Study of demand prediction using clustering methods

Application to the consumption of medications in hospitals

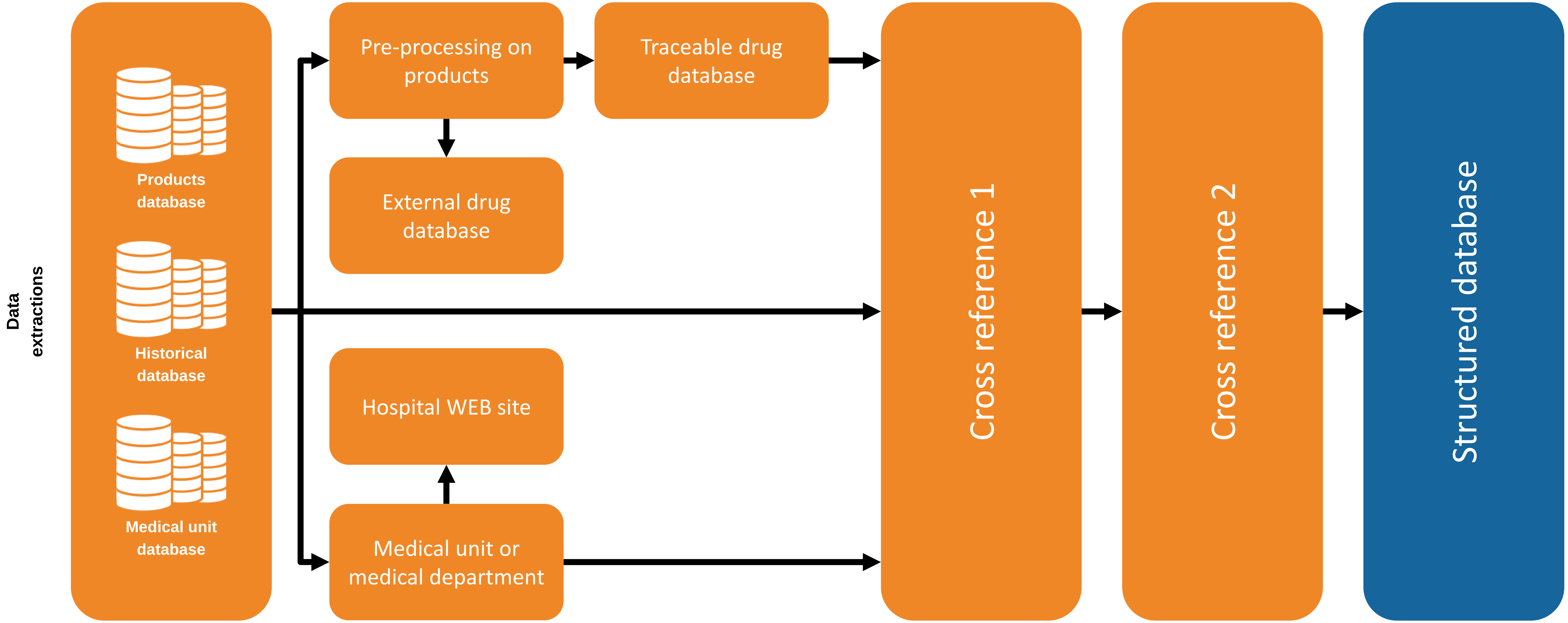
Presentation to KLS, September 7 2023

Summary

Author: Douglas MACHADO

Supervisors: Gülgün ALPAN | Zakaria YAHOUNI

1: Context and Methodology



1: Context and Methodology

- The main objective of the present research is to identify **similarities** among various **medications**, thereby reducing the number of required models for **forecasting** hospital **demand** and optimizing stock resource management.

ID_REF	ID_SITE_RATTACHE	CODE_ATC	HOSPI_CODE_UCD	DATE_MOUV	N_UFS	QUANTITY	WEEK	MONTH	YEAR	N_ETB	POPULATION	P_MEDICAL	PN_MEDICAL	LIT_HC	LIT_HP	SEJ_MCO	SEJ_HAD	SEJ_PSY	SEJ_SSR	SEJ_SLD
502829	HOSPI_3	C01CA03	3400892508566	2014-11-07	3	210.0	45.0	11	2014.0	50	1107398.0	1158	7129	2063.0	521.0	117781	594	2903	1302	97
502829	HOSPI_3	C01CA03	3400892508566	2015-01-22	4	340.0	4.0	1	2015.0	50	1120190.0	1239	7161	2053.0	493.0	118924	650	2878	1334	75
501463	HOSPI_3	J01CR05	3400893022634	2018-02-14	2	200.0	7.0	2	2018.0	50	1159220.0	1322	7439	2008.0	517.0	115376	1093	2481	1183	118
9220364	HOSPI_2	H02AB06	3400892203645	2017-06-30	4	55.0	26.0	6	2017.0	5	539067.0	714	5001	1157.0	187.0	75420	0	1236	261	0
9490	HOSPI_4	B01AC06	3400892065366	2018-03-26	1	630.0	13.0	3	2018.0	39	1859524.0	2627	15723	4477.0	486.0	255490	0	837	8416	209
890900	HOSPI_1	M03BX01	3400892697789	2015-10-06	1	20.0	41.0	10	2015.0	12	571879.0	684	5295	1411.0	94.0	74102	0	0	1140	57
9272958	HOSPI_2	N02AX02	3400892729589	2019-12-30	5	150.0	1.0	12	2019.0	5	542302.0	706	5013	1141.0	141.0	76593	0	1007	206	0
9387549	HOSPI_2	N02BE01	3400893875490	2018-05-21	2	30.0	21.0	5	2018.0	5	541454.0	703	5007	1159.0	139.0	74663	0	1193	237	0
503386	HOSPI_3	N02BE01	3400893875490	2015-03-18	18	410.0	12.0	3	2015.0	50	1120190.0	1239	7161	2053.0	493.0	118924	650	2878	1334	75
503129	HOSPI_3	N02BE01	3400891996128	2015-08-05	35	4350.0	32.0	8	2015.0	50	1120190.0	1239	7161	2053.0	493.0	118924	650	2878	1334	75

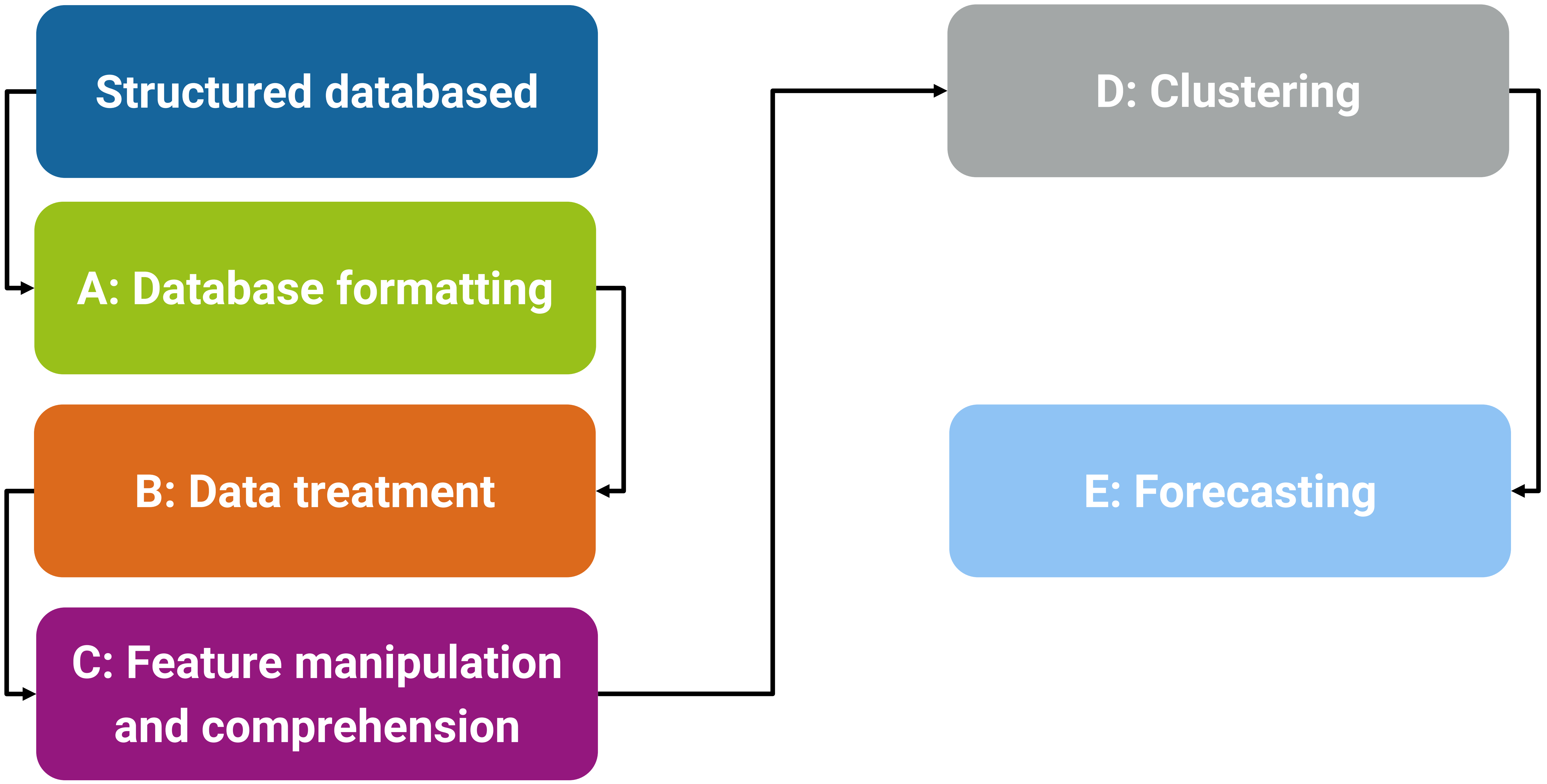
- Studying 21 medications across 4 different hospitals, we have a dataset comprising 75,684 data points with 21 features.

