Par Jean-baptiste
PERICAUD
Soutenance: 07.05.22

Projet 8 : Ventes en magasin - Prévisions de séries temporelles

CORPORACIÓN FAVORITA

Etapes de preprocessing

- Etapes de fusion des différents datasets
- Traitement de la variable Jour férié
- Méthode d'interpolation des valeurs manquantes du taux de pétrole

Variables	Types
Id	int64
Date	object
Type de produit (family)	Category
Etat	Category
Ville	Category
Prix du pétrole	Float64
Promotion	Float64
Ventes	Float64
Jours férié (ou évènement)	Category
Type de jours férié (National, local ou régional)	Category
Localisation du jours férié (Province)	Category
Description du jour férié	Objet
Jours férié transféré à une autre date	Booléen
Années	Période (D)
Nombre de magasin	int64
Type de magasin	Category
Cluster (regroupe les magasins similaires)	Category

Etapes de preprocessing

Variables	Types
Id	int64
Date	object
Type de produit (family)	Category
Etat	Category
Ville	Category
Prix du pétrole	Float64
Promotion	Float64
Ventes	Float64
Jours férié (ou évènement)	Category
Type de jours férié (National, local ou régional)	Category
Localisation du jours férié (Province)	Category
Description du jour férié	Objet
Jours férié transféré à une autre date	Booléen
Années	Période (D)
Nombre de magasin	int64
Type de magasin	Category
Cluster (regroupe les magasins similaires)	Category

Training Set : 2013 au 15 aout 2017

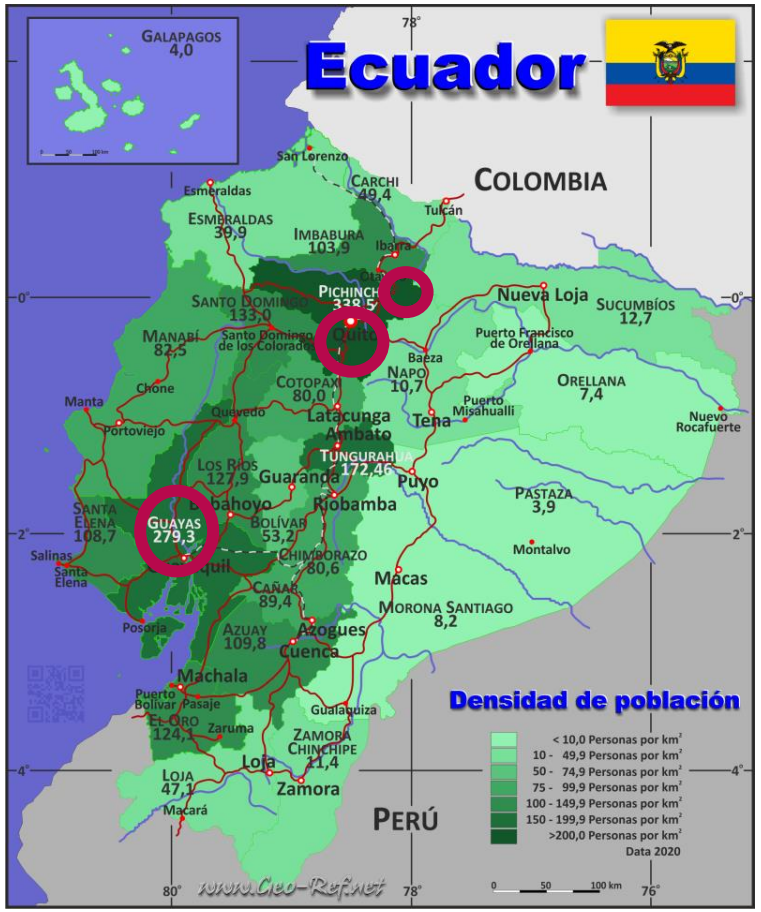
Test set : 16 aout au 31 aout 2017

Etapes de preprocessing

Variables	Types
Id	int64
Date	object
Type de produit (family)	Category
Etat	Category
Ville	Category
Prix du pétrole	Float64
Promotion	Float64
Ventes	Float64
Jours férié (ou évènement)	Category
Type de jours férié (National, local ou régional)	Category
Localisation du jours férié (Province)	Category
Description du jour férié	Objet
Jours férié transféré à une autre date	Booléen
Années	Période (D)
Nombre de magasin	int64
Type de magasin	Category
Cluster (regroupe les magasins similaires)	Category

