

Rapport projet Transverse

MBAESG

GROUPE:
TORTOSA HUGO
ROBO LUDOVICK
SKY TCHAGO LEWIS
BRUNEAU THOMAS

2021-2022

1 Table des matières

1	Table des matières	1
2	Contexte et mise en place de l'environnement de travail	2
3	Etude globale	2
3.1	Répartition Adhérant/VIP	2
3.2	Comportement du CA GLOBAL par client N-2 vs N-1	2
3.3	Répartition par âge x sexe	3
4	Etude par magasin	5
4.1	Résultat par magasin (+1 ligne Total)	5
4.2	Distance CLIENT / MAGASIN	7
5	Etude par univers	8
5.1	Evolution du CA par univers	8
5.2	Top 5 des familles les plus rentable par univers	10

2 Contexte et mise en place de l'environnement de travail

Après avoir créé une base de données PostgreSQL, nous avons décidé de découper le projet en deux parties indépendantes :

- Une partie basée sur la création des requêtes SQL répondant aux questions.
- Une autre partie basée sur la création des graphiques et des visuels via Power BI, un logiciel de Datavisualisation.

Également, nous mettons à disposition un [répertoire GitHub](#) où nous avons mis nos codes.

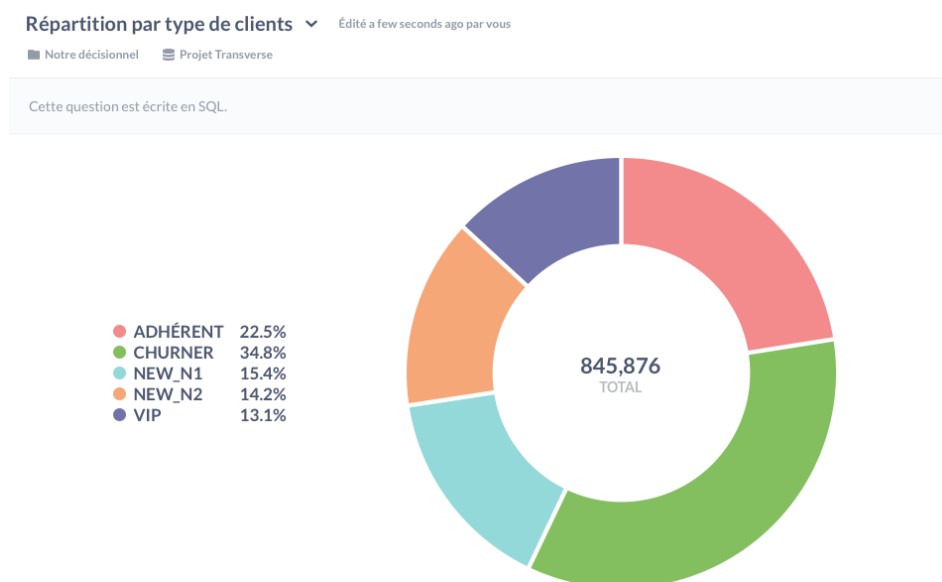
3 Etude globale

3.1 Répartition Adhérant/VIP

Afin de créer un camembert représentant la répartition des clients VIP, adhérents au cours de N-1, adhérent au cours de N-2 et les churner, il a fallu réaliser une segmentation des différents groupes :

1. VIP : nous avons filtré sur la colonne vip
2. NEW_N2 : nous avons filtré sur les clients ayant adhéré en 2016
3. NEW_N1 : nous avons filtré sur les clients ayant adhéré en 2017
4. ADERANT : nous avons filtré sur les clients dont la date d'adhésion est supérieure à Janvier 2018
5. CHURNER : nous avons filtré sur les clients dont la date d'adhésion est inférieure à Janvier 2018

