Master's Thesis

DATA ANALYSIS: CLUSTERING OF ELECTRICITY CONSUMPTION PROFILES

July 2015

Student: Josep Otal
Academic Supervisor: Salvatore Manco'
Company supervisor: Pep Salas

OUTLINE

1. OVERALL INTRODUCTION
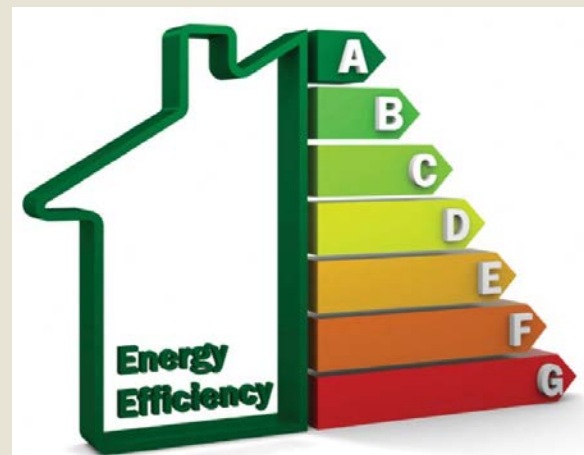

SMART METER DATA

CONSUMER ENGAGEMENT

CONSUMER SEGMENTATION

ENERGY EFFICIENCY

2. ENERBYTE

ELECTRIC UTILITIES
10.000 citizens

Small Energy Consumers

Residential

CITIES
1.500 citizens

ELECTRIC DATA COLLECTION

Utility Smart meter

Sub-metering device

ENERGY EFFICIENCY

Monitoring

GPS effect

Personalization   Multi-platform

Understandable, value added information   Gamification   Community

BUSINESS APPROACH

BtoBtoC

ENERBYTE (BUSINESS) → ELECTRIC UTILITY (BUSINESS) → END-CONSUMER (CUSTOMER)

ENERBYTE (BUSINESS) → City Council LOCAL OR REGIONAL GOVERMENT (BUSINESS) → CITIZEN (CUSTOMER)
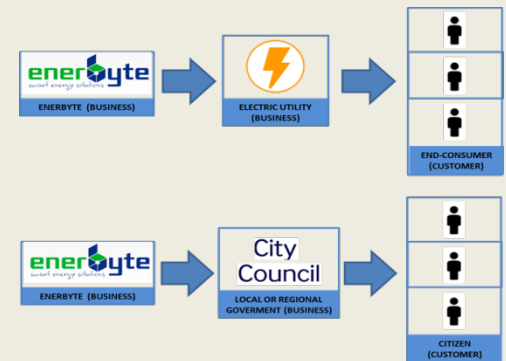
CASE OF STUDY:

ELECTRICITY LOAD PROFILES CLUSTERING

3.1. INTRODUCTION

RUBÍ BRILLA PROJECT

- PUBLIC, PRIVATE, PEOPLE

- RESIDENTIAL (SERVICE TO CITIZEN)

- SOFTWARE AND HARDWARE

CONSUMER BEHAVIOUR

- ENERGY EFFICIENY RECOMMENDATIONS

- DEMAND RESPONSE

- SAVINGS 10%

DATA ANALYTICS "R"

