**Data Market (Sales Data Mart) : E1 Note de cadrage**

### **1. Introduction**

### **1.1 Contexte du projet**

Le e-commerce est aujourd’hui un secteur très concurrentiel. Les entreprises doivent utiliser leurs données pour mieux comprendre le marché, suivre la performance et anticiper les changements.

Dans notre société, l’accès aux données n’est pas optimal. Les informations sont réparties dans plusieurs fichiers, avec des formats différents. Cela rend l’analyse complexe et lente. Les équipes métier ne peuvent pas toujours obtenir rapidement les informations dont elles ont besoin pour décider.

### **1.2 Problèmes rencontrés**

Nous faisons face à plusieurs difficultés :

* Les données sont dispersées entre plusieurs sources.
* Les formats sont parfois incompatibles.
* L’accès à l’information demande du temps et des compétences techniques.
* Il n’y a pas de référentiel unique pour les indicateurs clés.

### **1.3 Objectifs du Data Market**

Afin de surmonter ces obstacles, l’entreprise lance la création d’un **Data Market** interne :

* **Centraliser et fiabiliser la donnée :** constitution d’un socle de données propres, normalisées et documentées (« Sales Data Mart »).
* **Favoriser l’autonomie des équipes métier :** permettre aux utilisateurs non techniques d’accéder et d’exploiter la donnée en libre-service.
* **Réduire les délais d’accès à l’information :** accélérer la production de rapports, la prise de décision et la réactivité face au marché.
* **Préparer le terrain pour l’innovation analytique :** mise à disposition d’une base solide pour développer des cas d’usage avancés (prévision, segmentation, pricing…).

### **1.4 Perspectives d’évolution**

Avec cette base de données, l’entreprise pourra aller plus loin :

* Mettre en place des outils de machine learning pour anticiper les tendances ou les risques.
* Développer des modèles de pricing dynamique, par exemple pour ajuster les prix de certains produits déficitaires.
* Améliorer la gestion des risques, comme les retards de livraison ou la baisse de satisfaction client.

### **1.5 Valeur ajoutée**

Ce Data Market donnera à l’entreprise plus de réactivité et d’efficacité. Les équipes pourront travailler avec plus de confiance et prendre de meilleures décisions, même dans un environnement très concurrentiel.

### **2. Expression du besoin data**

### **2.1 Attentes des équipes métiers**

Les équipes métiers – produit, marketing, finance et service client – souhaitent disposer d’un accès direct, fiable et rapide à l’ensemble des données liées aux ventes, aux clients, aux produits, aux paiements et aux avis.

Elles veulent pouvoir :

* Suivre la performance des ventes par période, catégorie de produits, région ou canal.
* Identifier rapidement les segments de clients les plus rentables ou les plus à risque.
* Comprendre les causes des retards de livraison ou des insatisfactions clients.
* Analyser les tendances du marché et anticiper les évolutions de la demande.
* Prendre des décisions basées sur des indicateurs partagés et actualisés.

### **2.2 Limites des systèmes actuels**

Les systèmes et rapports existants présentent plusieurs limites :

* Les données sont dispersées entre différentes sources et formats, ce qui rend la consolidation complexe et chronophage.
* Les rapports sont souvent générés de manière manuelle, avec un risque d’erreur élevé.
* Il existe peu de documentation ou de référentiel unique, ce qui complique la compréhension pour de nouveaux utilisateurs ou pour des analyses transversales.
* L’accès aux données dépend des ressources techniques et peut entraîner des délais importants, freinant la réactivité des équipes.

### **2.3 Améliorations attendues avec un Data Mart de qualité**

La mise en place d’un Data Mart fiable et bien structuré va permettre :

* D’offrir une vision globale et cohérente de l’activité commerciale.
* De faciliter l’automatisation des analyses et la production de rapports standards ou ad hoc.
* De mettre à disposition des données prêtes à l’emploi pour des analyses avancées (segmentation, prévision, modèles de scoring, etc.).
* De préparer l’entreprise à intégrer des solutions de machine learning, comme la prévision des ventes, l’optimisation des stocks ou la détection des anomalies.
* De mieux piloter les politiques de prix, notamment sur les produits ou catégories qui ont généré des pertes, grâce à des données consolidées et actualisées.

### **2.4 Valeur ajoutée pour le business**

Un Data Mart de qualité apportera plusieurs bénéfices concrets à l’entreprise :

* Gain de temps pour les équipes grâce à un accès simplifié et autonome à l’information.
* Meilleure qualité et fiabilité des analyses, avec une réduction des erreurs et des incohérences.
* Accélération de la prise de décision, avec des indicateurs partagés et mis à jour en temps réel.
* Capacité à innover grâce à l’exploitation de nouvelles méthodes analytiques et à la préparation de la transition vers une organisation "data-driven".