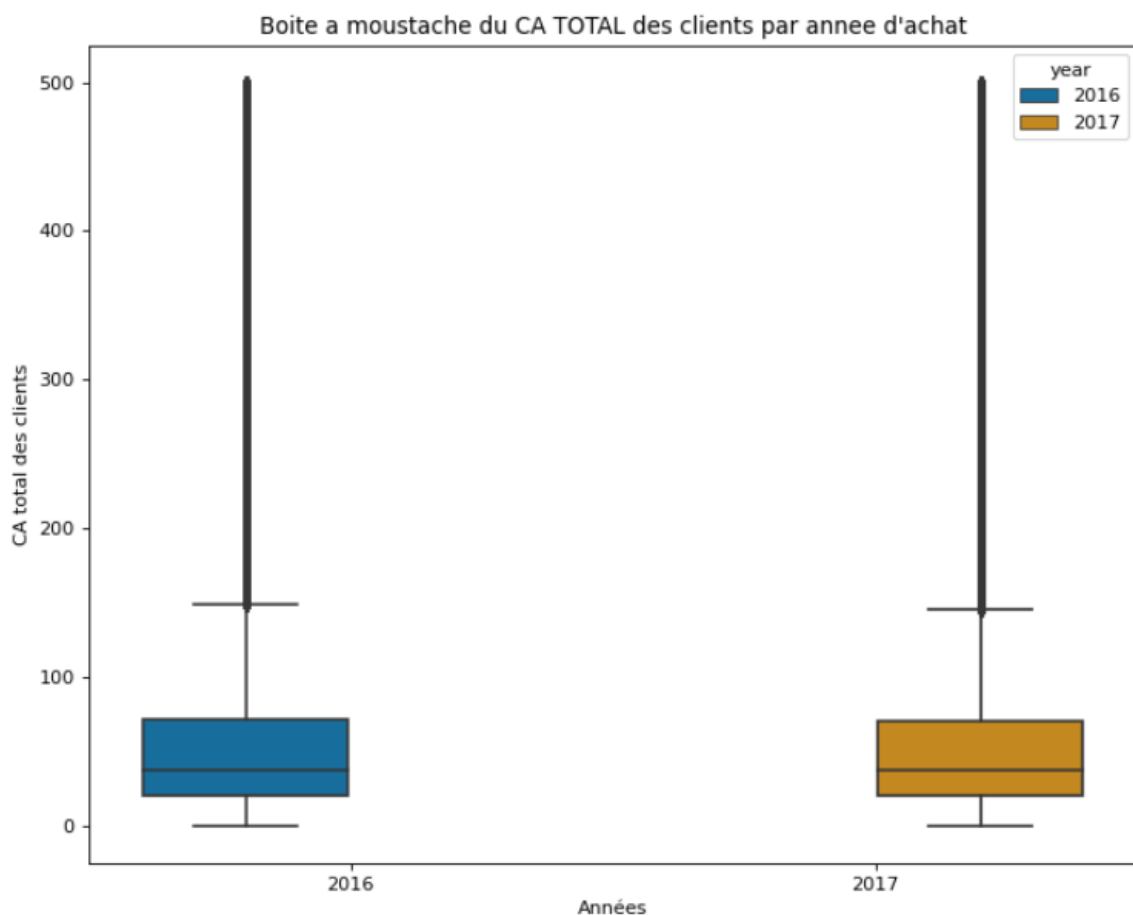
Nous obtenons le camembert suivant :



3.2 Comportement du CA GLOBAL par client N-2 vs N-1

Nous avons créé ce graphique qui représente deux boîtes à moustache représentant les chiffres d'affaires par client sur l'année N-1 (2017) et N-2(2016). Les graphiques ont été réalisés en Python car Power BI ne dispose pas de visuels de type boîte à moustache. En revanche, il a été possible d'intégrer le script Python au sein de Power BI. ([voir lien explicatif](#))

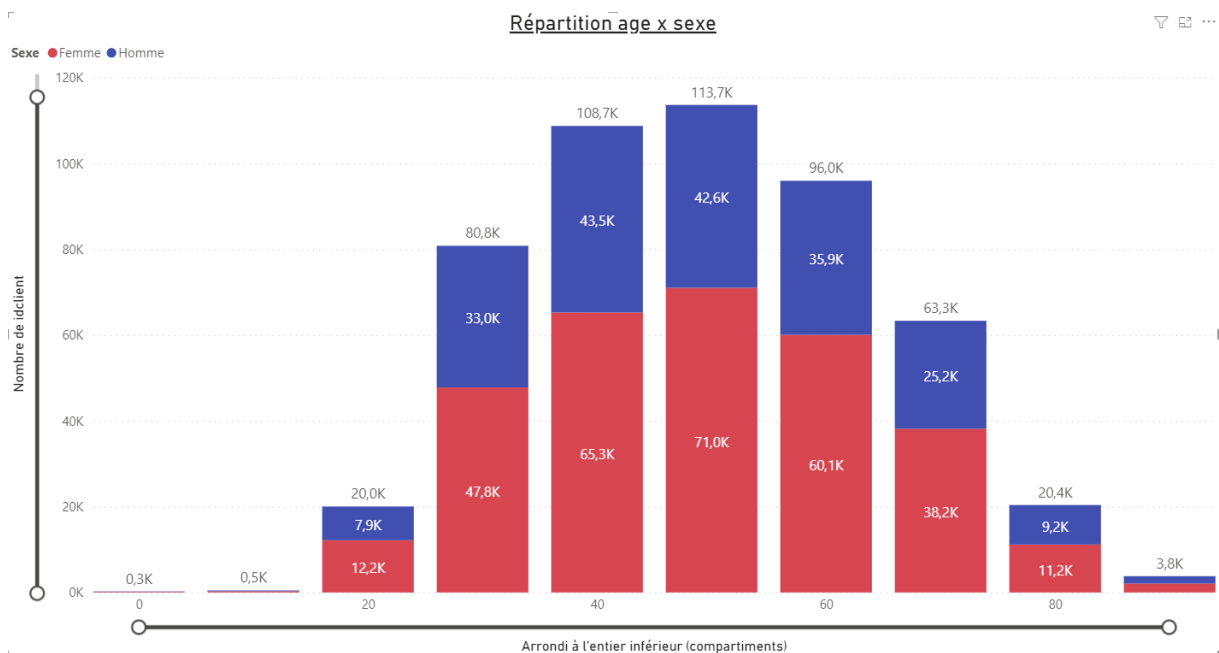
Également, nous avons filtré les valeurs de la colonne TICTOTAL_TTC entre 0 et 500 euros car certaines valeurs étaient négatives et d'autres étaient trop élevées. Ainsi, plus de 66,000 valeurs ont été supprimées