- LARGE AMOUNT OF DATA GENERATED

- DATA MINING AND ANALYTICS

- PROGRAMMING LANGUAGES

3.2. OBJECTIVES

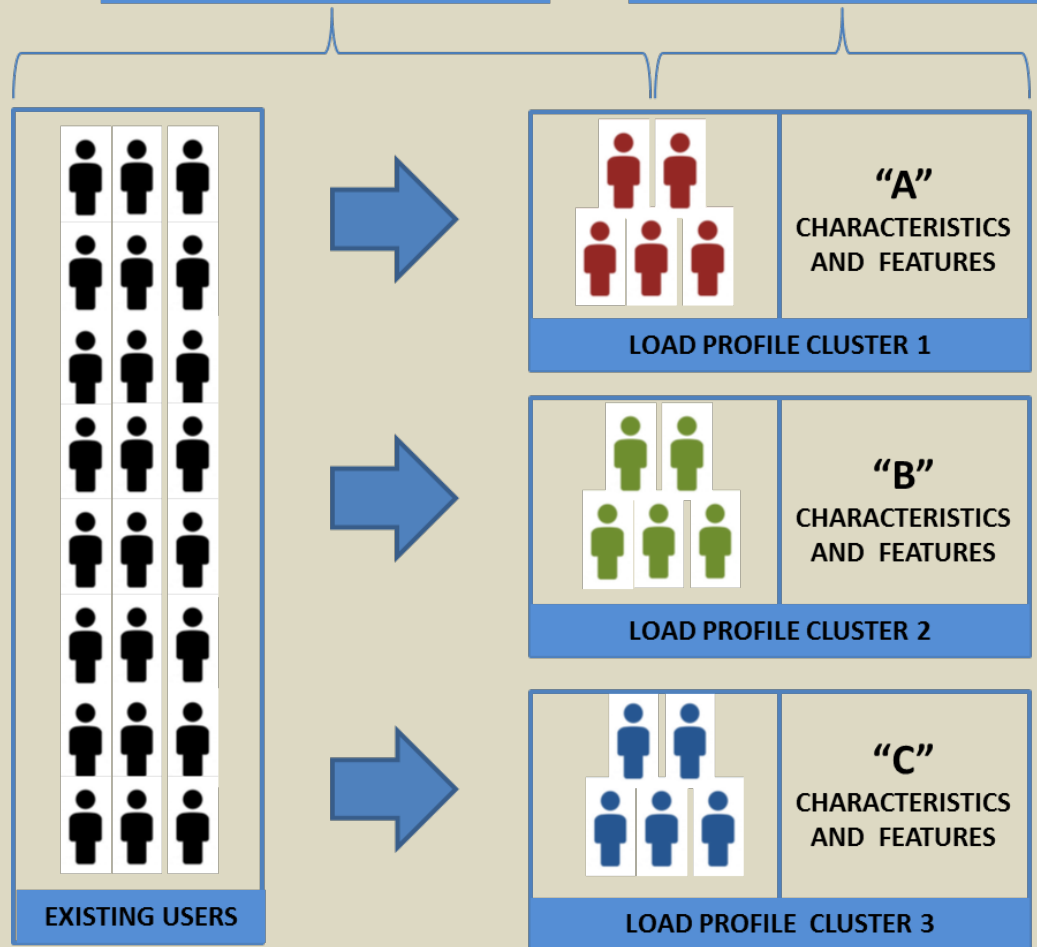DATA EXPLORATION AND VISUALIZATION

CONSUMERS SEGMENTATION BY LOAD PROFILE

PERSONALISED ENERGY EFFICIENCY RECOMMENDATIONS

HOUSEHOLD AND HOUSEHOLDERS CHARACTERISTICS

PHASE 1: LOAD PATTERN CLUSTER CREATION FROM EXISTING USERS

PHASE 2: FIND MORE LIKELY PROPERTIES FOR EACH CLUSTER

EXISTING USERS

"A" CHARACTERISTICS AND FEATURES

LOAD PROFILE CLUSTER 1

"B" CHARACTERISTICS AND FEATURES

LOAD PROFILE CLUSTER 2

"C" CHARACTERISTICS AND FEATURES

LOAD PROFILE CLUSTER 3

3.3. METHODOLOGY

I

- ➢ DATA COLLECTION
- ➢ DATA CLEANING
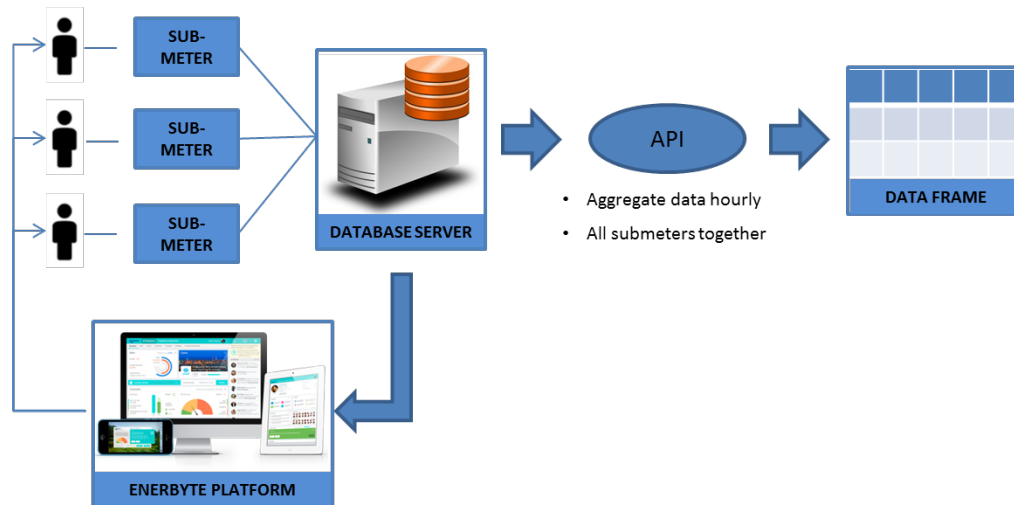- ➢ DATA EXPLORATION AND VISUALIZATION

II

- ➢ PRE-CLUSTERING
- ➢ CLUSTERING: HIERARCHICAL, K-MEANS AND SOM
- ➢ POST-CLUSTERING

III

- ➢ ANALYSE HOUSEHOLDS AND HOUSEHOLDERS DATA

3.4. DATA COLLECTION & CLEANING

➤ ELECTRICITY CONSUMPTION DATA COLLECTION

SUB-METER → SUB-METER → SUB-METER → DATABASE SERVER → API → DATA FRAME

- Aggregate data hourly
- All submeters together

ENERBYTE PLATFORM

- Sub-metering devices
- Small dataset 150 users
- Hourly measures (0-23h)
- Around 1 year period
- Missing values → Wi-Fi

## ➤ ELECTRICITY CONSUMPTION DATA CLEANING

- Discarded sub-meters
- ZEROS → Wi-Fi OFF
- FROZEN → Device issue

121 Valid users

	row.names	idmeter	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00
1	773	112695	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	774	112695	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	775	112695	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	776	112695	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	777	112695	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	778	112695	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	779	112695	149	152	156	186	149	142	142	148	231	160	139	458	826	247	192	238	201
