sexo	zip	age	w1	fo1	w2	fo2	w3	fo3	...	w26	\
3016	3017.0	Female	28012.0	117.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0

Age > 80 : 5

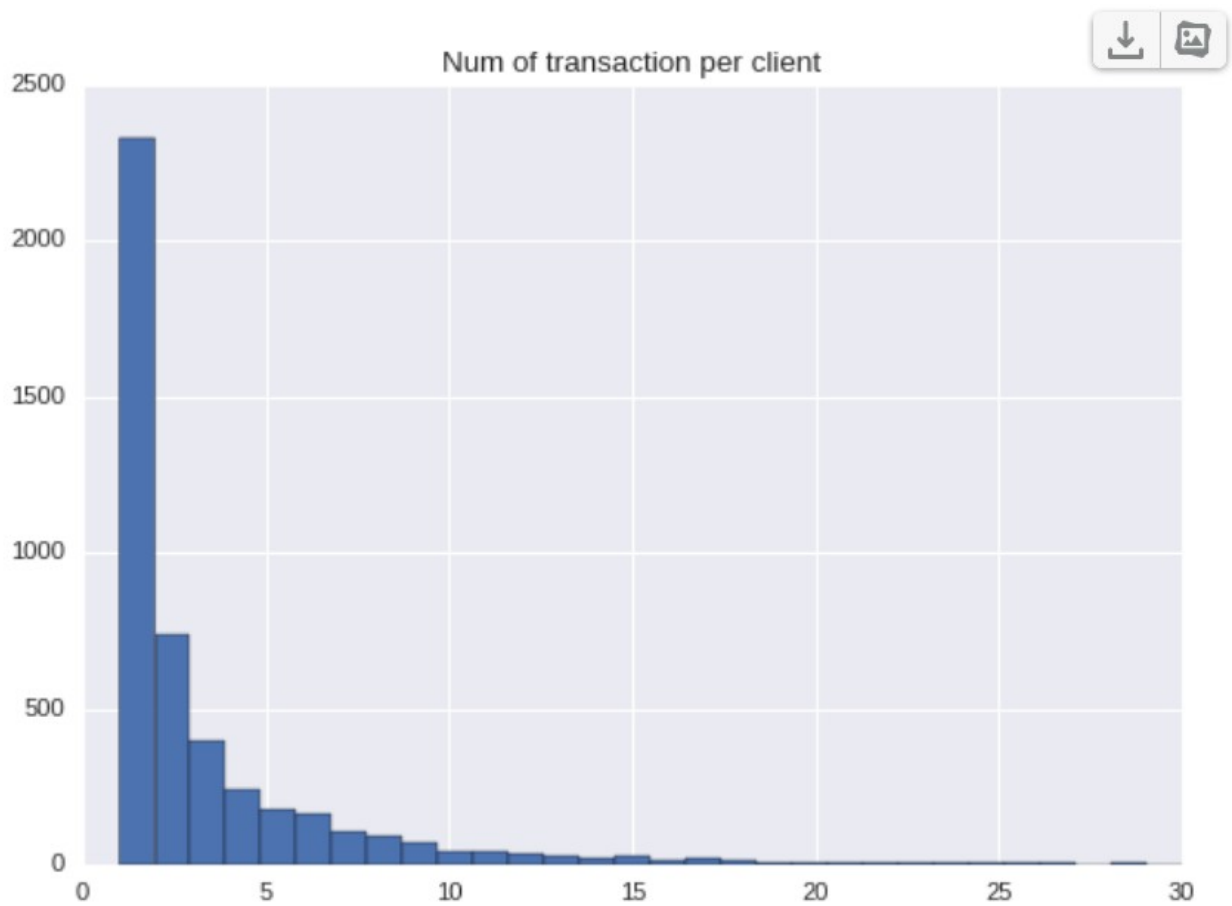
	id	sexo	zip	age	w1	fo1	w2	fo2	w3	fo3	...	w26	\
1187	1188.0	Female	28033.0	90.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
1669	1670.0	Female	8028.0	82.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
3016	3017.0	Female	28012.0	117.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
3241	3242.0	Female	46021.0	90.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
3284	3285.0	Female	28016.0	89.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	

