

Analyse de la performance commerciale de Deskly

Partie 1 : Définition des objectifs et modélisation dimensionnelle.....	2
T1 - Les objectifs du SID.....	2
Objectif A : Augmenter la rentabilité des ventes.....	2
Objectif B : Améliorer la gestion des stocks.....	3
T2 - Sources de données nécessaires.....	5
Objectif A : Augmenter la rentabilité des ventes.....	5
Objectif B : Améliorer la gestion des stocks.....	5
T3 - Modèle Dimensionnel.....	7
T4 - Classification des mesures.....	12
T5 - Opérations de transformation sous Python.....	14
Partie 2 : Nettoyage, transformation et structuration des données.....	17
Partie 3 : Tableaux de bord.....	18
Analyse et recommandations.....	20
Objectif A : Augmenter la rentabilité des ventes.....	20
Objectif B : Améliorer la gestion des stocks.....	21
Conclusion générale.....	22

Partie 1 : Définition des objectifs et modélisation dimensionnelle

T1 - Les objectifs du SID

Nous avons défini les objectifs du SID suivants :

- Objectif A : Augmenter la rentabilité des ventes
- Objectif B : Améliorer la gestion des stocks

Objectif A : Augmenter la rentabilité des ventes

1. Domaine concerné

- Ventes

2. Objectif métier du SID

Le SID doit permettre d'analyser les performances de ventes de Deskly (notamment identifier les tendances, les segments les plus rentables et les opportunités de croissance), afin d'améliorer les performances commerciales de l'entreprise.

3. Finalité stratégique

La stratégie consiste à identifier les produits les plus rentables ainsi que ceux qui génèrent peu ou pas de bénéfices, dans le but d'optimiser la rentabilité globale de l'entreprise et d'améliorer les marges bénéficiaires. Pour cela, une analyse détaillée sera menée selon plusieurs axes : les départements, les clients et les villes.

4. Acteurs utilisateurs du système

Dans le cadre de ce projet décisionnel visant à analyser la rentabilité des produits vendus par Deskly et à identifier les leviers d'optimisation commerciale, plusieurs acteurs internes seront directement concernés par l'utilisation du Système d'Information Décisionnel (SID).

- Directeur commercial
- Responsables des ventes régionaux
- Gestionnaires de produits
- Analystes business

5. Indicateurs nécessaire

Les indicateurs suivants seront mesurés et analysés afin de répondre à l'objectif principal d'optimisation de la rentabilité des ventes :

- Chiffre d'affaires (Sales)
- Profit (marge bénéficiaire)
- Quantité vendue
- Nombre de commandes
- Rentabilité par client
- Rentabilité par ville ou région
- prix moyen de vente

6. Axes d'analyse pertinents

Les indicateurs annoncés seront analysés selon les axes suivants afin de dégager des tendances pertinentes et orienter les décisions stratégiques :

- Temps (date de commande, année, mois, trimestre)
- Géographie (région, ville, état)
- Produit (catégorie/département, sous-catégorie, produit)
- Client (segment client, client individuel)
- Département (furniture, office supplies, Technology)

Objectif B : Améliorer la gestion des stocks

1. Domaine concerné

- Approvisionnement et logistique
- Gestion des stocks

2. Objectif métier du SID

Le SID doit permettre d'optimiser les niveaux de stock, de réduire les ruptures et les surstocks, d'améliorer la rotation des produits et de synchroniser l'approvisionnement avec la demande réelle pour minimiser les coûts de stockage tout en maintenant un niveau de service optimal.

3. Finalité stratégique

- Réduire les coûts de stockage
- Diminuer le taux de rupture de stock
- Améliorer la rotation des stocks
- Optimiser la répartition géographique des stocks

4. Acteurs utilisateurs du système

- Directeur des opérations
- Responsables d'entrepôt
- Gestionnaires de stock
- Supply chain analysts
- Responsable approvisionnement

5. Indicateurs nécessaire

Les indicateurs suivants seront mesurés et analysés afin de répondre à l'objectif principal d'améliorer la gestion des stocks :

- Taux de rotation des stocks
- Niveau de stock moyen
- Taux de rupture de stock
- Taux de surstock
- Délai moyen de réapprovisionnement

6. Axes d'analyse pertinents

Les indicateurs annoncés seront analysés selon les axes suivants afin de dégager des tendances pertinentes et orienter les décisions stratégiques :

