



HAI708I - Entrepôt de données et big data

Rapport du projet d'entrepôt de données

Apple Store



ALMALLOUHI Mohamad Satea (22011503)
DAYIOGLU Gurgun (22024990)
BARRERE Loreena (22206698)
RAZAFIMBAHOAKA Thomas (21912146)

Sommaire

1. Analyse du cas considéré.....	2
1.1. Les objectifs de l'entreprise Apple.....	2
1.2. La position d'Apple sur le marché.....	2
1.3. Les services ou produits proposés.....	2
1.4. Les formes de revenu.....	2
1.5. Les informations utiles pour la prise de décision au sein de l'entreprise.....	2
2. Actions et opérations.....	4
2.1. Gestion des ventes :.....	4
2.2. Gestion des stocks :.....	4
2.3. Gestion des livraisons :.....	5
3. Traitements possibles.....	5
3.1. Gestion des ventes :.....	5
3.2. Gestion des stocks ("capture d'écran" toutes les 24 heures).....	5
3.3. Gestion des livraisons.....	6
4. Ordre d'importance et rentabilité.....	6
5. Actions / opérations les plus importantes.....	7
6. Data-mart.....	7
6.1. L'action / opération la plus importante : vente.....	7
6.2. La deuxième action / opération : stock.....	8
7. Dimensions des modèles.....	10
7.1. Gestion des ventes.....	10
7.2. Gestion des stocks.....	11
8. Analyse de notre modèle en fonction des traitements.....	12
8.1. Modèle de la gestion des ventes.....	12
8.2. Modèle de la gestion des stocks.....	12
9. Test de la pertinence de notre modèle.....	13
10. Estimation de la taille des tables de l'entrepôt.....	15
11. Implémentation en ORACLE des tables des faits et des vues.....	17
11.1. Utilisation des vues virtuelles pour les dimensions partagées.....	17
12. Requêtes analytiques.....	17
Ressources / Annexes.....	25

L'analyse

1. Analyse du cas considéré

1.1. Les objectifs de l'entreprise Apple

Nous avons choisi de travailler sur l'entreprise Apple. C'est une entreprise américaine qui crée et commercialise des produits électroniques grand public. L'objectif de notre entreprise est de commercialiser un maximum de produits, afin de faire un maximum de chiffre d'affaires. La qualité des produits Apple a été prouvée de nombreuses fois, cette entreprise se trouve donc dans le haut du tableau dans son domaine. Sa mission principale est de garder sa stabilité sur le long terme en restant à la pointe de la technologie.

1.2. La position d'Apple sur le marché

Apple fait partie des 5 premiers fabricants et vendeurs de smartphone au monde depuis 2009. Le concurrent direct d'Apple est l'entreprise Samsung, une entreprise coréenne travaillant dans le même secteur d'activité. Depuis plusieurs années se poursuit un combat entre ces deux leaders afin d'avoir la première place sur le marché. En effet, dès qu'un nouveau produit sort dans une de ces entreprises, celle-ci passe facilement en tête du marché dans ce domaine. Apple se trouve régulièrement à la première ou à la deuxième place, suivie ou précédée de Samsung, en fonction des dernières sorties en termes d'appareil électronique.

1.3. Les services ou produits proposés

Parmi les produits qu'Apple propose sur le marché, nous retrouvons des téléphones portables (appelés "iPhone"), des ordinateurs personnels (appelés "Mac"), des tablettes tactiles (appelées "iPad"), et d'autres outils électroniques tels que des écouteurs sans fils (appelés "AirPods") ou bien des montres connectées (appelées "Apple Watch"). L'objectif de cette entreprise est donc de vendre ses produits, tout en ayant des hauts standards de qualité.

1.4. Les formes de revenu

Sa principale forme de revenu est liée à la vente de ses produits électroniques. Afin de les commercialiser, Apple utilise différentes stratégies. Les produits électroniques sont vendus de plusieurs façons. Il est possible de les acheter en ligne, sur le site officiel d'Apple mais également sur d'autres sites internet. Il est également possible d'acheter les produits Apple en physiques dans différents magasins, comme des magasins de matériel multimédia tel que la Fnac ou Darty. Mais encore dans des magasins de téléphonie mobile, comme orange ou free. Cependant, pour notre projet, nous allons nous intéresser aux produits Apple vendus directement dans les magasins Apple, appelés Apple Store. En 2019, l'entreprise compte 506 Apple Store basés sur 25 pays, et dans lesquels travaillent plus de 65 000 employés. Pour l'année 2020, l'entreprise compte un chiffre d'affaires atteignant les 111.4 milliards de dollars.

1.5. Les informations utiles pour la prise de décision au sein de l'entreprise

Afin d'évaluer au mieux les besoins de notre entreprise, voici quelques informations qui pourraient lui être utiles.

Tout d'abord, il est primordial d'avoir un visuel clair sur les ventes des produits électroniques.

Il serait intéressant de connaître le nombre de téléphones vendus par année, ou bien par magasin. Nous pourrions également regarder les ventes en fonction de la génération du téléphone (Iphone 14, Iphone 15). Lié à cette information, nous pouvons avoir une information concernant le chiffre d'affaires obtenu suite à la vente d'une génération d'Iphone. Nous avons pris comme exemple l'Iphone (les téléphones portables vendus par Apple), mais il en est de même pour les ordinateurs portables, les tablettes tactiles, et plus généralement, tous les produits vendus par Apple.

Nous pouvons également avoir une information concernant la date des ventes. En effet, savoir si nous avons plus de produits à une certaine période permet de prévoir les ventes pour l'année prochaine. Par exemple, lors des périodes de Noël, ou bien pendant les périodes de soldes, les appareils électroniques doivent se vendre davantage. Sur une plus longue période, il est possible d'analyser si au cours du temps les ventes d'Apple sont en croissance ou en déclin.

Pour augmenter ses ventes, il est important qu'Apple ait des informations sur ses clients, par exemple son âge, son sexe. Ainsi, les futurs appareils pourront être désignés en fonction de la tranche d'âge des futurs clients qui risquent d'être intéressés par ce produit. Il pourrait également y avoir des mises à jour pour certains produits en fonction de la tranche d'âge des utilisateurs. Le sexe permettrait de cibler les produits et ainsi augmenter les ventes de l'entreprise. Il sera également utile pour mettre en œuvre des campagnes de marketing ciblé.

De plus, il serait utile d'avoir des informations concernant les magasins (Apple Store). Nous pourrions notamment connaître le nombre de vente fait dans un magasin. De cette façon, si nous remarquons qu'un magasin vend peu, nous pourrions nous questionner sur l'utilité de ce magasin dans cette zone géographique. Dans le même état d'esprit, si nous remarquons qu'il y a beaucoup de vente dans un magasin, nous pourrions réfléchir à l'agrandir, ou bien en créer un deuxième dans la même zone géographique. Nous pourrions aussi regarder le nombre de ventes par rapport au nombre d'employés travaillant dans ce magasin, ceci par magasins.

