# PROJET TRANSVERSE SQL

MBDCDO

JANVIER 2021

**Emilie Groschêne**
**Clara Nussbaumer**
**Pedro Santiago Ugarte**
**David Velandia**

MBAESG

# SOMMAIRE

## Table des matières

# Avant-propos

## Objectifs

- Manipuler et analyser de la Data sous SQL
- Programmer en SQL
- A rendre avant le 31 Janvier 2021 avant 23h (un jour de retard = 2 points en moins)

## Rendu

Un fichier.sql commenté avec les numéros d'exercices ainsi qu'un rapport comprenant les graphiques effectués sur l'outil de Dataviz de notre choix.
Bonus : un lien vers le répertoire GIT avec le document .sql et le rapport.

## Contexte

Une société X a envoyé ses données clients ainsi que les achats sur l'année N-2 (2016) et N-1 (2017).

# 1) Etude globale

## a) Répartition Adhérent / VIP

***Enoncé :***

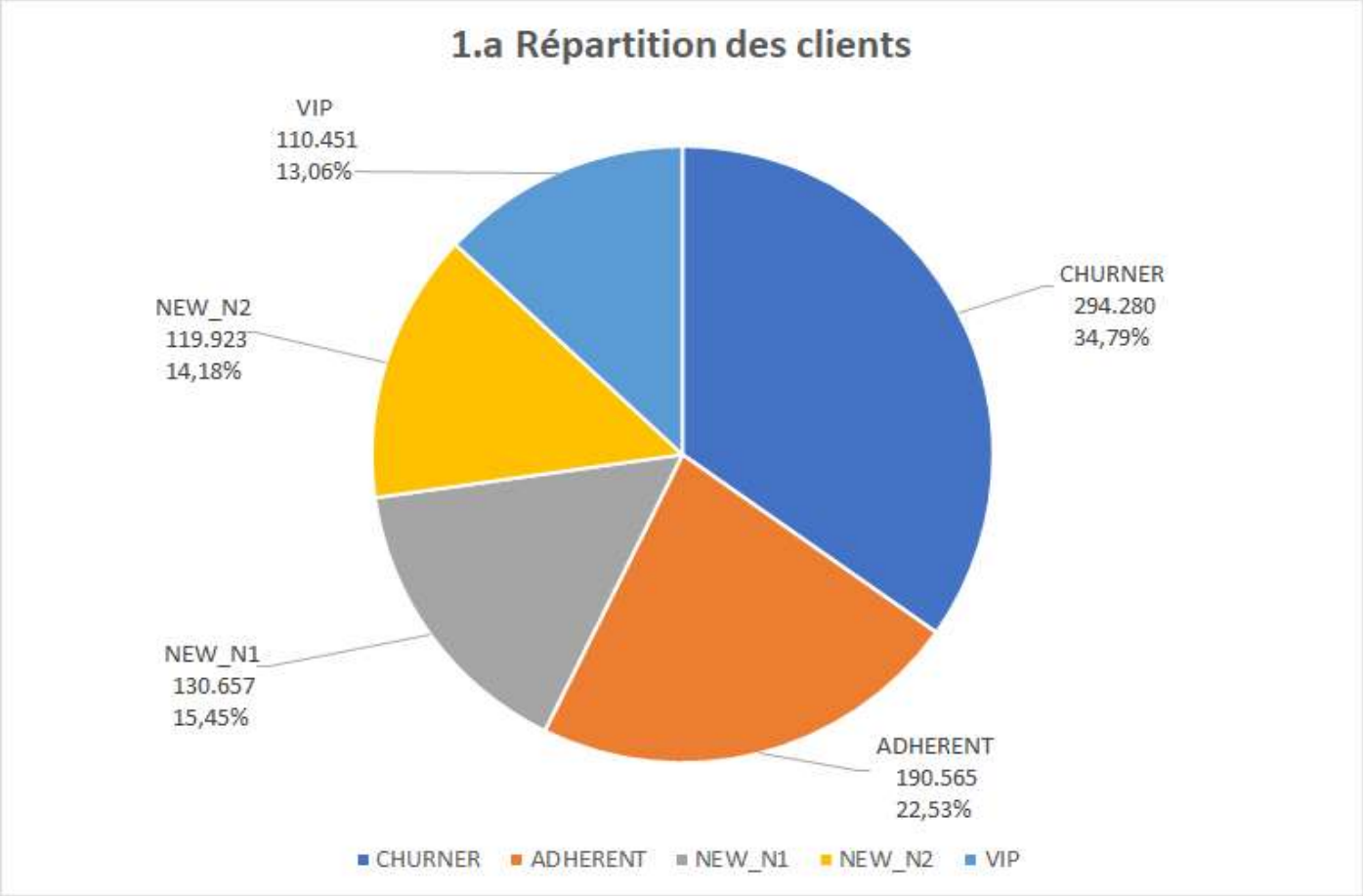
*Constituer un camembert suivant la répartition suivante :*

- *VIP : client étant VIP (VIP = 1)*
- *NEW_N2 : client ayant adhéré au cours de l'année N-2 (date début adhésion)*
- *NEW_N1 : client ayant adhéré au cours de l'année N-1 (date début adhésion)*
- *ADHÉRENT : client toujours en cours d'adhésion (date de fin d'adhésion > 2018/01/01)*
- *CHURNER : client ayant churné (date de fin d'adhésion < 2018/01/01)*

*Note : le critère le plus au-dessus est prioritaire.*

*Exemple : un client étant VIP, et ayant adhéré sur l'année N-1 sera compté comme étant VIP.*

Pour la création du graph sur excel à partir d'une table SQL, nous avons interrogé les données existantes (VIP, date de début et fin d'adhésion, churner) en suivant l'ordre de priorité indiqué dans l'énoncé.



