

Projet-Partie 2 (15 points)

Cette phase du projet compte pour 15% de votre note finale.

Date de remise : mardi 5 décembre 2023 avant 22h

Directives :

- Même équipe que pour la partie 1
- Pour réaliser les questions suivantes, vous devez utiliser les mêmes fichiers que vous avez utilisé pour la partie 1.
- Expliquez clairement ce que vous avez tenté de réaliser à chaque étape/sous-étape à l'aide de commentaires dans le code. Pour inscrire un commentaire, vous pouvez simplement écrire une phrase avec un # devant ou englober un paragraphe de trois guillemets anglais, au début et à la fin du paragraphe ("").
- Vous devez remettre soit :
 - un code (.py) commenté en expliquant vos étapes aux différentes questions clairement en commentaire dans le code pour chaque question.
 - un code (.ipynb) incluant vos explications.
- Veuillez déposer votre projet sur Moodle, au plus tard le mardi 5 décembre 2023 avant 22h.
- Inscrire votre numéro d'équipe dans le nom des documents remis (ex : Équipe 1 – Projet P2.py) et tous les prénoms, noms et matricules des participants de l'équipe à l'intérieur de chaque document.

Les bases du langage Python sont abordées en classe, mais n'hésitez pas à explorer par vous-même, par exemple sur le web, pour vous aider à résoudre les questions.

Dans votre script Python, identifiez les différentes sections par l'utilisation de zones de commentaires. Seul Python doit être utilisé pour réaliser le projet.

Pour les questions 1 à 3, répondez à l'aide de tests d'hypothèses :

1. (2 points) Existe-t-il une association significative entre la gravité de l'accident et la météo ? Considérez seulement les catégories de gravité suivantes:
 - Dommages matériels inférieurs au seuil de rapportage
 - Dommages matériels seulement
 - Léger
2. (2 points) Existe-t-il une corrélation entre la valeur des dommages pour les accidents dont la gravité est légère et l'âge du conducteur?
3. (2 points) Existe-t-il une différence significative entre la moyenne de la valeur des dommages en fonction du sexe du conducteur?

4. (1 point) Créez une Série nommée *time_serie* qui présente le nombre d'accident par jour
5. (1 points) Est-ce que *time_serie* est stationnaire ? Sinon, transformez-la en série stationnaire.
6. (1 points) Est-ce que *time_serie* présente une saisonnalité? Si oui, quelle est la période de celle-ci (journalière, hebdomadaire, mensuelle etc)?
7. (6 points) La SAAQ vous contacte afin de lancer une campagne de prévention des accidents de la route. Elle veut proposer différents messages aux MRCs en fonction des caractéristiques des accidents qui s'y produisent. Bien entendu, la SAAQ ne veut pas faire une campagne par MRC donc elle vous contacte pour que vous lui proposer une segmentation des MRCs en fonction des caractéristiques des accidents qui y sont survenus.

Vous devez pour mener à terme ce projet:

- Effectuez les pré-traitement requis à la segmentation (choix des variables, standardisation, gestion des valeurs extrêmes etc.)
- Proposez un nombre de groupes et les MRCs associées à chacun de ceux-ci.
- Décrire les caractéristiques de chacun des groupes.

Vous pouvez choisir la méthode de segmentation de votre choix. Justifiez chacune des étapes.

Bon travail!