

April 30<sup>th</sup> , 2020

# ELECTRIC POWER CONSUMPTION

Clustering and Analyzing the Consumes of our Clients

Predictive Analytics Course

Group 4 - Guida Giulia, Muttoni Davide, Ticozzi Paolo

# Introduction



Dataset

Measurements of electric power consumption in the last month



Region



Age



Calls



Payments



Kitchen (EPC1)



Laundry Room (EPC2)



Air System (EPC3)



Usage % energy consumption  
from 8:00 to 20:00

# Data Preparation

## Dataset Inspection

Columns inspection

- **Region** → households in the northern region tend to consume more in the nights; households in the north count of 45% of the whole dataset
- **Calls** (distribution of n° of calls) → 3 recent calls per customer are more concentrated in night usage and little consumes
- **Payment** → households in the north are the ones with more payment issues



## Preprocessing

- Converting Region in dummy variables
- StandardScaler on Consumes and Age

## Feature Engineering

- Aggregation of EPC1, EPC2, EPC3 consumes into a new column (Consumes)
- Focus on consumes instead of where the energy was consumed
- PCA: no better results

# Methods applied

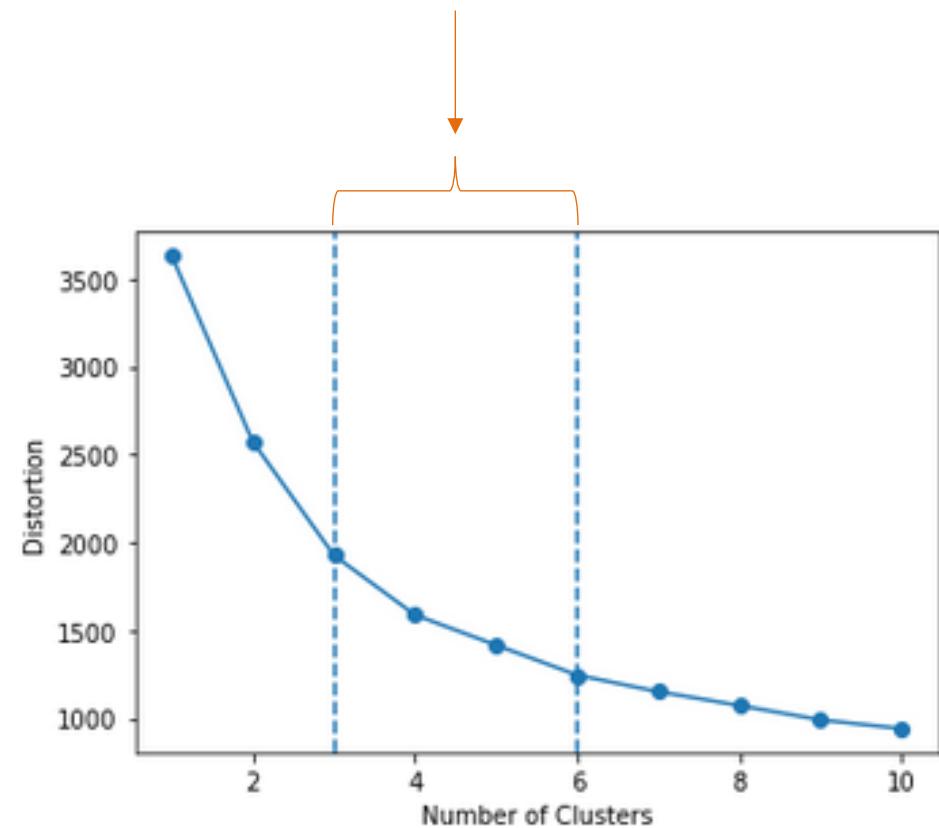
## K - Means

```
silh score: 2 0.46
silh score: 3 0.273
silh score: 4 0.28
silh score: 5 0.269
silh score: 6 0.314
silh score: 7 0.307
silh score: 8 0.304
silh score: 9 0.292
silh score: 10 0.265
```