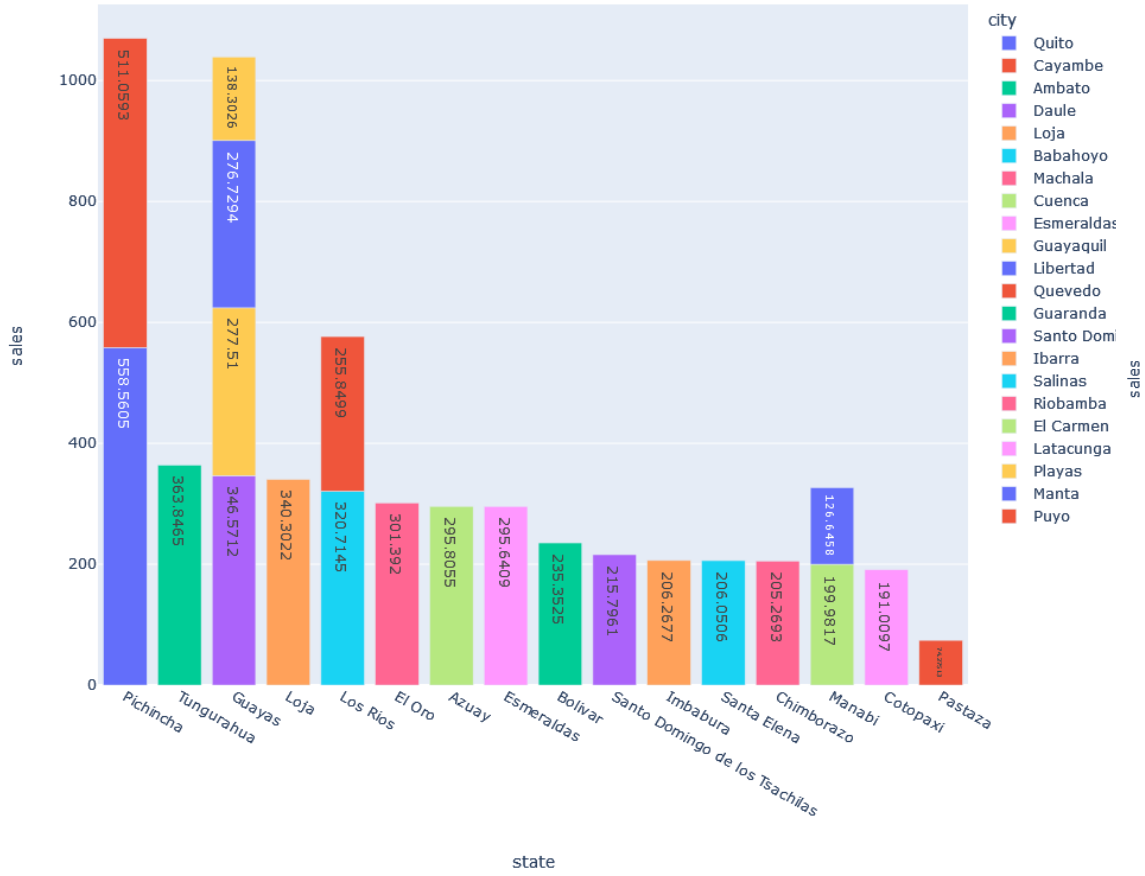
Analyse de données

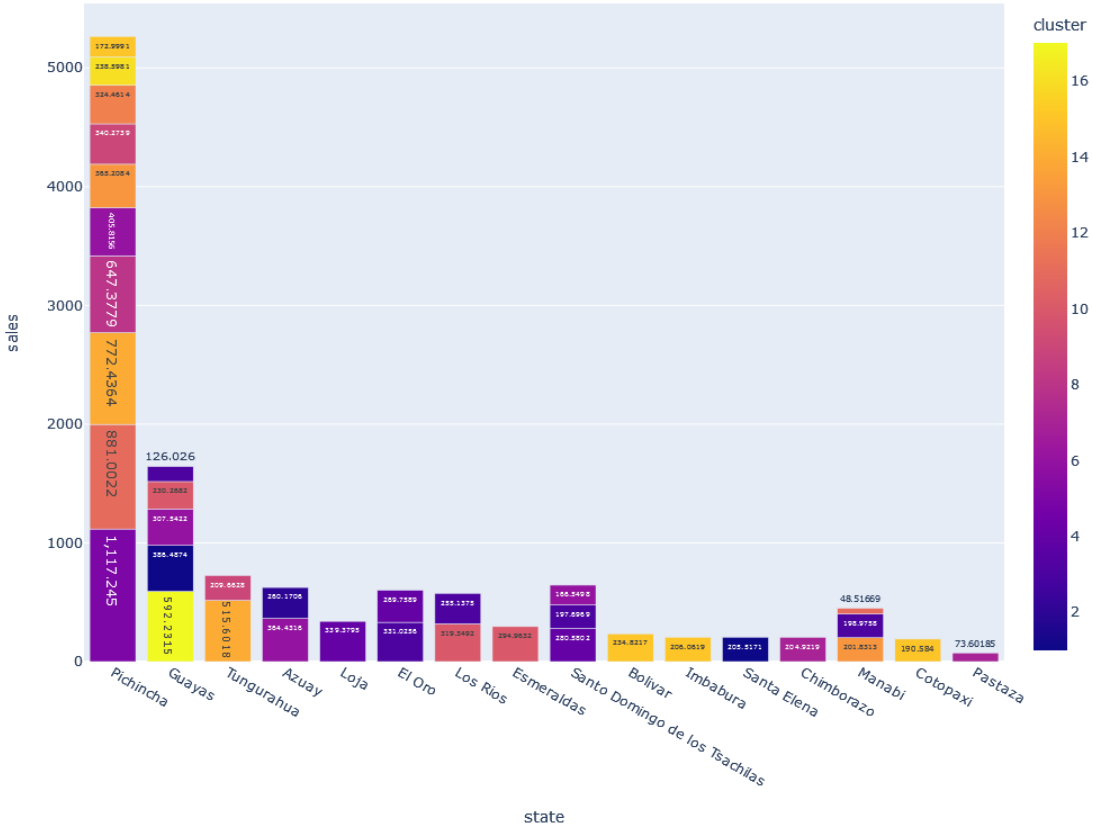
Analyse exploratoire des variables spatiales



Analyse de données

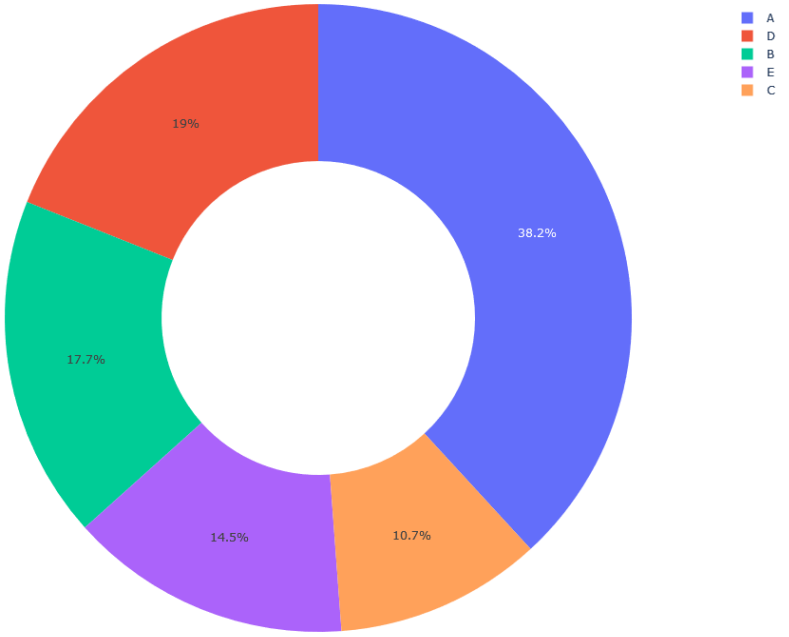
Analyse exploratoire des variables spatiales