8	780	112695	240	166	190	125	160	140	187	178	169	159	286	237	166	115	140	165	187
9	781	112695	228	215	144	187	134	161	179	146	220	187	169	560	594	149	148	163	219
10	782																		206
11	783																		148

|	row.names	idmeter	date	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00
1	19819	112769	24/04/2014	255	218	134	88	120	131	95	109	152	161	106	133	106	100	130
2	19820	112769	25/04/2014	296	248	128	83	128	127	80	132	162	121	168	150	135	170	158
3	19821	112769	26/04/2014	297	362	163	94	127	121	89	133	118	98	238	90	128	125	87
4	19822	112769	27/04/2014	142	381	305	217	135	98	103	133	134	99	154	128	85	123	128
5	19823	112769	28/04/2014	230	241	77	134	112	108	137	86	226	290	129	128	203	199	133
6	19824	112769	29/04/2014	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195
7	19825	112769	30/04/2014	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195
8	19826	112769	01/05/2014	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195
9	19827	112769	02/05/2014	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195
10	19828	112769	03/05/2014	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195
11	19829	112769	04/05/2014	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195
12	19830	112769	05/05/2014	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195
13	19831	112769	06/05/2014	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195
14	19832	112769	07/05/2014	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195
15	19833	112769	08/05/2014	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195