**Silhouette** - the best option is 6 clusters

**Distortion** - the best option is in between 3

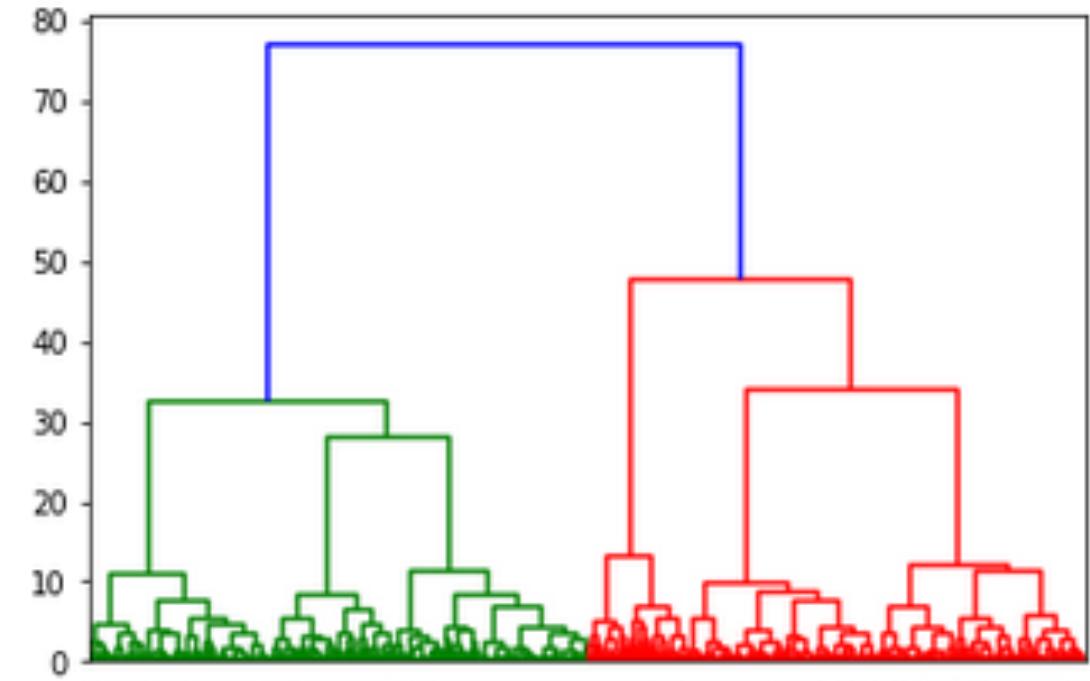
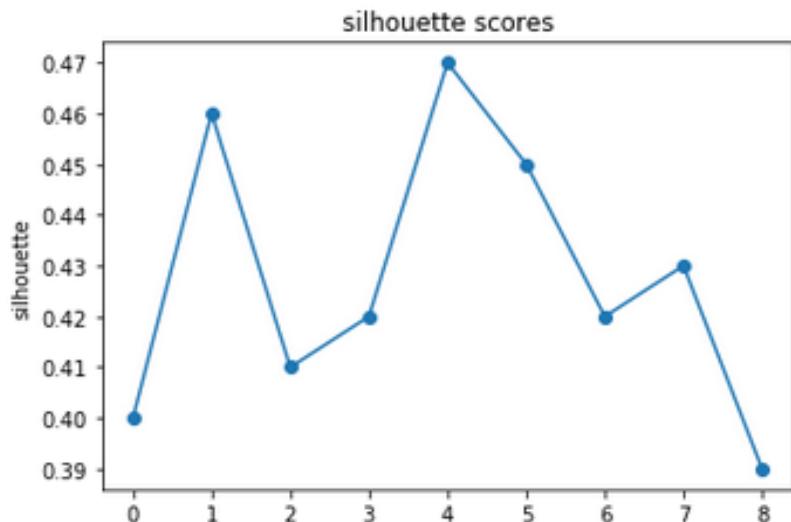
and 6 clusters



# Best technique

## Hierarchical Method: Ward distance

```
silh for 2 clusters: 0.401  
silh for 3 clusters: 0.459  
silh for 4 clusters: 0.414  
silh for 5 clusters: 0.417  
silh for 6 clusters: 0.467  
silh for 7 clusters: 0.449  
silh for 8 clusters: 0.417  
silh for 9 clusters: 0.434  
silh for 10 clusters: 0.393
```