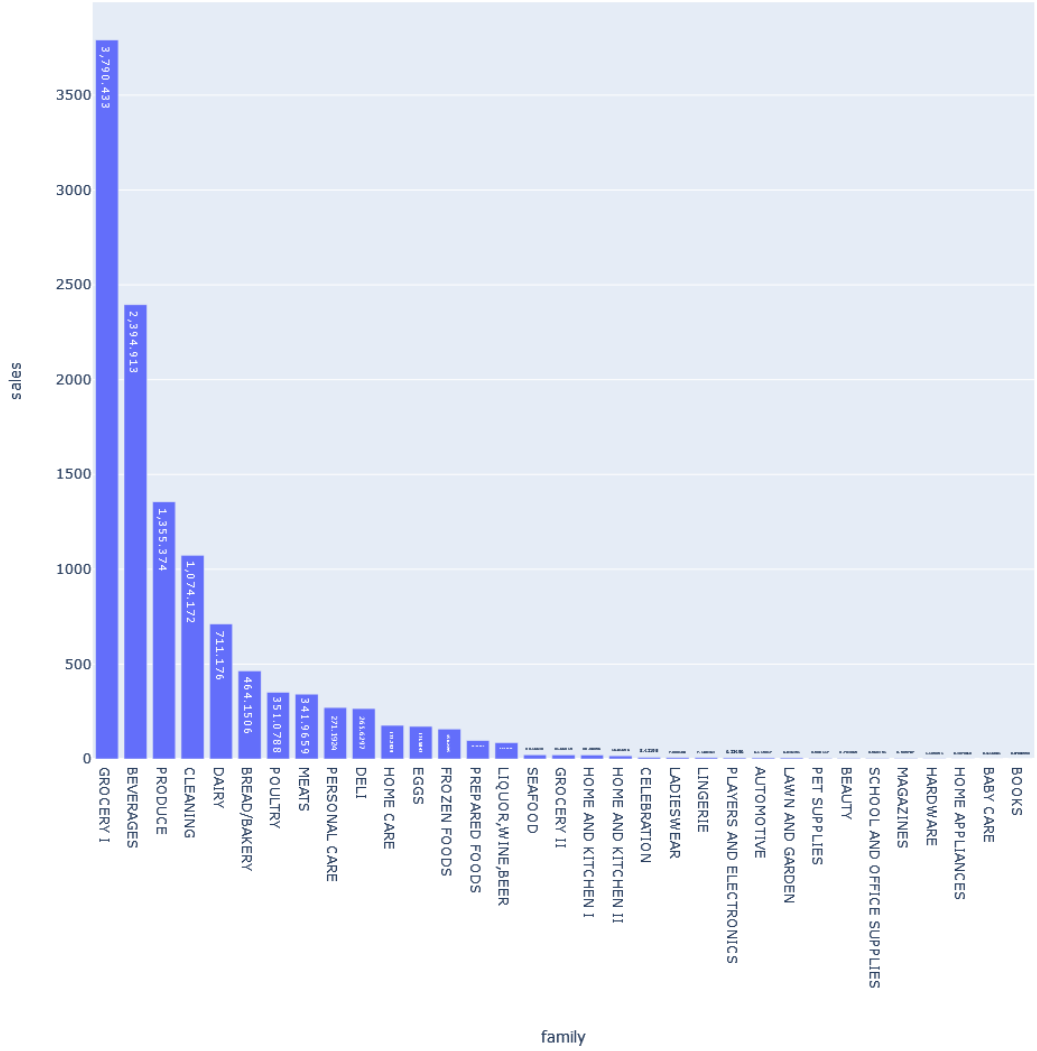




Analyse de données

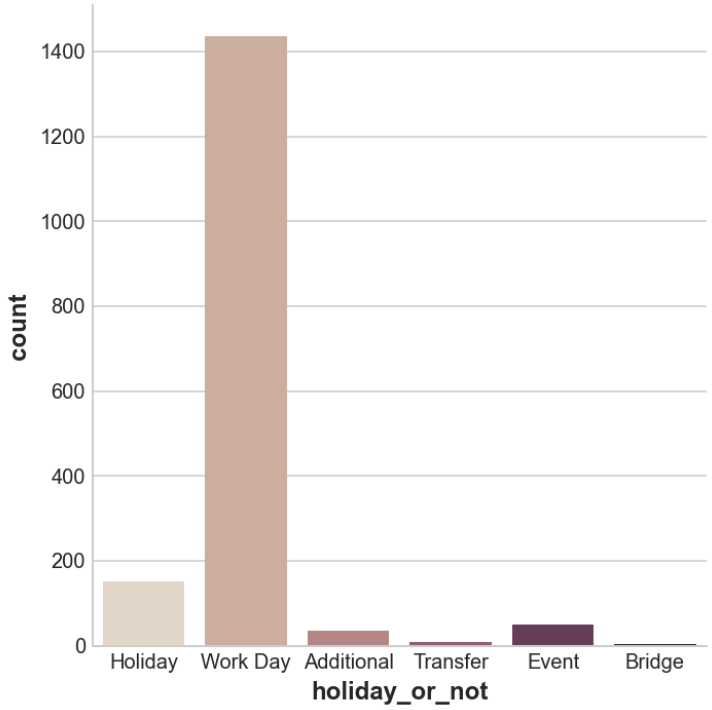
Analyse sur les types de magasin et de produits vendus

