



# Rapport TD6 2022

Abdourahmane NDIAYE  
Quentin MONTALAND  
Nathalie TANG  
Clara MORAIS

— université  
— lumière  
— LYON 2

INSTITUT  
de la  
communication

# Présentation des résultats

Question 1 - Importer les données et familiarisez vous avec les données

Lors de la première analyse des fichiers Hospi\_2017.2018.2019, nous avons tout de suite pensé qu'il serait intéressant de réunir chaque feuille et chaque fichier afin d'obtenir un seul et unique jeu de données sur lequel travailler. Avant d'entamer ces transformations, nous avons en premier lieu créé des librairies SAS pour stocker les fichiers et améliorer l'organisation du projet.

Question 2 - Décrivez les données (analyse, problèmes de qualité, outliers)

Concernant l'amélioration de la qualité des fichiers, nous avons commencé par pivoter les colonnes "Indicateur" et "Valeur" de la feuille Lits et Places de chaque fichier. En effet, cette étape sera nécessaire pour faire la jointure sur le champ "finess" entre chaque feuille.

	finess	Indicateur	Valeur
1	010007300	CI_AC1	0
2	010007300	CI_AC2	
3	010007300	CI_AC3	
4	010007300	CI_AC4	
5	010007300	CI_AC5	0
6	010007300	CI_AC6	0
7	010007300	CI_AC7	2
8	010007300	CI_AC8	0
9	010007300	CI_AC9	0
10	010007987	CI_AC1	30
11	010007987	CI_AC2	
12	010007987	CI_AC3	



	finess	CI_AC1	CI_AC2	CI_AC3	CI_AC4	CI_AC5	CI_AC6	CI_AC7	CI_AC8	CI_AC9
1	010007300	0				0	0	2	0	0
2	010007987	30				0				
3	010008407	68		6		7	26	7	21	1
4	010009132	20								
5	010780054	277	12	10	12	14	73	20	43	4
6	010780062	66		4		4	17	5	10	0
7	010780096	59								

Données pivotées

Ensuite, nous avons réunis pour chaque fichiers toutes les feuilles en une et ajouté une colonne "year" qui permettra de distinguer les fichiers. Puis, grâce à l'instruction "set" d'une procédure DATA, on obtient un dataset avec la forme suivante :

	finess	CI_AC1	CI_AC2	cat	year
1	010007300	0		CLI	2019
2	010007300	0		CLI	2017
3	010007300	0		CLI	2018
4	010007987	30		CH	2017
5	010007987	24		CH	2019
6	010007987	30		CH	2018
7	010008407	68		CH	2017

Forme finale du dataset

En terme de qualité de données, nous avons constaté que des valeurs indéfinies, mais que l'on pourrait considérées comme nulles faisaient leurs apparitions (cf. capture d'écran ci-dessous). Pour remédier à ce problème, nous avons simplement procédé à un changement de type pour nos données (cf. Bibliographie).

P6	P7	P8	P9
.Z	.Z		219163
4,5	30		.c
15,6	16,2	39914	252807
	26,5		.c
14,6	15,2	35119	272702
14,1	16,9		213991
1,1	25,3		.c
			.c
	16,8		.c

Données mal formatées

La procédure MEANS			
Variable	Libellé	N	Nbre manquant
year	year	4509	0
A7	A7	3943	566
A8	A8	3943	566
A9	A9	3943	566
A10	A10	521	3988
A11	A11	496	4013
A12	A12	3853	656
A13	A13	3568	941
A14	A14	2499	2010
A15	A15	1464	3045
CI_A1	CI_A1	4429	80
CI_A2	CI_A2	4429	80

Données manquantes (une fois les formats corrigés)

Question 3 - Etablissements qui ont disparu ou qui sont apparus au cours des années

Aucun établissement n'est apparu ou n'a disparu durant les trois années.

Question 4 - Etablissement qui ont changé de taille ou d'activité au cours des années

finess	distinct_MCO	distinct_M	distinct_C	distinct_O	max_code_différent
570000083	1	1	1	1	1
840017461	1	0	0	0	1
300780152	1	1	1	0	1
620103440	1	1	1	1	1
760780205	1	1	1	0	1
460780208	1	1	1	0	1

Aucun établissement n'a changé de taille au cours des trois années.  
(une valeur maximum trouvée pour chaque variable de taille).

finess	year	total_activite
030781116	2017	40979
030781116	2018	38969
030781116	2019	36453

Néanmoins, on peut constater quelques fluctuations au niveau des activités. On remarque que le total des variables concernant l'activité globale des hôpitaux (CI\_A1 à CI\_A15) varie en fonction des années.