Toujours en rapport avec les ventes d'un produit, il serait intéressant de connaître le vendeur ayant réalisé la vente. Avoir des employés qualifiés et professionnels permet d'augmenter les ventes de l'entreprise. Savoir donc si un employé ne vend pas permettrait d'organiser des formations ou des remises à niveaux pour les aider à s'améliorer. Nous pourrions aussi motiver les employés avec des primes, et ainsi savoir quel employé fait les meilleurs chiffres. Cela pourrait accroître la motivation et la productivité des employés, ce qui se refléterait dans la vente des produits.

Apple est attentif également à la qualité de ses produits. La satisfaction client est une priorité. Il est donc nécessaire pour l'entreprise de connaître les notes attribuées à chaque produit. En effet, lorsqu'un client achète un produit, il a la possibilité de déposer une note sur ce produit. Nous pourrions donc regarder quels sont les produits les plus appréciés par mois, ou bien par année, afin de savoir quels produits risquent d'être le plus vendus. Cela permettra de savoir également quelles sont les choses appréciées des consommateurs, et donc, diriger au mieux les prochaines sorties d'appareils électroniques.

Il est également primordial d'avoir des informations sur les stocks. Il est important de maintenir un stock minimum afin d'avoir toujours des produits à vendre. Si l'entreprise se trouve dans une situation de rupture de stock, cela voudra dire que l'entreprise ne peut pas vendre les produits qu'elle ne possède pas, elle perd donc du chiffre d'affaires. Nous pouvons donc nous demander quels produits sont en rupture de stock régulièrement, afin d'en commander plus régulièrement. Ou bien, nous pouvons nous demander quels produits restent longtemps dans la zone de stock, ainsi, il faudrait en commander moins et garder de la place dans le stock pour les produits qui se vendent plus.

Les stocks seront mis à jour une fois par jour. Nous pourrions nous demander s'il y a une certaine période dans chaque mois où les stocks sont au plus bas. Par exemple, nous pouvons imaginer que les clients achètent des appareils électroniques plutôt en début du

mois. Ainsi, nous pouvons prévoir d'augmenter les stocks durant cette période afin de ne pas être en rupture. Nous parlons ici de regarder les jours dans un mois, mais nous pouvons envisager la même idée en regardant les mois dans une année. Nous pouvons regarder à quelle période de l'année les stocks sont les plus vides, et ainsi, prévoir plus de stocks pour certains mois (probablement le mois de décembre).

Nous aurons aussi une information concernant les stocks des produits. De cette manière, nous pouvons savoir quels produits sont le plus souvent en rupture de stocks, et ainsi prévoir d'augmenter leur stocks, ou leur production. Il serait également intéressant de connaître le nombre de stocks en fonction du type de produit. En effet, un ordinateur portable ne prend pas la même place en guise d'espace dans l'entrepôt que des écouteurs.

Nous avons également connaissance du magasin dans lequel le stock est compté. Ainsi, si un magasin est trop souvent en rupture de stock, il serait judicieux d'augmenter ses stocks, en agrandissant la zone de stockage par exemple. A l'inverse, si un magasin a des difficultés à écouler ses stocks et donc ne peut pas recevoir des produits de dernière génération, il sera possible d'envoyer ses stocks à d'autres magasins ayant moins de difficulté à écouler ses stocks.

Enfin, il est possible d'acheter un appareil neuf ou reconditionné. Il pourrait être intéressant de connaître l'état des smartphones en stock. En effet, il faut s'assurer d'avoir toujours une certaine quantité d'appareils neufs, et d'appareils reconditionnés. Nous ne pouvons avoir que des appareils reconditionnés, car il est possible que certains clients souhaitent uniquement acheter du neuf, et vice-versa.

On peut également avoir des informations sur les livraisons, et donc, se questionner sur la livraison des produits depuis les entrepôts. Si un magasin voit que ses stocks se vident, elle va recevoir une livraison avec les produits qui lui manquent. De cette manière, elle aura plus de produits à vendre, et pourra augmenter son chiffre d'affaires.

2. Actions et opérations

Voici les actions (ou opérations) à tracer pour récupérer ces informations :

2.1. Gestion des ventes :

Le choix de l'action de gestion des ventes, démontre une approche stratégique axée sur la compréhension approfondie des habitudes d'achat et des performances de vente. En analysant les données de manière détaillée, l'entreprise pourra évaluer la dynamique des ventes au fil des années, identifier les variations de performance entre les magasins, et cibler plus efficacement ses segments de clientèle en fonction des tranches d'âge. De plus, le calcul du nombre de ventes par type de matériel chaque année offre une perspective sur la popularité des produits, facilitant ainsi les décisions stratégiques liées à l'offre et à la demande. Cette approche analytique enrichie permettra à l'entreprise d'optimiser ses opérations, d'anticiper les tendances du marché, et d'ajuster sa stratégie de vente pour répondre au mieux aux attentes des clients.

2.2. Gestion des stocks :

Le choix de surveiller les stocks à travers des captures d'écran quotidiennes est motivé par la nécessité d'obtenir une vue sur la disponibilité des produits. En évaluant les ruptures, l'entreprise peut affiner ses stratégies logistiques pour répondre efficacement aux fluctuations de la demande. La surveillance des produits en rupture fréquente permet d'optimiser les niveaux de stock, améliorant ainsi la satisfaction client et augmentant les opportunités de vente. En suivant l'évolution des stocks par type, cette approche offre un aperçu des préférences des clients, permettant à l'entreprise d'ajuster ses

approvisionnements de manière stratégique pour minimiser les coûts tout en maintenant des niveaux adaptés à la variabilité de la demande.

2.3. Gestion des livraisons :

La gestion des livraisons s'articule autour de plusieurs axes stratégiques. Tout d'abord, nous visons à évaluer la performance des fournisseurs par le biais du nombre de livraisons par magasin. Cela permet à l'entreprise de mieux évaluer la fiabilité des fournisseurs, favorisant ainsi des relations solides. Cela permet également de faire des ajustements en cas d'écart importants entre les prévisions et les livraisons réelles. En optimisant les distances de livraison en fonction des villes, des fournisseurs et des magasins, l'entreprise contribue à réduire son impact environnemental et ses coûts logistiques. Ces analyses intégrées soutiennent une gestion logistique efficace, renforcent les partenariats avec les fournisseurs et privilégient une approche durable des opérations

3. Traitements possibles

Nous avons identifié divers traitements possibles pour gérer les ventes, les stocks et les livraisons. Ces traitements sont cruciaux pour optimiser la performance commerciale et la logistique de l'entreprise. Voici quelques exemples spécifiques de traitements que nous voulons mettre en place en utilisant l'entrepôt de données :

3.1. Gestion des ventes :

- Calcul du chiffre d'affaires par année et par magasin : il est essentiel de connaître les ventes par année pour évaluer la performance des ventes et comparer la performance des magasins.
- Calcul du nombre des ventes effectuées en fonction des tranches d'âge des clients par année : permet d'analyser les achats réalisés par les clients appartenant à chaque tranche d'âge. Cette analyse aide à comprendre les habitudes d'achat et à identifier la tranche d'âge qui génère le plus de ventes chaque année. Cette information est cruciale pour mieux cibler les clients. Nous cherchons à calculer le nombre de produits vendus.
- Calcul du nombre des ventes par type de matériel par année : nous allons calculer le nombre de ventes par type de matériel par année : cette analyse nous permettra de comprendre quels produits sont populaires auprès des clients chaque année et si la sortie de certains produits pourrait influencer d'autres. Ici, nous nous intéressons au nombre de produits vendus.

