**)àç\_/uuuuu/uu/u/uu/Données récupérer la nuit**

Ava\_contact : id clients

Ava\_entent : référentiels

Ava\_ligne : tickets, montant

Ava\_magasin : référentiel nom du lagasin

Ava\_produit : référentiels produit

A\_Stage\_2024Dbo.AVA\_contact dedup selon le prénom, nom

L80 calculer une distance entre de chaine deux caractère

Anonymisation : plus de nouvelle de client , vidage de son prénom , adresse , garder que l’identifiant et transactions \*

Dbo.INFO\_decile : score et classement des clients

Dbo.REF : standardiser les données pays

Transcodage : standardiser toutes les données

**Stage2024 :**

**Id\_ava propre identifiant que avanci créer**

**Alim ava contact**

**Procédures stockées :**

**Data Alim\_ava\_contactgrip**

**Genrer une requête insert**

**Next value for**

**Id-ava\_contact\_intital**

**Id\_ava\_contact**

**Expl : Le 1 c’est celui qu’on veut garder id contact\_initial**

**Client1<-client 1**

**client2<- client 1**

**Client3<-client 1**

**Code\_source\_contact : identifiant qui provient des clients caractère ou autre chose**

**Source : exemple erp**

**Type\_contact : prospect- Compte sans achat – ‘client ‘**

**Src\_nom :nom avant retraitement**

**Ava\_histo\_email**

**Ava\_histo\_email**

**Retraitement d’emails et voir si ‘email existe**

**Domaine du regroupement yahoo**

**Date traitement externe : 0bounce**

**Created\_dt : plusieurs email sont créer**

**Is\_dernier = garder le dernier email remonté a la table contact**

GyuvBbuUCxUIu09ZJj2K

import streamlit as st

import folium

from streamlit\_folium import folium\_static

import pandas as pd

from geopy.geocoders import Nominatim

import time

# Titre de l'application

st.title("Carte de France")

df=pd.read\_csv(r'C:\Users\y.dir\Desktop\Stage\_2024\ava\_contact\_final.csv')

df1=df[:100]

initial\_lat, initial\_lon = df1.loc[0, 'geo\_point\_2d'].split(', ')

# Create a Folium Map centered on the initial coordinates

m = folium.Map(location=[float(initial\_lat), float(initial\_lon)], zoom\_start=6)

# Add markers for each point in the DataFrame

for index, row in df1.iterrows():

    lat, lon = row['geo\_point\_2d'].split(', ')

    folium.Marker([float(lat), float(lon)]).add\_to(m)

# Render the map

folium\_static(m)

[**https://callendar.medium.com/tuto-cr%C3%A9er-une-carte-interactive-avec-python-et-folium-a3d56001e75a**](https://callendar.medium.com/tuto-cr%C3%A9er-une-carte-interactive-avec-python-et-folium-a3d56001e75a)

[**https://zenn.dev/shimat/articles/4aa5062d78c483**](https://zenn.dev/shimat/articles/4aa5062d78c483)

[**https://github.com/okld/streamlit-elements?tab=readme-ov-file**](https://github.com/okld/streamlit-elements?tab=readme-ov-file)

[**https://docs.streamlit.io/develop/api-reference/layout**](https://docs.streamlit.io/develop/api-reference/layout)

avancisqldev01.domavanci.local)  
login : A\_avancisqldev01  
mdp : Y8tlSduh01L3CpsxzCvi

**8E\*VYuu#PvBpSC0l**

**serveur avancisql05  
base Score\_IA\_Telco  
user A\_Score\_IA\_Telco**

**mdp pHoC66NC1PiQNEr43EXR**

le deuxième tableau contient l’identifiant entête pour chaque client qui rassemble toute les transactions faite par ce dernier ,on y trouve également le code source contact, et la date de création du contact ,le type de canal de vente (Ecomerce ou Magasin), le montant total de ses achats qui fait la somme des achats (somme des achats reporteriez sur le troisième tableau ), le montant remise sur les livraisons , le montant net ttc sur des commandes