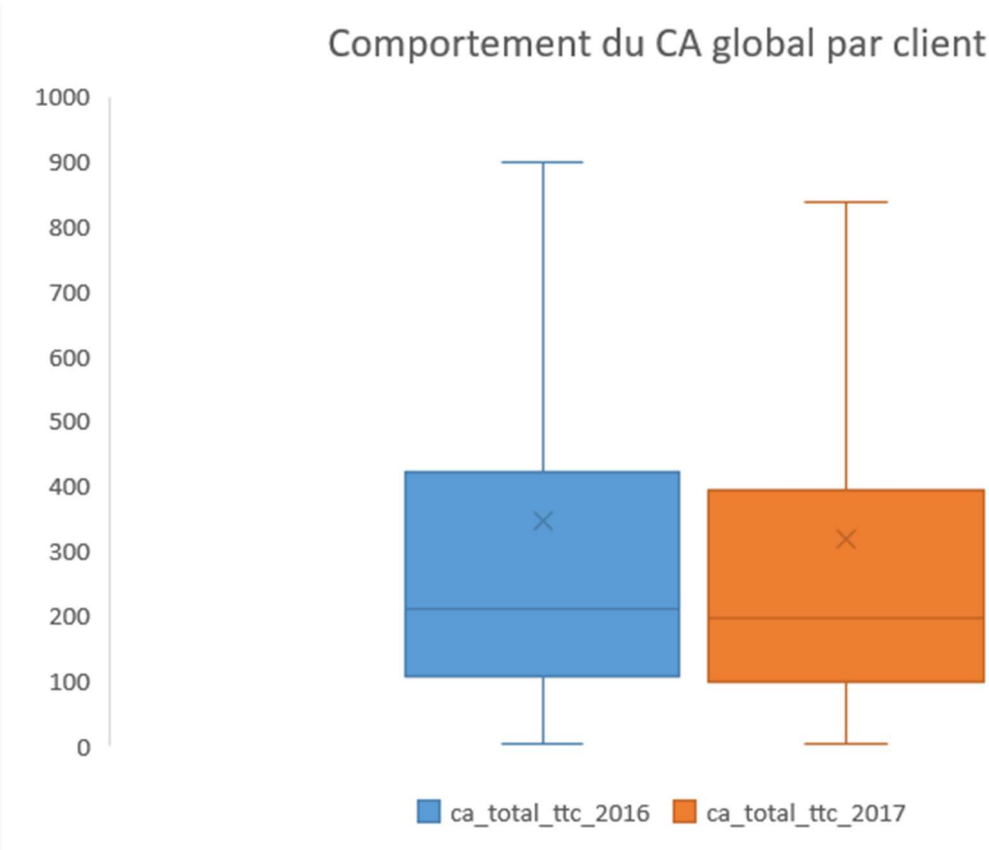

*Dataviz sous Excel*

## b) Comportement du CA Global par client N-2 vs N-1

***Enoncé :***

*Constituer une boîte à moustaches pour chaque année (N-2 et N-1) comparant le CA TOTAL (TTC) des clients (sommer les achats par client par années)*

Nous avons pris la décision de filtrer les données inférieures à 0 pour ne pas prendre en compte les remboursements, ainsi que d'arrondir le résultat de l'addition du chiffre d'affaires par client.