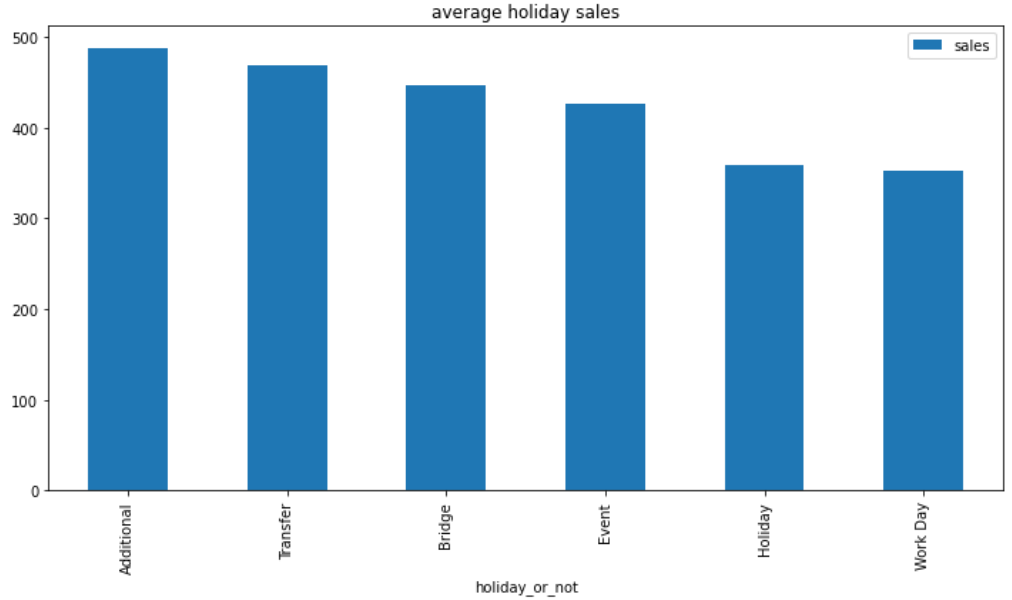


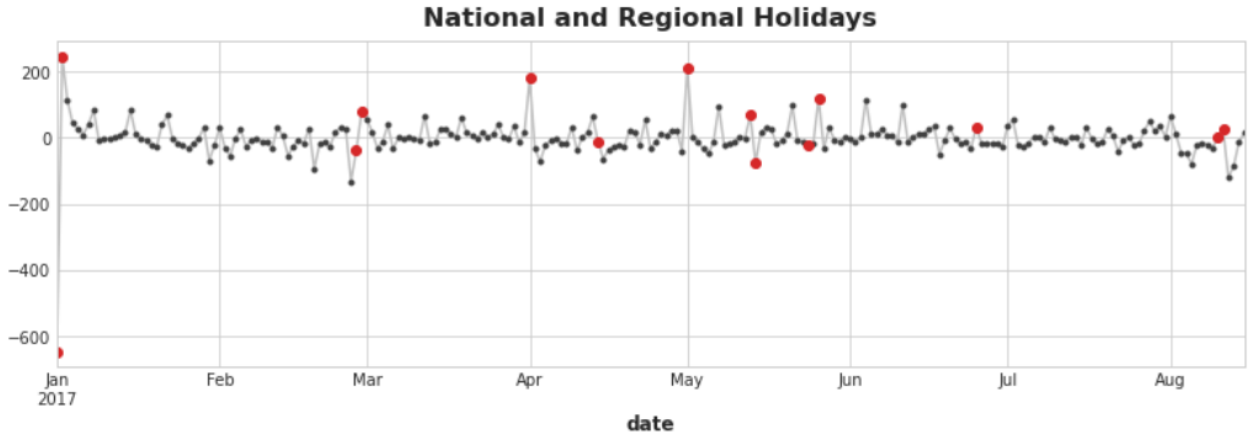


Analyse de données

Analyse des évènements sur la vente de produits

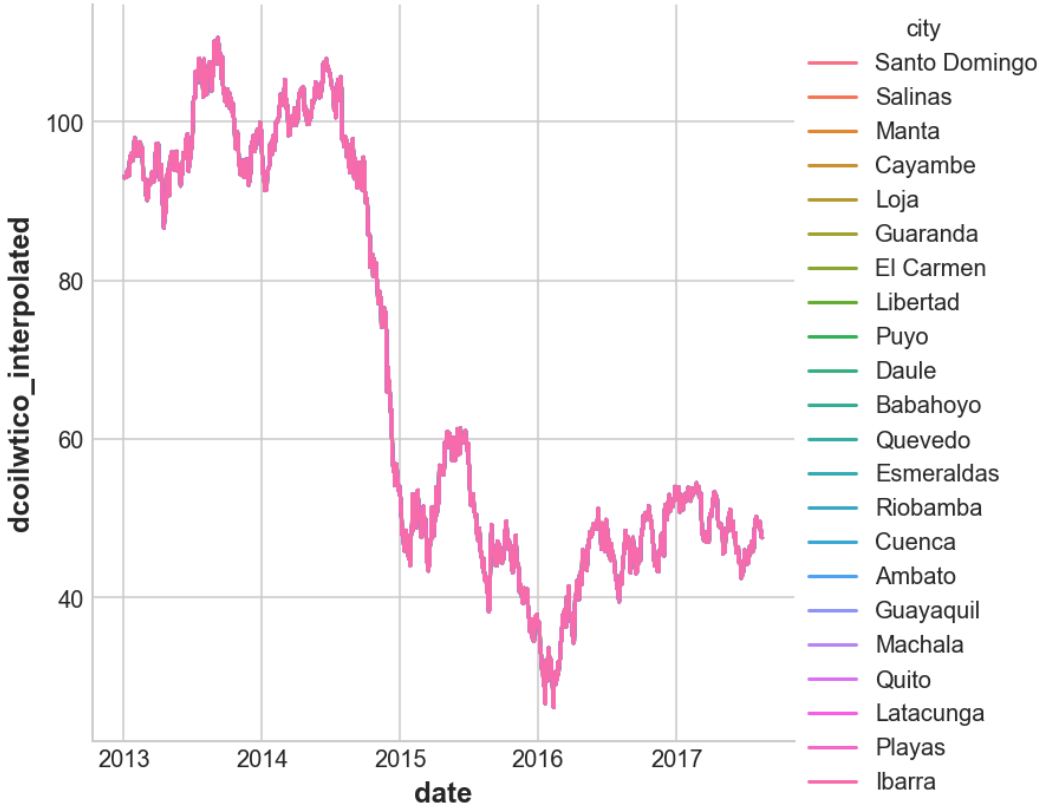


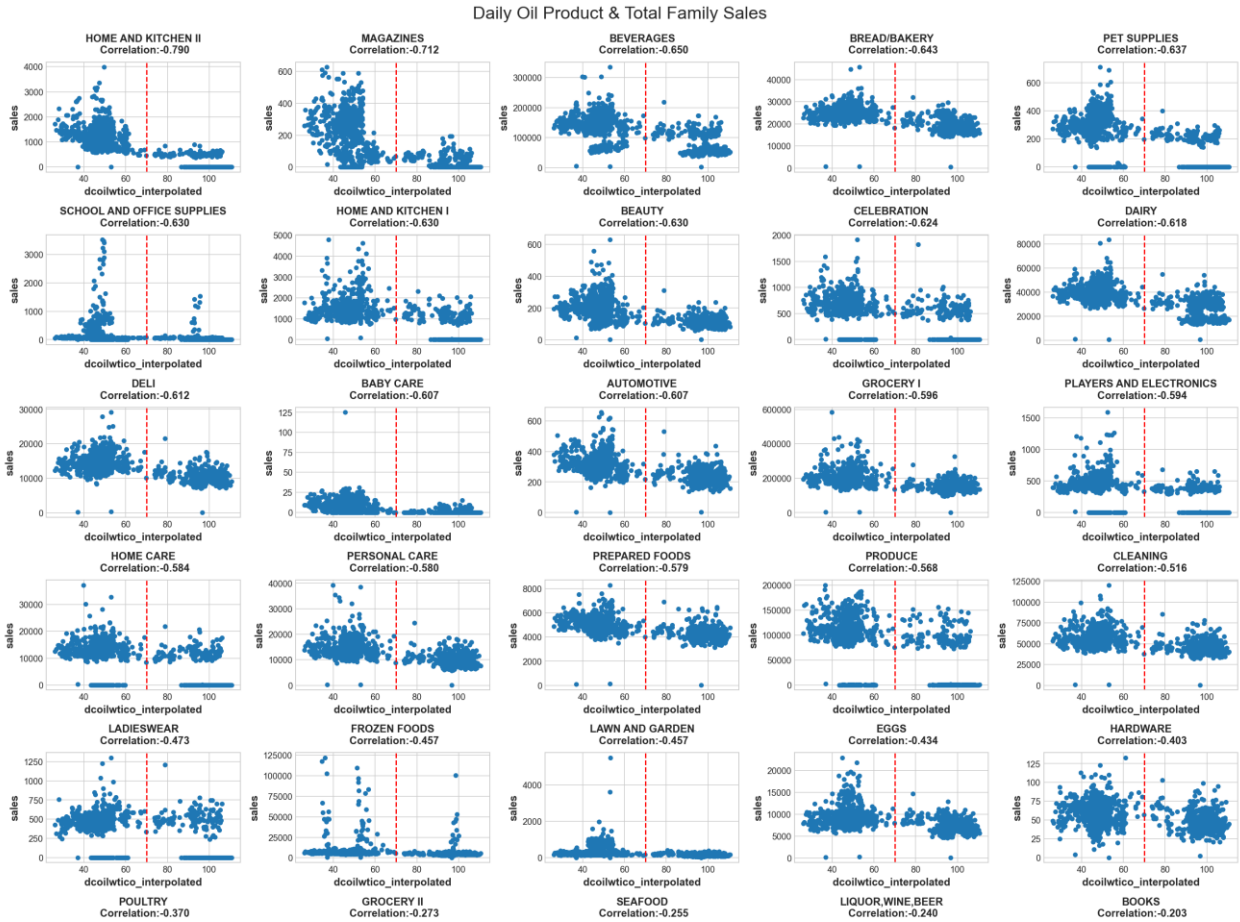




Analyse de données

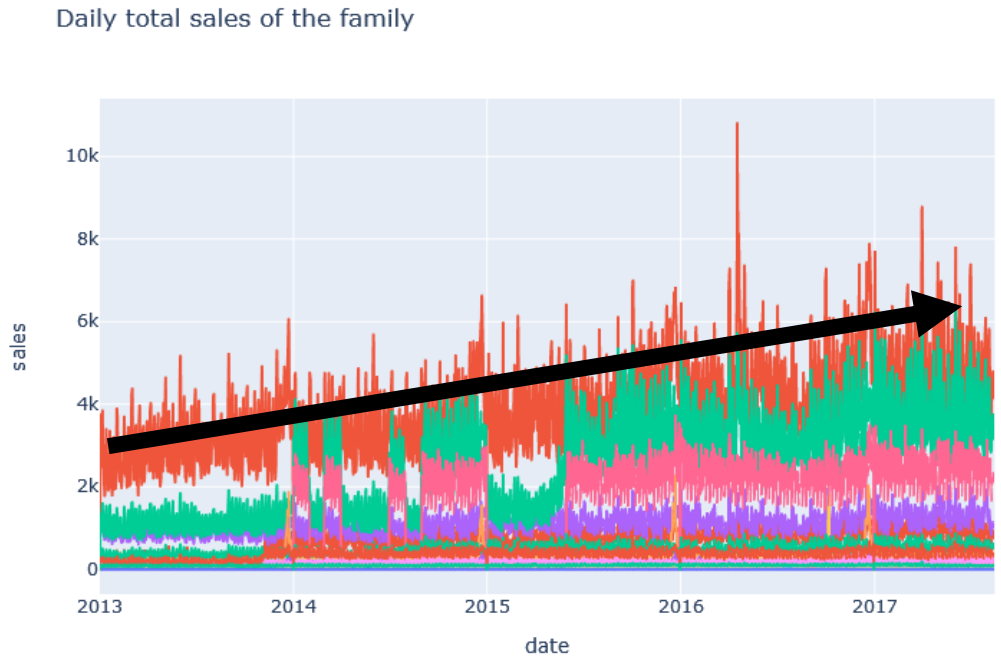
Analyse du taux de pétrole sur la vente des différents produits