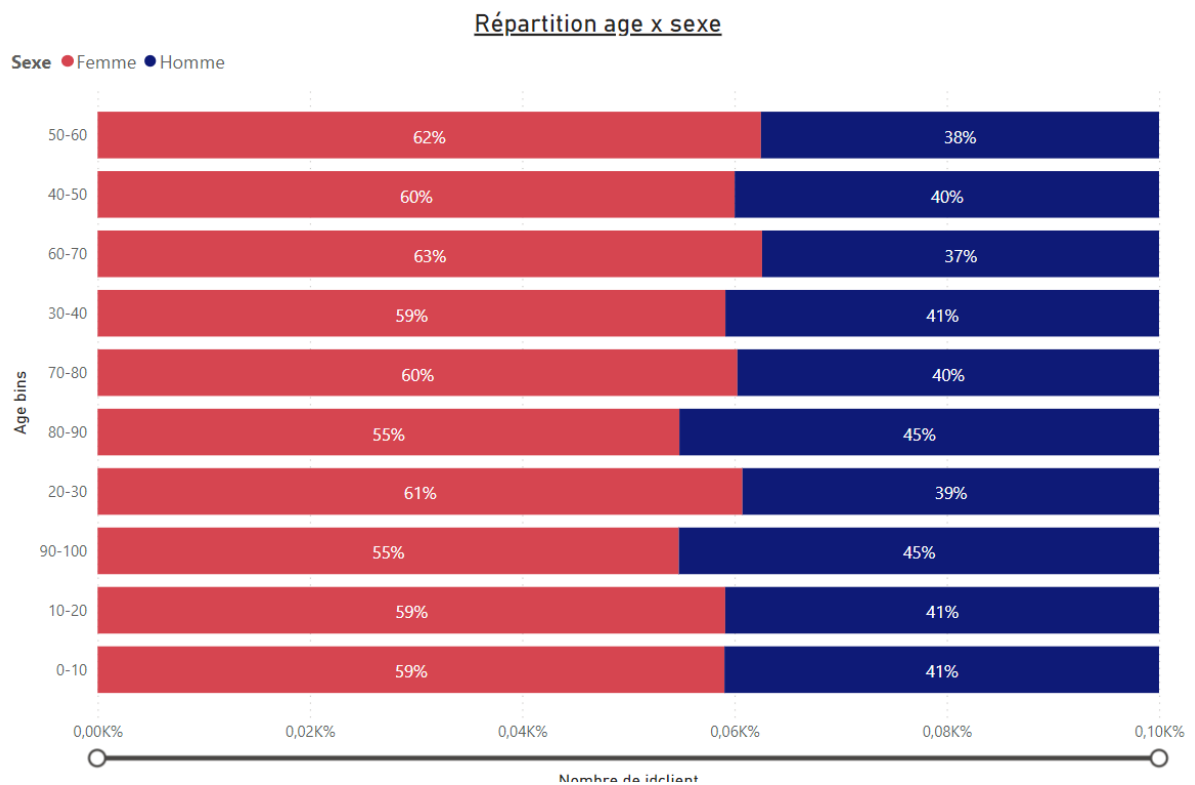
3.3 Répartition par âge x sexe

Avant de réaliser les graphiques pour cette réponse, nous avons dû nettoyer les données. Il a fallu exclure les valeurs anormales. Ainsi, les valeurs supérieures à 100 ans et celles non renseignées ont été supprimées. Ainsi, nous avons décidé de créer deux graphiques pour étudier la répartition des clients par âge et sexe :

- Le premier graphique représente le nombre de clients homme et femme par tranche d'âge de 10 ans. Les tranches d'âges commencent de 0 à 9 ans inclus, de 10 à 19 ans inclus, de 20 à 29 ans inclus etc.. jusqu'à 90 à 99 ans inclus.