- The distribution of data points among hospitals is as follows:
  - Hospital 1: 22, 725 | Hospital 2: 15, 439 Hospital 3: 27, 591 | Hospital 4: 9, 929

G-SCOP SCIENCES POUR LA CONCEPTION, L'OPTIMISATION ET LA PRODUCTION

Grenoble INP

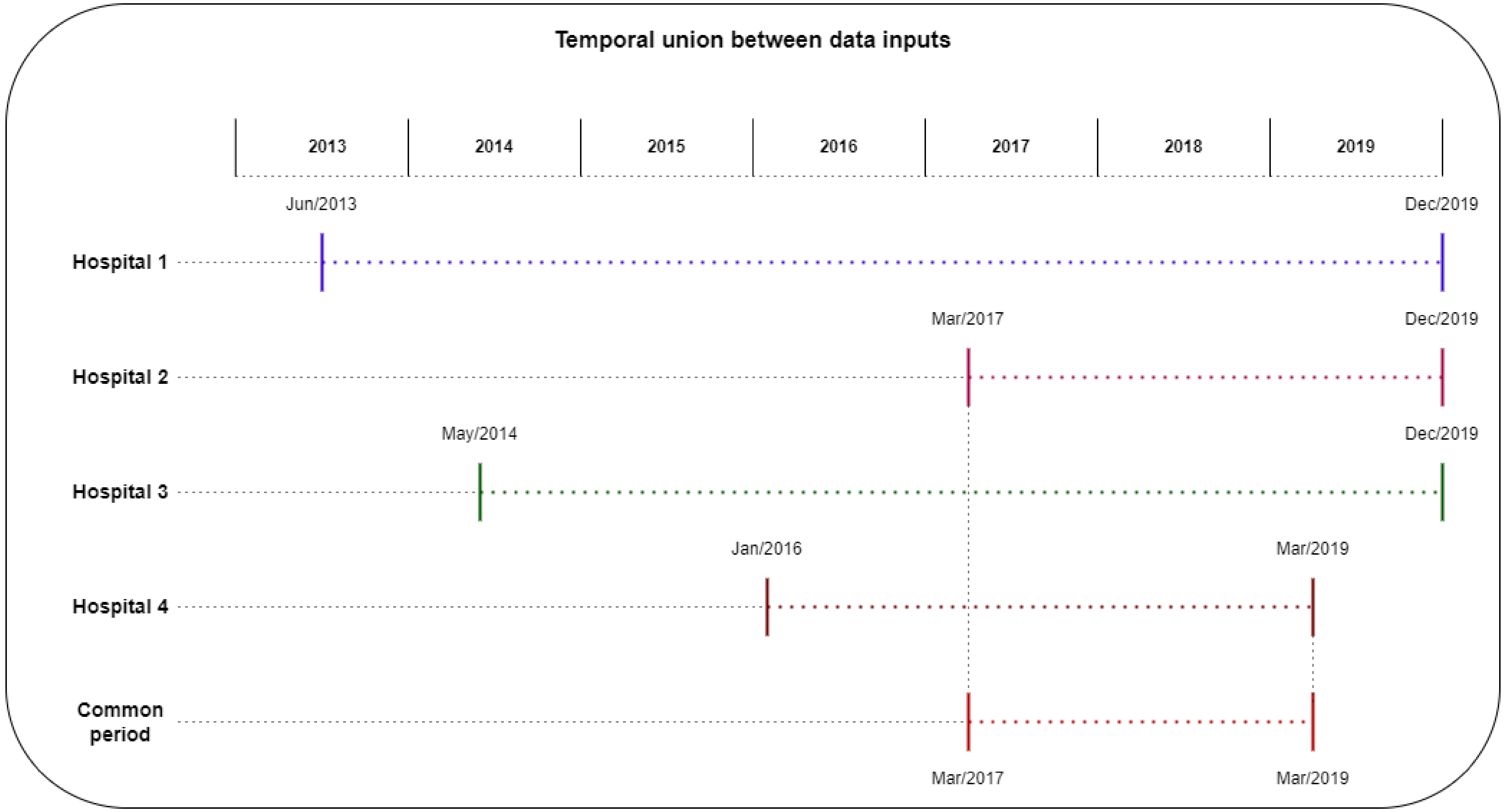
1: Context and Methodology



1: Context and Methodology

B: Data treatment

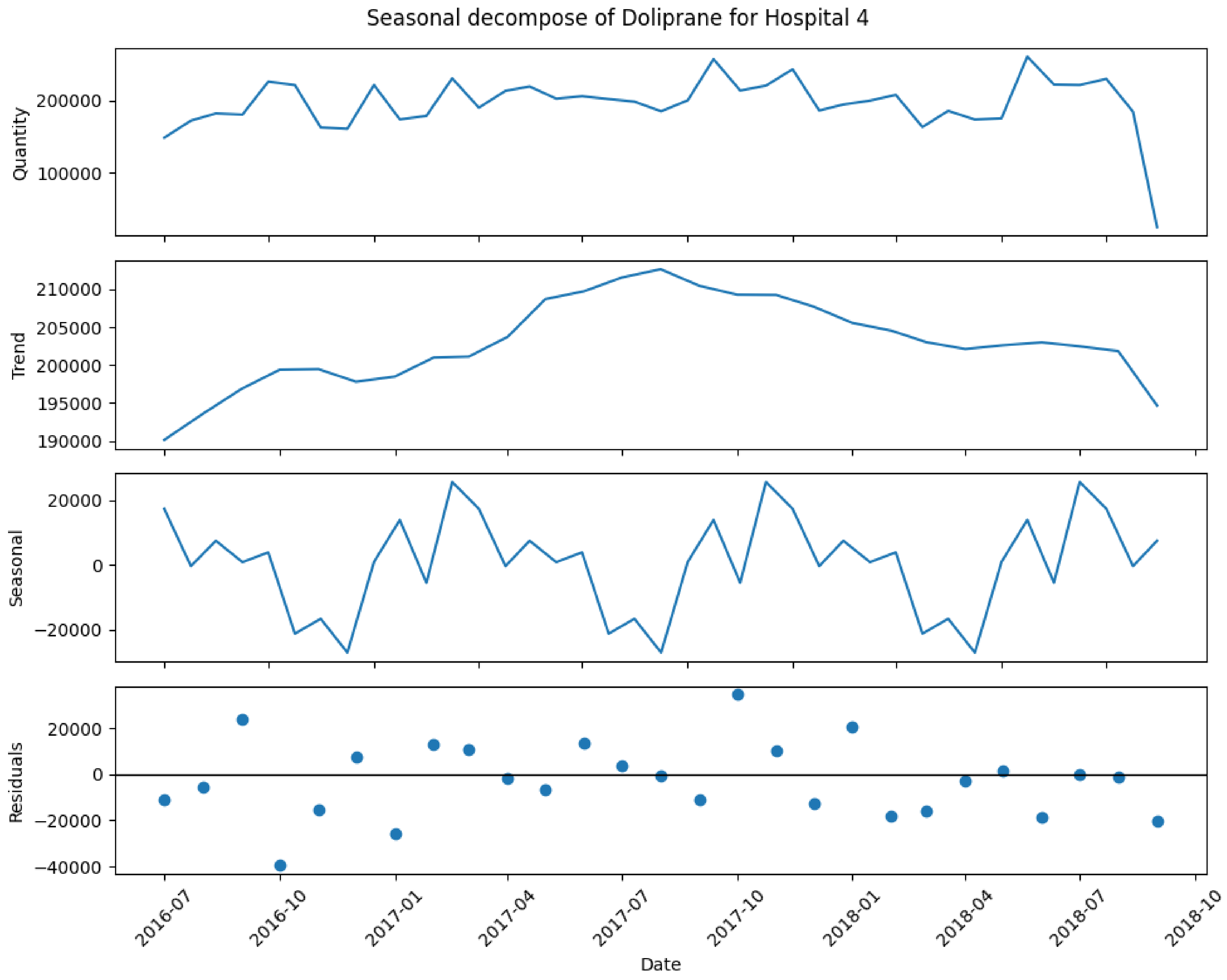
- Remove outliers and address missing data.
- Group data by month, year, UCD code, and hospital.
- Understand and manage the gaps for the different hospitals.
- For the time series approach, a common data period was chosen for all hospitals.



1: Context and Methodology

C: Feature manipulation and comprehension

- Seasonal decomposition.
- Calculation of the moving average.
- Principal Component Analysis (PCA) was tested but did not yield useful information.
- Trends in consumption, seasonality, and residuals were used as features to group different medicines.



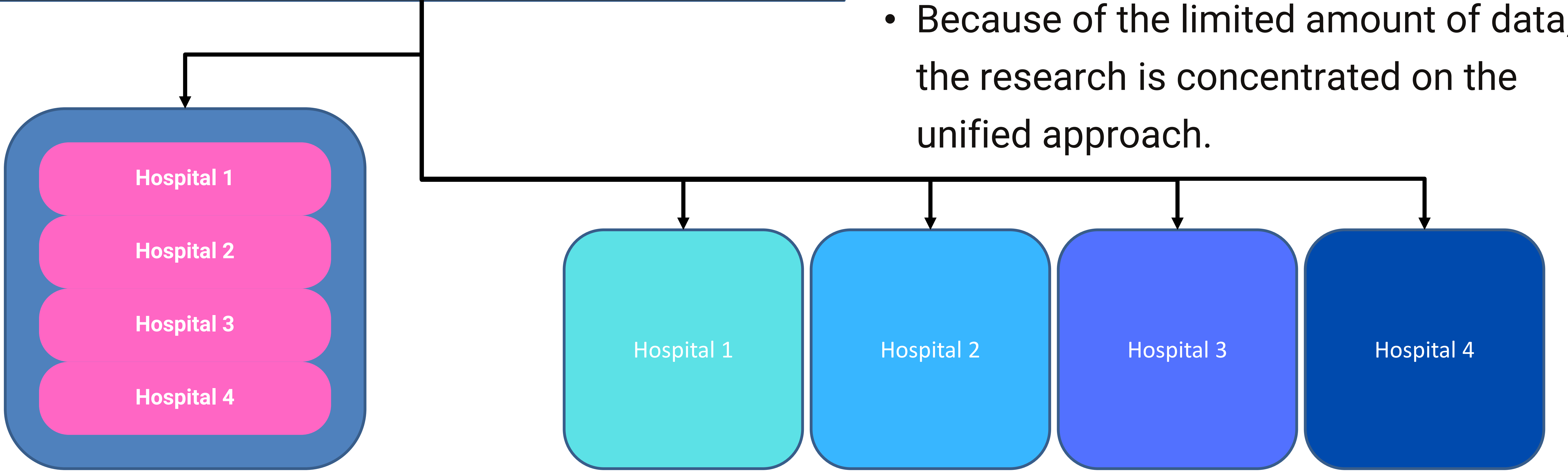
2: Clustering

- D: Clustering
  - k-means clustering
  - mini-batch k-means
  - agglomerative clustering
  - manual clustering