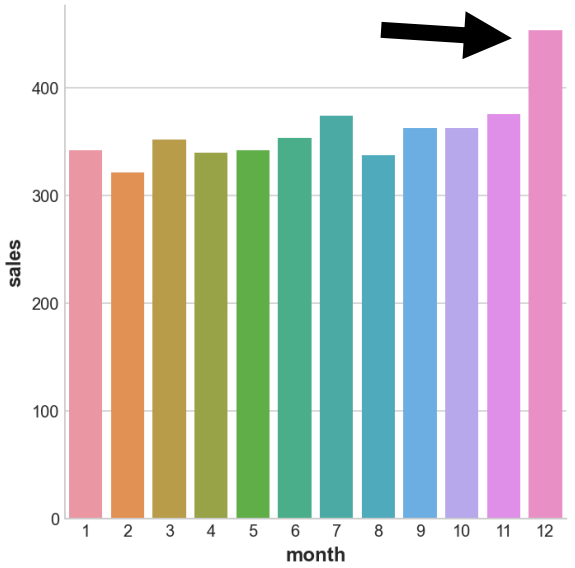


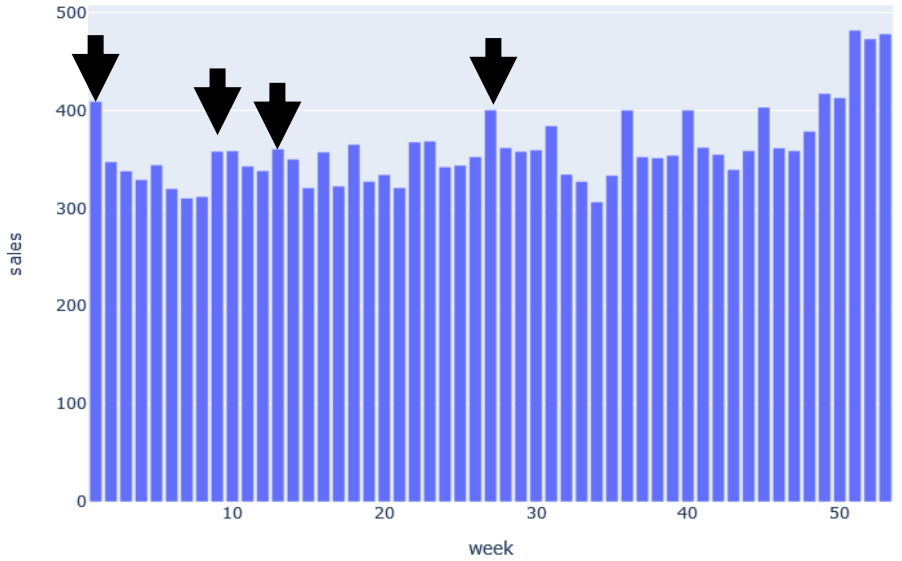


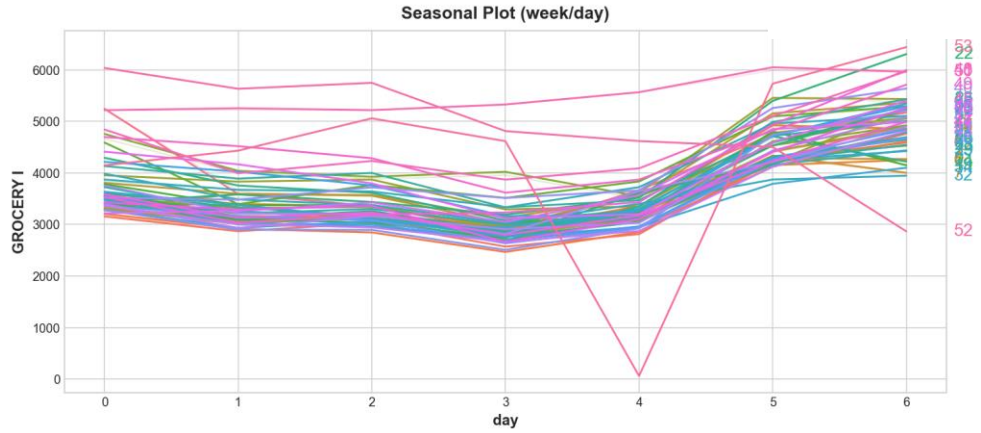
Analyse de données

Analyses temporelles de ventes des différents produits







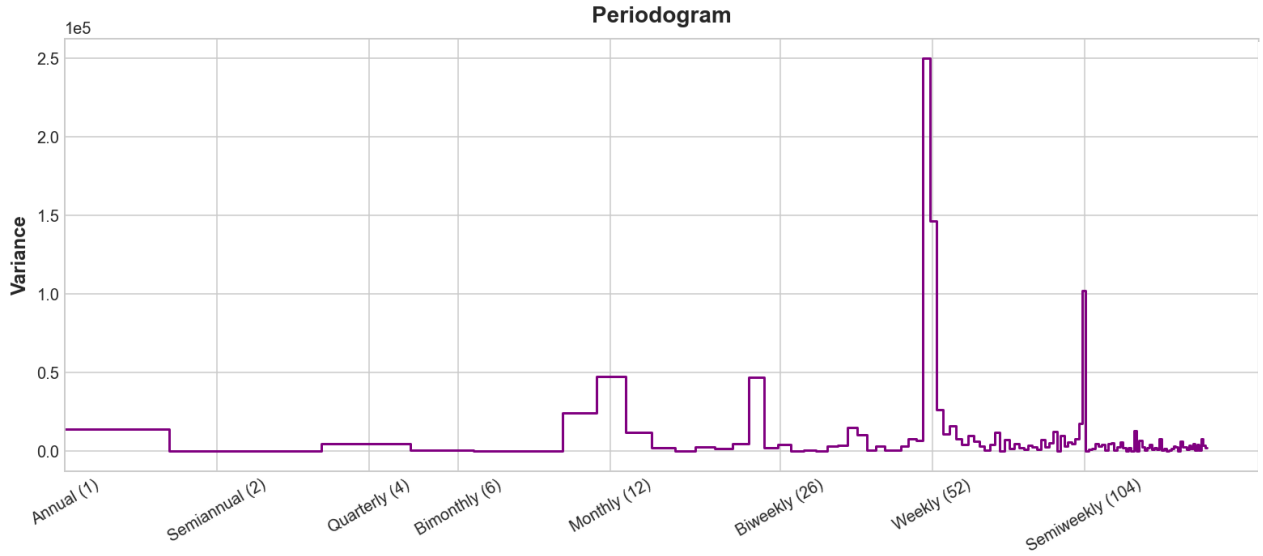


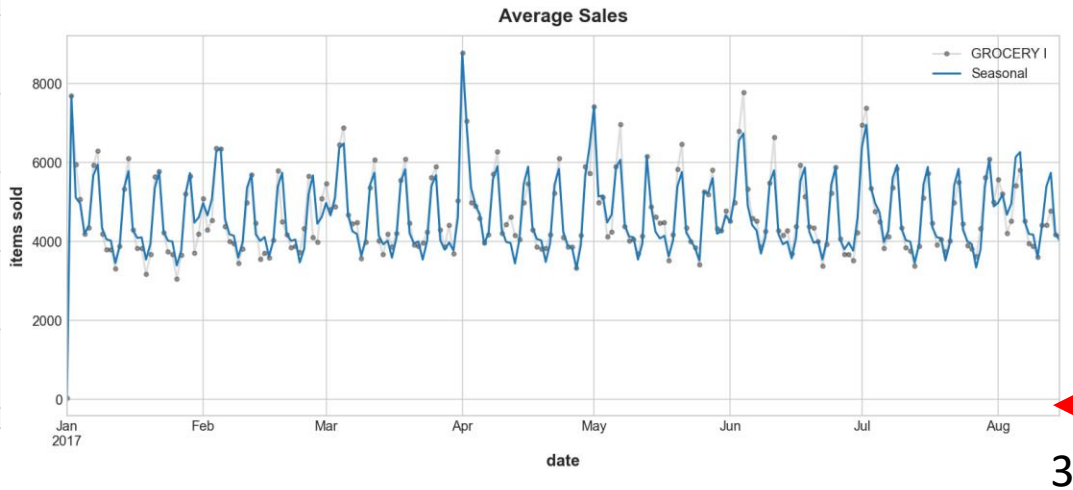
Partie modélisation :Quel métrique utilisée ?

$$\textbf{RMSLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(x_i+1) - \log(y_i+1))^2}$$