➤ VISUALIZATION "RUBÍ" LOAD PROFILES
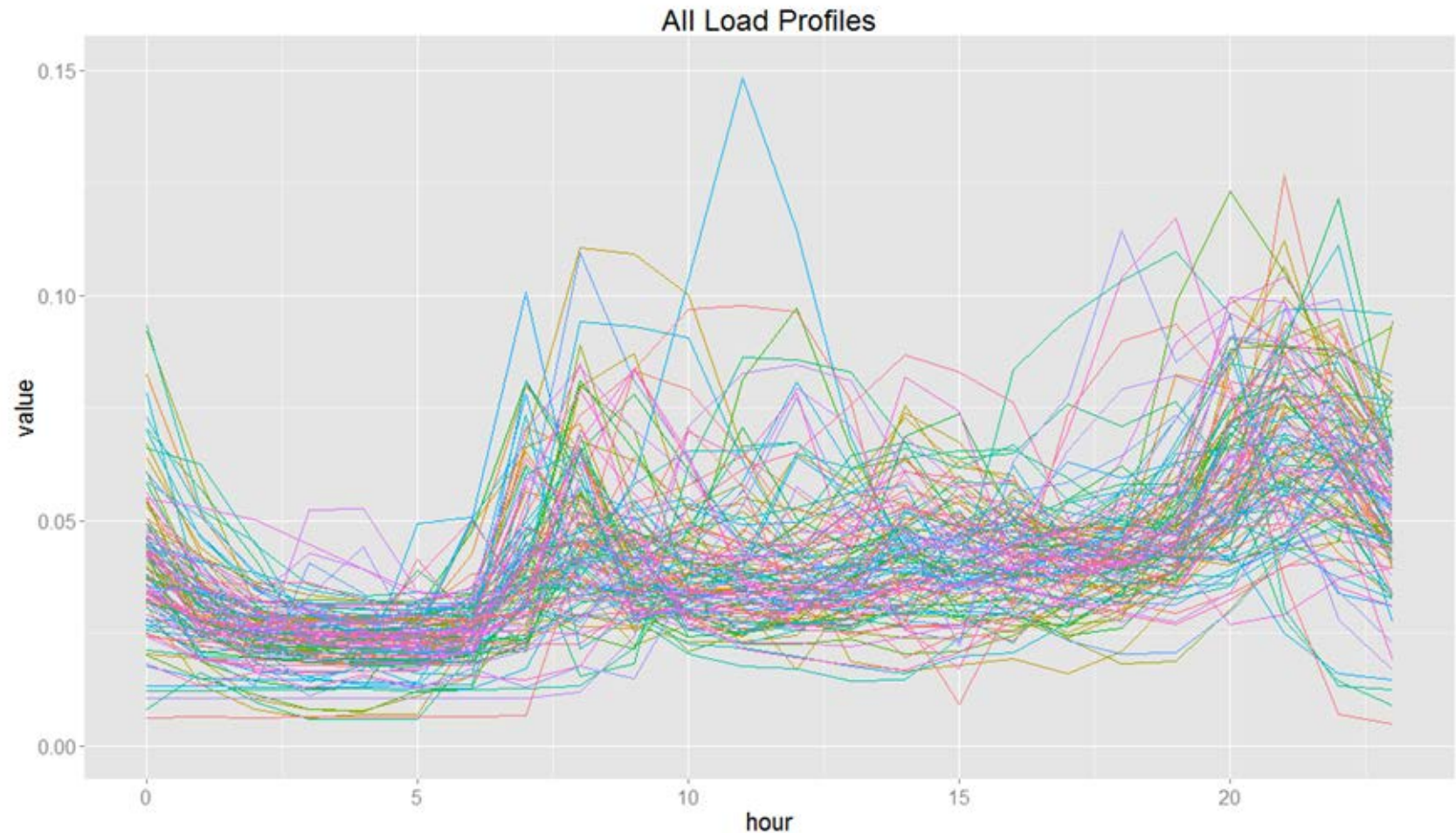
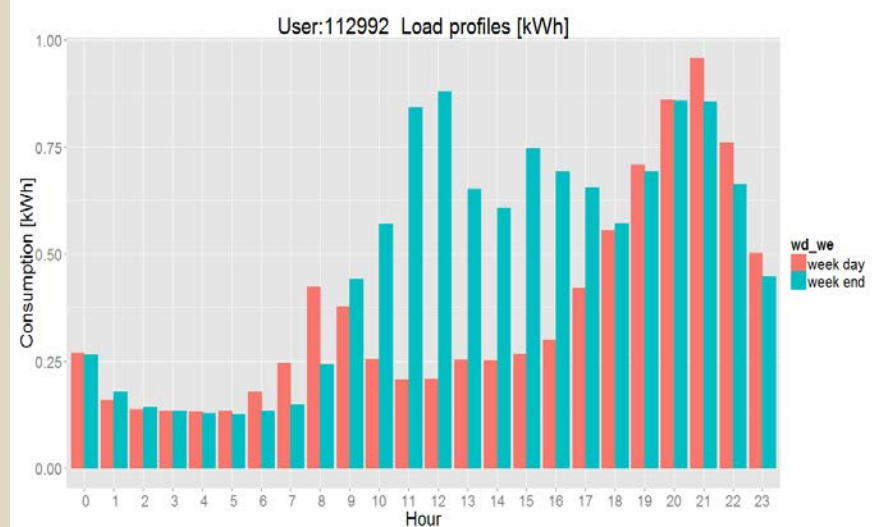# 3.5.1. PRE-CLUSTERING

INPUT DATA CALCULATION

➤ COMMON FRAMEWORK:
   FAIRLY COMPARISON

➤ PROPORTION OF HOURLY ENERGY USAGE
   (%)