### **3. Organisation et organigramme**

### **3.1 Présentation générale de l’organisation**

Le projet Data Market implique plusieurs équipes et acteurs clés de l’entreprise, chacun jouant un rôle spécifique dans la collecte, la gestion et l’exploitation des données.

* **Équipe IT / Data Engineering**Responsable de l’infrastructure technique, de l’intégration, du nettoyage et de la structuration des données.  
  Gère la mise en place des pipelines, l’automatisation et la qualité des flux de données.
* **Équipe Produit (Product Owners, Product Managers)** Définit les besoins métiers et les priorités en matière de données.  
   Participe à la validation des indicateurs et à la formalisation des cas d’usage.
* **Équipe Marketing**Utilise les données pour analyser les comportements clients, piloter les campagnes, segmenter la base client et optimiser l’acquisition ou la fidélisation.
* **Équipe Finance / Contrôle de gestion**Exploite les données pour suivre la rentabilité, analyser les marges, anticiper les risques financiers et piloter les performances.
* **Équipe Service Client**S’appuie sur les données pour suivre la satisfaction client, traiter les réclamations, anticiper les problèmes récurrents et améliorer la qualité de service.
* **Vendeurs / Fournisseurs**Sont à l’origine d’une partie des données (catalogues produits, statuts de livraison, informations sur les stocks), participent indirectement à la qualité et à l’exhaustivité des données collectées.

### **3.2 Exploitation des données : principaux utilisateurs**

|  |  |  |  |
| --- | --- | --- | --- |
| **Rôle** | **Production de données** | **Utilisation des données** | **Besoins principaux** |
| Data Engineer / IT | Intégration, nettoyage, structuration | Assurer la fiabilité et la qualité de la donnée | Automatisation, qualité, documentation |
| Product Owner / Manager | Spécification des besoins, validation | Suivi des indicateurs, pilotage produit | Accès aux KPI, vision transversale |
| Marketing | - | Analyse des clients, campagnes, segmentation | Données clients, historique, taux de conversion |
| Finance / Contrôle gestion | - | Analyse des ventes, marges, reporting financier | Données agrégées, fiabilité des chiffres |
| Service Client | - | Suivi satisfaction, gestion des incidents | Accès rapide aux retours et réclamations |
| Vendeurs / Fournisseurs | Mise à jour catalogue, livraison | Suivi performance, gestion du stock, reporting | Statistiques sur produits, livraison, ventes |

### **4. Activités métiers associées**

L’intégration d’un Sales Data Mart transforme la manière dont les équipes exploitent la donnée au quotidien. Quelques exemples d’activités et d’analyses rendues possibles ou fortement améliorées grâce à ce projet :

### **Exemple 1 : Pilotage quotidien et prise de décision rapide**

Le Sales Data Mart permet également :

* De suivre en temps réel les indicateurs clés (ventes, marges, délais de livraison, satisfaction client).
* De générer automatiquement des rapports opérationnels pour les équipes terrain.
* D’automatiser la détection d’anomalies (par exemple, une baisse soudaine des ventes ou une hausse des retours sur une région précise).

### **Exemple 2 : Analyse des modes de paiement et optimisation de l’offre**

Le Data Mart centralise toutes les informations liées aux paiements. Grâce à cette consolidation, les équipes peuvent :

* Comparer la répartition des modes de paiement (carte, virement, paiement en plusieurs fois…) par segment de clientèle ou par région.
* Détecter si certains moyens de paiement sont associés à un taux de conversion plus élevé ou à moins de litiges.
* Ajuster l’offre commerciale, par exemple en proposant le paiement en plusieurs fois sur des catégories où cela booste les ventes.

### **Exemple 3 : Suivi de la performance des vendeurs et gestion des risques**

En reliant les commandes, les avis clients et les délais de livraison, il devient possible :

* D’évaluer la performance des vendeurs sur le volume de ventes, la qualité de service (respect des délais, satisfaction client, taux de retour).
* Identifier rapidement les vendeurs à risque (taux de réclamation élevé, retards fréquents).
* D’accompagner les vendeurs dans l’amélioration de leur offre.