N est le nombre total d'observations dans l'ensemble des données.
Xi est la prédiction de la valeur cible
Yi est la valeur cible pour *i*
.*log*(*x*) est le logarithme naturel de *x* (*loge*(*x*).

Partie modélisation : prétraitement du modèle

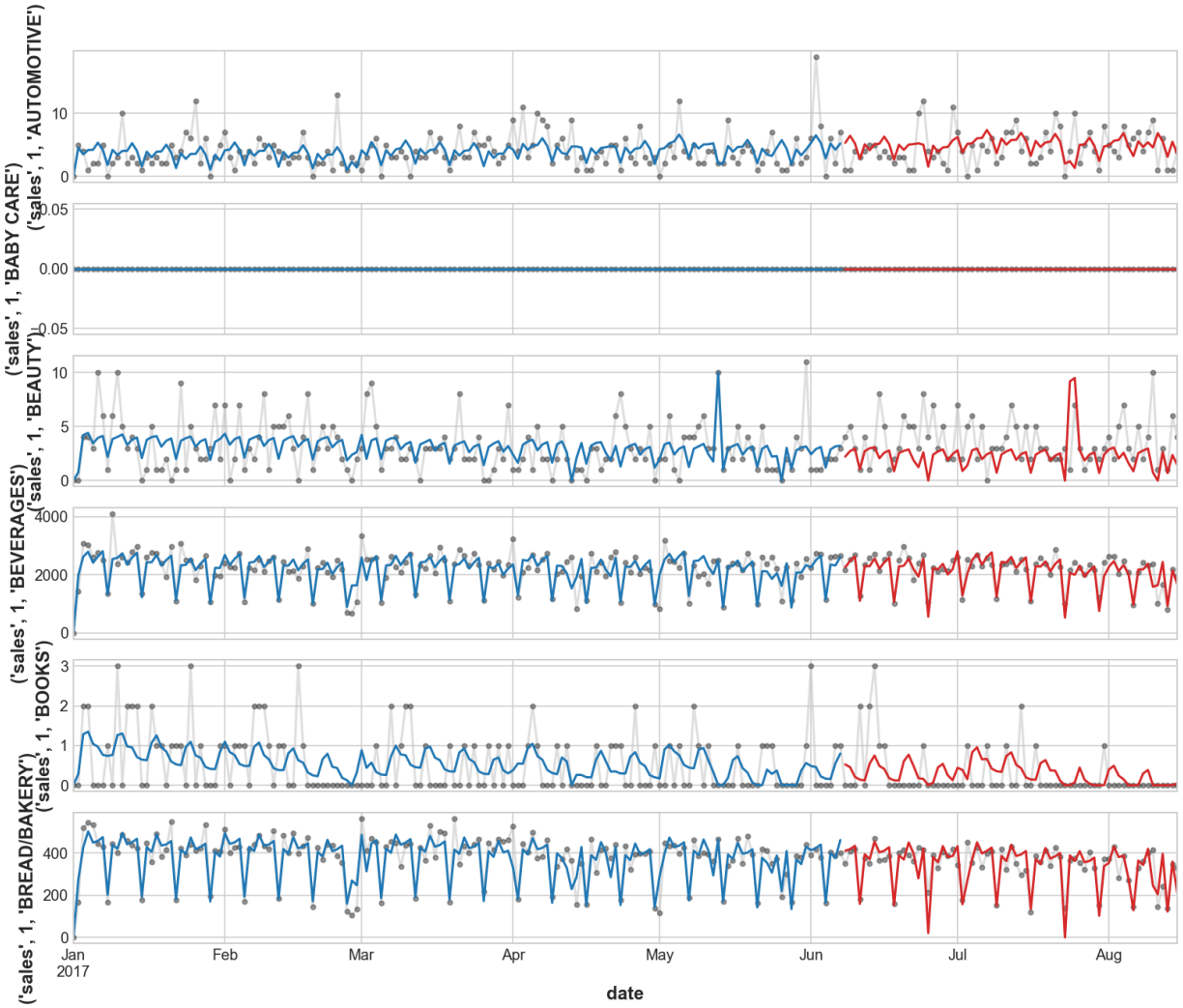




31 aout

Partie modélisation : Les modèles

Régressions Linéaire Simple :Training RMSLE: **0.53**
Validation RMSLE: **0.64**



Régressions Ridge (alpha 10): Training RMSLE: **0.62**
Validation RMSLE: **0.58**