*Dataviz sous Excel*

## c) Répartition par âge et sexe

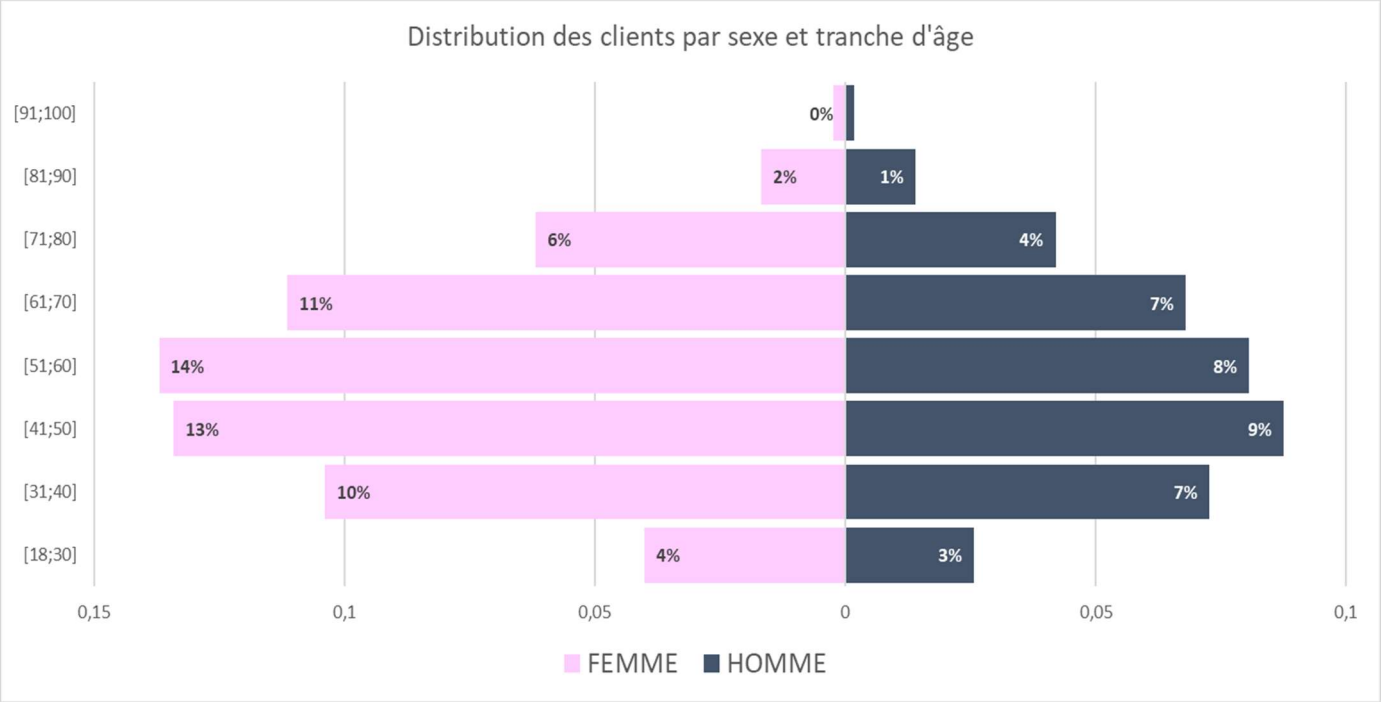
***Enoncé :***

*Constituer un graphique montrant la répartition par âge et sexe sur l'ensemble des clients.*

Afin de faciliter la lecture, nous avons sélectionné dans SQL les clients entre 18 et 100 ans.

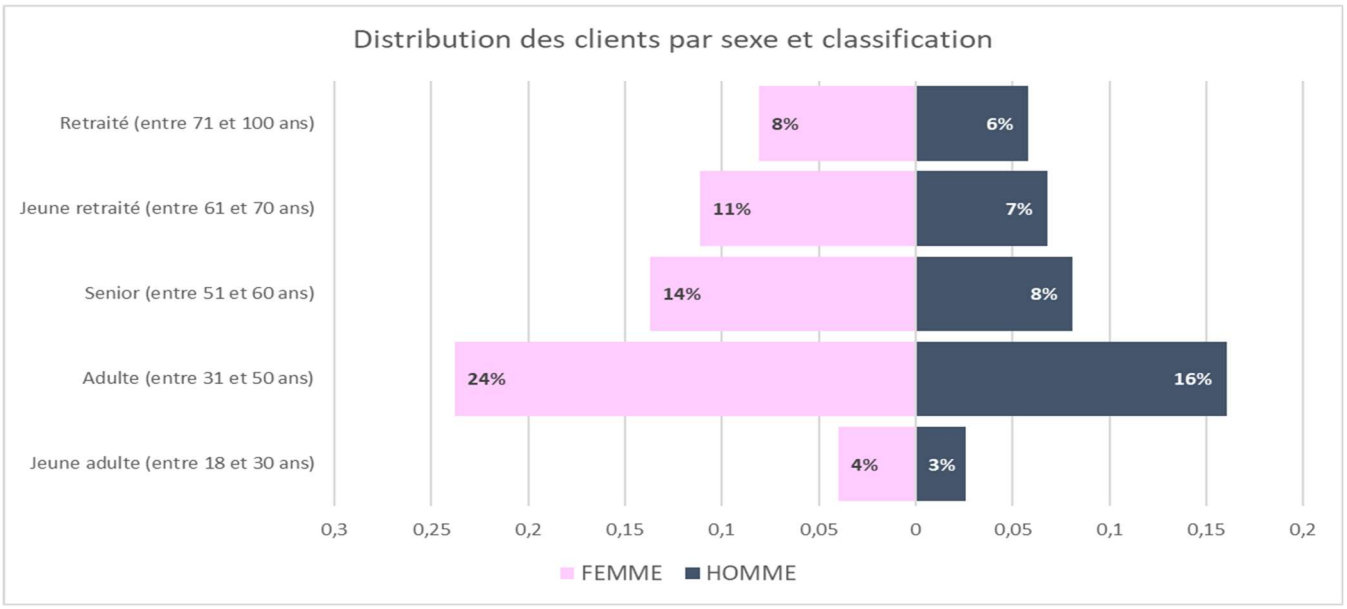
Nous avons choisi de représenter les données sous la forme d'une pyramide des âges. Ce graphique a l'avantage de nous fournir une double indication :

- Les femmes représentent une part plus importante de la clientèle totale
- La tranche d'âge la plus représentée chez les femmes est la tranche [51;60] et celle des hommes [41;50].




*Dataviz sous Excel*

Même représentation graphique en fonction non pas des tranches d'âge mais d'une autre classification :




*Dataviz sous Excel*

# 2) Etude par magasin

## a) Résultat par magasin (+1 ligne TOTAL)

***Enoncé :***

*Constituer un tableau reprenant les données suivantes :*

- *Magasin*
- *Nombre de clients rattachés au magasin (avec une color_bar en fonction de la quantité)*
- *Nombre de client actif sur N-2*
- *Nombre de client actif sur N-1*
- *% clients N-1 vs N-2 (couleur police : vert si positif, rouge si négatif)*
- *TOTAL_TTC N-2*
- *TOTAL_TTC N-1*
- *Différence entre N-1 et N-2 (couleur police : vert si positif, rouge si négatif)*
- *Indice évolution (icône de satisfaction : positif si % client actif évolue et total TTC aussi, négatif si diminution des 2 indicateurs, moyen seulement l'un des deux diminue)*
- *Note : on effectuera un tri sur l'indice d'évolution (les positifs en haut, les négatifs en bas.*

Ce tableau confirme la question 1) a) où nous avions déduit que le CA moyen par personne avait diminué entre 2016 et 2017. En effet le nombre de clients actifs a augmenté de 6% entre ces deux années alors que le CA a quant à lui diminué de 1,68%.