6 clusters, hierarchical method with Ward distance



# Clusters

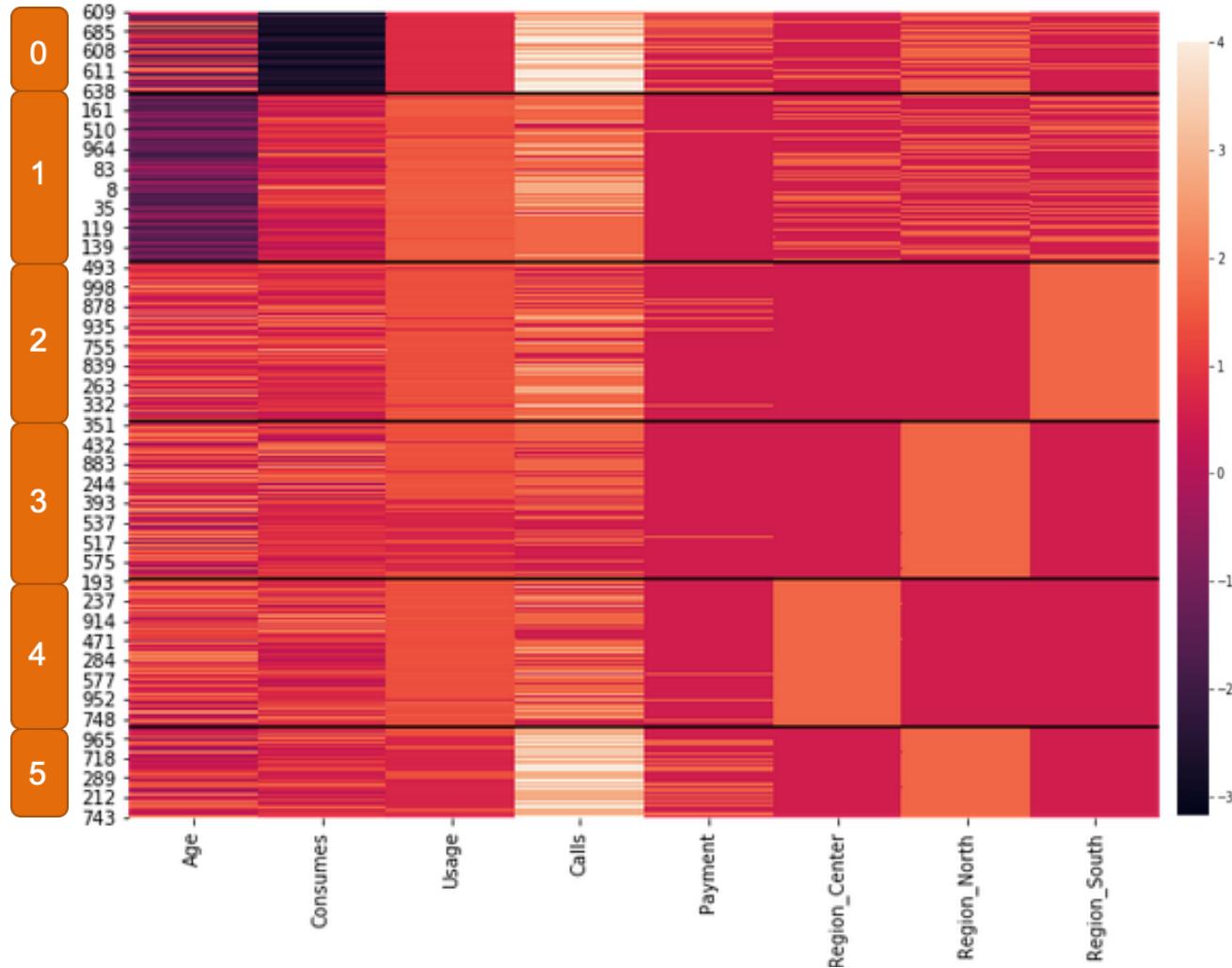
Clusters' dimension

Clusters' information

Clusters	
0	100
1	207
2	194
3	192
4	182
5	125

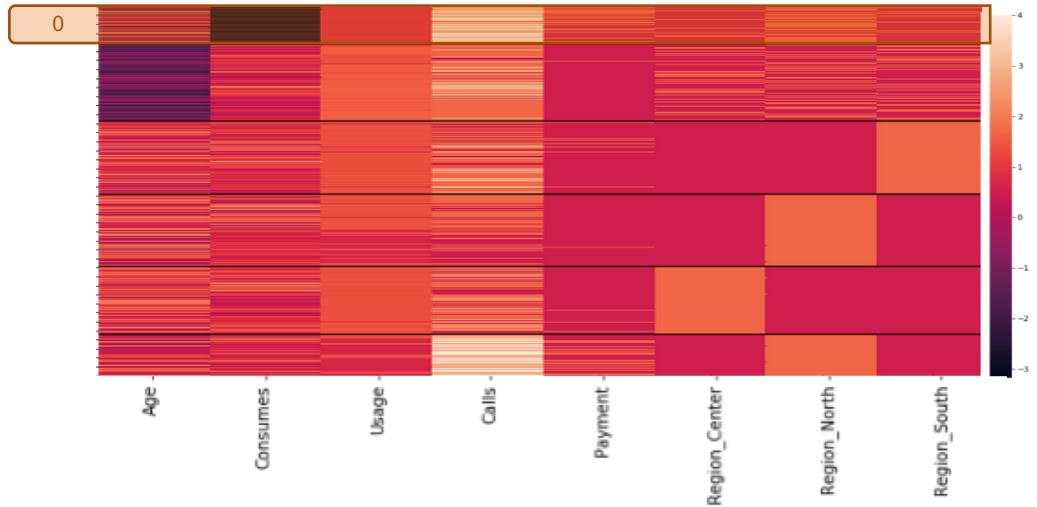
Clusters	Age*	Consumes*	Usage	Calls	Payment	Region_Center	Region_North	Region_South
0	-0.29	-2.71	0.25	2.53	0.51	0.17	0.72	0.11
1	-1.40	0.14	0.86	1.21	0.00	0.32	0.30	0.37
2	0.51	0.33	0.74	0.94	0.05	0.00	0.00	1.00
3	0.47	0.34	0.56	0.54	0.02	0.00	1.00	0.00
4	0.59	0.39	0.74	0.85	0.03	1.00	0.00	0.00
5	0.16	0.34	0.36	2.33	0.31	0.02	0.98	0.00