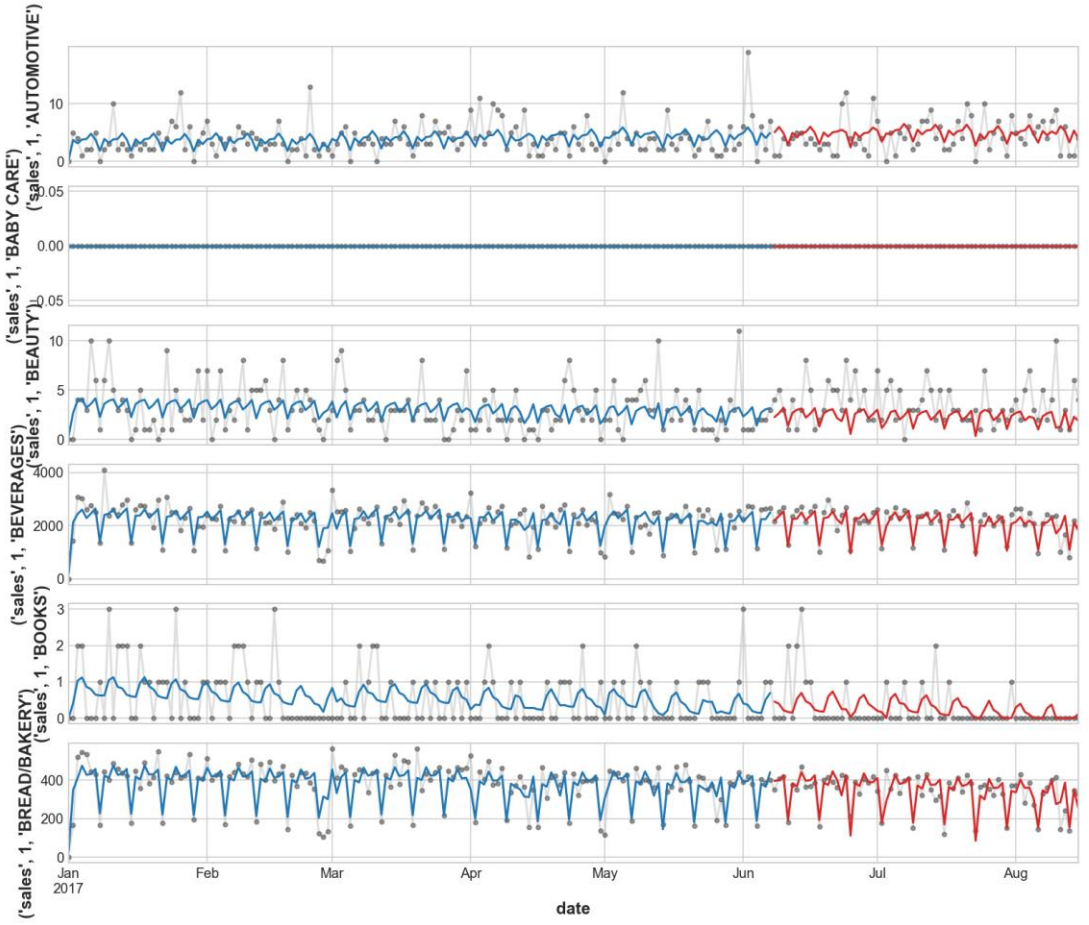
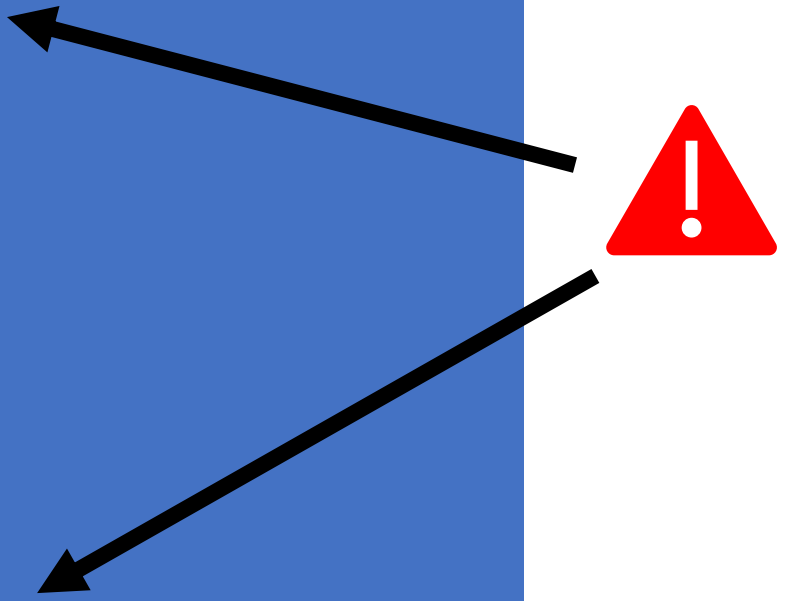