| Code Magasin | Ville | Departement | Region | Nombre d'adherents | Client actif N-1 | Client Actif N-2 | Evolution Client Actif | TTC total N-1 | TTC total N-2 | Evolution total TTC | Indice Evolution |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PRI | ST PRIEST | 69 | Rhône-Alpes | 26.935 | 24.231 | 22.413 | 8,11 | 5.557.473 € | 5.528.750 € | 28.723 € | |
| QUE | QUETIGNY | 21 | Alsace-Est | 11.699 | 11.364 | 9.924 | 14,51 | 1.898.024 € | 1.783.723 € | 114.301 € | |
| BEC | STRASBOURG | 67 | Alsace-Est | 7.977 | 6.683 | 6.092 | 9,70 | 1.485.816 € | 1.470.501 € | 15.315 € | |
| RMA | RUEIL MALMAISON | 92 | Centre-Paris | 6.499 | 8.182 | 4.620 | 77,10 | 2.656.634 € | 1.108.075 € | 1.548.559 € | |
| SAL | SALLANCHES | 74 | Rhône-Alpes | 3.124 | 3.065 | 2.621 | 16,94 | 664.169 € | 605.825 € | 58.344 € | |
| SNO | SEYNOD | 74 | Rhône-Alpes | 10.828 | 12.165 | 9.927 | 22,54 | 2.292.898 € | 2.071.763 € | 221.135 € | |
| SMR | ST MITRE LES REMPARTS | 13 | Littoral | 12.532 | 10.005 | 9.628 | 3,92 | 3.063.415 € | 3.044.484 € | 18.931 € | |
| SEY | SEYSSINS | 38 | Rhône-Alpes | 25.967 | 21.518 | 19.712 | 9,16 | 4.400.119 € | 4.336.407 € | 63.712 € | |
| SLM | SARGE LES LE MANS | 72 | Centre-Paris | 7.894 | 5.742 | 5.448 | 5,40 | 1.431.045 € | 1.366.816 € | 64.229 € | |
| SJV | ST JEAN DE VEDAS | 34 | Littoral | 11.065 | 9.430 | 8.658 | 8,92 | 2.414.626 € | 2.254.727 € | 159.899 € | |
| BLA | BLAGNAC | 31 | Littoral | 11.748 | 8.613 | 7.592 | 13,45 | 2.255.600 € | 2.143.925 € | 111.675 € | |
| BRE | BRETIGNY SUR ORGE | 91 | Centre-Paris | 10.106 | 7.539 | 6.429 | 17,27 | 2.121.890 € | 2.089.464 € | 32.426 € | |
| VIV | VIVIER AU COURT | 8 | Alsace-Est | 3.632 | 3.465 | 3.126 | 10,84 | 973.409 € | 923.416 € | 49.993 € | |
| HAG | HAGUENAU | 67 | Alsace-Est | 4.780 | 4.653 | 4.028 | 15,52 | 848.129 € | 801.071 € | 47.058 € | |
| CAG | CAGNES SUR MER | 6 | Littoral | 9.071 | 8.978 | 7.955 | 12,86 | 1.792.635 € | 1.697.422 € | 95.213 € | |
| IAB | L'ISLE D'ABEAU | 38 | Rhône-Alpes | 16.885 | 13.749 | 11.975 | 14,81 | 3.239.041 € | 3.222.142 € | 16.899 € | |
| BAR | BARCELONNETTE | 4 | Littoral | 1.321 | 1.486 | 1.331 | 11,65 | 511.120 € | 455.609 € | 55.511 € | |
| MAC | MACON | 71 | Centre-Paris | 14.632 | 11.829 | 10.776 | 9,77 | 2.821.782 € | 2.798.360 € | 23.422 € | |
| VIC | VILLECHETIF | 10 | Alsace-Est | 11.556 | 8.455 | 7.703 | 9,76 | 2.133.379 € | 2.114.767 € | 18.612 € | |
| VIT | VITROLLES | 13 | Littoral | 14.673 | 12.106 | 10.930 | 10,76 | 3.089.915 € | 3.007.212 € | 82.703 € | |
| MOB | MONTBONNOT ST MARTIN | 38 | Rhône-Alpes | 18.929 | 18.240 | 16.342 | 11,61 | 6.384.418 € | 5.977.005 € | 407.413 € | |
| ALM | LES MILLES | 13 | Littoral | 16.499 | 15.577 | 14.276 | 9,11 | 4.113.500 € | 4.036.789 € | 76.711 € | |
| DUM | ANNECY LE VIEUX | 74 | Rhône-Alpes | 6.761 | 9.279 | 8.058 | 15,15 | 2.458.297 € | 2.208.889 € | 249.408 € | |
| THO | THONON LES BAINS | 74 | Rhône-Alpes | 10.549 | 8.842 | 8.111 | 9,01 | 2.979.304 € | 2.897.265 € | 82.039 € | |
| OBE | OBERNAI | 67 | Alsace-Est | 9.709 | 8.122 | 7.552 | 7,55 | 1.796.716 € | 1.793.099 € | 3.617 € | |