- Le second graphique représente le pourcentage d'homme et de femme par tranche d'âge. Ce graphique est trié par rapport au nombre d'individu le plus élevé.



Observations : On constate que la clientèle est essentiellement féminine (proche des 60%) et que les produits ou services sont principalement destinées à la tranche d'âge 50-59 ans.

4 Etude par magasin

4.1 Résultat par magasin (+1 ligne Total)

Pour répondre à cette question, il a été important d'effectuer un travail de Data Cleaning.

Beaucoup de valeurs étaient non renseignées, ce qui pose un problème. Dans notre cas, nous avons pris le parti de supprimer les lignes comportant des valeurs non renseignées. Ainsi, nous obtenons un table 'Client' contenant 507,902 lignes.

Table : public client (507 902 lignes)

Ensuite, nous avons réalisé des mesures rapides permettant de créer les colonnes du tableau :

- Filtre sur les dates afin de créer deux colonnes « Clients actifs N-2 », Clients actifs N-1 », « Total TTC N-2 » et « Total TTC N-1 »
- Différence en % entre deux colonnes afin de créer « Evolution clients actifs » et « Indice évolution »

Enfin, nous avons créé une mise en forme conditionnelle (couleur police : vert si positif, rouge si négatif).

Nous obtenons donc ce tableau suivant :

magasin	Nombre adhérent	Client actifs N-2	Clients actifs N-1	Evolution client actifs	Total TTC N-2	Total TTC N-1	Evolution Total TTC	Indice évolution
ALB	3159	1157	2002	73,03 %	560 414,84	675 009,44	20,45 %	☆☆☆☆☆
ALM	5530	2253	3277	45,45 %	962 775,12	1 136 148,79	18,01 %	☆☆☆☆☆
AVI	2485	1024	1461	42,68 %	442 799,87	532 429,02	20,24 %	☆☆☆☆☆
BAR	35	11	24	118,18 %	6 741,02	9 640,63	43,01 %	☆☆☆☆☆
BEA	4996	1391	3605	159,17 %	932 316,67	1 158 548,19	24,27 %	☆☆☆☆☆
BEC	2652	1099	1553	41,31 %	380 090,83	451 308,21	18,74 %	☆☆☆☆☆
BLA	3793	1692	2101	24,17 %	553 192,10	625 812,44	13,13 %	☆☆☆☆☆
BRE	3181	1446	1735	19,99 %	490 942,70	563 622,63	14,80 %	☆☆☆☆☆
BSN	2019	953	1066	11,86 %	315 478,91	363 905,62	15,35 %	☆☆☆☆☆
CAG	4066	930	3136	237,20 %	692 127,30	798 161,98	15,32 %	☆☆☆☆☆
CLA	6389	2304	4085	77,30 %	1 031 142,71	1 207 762,53	17,13 %	☆☆☆☆☆
CLI	908	530	378	-28,68 %	110 444,45	113 713,30	2,96 %	☆☆☆☆☆
DIJ	4818	2111	2707	28,23 %	752 607,44	862 317,36	14,58 %	☆☆☆☆☆
DUM	1029	356	673	89,04 %	206 517,81	266 145,72	28,87 %	☆☆☆☆☆
ECU	1190	417	773	85,37 %	249 242,43	324 464,92	30,18 %	☆☆☆☆☆
EPN	2911	934	1977	111,67 %	440 766,06	522 410,21	18,52 %	☆☆☆☆☆
EST	323	120	203	69,17 %	47 839,14	56 404,94	17,91 %	☆☆☆☆☆
FEG	3030	1312	1718	30,95 %	433 532,52	491 169,47	13,29 %	☆☆☆☆☆
FRV	3913	1652	2261	36,86 %	647 609,32	769 404,00	18,81 %	☆☆☆☆☆
GAI	5287	2030	3257	60,44 %	1 177 861,53	1 381 142,34	17,26 %	☆☆☆☆☆
GAP	1972	656	1316	100,61 %	364 715,29	440 903,53	20,89 %	☆☆☆☆☆
GEX	5416	1858	3558	91,50 %	1 228 812,98	1 478 728,23	20,34 %	☆☆☆☆☆
HAG	1661	625	1036	65,76 %	223 970,58	261 750,87	16,87 %	☆☆☆☆☆
HEI	6994	2908	4086	40,51 %	1 139 476,60	1 305 726,29	14,59 %	☆☆☆☆☆
IAB	4151	1736	2415	39,11 %	618 580,97	697 664,33	12,78 %	☆☆☆☆☆
Total	233612	92509	141103	52,53 %	37 833 474,78	44 515 207,04	17,66 %	☆☆☆☆☆