\* Consumes and Age are standardized

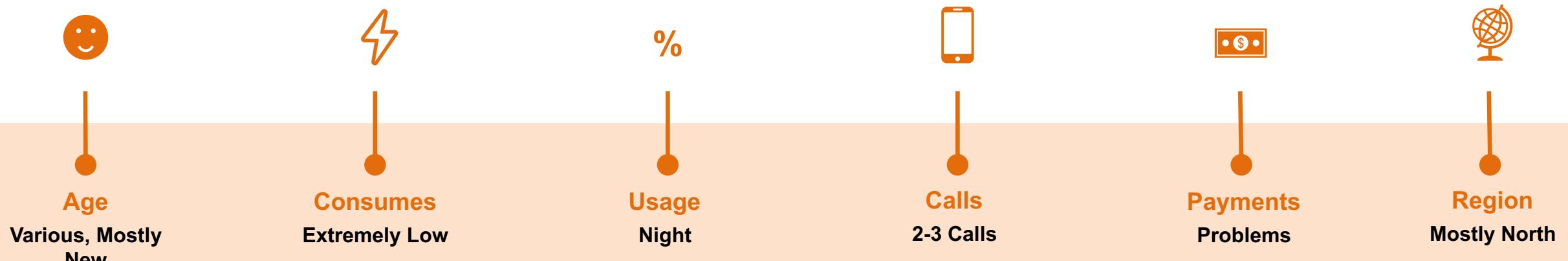


# Cluster 0

## “Trouble-Makers”



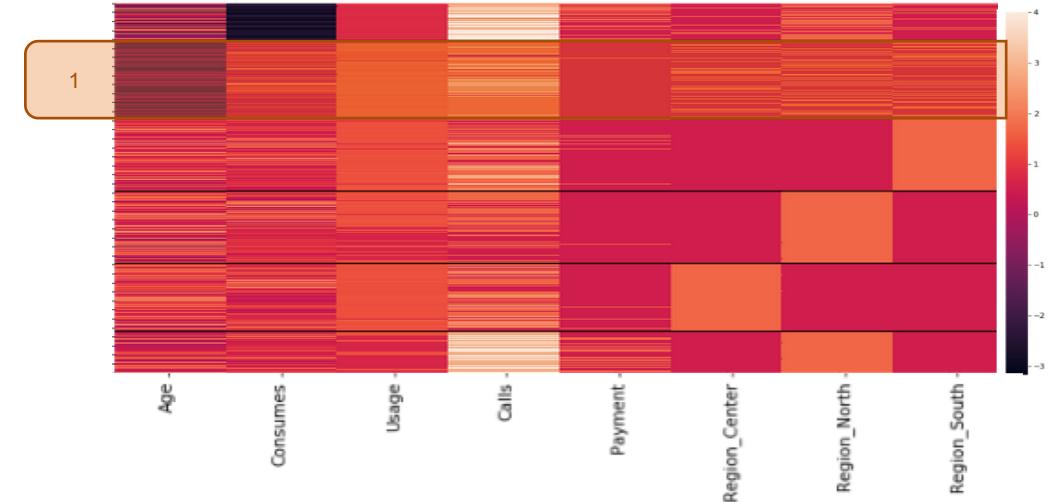
- Most 'Trouble-Makers' are based on the North, and we suppose they are **low-waged workers** that are **working during the day** and have **economic difficulties**.
- **PROPOSAL:** With serious cases, it will be necessary some warnings and drastic measures



Clusters	Age	Consumes	Usage	Calls	Payment	Region_Center	Region_North	Region_South
0	-0.29	-2.71	0.25	2.53	0.51	0.17	0.72	0.11

# Cluster 1

## “Newer Clients”



- Newer, delicate contracts. To keep into consideration because the risk of losing the client is higher.
- **PROPOSAL:** paying more attention about this cluster, offering a good customer service and maybe some bonus/special offers for this cluster



Age

New



Consumes

Well distributed



Usage

Mostly Day



Calls

More than standard



Payments

OK



Region

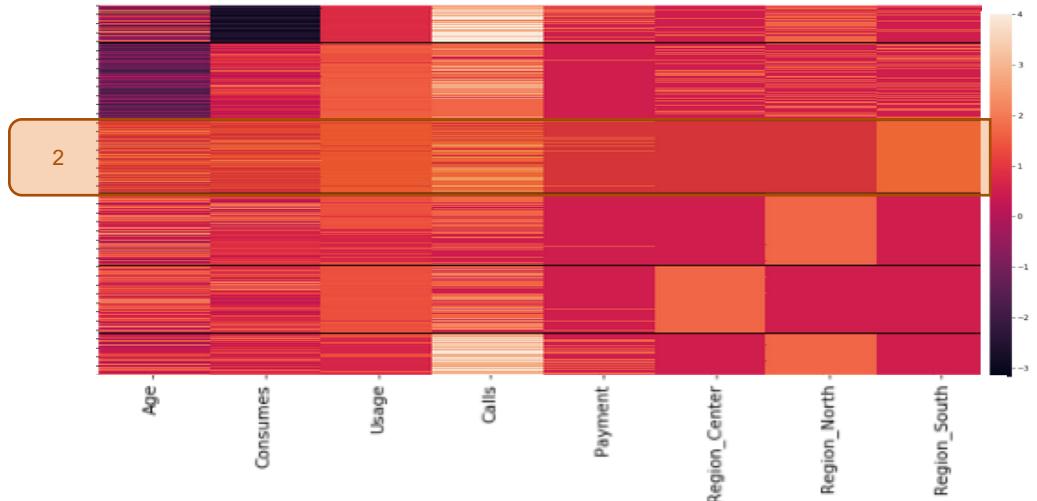
Variable

Clusters	Age	Consumes	Usage	Calls	Payment	Region_Center	Region_North	Region_South
1	-1.40	0.14	0.86	1.21	0.00	0.32	0.30	0.37

Well distributed

# Cluster 2

## “The South”



- Geographic-centered cluster in the South.
- It represents the **standard customer in this region**.