| Code Magasin | Ville | Departement | Region | Nombre d'adherents | Client actif N-1 | Client Actif N-2 | Evolution Client Actif | TTC total N-1 | TTC total N-2 | Evolution total TTC | Indice Evolution |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ECU | ECULLY | 69 | Rhône-Alpes | 15.410 | 15.813 | 14.485 | 9,17 | 4.274.238 € | 4.112.227 € | 162.011 € | ↑ |
| EPN | EPINAL | 88 | Alsace-Est | 8.800 | 7.123 | 6.283 | 13,37 | 1.821.401 € | 1.748.056 € | 73.345 € | ↑ |
| ALB | GILLY SUR ISERE | 73 | Rhône-Alpes | 12.828 | 10.466 | 9.717 | 7,71 | 3.049.735 € | 2.960.492 € | 89.243 € | ↑ |
| SUR | SURESNES | 92 | Centre-Paris | 23.784 | 15.674 | 14.926 | 5,01 | 4.966.888 € | 4.807.772 € | 159.116 € | ↑ |
| SSM | LA SEYNE SUR MER | 83 | Littoral | 23.601 | 18.358 | 16.400 | 11,94 | 6.842.403 € | 6.413.412 € | 428.991 € | ↑ |
| FEG | FEGERSHEIM | 67 | Alsace-Est | 10.561 | 8.741 | 8.570 | 2,00 | 2.161.776 € | 2.347.398 € | -185.622 € | → |
| AVI | VILLENEUVE LES AVIGNON | 30 | Littoral | 15.435 | 12.889 | 11.904 | 8,27 | 3.951.028 € | 4.065.940 € | -114.912 € | → |
| BEA | BEAUMONT | 63 | Centre-Paris | 20.714 | 16.021 | 14.792 | 8,31 | 5.566.070 € | 5.739.924 € | -173.854 € | → |
| CLA | CLAPIERS | 34 | Littoral | 23.744 | 17.439 | 16.444 | 6,05 | 5.639.066 € | 5.847.742 € | -208.676 € | → |
| DIJ | DIJON | 21 | Alsace-Est | 12.603 | 11.524 | 10.906 | 5,67 | 2.900.306 € | 3.043.839 € | -143.533 € | → |
| FRV | FRANCHEVILLE | 69 | Rhône-Alpes | 15.377 | 14.205 | 13.292 | 6,87 | 3.764.134 € | 3.922.133 € | -157.999 € | → |
| GAI | GAILLARD | 74 | Rhône-Alpes | 16.162 | 16.478 | 16.405 | 0,44 | 5.666.501 € | 6.143.169 € | -476.668 € | → |
| GAP | GAP | 5 | Littoral | 8.137 | 6.997 | 6.869 | 1,86 | 1.926.181 € | 2.080.513 € | -154.332 € | → |
| GEX | ST GENIS POUILLY | 1 | Rhône-Alpes | 19.278 | 15.865 | 15.544 | 2,07 | 6.957.686 € | 7.388.908 € | -431.222 € | → |
| HEI | HEILLECOURT | 54 | Alsace-Est | 23.500 | 18.558 | 16.999 | 9,17 | 5.796.841 € | 5.834.750 € | -37.909 € | → |
| LAB | LABEGE | 31 | Littoral | 11.058 | 8.448 | 7.684 | 9,94 | 2.021.050 € | 2.053.102 € | -32.052 € | → |
| MET | METZ TESSY | 74 | Rhône-Alpes | 17.997 | 18.655 | 18.161 | 2,72 | 4.633.845 € | 4.826.410 € | -192.565 € | → |
| MOU | MOUANS SARTOUX | 6 | Littoral | 24.389 | 18.111 | 16.780 | 7,93 | 6.045.146 € | 6.349.088 € | -303.942 € | → |
| MUL | MULHOUSE | 68 | Alsace-Est | 14.240 | 10.333 | 10.105 | 2,26 | 3.172.042 € | 3.332.191 € | -160.149 € | → |
| ORL | ORLEANS | 45 | Centre-Paris | 9.891 | 7.229 | 6.981 | 3,55 | 2.467.712 € | 2.695.003 € | -227.291 € | → |
| PEG | PERRIGNY | 89 | Alsace-Est | 11.012 | 8.164 | 7.778 | 4,96 | 2.044.230 € | 2.118.864 € | -74.634 € | → |
| PEP | PERPIGNAN | 66 | Littoral | 7.609 | 4.952 | 4.912 | 0,81 | 1.225.989 € | 1.315.595 € | -89.606 € | → |
| RAV | LA RAVOIRE | 73 | Rhône-Alpes | 11.396 | 9.524 | 9.425 | 1,05 | 2.302.864 € | 2.625.701 € | -322.837 € | → |
| VEN | VENELLES | 13 | Littoral | 10.544 | 11.742 | 11.411 | 2,90 | 3.182.816 € | 3.354.293 € | -171.477 € | → |
| VIB | VILLEURBANNE | 69 | Rhône-Alpes | 24.640 | 20.604 | 19.105 | 7,85 | 4.323.459 € | 4.402.350 € | -78.891 € | → |
| VIF | VILLEFRANCHE SUR SAONE | 69 | Centre-Paris | 14.180 | 10.526 | 10.052 | 4,72 | 2.586.192 € | 2.709.179 € | -122.987 € | → |
| SCH | SCHWEIGHOUSE SUR MODER | 67 | Alsace-Est | 5.526 | 3.530 | 4.555 | -22,50 | 624.614 € | 979.935 € | -355.321 € | ↓ |
| PON | LE PONTET | 84 | Littoral | 11.788 | 9.625 | 9.882 | -2,60 | 2.363.868 € | 2.577.363 € | -213.495 € | ↓ |
| STE | LA FOUILLOUSE | 42 | Rhône-Alpes | 12.001 | 9.388 | 9.519 | -1,38 | 2.279.578 € | 2.526.952 € | -247.374 € | ↓ |
| STR | SISTERON | 4 | Littoral | 2.946 | 3.048 | 3.101 | -1,71 | 642.992 € | 670.530 € | -27.538 € | ↓ |
| CLI | CLIRON | 8 | Alsace-Est | 4.454 | 3.072 | 3.297 | -6,82 | 794.102 € | 918.076 € | -123.974 € | ↓ |
| POC | PONTAULT COMBAULT | 77 | Centre-Paris | 10.182 | 6.658 | 6.688 | -0,45 | 2.235.014 € | 2.406.910 € | -171.896 € | ↓ |
| NEV | NEVERS | 58 | Centre-Paris | 7.007 | 9.925 | 10.268 | -3,34 | 1.769.607 € | 1.987.412 € | -217.805 € | ↓ |
| VAL | VALENCE | 26 | Rhône-Alpes | 16.685 | 11.776 | 11.995 | -1,83 | 2.986.766 € | 3.459.988 € | -473.222 € | ↓ |
| VAR | VARENNES VAUZELLES | 58 | Centre-Paris | 12.948 | 11.756 | 12.025 | -2,24 | 2.701.641 € | 3.150.352 € | -448.711 € | ↓ |
| BSN | CESSON | 77 | Centre-Paris | 10.302 | 6.082 | 7.137 | -14,78 | 1.665.793 € | 2.020.922 € | -355.129 € | ↓ |
| MAN | MANOSQUE | 4 | Littoral | 11.343 | 8.749 | 8.905 | -1,75 | 2.461.591 € | 2.719.688 € | -258.097 € | ↓ |
| VLG | VILLE LA GRAND | 74 | Rhône-Alpes | 22.541 | 19.603 | 19.855 | -1,27 | 5.653.658 € | 6.212.517 € | -558.859 € | ↓ |
| SGL | ST GENEST LERPT | 42 | Rhône-Alpes | 8.846 | 6.652 | 6.682 | -0,45 | 1.450.191 € | 1.575.548 € | -125.357 € | ↓ |
| SMA | STE MAXIME | 83 | Littoral | 5.413 | 4.489 | 4.783 | -6,15 | 1.490.880 € | 1.741.264 € | -250.384 € | ↓ |
| SEM | FEVES | 57 | Alsace-Est | 10.360 | 6.440 | 6.875 | -6,33 | 1.781.800 € | 1.959.365 € | -177.565 € | ↓ |
| TOTAL | FRANCE | 0 | FR | 844.603 | 708.550 | 666.724 | 6,27 | 193.575.048 € | 196.882.379 € | -3.307.331 € | → |