## 2.2 Spendings by client

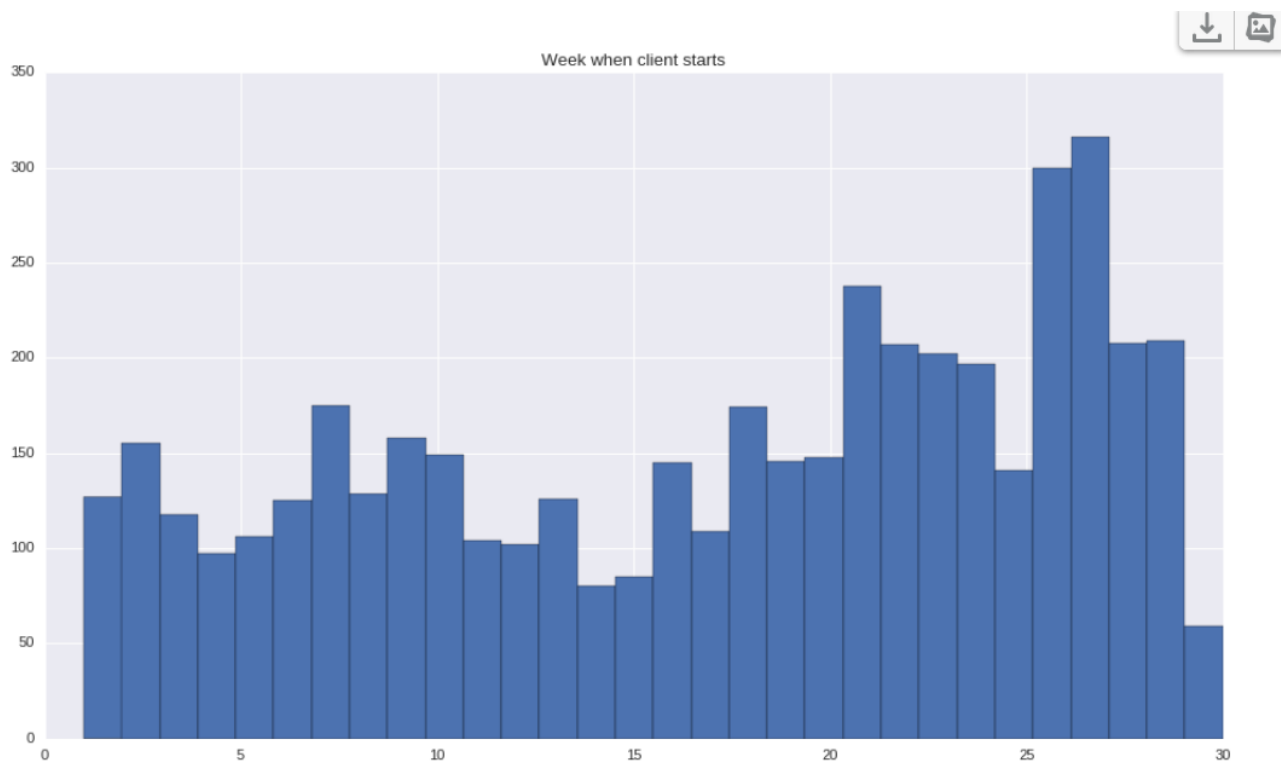




### 2.3 Number of transaction realized by client



2.4 Week when client starts



2.5 Total money per week