2: Clustering

ID_REF	ID_SITE_RATTACHE	CODE_ATC	HOSPI_CODE_UCD	DATE_MOUV	N_UFS	QUANTITY	WEEK	MONTH	YEAR	N_ETB	POPULATION	P_MEDICAL	PN_MEDICAL	LIT_HC	LIT_HP	SEJ_MCO	SEJ_HAD	SEJ_PSY	SEJ_SSR	SEJ_SLD
502829	HOSPI_3	C01CA03	3400892508566	2014-11-07	3	210.0	45.0	11	2014.0	50	1107398.0	1158	7129	2063.0	521.0	117781	594	2903	1302	97
502829	HOSPI_3	C01CA03	3400892508566	2015-01-22	4	340.0	4.0	1	2015.0	50	1120190.0	1239	7161	2053.0	493.0	118924	650	2878	1334	75
501463	HOSPI_3	J01CR05	3400893022634	2018-02-14	2	200.0	7.0	2	2018.0	50	1159220.0	1322	7439	2008.0	517.0	115376	1093	2481	1183	118
9220364	HOSPI_2	H02AB06	3400892203645	2017-06-30	4	55.0	26.0	6	2017.0	5	539067.0	714	5001	1157.0	187.0	75420	0	1236	261	0
9490	HOSPI_4	B01AC06	3400892065366	2018-03-26	1	630.0	13.0	3	2018.0	39	1859524.0	2627	15723	4477.0	486.0	255490	0	837	8416	209
890900	HOSPI_1	M03BX01	3400892697789	2015-10-06	1	20.0	41.0	10	2015.0	12	571879.0	684	5295	1411.0	94.0	74102	0	0	1140	57
9272958	HOSPI_2	N02AX02	3400892729589	2019-12-30	5	150.0	1.0	12	2019.0	5	542302.0	706	5013	1141.0	141.0	76593	0	1007	206	0
9387549	HOSPI_2	N02BE01	3400893875490	2018-05-21	2	30.0	21.0	5	2018.0	5	541454.0	703	5007	1159.0	139.0	74663	0	1193	237	0
503386	HOSPI_3	N02BE01	3400893875490	2015-03-18	18	410.0	12.0	3	2015.0	50	1120190.0	1239	7161	2053.0	493.0	118924	650	2878	1334	75
503129	HOSPI_3	N02BE01	3400891996128	2015-08-05	35	4350.0	32.0	8	2015.0	50	1120190.0	1239	7161	2053.0	493.0	118924	650	2878	1334	75

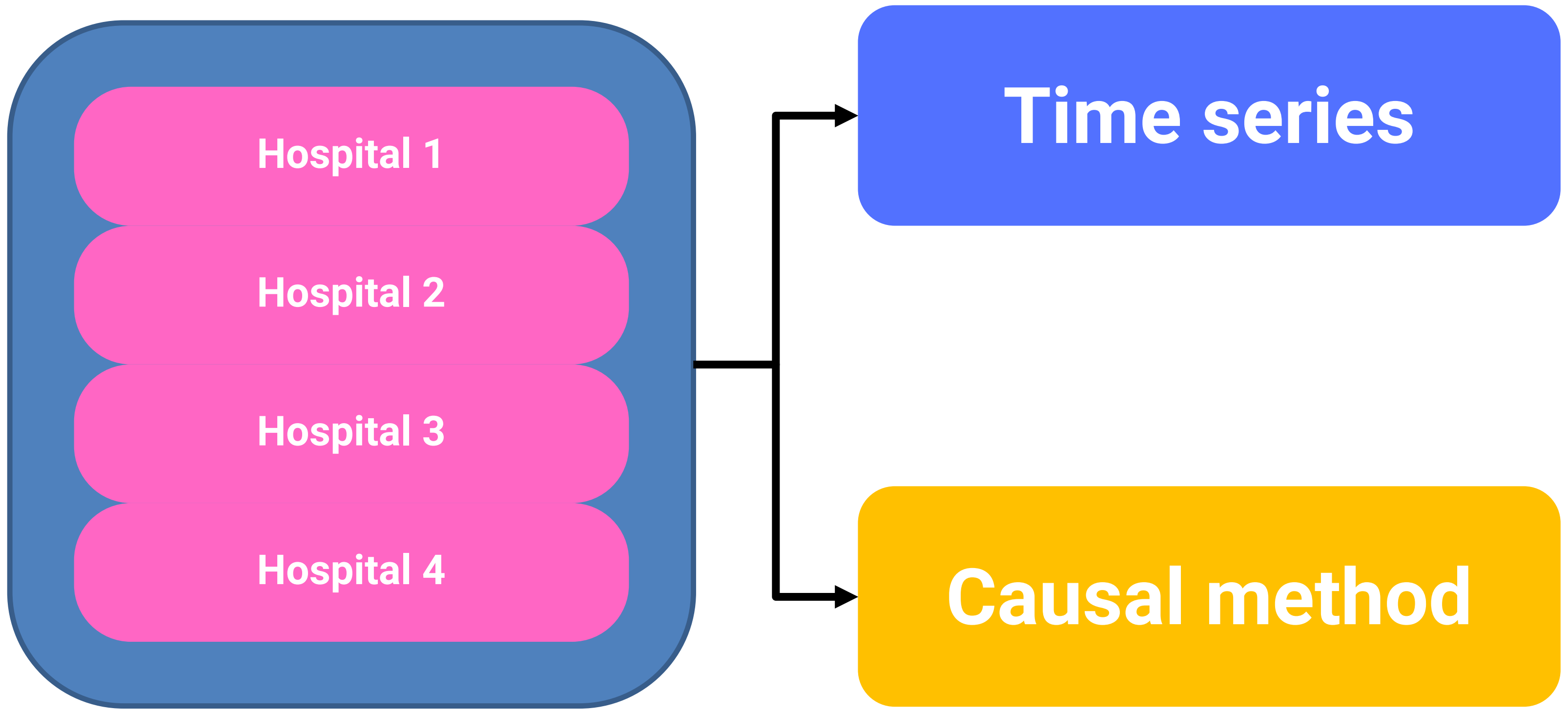
- The raw dataset contains information about various medicines and different hospitals.
- Because of the limited amount of data, the research is concentrated on the unified approach.



Unified approach

Division approach

2: Clustering



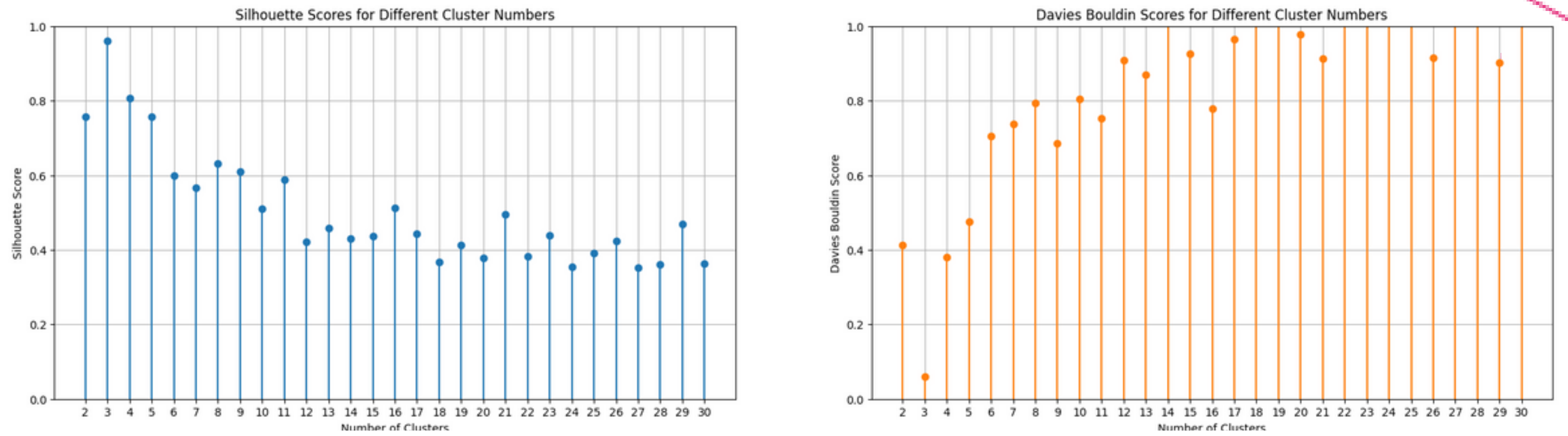
Unified approach

- Various time series are available for each medicine and hospital, taking into account the seasonal decomposition features (trend, seasonality, and residuals).
- All data from different periods, regardless of time factors and seasonality, were considered, taking into account the 21 features presented earlier.
- After grouping the data as presented in the data treatment step, there was an average of 50 data points for each combination of hospital and medicine. The minimum was 8 data points, and the maximum was 68.

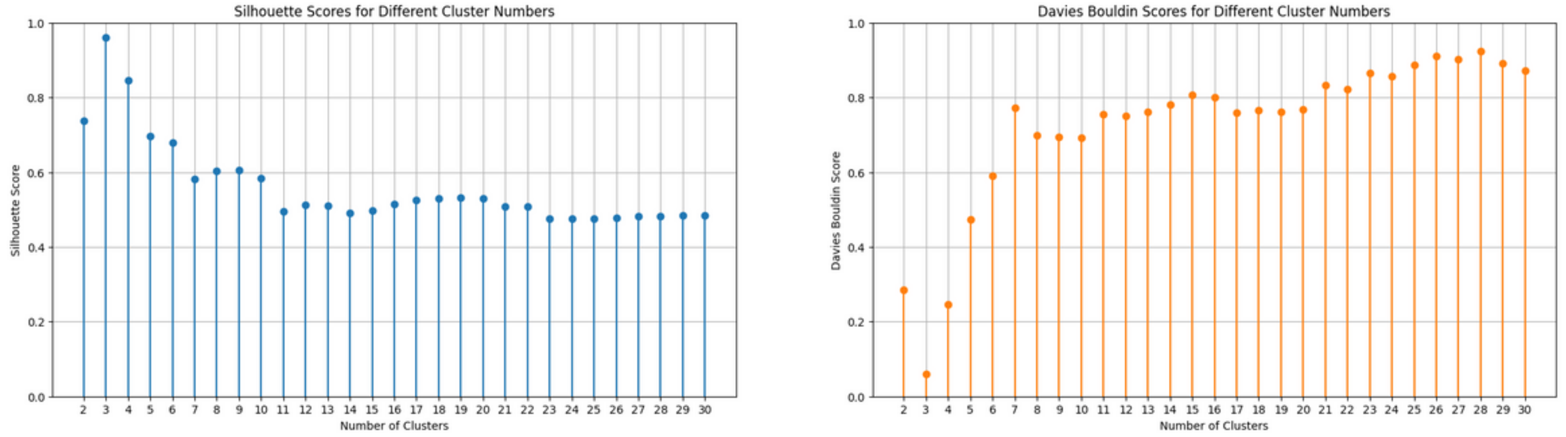
2: Clustering

Regarding the silhouette metric, all three different clustering methods yielded their maximum score when using **3 clusters.**