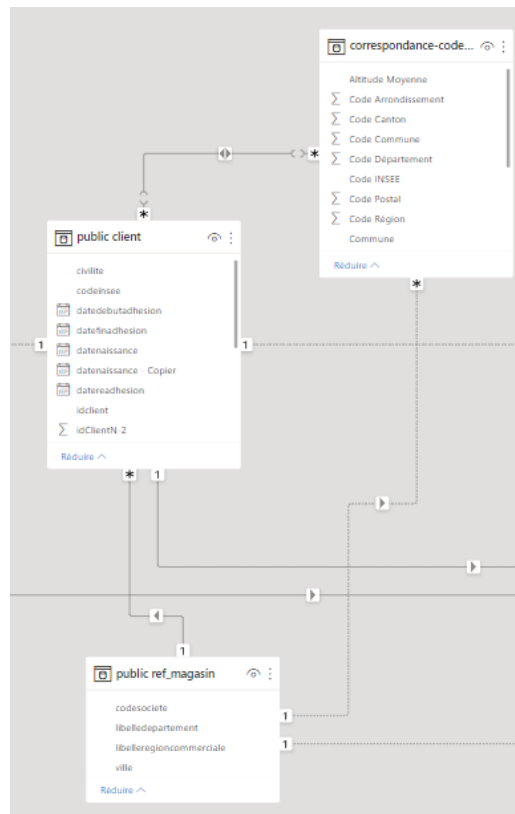
magasin	Nombre adhérent	Client actifs N-2	Clients actifs N-1	Evolution client actifs	Total TTC N-2	Total TTC N-1	Evolution Total TTC	Indice évolution
LAB	4118	1778	2340	31,61 %	616 014,74	702 179,88	13,99 %	☆☆☆☆☆
MAC	4754	2025	2729	34,77 %	663 248,85	771 587,21	16,33 %	☆☆☆☆☆
MAN	3737	1665	2072	24,44 %	611 687,31	697 673,72	14,06 %	☆☆☆☆☆
MET	5446	2122	3324	56,64 %	925 548,52	1 094 190,86	18,22 %	☆☆☆☆☆
MOB	4345	1446	2899	100,48 %	811 508,12	1 019 555,65	25,64 %	☆☆☆☆☆
MOU	6391	2444	3947	61,50 %	1 226 737,76	1 374 662,42	12,06 %	☆☆☆☆☆
MUL	1858	869	989	13,81 %	295 437,05	352 284,21	19,24 %	☆☆☆☆☆
NEV	372	53	319	501,89 %	74 374,04	96 167,72	29,30 %	☆☆☆☆☆
OBE	3600	1362	2238	64,32 %	513 608,86	626 585,14	22,00 %	☆☆☆☆☆
ORL	3363	1329	2034	53,05 %	529 791,95	621 117,89	17,24 %	☆☆☆☆☆
PEG	4228	1836	2392	30,28 %	637 749,97	747 722,75	17,24 %	☆☆☆☆☆
PEP	3042	1404	1638	16,67 %	397 565,64	435 413,52	9,52 %	☆☆☆☆☆
POC	3375	1424	1951	37,01 %	478 304,17	543 810,87	13,70 %	☆☆☆☆☆
PON	3937	1590	2347	47,61 %	617 625,35	707 383,24	14,53 %	☆☆☆☆☆
PRI	6909	2899	4010	38,32 %	1 054 688,74	1 235 018,12	17,10 %	☆☆☆☆☆
QUE	3880	1718	2162	25,84 %	513 280,04	591 091,83	15,16 %	☆☆☆☆☆
RAV	3520	1495	2025	35,45 %	531 700,25	608 642,92	14,47 %	☆☆☆☆☆
RMA	1568	2	1566	78200,00 %	274 884,27	449 764,12	63,62 %	☆☆☆☆☆
SAL	1103	407	696	71,01 %	190 417,10	221 897,05	16,53 %	☆☆☆☆☆
SCH	1786	639	1147	79,50 %	231 548,55	275 333,60	18,91 %	☆☆☆☆☆
SEM	3181	1426	1755	23,07 %	465 836,87	501 042,91	7,56 %	☆☆☆☆☆
SEY	7599	3000	4599	53,30 %	981 789,58	1 184 907,29	20,69 %	☆☆☆☆☆
SGL	1628	759	869	14,49 %	228 489,09	250 441,10	9,61 %	☆☆☆☆☆
SJV	2345	874	1471	68,31 %	359 417,59	449 972,45	25,19 %	☆☆☆☆☆
SLM	2459	1005	1454	44,68 %	303 350,61	360 574,23	18,86 %	☆☆☆☆☆
SMA	1075	688	387	-43,75 %	155 315,13	148 001,41	-4,71 %	☆☆☆☆☆
Total	233612	92509	141103	52,53 %	37 833 474,78	44 515 207,04	17,66 %	☆☆☆☆☆