2.6 Number of new people per week

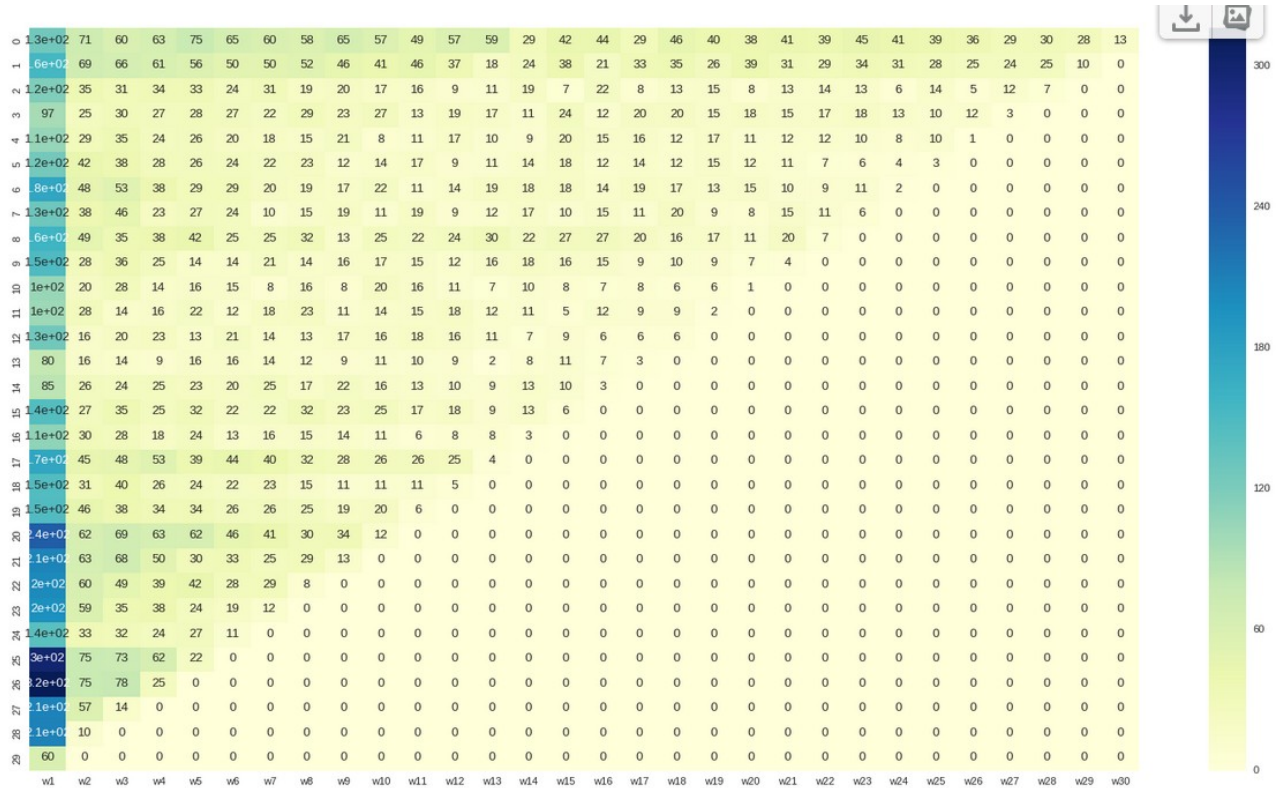


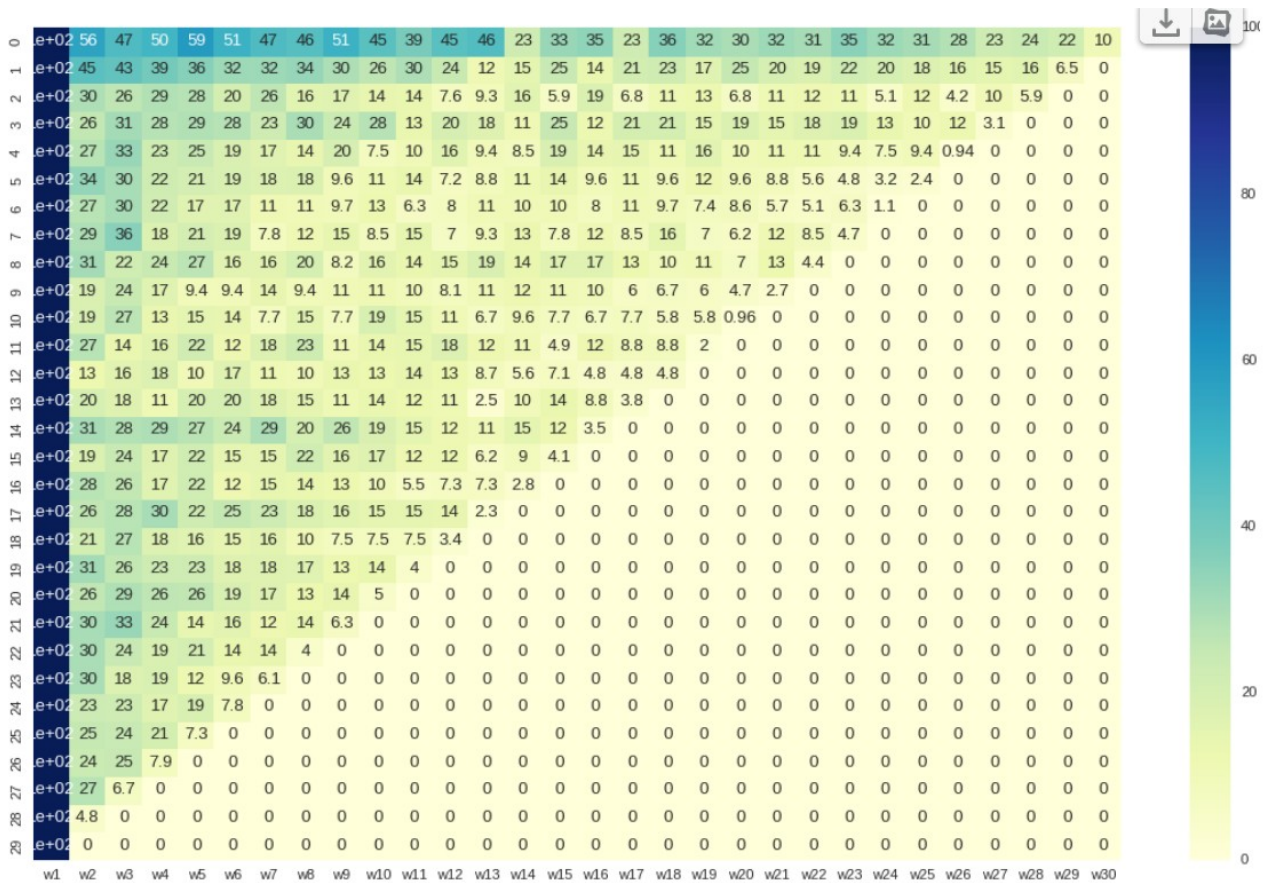
### 3 Cohort Analysis

I cannot open excel file in the provided paper, I