- Temps (date de commande, année, mois, trimestre)
- Géographie (région, ville, état)
- Produit (catégorie/département, sous-catégorie, produit)
- Entrepôt / Site logistique
- Fournisseur

T2 - Sources de données nécessaires

Objectif A : Augmenter la rentabilité des ventes

Les sources de données pour les indicateurs sont les suivantes :

- Chiffre d'affaires : Sales (`order_details.csv`)
- Profit (marge bénéficiaire) : Profit (`order_details.csv`)
- Quantité vendue : Quantity (`order_details.csv`)
- Nombre de commandes : Order ID (`orders.csv`)
- Rentabilité par client : via Customer ID et Segment (`customers.csv`)
- Rentabilité par région/ville : via City, Region, State (`addresses.csv`)
- Prix moyen : calculé à partir de Sales / Quantity

Les sources de données pour les axes d'analyse sont les suivantes :

- Temps : Order Date (`orders.csv`)
- Produit : Department, Sub-Category, Product Name (`products.csv`)
- Client : Customer Name, Segment (`customers.csv`)
- Géographie : City, State, Region (`addresses.csv`)
- Département : Department (`products.csv`)

Objectif B : Améliorer la gestion des stocks

Les sources de données pour les indicateurs sont les suivantes :

- Taux de rotation : via Quantity, Sales par période (`order_details.csv`)
- Niveau de stock moyen : à estimer via Storage Capacity (`departments.csv`)
- Taux de rupture / surstock : estimations par croisement de volume vs. capacité
- Délai moyen de réapprovisionnement : basé sur Order Date vs Ship Date (`orders.csv`)

Les sources de données pour les axes d'analyse sont les suivantes :

- Temps : Order Date, Ship Date (`orders.csv`)
- Produit : Product ID, Department, Sub-Category (`products.csv`)
- Géographie : City, State, Region (`addresses.csv`)
- Entrepôt / site : City + Storage Capacity (`departments.csv`)
- Fournisseur : (non présent, à simuler ou enrichir éventuellement)

T3 - Modèle Dimensionnel

Pour répondre aux deux objectifs stratégiques (rentabilité des ventes et optimisation des stocks), un modèle dimensionnel en étoile a été construit. Il repose sur deux tables de faits principales, chacune associée à un ensemble de dimensions analytiques.

Identification des tables de faits

Nous avons identifié deux tables de faits d'après les objectifs définis :

1. Table de faits VENTES (FACT_SALES)

Cette table contiendra les métriques liées aux transactions de vente :

- Chiffre d'affaires (Sales)
- Profit
- Quantité vendue (Quantity)
- Montant de la remise (Discount)
- Délai de livraison (calculé à partir de Order_Date et Ship_Date)

2. Table de faits STOCKS (FACT_INVENTORY)

Cette table contiendra les métriques liées à la gestion des stocks :

- Quantité vendue par département et produit (agrégée depuis FACT_SALES)
- Capacité de stockage (depuis departments.csv)
- Taux d'utilisation estimé (calculé)

Identification des dimensions

Nous avons également identifié plusieurs dimensions :

1. Dimension TEMPS (DIM_TIME)

- ID_Temps (clé primaire)
- Date
- Jour de la semaine
- Mois
- Trimestre
- Année
- Saison

2. Dimension PRODUIT (DIM_PRODUCT)

- ID_Produit (clé primaire)
- Nom du produit (Product Name)
- Sous-catégorie (Sub-Category)
- Département (Department)

3. Dimension CLIENT (DIM_CUSTOMER)

- ID_Client (clé primaire)
- Nom du client (Customer Name)
- Segment (Consumer, Corporate, Home Office)
- Date de première commande (calculée à partir de orders.csv)
- Nombre total de commandes (calculé)