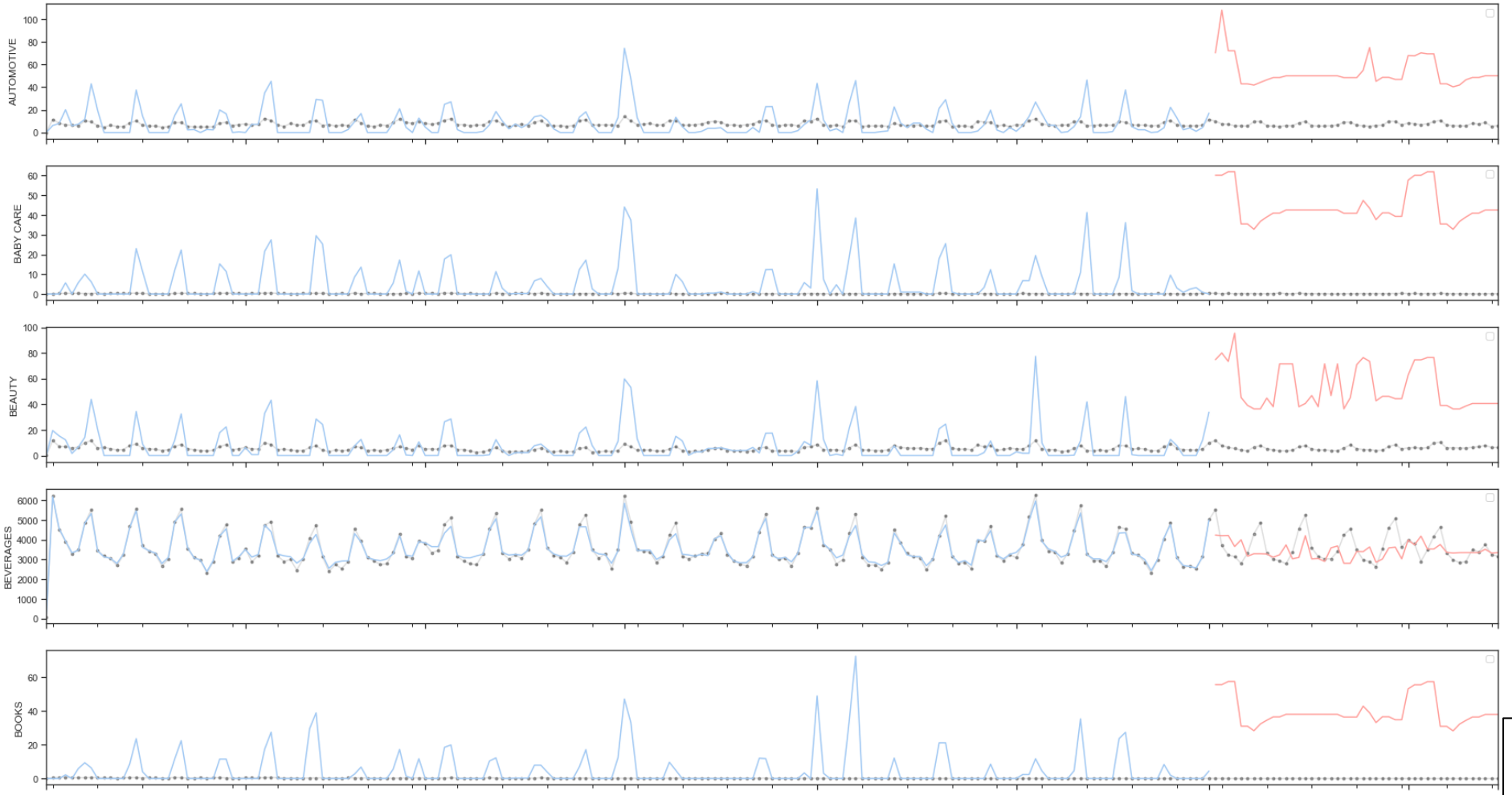
Tableau représentant les valeurs prédites par les données de validations dans le magasin 0

Type de produit	RMSLE
AUTOMOTIVE	0.347148
BABY CARE	0.000000
BEAUTY	0.289788
FROZEN FOODS	0.073438
GROCERY I	0.026008
GROCERY II	0.215336
HARDWARE	0.393565
HOME AND KITCHEN I	0.527739
HOME AND KITCHEN II	0.204982
HOME APPLIANCES	0.193820
LINGERIE	0.972237
SCHOOL AND OFFICE SUPPLIES	0.133763
SEAFOOD	0.137892



Mauvaise bonne Idée : Modèle Hybride

```
model = BoostedHybrid(
    model_1=LinearRegression(),
    model_2=XGBRegressor())
```

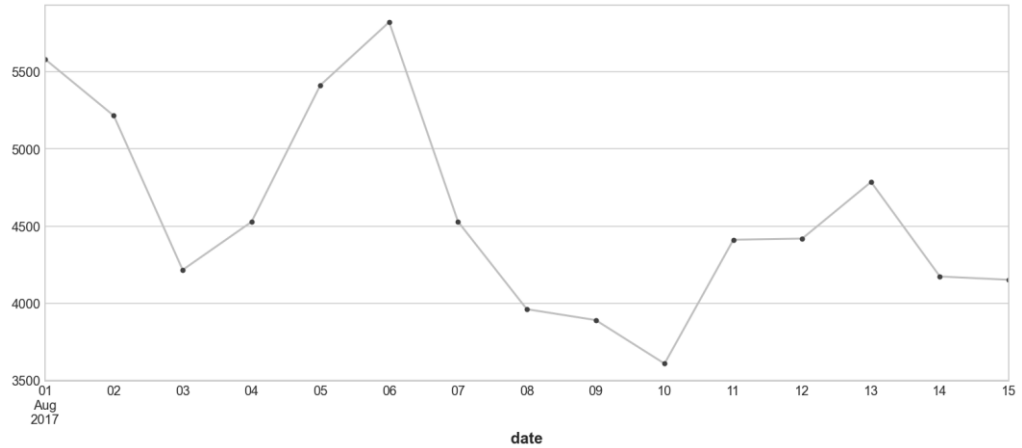


Notebook de Maryna Antonevychv 03/05/2022

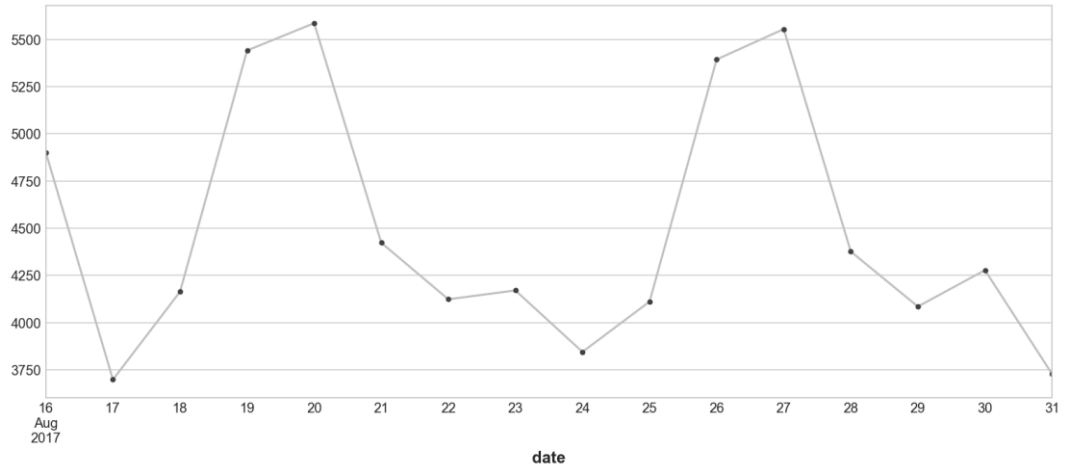
Partie Prédiction : valeurs de ventes sur 15 jours

Moyenne des ventes (Grocery 1) de tout les magasins

Partie Entrainement



Partie Prédiction



Conclusion

- Sur de la prédiction temporelle : Les meilleurs modèles sont souvent les plus simples.
- Pour améliorer la précision du modèle:
  - Retravailler les variables d'évenements
  - Rajouter des Séries décalées (Lag series and Lag plots
  - Réajuster les hyper parametres XGBOOST
  - Travailler sur des réseaux neuronneaux (LSTM)