- Sum all daily values (0-23h)
- Divide each hour value by the sum of each row
- Mean hourly percentage values per user (%)
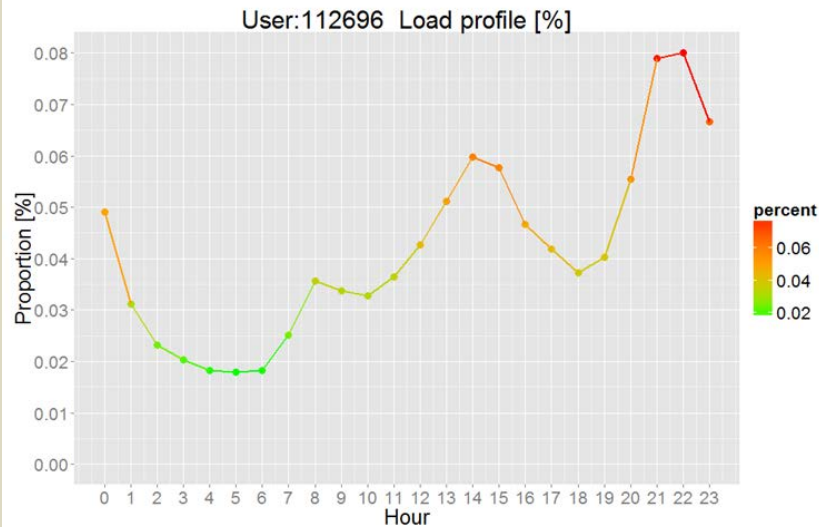
INPUT DATA FOR CLUSTERING

➤ ONLY WEEKDAYS (MONDAY-FRIDAY)

➤ 24 HOUR'S VECTOR (00:00-23:00H)

User:112696 Load profile [%]

User:112992 Load profiles [kWh]

3.5.2. CLUSTERING

PROCEDURE

Step 1:	Step 2:	Step 3:	Step 4:
Define Similarity	Choose cluster techniques	Clusters Visualization	Clusters interpretation

DISTANCES AND LINKAGES

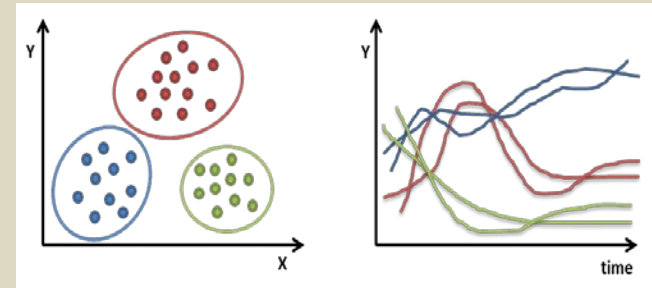EUCLIDEAN DISTANCE
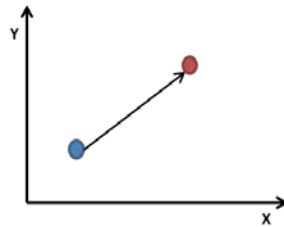
MANHATTAN DISTANCE
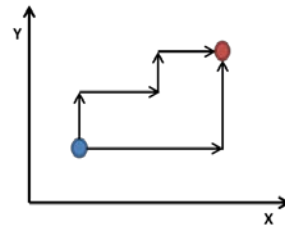
SINGLE LINKAGE
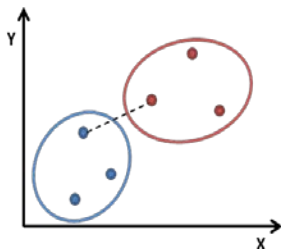
COMPLETE LINKAGE
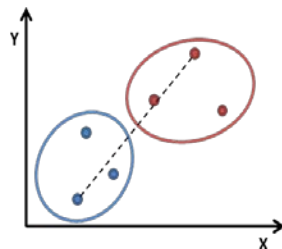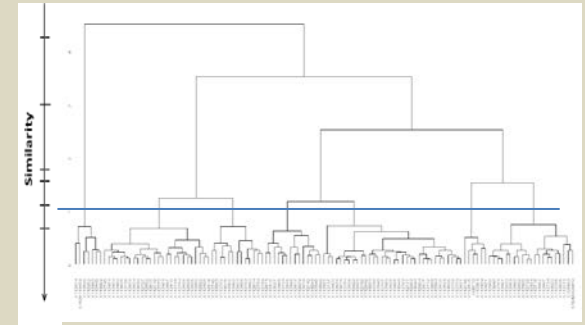
UPGMC LINKAGE
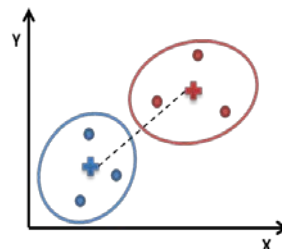
TIMES SERIES DATA