magasin	Nombre adhérent	Client actifs N-2	Clients actifs N-1	Evolution client actifs	Total TTC N-2	Total TTC N-1	Evolution Total TTC	Indice évolution
RAV	3520	1495	2025	35,45 %	531 700,25	608 642,92	14,47 %	☆☆☆☆☆
RMA	1568	2	1566	78200,00 %	274 884,27	449 764,12	63,62 %	☆☆☆☆☆
SAL	1103	407	696	71,01 %	190 417,10	221 897,05	16,53 %	☆☆☆☆☆
SCH	1786	639	1147	79,50 %	231 548,55	275 333,60	18,91 %	☆☆☆☆☆
SEM	3181	1426	1755	23,07 %	465 836,87	501 042,91	7,56 %	☆☆☆☆☆
SEY	7599	3000	4599	53,30 %	981 789,58	1 184 907,29	20,69 %	☆☆☆☆☆
SGL	1628	759	869	14,49 %	228 489,09	250 441,10	9,61 %	☆☆☆☆☆
SJV	2345	874	1471	68,31 %	359 417,59	449 972,45	25,19 %	☆☆☆☆☆
SLM	2459	1005	1454	44,68 %	303 350,61	360 574,23	18,86 %	☆☆☆☆☆
SMA	1075	688	387	-43,75 %	155 315,13	148 001,41	-4,71 %	☆☆☆☆☆
SMR	3092	1217	1875	54,07 %	539 571,12	634 608,98	17,61 %	☆☆☆☆☆
SNO	2926	928	1998	115,30 %	447 317,88	583 492,84	30,44 %	☆☆☆☆☆
SSM	6307	2422	3885	60,40 %	1 030 663,44	1 255 759,71	21,84 %	☆☆☆☆☆
STE	3738	1585	2153	35,84 %	576 592,57	651 413,32	12,98 %	☆☆☆☆☆
STR	1073	425	648	52,47 %	155 861,59	184 327,99	18,26 %	☆☆☆☆☆
SUR	5849	2611	3238	24,01 %	967 684,15	1 086 512,06	12,28 %	☆☆☆☆☆
THO	3626	1362	2264	66,23 %	700 091,46	837 974,15	19,69 %	☆☆☆☆☆
VAL	1586	624	962	54,17 %	279 814,39	335 840,72	20,02 %	☆☆☆☆☆
VAR	2890	1164	1726	48,28 %	482 291,69	569 092,51	18,00 %	☆☆☆☆☆
VEN	2489	972	1517	56,07 %	428 619,80	513 539,53	19,81 %	☆☆☆☆☆
VIB	8060	3516	4544	29,24 %	1 155 137,73	1 330 003,76	15,14 %	☆☆☆☆☆
VIC	3842	1540	2302	49,48 %	564 641,27	658 796,72	16,68 %	☆☆☆☆☆
VIF	3803	1357	2446	80,25 %	583 305,10	697 290,31	19,54 %	☆☆☆☆☆
VIT	4831	1865	2966	59,03 %	784 241,40	926 572,63	18,15 %	☆☆☆☆☆
VIV	1310	516	794	53,88 %	209 203,20	248 878,95	18,97 %	☆☆☆☆☆
VLG	6663	2641	4022	52,29 %	1 228 518,65	1 439 785,76	17,20 %	☆☆☆☆☆
Total	233612	92509	141103	52,53 %	37 833 474,78	44 515 207,04	17,66 %	☆☆☆☆☆

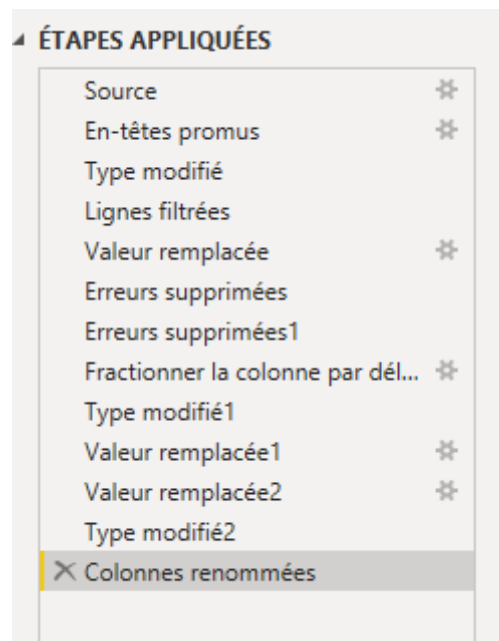
4.2 Distance CLIENT / MAGASIN

Afin de répondre à cette question, il a fallu réaliser un travail important en amont.