m using linux – Ubuntu. It means I had to write my own program for C.A.

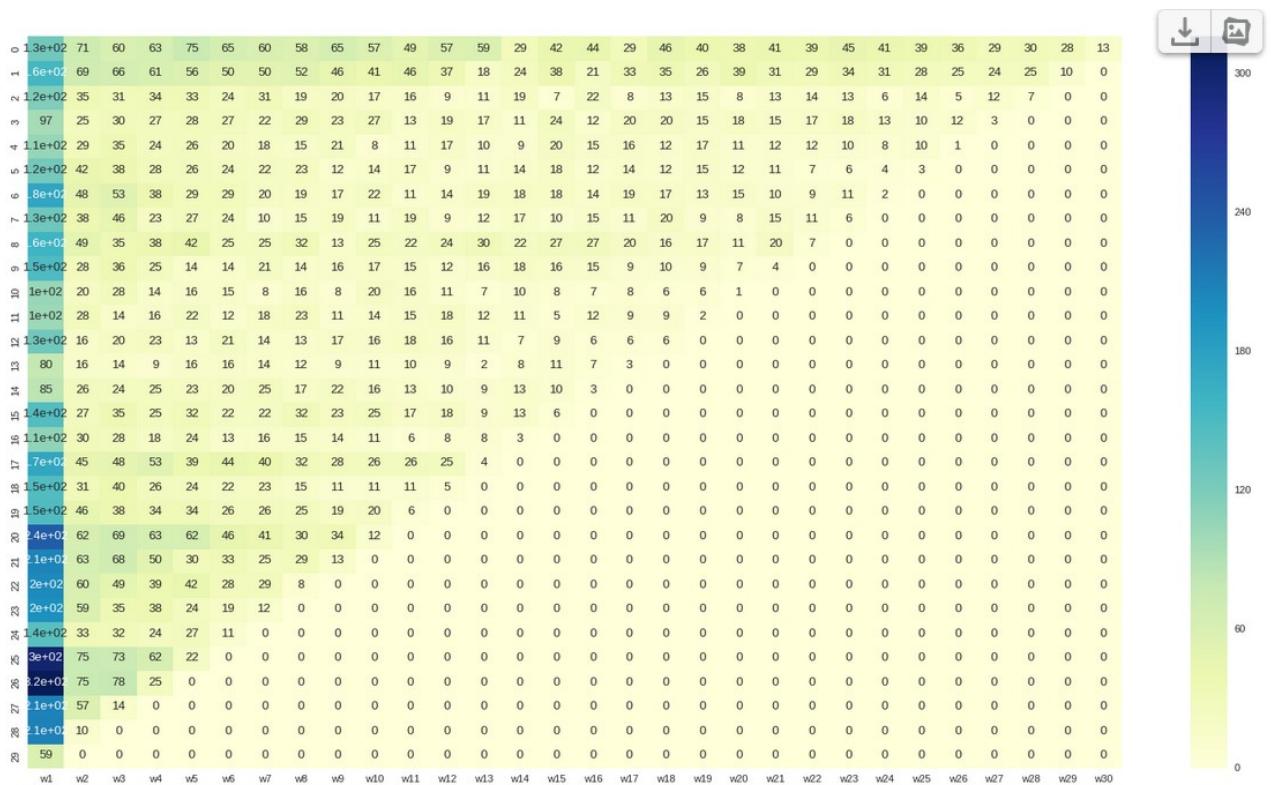
### 3.1 Evolution of clients by week , 30 weeks have been considered

