



# Classification des consommateurs d'électricité dans le système Smart Grid

---

PRÉSENTÉ PAR:

YOSR TAKOUTI

MOHAMED ALI GUIRAT

INDP3\_SYSTIC

ENCADRÉ PAR:

MME ONS BEN RHOUMA

MME MANEL BEN ROMDHANE

**04/01/2022**



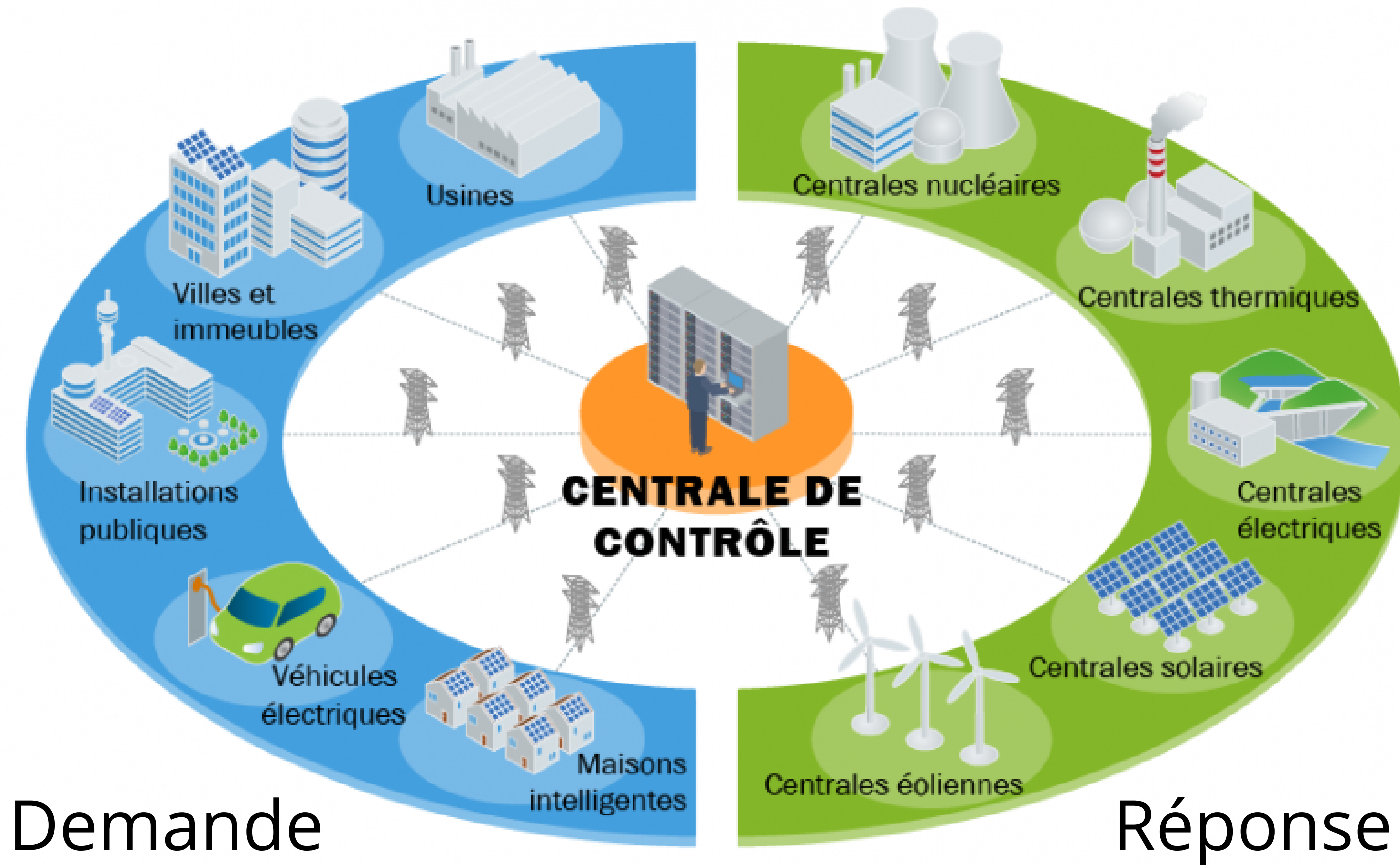
# Plan

- 1- Cadre du projet
  - 2- Base de données
  - 3- Préparation et visualisation des données
  - 4- Classification des données
  - 5- Visualisation des résultats
  - 6- Conclusion
-

