

Table des matières

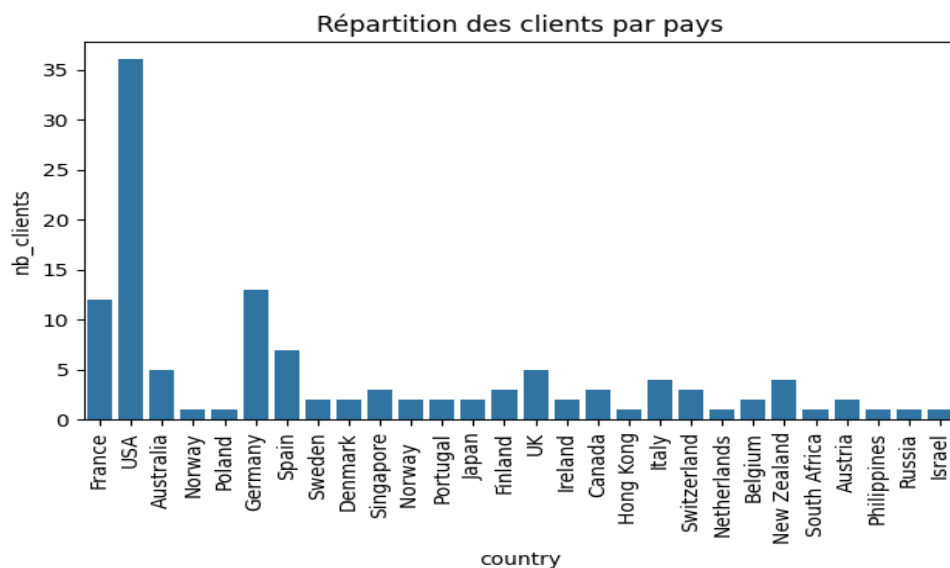
I.	Introduction :	2
II.	Analyse exploratoire	2
III.	Analyse temporelle des ventes :	3
IV.	Analyse par gamme de produits	4
V.	Analyse RFM (Segmentation client) :	4
VI.	Analyse de corrélation :	5
VII.	Analyse panier (Market Basket Analysis) :	5
VIII.	Performance des employés :	6
	Conclusions :	6

I. Introduction :

Dans un contexte de transformation numérique des entreprises, l'exploitation stratégique des données constitue un levier central de performance. L'objectif de ce projet est d'analyser la base de données relationnelle ClassicModels afin d'identifier des axes d'optimisation commerciale, financière et logistique. La base contient des informations relatives : aux clients, aux commandes, aux produits, aux employés, aux paiements. L'étude combine : des requêtes SQL avancées (jointures, agrégations, CTE, procédures), des analyses statistiques sous Python, des visualisations graphiques, une segmentation RFM. Comment exploiter les données transactionnelles afin : d'identifier les déterminants du chiffre d'affaires, de segmenter efficacement la clientèle, d'optimiser la performance commerciale, de réduire les risques liés à la concentration des ventes ?

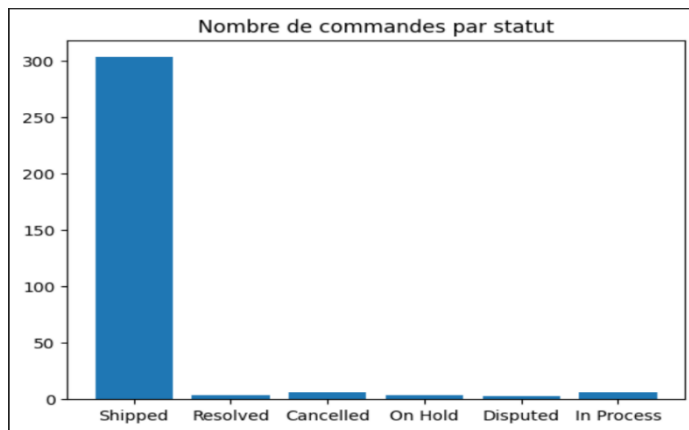
II. Analyse exploratoire

- Répartition géographique :



La distribution spatiale des clients révèle l'étendue du rayonnement de l'entreprise et ses dépendances de marché. Les **USA** dominent largement le portefeuille avec environ 36 clients, soit une concentration nettement supérieure à n'importe quelle autre région. L'Europe constitue le second pilier stratégique, portée par l'**Allemagne** (13 clients) et la **France** (12 clients). Ces marchés matures offrent une stabilité relative face à la domination américaine. L'entreprise affiche une présence dans 27 pays, mais avec une "longue traîne" de marchés émergents ou marginaux (ex: Norvège, Pologne, Hong Kong) ne comptant souvent qu'un seul client.