Age

Medium-High

Medium-long term clients



Consumes

Standard



Usage

Mostly in the Day



Calls

Medium-low



Payments

OK



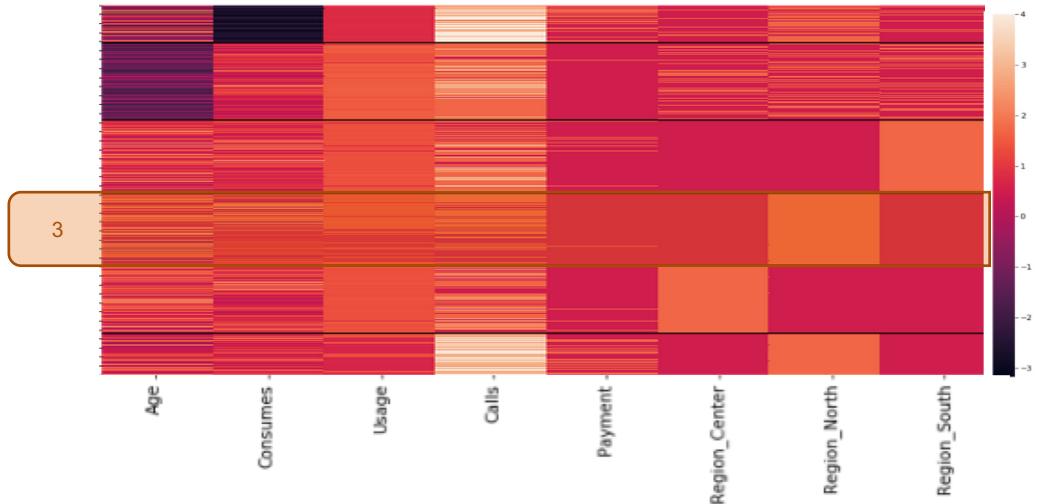
Region

South

Clusters	Age	Consumes	Usage	Calls	Payment	Region_Center	Region_North	Region_South
2	0.51	0.33	0.74	0.94	0.05	0.00	0.00	1.00

# Cluster 3

## “The North”



- Geographic-centered cluster in the North.
- Very similar to the previous “geographical” cluster, but this cluster uses the services more during the night.
- **PROPOSAL:** a “Night Plan”, which lowers the cost for electricity usage during the night (which can have interesting synergies with Cluster 0) promoted mostly in the Northern zone.