CLUSTERING TECHNIQUES

- ➢ Hierarchical
- ➢ K-means
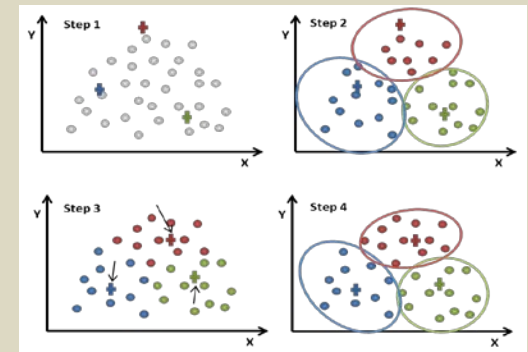- ➢ Self-Organizing Maps (SOM)

3.5.2. CLUSTERING

Hierarchical

➢ Bottom-up
➢ Dendrogram (tree)
➢ Manhattan distance & Ward linkage
➢ Cut the tree (a posteriori) → 7 groups

K-means

➢ k-means algorithm
  ▪ K defined a priori
  ▪ Not global minimum (WCSS)
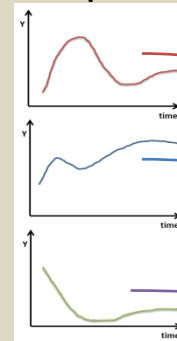➢ Hartigan-Wong's algorithm
➢ K= 7 groups

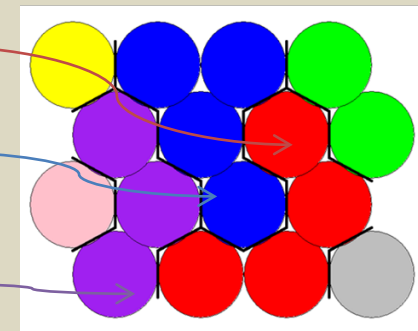$$arg\ min \sum_{1}^{k} \sum_{1}^{n} \|x_i - \bar{\mu}\|^2$$