## Question 5 - Déterminer les seuils utilisés pour les variables taille\_M, taille\_C, taille\_O

taille_M	min_lits	max_lits
	31	35
M0	0	2
M1	0	82
M2	0	235
M3	4	354
M4	47	1047
M5	904	8568

  

taille_C	min_lits	max_lits
	0	21
C0	0	0
C1	0	140
C2	0	114
C3	4	221
C4	21	342
C5	62	3052

  

taille_O	min_lits	max_lits
	0	45
O0	0	20
O1	0	18
O2	7	50
O3	14	63
O4	22	163
O5	65	866

  

Quantiles (Définition 5)	
Niveau	Quantile
100Max 100%	8568
99%	1380
95%	696
90%	485
75% Q3	265
50% Médiane	130
25% Q1	59
10%	23
5%	10
1%	4
0% Min	0

  

Quantiles (Définition 5)	
Niveau	Quantile
100Max 100%	3052
99%	605
95%	267
90%	176
75% Q3	97
50% Médiane	57
25% Q1	34
10%	20
5%	11
1%	0
0% Min	0

  

Quantiles (Définition 5)	
Niveau	Quantile
100Max 100%	866
99%	148
95%	90
90%	71
75% Q3	47
50% Médiane	28
25% Q1	20
10%	15
5%	13
1%	9
0% Min	4

Analyse de distribution des variables CI\_AC1, CI\_AC6 et CI\_AC8

Plusieurs méthodes existent pour catégoriser des variables numériques. On peut décider manuellement des intervalles, utiliser les quantiles pour diviser la distribution en groupes, ou encore se baser sur des tests statistiques pour valider ou non le découpage de nos données (Student, ANOVA, etc). Dans le cas des variables de taille, on aurait pu supposer que la classification ait été faite en utilisant les quantiles. En effet, si l'on observe les résultats plus haut, on constate que les valeurs des quantiles se rapprochent des bornes observées dans la classification. Cependant, on remarque également que certaines classes se chevauchent, ce qui peut signifier que : soit la classification est erronée, soit d'autres variables en plus de la taille des services ont été prises en compte dans la classification.

## Question 6

	2017				2018				2019			
	Nb_lits_M	Nb_lits_C	Nb_lits_O	Total	Nb_lits_M	Nb_lits_C	Nb_lits_O	Total	Nb_lits_M	Nb_lits_C	Nb_lits_O	Total
CH	63514.00	15543.00	8627.00	87684.00	63232.00	14699.00	8377.00	86308.00	62832.00	14173.00	8129.00	85134.00
CHR	44322.00	18429.00	4883.00	67634.00	44561.00	18113.00	4853.00	67527.00	45254.00	17766.00	4857.00	67877.00
CLCC	1983.00	853.00	0.00	2836.00	1940.00	843.00	0.00	2783.00	1907.00	830.00	0.00	2737.00
CLI	15350.00	29952.00	4562.00	49864.00	15474.00	28474.00	4294.00	48242.00	15628.00	27574.00	4159.00	47361.00
EBNL	7576.00	3904.00	1022.00	12502.00	7639.00	3841.00	982.00	12462.00	7977.00	3718.00	987.00	12682.00

## Question 7

	CH	CHR	CLCC	CLI	EBNL	Total
	Nb Etab	Nb Etab	Nb Etab	Nb Etab	Nb Etab	
01	24	.	.	12	.	36
02	30	.	.	6	.	36
03	15	.	.	9	.	24
04	18	.	.	6	.	24
05	15	.	3	3	.	21
06	30	3	3	48	9	93

## Question 8 - Tableau

		Nb_Accouchements	Nb_Lits_Obstétriques
Catégorie	Etape Maternité		
CH	.	20.00	4.00
	0	312.00	4.00
	1	167088.00	4683.00
	2	678739.00	15122.00
	3	249415.00	5320.00
CHR	.	3.00	63.00
	1	3490.00	80.00
	2	51121.00	1257.00
	3	545697.00	13193.00
CLI	.	2260.00	11.00
	0	218.00	0.00
	1	191643.00	5583.00
	2	303205.00	7406.00
	3	1387.00	15.00
EBNL	.	7.00	8.00
	0	3.00	22.00
	1	67506.00	1271.00
	2	92791.00	1690.00

## Question 8 - TOP 5

finess	nb_accouchement	nb_lits_obstetrique
750712184	114402	2560
690781810	32488	778
770021145	21971	333
970408589	18947	471
670780055	18189	446

## Question 9

dep	nb_accouchement	nb_lits_obstetrique	min_nb_accouchement	max_accouchement
75	241397	5207	0	38519
59	100838	2572	0	5500
13	96226	2483	0	5391
97	88599	2298	0	6489
69	84412	1799	0	10901
92	75781	1415	0	3737
93	72259	1482	0	4630
94	59425	1146	0	3593

## Question 10 - Score de qualité pour l'obstétrique

Pour le calcul de notre score de qualité, nous avons décidé de nous baser sur les variables suivantes : nombre d'accouchements, nombre de lits obstétriques et le nombre de gynécologues. À partir de ces variables, nous avons calculé le taux de lits par accouchement ainsi qu'un taux représentant le nombre de gynécologues par accouchement. Nous avons réalisé ce score uniquement sur les établissements ayant des données renseignées sur les accouchements et ayant réalisés plus de 12 accouchements par an.

