

Question 1 - Enjeux 2

Enjeu #2 - Granularité au niveau du magasin

L'enjeu #2 concerne la **granularité des données au niveau du magasin** et l'impact de la présence de nombreuses valeurs nulles sur la qualité des analyses et des prévisions. Une forte proportion de zéros est observée dans la série de ventes au niveau le moins granulaire. Lorsque l'on descend à des niveaux encore plus fins (par exemple, au niveau d'un magasin), la proportion de zéros augmente encore davantage, comme l'illustre le tableau ci-dessous :

Région	Pourcentage de zéros
Centre	77,1%
Est	73,8%
Sud	83,0%
Ouest	68,3%

Nous pourrions être tentés de **rééchantillonner** les données de la série globale à une fréquence hebdomadaire afin d'obtenir une série plus homogène et sans intermittence. Toutefois, il ne serait **probablement pas pertinent de réaliser une prévision uniquement au niveau global**, sans distinction par magasin ou région. Une telle approche risquerait d'être **peu exploitable** pour les gestionnaires, qui cherchent à comprendre comment optimiser les performances individuelles de leurs magasins.

De plus, lors du **rééchantillonnage à une fréquence hebdomadaire ou mensuelle**, la perte d'identification des **magasins individuels** limite la capacité à modéliser les ventes de manière conditionnelle à un magasin spécifique, ou même à un segment ou une région. Or, les segments et les régions peuvent présenter des **dynamiques saisonnières distinctes**, ce qui impacte la qualité des prévisions et réduit leur pertinence pour la prise de décision.

Insight pour l'apprentissage

Pour atténuer cet enjeu, nous pourrions envisager une **segmentation alternative**, où la segmentation serait effectuée **par région, qui jouerait alors le rôle de regroupement des magasins**. Ensuite, les données pourraient être **rééchantillonnées à une fréquence hebdomadaire**, permettant ainsi d'avoir des séries plus exploitables tout en conservant une granularité pertinente.