

BASKARA Havar

BIDABI M1

2025/2026

Partie 1 – Analyse exploratoire de la base de données

Dans le cadre de ce projet, nous avons été amenés à analyser la base de données ClassicModels, qui modélise l'activité d'un distributeur international spécialisé dans la vente de modèles réduits de véhicules de collection. L'objectif de cette première phase était de comprendre la structure globale de la base, d'identifier les principaux indicateurs de performance et de dégager des premières tendances stratégiques à partir de requêtes SQL.

La phase est composée de huit tables principales, les clients, les employés, les bureaux, les commandes, les détails de commandes, les produits, les gammes de produits et les paiements. Cette organisation relationnelle permet de suivre l'ensemble du cycle commercial, depuis l'acquisition du client jusqu'au paiement final, en passant par la gestion des produits et des équipes commerciales.

L'analyse du volume de données montre que l'entreprise travaille avec **122 clients**; gérés par **23 employés** répartis dans **7 bureaux**. Au total, **326 commandes** ont été enregistrées, représentant près de **3000 lignes de détail**. Cela signifie que chaque commande contient en moyenne plusieurs produits, ce qui suggère un modèle orienté vers des achats en volume. Le catalogue comprend **110 produits** répartis en **7 gammes** différentes. Ces chiffres indiquent une structure relativement concentrée, avec une clientèle limitée mais active.

L'étude de la **répartition géographique des clients** révèle une forte concentration aux **États-Unis**, qui représentent environ un tiers de la clientèle totale. **L'Allemagne et la France** suivent, avec respectivement **13 et 12 clients**. Cette répartition montre que l'entreprise possède une présence internationale, mais reste fortement dépendante du marché américain. Cette concentration géographique peut constituer un risque stratégique en cas de ralentissement économique local, mais elle reflète également un positionnement solide sur un marché clé.

Concernant la **gestion des commandes**, l'analyse des statuts montre que plus de **93% des commandes** sont livrées avec succès. Les annulations, litiges et mises en attente restent très faibles. Cela traduit une organisation logistique performante et une bonne maîtrise du processus de vente. Le faible taux de commandes annulées est un indicateur positif de satisfaction client et de fiabilité opérationnelle.

Le **chiffre d'affaires** total généré par l'entreprise s'élève à **9,6 millions**. Ce montant témoigne d'une activité commerciale soutenue. Le **panier moyen**, estimé à environ **29 460 par commande**, est particulièrement élevé. Ce niveau suggère que l'entreprise travaille principalement en **B2B**, avec des distributeurs ou détaillants qui commandent en grande quantité. Le modèle économique repose donc davantage sur la valeur des transactions que sur leur fréquence.

L'analyse des **produits les plus vendus** met en évidence une forte domination des modèles **automobiles classiques**, notamment la **Ferrari 360 Spider**, qui arrive largement en tête. Cette tendance confirme que la gamme des voitures classiques constitue le **cœur de l'activité**.

commerciale. Les modèles vintage et iconiques semblent particulièrement appréciés, ce qui peut s'expliquer par une clientèle composée de collectionneurs ou de boutiques spécialisées.

L'étude des meilleurs clients montre que 2 entreprises génèrent à elles seules une part significative du chiffre d'affaires.

Enfin, l'analyse des **performances de employés** révèle qu'un commercial se distingue nettement des autres en termes de chiffre d'affaires généré. Cette information peut servir de base à une réflexion sur la politique de rémunération variable ou sur le partage des bonnes pratiques commerciales au sein des équipes.

Globalement, cette première analyse exploratoire met en évidence une entreprise rentable, structurée et performante, mais présentant certaines dépendances stratégiques, notamment vis-à-vis du marché américain et de quelques clients majeurs. Ces premières constats servent de base pour approfondir l'analyse à travers des requêtes SQL plus avancées.

Partie 2 – Requêtes SQL avancées

La seconde partie du projet avait pour objectif d'exploiter des techniques SQL plus avancées afin d'obtenir une vision plus fine et stratégique de l'activité de l'entreprise.

La première analyse avancée a porté sur **l'évolution du chiffre d'affaires mensuel**, accompagnée d'une **moyenne mobile sur 3 mois**. Les résultats montrent une **forte saisonnalité des ventes**, avec des pics très marqués au **quatrième trimestre**, en particulier en **fin d'année**. Cette concentration des ventes suggère une forte influence des périodes festives ou des cycles budgétaires des clients professionnels. L'entreprise doit donc anticiper ses besoins logistiques et ses stocks avant cette période clé.

L'analyse de la **hiérarchie organisationnelle** à l'aide d'un CTE récursif a permis de reconstruire la structure interne de l'entreprise. On observe une organisation pyramidale classique, avec un dirigeant au sommet, plusieurs managers intermédiaires et des commerciaux rattachés à ces responsables. Cette structure facilite la supervision et la répartition des responsabilité. Elle pourrait également servir de base pour analyser la performance par niveau hiérarchique.