# I-Cadre du projet:

# Réseau intelligent

4

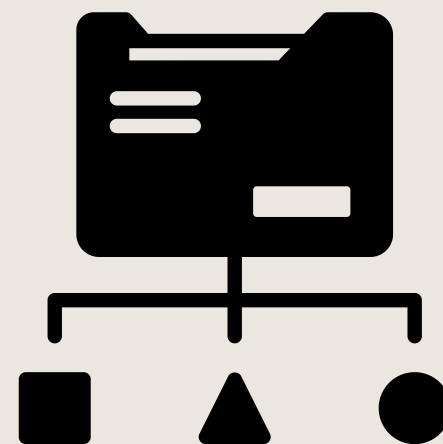


# OBJECTIF



5

L'objectif principal consiste donc à regrouper des observations similaires afin de faciliter l'examen de l'ensemble de données. Par conséquent, on procédera à l'identification des modèles de consommation de l'énergie avec l'algorithme de « clustering » K-means



2-Base de données:

# La base de données

- 15 Fichiers CSV
- un million de mesures
- 4 Attributs

```
1 LCLid,stdorToU,DateTime,KWH/hh (per half hour)
2 MAC005007,Std,2012-07-09 02:30:00.0000000, 0.04
3 MAC005007,Std,2012-07-09 03:00:00.0000000, 0.042
4 MAC005007,Std,2012-07-09 03:30:00.0000000, 0.039
5 MAC005007,Std,2012-07-09 04:00:00.0000000, 0.028
6 MAC005007,Std,2012-07-09 04:30:00.0000000, 0.032
7 MAC005007,Std,2012-07-09 05:00:00.0000000, 0.043
8 MAC005007,Std,2012-07-09 05:30:00.0000000, 0.04
9 MAC005007,Std,2012-07-09 06:00:00.0000000, 0.029
10 MAC005007,Std,2012-07-09 06:30:00.0000000, 0.03
11 MAC005007,Std,2012-07-09 07:00:00.0000000, 0.175
12 MAC005007,Std,2012-07-09 07:30:00.0000000, 0.157
13 MAC005007,Std,2012-07-09 08:00:00.0000000, 0.064
14 MAC005007,Std,2012-07-09 08:30:00.0000000, 0.07
15 MAC005007,Std,2012-07-09 09:00:00.0000000, 0.294
16 MAC005007,Std,2012-07-09 09:30:00.0000000, 0.04
17 MAC005007,Std,2012-07-09 10:00:00.0000000, 0.033
18 MAC005007,Std,2012-07-09 10:30:00.0000000, 0.028
19 MAC005007,Std,2012-07-09 11:00:00.0000000, 0.04
20 MAC005007,Std,2012-07-09 11:30:00.0000000, 0.042
21 MAC005007,Std,2012-07-09 12:00:00.0000000, 0.037
22 MAC005007,Std,2012-07-09 12:30:00.0000000, 0.025
23 MAC005007,Std,2012-07-09 13:00:00.0000000, 0.041
24 MAC005007,Std,2012-07-09 13:30:00.0000000, 0.041
25 MAC005007,Std,2012-07-09 14:00:00.0000000, 0.038
26 MAC005007,Std,2012-07-09 14:30:00.0000000, 0.024
27 MAC005007,Std,2012-07-09 15:00:00.0000000, 0.04
28 MAC005007,Std,2012-07-09 15:30:00.0000000, 0.041
29 MAC005007,Std,2012-07-09 16:00:00.0000000, 0.038
30 MAC005007,Std,2012-07-09 16:30:00.0000000, 0.035
```

3-réparation et visualisation des données:



# Concaténation

9

```
file.head(10)
```

	LCLid	stdorToU	DateTime	KWH/hh (per half hour)
0	MAC005007	Std	2012-07-09 02:30:00.0000000	0.04
1	MAC005007	Std	2012-07-09 03:00:00.0000000	0.042
2	MAC005007	Std	2012-07-09 03:30:00.0000000	0.039
3	MAC005007	Std	2012-07-09 04:00:00.0000000	0.028
4	MAC005007	Std	2012-07-09 04:30:00.0000000	0.032
5	MAC005007	Std	2012-07-09 05:00:00.0000000	0.043
6	MAC005007	Std	2012-07-09 05:30:00.0000000	0.04
7	MAC005007	Std	2012-07-09 06:00:00.0000000	0.029
8	MAC005007	Std	2012-07-09 06:30:00.0000000	0.03
9	MAC005007	Std	2012-07-09 07:00:00.0000000	0.175

```
file.tail(10)
```