L'écart de 1273 clients entre notre tableau et le nombre total de clients de la table client correspond aux clients du magasin 'EST', où nous n'avons pas un chiffre d'affaires pour les 2 années étudiées.
Nous avons décidé de ne pas prendre en compte ce magasin sur le tableau pour ne pas impacter l'évolution totale.

## b) Distance CLIENT / MAGASIN

***Enoncé :***

*Le but étant de calculer la distance qui existe entre le magasin et le client.*
*Les infos disponibles pour le moment sont :*

- *La ville du magasin,*
- *Le code insee du client*

*Il faut donc télécharger les données GPS des villes et code-insee pour pouvoir calculer la distance :*
*https://public.opendatasoft.com/explore/dataset/correspondance-code-insee-code-postal/table/*

*Une fois les données acquises, il faut lier les données GPS composés de la latitude et de la longitude au client et au magasin (constituer pour chaque client et chaque magasin 2 colonnes : latitude et longitude).*

*Créer une fonction qui détermine la distance entre 2 points. La fonction doit prendre 4 variables en compte : latitude1, longitude1, latitude2, longitude2.*

*Pour savoir si la fonction est correcte : http://www.lexilogos.com/calcul_distances.htm*

*Constituer une représentation (tableau ou graphique --> au choix) représentant le nombre de clients par distance : 0 à 5km, 5km à 10km, 10km à 20km, 20km à 50km, plus de 50km*

Pour cette question nous avons créé 3 tables : une première avec les données INSEE et GPS triées, une deuxième table avec les données clients et une troisième table avec les données magasin. Nous avons ensuite mis ces tables ensemble dans une dernière table finale où nous avons appliqué notre formule.

- 1ère table : données INSEE et GPS.

Pour la création de cette table nous avons téléchargé les données CODEINSEE, CODE_POSTAL, COMMUNE, GEO_POINT2. Nous avons harmonisé ces données pour pouvoir appliquer notre formule (latitude et longitude dans 2 colonnes séparées) et pout faire les jointures avec les tables client et magasin : code INSEE à 5 chiffres ; code postal à 2 chiffres ; suppression des « - » et « CEDEX » dans les noms de communes ; remplacement des « SAINT » par « ST ».

- 2ème table : données clients dont nous avons besoin.

Pour cette table nous avons récupéré les IDCLIENT, MAGASIN (pour jointure avec la 3ème table), CODEINSEE (pour jointure avec la 1ère table), VILLE, LATITUDE et LONGITUDE.

Nous constatons qu'il y a un total de 35.221 clients sans coordonnées GPS, cela est dû aux 27.491 clients qui n'ont pas un code INSEE renseigné et aux 7800 clients résidant en Suisse, donc il n'y a pas de code INSEE ou d'adresse connue. Nous aurions pu associer ces clients au code INSEE 74208 (en Haute-Savoie) pour avoir une distance approximative.