**dans le troisième tableau sont répertoriés les achats détaillés des clients par commandes, chaque ligne contient le numéro de la ligne , le code source de l’entête , le code source du produit, le montant de l’achat réalisé et la quantité achetée , il est important de noter que un entête peut correspondre à plusieurs lignes dans le tableau des lignes .**

**Implémentation sur le server au sein de MYSQL**

**Une fois que les données sont implémentés, des procédures stocké doivent être exécutée pour le traitement et le nettoyage des données ainsi que la génération de nouveaux tableaux comme la segmentation et le comportement d‘achat.**

**Parmi ces Tableaux, on trouve celui des adresses postales qui traite les cordonnées telle que de le code postal ,la ville et les cordonnés x et y , un tableau es dédiée au traitement des email incluant l’email retraité , un champ indiquant le domaine , ces données sont envoyé vers un organisme externe indiquant l’email si l’email est optin et joignable ainsi qu’une indication s’il s’agit du dernier email utilisé en cas de défaillance de l’email un autre tableau traite les emails qui sont pas valide en indiquant le detail.**



**un autre tableau est consacré pour le traitement des numéro de téléphones mobile contenant le numéro de téléphone retraité le type de téléphone (fixe ou mobile ) , la validité et la joignabilité du numéro de téléphone qui sont traité par un organisme externe .**

**Tableau de segmentation :**

Une image contenant texte, capture d’écran, nombre, Police

Description générée automatiquement

**C’est un tableau de segmentation qui contient la période, l’identifiant contact , le score du premier semestre et deuxième , le libellé de segment (client VIP, bon client…) , si le client est actif dans la période.**

**La segmentation est basé sur le semestre et les critères du comportement d’achat des client parmi ces critères on trouve le classement des clients actifs par semestre et par ordre décroissant selon le chiffre d’affaire. ensuite ces clients sont regroupée de 1 à 10 selon le CA , 1 étant ceux qui ont le plus fort CA ensuite chaque client est rattaché avec un score 1,2,3 , les clients avec un score 3 sont ceux qui ont le plus de CA , pour les deux semestre dans l’année un score est attribué pour permettre de prédire le comportement d’achats de l’année suivante enfin les segments sont déterminée à partir du regroupement des scores.**

Une image contenant texte, capture d’écran, logiciel, nombre

Description générée automatiquement

**La méthode de ségmentation RFM**

La méthode de segmentation RFM, pour Récence-Fréquence-Montant, a été popularisée dans les années 60 avec l’émergence de la vente par correspondance.

Les variables prises en compte sont immuables depuis 60 ans et se fondent sur le **comportement transactionnel des clients**. Mon client a-t-il acheté récemment, à quelle fréquence achète-t-il, quels sont ses critères, quel type de produit, et pour quel montant ?

**On obtient toujours une bonne grille de lecture avec des sous-ensembles de clients à comportements d'achat homogènes**. Les données clients sont faciles d’accès car logées dans un système de caisse, ou un seul back-office, le plus souvent unique, avec des données souvent bien structurées car peu nombreuses et très standardisées.

Cette première approche permet de se familiariser avec sa base de données (BDD), de **comprendre les ressorts de fidélisation, d’imaginer le cycle de vie de ses clients, et les moments-clefs.** En revanche, il manquera les**données comportementales des différents canaux d’achat online et offline**, ainsi que les**données démographiques**. Celles-ci sont essentielles pour rajouter de l'humain, du profiling, du prédictif dans un schéma d’achat par segments RFM parfois stéréotypé ou trop simpliste.

Chez AVANCI la segmentation à été améliorer pour prendre en compte la dimension démographique, omnicanale, digitale, et sociale des clients ou des prospects.

Les pré-requis à l’hyper-segmentation . : L’hyper-segmentation se fonde sur une donnée bien structurée. En effet, les nombreux canaux de communication, mettent en évidence une difficulté pour les marques à unifier la donnée client dans un référentiel client unifiée avec des données clients propres et exploitables par une entreprise.

Cette problématique data peut se résumer autour des 4V

La vitesse : ma BDD est-elle capable de gérer la vitesse d'acquisition des données ?

La variété : ma BDD est-elle capable de gérer une telle variété de données ?

Le volume : ma BDD est-elle capable de gérer un tel volume de données ?