Total number vs %

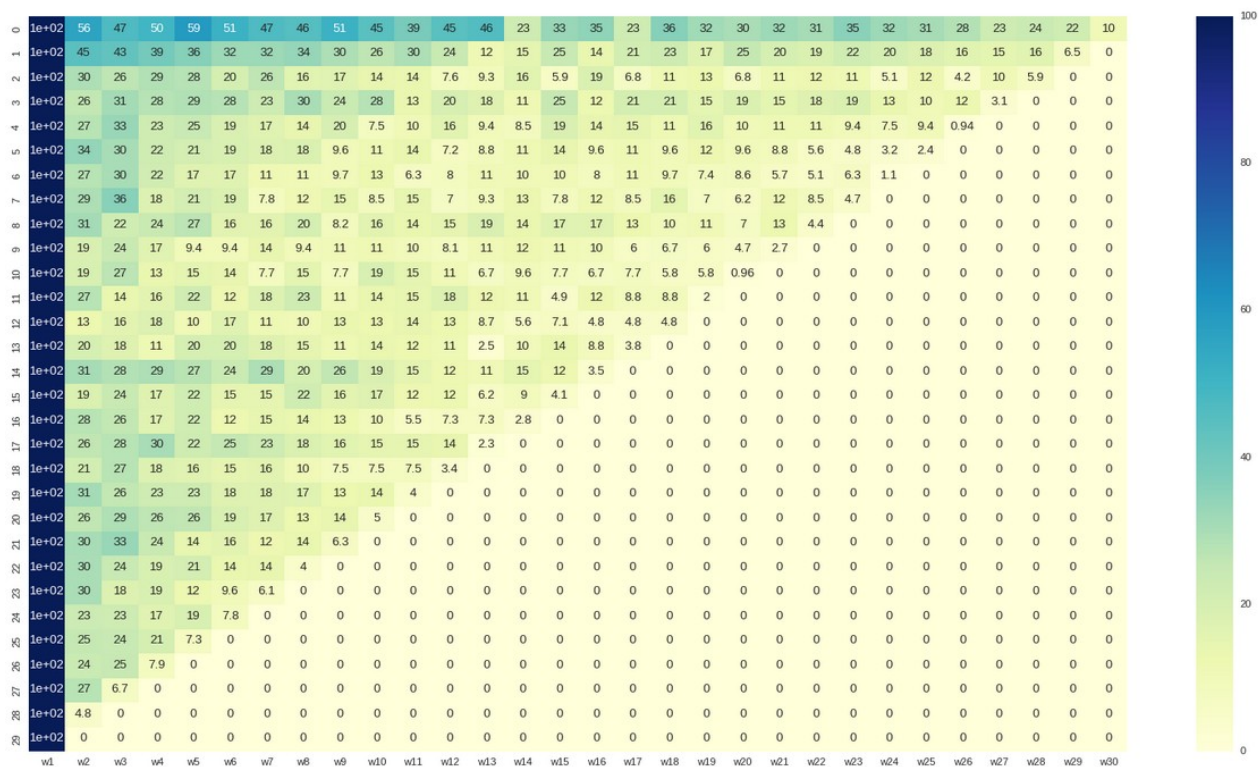




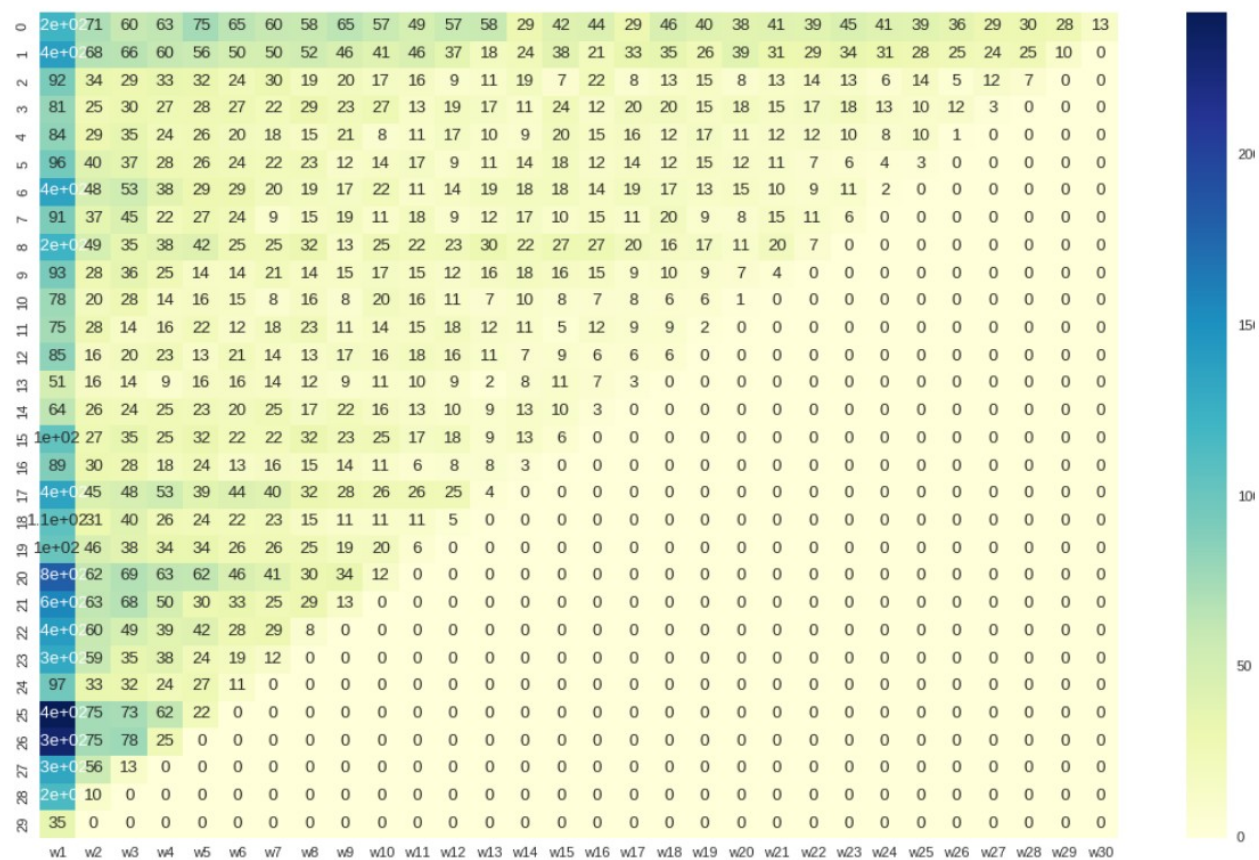
### 3.2 Cleaned Data without mistakes as described in 2.1







### 3.3 Conditinal cohorts : gastos >30 ( provided CAC)















2. Clasifica a tu juicio los clientes en función de la calidad en 4 grupos.

I guess

a) From money point of view

1. group is the people with 0 transactions
2. group with less than CAC transactions in money
3. group with spendings 50 to 100 money in total
4. group with spendings > 100 money in total

as well it's possible to divide quartiles of transaction per client or quartiles of grand total in money

b) from geography point of view

1. Madrid
2. Valencia
3. Barcelona
4. The rest

3 Realiza un análisis de cohortes como el descrito en este artículo  
is realized, see above

4 ¿Cómo lo podrías mejorar?

I can make additional conditional cohort analysis, depending on geography f.e. or when I see the groups like



I can cluster it.

4 ¿Qué conclusiones de utilidad para el negocio puedes sacar de estos datos? Cita al menos 6.

4.1 Something is happening in weeks, marketing influence? Here I have no data. Having the data I could estimate marketing impact. Usually as it is described in literature the cohort analysis is used to check some influences... behaviour influences, marketing influences, promotion influences ... but I do not have them

4.2 Loyalty of clients f.e. I see that ppl with spending over 100 are more loyal than from 50 to 100

4.3 Better segmentation of clients

4.4 Prediction: How many incomes in money company may have or how money clients are going to be involved buying product of company in next weeks

4.5 See evolution of clients by time in different regions

4.6 Use combination of conditions to understand better clients of company.. as age and region and gastos and frequency of transactions.. playing with those parameters may help to segment the clients of company to loyal, frequent, with money better.