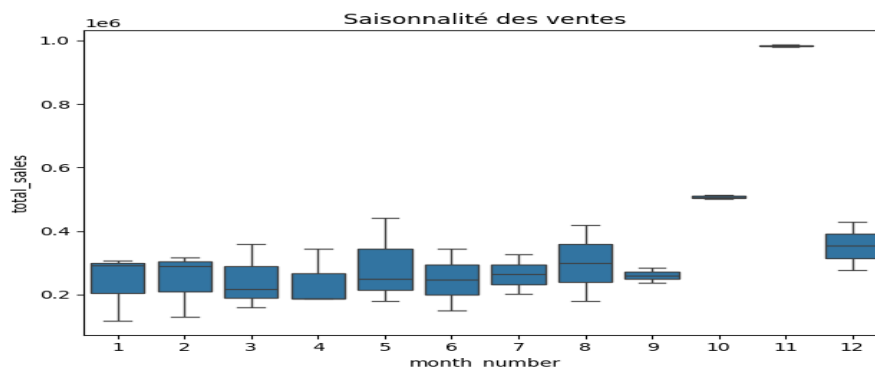
- Distribution des commandes par statuts :



Le graphique relatif au statut des commandes permet de mesurer l'efficacité de la chaîne logistique et le niveau de satisfaction client. La grande majorité des commandes (plus de 300 unités) affiche le statut Shipped (Expédiée). Cela démontre une fluidité opérationnelle satisfaisante et une capacité à traiter les volumes de vente actuels. Les statuts critiques tels que Cancelled (Annulée), Disputed (Litige) ou On Hold (En attente) sont extrêmement minoritaires, ne dépassant pas la dizaine d'unités chacun. Le faible taux de litiges est un indicateur de fiabilité du service. Les commandes In Process et Resolved sont également marginales, ce qui indique un temps de cycle court entre la commande et l'expédition finale.

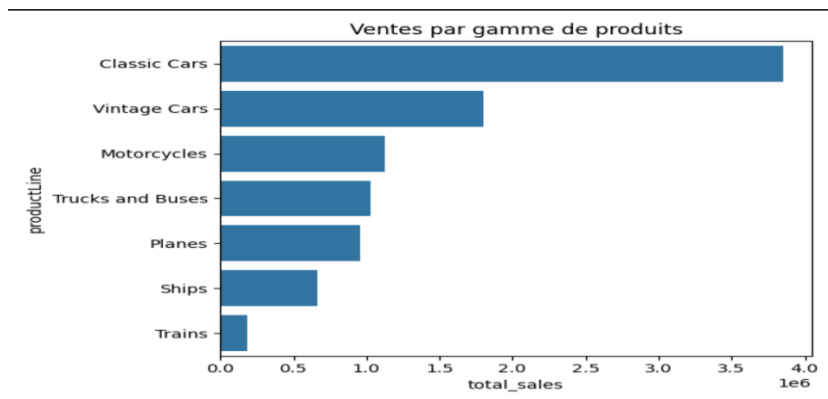
III. Analyse temporelle des ventes :

- Saisonnalité :



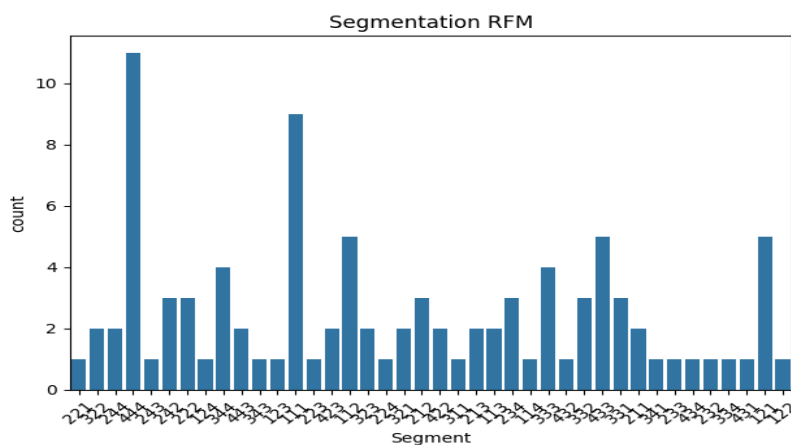
Le diagramme en boîtes à moustaches (Figure 3) illustre la variabilité du chiffre d'affaires sur un cycle annuel. De janvier (mois 1) à septembre (mois 9), les ventes totales oscillent de manière constante entre 200 000 et 400 000 unités monétaires. On observe une accélération majeure en fin d'année. Le mois d'octobre (10) marque un palier à environ 500 000, suivi d'un pic exceptionnel en novembre (11) avoisinant 1 000 000. Le mois de décembre (12) montre une décrue par rapport à novembre, tout en restant supérieur à la moyenne annuelle, avec une médiane proche de 350 000.