3.2. Gestion des stocks ("capture d'écran" toutes les 24 heures)

- Nombre de produits en stock par magasin, par mois : cela nous permet d'évaluer le travail des équipes et la stratégie logistique de chaque magasin afin de prévoir les possibles ruptures de stocks à venir.
- Nombre de produits en stock, par produit, par magasin, par mois : en ciblant les produits en rupture fréquente, cette analyse permet d'optimiser les niveaux de stock.

En augmentant le stock des articles les plus demandés, l'entreprise peut maximiser les opportunités de vente, améliorer la satisfaction client et augmenter les revenus.

- Moyenne du stock par produit, par année, par magasin, par état : suivre l'évolution des stocks par type (reconditionné, neuf) offre une compréhension approfondie des préférences des clients. En comparant la moyenne du stock avec les ventes, l'entreprise peut ajuster sa stratégie d'approvisionnement, minimisant ainsi les dépenses de stockage tout en maintenant des niveaux adéquats pour répondre à la demande variée.

3.3. Gestion des livraisons

- Calculer le nombre des livraisons faites par fournisseur par magasin : la requête permet à l'entreprise d'évaluer la performance de chaque fournisseur en termes de nombre de livraisons effectuées et de minimiser les risques opérationnels. Cela aide à identifier les magasins qui sont dépendants sur un fournisseur, qui pourrait mettre le magasin en rupture si le fournisseur n'effectue plus de livraison.
- Calculer le nombre de livraisons par magasin, par fournisseur, par année : afin d'évaluer la fiabilité de chaque fournisseur et de prendre des mesures correctives en cas de différences significatives entre les livraisons prévues et réelles. En comprenant les performances individuelles de chaque fournisseur, l'entreprise peut développer des relations plus solides avec les fournisseurs fiables et travailler sur des améliorations avec ceux qui présentent des écarts.
- Livraison effectuée par ville de fournisseur, par magasin (en voyant sa ville), par année : optimiser la distance entre le fournisseur et le magasin. En minimisant les distances de livraison, l'entreprise contribue à la réduction de son impact environnemental et minimise le coût de livraisons. Moins de kilomètres parcourus signifient une empreinte carbone ainsi que des frais réduits, ce qui s'aligne avec les objectifs de durabilité.

4. Ordre d'importance et rentabilité

Il est nécessaire de se rappeler que l'objectif principal d'Apple est d'augmenter son chiffre d'affaires, et donc, de faire un maximum de vente possible. Il paraît donc évident que l'action vente est la plus importante. Avoir des informations et pouvoir faire des statistiques sur les ventes nous permettra d'augmenter les futures ventes, et donc, d'augmenter le chiffre d'affaires.

Nous pouvons dire que la gestion des stocks est la deuxième idée la plus importante. En effet, une utilisation optimale de l'espace de stockage dans le magasin permet de ne pas être en rupture de stock. Être en rupture de stock est synonyme de perte de chiffre d'affaires pour l'entreprise, garder un œil sur les stocks est donc primordial.

Pour finir, nous pouvons parler des livraisons. La livraison est importante car il est nécessaire de recevoir les produits à temps pour ne pas être en rupture de stock. Une bonne logistique est primordiale pour maximiser le chiffre d'affaires. Cependant, elle vient après l'information sur les stocks car sans l'information sur les stocks, nous ne pouvons pas faire de livraison intéressante.

La conception

5. Actions / opérations les plus importantes

Comme expliqué précédemment, les deux actions les plus importantes sont la vente et la gestion des stocks.

La première action va se baser sur la vente d'un produit, par un client et un employé, dans un magasin à une date précise. Ce sont des données primordiales à analyser afin de faire des statistiques et de maximiser les ventes. Ce sera également la table la plus remplie de notre modèle. Le bon fonctionnement d'Apple est centré sur la vente de ses produits. Avec ces informations, il est possible d'analyser différentes requêtes permettant de mieux comprendre et d'améliorer les ventes par des choix stratégiques. Cela permettrait également de répondre à d'autres objectifs comme cibler les publicités ou encore améliorer les futurs produits.

La deuxième action est la gestion des stocks. Une gestion précise des stocks garantit que les produits sont disponibles lorsque les clients veulent les acheter, évitant les ruptures de stock qui peuvent entraîner une perte de ventes et de clientèle. Elle minimise aussi les coûts associés au sur-stockage, tels que le stockage, la dépréciation, et l'obsolescence, en particulier pour les produits technologiques qui évoluent rapidement.

6. Data-mart

6.1. L'action / opération la plus importante : vente

Voici le data-mart créé pour l'opération la plus importante, qui est la vente. On remarque ici que l'on a un idVente dans la clé primaire. Il s'agit d'une dimension dégénérée. Sans cette dimension, nous aurions pu avoir certains soucis. En effet, sans cette dimension, si un client achète plusieurs produits dans le même magasin le même jour, nous perdons cette information. Or, celle-ci pourrait être intéressante. Si on remarque qu'un produit se vend régulièrement avec un autre, il serait possible de faire des offres promotionnelles ciblées. IdVente fait donc partie de la clé primaire.

À présent, nous allons parler des mesures que l'on retrouve dans la table vente. Nous en avons uniquement trois : le prix HT (hors taxe), le prix TTC (toutes taxes comprises) et la note.