	LCLid	stdorToU	DateTime	KWH/hh (per half hour)
14999990	MAC000146	ToU	2012-08-16 21:00:00.0000000	0.188
14999991	MAC000146	ToU	2012-08-16 21:30:00.0000000	0.137
14999992	MAC000146	ToU	2012-08-16 22:00:00.0000000	0.116
14999993	MAC000146	ToU	2012-08-16 22:30:00.0000000	0.138
14999994	MAC000146	ToU	2012-08-16 23:00:00.0000000	0.197
14999995	MAC000146	ToU	2012-08-16 23:30:00.0000000	0.155
14999996	MAC000146	ToU	2012-08-17 00:00:00.0000000	0.125
14999997	MAC000146	ToU	2012-08-17 00:30:00.0000000	0.069
14999998	MAC000146	ToU	2012-08-17 01:00:00.0000000	0.101
14999999	MAC000146	ToU	2012-08-17 01:30:00.0000000	0.071

# Nettoyage

10

	LCLid	DateTime	KWH/hh (per half hour)
0	MAC005007	2012-07-09 02:30:00.0000000	0.04
1	MAC005007	2012-07-09 03:00:00.0000000	0.042
2	MAC005007	2012-07-09 03:30:00.0000000	0.039
3	MAC005007	2012-07-09 04:00:00.0000000	0.028
4	MAC005007	2012-07-09 04:30:00.0000000	0.032
5	MAC005007	2012-07-09 05:00:00.0000000	0.043
6	MAC005007	2012-07-09 05:30:00.0000000	0.04
7	MAC005007	2012-07-09 06:00:00.0000000	0.029
8	MAC005007	2012-07-09 06:30:00.0000000	0.03
9	MAC005007	2012-07-09 07:00:00.0000000	0.175

# Nombre des compteurs & Nombre des mesures Totales

II

```
print("Le nombre de compteurs:")  
print(file["LCLid"].value_counts())
```

```
Le nombre de compteurs:  
MAC000017      39114  
MAC000015      39089  
MAC000031      39059  
MAC000048      39028  
MAC000046      39025  
...  
MAC005062       9726  
MAC005559         1  
MAC005556         1  
MAC005560         1  
MAC005563         1  
Name: LCLid, Length: 461, dtype: int64
```

```
mesure = len(df)  
print("nombre de mesure est ", mesure)
```

```
nombre de mesure est 14999541
```

**461 Compteurs**

# Détermination des paramètres de consommation énergétique par compteur

LCLid	consommation			
	sum	min	max	mean
MAC000005	2911.808000	0.0	1.979	0.095385
MAC000014	3132.670999	0.0	2.132	0.130332
MAC000015	8957.714000	0.0	2.613	0.229162
MAC000017	5202.072001	0.0	2.331	0.132998
MAC000031	16295.187997	0.0	4.247	0.417194
...	...	...	...	...
MAC005561	4193.402000	0.0	1.886	0.151995
MAC005562	6910.988999	0.0	1.879	0.218012
MAC005563	0.000000	0.0	0.000	0.000000
MAC005566	8946.788998	0.0	3.936	0.363085
MAC005567	2268.484999	0.0	1.528	0.108556

461 rows x 4 columns

Consommation  
moyenne par jour  
de semaine,  
moyenne par jour  
de week-end

LCLid	consommation					
	mean					
	Friday	Monday	Thursday	Tuesday	Wednesday	Weekend
MAC000005	0.096364	0.104157	0.089285	0.093749	0.093253	0.095452
MAC000014	0.122962	0.127625	0.133138	0.138966	0.124083	0.132754
MAC000015	0.279244	0.206488	0.224172	0.213075	0.245149	0.217960
MAC000017	0.137118	0.129727	0.131517	0.118622	0.139714	0.137141
MAC000031	0.353874	0.379685	0.363709	0.397497	0.452044	0.486887
...	...	...	...	...	...	...
MAC005561	0.143070	0.154062	0.155568	0.157067	0.153036	0.150564
MAC005562	0.211326	0.216657	0.213150	0.223063	0.218843	0.221510
MAC005563	NaN	NaN	NaN	NaN	0.000000	NaN
MAC005566	0.370457	0.368846	0.357546	0.376491	0.356384	0.356017
MAC005567	0.118291	0.111119	0.098492	0.106132	0.103454	0.111280
461 rows × 6 columns						