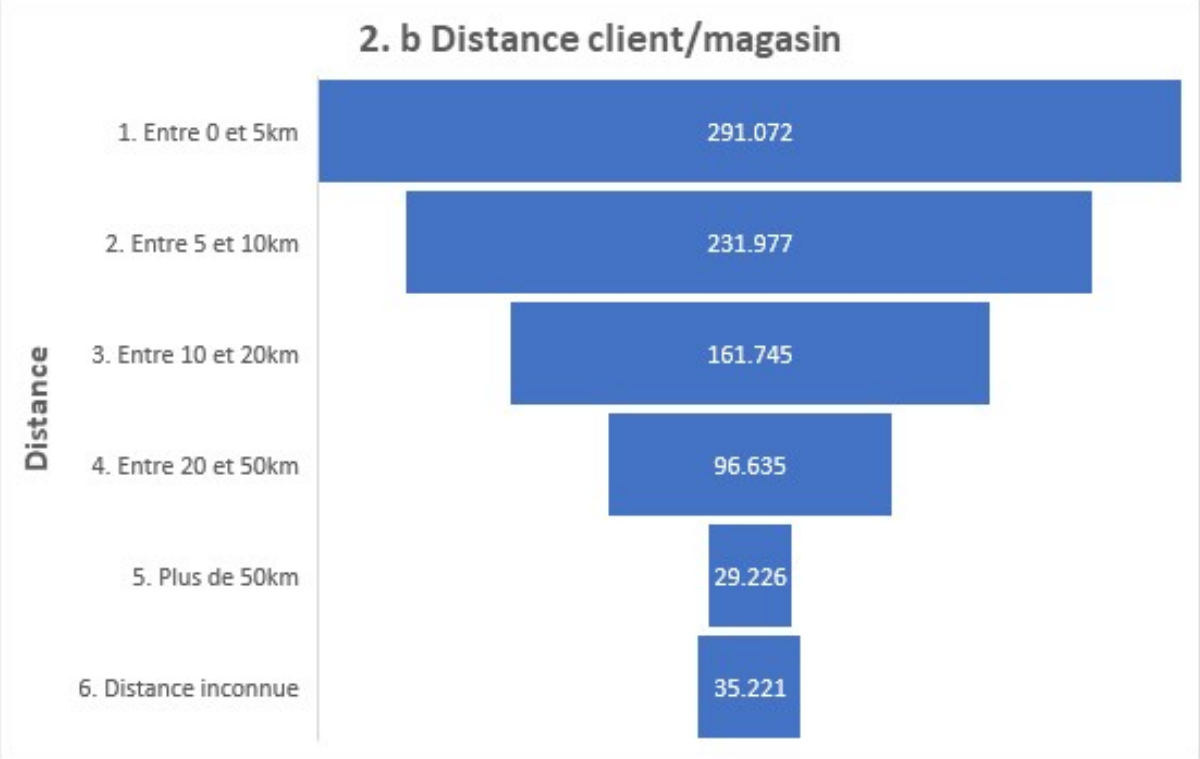
- 3ème table : données magasin dont nous aurons besoin.

Pour cette table nous avons récupéré les CODESOCIETE (pour jointure avec la 2ème table), VILLE(pour jointure avec la 1ère table), MAG_libelledepartement, COMMUNE(pour jointure avec la 1ère table), LATITUDE, LONGITUDE et GPS_libelledepartement.

Nous avons gardé les libellés département des deux tables (1ère et 3ème) pour pouvoir avoir les bonnes coordonnées GPS des magasins qui se trouvent dans des communes qui ont plusieurs libellés département (code postal) sur la 1ère table (INSEE et GPS).

Une fois ces trois tables créées et les données harmonisées, nous avons fait une jointure finale en gardant les données IDCLIENT et ses coordonnées ainsi que MAGASIN et ses coordonnées, dans laquelle nous avons appliqué la formule que nous avons codée, vérifiée au préalable.

| Distance | Nombre de Clients | Pourcentage |
|---|---|---|
| 1. Entre 0 et 5km | 291.072 | 34% |
| 2. Entre 5 et 10km | 231.977 | 27% |
| 3. Entre 10 et 20km | 161.745 | 19% |
| 4. Entre 20 et 50km | 96.635 | 11% |
| 5. Plus de 50km | 29.226 | 3% |
| 6. Distance inconnue | 35.221 | 4% |
| **TOTAL** | **845.876** | **100%** |