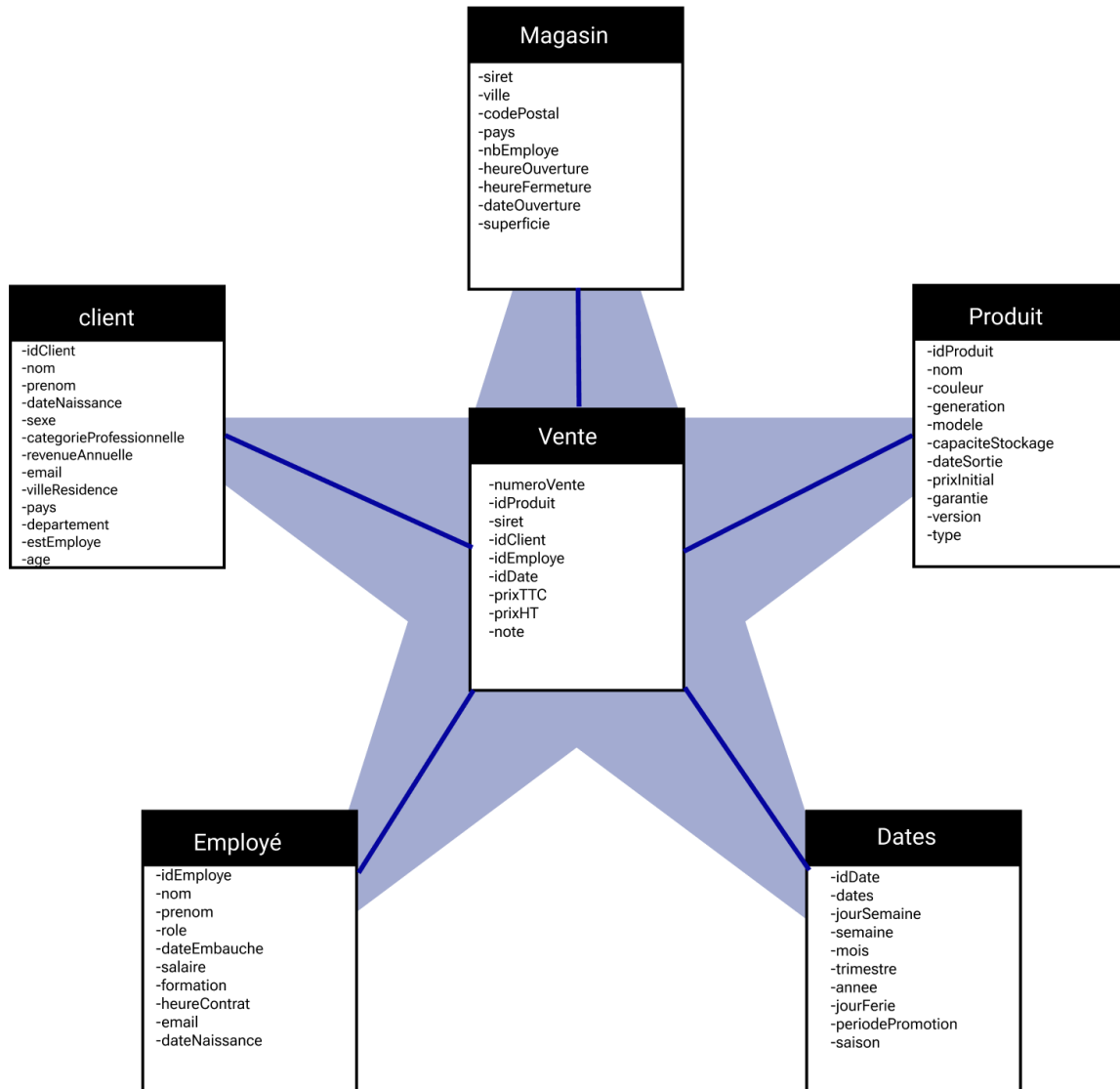
La note associée à la vente d'un produit est non additif. En effet, sommer cette valeur n'aurait pas de sens, car cette mesure est qualitative. Cependant, nous pouvons sommer les notes pour et diviser le résultat par le nombre total de notes, et ainsi obtenir une note moyenne.

Les prix (HT et TTC) sont des mesures additives. En effet, elles peuvent être sommées par toutes les dimensions :

- Possibilité de sommer les prix de plusieurs produits pour un magasin, un client, une date et un vendeur.
- Possibilité de sommer les prix de plusieurs vendeurs pour un magasin, un client, une date et un produit.
- Possibilité de sommer les prix de plusieurs magasins pour un produit, un client, une date et un vendeur.
- Possibilité de sommer les prix de plusieurs clients pour un magasin, un produit, une date et un vendeur.
- Possibilité de sommer les prix de plusieurs dates pour un magasin, un client, un produit et un vendeur.

Nous pouvons sommer cette donnée afin de connaître le chiffre d'affaires et de ainsi créer des statistiques.

Voici le modèle en étoile créé pour la table de faits Vente:



6.2. La deuxième action / opération : stock

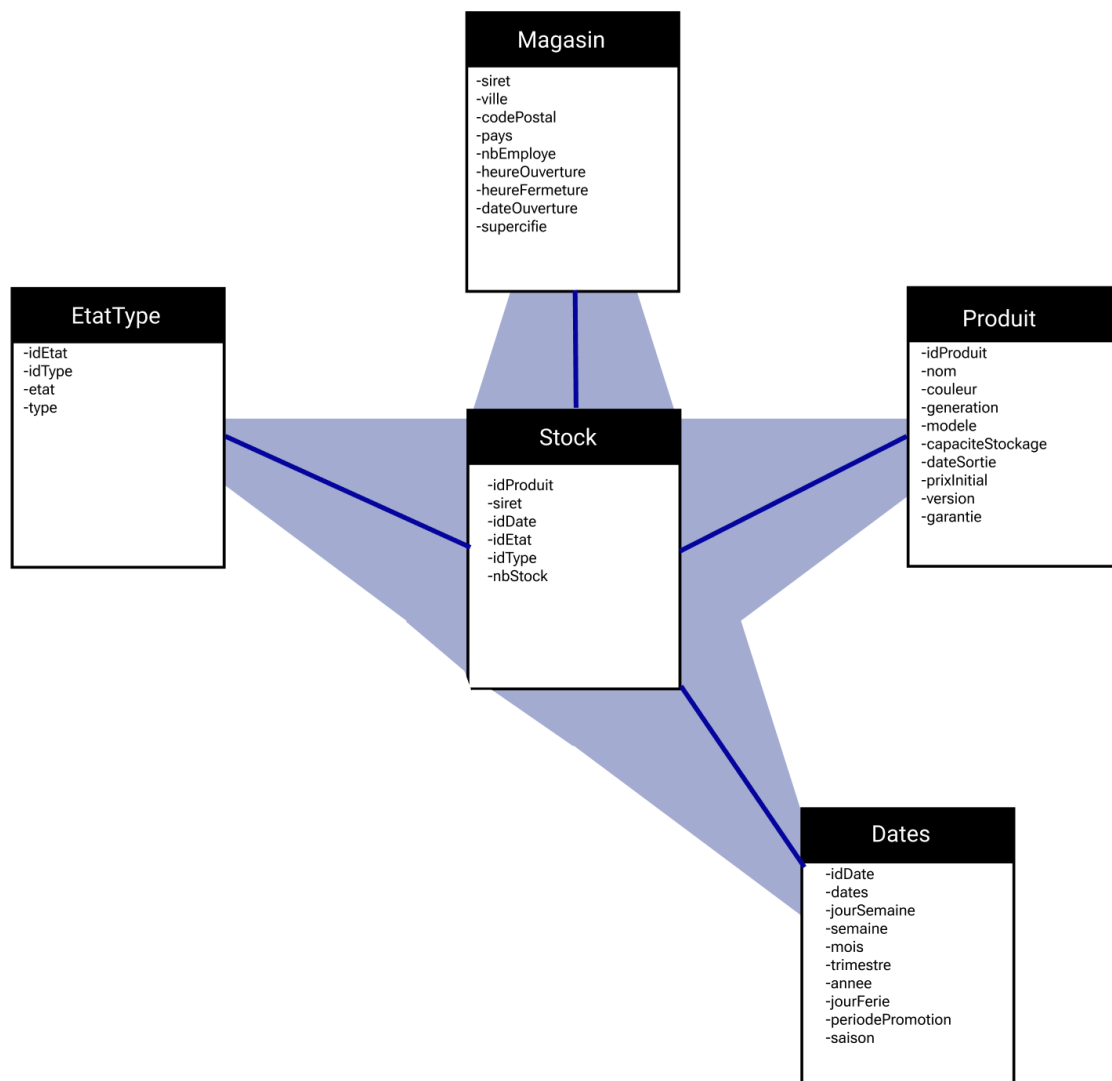
Pour le modèle en étoile de la gestion des stocks, il est nécessaire d'ajouter quelques informations. Tout d'abord, il s'agit d'un snapshot, une copie des données réalisée à un moment donnée, ici, on estime que les données seront rentrées une fois par jour, en fin de journée. De plus, dans la dimension produit défini dans le data-mart de "vente", on retrouve le type d'un produit. Cependant, pour les stocks, il nous est intéressant de connaître le stockage des produits par type. En effet, un produit de type ordinateur ne prend pas la même place dans les stocks qu'un produit de type écouteurs sans fils. Donc, dans ce data-mart, nous allons faire de "type" une dimension à part, qui sera donc une mini dimension. De cette façon, les calculs comprenant un regroupement par type de produit seront plus rapides.