# Consommation par mois

month	mean											
	April	August	December	February	January	July	June	March	May	November	October	September
LCLid												
MAC000005	0.091192	0.081748	0.120143	0.111776	0.115403	0.078975	0.072775	0.112075	0.076702	0.110640	0.093258	0.077257
MAC000014	0.121143	0.109840	0.148667	0.142179	0.149668	0.084555	0.110174	0.111492	0.113759	0.136778	0.138323	0.139617
MAC000015	0.172663	0.187595	0.261687	0.284527	0.278502	0.183657	0.189356	0.281328	0.181244	0.252982	0.214964	0.200658
MAC000017	0.126987	0.123031	0.138492	0.149389	0.166431	0.130255	0.109097	0.126562	0.106177	0.140711	0.134822	0.118020
MAC000031	0.350321	0.178974	0.454380	0.494584	0.550488	0.340201	0.364273	0.476084	0.411682	0.454637	0.425191	0.392951
...	...	...	...	...	...	...	...	...	...	...	...	...
MAC005561	0.143615	0.091422	0.199011	0.210832	0.192797	0.114964	0.137646	0.176398	0.107396	0.176110	0.120618	0.117057
MAC005562	0.260709	0.142179	0.262332	0.273900	0.262678	0.168384	0.169062	0.282420	0.178002	0.239065	0.224016	0.207226
MAC005563	NaN	NaN	0.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MAC005566	0.356997	0.607306	0.126849	0.185331	0.110057	0.854140	0.502741	0.087798	0.437714	0.433770	0.541430	0.516285
MAC005567	0.085281	0.118277	0.117174	0.109162	0.115221	0.111548	0.095264	0.111419	0.090206	0.120120	0.106475	0.111298

# Consommation par période de jour

periode LCLid	periode 1	periode 2	periode 3
MAC000005	0.044894	0.091721	0.149583
MAC000014	0.080308	0.187521	0.123193
MAC000015	0.127207	0.268949	0.291527
MAC000017	0.070333	0.138637	0.190092
MAC000031	0.205624	0.489440	0.556824
...	...	...	...
MAC005561	0.075379	0.185921	0.194765
MAC005562	0.112731	0.252142	0.289393
MAC005563	NaN	0.000000	NaN
MAC005566	0.250454	0.429897	0.409116
MAC005567	0.065945	0.106982	0.152811

461 rows x 3 columns

# Consommation par saison

season	season 1	season 2	season 3	season 4
LCLid				16
MAC000005	0.115930	0.093347	0.077928	0.093711
MAC000014	0.147020	0.115403	0.101427	0.137932
MAC000015	0.274848	0.212183	0.186842	0.222772
MAC000017	0.151782	0.119833	0.120920	0.131224
MAC000031	0.501116	0.413466	0.293717	0.424273
...	...	...	...	...
MAC005561	0.200493	0.142463	0.108924	0.137740
MAC005562	0.266006	0.226883	0.159673	0.223442
MAC005563	0.000000	NaN	NaN	NaN
MAC005566	0.139007	0.293534	0.656380	0.493330
MAC005567	0.113664	0.095749	0.108505	0.112563