La **segmentation des clients en catégories VIP et Standard**, réalisée à partir de jointures multiples entre 5 tables, met en évidence une concentration importante du chiffre d'affaires sur un nombre restreint de clients. Seuls **2 clients** se distinguent nettement comme VIP, générant des montants bien supérieurs aux autres. Cette concentration renforce l'idée d'une dépendance commerciale. Une stratégie ciblée pourrait consister à développer des offres spécifiques pour ses clients clés tout en travaillant à l'augmentation du chiffre d'affaires des clients standards.

L'analyse **temporelle des nouveaux clients**, réalisée grâce à une sous-enquête corrélée permettant d'identifier la première commande de chaque client, révèle une **forte acquisition en 2003** suivi d'un **ralentissement en 2004**. Cette baisse peut indiquer un changement de stratégie, un ralentissement du marché ou une stabilisation du portefeuille client.

Enfin, le rapport de type pivot comparant les ventes par gamme de produits et par trimestre confirme la domination de la gamme Classic cars et l'importance stratégiques du 4^{ème} trimestre pour toutes les gammes.

Dans l'ensemble, ces requêtes avancées permettent de mieux comprendre non seulement les performances actuelles de l'entreprise, mais également ses dynamiques internes et saisonnières.

Partie 3 – Analyse exploratoire et visualisation avec Python

Dans cette troisième partie, l'objectifs était d'exploiter les données extraites en SQL afin de produire une analyse visuelle et interprétative à l'aide de Python (Pandas & Matplotlib). Cette approche permet de transformer des résultats bruts en indicateurs clairs et exploratoires pour la prise de décision.

3.1 Évolution mensuelle du chiffre d'affaires

L'analyse temporelle met en évidence une forte variabilité du chiffre d'affaires entre 2003 et 2005. On observe une progression globale du chiffre d'affaires, avec 2 pics particulièrement marqués, une en fin **2003** avec environ **980 000€** et une en fin **2004** avec près de **1 000 000€**. Ces pics traduisent une forte saisonnalité, probablement liée à des périodes commerciales stratégiques (fin d'année, fêtes...). En dehors de ces périodes, le chiffre d'affaires mensuel oscille généralement entre 150 000€ et 350 000€ en 2003 et 250 000€ et 450 000€ en 2004 et 2005. Donc l'activité est saisonnière, mais globalement en croissance.

3.2 Répartition du chiffre d'affaires par pays

L'analyse géographique montre une concentration très forte du chiffre d'affaires. Les **États-Unis** dominent largement avec environ **3 000 000€**. Ce montant représente une part significative du chiffre d'affaires total. Les pays suivants sont loin derrière, **l'Espagne** avec environ 1 100 000€, la **France** avec 1 000 000€, **l'Australie** avec 550 000€ et enfin le **Royaume-Uni** avec environ 500 000€. Les autres pays se situent majoritairement sous les 300 000€. Donc l'entreprise est très dépendante du marché américain. Une diversification géographique pourrait réduire le risque commercial.

3.3 Ventes par gamme de produits

L'analyse par gamme révèle une forte disparité entre les catégories. La gamme la plus performante est le **Classic cars** avec environ à lui seul **3 800 000€**. Cette gamme représente à elle seule une part majeure du chiffre d'affaires total.

| Gamme | Chiffre d'affaires |
|----------------|--------------------|
| Classic Cars | 3 800 000€ |
| Vintage cars | 1 800 000€ |
| Motorcycles | 1 100 000€ |
| Trucks & Buses | 1 000 000€ |
| Planes | 950 000€ |
| Ships | 650 000€ |
| Trains | 200 000€ |

La gamme « **Classic Cars** » est clairement le cœur stratégique de l'activité. À l'inverse, la gamme « **Trains** » est marginale et pourrait faire l'objet d'une réflexion stratégique.

3.4 Top 10 des produits les plus vendus

L'analyse des quantités vendues montre que certains produits dominent nettement. Comme en 1992, où il y a eu 1 808 unités vendus de la Ferrari 360 Spider red. Les autres produits du top 10 se situent entre 1 050 et 1 150 unités. Cela montre que le produit phare dépasse significativement les autres références.

3.5 Analyse de l'état des stocks

L'analyse du stock montre une situation globalement saine. Plus de 100 produits sont en stock suffisant. Seulement quelques produits sont en stocks faible. Exemple de produits à surveiller on a :

| Produit à surveiller | Unité |
|--------------------------|-------|
| 1960 BSA Gold Star DBD34 | 15 |
| 1968 Ford Mustang | 68 |
| 1928 Ford Phaeton Deluxe | 136 |
| 1997 BMW F650 ST | 178 |

Même si la majorité des produits ont un stock suffisant, certains modèles à forte valeur pourraient rapidement devenir problématique.

3.6 Top 10 des clients (chiffre d'affaires)

L'analyse client révèle une forte concentration du chiffre d'affaires. Euro+ Shopping Channel à 820 689€ et Mini Gifts Distributors Ldt à 591 827€. Les autres clients du top10 se situent entre 140 000€ et 180 000€. Donc on remarque que l'entreprise dépend fortement de quelques gros clients.