Age

Medium-High

Medium-long term clients



Consumes

Standard



Usage

Well-balanced



Calls

Low



Payments

OK



Region

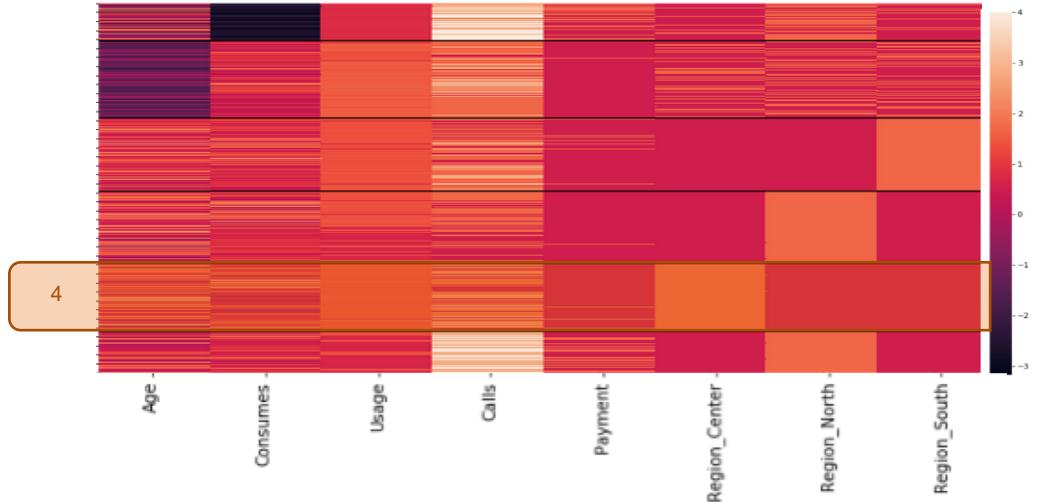
North

Clusters	Age	Consumes	Usage	Calls	Payment	Region_Center	Region_North	Region_South
3	0.47	0.34	0.56	0.54	0.02	0.00	1.00	0.00

# Cluster 4

## “The Center”

- The third geographic-centered cluster in the Central region.
- It represents the standard customer in this region.



Age

Medium-High



Consumes

Standard



Usage

Mostly in the Day



Calls

Medium-low



Payments

OK



Region

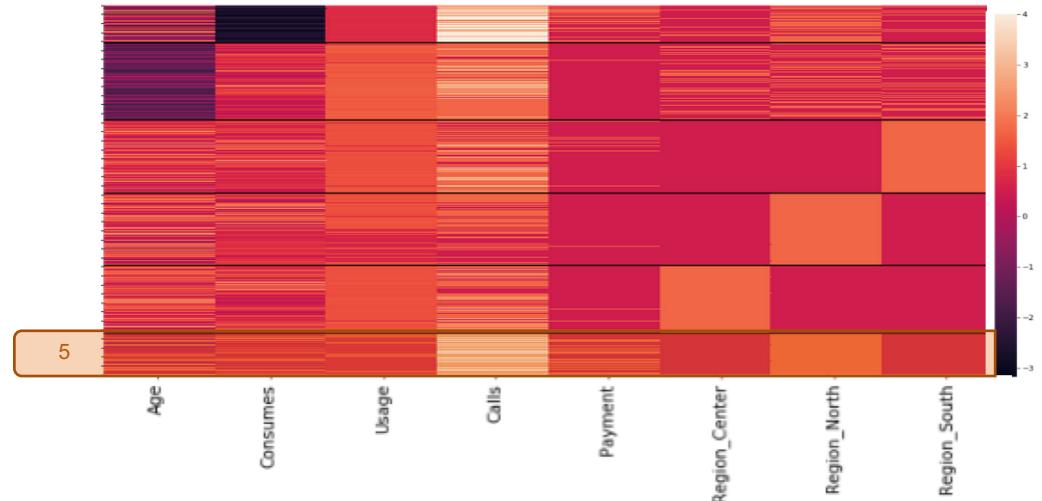
Center

Medium-long term clients

Clusters	Age	Consumes	Usage	Calls	Payment	Region_Center	Region_North	Region_South
4	0.59	0.39	0.74	0.85	0.03	1.00	0.00	0.00

# Cluster 5

## “High Risk”



- High number of calls is a warning signal: these clients have problems with the service and/or payments.
- This final cluster represents the contracts we are risking to lose, focusing in the northern region
- **PROPOSAL:** ① a low-cost plan, ② increase the quality of the call center ③ adopt a problem-solving strategy to understand the reason why this cluster exists.



Clusters	Age	Consumes	Usage	Calls	Payment	Region_Center	Region_North	Region_South
5	0.16	0.34	0.36	2.33	0.31	0.02	0.98	0.00

# Conclusion



**North, Center and South**

No action needed

Geeographical and  
Behavioural

**Boost Consumption**

- 1. Trouble-Makers -**  
Severe Measures
- 2. High-Risk - Facilitate**  
with low-cost plan



April 30<sup>rd</sup> , 2020

THANK YOU!

Predictive Analytics Course

Group 4 - Guida Giulia, Muttoni Davide, Ticozzi Paolo