**Plan prévisionnel**

## **Dataset retenu**

Le dataset utilisé est intitulé [Walmart Sales Forecast Datasets](https://www.kaggle.com/datasets/iamprateek/wallmart-sales-forecast-datasets). Ce dataset contient des données hebdomadaires de ventes pour 45 magasins Walmart, regroupant plusieurs départements sur une période allant de 2010 à 2012.

Les principales caractéristiques du dataset incluent :

* **Store** : Identifiant unique de chaque magasin.
* **Dept** : Identifiant unique pour chaque département dans un magasin.
* **Date** : Date correspondant à chaque semaine de collecte des données.
* **Weekly\_Sales** : Montant total des ventes hebdomadaires pour un département donné.
* **IsHoliday** : Indicateur binaire précisant si la semaine inclut une période promotionnelle ou un jour férié.

Ce dataset est cependant déséquilibré, certaines semaines ou certains départements ayant des volumes de ventes significativement plus élevés que d'autres. Cela en fait un cas d'étude réaliste pour des problématiques de prévision des ventes dans un environnement réel.

***Feature engineering :*** ***https://www.kaggle.com/code/aslanahmedov/walmart-sales-forecasting***

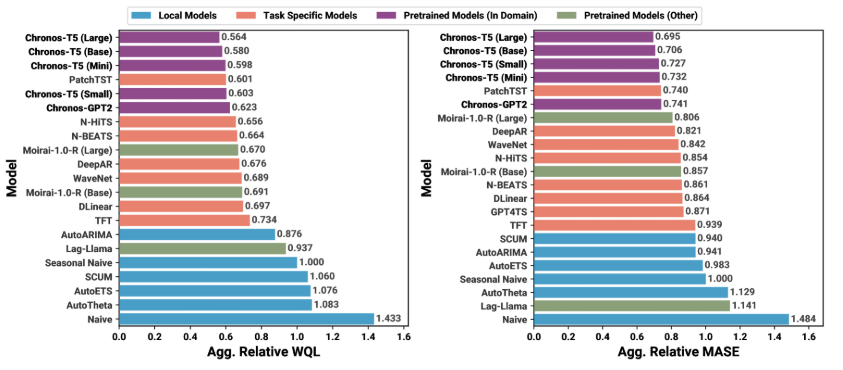
## **Modèle envisagé**

#### *1. Objectif de l’algorithme et contexte d’utilisation :*

L'algorithme **Chronos** vise à fournir une méthode générale et efficace pour la prévision probabiliste des séries temporelles. Il utilise des architectures de modèles de langage (basées sur des transformeurs) pour traiter les séries temporelles en les transformant en séquences de "tokens" discrets. L'objectif est d'améliorer les performances en exploitant des modèles préentraînés sur des corpus vastes et diversifiés de données publiques, complétés par des données synthétiques. Le cadre de **Chronos** peut être utilisé dans divers contextes, notamment dans les domaines de la vente au détail pour soutenir la prise de décision.

#### 2. Arguments justifiant le choix de l’algorithme et sa performance :

* **Adaptation minimale mais efficace** : Chronos se distingue par une approche minimaliste qui utilise des modèles de langage existants sans modifications majeures de leur architecture, facilitant ainsi leur application directe aux séries temporelles.
* **Résultats démontrés sur 42 ensembles de données** : Dans des benchmarks exhaustifs, Chronos surpasse significativement les méthodes traditionnelles et certaines approches d'apprentissage profond sur les ensembles de données inclus dans le corpus d'entraînement. Il démontre également des performances comparables, voire supérieures, en apprentissage "zero-shot" sur des ensembles de données non vues auparavant.
* **Flexibilité et robustesse** : L'utilisation de techniques d'augmentation de données, telles que **TSMixup** (mélange de séries temporelles réelles) et des séries temporelles synthétiques générées par processus gaussiens, améliore la robustesse et la généralisation du modèle.
* **Efficacité computationnelle** : En comparaison avec d'autres modèles nécessitant des ajustements spécifiques ou des ressources intensives (comme GPT-3), Chronos offre une alternative moins coûteuse avec une taille de modèle modeste.

Figure 1: Benchmark 5.5.1 : https://arxiv.org/html/2403.07815v3

## **Références bibliographiques**

* Chronos paper : https://arxiv.org/abs/2403.07815
* Comment installer Chronos : [lien youtube](https://www.youtube.com/watch?v=WxazoCVkBhg)
* Article de blog : [lien](https://arize.com/blog/demystifying-chronos-learning-the-language-of-time-series/)
  + Concept fondamentaux des séries temporelles
  + Présentation de Chronos
  + Application et cas d’utilisation
  + Avantages et limites

## **Explication de votre démarche de test du nouvel algorithme (votre preuve de concept)**

1. Utilisation du [kernel](https://www.kaggle.com/code/aslanahmedov/walmart-sales-forecasting)de Monsieur ASLAN AHMEDOV comme point de départ.
2. Évaluation du modèle AUTOARIMA (calcul MAPE).
3. Utilisation de autogluon.
4. Évaluation du modèle Amazon Chronos
5. Optimisation du modèle Amazon Chronos
6. Interface pour le métier qui sert à évaluer les performances des différents magasins.