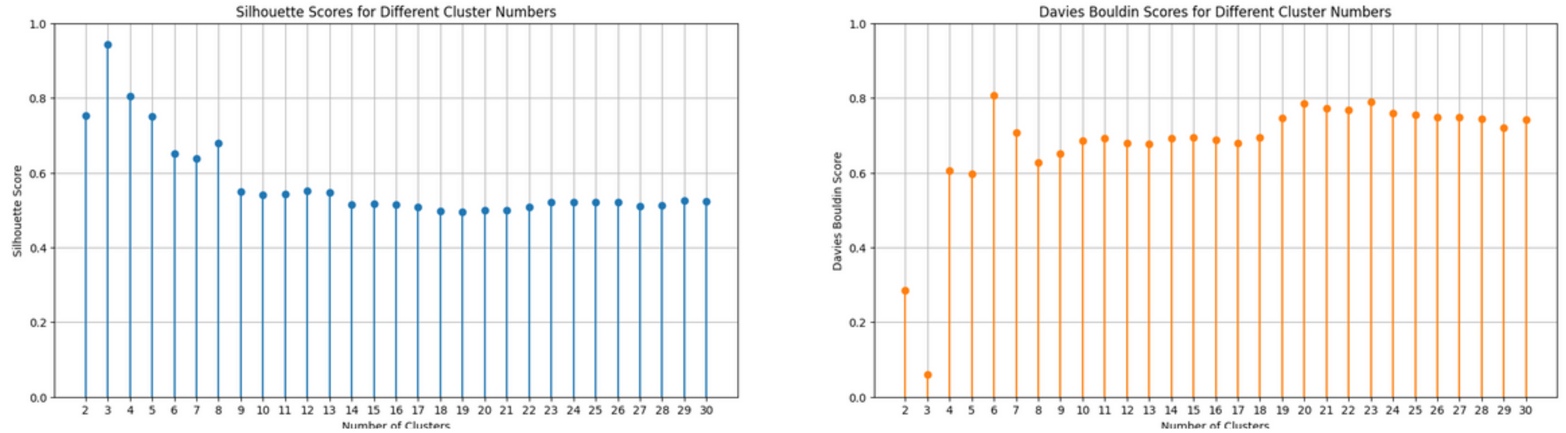
mini-batch k-means



k-means clustering



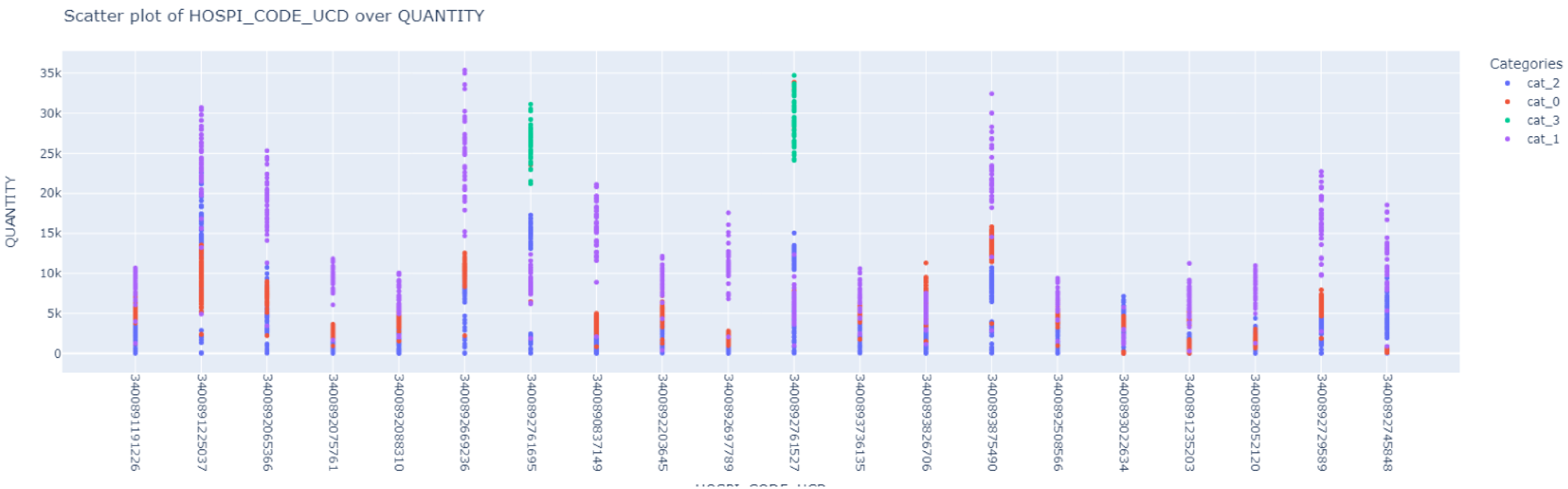
agglomerative clustering



G-SCOP SCIENCES POUR LA CONCEPTION, L'OPTIMISATION ET LA PRODUCTION

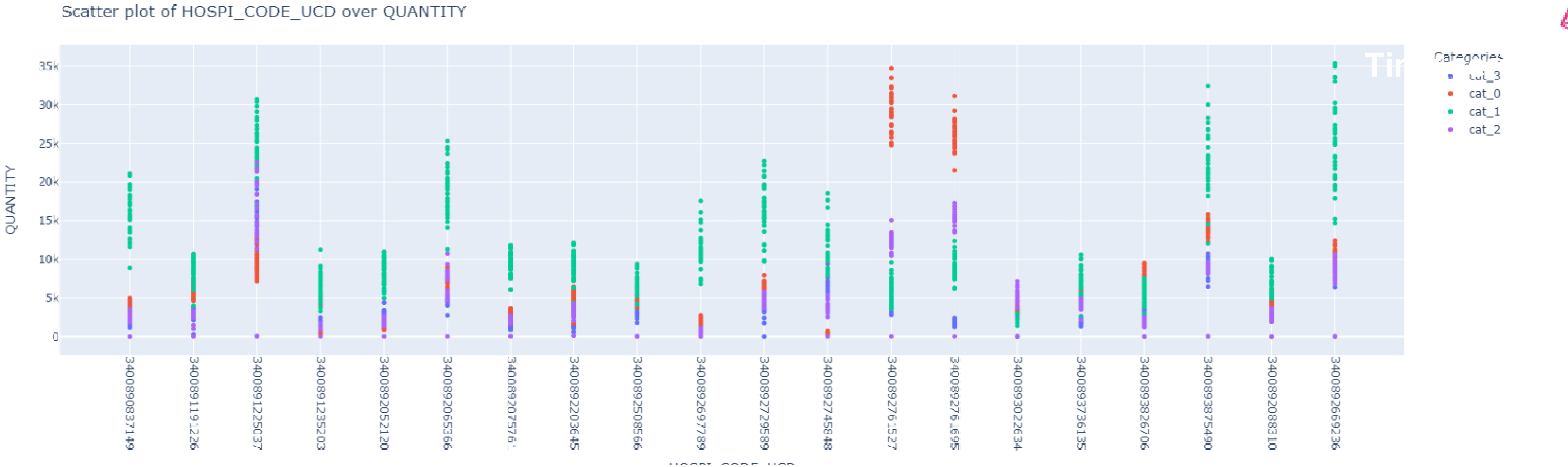
Grenoble INP

2: Clustering



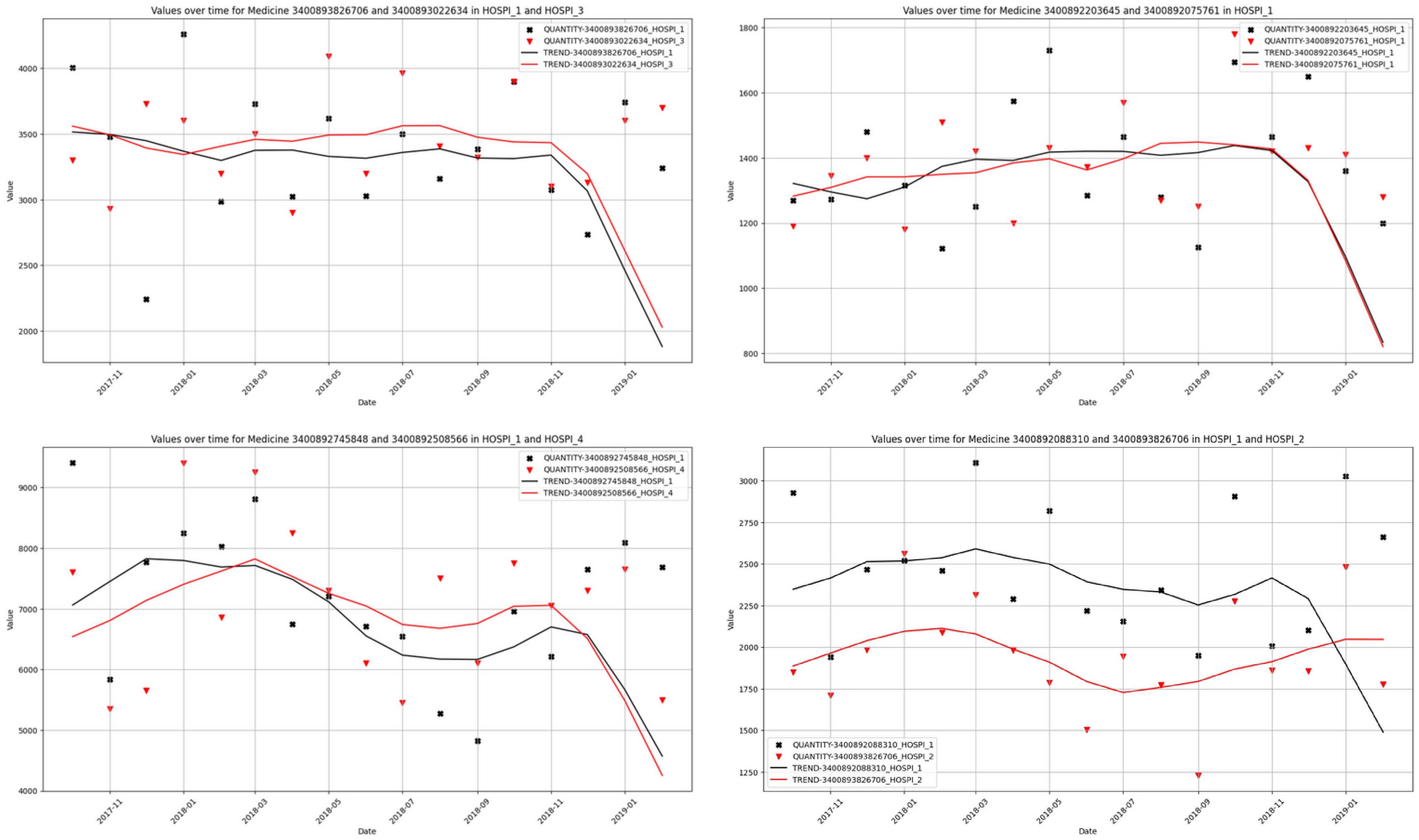
- All medicines were placed in separate clusters.
- There is an issue concerning the distribution of data points in each cluster.
- This problem becomes more pronounced in subsequent sections where there is less data available for training models.
- The quantity factor plays a predominant role in the clustering decisions.

