



Mini projet

Consommation de carburants en France
Programmation pour la science des données 2024-2025



Objectifs

Livrable : un notebook Python implémentant les fonctionnalités requises comme spécifié ci-après

Présentation : Charger et traiter les données de prix des carburants :

- ☐ Collecter des données
- ☐ Préparer les données
- ☐ Visualiser les prix de l'essence
- ☐ (Option) Représenter une carte des indices de prix moyens des carburants par département
- ☐ Modéliser l'évolution des prix de l'essence

Collecte de données

Attendu #1

Télécharger les données des prix des carburants de 2023 (1 an) depuis Github (<https://github.com/rvm-courses/GasPrices>).

Il s'agit de fichiers préparés à partir du site Web d'origine et ne présentant pas de difficultés majeures pour être lu.

Quelques opérations de préparation sont néanmoins nécessaires : les documenter dans le notebook.

Fichier pour 2023 : `Prix2023.csv.gz`

Préparation des données – étape 1

Attendu #2

En Python :

- Lire le fichier des prix des carburants
- Extraire de la colonne date l'année, le mois, et la semaine de l'année
- Préparer la latitude et la longitude pour la cartographie (à diviser par la bonne puissance de 10)

Attendu #3

À l'aide de statistiques de base, déterminer quels types de carburants présentent un intérêt pour le reste du projet. Deux d'entre eux présentent peu d'intérêt et peuvent être filtrés pour le reste du projet.

Présenter l'analyse et indiquer ses conclusions.

Préparation des données – étape 2

Attendu #4

Calculer l'indice de prix pour chaque station par semaine :

- Calculer une nouvelle variable appelée « Indice de prix » pour chaque type de carburant vendu dans une station, ainsi :

$$\text{Price Index} = 100 \times \left(\frac{\text{Day Price in station} - \text{Average Day Price in France}}{\text{Average Day Price in France}} + 1 \right)$$

Attendu #5

Calculer l'index de la semaine :

- Calculez une nouvelle variable appelée « Index de la semaine » pour chaque enregistrement indiquant le numéro de la semaine dans l'année pour l'enregistrement. Cette variable sert aux agrégations hebdomadaires.

Attendu #6

En utilisant matplotlib/seaborn ou plotly :

- Représenter l'évolution hebdomadaire des indices de prix moyen de l'essence sur la France ainsi :
 - Chaque type de carburant est une ligne
 - Les coordonnées X sont l'index de la semaine (attendu #5)
 - Les coordonnées Y représentent l'indice de prix moyen du type de carburant en France sur la semaine (à calculer à partir de l'attendu #4).

Astuce : considérer l'utilisation de Seaborn FacetGrid pour un graphique à plusieurs lignes.

Visualisation des données – Question bonus

Option #1

Cette question donne droit à des points bonus (+10 %, max à 100 %)

- Représenter une carte pour chaque type de carburant des cartes géographiques de chaleur des indices de prix de la France, par exemple au niveau départemental

Références : le notebook suivant contient un exemple d'utilisation de Folium pour dessiner une carte au niveau du département.

https://github.com/HerveMignot/geopython/blob/main/notebooks/Cartographie_avec_geopandas_%26_folium.ipynb

Modélisation – Prédiction du prix du jour suivant

Attendu #7

- Agréger le prix moyen par carburant par jour de l'année. Travailler au niveau national de façon à n'avoir qu'une série par type de carburant
- Entraîner un modèle de prédiction du prix du lendemain pour chaque carburant le lendemain : organiser le code pour utiliser le même code sur les différentes séries
- Fournir les mesures de précision pertinentes et un graphique de dispersion pertinent entre les données réelles et prévues
- **Astuces** : utiliser des fonctionnalités de décalage (comme le prix de la veille, et le jour d'avant, etc.) pour construire le modèle.
Voir la référence : <https://arxiv.org/abs/2101.02118>
- Regarder des tutoriels comme :
<https://tichmangono.github.io/tutorials/2018/05/04/time-series-data-munging-lagging-variables-across-multiple-groups>
- **Remarque** : il n'est pas attendu que le modèle soit très optimisé, c'est la démarche que sera évaluée

Livraison du projet

Seul un notebook documenté doit être rendu.

Le code sera revu et évalué avec les commentaires dans le notebook.

Assurez-vous de documenter votre code, c'est important pour l'évaluation !

Aucun livrable supplémentaire n'est prévu.

Des questions ?

herve.mignot@equancy.com