Links for better graphs

1 <http://www.evernote.com/l/AKweqbn0eeRONKKn3SKlI6Ql6-ohIJXJkus/>

2 [http://www.evernote.com/l/AKzoVwIO8yJCiKklVYftP\\_a6mwJzsZf5NsA/](http://www.evernote.com/l/AKzoVwIO8yJCiKklVYftP_a6mwJzsZf5NsA/)

## Appendix

1.a Bad ids 28 out of 4748

167 168.0

227 228.0

280 281.0

309 310.0

329 330.0

374 375.0

585 586.0

636 637.0

768 769.0

797 798.0

983 984.0

1001 1002.0

1014 1015.0

1035 1036.0

1050 1051.0

1185 1186.0

1238 1239.0

1569 1570.0

1632 1633.0

1697 1698.0

1770 1771.0

1919 1920.0

2059 2060.0

2090 2091.0

2662 2663.0

2749 2750.0

3235 3236.0

3411 3412.0

1.b

Gastis total = 0 :84 out of 4748

378 379.0 1

439 440.0 2

472 473.0 3  
554 555.0 4  
604 605.0 5  
629 630.0 6  
647 648.0 7  
671 672.0 8  
672 673.0 9  
684 685.0 10  
688 689.0 11  
692 693.0 12  
702 703.0 13  
728 729.0 14  
730 731.0 15  
780 781.0 16  
848 849.0 17  
849 850.0 18  
974 975.0 19  
987 988.0 20  
992 993.0 21  
1182 1183.0 22  
1405 1406.0 23  
1421 1422.0 24  
1457 1458.0 25  
1463 1464.0 26  
1509 1510.0 27  
1629 1630.0 28  
1634 1635.0 29  
1635 1636.0 30  
1652 1653.0 31  
1732 1733.0 32  
1747 1748.0 33  
1758 1759.0 34  
1771 1772.0 35  
1816 1817.0 36  
1844 1845.0 37  
1853 1854.0 38  
1869 1870.0 39  
1885 1886.0 40  
1886 1887.0 41  
1901 1902.0 42  
1945 1946.0 43  
1979 1980.0 44  
1981 1982.0 45  
2073 2074.0 46  
2162 2163.0 47  
2176 2177.0 48  
2178 2179.0 49  
2187 2188.0 50  
2199 2200.0 51  
2237 2238.0 52  
2304 2305.0 53  
2305 2306.0 54



2308 2309.0 55  
2310 2311.0 56  
2312 2313.0 57  
2370 2371.0 58  
2380 2381.0 59  
2481 2482.0 60  
2522 2523.0 61  
2538 2539.0 62  
2599 2600.0 63  
2710 2711.0 64  
2737 2738.0 65  
2936 2937.0 66  
2980 2981.0 67  
3120 3121.0 68  
3158 3159.0 69  
3199 3200.0 70  
3200 3201.0 71  
3206 3207.0 72  
3307 3308.0 73  
3341 3342.0 74  
3342 3343.0 75  
3352 3353.0 76  
3358 3359.0 77  
3477 3478.0 78

3512 3513.0 79  
3534 3535.0 80  
3648 3649.0 81  
3789 3790.0 82  
3798 3799.0 83  
4045 4046.0 84

1.c

Madrid 3548  
Barcelona 645  
Valencia 116  
A Coruna 40  
Alicante 37  
Malaga 33  
Seville 25  
Asturias(formerly Oviedo) 23  
Bizkaia 21  
Zaragoza 20  
Pontevedra 17  
Murcia 17  
Cantabria(formerly Santander) 15  
Valladolid 14  
Gipuzkoa 13  
Cadiz 13  
Araba 11  
Navarre 11

Toledo 11  
Tarragona 10  
Granada 9  
Lleida 8  
Leon 8  
Guadalajara 8  
La Rioja(formerly Logrono) 8  
Lugo 7  
Badajoz 5  
Girona 5  
Burgos 5  
Caceres 5  
Albacete 4  
Huelva 4  
Ciudad Real 4  
Cordoba 4  
Huesca 4  
Castellon 3  
Salamanca 3  
Palencia 2  
Segovia 2  
Ourense 2  
Zamora 1  
Teruel 1  
Balearic Islands 1  
Avila 1  
Jaen 1  
Almeria 1  
Cuenca 1