

Rapport explicatif sur les démarches suivies et les difficultés rencontrées liées à la qualité des données :

Lors de l'analyse des résultats de certaines requêtes, des incohérences ont été identifiées, soulignant des problèmes de qualité de données. J'ai résolu certaines de ces anomalies dans le script 'beev-test' avant de les charger dans la base de données. J'ai corrigé les noms de colonnes dans la table 'consumer' et ajusté certaines valeurs en me basant sur la table 'car', qui spécifie les caractéristiques des véhicules pour notre étude.

Cependant, certaines parties des données dans les deux tables semblent ambiguës. Par exemple, la colonne 'Year' dans la table 'car' représente l'année de fabrication d'un modèle précis. Pour le modèle Jaguar XF, nous observons quatre versions : deux thermiques (2010 - prix : 60000, 2013 - prix : 23000) et deux électriques (2014 - prix : 72000, 2017 - prix : 22000). Le problème survient avec la colonne 'Year' dans la table 'consumer', qui représente les années de vente de chaque modèle dans différents pays de 2010 à 2022. Il devient difficile de déterminer quelle version de la Jaguar XF est commercialisée en 2018, par exemple.

J'ai tenté d'appliquer un modèle commercial où chaque version est commercialisée jusqu'à la sortie de la nouvelle version. Cependant, ce modèle rencontre des difficultés, notamment dans le cas de la Peugeot 508 en 2019, où deux versions (thermique à 54000 et électrique à 58000) coexistent. Cette approche s'est avérée problématique pour déterminer quelle version sera commercialisée en 2019 dans différents pays.

J'ai également envisagé d'enrichir les données de la table 'car' pour qu'elle couvre les différentes versions des modèles, mais les mêmes problèmes persistent. Ainsi, pour obtenir des résultats fiables, **une intervention d'experts métiers** est nécessaire afin de résoudre ces problèmes spécifiques et d'assurer la qualité des données qui couvre le bon **business model**.

Beev