### **5. Identification des sources de données**

### **5.1 Tableau des sources de données**

|  |  |  |  |
| --- | --- | --- | --- |
| **Source** | **Description** | **Utilité principale dans le projet** | **Clé(s) de jointure** |
| olist\_orders\_dataset.csv | Liste des commandes (statuts, dates, clients associés…) | Base pour l’analyse des ventes et du cycle de commande | order\_id, customer\_id |
| olist\_order\_items\_dataset.csv | Détail de chaque article dans une commande (produit, vendeur…) | Analyse des paniers, croisement produits-vendeurs | order\_id, product\_id, seller\_id |
| olist\_customers\_dataset.csv | Informations clients (ville, état, code postal…) | Segmentation, analyse géographique des clients | customer\_id |
| olist\_products\_dataset.csv | Catalogue produits (catégorie, dimensions, poids…) | Enrichissement des analyses produits, catégories | product\_id, product\_category\_name |
| product\_category\_name\_translation.csv | Table de traduction des catégories produits (PT/EN) | Standardisation des catégories pour l’analyse | product\_category\_name |
| olist\_order\_payments\_dataset.csv | Informations sur les paiements (type, montant, nombre d’échéances) | Analyse des moyens de paiement, revenus, anomalies | order\_id |
| olist\_order\_reviews\_dataset.csv | Avis clients sur les commandes (note, commentaire, date…) | Analyse de la satisfaction et du service client | order\_id |
| olist\_sellers\_dataset.csv | Informations vendeurs (localisation, identifiant…) | Analyse performance vendeur, géographie de l’offre | seller\_id |
| olist\_geolocation\_dataset.csv | Coordonnées géographiques (code postal, latitude, longitude…) | Cartographie, enrichissement géographique | zip\_code\_prefix |

### **5.2 Problèmes et limitations identifiés**

* **Valeurs manquantes :** Champs vides fréquents (dates, dimensions, commentaires, etc.) nécessitant un traitement spécifique.
* **Doublons :** Possibilité de lignes dupliquées pour certains identifiants (clients, vendeurs).
* **Clés absentes ou non correspondantes :** Certains IDs présents dans un fichier peuvent manquer dans d’autres (ex : product\_id).
* **Formats hétérogènes :** Variabilité dans le format des dates, langues (catégories en portugais), etc.

### **6. Contraintes et risques**

### **Contraintes techniques**

Dans la mise en place du Data Mart, plusieurs difficultés techniques peuvent apparaître :

* Harmonisation des formats de données venant de sources variées.
* Gestion des valeurs manquantes, des doublons et des incohérences dans les fichiers.
* Traitement de gros volumes de données qui peuvent allonger les temps d’exécution.

### **Contraintes organisationnelles**

Dans la pratique, ce type de projet rencontre aussi des obstacles organisationnels courants :

* Coordination nécessaire entre plusieurs équipes pour valider les besoins et les priorités.
* Évolution des attentes métiers pendant la réalisation du projet.
* Risques de mauvaise communication ou de compréhension des objectifs entre métiers et technique.
* Disponibilité limitée de certains interlocuteurs pour tester ou valider les livrables.

### **7. Synthèse des besoins utilisateurs**

### **7.1 Grille d’interview pour les producteurs de données**

|  |  |  |
| --- | --- | --- |
| **Question** | **Objectif** | **Utilité pour le projet** |
| Quels systèmes ou outils utilisez-vous pour saisir ou gérer les données ? | Comprendre le flux de création des données | Identifier les sources et points de contrôle |
| À quelle fréquence les données sont-elles mises à jour ? | Évaluer la fraîcheur des informations | Adapter la collecte et l’automatisation |
| Quels types d’erreurs ou de problèmes rencontrez-vous dans les données ? | Repérer les risques de qualité des données | Anticiper les traitements de nettoyage |
| Comment validez-vous la qualité ou la cohérence des données ? | Connaître les étapes de vérification | Ajuster le processus de contrôle qualité |

### **7.2 Grille d’interview pour les utilisateurs de données**

|  |  |  |
| --- | --- | --- |
| **Question** | **Objectif** | **Utilité pour le projet** |
| Pour quels usages ou décisions avez-vous besoin d’accéder aux données ? | Comprendre les besoins métiers principaux | Prioriser les indicateurs et les analyses |
| Quelles difficultés rencontrez-vous avec les rapports ou données actuels ? | Identifier les limites des solutions existantes | Corriger ou améliorer dans le Data Mart |
| Quels indicateurs ou analyses vous manquent aujourd’hui ? | Recueillir les attentes insatisfaites | Enrichir la solution de nouveaux KPIs |
| À quelle fréquence souhaitez-vous accéder aux données ou aux rapports ? | Saisir la périodicité et l’urgence des besoins | Adapter l’automatisation et la fréquence |

### **8. Conclusion**

Le lancement du Data Market arrive à un moment où l’entreprise doit mieux valoriser ses données pour accompagner son développement.

Cette première étape va simplifier l’accès à l’information, rendre les analyses plus fiables et poser les bases nécessaires pour de futurs projets innovants au service des équipes métiers.