2: Clustering



- The time series approach, which exclusively utilized features related to quantity, produced a comparable result to the clustering approach. Each medicine appears to be included in more than one cluster.
- However, a significant challenge remains, particularly for medicines with minimal data points. This limitation could have a negative impact on the forecasting models.

2: Clustering

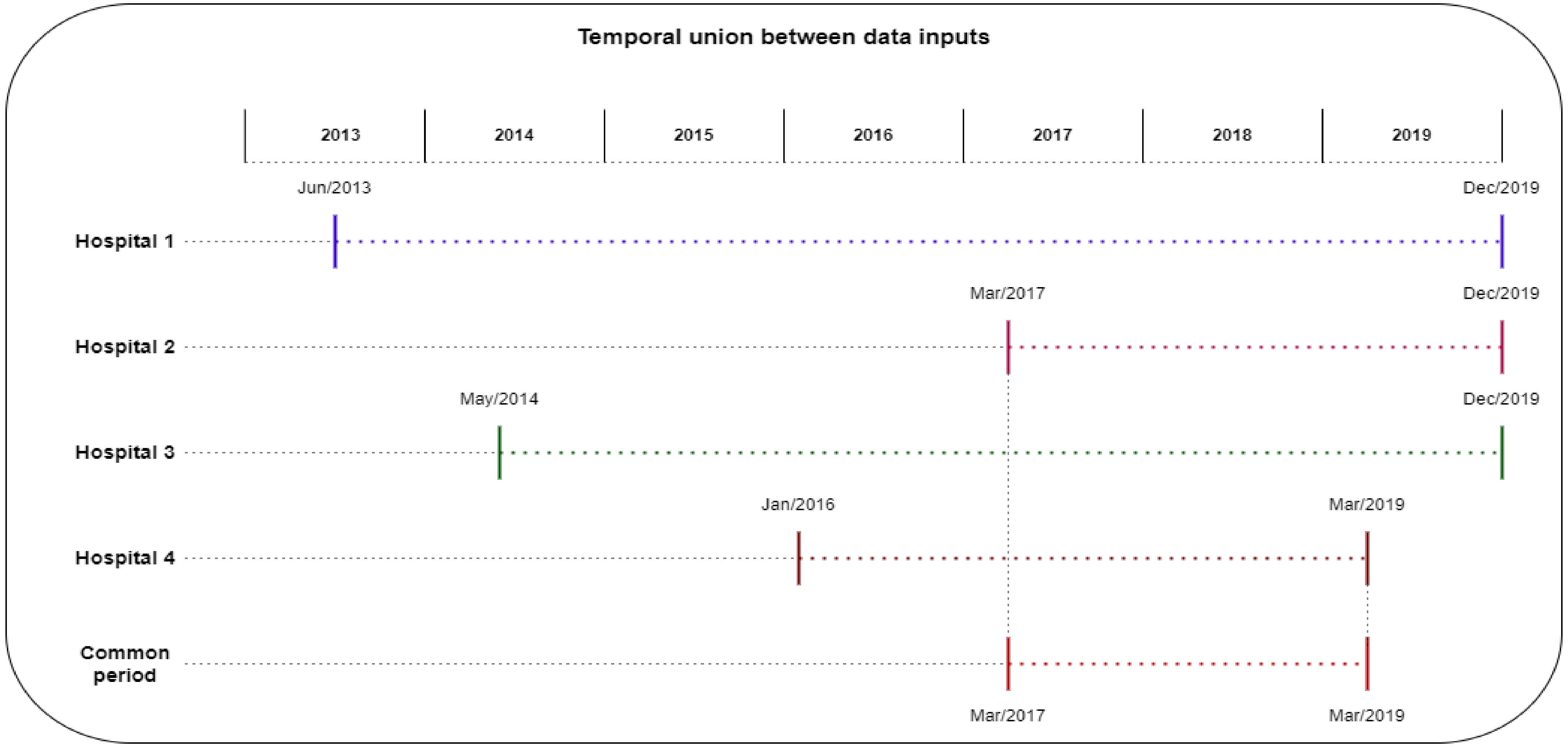


- Manual clustering was performed on pairs of medicines.
- The pairs consisted of combinations of hospitals and medicines.
- These pairs were grouped based on similarities in trend.

3: Forecasting

E: Forecasting → Random forest regressor

- Baseline
- Clusters with causal method
- Clusters with Time series
- Forecasting with Manual clusters



G-SCOP SCIENCES POUR LA CONCEPTION, L'OPTIMISATION ET LA PRODUCTION
Grenoble INP

3: Forecasting

- ***The primary objective of this research is to identify similarities among different medicines, ultimately reducing the number of models required for forecasting hospital demand, thus optimizing stock resources.***
- The research is divided into two main streams: the causal and time series approaches. Additionally, it is categorized into the unified and division approaches. However, the unified approach proves to be more suitable for addressing the current problem.
- Based on the work of D. KOALA, the random forest regressor outperforms other tested models in the current context.
- To establish performance benchmarks for the new models, a baseline must be calculated for the tested methodologies.

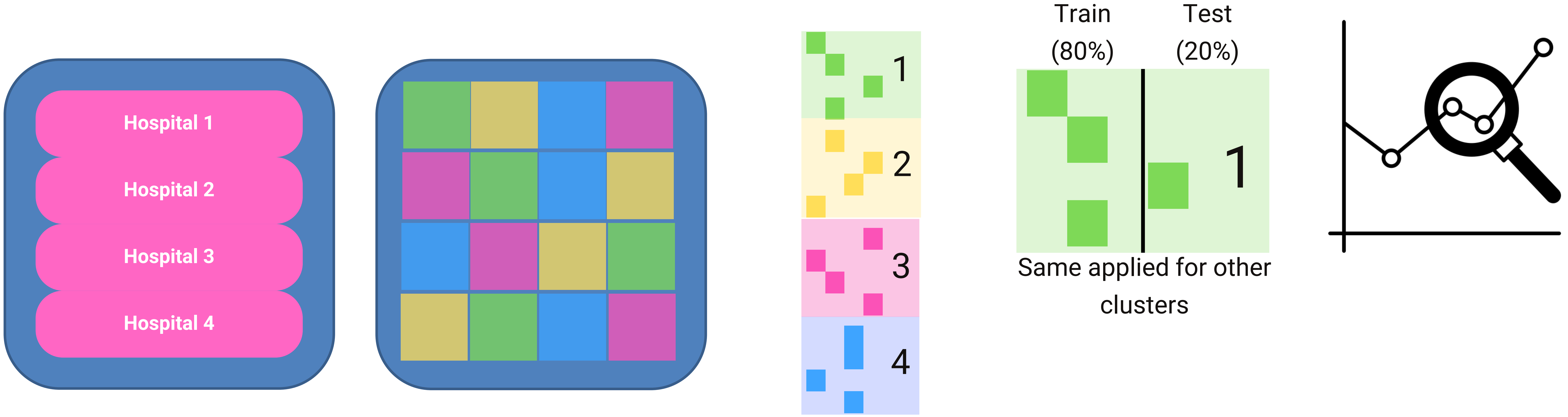
3: Forecasting

Id	CODE UCD	Unified dataset - Time series	Unified dataset - Causal
		MAPE	MAPE
1	CODE_UCD_3400892088310	0.10	2.59
2	CODE_UCD_3400892075761	0.11	0.64
3	CODE_UCD_3400892203645	0.24	1.35
4	CODE_UCD_3400892065366	0.15	0.29
5	CODE_UCD_3400892052120	0.16	5.87
6	CODE_UCD_3400891996128	28.48	28.48
7	CODE_UCD_3400893826706	0.22	74.65
8	CODE_UCD_3400893736135	0.14	0.48
9	CODE_UCD_3400893875490	0.07	11.84
10	CODE_UCD_3400890837149	55.67	0.58
11	CODE_UCD_3400891235203	0.18	2.41
12	CODE_UCD_3400891225037	6.75	3.34
13	CODE_UCD_3400891191226	0.24	0.39
14	CODE_UCD_3400892729589	15.88	11.19
15	CODE_UCD_3400892745848	0.20	6.99
16	CODE_UCD_3400892697789	0.26	1.54
17	CODE_UCD_3400892761527	0.25	7.02
18	CODE_UCD_3400893022634	12.41	26.69
19	CODE_UCD_3400892761695	0.39	1.82
20	CODE_UCD_3400892669236	0.12	0.11
21	CODE_UCD_3400892508566	3.20	0.39

- Considering the practical requirements of the current project, the Mean Absolute Percentage Error (MAPE) metric was chosen to assess the quality of each model.
- A separate model was developed for each medicine, utilizing data from all four hospitals.
- It is observed that the time series approach outperforms the causal approach in terms of model performance.
- These results can be attributed to variations in dataset composition between the two approaches and specific periods in the data that pose challenges for the models.

3: Forecasting

- To execute the training of our models, followed by the testing and validation phases, the following policy was implemented:



Unified approach | Clustering | Separate by cluster | Divide by train test | Predict and analyze by drug

3: Forecasting

Unified dataset - Causal Method							
ID	CODE UCD	CLUSTER	MAPE	ID	CODE UCD	CLUSTER	MAPE
	CODE_UCD_3400890837149	0	0.38		CODE_UCD_3400892669236	0	0.77
1	CODE_UCD_3400890837149	1	0.19	11	CODE_UCD_3400892669236	1	0.51
	CODE_UCD_3400890837149	2	1.81		CODE_UCD_3400892669236	2	54.27
	CODE_UCD_3400891191226	0	0.66		CODE_UCD_3400892697789	0	0.49
2	CODE_UCD_3400891191226	1	0.68	12	CODE_UCD_3400892697789	1	0.27
	CODE_UCD_3400891191226	2	0.22		CODE_UCD_3400892697789	2	13.61
	CODE_UCD_3400891225037	0	0.74		CODE_UCD_3400892729589	0	0.65
3	CODE_UCD_3400891225037	1	0.48	13	CODE_UCD_3400892729589	1	0.30
	CODE_UCD_3400891225037	2	11.33		CODE_UCD_3400892729589	2	0.47
	CODE_UCD_3400891235203	0	1.01		CODE_UCD_3400892745848	0	5.26
4	CODE_UCD_3400891235203	1	1.02	14	CODE_UCD_3400892745848	1	0.22
	CODE_UCD_3400891235203	2	161.15		CODE_UCD_3400892745848	2	0.51
	CODE_UCD_3400892052120	0	0.71		CODE_UCD_3400892761527	0	0.92
5	CODE_UCD_3400892052120	1	0.18	15	CODE_UCD_3400892761527	1	1.04
	CODE_UCD_3400892052120	2	9.63		CODE_UCD_3400892761527	2	1.17
	CODE_UCD_3400892065366	0	0.67		CODE_UCD_3400892761527	3	0.10
6	CODE_UCD_3400892065366	1	0.47		CODE_UCD_3400892761695	0	0.91
	CODE_UCD_3400892065366	2	0.39		CODE_UCD_3400892761695	1	0.56
	CODE_UCD_3400892075761	0	0.22	17	CODE_UCD_3400892761695	2	4.63
7	CODE_UCD_3400892075761	1	0.70		CODE_UCD_3400892761695	3	0.06
	CODE_UCD_3400892075761	2	8.30		CODE_UCD_3400893022634	0	13.12
	CODE_UCD_3400892088310	0	0.58	18	CODE_UCD_3400893022634	1	8.73
8	CODE_UCD_3400892088310	1	0.58		CODE_UCD_3400893022634	2	0.44
	CODE_UCD_3400892088310	2	30.55		CODE_UCD_3400893736135	0	0.59
	CODE_UCD_3400892203645	0	0.58	19	CODE_UCD_3400893736135	1	0.67
9	CODE_UCD_3400892203645	1	0.46		CODE_UCD_3400893736135	2	3.24
	CODE_UCD_3400892203645	2	1.51		CODE_UCD_3400893826706	0	0.64
	CODE_UCD_3400892508566	0	0.57	20	CODE_UCD_3400893826706	1	1.69
10	CODE_UCD_3400892508566	1	0.59		CODE_UCD_3400893826706	2	0.80
	CODE_UCD_3400892508566	2	23.18		CODE_UCD_3400893875490	0	0.84
				21	CODE_UCD_3400893875490	1	0.50
					CODE_UCD_3400893875490	2	0.68