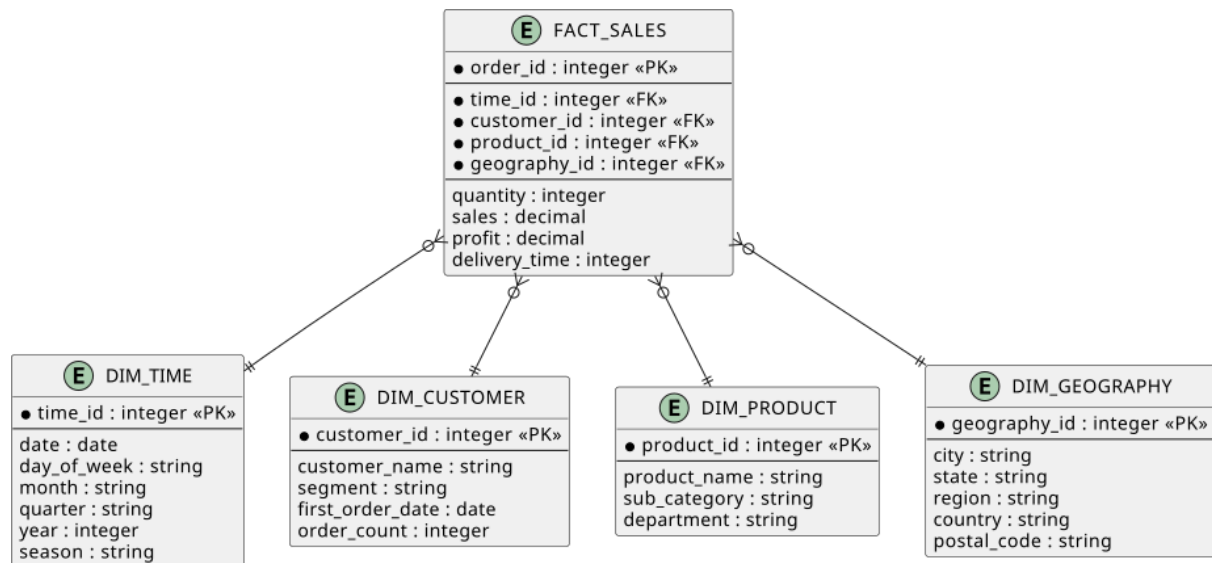
4. Dimension GÉOGRAPHIE (DIM_GEOGRAPHY)

- ID_Géographie (clé primaire)
- Ville (City)
- État/Province (State)
- Région (Region)
- Pays (Country)
- Code postal (Postal Code)

5. Dimension DÉPARTEMENT/ENTREPÔT (DIM_DEPARTMENT)

- ID_Département (clé primaire)
- Nom du département (Department)
- Ville (City)
- Capacité de stockage (Storage Capacity)

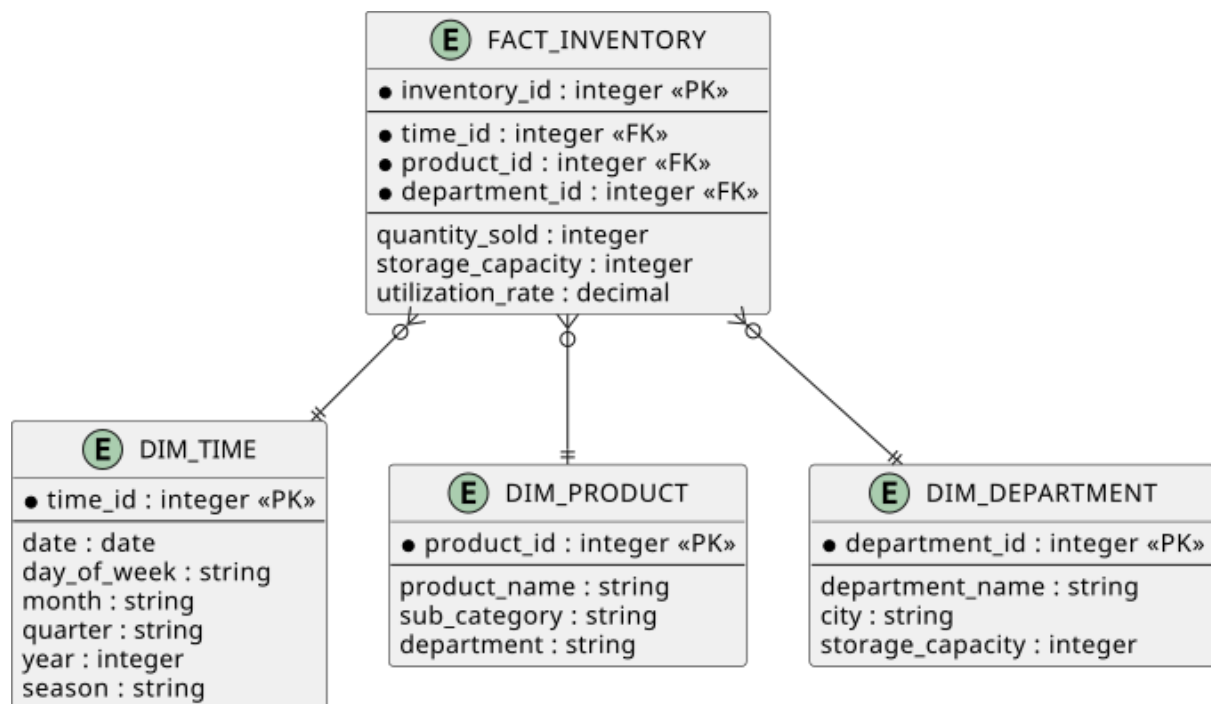
Schéma en étoile pour l'objectif A : Augmenter la rentabilité des ventes