La véracité : ma BDD me donne-t-elle des information vraies ?

Et c’est là tout l’enjeu de la [**mise en place d’une CDP**](https://www.avanci.fr/blog/actualites-data/pourquoi-mon-entreprise-a-t-elle-besoin-dune-cdp/). En effet, les données marketing ne se gèrent pas comme des données de gestion classiques Il faut des systèmes souples, adaptables, capable de gérer de grande quantités de données et intégrant des chaînes de traitement garantissant que les groupes de segment de données soient valides, fiables, pertinentes, utiles et permettent une vision 360° du client. Ce domaine est souvent négligé et compromet la mise en place d’une bonne segmentation et [**la réussite des projets CRM**](https://www.avanci.fr/expertise/strategie-crm/).

Chez AVANCI ils ont utilisé des chaînes de traitements de données complexes en utilisant le machine learning.  Ils sont destinés à garantir des modèles statistiques robustes, , une segmentation fine et des modèles de données adaptés à chaque marque et chaque marché .

L’hyper-segmentation Avanci est une **analyse comportementale fondée sur l’étude du parcours client omnicanal, du premier point de contact à l’achat final, enrichie par une grille de lecture démographique.**

Cette segmentation enrichie de nouvelles variables permet alors une analyse plus fine et plus granulaire. Elle donne des clés de lecture fondamentales à la compréhension des parcours prospects et clients, en ajoutant de l’humain.

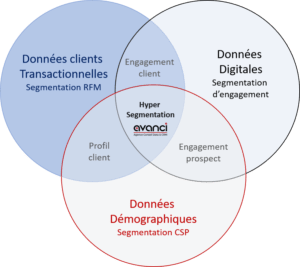
L’hyper -segmentation peut s’expliquer à travers un diagramme de Venn

Qui comprend 3 ensemble :

Les données transactionnelles offrant une segmentation RFM fondées sur un comportement d’achat.

Les données digitale : offrant une segmentation d’engagement , fondées sur un comportement d’engagement

3. les données démographiques offrant une segmentation CSP , fondées sur un comportement socio démographique , avec des socio-types



C’est un tableau de segmentation qui est basé sur le semestre et les critères du comportement d’achat des client parmi ces critères on trouve le classement des clients actifs par semestre et par ordre décroissant selon le chiffre d’affaire. ensuite ces clients sont regroupée de 1 à 10 selon le CA , 1 étant ceux qui ont le plus fort CA ensuite chaque client est rattaché avec un score 1,2,3 , les clients avec un score 3 sont ceux qui ont le plus de CA , pour les deux semestre dans l’année un score est attribué pour permettre de prédire le comportement d’achats de l’année suivante enfin les segments sont déterminée à partir du regroupement des scores.

Tableau media\_invalide \_etail verification email et detail de email qui ont pas été retenu abuse ou don email ou null

Cette hyper-segmentation est parfaitement utilisée par les responsables d’études CRM car ses résultats sont à vocation directement opérationnels ils permettent de mettre en place des actions marketing et commerciales, individualisées et prédictives à destintion du client ou du prospect . grâce aux données scorées , parfois en temps réel , l’activation des données devient naturellement plus simple .

**Les avantages du scoring et de la segentation**

**.**

**Le scoring client :** est une technique consistant un score à chaque client **.**

c’est une méthode qui permet d’établir un classement ou une hiérarchie claire au sein d’une clientèle , en fonction de critères et d’objectifs variables selon les secteurs des entreprises :

dans le domaine du marketing . le score a généralement pour fonction d’évaluer la probabilité d’achat de chaque client. Il se base entre autres sur l’historique d’achat et sur des données socio-démographiques.

Dans le domaine de finance de la banque, le score est plutôt destiné à estimer le niveau de risque représenté par un client.<< Le crédit scoring >> par exemple détermine la qualité et la solvabilité d’un emprunteur , un assureur , de même peut classer ses clients en fonction du taux estimé de sinistralité (probabilité de réalisation du risque)

Le principal intérêt du scoring est d’intégrer la notion de la valeur client dans la stratégie marketing , il permet d’identifier les clients à plus fort potentiel pour concentrer les outils, les efforts, les dépenses sur eux , en d’autre terme , le scoring est un outil puissant pour optimiser le roi (retour sur investissement ) des différentes compagnes,

Un scoring de bonne qualité doit évidemment s’appuyer sur une bonne base de connaissance client bien organisée, et sur une segmentation fine .

