Fusionner et nettoyer les fichiers d'Altitrek Diois

Altitrek Diois, magasin de sports de montagne de 9 salariés installé à Die (Drôme), vend en boutique et sur son site marchand, et loue du matériel (raquettes, skis de randonnée, vélos à assistance électrique). Ses données clients sont dispersées dans quatre fichiers qui ne communiquent pas.

Avant d'installer une plateforme de données clients (CDP), qui rassemblera toutes les sources en un profil unique, le gérant demande à Elouan Martel, alternant en BTS NDRC, de mesurer la qualité des données et de préparer leur nettoyage au 1er octobre 2026.

Complétez l'annexe 1, puis répondez aux questions d'analyse.

Document 1 : les quatre fichiers à rassembler (septembre 2026)

Fichier	Contenu	Fiches
Fichier fidélité du logiciel de caisse	Clients de la boutique (carte de fidélité)	4 860
Comptes clients du site marchand	Clients ayant commandé en ligne	3 240
Logiciel de location	Clients ayant loué du matériel	1 380
Abonnés à la lettre d'information	Inscrits à la lettre, clients ou non	1 120

Document 2 : résultats du test de la plateforme de données clients

- Fiches en double (copies en trop d'une personne déjà présente dans un autre fichier, à fusionner avec sa fiche) : 2 438.
- Profils complets après dédoublonnage (nom, e-mail valide, téléphone, code postal, date du dernier achat ou du dernier contact) : 6 285.

Document 3 : prospects (profils sans achat ni location, après dédoublonnage) selon la date du dernier contact émanant d'eux

Dernier contact (inscription, clic dans la lettre, message)	Profils
Avant le 1er octobre 2023	286
Du 1er octobre 2023 au 30 septembre 2025	471
Depuis le 1er octobre 2025	283
Total	**1 040**

Document 4 : volume des données stockées

- Données structurées (tickets de caisse, commandes, contrats de location, fiches clients) : 6,4 Go.
- Données non structurées (photos jointes aux avis et aux demandes de réparation, e-mails du service après-vente, messages reçus sur les réseaux sociaux) : 41,6 Go.

Document 5 : offre d'un prestataire spécialisé

Vérification des adresses e-mail et normalisation des adresses postales : 0,045 € HT par profil traité (TVA à 20 %).

Annexe 1 (arrondir les pourcentages à une décimale, les montants au centime)

Question	Réponse
Nombre total de fiches avant fusion	
Taux de doublons (en % des fiches avant fusion)	
Nombre de profils uniques après dédoublonnage	
Taux de complétude des profils (%)	
Nombre de profils de prospects à supprimer au 1er octobre 2026	
Nombre de profils conservés après cette suppression	
Part des données non structurées dans le volume stocké (%)	
Coût TTC de la vérification des profils conservés (€)	

Questions d'analyse

Doublons, adresses e-mail invalides, fiches incomplètes : lequel des « 5 V » du big data ces défauts mettent-ils en cause ?

☐ La vélocité
☐ Le volume
☐ La véracité

Justification :

Pour stocker ensemble les tickets de caisse, les commandes, les photos et les messages des clients, puis les analyser, quelle solution convient ?

☐ Un tableur partagé en ligne, qui rassemble les fichiers tenus par chaque vendeur
☐ Un lac de données (data lake), qui gère aussi les données non structurées
☐ Un entrepôt de données (data warehouse), conçu pour les données structurées

Justification :