FACT_SALES

- ID_Temps (FK)
- ID_Client (FK)
- ID_Produit (FK)
- ID_Géographie (FK)
- Order ID (Numéro de commande)
- Quantité vendue (Quantity)
- Chiffre d'affaires (Sales)
- Profit
- Délai de livraison (Ship_Date - Order_Date)

Schéma en étoile pour l'objectif B : Améliorer la gestion des stocks



FACT_INVENTORY

- ID_Temps (FK)
- ID_Produit (FK)
- ID_Département (FK)
- Quantité vendue (agrégée depuis FACT_SALES)
- Capacité de stockage (depuis DIM_DEPARTMENT)
- Taux d'utilisation estimé (calculé)

Mesures calculées

Pour l'objectif A (Rentabilité des ventes) :

- Marge par produit : Profit / Chiffre d'affaires
- Taux de rentabilité : Profit / Chiffre d'affaires
- Panier moyen : Chiffre d'affaires / Nombre de commandes
- Délai de livraison moyen : Date d'expédition - Date de commande

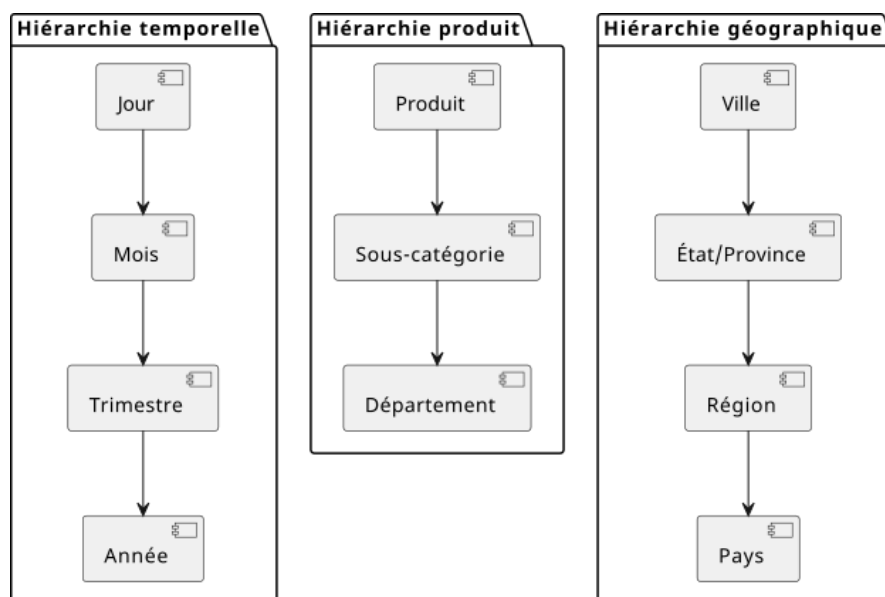
Pour l'objectif B (Gestion des stocks) :

- Taux d'utilisation de la capacité : Quantité vendue / Capacité de stockage (estimation)
- Répartition des ventes par département : % des ventes totales par département
- Fréquence de réapprovisionnement estimée : basée sur les dates de commande

Hiérarchies dimensionnelles

Pour faciliter les analyses à différents niveaux de granularité, nous définissons les hiérarchies suivantes :

- **Hiérarchie temporelle** : Jour → Mois → Trimestre → Année
- **Hiérarchie produit** : Produit → Sous-catégorie → Département
- **Hiérarchie géographique** : Ville → État/Province → Région → Pays



Cette modélisation dimensionnelle permettra aux utilisateurs d'analyser les données selon différentes perspectives et à différents niveaux de détail. Cela facilitera la prise de décision pour atteindre les objectifs d'amélioration de la rentabilité des ventes et d'optimisation de la gestion des stocks.