IV. Analyse par gamme de produits



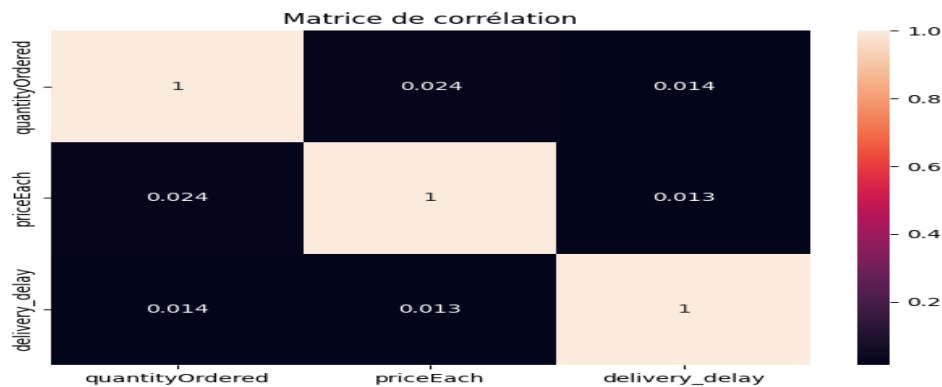
Ce graphique révèle une spécialisation marquée. L'entreprise semble bénéficier d'un fort avantage concurrentiel ou d'une demande de marché supérieure sur le segment des voitures classiques. Le déséquilibre entre les « Classic Cars » et le reste du portefeuille suggère deux scénarios possibles : Les voitures classiques financent potentiellement le développement des autres gammes, La performance globale de l'organisation est fortement corrélée au succès d'une seule catégorie, ce qui peut représenter un risque si le marché des voitures de collection venait à stagner.

V. Analyse RFM (Segmentation client) :



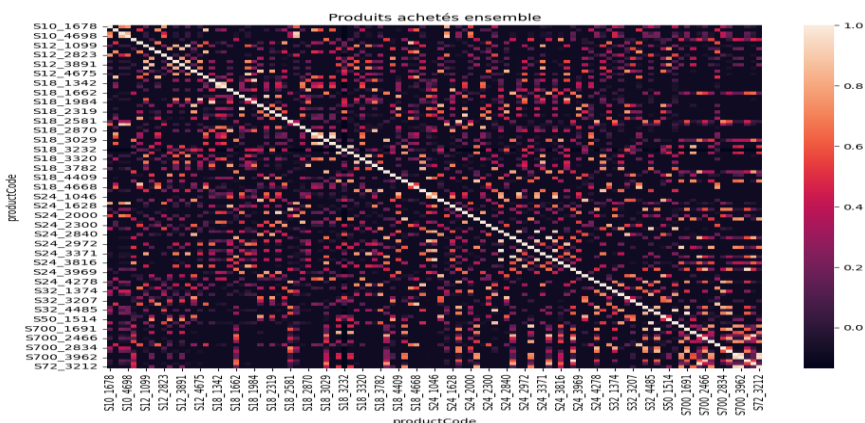
L'organisation s'appuie sur une base de clients fidèles ("444") et une domination claire sur le marché américain. La robustesse du système logistique (taux élevé de commandes expédiées) permet de soutenir des volumes de ventes exceptionnels durant le pic de novembre. Pour améliorer on doit planifier les stocks et la capacité logistique dès le troisième trimestre pour absorber le volume de ventes qui triple en fin d'année, mettre en œuvre des campagnes ciblées sur le segment "111" pour convertir ces clients inactifs en profils intermédiaire, renforcer la présence en Allemagne et en France pour équilibrer la dépendance au marché américain.