Self-Organizing Maps

➢ Dimension reduction in a Map
➢ From 121 users to 4x4 nodes Map
➢ Apply clustering techniques → Hierarchical
➢ Manhattan distance, Ward linkage, 7 groups
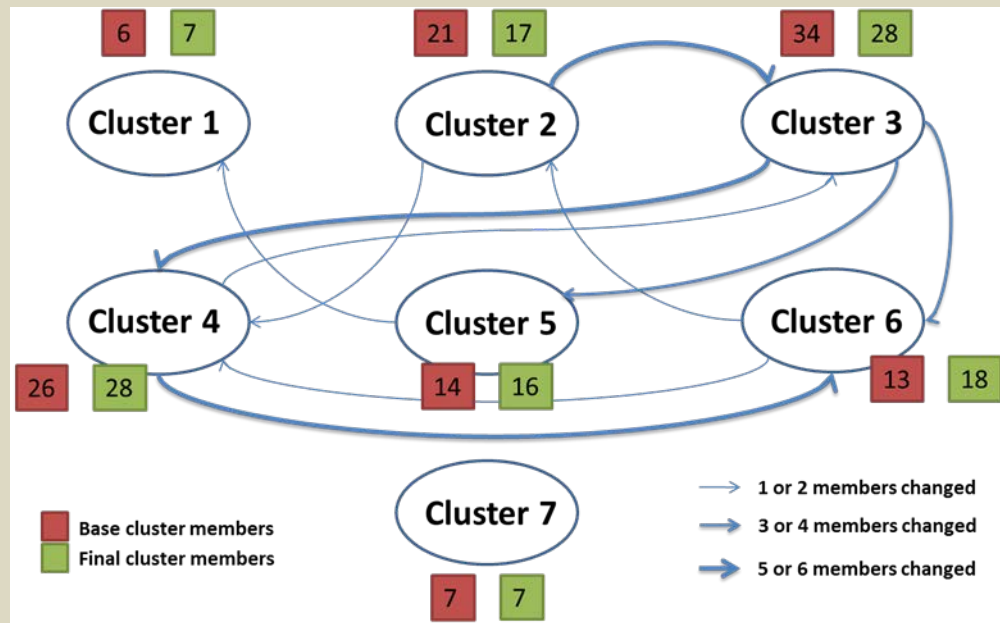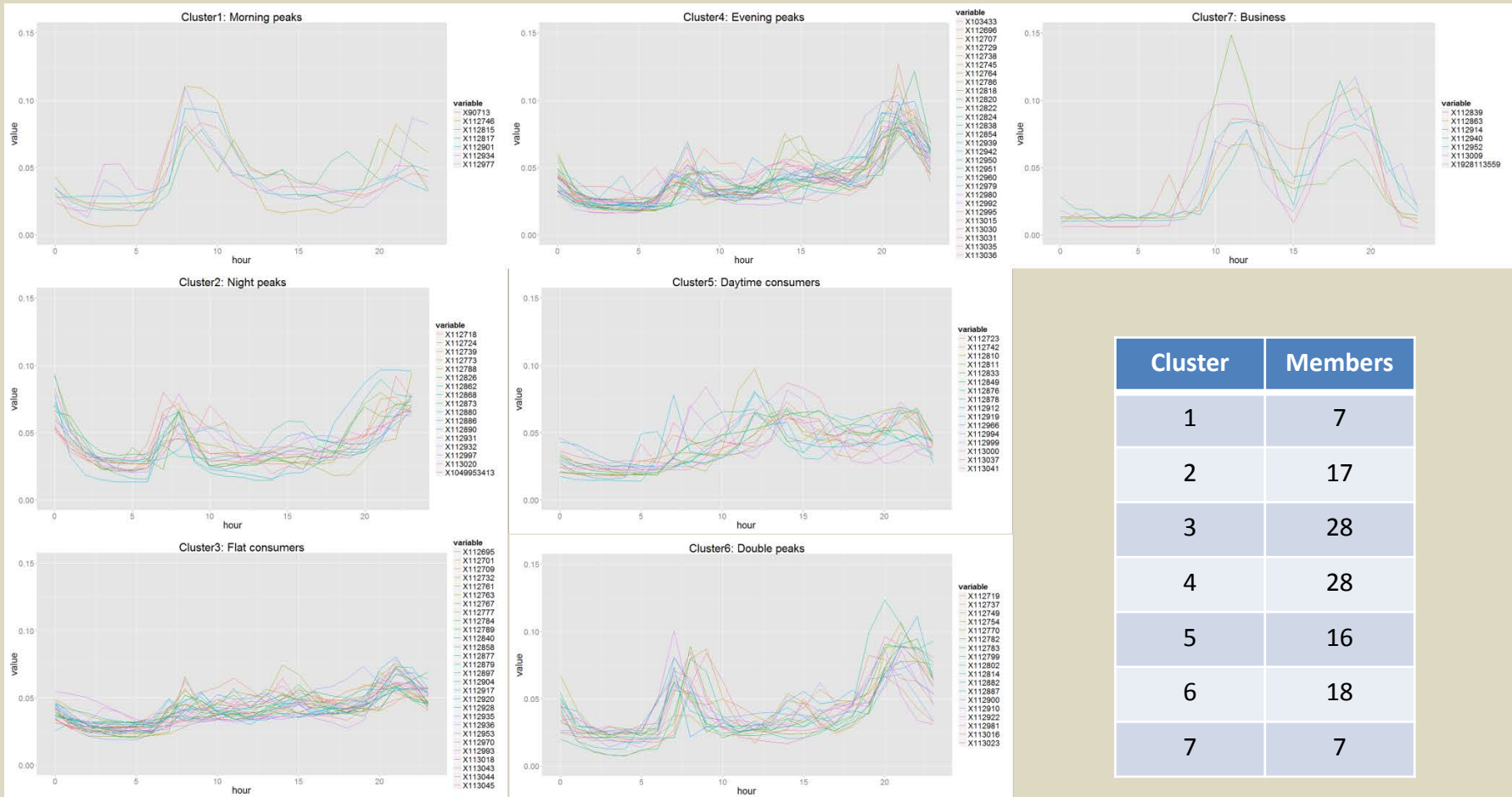

Load profiles   **Clusters in the Map**

Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7

- Base solution: Hierachical Clustering
    - 7 groups and Manhattan distance and Ward linkage