461 rows x 4 columns



# Regrouper les dataframes

17

df\_finale

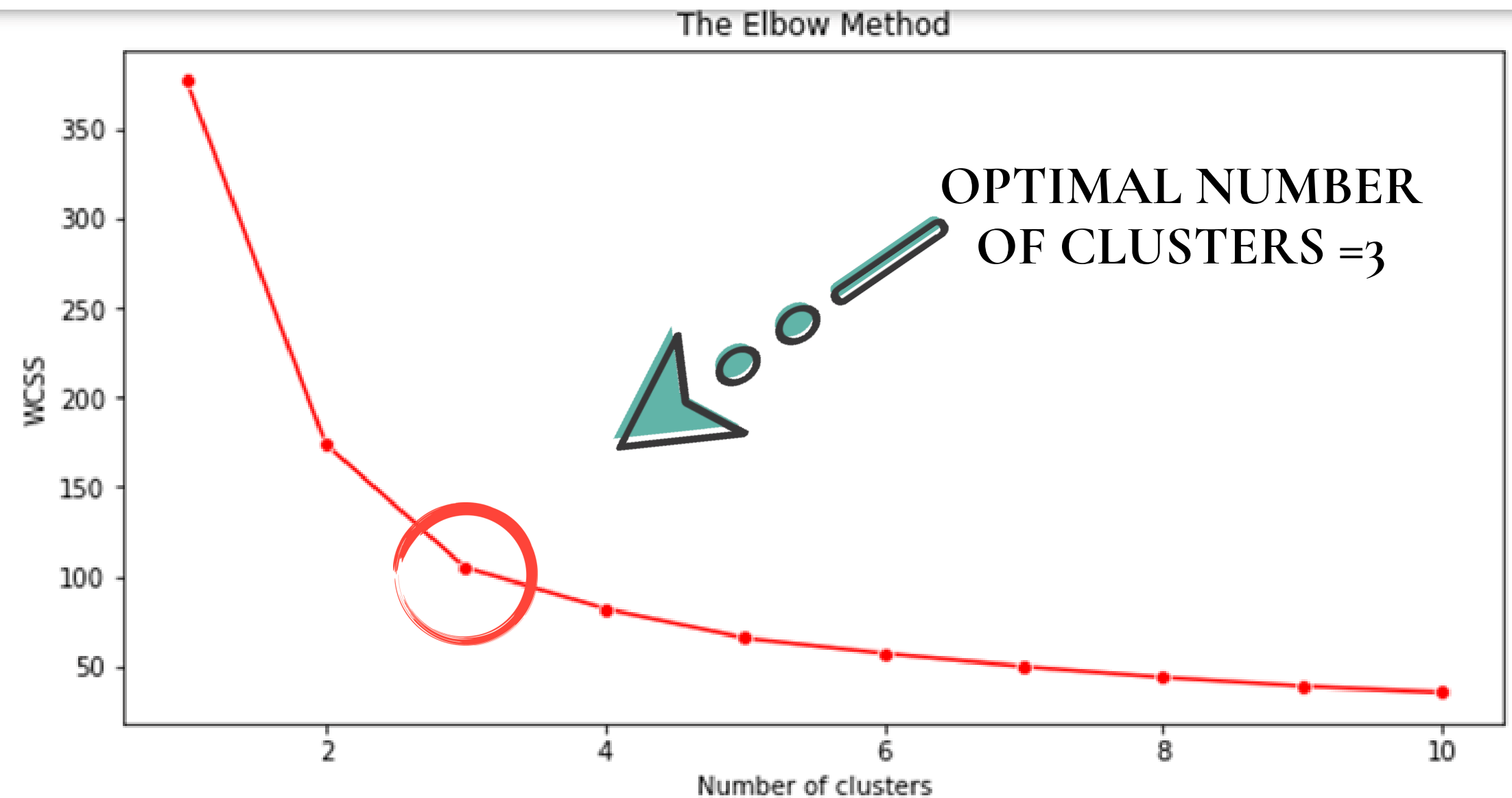
Friday	Monday	Thursday	Tuesday	Wednesday	...	November	October	September	season 1	season 2	season 3	season 4	periode 1	periode 2	periode 3
0.096364	0.104157	0.089285	0.093749	0.093253	...	0.110640	0.093258	0.077257	0.115930	0.093347	0.077928	0.093711	0.044894	0.091721	0.149583
0.122962	0.127625	0.133138	0.138966	0.124083	...	0.136778	0.138323	0.139617	0.147020	0.115403	0.101427	0.137932	0.080308	0.187521	0.123193
0.279244	0.206488	0.224172	0.213075	0.245149	...	0.252982	0.214964	0.200658	0.274848	0.212183	0.186842	0.222772	0.127207	0.268949	0.291527
0.137118	0.129727	0.131517	0.118622	0.139714	...	0.140711	0.134822	0.118020	0.151782	0.119833	0.120920	0.131224	0.070333	0.138637	0.190092
0.353874	0.379685	0.363709	0.397497	0.452044	...	0.454637	0.425191	0.392951	0.501116	0.413466	0.293717	0.424273	0.205624	0.489440	0.556824
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
0.143070	0.154062	0.155568	0.157067	0.153036	...	0.176110	0.120618	0.117057	0.200493	0.142463	0.108924	0.137740	0.075379	0.185921	0.194765
0.211326	0.216657	0.213150	0.223063	0.218843	...	0.239065	0.224016	0.207226	0.266006	0.226883	0.159673	0.223442	0.112731	0.252142	0.289393
NaN	NaN	NaN	NaN	0.000000	...	NaN	NaN	NaN	0.000000	NaN	NaN	NaN	NaN	0.000000	NaN
0.370457	0.368846	0.357546	0.376491	0.356384	...	0.433770	0.541430	0.516285	0.139007	0.293534	0.656380	0.493330	0.250454	0.429897	0.409116
0.118291	0.111119	0.098492	0.106132	0.103454	...	0.120120	0.106475	0.111298	0.113664	0.095749	0.108505	0.112563	0.065945	0.106982	0.152811