*Dataviz sous Excel*

# 3) Etude par univers

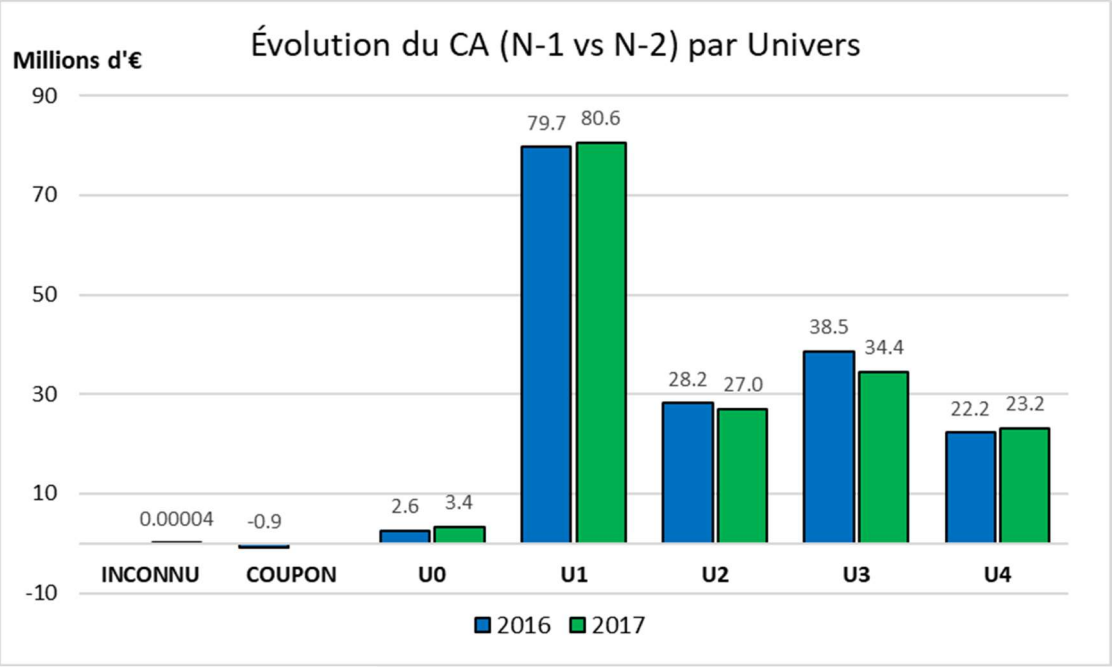
## a) Etude par univers

***Enoncé :***

*Constituer un histogramme N-2 / N-1 évolution du CA par univers.*

Nous avons souhaité récupérer les codes univers mêmes manquants de tous les articles achetés.
Nous avons donc fusionné la table lignes_ticket avec ref_article à l'aide d'un LEFT JOIN.
Les codes univers tagués « INCONNU » correspondent donc aux codes univers non liés à un article mais que nous avons souhaité garder dans notre analyse.
Seuls les articles vendus et présents dans lignes_tickets sont pris en compte.

Dans le graphique, nous pouvons voir que le CA ne varie pas beaucoup d'une année à l'autre.
L'Univers 1 est le plus grand parmi les autres.
Dans l'année 2016 (N-2), l'Univers Coupon atteint environ -0,9 millions €, ce qui a notre avis peut représenter les remboursements et les réductions.
Pour l'année 2017 (N-1), nous avons identifié une nouvelle catégorie dans l'Univers que nous appelons INCONNU, car elle n'appartient pas à aucun autre code. Ce code Univers ne représente que 39€.



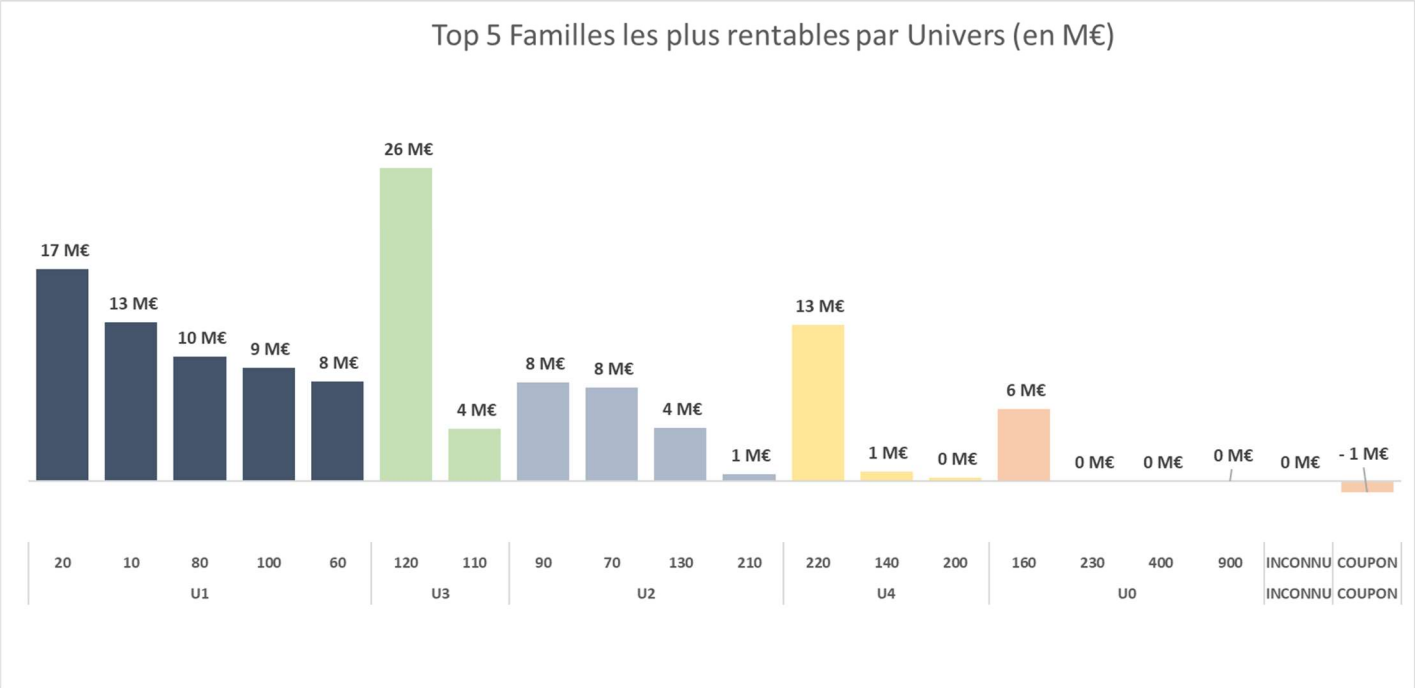

*Dataviz sous Excel*

## b) Top par univers

***Enoncé :***

*Afficher le top 5 des familles les plus rentables par univers (en fonction de la marge obtenue) (tableau ou graphique -> au choix).*

Nous avons souhaité récupérer les codes univers et famille mêmes manquants de tous les articles achetés et figurant dans la table lignes_ticket.
Nous avons donc fusionné la table lignes_ticket avec ref_article à l'aide d'un left join.
Les codes univers et famille tagués « INCONNU » correspondent donc aux codes univers et codes famille non liés à un article mais que nous avons souhaité garder dans notre analyse.
Seuls les articles vendus et présents dans lignes_tickets sont pris en compte.




*Dataviz sous Excel*