Nous avons également une autre mini dimension, la dimension État. Nous les avons donc regroupés en faisant le produit cartésien des 2 mini dimensions. Ce choix a été fait pour optimiser notre data-mart.

Nous retrouvons une seule mesure qui est le nombre de stocks (nbStock).

- Le mesure nbStock est semi-additive. En effet, on peut sommer cette valeur avec certaines dimensions, mais pas toutes. On peut sommer le stock de tous les produits pour un magasin à un jour précis pour un état précis et un type de produit. On peut aussi sommer le stock de tous les magasins, pour un jour précis et pour un article précis, pour un état précis et un type de produit précis. Cependant, on ne peut pas sommer le stock de tous les jours pour un magasin donné, un produit donné, un type de produit et un état précis. Cette valeur n'a pas de sens.

Voici le datamart créé pour la deuxième action la plus importante : la gestion des stocks.



7. Dimensions des modèles

7.1. Gestion des ventes

Table de Faits : Vente

La table de faits 'Vente' est au cœur de l'entrepôt de données et contient les métriques clés des transactions effectuées.

- idVente : L'identifiant de la vente est important au cas où un client achète le même produit plusieurs fois au même jour, du même magasin et par le même employeur.
- idProduit, siret, idEmploye, idClient, idDate : Clés étrangères liant aux dimensions correspondantes.
- prixTTC, prixHT : Le montant de la transaction est essentiel pour les analyses de chiffre d'affaires.
- note : Un retour client qui peut être utilisé pour l'amélioration de la qualité du produit.

Dimension Produit

Cette dimension offre une vue complète des produits vendus, enrichissant l'analyse des ventes avec des détails sur les articles achetés.

- idProduit : L'identifiant unique de chaque produit.
- nom, couleur, generation, type, modele : Descriptions détaillées qui facilitent le suivi des tendances de produits spécifiques.
- capaciteStockage, dateSortie, prixInitial, garantie, version : Permettent d'analyser l'évolution des ventes en fonction des caractéristiques techniques et du cycle de vie du produit. Nous précisons que la capaciteStockage sera NULL pour certains produits, par exemple les écouteurs. l'attribut "garantie" est un entier représentant le nombre de mois pendant lequel le produit est sous garantie.

Dimension Magasin

Chaque magasin est unique et ses caractéristiques peuvent affecter les ventes.

- siret : Identifiant unique pour chaque magasin, essentiel pour des rapports par emplacement.
- ville, codePostal, pays : Localisation précise pour des analyses régionales.
- nbEmployes, superficie : Caractéristiques opérationnelles pour évaluer l'impact sur la performance.
- heureOuverture, heurefermeture, dateOuverture : Données temporelles pour analyser l'efficacité des heures et jours d'opération.

Dimension Employé

Le personnel joue un rôle crucial dans les ventes et la satisfaction client.

- idEmploye, nom, prenom : Identification et suivi personnel.
- role, dateEmbauche, salaire : Données sur la carrière pour corréliser les ventes avec la performance du personnel. Le rôle sera compris entre "CHEF", "MANAGER" et "POLYVALENT"
- formation, heureContrat : Pour évaluer l'efficacité de la formation et l'engagement des employés. La formation sera un entier représentant le nombre d'années d'études supérieures (par exemple "2" pour un BTS). Ensuite, "heureContrat" sera un entier représentant le nombre d'heures effectuées par semaine pour un employé (exemple "35", "20"...).

- email, dateNaissance: Communications et démographie pour la gestion des ressources humaines.

Dimension Client

Les données sur les clients nous aident à personnaliser les offres et à comprendre le marché.

- idClient, nom, prenom, dateNaissance, sexe, age : Profilage pour des campagnes marketing ciblées.
- categorieProfessionnelle, revenuAnnuel : Stratification pour des offres adaptées aux segments de revenu.
- email : Engagement client pour des stratégies de fidélisation.
- villeResidence, pays, departement : Préférences et localisation pour l'ajustement des stocks et des stratégies de vente locale.
- estEmployé : permet de savoir si le client est un employé d' Apple.

Dimension Date

La temporalité des ventes est cruciale pour toute analyse de tendance ou saisonnière.

- idDate : Clé primaire, utilisée pour transformer une date en entier. Les requêtes sont plus performantes sur des données de type entier que des données de type date. Ainsi, la date 20/04/2023 aura comme id l'entier 20042023.
- dates : Liaison claire avec la table de faits.
- jourSemaine, semaine, mois, trimestre, annee : Découpage temporel pour différentes granularités d'analyse.
- jourFerie, periodePromotion, saison : Pour évaluer l'impact des vacances et des promotions sur les ventes. jourFerie et periodePromotion seront des booléen.

7.2. Gestion des stocks

Pour simplifier le processus de gestion et stockage des données, nous allons utiliser les mêmes tables pour les dimensions Produit, Magasin et Dates, en créant des vues de celles-ci. La seule différence se trouvera dans la vue produit. En effet, comme dit précédemment, il nous est intéressant de connaître le stock en fonction du type de produit. Nous allons donc retirer l'attribut "type" de la vue correspondant à la dimension Produit pour en faire une mini-dimension à part. Cette dimension sera combinée avec la mini-dimension "Etat", afin de réduire le nombre de dimensions et optimiser notre entrepôt de donnée.

Table de Faits : Stock

La table de faits 'Stock' enregistre les niveaux de stock des produits dans les magasins.

- idProduit, siret, idEtat, idType, idDate : Clés étrangères reliant aux dimensions correspondantes pour une intégration transparente.
- nbStock : Quantité actuelle en stock, vital pour le suivi des niveaux d'inventaire.

Dimension: EtatType

La condition des produits en stock affecte directement leur gestion et leur vente.

- idEtat, etat : Description de la condition actuelle des stocks (neuf, reconditionné).
- idType, type : Description du type de produit (smartphone, ordinateur, écouteurs, tablette)

8. Analyse de notre modèle en fonction des traitements

8.1. Modèle de la gestion des ventes

Pour calculer les ventes (en termes de chiffre d'affaires) par année et par magasin, nous servons des dimensions suivantes :

- La dimension Dates permet de filtrer les ventes par année.
- La dimension Magasin permet d'agréger les ventes par magasin. Si nous cherchons uniquement à classer les magasins selon le numéro de siret, il est possible de ne pas faire de jointure avec la dimension Magasin. En effet, le numéro de siret est une clé étrangère de "Stock", lié à la dimension "Magasin"

On calcule le chiffre d'affaires en faisant la somme de toutes les valeurs de la mesure "prixTTC" dans la table vente.

