

## Projet-Partie 1 (15 points)

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

Date de remise : mardi 7 novembre 2023 avant 22h

### Directives :

- Projet en équipe de 3 personnes (inscrire vos équipes avec le lien suivant) :  
[https://docs.google.com/spreadsheets/d/1ju853NBzxO9N5Yl3fB56cfLI8f5TZMSiAQf9\\_UuGQ-c/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1ju853NBzxO9N5Yl3fB56cfLI8f5TZMSiAQf9_UuGQ-c/edit?usp=sharing)
- Pour répondre aux questions suivantes, vous aurez besoin du fichier de données qui se trouve sur Moodle dans le dossier *data\_projet.zip* et devez utiliser python.
- 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 7 novembre 2023 22h.
- Inscrire votre numéro d'équipe dans le nom des documents remis (ex : Équipe 1 – Projet P1.py) et tous les prénoms, noms et matricules des participants de l'équipe à l'intérieur de chaque document.

Dans le cadre de ce travail, vous effectuerez des manipulations de données et produirez des visualisations en Python à partir des données contenues dans les fichiers :

- environnement.csv
- genre.csv
- meteo.csv
- population.csv
- surface.csv
- rapports-accident-2021.xlsx

déposés sur Moodle dans le dossier *data\_projet.zip*.

Ces données contiennent les informations concernant les accidents de la route au Québec au cours de l'année 2021. À l'onglet **Variable** du fichier *rapports-accident-2021.xlsx*, vous trouverez une description sommaire des variables.

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.

1. (2 points) Importez les données des fichiers dans des objets de classe *dataframe* de mêmes noms. Effectuez des vérifications sur les variables afin de vous assurer que les données pourront être utilisées sans problème pour la suite. Voici quelques suggestions de vérifications :
  - Valeurs manquantes
  - Classe des colonnes
  - Pour les variables catégorielles, nombre de catégories et libellés
  - ...

Si vous pensez qu'il faut faire un certain nettoyage des données, vous pouvez prendre les décisions que vous jugez les plus pertinentes, tant que vous justifiez vos modifications à l'aide de commentaires.

Pour avoir le point pour cette question, vous devez démontrer que vous comprenez bien les données que vous travaillerez ensuite dans les autres questions. Vous pouvez utiliser les méthodes que vous préférez pour faire l'analyse. Les commentaires seront pertinents pour me permettre de suivre votre raisonnement et vos constats.

2. (1 points) Créez une variable ***EstCongéFérié*** qui contient ***True*** lorsque l'accident a eu lieu lors d'un congé férié (au Québec) et ***False*** lorsque c'était une journée régulière.

Voici la liste des congés fériés en 2021:

- le 1er janvier (jour de l'An)
- le Vendredi saint (2 avril)
- le 24 mai (Journée nationale des patriotes)
- le 24 juin (Fête nationale du Québec)
- le 1er juillet (Fête du Canada)
- le 6 septembre (fête du Travail)
- le 11 octobre (Action de grâces)
- le 25 décembre (jour de Noël)

3. (1 points) Créez une variable **EstWeekEnd** qui retourne **True** lors que l'accident est survenu durant le week-end (samedi ou dimanche) et **False** sinon.
4. (2 points) Produisez un nouvel objet de classe *DataFrame* nommé **tx\_mortalité**, qui présentera le taux de mortalité selon les catégories suivantes:
  - a) Véhicules: Accident impliquant **seulement des véhicules**
  - b) Piéton: Accident impliquant **au moins un piéton**
  - c) Moto: Accident impliquant **au moins un motocycliste**
  - d) Cycliste: Accident impliquant **au moins un cycliste**

Le taux de mortalité doit être calculé avec la formule suivante :

$$\frac{\text{Nombre d'accidents où GRAVITÉ} = \text{"Mortel"}}{\text{Nombre d'accidents total}}$$

Le taux de mortalité doit être présenté en pourcentage, doit être arrondi à deux (2) décimales et présenté avec le symbole "%".

5. (3 points) Produisez un nouvel objet de classe *DataFrame* nommé **agg\_acc\_genre**, qui présentera pourcentage du nombre d'accidents total par grandes catégories de genre d'accident et selon les mois de l'année:
  - a) Collision
  - b) Objet fixe
  - c) Sans collision

Les colonnes doivent être triées par ordre chronologique des mois.

Le pourcentage indiqué doit être arrondi à une seule décimale et présenté avec le symbole "%". Votre tableau doit contenir une ligne qui contient le total de toutes les colonnes et une colonne qui contient le total de toutes les lignes. Cette ligne et cette colonne doivent se nommer « Total ».

6. (3 points) Calculez le taux d'accidents par nombre d'habitant dans chacune des MRC (Nombre d'accident/Population). Produisez un nouvel objet de type *DataFrame* nommé **agg\_tx\_acc** qui présentera ce taux par gravité d'incident pour les 5 MRCs avec le plus haut taux d'accident global.

Le pourcentage indiqué doit être arrondi à deux (2) décimales et présenté avec le symbole "%".

Votre tableau doit contenir une colonne qui contient le total de toutes les lignes. Cette colonne doit se nommer 'Total'.

7. (3 points) Créez un nouvel objet de type *Series* nommé **temps** qui présentera le maximum d'heures écoulées entre chacun des accidents selon le mois de l'année.

8. (5 points) À partir des données fournies sur les accidents de la route, réalisez cinq (5) visualisations présentant les informations que vous jugez les plus pertinentes.

Commentez les graphiques et décrivez vos constats.

Bon travail!