T4 - Classification des mesures

Voici la classification des mesures identifiées dans notre modèle selon leur caractère additif, semi-additif ou non-additif.

Mesures additives

Les mesures additives peuvent être agrégées (sommées) selon toutes les dimensions sans restriction. Elles produisent des résultats cohérents quelle que soit la dimension d'analyse.

Dans **FACT_SALES** :

- **Quantité vendue (Quantity)** : Les quantités peuvent être sommées par produit, par client, par période, etc.
- **Chiffre d'affaires (Sales)** : Le montant des ventes peut être additionné sur toutes les dimensions.
- **Profit** : Les bénéfices peuvent être sommés sur toutes les dimensions.
- **Nombre de commandes** : Le décompte des commandes peut être additionné sur toutes les dimensions.

Dans **FACT_INVENTORY** :

- **Quantité vendue (agrégée depuis FACT_SALES)** : Les quantités vendues peuvent être sommées sur toutes les dimensions.

Les mesures additives peuvent être pré-agrégées à différents niveaux pour améliorer les performances des requêtes, sans risque d'erreur d'interprétation.

Mesures semi-additives

Les mesures semi-additives peuvent être agrégées selon certaines dimensions mais pas toutes. Typiquement, elles ne peuvent pas être additionnées sur la dimension temporelle.

Dans **FACT_INVENTORY** :

- **Capacité de stockage** : Cette mesure représente un état à un moment donné. Elle peut être additionnée par département, mais pas à travers le temps.

Pour les mesures semi-additives, des précautions doivent être prises lors de l'agrégation temporelle. Par exemple, pour le niveau de stock, on pourrait utiliser des moyennes, des valeurs de fin de période, ou des minimums/maximums selon le contexte d'analyse.

Mesures non-additives

Les mesures non-additives ne peuvent pas être agrégées par simple addition, quelle que soit la dimension. Elles nécessitent généralement d'autres types d'agrégation (moyenne, minimum, maximum) ou des calculs spécifiques.

Dans FACT_SALES :

- **Marge par produit (Profit/Sales)** : Ce ratio ne peut pas être additionné directement. Pour calculer la marge globale, il faut d'abord additionner les profits et les ventes séparément, puis calculer le ratio.
- **Taux de rentabilité (Profit/Sales)** : Comme la marge, ce ratio ne peut pas être additionné directement.
- **Panier moyen (Sales/Nombre de commandes)** : Cette moyenne ne peut pas être additionnée. Pour calculer le panier moyen global, il faut additionner les ventes et le nombre de commandes séparément, puis diviser.
- **Délai de livraison moyen (Ship_Date - Order_Date)** : Les moyennes ne peuvent pas être additionnées directement.

Dans FACT_INVENTORY :

- **Taux d'utilisation estimé (Quantité vendue/Capacité de stockage)** : Ce pourcentage ne peut pas être additionné directement.
- **Répartition des ventes par département** : Ce pourcentage ne peut pas être additionné directement.

Les mesures non-additives doivent être calculées à partir des mesures additives de base. Par exemple, pour calculer le taux de rentabilité global, il faut d'abord additionner tous les profits et toutes les ventes, puis effectuer la division.

Cette classification guidera également la conception des rapports et tableaux de bord, en s'assurant que les agrégations utilisées sont appropriées pour chaque type de mesure, évitant ainsi les erreurs d'interprétation qui pourraient conduire à de mauvaises décisions commerciales.

T5 - Opérations de transformation sous Python

Pour alimenter notre modèle dimensionnel à partir des sources de données identifiées, plusieurs opérations de transformation seront nécessaires. Voici la liste des opérations de transformation à effectuer sous Python, organisées par dimension et table de faits.