Pour calculer les ventes (en termes de nombre de produits) effectuées en fonction des tranches d'âge des clients par année, nous allons utiliser les dimensions suivantes :

- La dimension Dates peut être utilisée pour déterminer l'année des ventes.
- La dimension Client peut être utilisée pour segmenter les ventes par tranche d'âge.

En effet, nous aurions besoin de faire le calcul entre la date de naissance du client et la date de l'achat du produit pour calculer l'âge du client quand il a acheté le produit. Nous n'oublierons pas de faire le nécessaire pour faire la somme du nombre des ventes en utilisant "COUNT".

Afin de calcul le nombre des ventes par type de matériel par année, nous travaillons sur les dimensions suivantes :

- La dimension Dates permet de sélectionner l'année des ventes.
- La dimension Produit permet d'identifier le type de matériel.

Nous finissons par compter avec la fonction "COUNT" le nombre de ventes pour avoir le résultat.

8.2. Modèle de la gestion des stocks

Nous cherchons à calculer le nombre de produits en stocks par magasin, par mois. Nous aurons besoin de ces deux dimensions :

- La dimension Magasin permet d'agréger les ruptures par magasin. (Nous avons le numéro de siret du magasin directement dans la table Stock)
- La dimension Dates est essentielle pour filtrer les ruptures par mois.

On utilisera le fait "nbStock" dans notre table des faits.

Afin de calculer le nombre de produits en stocks par produit, par magasin, par mois, nous utilisons ces trois dimensions :

- La dimension Produit est utilisée pour identifier les produits en rupture.
- La dimension Magasin permet de segmenter les ruptures par magasin.
- La dimension Dates est utilisée pour déterminer le mois des ruptures.

Nous ne pouvons pas additionner les stock, cela n'a pas de sens. Pour connaître le nombre de stock par mois, il faudrait soit faire une moyenne, ou bien regarder uniquement le nombre de stock que nous avons le dernier jour du mois.

Pour faire la moyenne du stock par produit, par année, par magasin, par état, nous nous serons des dimensions suivantes :

- La dimension Produit permet d'agréger les stocks par type.
- La dimension Dates est cruciale pour suivre l'évolution du stock par année.
- La dimension Magasin est utilisée pour segmenter les stocks par magasin.
- La dimension État du Produit offre une compréhension approfondie des conditions des produits.

On terminera par faire la moyenne en utilisant le fait "nbStock" dans notre table des faits, puis en appliquant la fonction avg().

9. Test de la pertinence de notre modèle

Voici quelques exemples que nous pouvons retrouver dans notre modèle. Commençons par se concentrer sur les dimensions qui apparaissent dans les deux tables Ventes et Stocks, c'est-à-dire les dimensions date, magasin et produit.

	Date								
idDate	date	jour	mois	année	trimestre	semaine	jourFerie	prom o	saison
31082023	31/08/2023	31	08	2023	3	35	N	N	ete
20122022	20/12/2022	20	12	2022	4	51	N	Y	hivers
18112023	18/11/2023	18	05	2023	2	20	Y	N	printemps

Magasin								
siret	ville	codePostal	pays	nbEmploye	superficie	heureOuverture	heureFermeture	dateOuverture
15	Montpellier	34000	France	17	1.400	8	18	02/05/2008
49	Paris	75000	France	38	2.900	8	20	20/09/2004
36	Bilbao	48001	Espagne	13	950	10	18	26/07/2014

Produit										
idProduit	nom	couleur	generation	type	modele	capacite Stockage	dateSortie	Prix Initial	garantie	version
P1501	Iphone15	noir	15	smartphone	NULL	128	22/08/2023	969	24	iOS 14
M1301	MacBook13	gris	13	ordinateur	Air	256	02/08/2023	1199	36	mac OS
AP02	AirPods2	blanc	2	ecouteurs	pro	NULL	10/11/2023	300	12	1ere gen

Ensuite, nous pouvons passer aux dimensions que nous retrouvons uniquement sur la table ventes, c'est-à-dire les dimensions Employé et Client.

Employe									
idEmploye	nom	prenom	role	dateEmbauche	salair e	formation	heure Contrat	email	dateNaissance
0	Mikhaylov	Dimitri	Chef	16/05/2023	2.500	5	35	dimitri.C@gmail.com	14/03/1985
1	Giovanni	Lombardi	Manager	16/05/2021	2.000	3	35	Giovanni@live.com	25/03/1992
2	Monet	Ayden	Polyvalent	22/04/2020	1.400	0	10	Aayden@yahoo.com	12/09/2001

Client												
idCli	nom	pre nom	age	dateNaissance	sexe	categorie Professionnelle	revenu Annuel	email	ville Residence	departement	pays	estEmploye
0	Schne	Théo	21	14/07/2001	H	développeurs de jeux	36.000	Schneider.th@gmail.com	Paris	75	France	N
1	Matar	Maya	29	20/06/1994	F	ingénieur	75.000	Maya.matar@yahoo.com	Marseille	13	France	N
2	Ogun	Taner	40	02/07/1983	H	economiste	45.000	Ogun.Taner@live.com	Montpellier	34	France	Y

A présent, nous pouvons avoir un aperçu de notre table Vente :

Ventes								
idVente	idProduit	siret	idEmploye	idClient	idDate	note	prixTTC	prixHT
21853664	P1501	15	1	2	31082023	5	969	807
44815487	AP02	49	2	0	18112023	4	240	200
42488419	P1501	49	0	0	10112023	4	969	807

Passons maintenant à la table des faits Stocks, en commençant par donner un exemple de la dernière dimension qu'il nous manque, c'est-à-dire la dimension EtatType

EtatType			
idEtat	Etat	idType	type
0	neuf	0	smartphone
0	neuf	1	ordinateur
1	reconditionne	0	smartphone

Pour finir, voici un exemple pour la table des faits stock:

Stock					
idProduit	siret	idEtat	idDate	idType	nbStock
P1501	25	1	25052023	1	102
AP02	30	1	25052021	2	268
AP02	445	2	25052022	2	0

10. Estimation de la taille des tables de l'entrepôt

Nous allons commencer par estimer la taille de notre modèle de gestion des ventes, en estimant la taille de notre table des faits "vente" pour une année.

En 2022, Apple a vendu 232 millions de smartphones (Iphones), 61 millions de tablettes tactiles (Ipad) et 26 millions d'ordinateurs (Mac). Pour notre projet, nous nous concentrons sur les ventes effectuées en magasin. Parmi les ventes de smartphones, seulement 10 % ont été faites en magasin. Pour les tablettes tactiles, nous en comptons 20 %. 40 % des ventes d'ordinateurs ont été faites en boutique. On peut ajouter les données pour les ventes des écouteurs sans fils et des montres connectées. 82 millions d'écouteurs sans fils ont été vendus et 25 % de ses ventes ont été effectuées dans les magasins d'appel et 53 millions d'Apple watches avec un taux de 30 % de ventes dans les magasins d'appel.

