

Rapport de Data Refinement

Ce rapport présente les différentes étapes de data refinement réalisées sur le dataset de transactions, depuis l'exploration initiale jusqu'à la préparation finale des données pour l'analyse et la visualisation.

1. Exploration des données

Une analyse exploratoire a été menée afin d'identifier la structure du dataset, les types de variables ainsi que la présence de valeurs manquantes.

2. Nettoyage des données

Les valeurs manquantes ont été traitées selon la nature des variables. Les variables numériques ont été imputées par la médiane, tandis que les variables catégorielles ont été complétées soit par la valeur la plus fréquente (pour la variable Item), soit par la catégorie 'Unknown' lorsque le taux de données manquantes était élevé.

3. Transformation des données

La variable Transaction Date a été convertie en format datetime afin de permettre l'extraction de composantes temporelles telles que l'année, le mois et le jour. Des variables supplémentaires ont été créées pour faciliter les analyses futures.

4. Organisation du projet

Le projet a été structuré en plusieurs notebooks correspondant aux différentes étapes du data refinement. Les datasets bruts et intermédiaires n'ont pas été poussés sur GitHub et ont été ajoutés au fichier .gitignore afin de respecter les bonnes pratiques liées à la gestion des données.

Conclusion

À l'issue de ces étapes, le dataset est propre, cohérent et prêt à être utilisé pour l'analyse exploratoire avancée et la visualisation, notamment à l'aide de Power BI.