Transformations pour les dimensions

DIM_TIME (Dimension Temps)

1. Extraction des dates uniques à partir des champs Order Date et Ship Date du fichier `orders.csv`
2. Génération des attributs temporels pour chaque date :
 - Jour de la semaine
 - Mois
 - Trimestre
 - Année
 - Saison

DIM_PRODUCT (Dimension Produit)

1. Extraction des données produits depuis `products.csv`
2. Normalisation des noms de produits
3. Création d'une hiérarchie simple : Produit → Sous-catégorie → Département
4. Gestion des valeurs manquantes pour les attributs produits

DIM_CUSTOMER (Dimension Client)

1. Extraction des données clients depuis `customers.csv`
2. Normalisation des noms de clients
3. Calcul de la date de première commande pour chaque client (à partir de `orders.csv`)
4. Calcul du nombre total de commandes par client
5. Gestion des doublons potentiels dans les données clients

DIM_GEOGRAPHY (Dimension Géographie)

1. Extraction des données géographiques depuis `addresses.csv`
2. Normalisation des noms de villes, états et régions
3. Correction des erreurs typographiques dans les noms géographiques
4. Standardisation des codes postaux

DIM_DEPARTMENT (Dimension Département/Entrepôt)

1. Extraction des données de département depuis `departments.csv`
2. Normalisation des noms de département

Transformations pour les tables de faits

FACT_SALES (Table de faits Ventes)

1. Jointure entre `orders.csv` et `order_details.csv` sur `Order_ID`
2. Création des clés étrangères vers les dimensions :
 - Recherche des `ID_Temps` correspondant aux dates de commande
 - Recherche des `ID_Client` pour chaque `Customer_ID`
 - Recherche des `ID_Produit` pour chaque `Product_ID`
 - Recherche des `ID_Géographie` basés sur les adresses
3. Calcul du délai de livraison (`Ship_Date - Order_Date`)
4. Vérification et correction des valeurs aberrantes dans les quantités et montants
5. Agrégation des lignes de commande si nécessaire (selon la granularité choisie)

FACT_INVENTORY (Table de faits Stocks)

1. Agrégation des données de vente par produit et département
2. Jointure avec les données de capacité de stockage depuis `departments.csv`
3. Création des clés étrangères vers les dimensions :
 - Recherche des `ID_Temps` pour les périodes d'analyse
 - Recherche des `ID_Produit` pour chaque produit
 - Recherche des `ID_Département` pour chaque département
4. Calcul des métriques dérivées :
 - Taux d'utilisation estimé (basé sur les ventes et la capacité de stockage)
 - Répartition des ventes par département

Opérations générales de transformation

Nettoyage des données :

1. Suppression des doublons
2. Traitement des valeurs nulles ou manquantes
3. Correction des erreurs typographiques
4. Standardisation des formats (dates, nombres, textes)

Validation des données :

1. Vérification des contraintes d'intégrité
2. Détection des valeurs aberrantes
3. Contrôle de cohérence entre les différentes sources

Enrichissement des données :

1. Calcul de métriques dérivées
2. Ajout d'indicateurs de performance

Optimisation :

1. Agrégation préalable de certaines métriques
2. Création d'index pour améliorer les performances

Cette liste d'opérations de transformation servira de base pour développer le processus ETL (Extract, Transform, Load) qui alimentera notre entrepôt de données et permettra de répondre efficacement aux objectifs d'analyse définis pour le SID de Deskly.

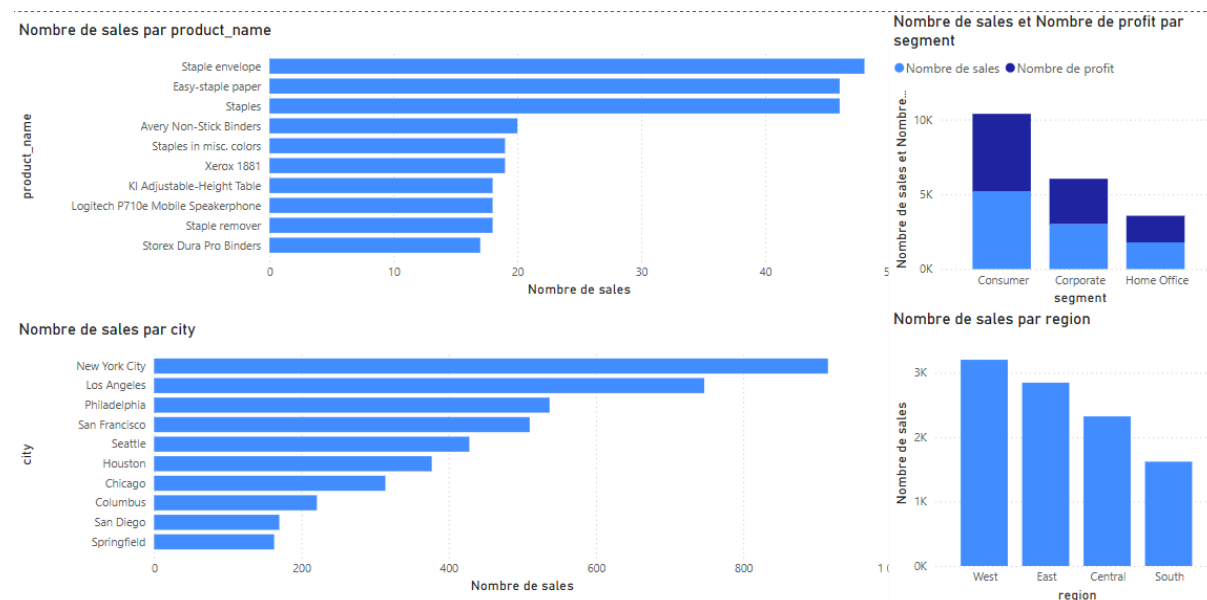
Partie 2 : Nettoyage, transformation et structuration des données