- Visualization and Statistical

- Find outliers:
    - Cluster of a cluster
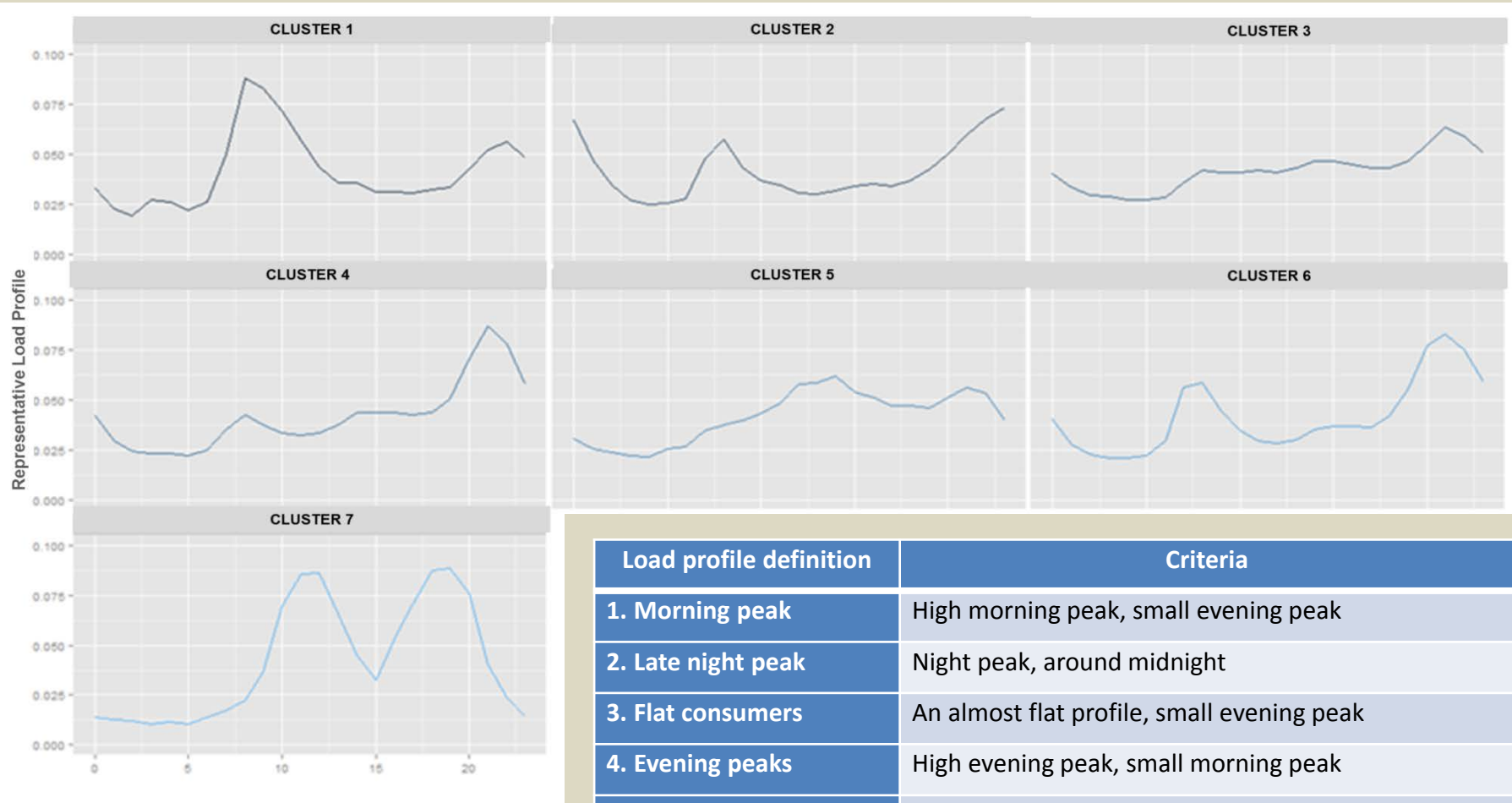   - Distance to the mean

- Members reallocation

Cluster	Members
1	7
2	17
3	28
4	28
5	16
6	18
7	7

Load profile definition	Criteria
1. Morning peak	High morning peak, small evening peak
2. Late night peak	Night peak, around midnight
3. Flat consumers	An almost flat profile, small evening peak
4. Evening peaks	High evening peak, small morning peak
5. Daytime consumers	Highest period of consumption during the day
6. Double peak	Morning and evening peaks
7. Business	Not residential. Small business

➤ **Energy efficiency in Residential sector**
- Translate technical data into call -to-action measures
- GPS effect

➤ **Benefits or applications of consumer segmentation**
- Personalised advices
  - Increase effectiveness
  - Test different energy efficiency strategies to same group
- Reference or baseline groups
- Keep track of users
  - Change the load profile group or not
  - Quantify the savings for each consumer (in kWh)

➤ **Further steps**
- Larger datasets and better data quality   → Smart meter data
- Improve Households and householders analysis

THANK YOU !!!