

SAÉ Analyse de données, Reporting et Datavisualisation



Bilan individuel : **Alioune DIOP**

Travail effectué :

Dans le cadre de notre projet de SAE finale, j'étais responsable de la création d'un script pour l'analyse des données. Le script commence par importer les bibliothèques nécessaires, notamment pandas et chi2_contingency de scipy.stats. Ensuite, j'ai chargé les données depuis un fichier Excel qui nous était transmis par MAVIE contenant plusieurs feuilles de données, notamment les feuilles "Accident" et "BD_3quest". La première étape du processus consistait à nettoyer ces données en supprimant les colonnes non pertinentes pour notre analyse. J'ai également transformé et normalisé certaines valeurs pour assurer la cohérence des données, par exemple en regroupant les accidents survenus dans différents types de lieux sous une seule catégorie et en traitant les heures d'accidents pour les transformer en plages horaires.

```
def extraire_apres_deux_points(cellule):  
    try:  
        # Extraire la valeur après ":"  
        valeur_apres_deux_points = cellule.split('"Type d\\'accident":')[1].strip('"')  
        valeur_apres_deux_points = valeur_apres_deux_points[:-3]  
        valeur_apres_deux_points = valeur_apres_deux_points.encode().decode('unicode_escape')  
        # Retourner la valeur extraite  
        return valeur_apres_deux_points  
    except :  
        # En cas d'erreur de décodage JSON ou d'indice, retourner None  
        return None
```

J'ai ensuite remplacé les types d'accidents non répertoriés par "Autres accidents". J'ai aussi catégorisé les heures d'accidents en trois périodes : matin, après-midi et soir. Après avoir nettoyé et transformé les données, j'ai fusionné les différents ensembles de données en une seule Data Frame finale, que j'ai exportée en fichier Excel.

```
def remplacer_type_accident(type_acc):  
    if type_acc not in Liste_typeacc:  
        return "Autres accidents"  
    else:  
        return type_acc
```

La deuxième partie du script concerne l'analyse statistique. J'ai effectué des tests d'indépendance du chi-carré pour examiner les relations entre différentes variables, telles que la fatigue et le lieu de l'accident, le sexe et la survenue d'accidents, le type d'accident et la plage horaire, etc. Pour chaque test, j'ai calculé la statistique du chi-carré, la p-value, les degrés de liberté, et les fréquences attendues, puis j'ai déterminé l'importance statistique de chaque relation en utilisant des seuils de p-value de 0,05 et 0,1.

Enfin, j'ai consolidé les résultats de ces analyses dans une liste de décisions basées sur les p-values. Mon rôle dans ce projet a été crucial pour la préparation, le nettoyage, la transformation et l'analyse des données, permettant ainsi d'obtenir des résultats significatifs et exploitables pour notre étude.

```
Statistique Chi-2: 8.688905688389951  
P-value: 0.012978607625861446  
Degrés de liberté: 2  
Fréquences attendues :  
[[9.44908180e-01 2.38116861e+02 3.26938230e+02]  
 [5.50918197e-02 1.38831386e+01 1.90617696e+01]]  
Heure_cat      Après-midi  Matin  Soir  
Type_acc  
Autres accidents      2         0         2  
Brûlure, refroidissement ou autre effet thermique      3         0         0  
Choc (coup, heurt par contact avec un objet, un...      2         1         0  
Chute                 5         5         3  
Surmenage physique (sur-sollicitation du corps,...      3         0         0  
Ecrasement, coupure, perforation      3         3         1
```

Mon ressenti sur ce Projet :

En travaillant sur ce projet, j'ai souvent ressenti de la frustration et du stress face aux données manquantes et aux défis techniques. Chaque fois que je devais déboguer un problème, cela semblait parfois décourageant et éprouvant. De plus, la concentration intense nécessaire pour accomplir des tâches techniques complexes a souvent entraîné une fatigue mentale considérable.

Travailler sur des détails minutieux, tels que la manipulation de chaînes de caractères ou la gestion des types de données, a été particulièrement épuisant. Ces moments de stress et de fatigue mentale ont été des défis constants tout au long du projet.

La partie analyse statistique avec les tests du chi-carré était particulièrement intéressante. Elle m'a permis de mettre en évidence des relations potentiellement significatives entre différentes variables. Cela m'a procuré un sentiment d'accomplissement, car cela démontrait l'impact direct de notre nettoyage et préparation des données sur la capacité à obtenir des résultats analytiques significatifs.

Globalement, la réalisation de ce script a été une expérience enrichissante. Elle a renforcé mes compétences en manipulation de données, en nettoyage, en transformation et en analyse statistique. Voir le script s'exécuter sans erreur et produire des résultats analytiques clairs et exploitables a été extrêmement gratifiant.

Conclusion :

En conclusion, bien que le processus ait été rempli de défis techniques et de moments de frustration, le résultat final a prouvé la valeur de l'effort et du détail investis. Ce projet a non seulement permis d'améliorer mes compétences techniques, mais aussi de comprendre l'importance d'une bonne préparation des données pour toute analyse ultérieure.