Puis, afin de donner la même importance aux deux indicateurs calculés précédemment, nous avons normalisé les deux nouvelles variables (moyenne de 0 et variance de 1). Nous avons fait la moyenne de ces deux valeurs que nous avons ensuite multiplié par 100 pour obtenir notre score final. Voici notre TOP 5 des meilleurs et des pires établissements en matière d'activité obstétrique :

finess	rs	temps_plein_obst	nb_lits	nb_accouchement	score_qual
100010818	GCS ES CLINIQUE DE CHAMPAGNE	11.5	15	284	725.64
920300787	HÔPITAL AMERICAINE	20.3	15	612.6667	375.6904
970107249	CLINIQUE LES EAUX CLAIRES	3.666667	26	811.3333	331.3087
750300840	CLINIQUE DE LA MUETTE	35	24	1115	331.0391
070002878	CH DES VALS D'ARDECHE	2.093333	9	168.6667	327.8291

**TOP**  
5

**FLOP**  
5

finess	rs	temps_plein_obst	nb_lits	nb_accouchement	score_qual
690805361	CH ST JOSEPH ST LUC	7.973333	27	2339.667	-113.974
130785652	HÔPITAL SAINT JOSEPH	8.666667	81.66667	4827.667	-110.84
750100208	HÔPITAL NECKER ENFANTS MALADES	9.51	45	3201.333	-108.002
940110042	CHIC DE VILLENEUVE ST GEORGES	9.476667	48.66667	3347.333	-107.585
940016819	LES HÔPITAUX DE SAINT-MAURICE	8.476667	53.33333	3363.667	-105.792



## Question 11 - Evolution du score de qualité

Nous avons observé les établissements avec les évolutions les plus importantes au cours des années, on est alors tombés sur 5 établissements.

finess	diff
970107249	1185.642
070002878	727.8255
100010818	634.623
310026927	384.1284
970100178	229.8941

finess	rs	year	CI_AC8	CI_A11	CI_ACTot	Lit_par_acc	Temps_plein_par_acc	score_qual
970107249	CLINIQUE LES EAUX CLAIRES	2017	26	131	151	16.598535568	5.790718636	1119.4627102
970107249	CLINIQUE LES EAUX CLAIRES	2018	26	1154	151	-0.407666086	-0.91591817	-66.17921282
970107249	CLINIQUE LES EAUX CLAIRES	2019	26	1149	151	-0.398189477	-0.788958953	-59.35742152

L'établissement qui fluctue le plus semble être la cause d'une erreur de saisie de données, puisque nous avons 131 accouchements en 2017, suivis de 1154 et 1149 en 2018 et 2019 pour le même nombre de lits obstétriques.

finess	rs	year	CI_AC8	CI_A11	CI_ACTot	Lit_par_acc	Temps_plein_par_acc	score_qual
970100178	CH DE LA BASSE TERRE	2017	40	634	213	3.5128725943	1.1551651783	233.40188863
970100178	CH DE LA BASSE TERRE	2018	20	734	198	0.0483278404	0.231332488	13.983016418
970100178	CH DE LA BASSE TERRE	2019	20	731	176	0.0591365354	0.0110197304	3.5078132892

En revanche, pour un autre établissement, on peut voir que le nombre d'accouchements a augmenté de 100 entre 2017 et 2018, alors que le nombre de lits est passé de 40 à 20.



## Question 12 - Score de qualité par département

Pour obtenir un score de qualité par région, il suffit de calculer la moyenne des scores de qualité des établissements de chaque région.

	2017	2018	2019
	Score moyen	Score moyen	Score moyen
01	-15.70	-36.44	-19.64
02	32.07	29.51	20.43
03	-3.05	9.96	2.31
04	51.93	91.49	72.39
05	-1.47	-3.86	-6.72
06	-30.54	-22.13	-0.31
07	34.55	54.93	227.14
08	6.56	33.17	3.91
09	27.47	40.66	23.66
10	-22.11	333.98	186.86
11	3.53	-9.70	12.47
12	16.35	31.79	54.08
13	-9.89	-17.94	-15.42
14	-1.32	-7.40	-3.29
15	100.29	85.00	103.03
16	20.96	17.68	-26.30

Notre score a des limites, car il ne repose que sur 2 indicateurs liés au nombre d'accouchements. Une erreur de saisie au niveau de cette variable peut donc fausser notre score. Pour l'améliorer, il faudrait se baser sur plus de variables.

## Bibliographie

1. *Convert multiple variables between character and numeric formats in SAS.* (2022, 9 février). The Chemical Statistician.

<https://chemicalstatistician.wordpress.com/2018/04/27/convert-multiple-variables-between-character-and-numeric-formats-in-sas/comment-page-1/>