Après avoir fait le calcul, nous pouvons en conclure que le nombre des ventes des produits Apple dans les Apple Store partout dans le monde en 2007 a atteint 78 millions donc les ventes par année doivent être d'environ 80 millions.

Si nous voulons aller plus loin, nous pouvons estimer que notre entrepôt de donnée est utilisé depuis 2010, cela voudrait dire que notre table de vente est composée de plus d'un milliard de lignes ($80\,000\,000 * 14 = 1\,120\,000\,000$).

Nous allons maintenant passer à une estimation de la table des faits Stocks.

- Pour la dimension produit, nous pouvons estimer un total de 2 069 lignes. Voici le détail de ce calcul :
 - En ce moment, nous avons sur le site Apple la possibilité d'acheter 5 iPhone différents. Pour chaque iPhone, nous avons au maximum 3 générations, 4 capacités de stockage, 5 couleurs et 3 versions. Nous pouvons estimer 900 lignes pour stocker tous les téléphones.
 - Passons à présent aux tablettes tactiles. Quatre tablettes sont actuellement disponibles. Deux couleurs sont disponibles, ainsi que 5 capacités de stockage et 5 versions. Il n'y a pas de génération pour les tablettes, nous sommes donc sur une estimation de 200 lignes.
 - Pour les ordinateurs, nous avons 4 modèles disponibles, 4 couleurs, 3 versions, 4 générations, et 5 capacités de stockage, ce qui nous fait une estimation de 960 lignes.
 - Pour les écouteurs, nous avons 1 modèle, 3 versions, 3 générations et 1 couleur, ce qui nous fait une estimation de 9 lignes.
- Pour la dimension magasin, nous aurons un total de 527 lignes, car il existe un total de 527 magasins Apple dans le monde.
- Pour la dimension dates, nous aurons 367 lignes (une année est composée au maximum de 367 jours).
- Pour la dimension étatType, il nous faut un total de 8 lignes. En effet, il s'agit du produit cartésien des 2 mini dimensions Etat et Type :
 - Nous avons 2 états possibles : (neuf, reconditionné)
 - Nous avons 4 types possibles : ordinateur, telephones, ecoouteurs et tablette

Pour un jour, nous avons autant de ligne que de produit possible dans un magasin possible pour un Etat / Type possible. Nous pouvons donc estimer un total de plus de 3 milliards de lignes au bout d'un an pour la table des faits Stocks. ($367 * 2\,069 * 527 * 8 = 3\,201\,305\,768$)

L'implémentation et le requêtage

11. Implémentation en ORACLE des tables des faits et des vues

11.1. Utilisation des vues virtuelles pour les dimensions partagées

La créations des tables se trouve dans les fichiers Ventes.sql et Stocks.sql. Ici, nous ne reviendrons uniquement sur les vues virtuelles. En effet, les dimensions “Magasin” “Produit” et “Dates” du data-mart “Ventes” sont utilisées pour le data-mart “Stocks” sous formes de vues. Nous allons commencer par la création de la vue “Produit”, tout en prenant soin d'enlever l'attribut “Type” :

```
CREATE OR REPLACE VIEW ProduitStocks AS
SELECT
    idProduit,
    nom,
    couleur,
    generation,
    modele,
    capaciteStockage,
    dateSortie,
    prixInitial,
    garantie,
    version
FROM Produit;
```

A présent, voici la créations des vues Magasins et Dates, identiques aux dimensions précédemment créés :

```
CREATE OR REPLACE VIEW MagasinStocks AS
SELECT *
FROM Magasin;

CREATE OR REPLACE VIEW DatesStocks AS
SELECT *
FROM Dates;
```

12. Requêtes analytiques

Voici 12 requêtes analytiques correspondant aux traitements que nous avons imaginé précédemment. Nous allons commencer par les requêtes sur notre action principale : les ventes.

- Calculer le chiffre d'affaires par année et par magasin.

- Cette requête nous permet d'avoir une évolution au fil du temps des statistiques de ventes du magasin. Il sera possible d'analyser si certains magasins sont en déclin au fil du temps.
- Requête SQL :

```
SELECT
    M.siret,
    D.annee,
    SUM(V.prixTTC)
FROM
    Vente V
JOIN
    Magasin M ON V.siret = M.siret
JOIN
    Dates D ON V.idDate = D.idDate
GROUP BY
    M.siret,
    D.annee
ORDER BY
    M.siret,
    D.annee;
```

- Calculer le nombre de ventes total par tranche d'âge du client et par année.
 - Cette requête nous permet d'analyser les ventes en fonction de l'âge du client. Ainsi, il sera plus simple de cibler les publicités et les futures ventes. Il est intéressant de les classer par année pour analyser les ventes en fonction de la tranche d'âge au fil du temps : savoir si nos clients sont de plus en plus jeunes par exemple.
 - Requête SQL :

```
SELECT
    CASE
        WHEN C.age <= 20 THEN 'Under 21'
        WHEN C.age BETWEEN 21 AND 30 THEN '21-30'
        WHEN C.age BETWEEN 31 AND 40 THEN '31-40'
        WHEN C.age BETWEEN 41 AND 50 THEN '41-50'
        WHEN C.age BETWEEN 51 AND 60 THEN '51-60'
        ELSE 'Over 60'
    END AS AgeGroup,
    EXTRACT(YEAR FROM D.dates) AS Year,
    COUNT(*) AS TotalSales
FROM
    Vente V
JOIN
    Client C ON V.idClient = C.idClient
JOIN
```

```

    Dates D ON V.idDate = D.idDate
GROUP BY
    CASE
        WHEN C.age <= 20 THEN 'Under 21'
        WHEN C.age BETWEEN 21 AND 30 THEN '21-30'
        WHEN C.age BETWEEN 31 AND 40 THEN '31-40'
        WHEN C.age BETWEEN 41 AND 50 THEN '41-50'
        WHEN C.age BETWEEN 51 AND 60 THEN '51-60'
        ELSE 'Over 60'
    END,
    EXTRACT(YEAR FROM D.dates)
ORDER BY
    Year,
    AgeGroup;

```

- Afficher le nombre de ventes par produit, par magasin, et par année.
 - Cette requête permet de gérer le stock d'un produit, par magasin et par année. Elle est généraliste, mais permettra de faire un lien avec la moyenne du nombre de stocks par produit, par magasin et par année.
 - Requête SQL :

```

SELECT
    P.nom,
    M.siret,
    D.annee,
    COUNT(*) AS Ventes
FROM
    Vente V
JOIN
    Produit P ON V.idProduit = P.idProduit
JOIN
    Magasin M ON V.siret = M.siret
JOIN
    Dates D ON V.idDate = D.idDate
GROUP BY
    P.nom,
    M.siret,
    D.annee
ORDER BY
    P.nom,
    M.siret,
    D.annee;

```

- Afficher le nom des produits et le nombre total d'achats de ces produits, des clients qui sont employés d'Apple. Afficher les résultats par produit, classés par ordre descendant de ventes. (le produit le plus acheté par un client qui est aussi employé, au moins acheté).