**Segmentation client : un préalable indispensable au scorin**

La segmentation de base client consiste à diviser une base clientèle en plusieurs catégorie présentant des caractéristiques homogènes. Elle a pour objectif d’améliorer

La connaissance client de l’entreprise , en distinguant mieux les besoins et les attentes de chaque cible marketing.

En pratique , il est possible de distinguer quatre grandes méthodologie de segmentation client :

**La segmentation socio-démographique** réparti la base clientèle par tranches d’âge, par genre, par composition familiale .

**La segmentation psychologique** privilégie une classification des profils en fonction de leur centres d’intérêt , de leurs style de vie , de leur valeurs personnelles ou encore

De leurs traits de caractère.

**La segmentation comportementale et transactionnelle :**

S’intéresse essentiellement aux habitudes d’achat cela peut inclure la fréquence des transactions , le montant moyen du panier, le canal d’achat privilégié , le mode de paiement préférer ou le type de livraison .

**La segmentation géographique** : qui permet de catégoriser les clients en fonction de leur zone de résidence(pays, région,ville). Cette dernière , en effet peut avoir un impact sur les besoins du client sur ses habitudes d’achat ou encore su sa fidélisation.

**Quelle méthode pour segmenter une base et définir un score**

Plusieurs méthodes coexistant et peuvent présenter des spécifités , avec l’application de formules mathématiques et de règles de pondération différentes.

La plus connue est probablement le scoring **RFM** , dont le principe et d’agréger trois taux :

La récence (R) : c’est-à-dire le caractère plus ou moins du dernier achat effectué par le client

La fréquence : c’est-à-dire le nombre de commandes réalisées sur une période de temps donnée

Le montant , c’est-à-dire la valeur du panier moyen ou celle du dernier achat réalisé

Le score **RFM**  obtenu permet de classer les clients en neuf grandes catégories , soit dans un ordre de valeur croissant :

**Les clients perdus depuis longtemps : ou client inactif**

**Les clients on convertis (un seul achat non renouvelé)**

**Les clients régulier mais perdu récemment (interruption des achats ) : ou décrocheurs**

**Les clients acquis récemment et à faible pouvoir d’achat**

**Les clients acquis récemment et à fort pouvoir d’achat**

**Les clients réguliers dont la fréquence d’achat diminue**

**Les clients réguliers dont la fréquence d’achat augmente**

**Les clients réguliers à fort pouvoir d’achat**

Le scoring des clients est l’une des applications concrète les plus intéressantes d’une bonne stratégie CRM . il permet d’identifier les segments de profils avec certains critère : les plus actifs et les plus rémunérateurs .

**Streamlit Interface**

**Pour commencer le développement de l’application il était nécessaire d’ouvrir un environnement de travail et d’installer les bibliothéques nécessaire telles que streamlit,plotly , matplotlib, numpy pandas,pyodbc afin d’ assurer le bon fonctionnement de l’application .**

**Extraction des donnée du server**

**Pour commencer L’interface streamlit j’ai créer un fichier dédié à l’extraction des donnée en utilisant le décorateur st.cache , ce décorateur , propre à streamlit , permet de stocker les résultats d’une fonction en cache afin d’optimiser le temp**

**de traitement ainsi que les résultats sont renvoyés immédiatement sans avoir exécuté à nouveau la fonction**

**Pour se connecter a la base sql avec python j’ai utilisé pyodbc ,un module python qui permet d’accéder aux bases de données via ODBC (Open Data base connectivity ) sur MYSQL , une fois connecté à la base de donnée j’ai récupérer les donnée de maniéré optimisée .parmi ces données on trouve le tableau des contacts , qui contient l’identifiant contact ,le pénom , nom, sexe, genre, le type de contact (prospect ou client ), la date de naissance , Age**

**Le tableau de la segmentation .Le tableau des adresses : contenant l’identifiant contact , le code postale retraité , la ville retraité , les coordonnées x et y**