VI. Analyse de corrélation :



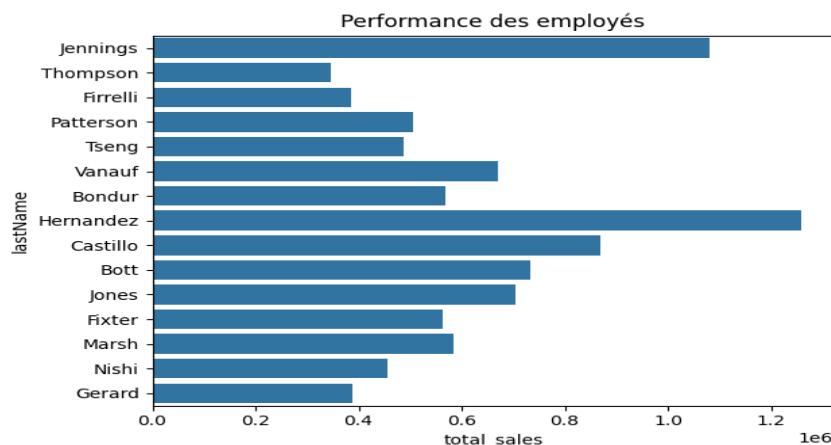
La matrice de corrélation révèle une indépendance quasi totale entre la quantité commandée, le prix unitaire et le délai de livraison, avec des coefficients extrêmement faibles (tous $< 0,025$). Cela indique que le volume d'une commande n'influence pas les délais de livraison, et que la tarification ne semble pas corrélée au volume des unités achetées.

VII. Analyse panier (Market Basket Analysis) :



L'analyse de la densité de couleur permet d'identifier des comportements d'achat spécifiques. Les zones de couleur claire en dehors de la diagonale indiquent des produits qui sont fréquemment achetés simultanément. Par exemple, certains produits de la série **S18** et **S24** présentent des points d'intersection orange vif, suggérant une forte complémentarité. La majorité de la matrice est sombre, ce qui indique que la plupart des produits n'ont pas de lien statistique fort entre eux. Cela signifie que l'achat de l'un n'influence pas nécessairement l'achat de l'autre. On observe visuellement des "blocs" de corrélation, notamment parmi les codes commençant par les mêmes préfixes (comme la famille **S24**), ce qui suggère que les clients ont tendance à acheter plusieurs produits au sein d'une même gamme ou catégorie technique.

VIII. Performance des employés :



D'un point de vue statistique et managérial, on observe une hétérogénéité marquée dans les performances. Il existe un ratio de près de 1 pour 4 entre le moins performant (Thompson) et le plus performant (Hernandez). Cet écart suggère soit une disparité dans les portefeuilles clients, soit des différences significatives dans les compétences de négociation ou l'ancienneté. Une majorité d'employés (comme Vanauf, Bott, Jones, Marsh et Fixter) se situe dans une fourchette médiane allant de 0,55 à 0,75 million. Ce groupe constitue le "moteur" stable de l'entreprise.

Conclusions :

L'exploitation de la base ClassicModels a permis de mettre en évidence l'apport décisif d'une analyse quantitative appliquée à la gestion commerciale. Les résultats révèlent une concentration significative du chiffre d'affaires sur un nombre restreint de clients et de produits, traduisant une structure asymétrique de la création de valeur. Si cette configuration constitue un levier de performance, elle expose également l'entreprise à un risque de dépendance stratégique. La segmentation RFM confirme l'hétérogénéité du portefeuille client et souligne la nécessité d'une allocation différenciée des ressources commerciales. L'identification d'une saisonnalité des ventes ainsi que de complémentarités entre produits met en lumière des opportunités d'optimisation logistique et marketing. Au-delà des constats empiriques, ce travail illustre l'importance d'une démarche analytique rigoureuse dans la prise de décision. La maîtrise des outils SQL et Python ne relève pas uniquement d'une compétence technique : elle constitue un instrument fondamental d'aide au pilotage, permettant de réduire l'incertitude, d'améliorer l'allocation des ressources et de soutenir une croissance durable fondée sur l'analyse économique des données.