- The utilization of clustering to group different medicines and reduce the number of models has negatively impacted the prediction quality in the causal method.
- This outcome can be attributed to the insufficient amount of data and the subdivision of already limited data into even smaller segments.
- Some clusters exhibit favorable outcomes, but it's notable that different medicines within the same cluster do not yield positive results, exemplified by **Cluster 1**.

3: Forecasting

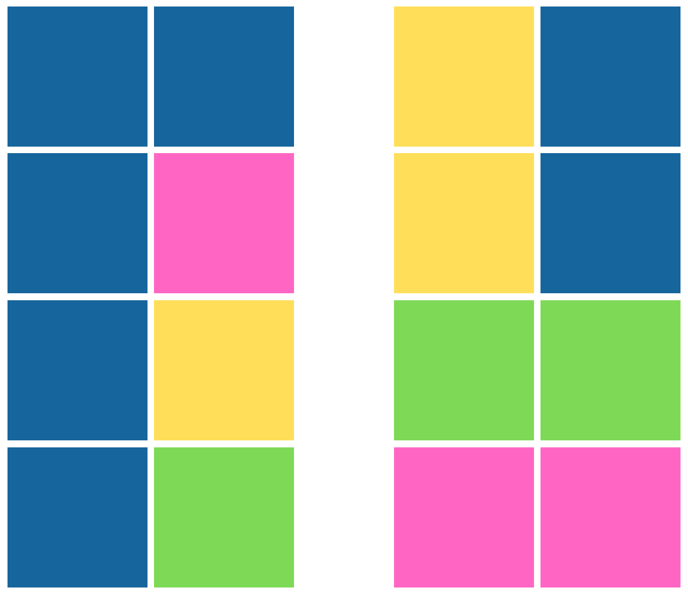
Unified dataset - Time series - 4 clusters							
ID	HOSPI_CODE_UCD	CLUSTER	MAPE	ID	HOSPI_CODE_UCD	CLUSTER	MAPE
1	CODE_UCD_3400890837149	0	0.20	11	CODE_UCD_3400892669236	0	0.62
	CODE_UCD_3400890837149	1	0.32		CODE_UCD_3400892669236	1	0.49
	CODE_UCD_3400890837149	2	28.41		CODE_UCD_3400892669236	2	8.17
	CODE_UCD_3400890837149	3	0.74		CODE_UCD_3400892669236	3	0.86
2	CODE_UCD_3400891191226	0	0.28	12	CODE_UCD_3400892697789	0	1.32
	CODE_UCD_3400891191226	1	0.52		CODE_UCD_3400892697789	1	0.26
	CODE_UCD_3400891191226	2	0.30		CODE_UCD_3400892697789	2	1.63
	CODE_UCD_3400891191226	3	0.36		CODE_UCD_3400892697789	3	0.91
3	CODE_UCD_3400891225037	0	0.61	13	CODE_UCD_3400892729589	0	0.36
	CODE_UCD_3400891225037	1	0.51		CODE_UCD_3400892729589	1	0.28
	CODE_UCD_3400891225037	2	0.88		CODE_UCD_3400892729589	2	0.63
	CODE_UCD_3400891225037	3	0.87		CODE_UCD_3400892729589	3	22.21
4	CODE_UCD_3400891235203	0	13.37	14	CODE_UCD_3400892745848	0	8.06
	CODE_UCD_3400891235203	1	0.92		CODE_UCD_3400892745848	1	0.16
	CODE_UCD_3400891235203	2	0.31		CODE_UCD_3400892745848	2	17.43
	CODE_UCD_3400891235203	3	1.09		CODE_UCD_3400892745848	3	0.76
5	CODE_UCD_3400892052120	0	1.63	15	CODE_UCD_3400892761527	0	0.86
	CODE_UCD_3400892052120	1	0.52		CODE_UCD_3400892761527	1	0.93
	CODE_UCD_3400892052120	2	0.13		CODE_UCD_3400892761527	2	0.85
	CODE_UCD_3400892052120	3	0.32		CODE_UCD_3400892761527	3	0.60
6	CODE_UCD_3400892065366	0	0.47	16	CODE_UCD_3400892761695	0	0.84
	CODE_UCD_3400892065366	1	0.37		CODE_UCD_3400892761695	1	0.17
	CODE_UCD_3400892065366	2	0.68		CODE_UCD_3400892761695	2	0.90
	CODE_UCD_3400892065366	3	0.63		CODE_UCD_3400892761695	3	0.13
7	CODE_UCD_3400892075761	0	0.58	17	CODE_UCD_3400893022634	0	0.19
	CODE_UCD_3400892075761	1	0.20		CODE_UCD_3400893022634	1	0.79
	CODE_UCD_3400892075761	2	0.20		CODE_UCD_3400893022634	2	0.68
	CODE_UCD_3400892075761	3	0.49		CODE_UCD_3400893022634	3	0.48
8	CODE_UCD_3400892088310	0	0.19	18	CODE_UCD_3400893736135	0	0.20
	CODE_UCD_3400892088310	1	0.43		CODE_UCD_3400893736135	1	0.51
	CODE_UCD_3400892088310	2	0.43		CODE_UCD_3400893736135	2	0.59
	CODE_UCD_3400892088310	3	0.39		CODE_UCD_3400893736135	3	0.15
9	CODE_UCD_3400892203645	0	0.26	19	CODE_UCD_3400893826706	0	0.46
	CODE_UCD_3400892203645	1	0.17		CODE_UCD_3400893826706	1	1.14
	CODE_UCD_3400892203645	2	0.48		CODE_UCD_3400893826706	2	7.80
	CODE_UCD_3400892203645	3	0.20		CODE_UCD_3400893826706	3	0.44
10	CODE_UCD_3400892508566	0	0.09	20	CODE_UCD_3400893875490	0	0.71
	CODE_UCD_3400892508566	1	0.35		CODE_UCD_3400893875490	1	0.51
	CODE_UCD_3400892508566	2	40.88		CODE_UCD_3400893875490	2	0.79
	CODE_UCD_3400892508566	3	0.48		CODE_UCD_3400893875490	3	0.83

- Employing clustering on the time series dataset for forecasting did not yield the anticipated improvement in MAPE scores when compared to the baseline.
- While the scores obtained through this approach are superior to those of the causal method, they remain unacceptable due to the unique characteristics of the problem.

SCIENCES POUR LA CONCEPTION, L'OPTIMISATION ET LA PRODUCTION

3: Forecasting

- In general, the results using clustering techniques are not good and can't replace the techniques of forecasting by medicine
- The main hypothesis for the poor results is the lack of data present in the dataset.
- Taking, for example, the data points presented in the previous sections where there are 8 registered points that are not zero, if the medicine is split into 4 clusters, the best scenario would be 2 data points for each cluster, the worst, one cluster with 5 data points, and the others with 1, making it impossible to conduct effective training and evaluation tests with the forecasting models.
- Taking this into consideration, one way to solve this problem would be to identify similarities in the entire time series, where we could create pairs of medicines and test them together.



Worst Best

Both are not enough

3: Forecasting

E: Forecasting → Random forest regressor → Baseline / Clusters with causal method / Clusters with Time series / **Forecasting with Manual clusters**

- To recap, we explored various approaches to forecast the demand for medicines in hospitals.
- The most recent attempt involves identifying similarities between different time series, each composed of a combination of a medicine and a hospital, to predict the demand for each medicine in these pairs.
- As a result of manual clustering, specific pairs were chosen for testing. Data points were split into training and testing sets, and each medicine in the pair was evaluated individually.

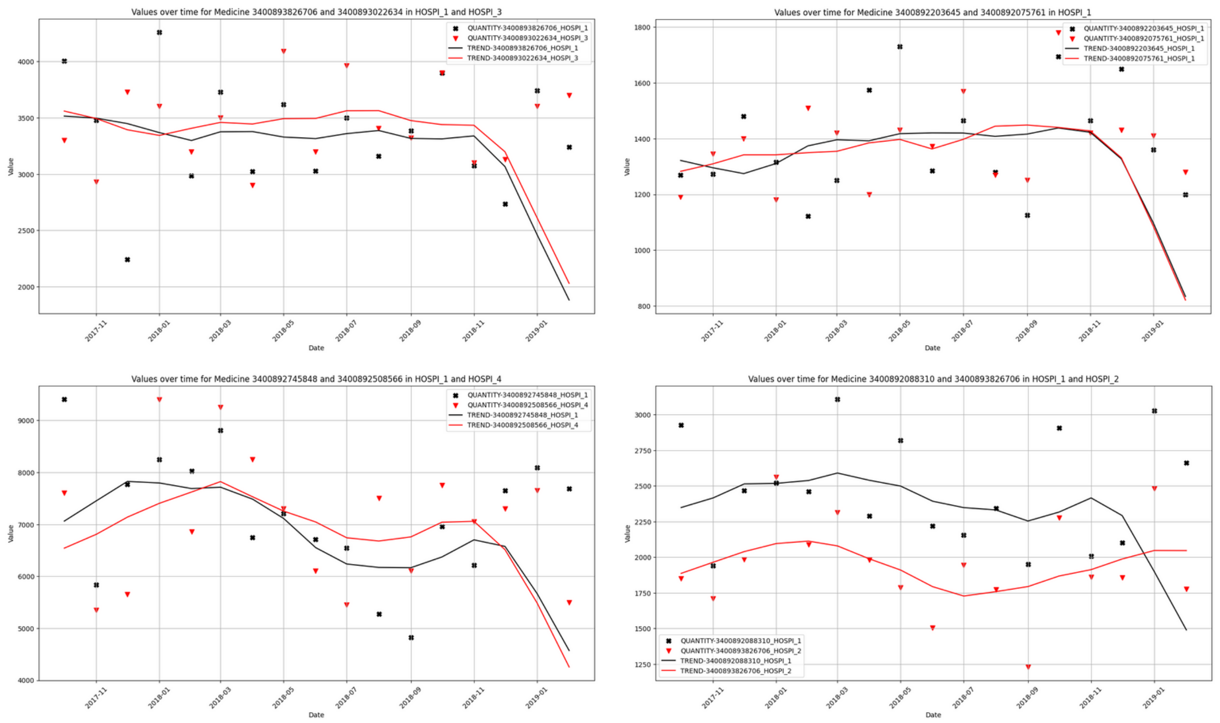
3: Forecasting

E: Forecasting → Random forest regressor → Baseline

→ Clusters with causal method

→ Clusters with Time series

→ **Forecasting with Manual clusters**



- The selected pairs to test in the forecasting part represent 10% (206) of the total number of possibilities (2310).