- Cette requête permet de voir si les employés Apple (qui sont donc connaisseurs) achètent les produits Apple. Ainsi, il est possible de voir les préférences d'achats des clients-employés qui connaissent parfaitement les produits.
- Requête SQL :

```
SELECT
    P.nom,
    COUNT(*) AS Achats
FROM
    Vente V
JOIN
    Client C ON V.idClient = C.idClient
JOIN
    Produit P ON V.idProduit = P.idProduit
WHERE
    C.estEmploye = 'Y'
GROUP BY
    P.nom
ORDER BY
    Achats DESC;
```

- Afficher la somme totale hors taxes de ventes générée en France par note et département du client.

- Cette requête permet d'analyser le chiffre d'affaires réalisé en France en fonction des notes attribuées et du département du client.
- On utilise le département pour avoir une analyse plus fine des notes. Car on suppose que la note varie en fonction de l'humeur et l'exigence du client, et bien souvent le département influe sur l'humeur générale et l'exigence de sa population. La région parisienne a la réputation d'avoir des clients de moins bonne humeur et plus exigeants que l'Hérault par exemple.
- Requête SQL :

```
SELECT
    C.departement,
    V.note,
    SUM(V.prixHT) AS HorsTaxe
FROM
    Vente V
JOIN
    Magasin M ON V.siret = M.siret
JOIN
    Client C ON V.idClient = C.idClient
```

```

WHERE
    M.pays = 'France'
GROUP BY
    C.departement,
    V.note
ORDER BY
    C.departement,
    V.note;

```

- Afficher le nombre total de ventes, la superficie et le nombre d'employés par magasin et année dans un ordre descendant de ventes.
 - Cette requête permet d'analyser par magasin le nombre de vente, et de faire des liens entre leur superficie et leur nombre d'employés. Il est également possible d'analyser les tendances en regardant les différences entre les années.
 - Requête SQL :

```

SELECT
    COUNT(*) AS NbrTotaleVente,
    M.siret,
    D.annee,
    M.superficie,
    M.nbEmployes
FROM
    Vente V
JOIN
    Magasin M ON V.siret = M.siret
JOIN
    Dates D ON V.idDate = D.idDate
GROUP BY
    M.siret,
    D.annee,
    M.superficie,
    M.nbEmployes
ORDER BY
    M.superficie DESC,
    NbrTotaleVente DESC;

```

- Calculez le chiffre d'affaires total et le nombre de ventes par année et par magasin, avec des sous-totaux pour chaque année et un total général pour tous les magasins et toutes les années.
 - Cette requête permet de faire des analyses précises sur les performances des magasins. Il est également possible de voir les tendances d'une année à l'autre.
 - Requête SQL :

```

SELECT
    COALESCE(TO_CHAR(D.annee), 'Toutes les années') AS Annee,
    COALESCE(TO_CHAR(M.siret), 'Tous les magasins') AS
Magasin,
    SUM(V.prixTTC) AS TotalRevenue,
    COUNT(*) AS TotalSales
FROM
    Vente V
JOIN
    Magasin M ON V.siret = M.siret
JOIN
    Dates D ON V.idDate = D.idDate
GROUP BY
    ROLLUP(D.annee, M.siret)
ORDER BY
    D.annee,
    M.siret;

```

- Calculer le nombre total de vente d'un produit par couleur du produit et par sexe du client.
 - Cette requête permet d'analyser les préférences des clients en fonction de leur sexe, et permettra de faire des futures campagnes de publicité plus ciblées.
 - Requête SQL :

```

SELECT
    COUNT(*) AS NbrVente,
    P.couleur AS CouleurProduit,
    C.sexe AS SexeClient
FROM
    Vente V
JOIN
    Produit P ON V.idProduit = P.idProduit
JOIN
    Client C ON V.idClient = C.idClient
GROUP BY
    P.couleur,
    C.sexe
;

```

- Analyser les ventes totales et le chiffre d'affaires dans différentes dimensions : par année, par magasin et par type de produit.
 - Requête SQL :

```
SELECT
    COALESCE(TO_CHAR(D.annee), 'Toutes les années') AS Annee,
    COALESCE(TO_CHAR(M.siret), 'Touts les magasins') AS
Magasin,
    COALESCE(P.type, 'Tous les types') AS TypeProduit,
    SUM(V.prixTTC) AS ChiffresAffaires,
    COUNT(*) AS VentesTotales
FROM
    Vente V
JOIN
    Magasin M ON V.siret = M.siret
JOIN
    Dates D ON V.idDate = D.idDate
JOIN
    Produit P ON V.idProduit = P.idProduit
GROUP BY
    CUBE(D.annee, M.siret, P.type)
ORDER BY
    D.annee,
    M.siret,
    P.type;
```

Maintenant, nous allons traiter des requêtes sur l'action stock :

- Afficher le nombre moyen de stocks par produit, par magasin pendant les périodes de promotion
 - Cette requête permet d'analyser l'impact de la période de promotion sur les stocks des magasins.
 - Requête SQL :

```
SELECT
    s.idProduit, s.siret, AVG(s.nbStock) AS MoyenneStocks
FROM
    Stocks s
JOIN
    Dates d ON s.idDate = d.idDate
WHERE
    d.pperiodePromotion = 'Y'
GROUP BY
    s.idProduit, s.siret;
```


- Afficher la moyenne du nombre de stocks par produit, par magasin et par mois pour des produits neufs.
 - Cette requête nous permet d'optimiser la gestion des produits neufs. Il est important d'avoir toujours un stock de produits neuf car c'est un gage de qualité.
 - Requête SQL :

```
SELECT
  s.idProduit,
  s.siret,
  d.mois,
  AVG(s.nbStock) AS MoyenneStocks
FROM
  Stocks s
JOIN
  Dates d ON s.idDate = d.idDate
JOIN
  EtatType et ON s.idEtat = et.idEtat
WHERE
  et.etat = 'neuf'
GROUP BY
  s.idProduit, s.siret, d.mois;
```

- Afficher les stocks moyens par produit, par année, par magasin.
 - Cette requête est généraliste, mais permet de faire un lien avec le nombre de ventes par produit, par année et par magasin.
 - Requête SQL :

```
SELECT
  COUNT(*) AS NbrStock,
  P.nom AS ProductName,
  M.siret AS StoreID,
  D.annee AS Annee
FROM
  Stocks S
JOIN
  Produit P ON S.idProduit = P.idProduit
JOIN
  Magasin M ON S.siret = M.siret
JOIN
  Dates D ON S.idDate = D.idDate
GROUP BY
  P.nom, M.siret,
  D.annee
ORDER BY
  M.siret;
```

Ressources / Annexes

https://fr.wikipedia.org/wiki/Apple_Store

<https://fr.wikipedia.org/wiki/Apple>

<https://www.businessofapps.com/data/apple-statistics/>

<https://www.apple.com/fr/newsroom/2023/08/apple-reports-third-quarter-results/>

<https://www.blogdumoderateur.com/chiffres-apple/>

Github : <https://github.com/SateaMall/AppleDW/tree/main>