Conclusion de la partie 3 :

L'analyse Python confirme plusieurs éléments clés comme une forte saisonnalité du chiffre d'affaires, une dépendance au marché américain et une domination écrasante de la gamme Classic Cars et une concentration du chiffre d'affaires sur quelques clients majeurs et enfin une situation de stock globalement saine mais nécessitant une surveillance ciblée.

Partie 4 – Analyse avancées et indicateurs stratégiques

Dans cette dernière partie, l'objectif était d'aller plus loin que la simple analyse descriptive. À partir des données exploitées en SQL et traitées sous Python, plusieurs indicateurs stratégiques ont été construits afin d'évaluer la performance commerciale, la qualité du portefeuille clients et certains leviers d'optimisation.

4.1 Analyse RFM (Recency, Frequency, Monetary)

L'analyse RFM permet d'évaluer la valeur stratégique des clients à partir de 3 dimensions, Recency (date de la dernière commande), Frequency (nombre de commande) et Monetary

(chiffre d'affaires). En observant les résultats, on remarque immédiatement la présence de client très dominants. Par exemple, Euro+ Shopping Channel affiche un chiffre d'affaires supérieur à 820 000€ avec une fréquence extrêmement élevée (259 commandes). D'autres clients comme Mini Gifts Distributors Ltd dépassent également les 590 000€. À l'inverse, la majorité des clients se situent entre 50 000€ et 150 000€ de chiffre d'affaires. Le graphique Recency vs Monetary montre clairement une concentration importante, quelques points très élevé (gros clients), et une masse plus compacte de clients moyens.

Donc l'entreprise possède un petit noyau de clients à très forte valeur qui génèrent une part significative du chiffre d'affaires. Cela représente une opportunité de fidélisation...mais aussi un risque en cas de perte d'un de ces comptes majeurs.

4.2 Corrélation prix/quantité

L'analyse de corrélation entre le prix unitaire et la quantité commandée donne un coefficient d'environ 0,025. Ce chiffre est extrêmement proche de 0, cela signifie qu'il existe pratiquement aucune corrélation linéaire entre le prix d'un produit et la quantité vendue. En d'autres termes, les produits plus chers ne se vendent pas nécessairement moins et les produits moins chers ne se vendent pas automatiquement davantage. Cette information est stratégique, elle suggère que la demande n'est pas uniquement déterminée par le prix, mais probablement par la rareté du modèle, la valeur perçus et l'effet collection.

4.3 tendance temporelle des ventes

L'analyse de la tendance mensuelle conforme ce qui avait été observé précédemment, une saisonnalité très marquée. On observe grâce aux résultats du codes 2 pics majeurs, fin 2003 (980 000€) et fin 2004 (970 000€). Entre ces pics, le chiffre d'affaires mensuel oscille généralement entre 200 00€ et 400 000€. On constate également une progression globale entre 2003 et 2005, malgré les fluctuations.

4.4 Analyse du panier moyen

Le calcul du panier moyen donne un résultat très intéressant, le panier moyen est de 2 946,71 €, cela signifie qu'en moyenne, chaque commande représente près de 3 000€ de chiffre d'affaires. Ce montant est relativement élevé et confirme que l'entreprise évolue sur un marché spécialisé, avec des produits à forte valeur unitaire.

4.5 Performance des employés

L'analyse des performances commerciales met en évidence des écarts significatifs entre les employés. Gerard Hernandez, environ 1 300 000€ suivi de Leslie Jennings avec environ 1 100 000€ et elle suivi de Pamela Castillo avec environ 800 000€. À l'autre extrémité certains employés se situent autour de 350 000€ et 450 000€. La performance n'est pas homogène au sein de l'équipe commerciale.

Conclusion de la partie 4 :

Cette analyse avancée met en lumière plusieurs points clés, une forte dépendance à un petit nombre de clients à très forte valeur, une absence de relation significative entre prix et quantité vendue, une saisonnalité marquée avec des pics majeurs en fin d'année, un panier moyen élevé (environ 2950€) et une disparité notables dans la performance commerciale des employés.

CONCLUSION :

Ce projet a permis de mettre en évidence l'intérêt stratégique de l'analyse de données dans la compréhension de la performance d'une entreprise. À travers l'exploitation de la base Classic Models, nous avons pu identifier des tendances claires, notamment une forte saisonnalité des ventes avec des pics importants fin d'année, ainsi qu'une concentration du chiffre d'affaire sur certains marchés et certaines gammes de produits.

L'analyse des clients montre que quelques comptes génèrent une part significative des revenus, ce qui souligne l'importance de leur fidélisation. Le panier moyen relativement élevé confirme le positionnement orienté vers des produits à forte valeur. De plus, l'étude des performances commerciales met en évidence des écarts entre employés, ouvrant des pistes d'amélioration organisationnelle.

Globalement, l'entreprise présente une situation solide, mais certaines dépendances (clients majeurs, marchés dominants, produits phares) constituent des points de vigilance. Ce travail démontre ainsi que l'analyse des chiffres, mais qu'elle permet surtout d'éclairer la prise de décision et d'orienter la stratégie future.