Après avoir téléchargé les données de l'INSEE, nous avons lié les tables entre elles comme le montre l'image suivante



Ensuite nous les avons travaillés en appliquant une suite d'étape



Finalement, nous obtenons 2 tables :

- L'une contenant les longitudes et latitudes de la ville
- L'autre contenant les longitudes et latitudes du lieu de résidence du client

1.2 Lat_client	1.2 Long_client
45,61362895	5,155476734
45,61362895	5,155476734
45,61362895	5,155476734
45,61362895	5,155476734
45,61362895	5,155476734
45,61362895	5,155476734
45,61362895	5,155476734
45,61362895	5,155476734
45,61362895	5,155476734
45,61362895	5,155476734

Ensuite, à partir des deux latitudes et des deux longitudes, nous cherchons à calculer la distance en kilomètre en appliquant la formule de la **distance orthodromie** sur Power BI :

[RadiusEarth] = 6371

$$\text{Distance_km} = [\text{RadiusEarth}] * ((2 * \text{ASIN}(\text{SQRT}((\text{SIN}((\text{RADIANS}(\text{Distance}[\text{Lat_1}]) - \text{RADIANS}(\text{Distance}[\text{Lat_2}]))/2)^2 + \text{COS}(\text{RADIANS}(\text{Distance}[\text{Lat_1}]) * \text{COS}(\text{RADIANS}(\text{Distance}[\text{Lat_2}])) * (\text{SIN}((\text{RADIANS}(\text{Distance}[\text{Long_1}]) - \text{RADIANS}(\text{Distance}[\text{Long_2}]))/2)^2))))))$$

En revanche, nous ne sommes pas parvenus à appliquer la formule aux tables Power BI.

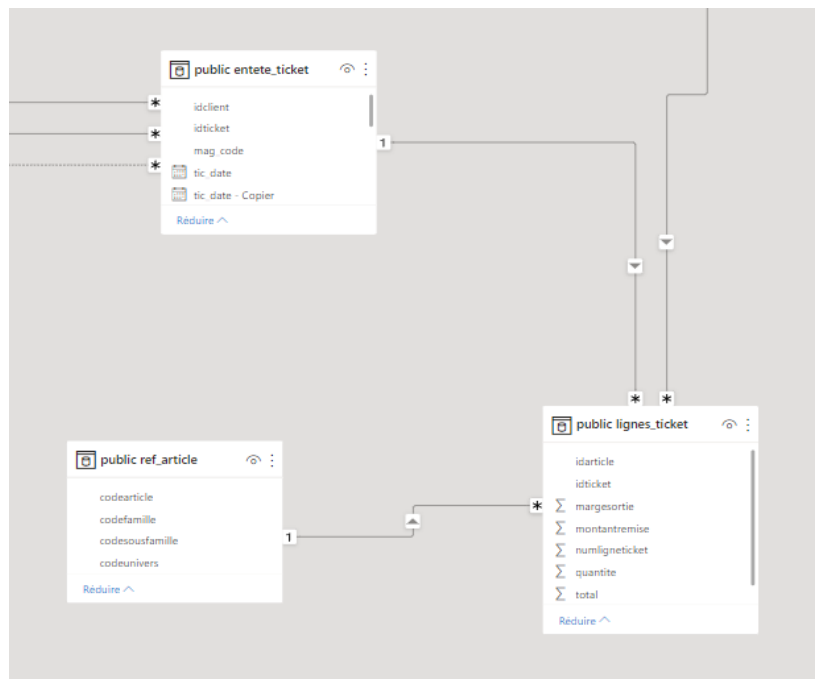
5 Etude par univers

5.1 Evolution du CA par univers

Pour réaliser cet histogramme, nous avons dû réaliser des jointures entre les tables, notamment entre les tables entete_ticket, lignes_ticket et ref_article.