Dans cette deuxième phase du projet, nous avons vérifié la cohérence du modèle défini précédemment, puis importé les fichiers CSV dans des dataframes pour analyser leur qualité.

Un nettoyage des données a été effectué (corrections, suppressions, normalisation). Ensuite, les jointures nécessaires ont permis de construire une structure en étoile avec une table de faits et des tables de dimensions. Les tables finales ([objectif A](#) et [objectif B](#)) ont été exportées en CSV et ont été enregistrées dans le dossier 'output_tables'. Enfin, toutes les étapes ont été documentées dans le notebook Python fourni.

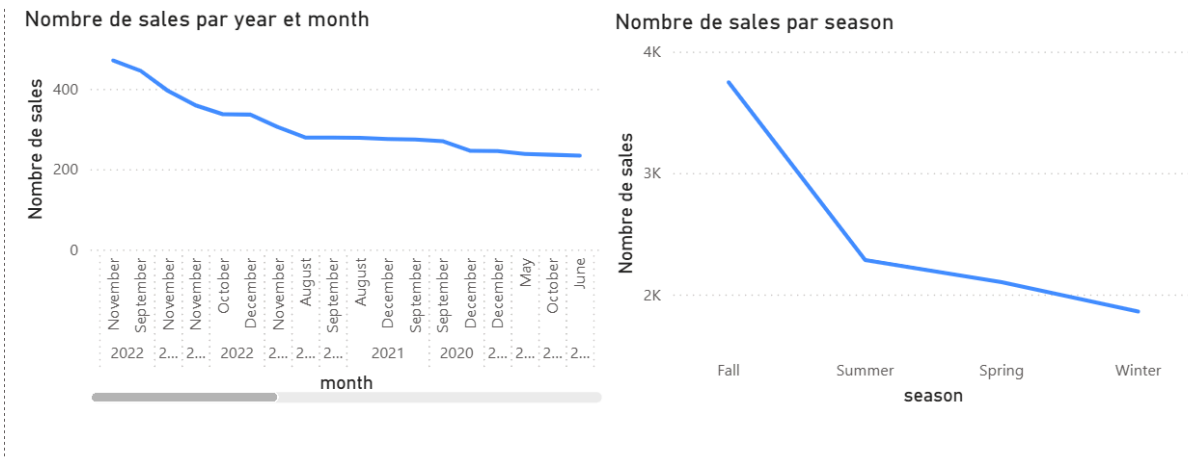
Partie 3 : Tableaux de bord

Dans cette dernière étape du projet, les données nettoyées ont été importées dans Power BI, ce qui nous a permis de créer un ensemble de tableaux de bord interactifs. Cela facilite la visualisation des indicateurs clés et l'analyse des performances.