G-SCOP SCIENCES POUR LA CONCEPTION, L'OPTIMISATION ET LA PRODUCTION

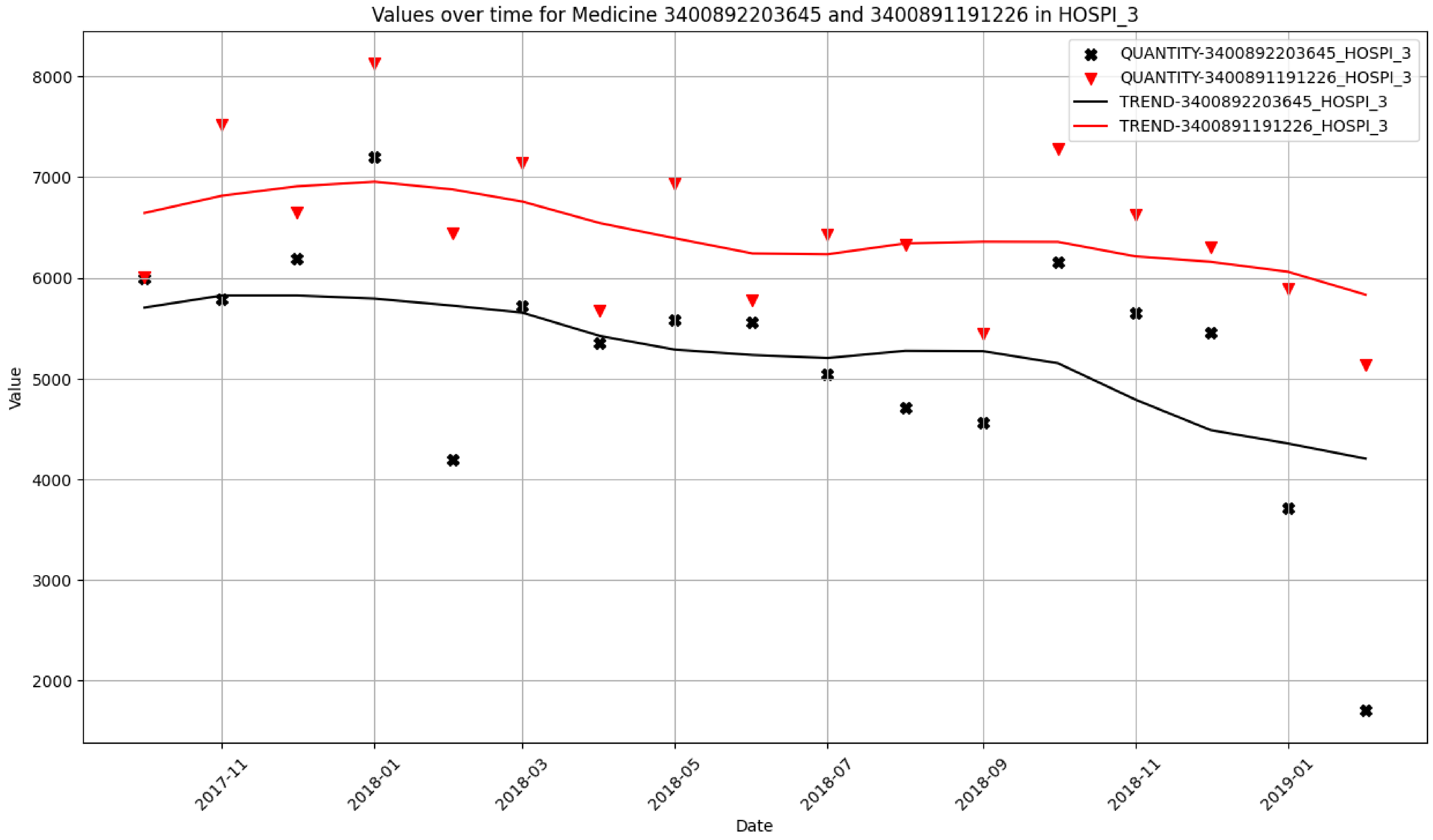
Grenoble INP

3: Forecasting

ID	PAIR	CODE UCD	MAPE
1	65366_91226_1_3	CODE_UCD_3400892065366	7%
	65366_91226_1_3	CODE_UCD_3400891191226	3%
2	03645_91226_3_3	CODE_UCD_3400892203645	2%
	03645_91226_3_3	CODE_UCD_3400891191226	6%
3	03645_91226_4_4	CODE_UCD_3400892203645	9%
	03645_91226_4_4	CODE_UCD_3400891191226	7%
4	03645_75761_1_1	CODE_UCD_3400892203645	5%
	03645_75761_1_1	CODE_UCD_3400892075761	8%
5	69236_75761_1_4	CODE_UCD_3400892669236	6%
	69236_75761_1_4	CODE_UCD_3400892075761	10%
6	69236_75761_2_4	CODE_UCD_3400892669236	5%
	69236_75761_2_4	CODE_UCD_3400892075761	9%
7	69236_75761_3_4	CODE_UCD_3400892669236	7%
	69236_75761_3_4	CODE_UCD_3400892075761	10%
8	69236_75490_2_2	CODE_UCD_3400892669236	6%
	69236_75490_2_2	CODE_UCD_3400893875490	6%
9	29589_35203_1_4	CODE_UCD_3400892729589	7%
	29589_35203_1_4	CODE_UCD_3400891235203	8%
10	29589_03645_1_4	CODE_UCD_3400892729589	7%
	29589_03645_1_4	CODE_UCD_3400892203645	8%

- The 10 pairs with the best metrics were selected to illustrate the quality of the forecasting models.
- It is evident that some combinations of medicines and hospitals perform better than others.
- Some medicines, in different combinations, also perform well, indicating that clusters with more than 2 medicines could be effective.
- After testing with 2 medicines, we also tested with 3 or more medicines, but the positive results did not repeat as expected.
- A better technique for clustering time series could be applied, such as DTW (Dynamic Time Warping).

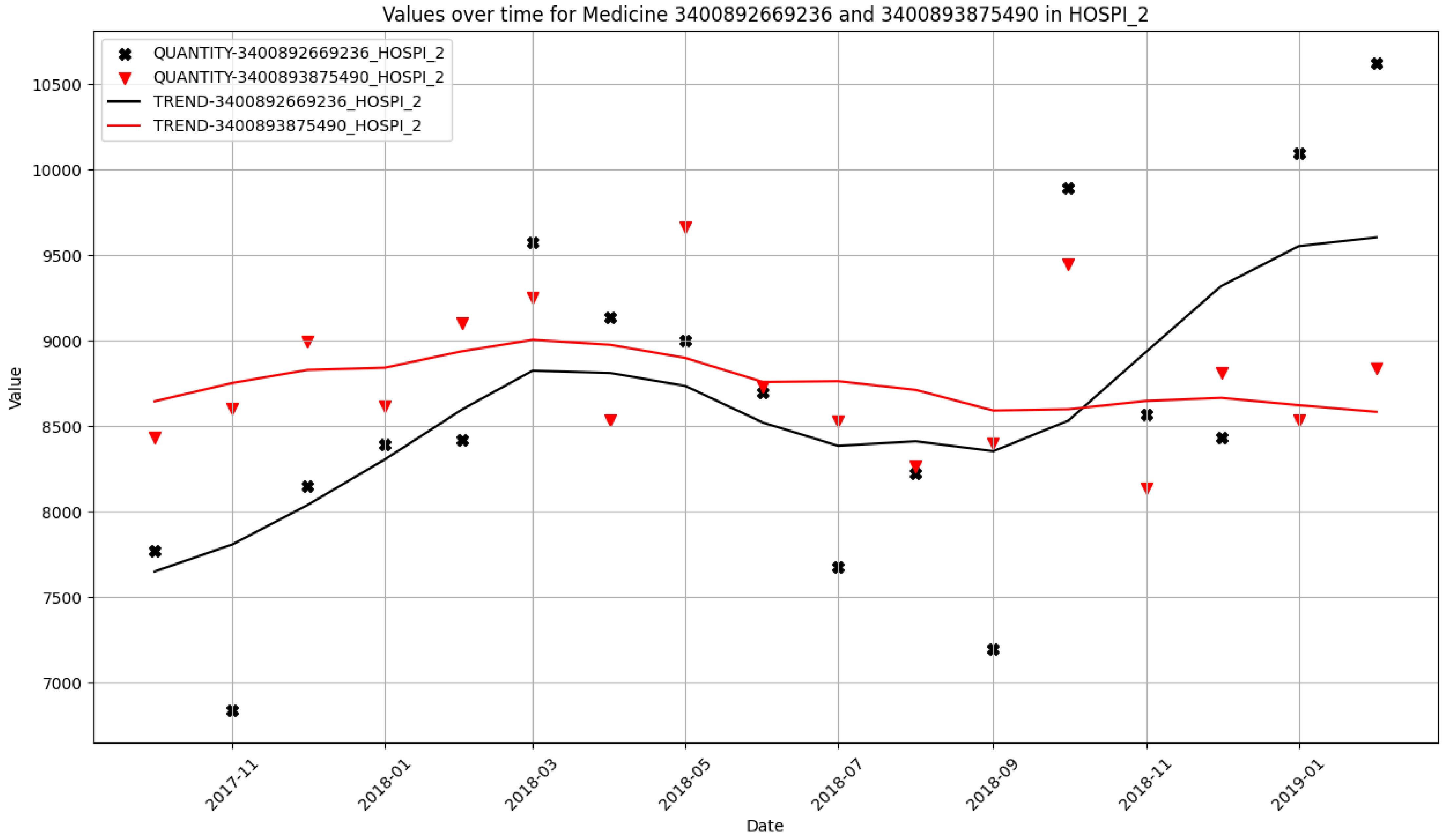
3: Forecasting



ID	PAIR	CODE UCD	MAPE
2	03645_91226_3_3	CODE_UCD_3400892203645	2%
	03645_91226_3_3	CODE_UCD_3400891191226	6%

- Even with different magnitudes, similarities were still present, which resulted in favorable forecasting evaluations.