**Le tableau des lignes ,le tableau des emails et téléphone retraité Et enfin le tableau des media\_invalide .**

**les entêtes , le tableau de la segmentation, le tableau des adresses,**

**Le tableau des lignes , le tableau des emails , le tableau des téléphone , le tableau des media invalide .**

**L’application comprend six onglet disposés sur bar de navigation de streamlit et affiche un tableau de bord interactif, Home , carte géographique , Audience , Segmentation , Data quality , IA**

**Pour améliorer l’aspect visuel j’ai intégré du css afin de personnaliser les couleurs du text-color et de l’arrière ,de plus, j’ai ajouté le logo avanci sur la bare latérale garantissant une expérience utilisateur agréable.**

**Métriques affichées :**

**Total CA (chiffre d’affaires) : Affiche le chiffre d'affaires total actuel qui égale la somme des montant ttc . La valeur est mise en forme en euros (€{Total\_CA\_actu}), et un changement de pourcentage par rapport à l'année précédente est également affiché. Si ce changement est positif, il est affiché en vert, sinon en rouge.**

**Total Clients : Montre le nombre total de clients actifs actuels. Le changement de pourcentage par rapport au semestre précédent est aussi affiché.**

**Panier Moyen : Représente la valeur moyenne des commandes effectuéesv par les clients, total de chiffre d’affaire divisé par le total des commandes exprimée en euros. Le pourcentage de changement sur le semestre est indiqué.**

**Fréquence : Indique la fréquence d'achat, c'est-à-dire le nombre moyen de transactions par client sur une période donnée , le total des commande sur le total d’actifs . Le pourcentage de variation semestrielle est également mentionné.**

**Graphiques**

les graphiques en barres et les jauges sont des outils visuels puissants pour représenter les données de manière claire et intuitive. Ils permettent aux utilisateurs de comprendre rapidement les tendances, les performances et les écarts par rapport aux objectifs.

Ces graphiques sont travaillé avec la library plootly et comprennes un seule filtre pour les trois pour filtrer l’année et le semestre.

**Évolution du Chiffre d’Affaires (CA) et la Fréquence :** Ce graphique en barres illustre l’évolution de la fréquence et le total des commandes sur une période définie. Cette visualisation permet d’identifier facilement les périodes de forte ou faible activité, facilitant ainsi l'analyse des tendances et des performances financières.

 **Évolution du Panier Moyen :** Ce graphique montre comment le panier moyen, défini comme le chiffre d'affaires total divisé par le nombre total de commandes, varie au fil du temps en fonction du chiffre d’affaire par année et par semestre . En observant ce graphique, les utilisateurs peuvent détecter des tendances dans les comportements d'achat des clients, comme une augmentation ou une diminution du montant moyen dépensé par transaction.

 **Évolution du CA par Semestre :** Ce graphique présente le chiffre d'affaires réparti par semestre. Il permet de comparer les performances financières d'un semestre à l'autre et d’évaluer l’impact des différentes stratégies ou événements commerciaux sur les revenus.

Le graphique de la jauge évalue la proportion du chiffre d'affaires total générée par un certain pourcentage des meilleurs clients. La méthodologie commence par trier les clients par montant net TTC en ordre décroissant, permettant ainsi de focaliser l'analyse sur ceux qui génèrent les revenus les plus élevés. Ensuite, un pourcentage spécifique, défini par l'utilisateur via un filtre, est utilisé pour déterminer combien des meilleurs clients doivent être inclus dans l'analyse. Le nombre de clients correspondant à ce pourcentage est calculé en multipliant le pourcentage par le total des clients dans la liste triée. Le chiffre d'affaires total généré par ces clients est ensuite calculé en additionnant les montants nets TTC de ces clients ciblés. Enfin, la proportion de ce chiffre d'affaires par rapport au chiffre d'affaires total est déterminée, et cette proportion est affichée sur la jauge, offrant une visualisation claire de la contribution financière des meilleurs clients.

**Onglet Home**

**C’est un onglet qui a pour but de faciliter l’accès et la vision aux clients pour voir Leur résultat de business d’une manière simple avec des graphiques automatisé et filtrable à temps réel .**