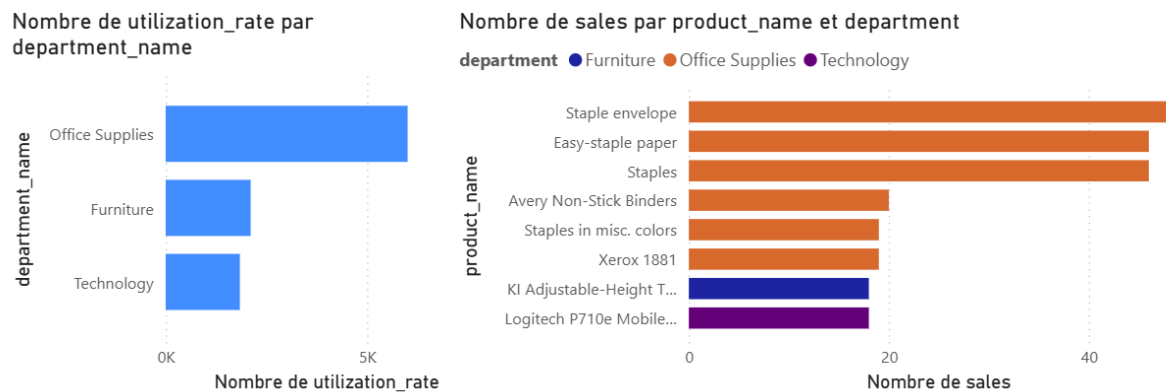
Le premier tableau de bord ci-dessus permet de visualiser les ventes et bénéfices de l'entreprise :

- Le premier graphique en haut à gauche représente le prix de chaque produit, ce qui permet de voir les plus vendus.
- Le graphique en bas à gauche permet de visualiser le nombre de ventes en fonction de la ville.
- Le graphique en haut à droite permet de voir la rentabilité et le nombre de ventes en fonction du segment de marché.
- Le graphique en bas à droite indique le nombre de ventes par région.



Ce second tableau de bord est axé sur le temps :

- La courbe de gauche permet de visualiser le nombre de ventes par année et par mois.
- La courbe de droite montre la quantité de vente en fonction de la saisonnalité.



Ce troisième et dernier tableau de bord se concentre sur l'aspect du stockage :

- A gauche, le graphique affiche la quantité de produits stockés selon les départements.
- A droite, le graphique affiche le nombre de ventes par produit et le département auquel appartient ce produit.

Analyse et recommandations

Cette section présente une synthèse des principaux résultats issus des visualisations Power BI, accompagnée de recommandations concrètes pour orienter les décisions stratégiques et opérationnelles de Deskly.

Objectif A : Augmenter la rentabilité des ventes

Les produits les plus vendus

Les visualisations montrent que certaines références comme "Staple Envelope", "Easy-staple Paper" et "Staples" génèrent une part importante du chiffre d'affaires.

Recommandation :

- Renforcer la promotion des produits les plus rentables
- Analyser les marges de ces produits pour ajuster les prix ou les volumes

Les villes les plus attractives

Les villes comme New York City, Los Angeles et Philadelphia concentrent la majorité des ventes.

Recommandation :

- Cibler ces régions avec des opérations marketing prioritaires
- Analyser les zones peu actives pour détecter des opportunités ou corriger des problèmes de distribution

Les performance selon le segment client

Le segment Consumer surpasse nettement les segments Corporate et Home Office en volume et en rentabilité.

Recommandation :

- Adapter l'offre aux besoins de chaque segment
- Envisager des campagnes spécifiques pour relancer les segments moins actifs

L'impact des saisons

Une baisse des ventes est visible entre l'automne et l'hiver.

Recommandation :

- Lancer des promotions ou offres ciblées pendant les périodes creuses

Objectif B : Améliorer la gestion des stocks

L'utilisation de la capacité

Le taux d'utilisation est nettement plus élevé dans les départements comme Office Supplies comparé à Furniture ou Technology.

Recommandation :

- Rééquilibrer les volumes stockés entre les départements
- Réévaluer les besoins réels de chaque secteur

La synchronisation des stocks avec la demande

Certains produits à faible rotation occupent un espace important.

Recommandation :

- Améliorer les prévisions de vente
- Adapter la politique de réapprovisionnement selon la rotation des stocks

La saturation des capacités locales

Certaines localisations montrent un taux de stockage proche de la saturation.

Recommandation :

- Optimiser la logistique inter-entrepôts
- Prévoir des débordements saisonniers avec des espaces de stock temporaires

Conclusion générale

Les tableaux de bord développés avec Power BI aident à repérer les produits les plus rentables, les clients et régions les plus importants, ainsi que les points à améliorer dans la gestion des stocks. Grâce à ces informations, Deskly peut augmenter sa rentabilité et mieux adapter ses stocks à la demande réelle.