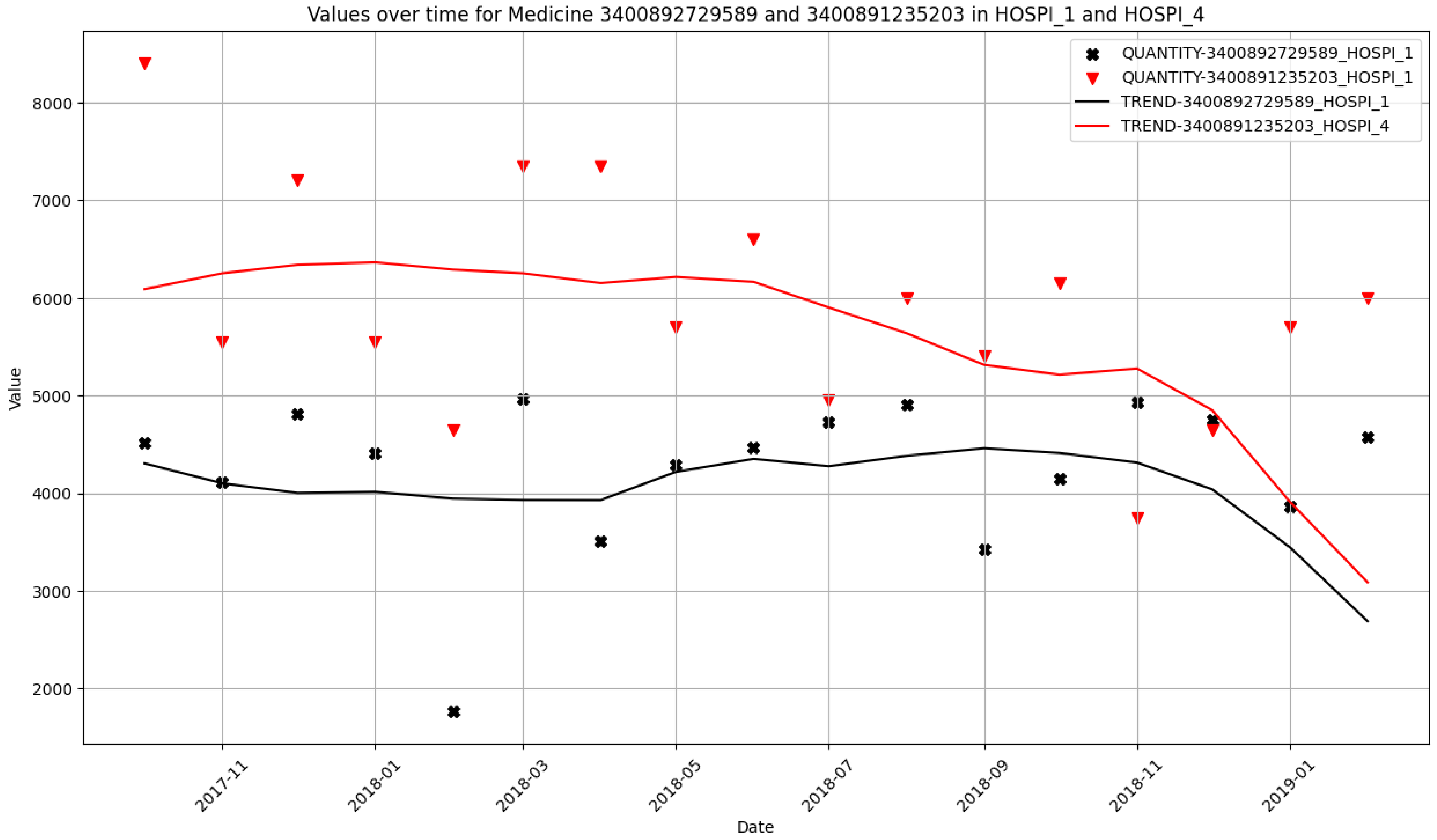
3: Forecasting



ID	PAIR	CODE UCD	MAPE
8	69236_75490_2_2	CODE_UCD_3400892669236	6%
	69236_75490_2_2	CODE_UCD_3400893875490	6%

- The trends and magnitudes of both series are similar, differing only at the end. However, these similarities were sufficient to produce a prediction with less than a 10% error rate.

3: Forecasting



ID	PAIR	CODE UCD	MAPE
9	29589_35203_1_4	CODE_UCD_3400892729589	7%
	29589_35203_1_4	CODE_UCD_3400891235203	8%

- Continuing from the previously observed patterns, this pair exhibits an error rate of less than 10%. Despite differences in magnitude, it produced favorable results in terms of MAPE and outperformed the baseline and other clustering techniques.

4: Conclusion

- Conventional clustering methods failed to produce satisfactory MAPE metric scores in forecasting tests.
- Discrepancies in the data across various hospitals complicated the development of robust models and accurate predictions.
- Manual clustering, however, led to favorable forecasting scores.
- Certain medicines achieved predictions with errors below 10% in manual clustering.
- In manual clustering, specific combinations of medicines and hospitals outperformed the baseline and other clustering techniques.

4: Conclusion

- It is advisable to explore advanced time series clustering techniques, like Dynamic Time Warping.
- Gathering additional data is essential to determine whether the subpar performance of clustering techniques can be mitigated with larger datasets.
- Addressing data gaps can be achieved through established methods found in the literature, including extrapolation, SARIMA, or other machine learning approaches.
- The current findings should undergo testing through simulations using synthetic data to emulate a new hospital scenario.

References

François Hada*. Les sources d'informations et de données sur le médicament. Revue française des affaires sociales, 1(3) :087–098, 2007.

Denis Koala, Zakaria Yahouni, Gülgün Alpan, and Yannick Frein. Factors influencing drug consumption and prediction methods, 2021

Denis Koala, Zakaria Yahouni, Gülgün Alpan, and Djamal Si Mohand. Correlation analysis of factors impacting health product consumption in French hospitals. IFAC-PapersOnLine, 55(10) :895–900, 2022. ISSN 24058963. doi : 10.1016/j.ifacol.2022.09.415.

Denis Koala, Zakaria Yahouni, Gülgün Alpan, and Djamal S I Mohand. Machine learning versus consumption-based models for predicting medicine demand in french hospitals : a case study. -, 2023

Stanley Suan You Lim, Siao-Leu Phouratsamay, Zakaria Yahouni, and Eric Gascard. Medicine consumption demand forecasting in french hospitals using SARIMA model. Master's thesis, Grenoble INP, 2023.

Wes McKinney, Josef Perktold, and Skipper Seabold. Time series analysis in python with statsmodels. Jarrodmillman Com, pages 96–102, 2011

A.C. Muller and S. Guido. Introduction to Machine Learning with Python : A Guide for Data Scientists. O'Reilly Media, Incorporated, 2018. ISBN 9789352134571. URL https://books.google.fr/books?id=jGdXswEACAAJ.

Meinard Müller. Dynamic time warping. Information retrieval for music and motion, pages 69–84, 2007

References

Kevin P. Murphy. Probabilistic Machine Learning : An Introduction. MIT Press, 2022. URL probml.ai.

Tim Oates, Laura Firoiu, and Paul R Cohen. Clustering time series with hidden markov models and dynamic time warping. In Proceedings of the IJCAI-99 workshop on neural, symbolic and reinforcement learning methods for sequence learning, pages 17–21. Citeseer, 1999

OECD. Health at a Glance 2013 : OECD Indicators. OECD Publishing, 2013. URL http://dx.doi.org/10.1787/health_glance-2013-en. DOI : 10.1787/health_glance-2013-en.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn : Machine learning in Python. Journal of Machine Learning Research, 12 : 2825–2830, 2011.

Skipper Seabold and Josef Perktold. statsmodels : Econometric and statistical modeling with python. In 9th Python in Science Conference, 2010.

The pandas development team. pandas-dev/pandas : Pandas. Zenodo, Feb 2020. URL https://doi.org/10.5281/zenodo.3509134

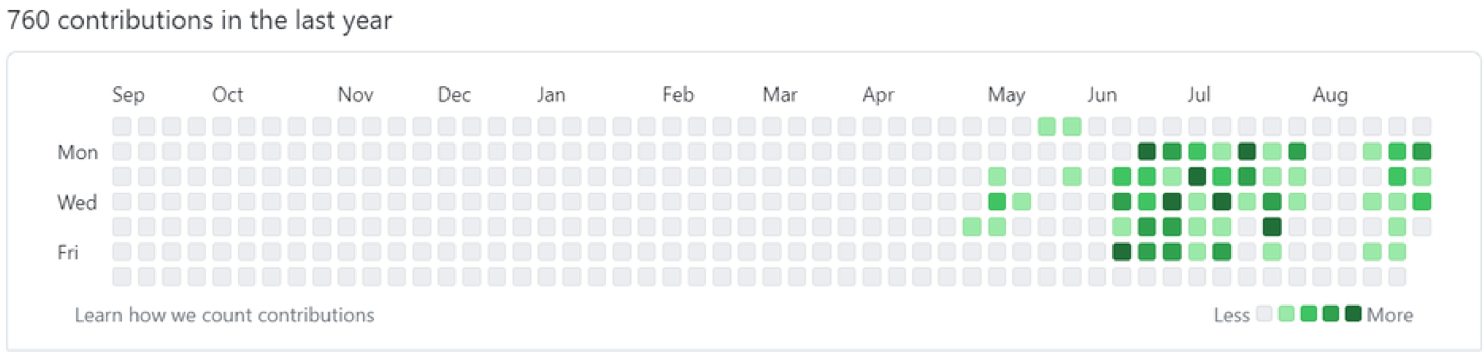
Daniel Vélez, Siao-Leu Phouratsamay, Zakaria Yahouni, and Gülgün Alpan. Predicting medicine demand fluctuations through markov chain. In 12th International Workshop on Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future, pages 329–340, Bucharest, Romania, September 2022. doi : 10.1007/978-3-031-24291-5_26. URL https://hal.archives-ouvertes.fr/hal-03951534.33

G-SCOP SCIENCES POUR LA CONCEPTION, L'OPTIMISATION ET LA PRODUCTION

Grenoble INP

Contact info

Douglas Mateus Machado
douglasmmachado



SCAN ME

douglas.mateus-machado@grenoble-inp.org

+33 07 49 78 00 55

[douglas-machado]

[douglasmmachado]

Thanks

Gülgün ALPAN | Zakaria YAHOUNI | Denis KOALA