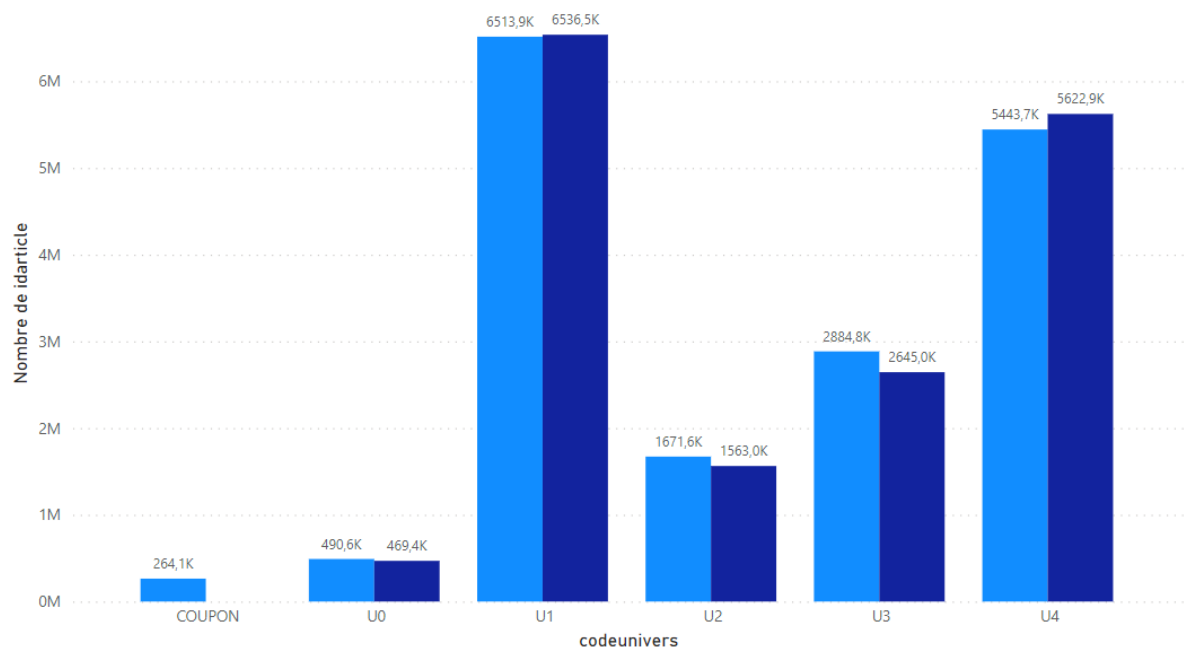
Power BI offre la possibilité de créer des jointures entre les tables.

Ainsi, voici ci-dessous le métamodèle que nous avons créé avec les relations entre les tables entete_ticket, lignes_ticket et ref_article



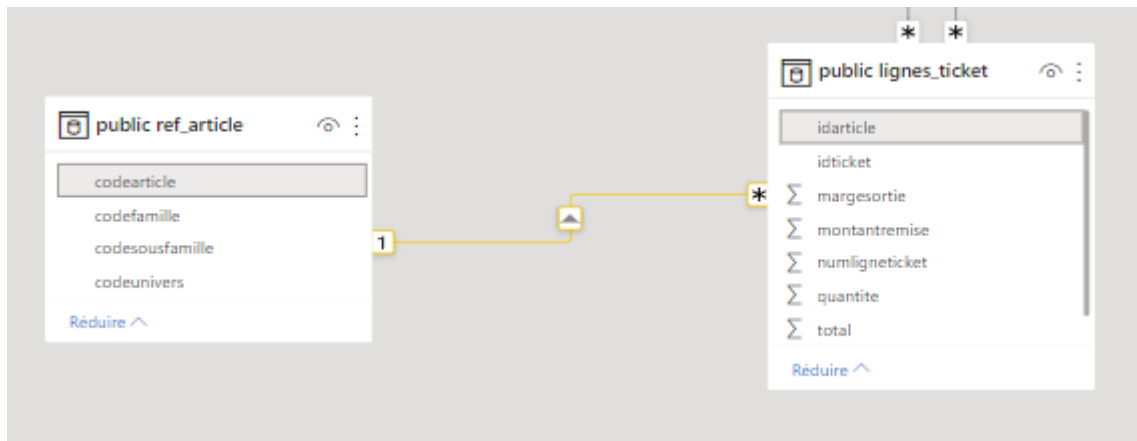
Histogramme N-2 / N-1 évolution du CA par univers

Année ● 2016 ● 2017



5.2 Top 5 des familles les plus rentable par univers

De manière similaire à la précédente question, il a fallu réaliser des jointures entre les tables `ref_article` et `lignes_ticket` comme le montre l'image suivante.



Ensuite, il a fallu faire un travail de « Data Cleaning » en supprimant les valeurs aberrantes. Ainsi, la valeur « `margesortie` » a été filtrée pour ne garder que les valeurs supérieures à 0.

On obtient les résultats suivants :

- Un histogramme des familles les plus rentables par univers



- Un tableau qui reprend les résultats de l'histogramme précédent

Tableau des familles les plus rentables par univers

codeunivers	codefamille	margesortie
U0	160	5 917 722,98
U1	010	13 094 263,08
U1	020	17 461 174,60
U1	050	6 141 795,40
U1	060	8 230 891,96
U1	080	10 245 218,21
U1	100	9 370 663,66
U1	190	2 339 395,83
U2	070	7 742 154,86
U2	090	8 140 625,82
U2	130	4 425 300,30
U2	210	568 901,34
U3	110	4 300 843,68
U3	120	25 890 826,74
U4	140	753 928,02
U4	200	282 383,97
U4	220	12 876 658,43
Total		137 782 748,88