## 4-Classification des données:

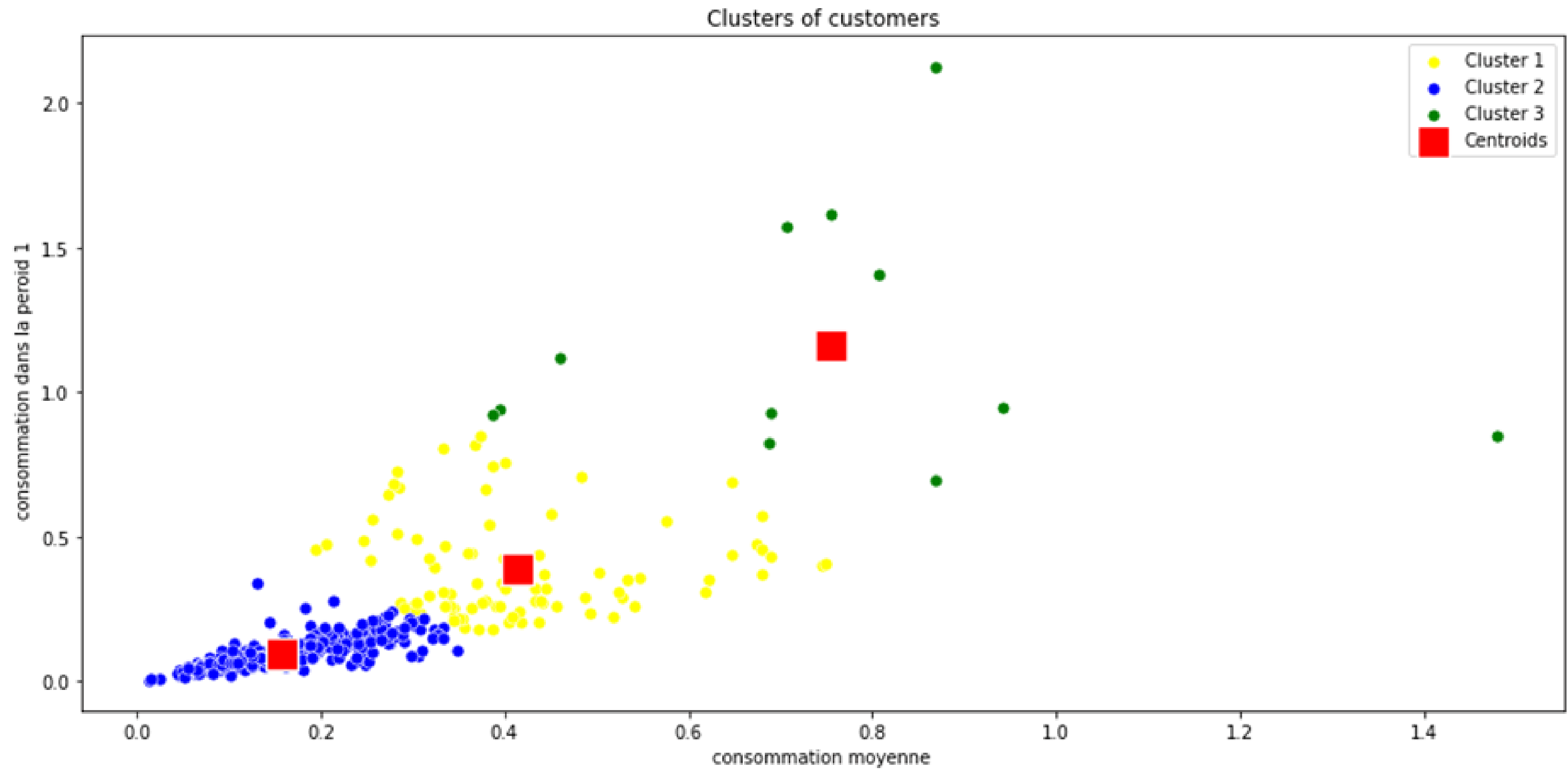
# Algorithme de K-means:

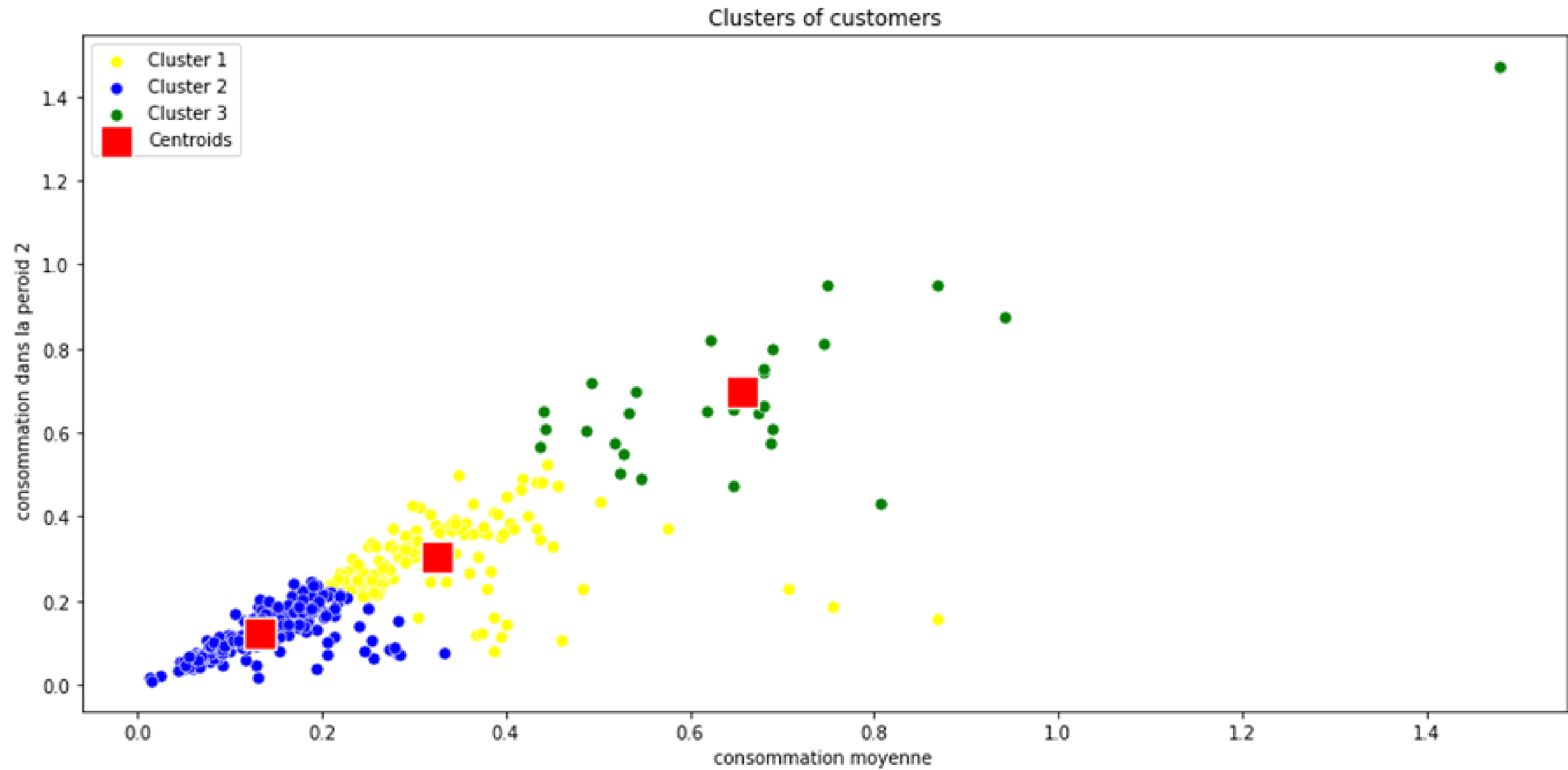
# Elbow Method:

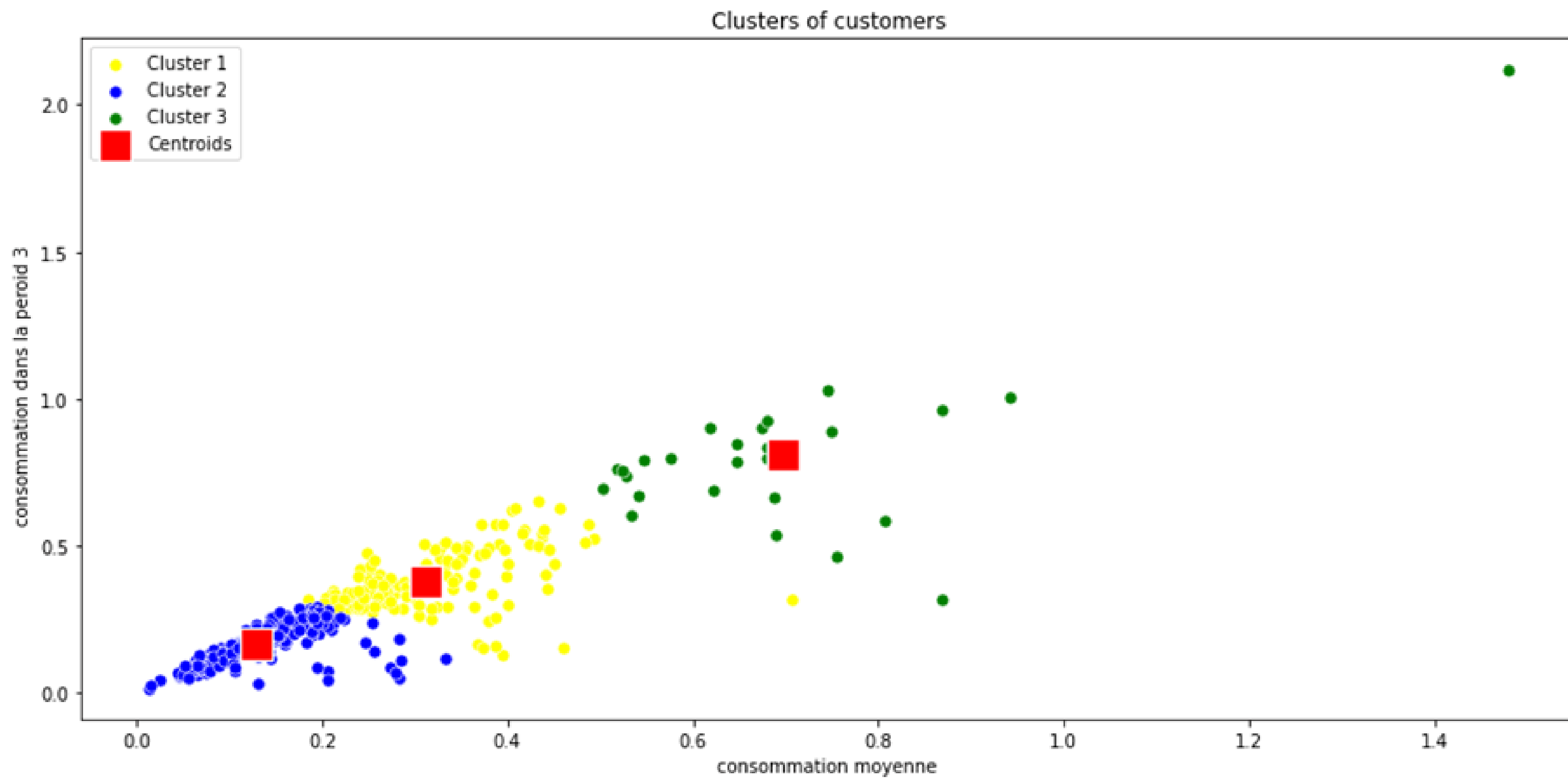
- ✓ Il excelle à identifier des groupes de formes non sphériques.
- ✓ Il résiste aux valeurs aberrants



## 5-Visualisation des résultats:







## 6-Conclusion:



- K-means est le plus adéquat pour notre application car il a pour objectif de regrouper des éléments similaires dans des clusters.
- Pour les perceptive et après l'exploration de données et décilage des individus similaires, on peut passer à la prédiction ou application des algorithmes d'optimisation de la consommation énergétique.



Merci pour votre  
attention

