Calyxis
PÔLE D'EXPERTISE DU RISQUE

Le 06/06/2024

Rapport du projet : Analyse de données, reporting et datavisualisation

Réalisé par :

- Leslie PLANET
- Xavier BARBEAU
- Hélène VIZOSO
- Rémi PIERRON
- Chloé DECOUST

Encadré par :

- M. BUREAU
- M. GARNIER
- Mme. CANONNE

Table des matières

Introduction :

Nous avons été chargés par MaVie observatoire de Calyxis de créer une visualisation de leurs données automatisées. Calyxis est une entreprise qui a pour but de prévoir divers risques comme les accidents de la vie courante, les risques naturels ou encore la santé et la nutrition. MaVie fait partie intégrante de Calyxis puisqu'elle est son laboratoire de recherche. Ils nous ont chargés de traiter les réponses d'un de leurs questionnaires qui traite des risques d'accidents domestiques. Tous les deux mois, les volontaires qui se sont inscrits reçoivent un questionnaire leur demandant s'ils ont eu un accident. Si c'est le cas alors ils répondent à un autre questionnaire plus poussé leur demandant des détails sur leur accident. Ils nous ont laissé libre choix de problématique, nous avons donc choisi : Comment les caractéristiques des volontaires et de leur logement permettent de déterminer le risque d'accident ? Pour pouvoir répondre à cette problématique, nous avons fourni un outil dynamique sur TKinter et un visuel sur Power BI.

I. Collecte et traitement des données :

1. La collecte :

Pour la collecte des données, nous avons commencé par récupérer nos fichiers XLSX puis par les convertir en fichier CSV. Au début, lors de la conversion de nos fichiers avec la procédure convertisseur, sur les colonnes qui contenaient du numérique comme année_nais ou Nb_pers, cela nous rajoutait à la fin ' .0'. C'était un problème parce que lors de l'import sur Power Bi, nous avions des erreurs parce qu'il ne reconnaissait pas le bon type de la variable. Pour résoudre ce problème, nous avons modifié la procédure pour pas qu'il rajoute le '.0'. À part ce problème que nous avons réglé, nous n'en avons pas rencontré d'autres.

2. Le traitement :

Pour le traitement des données ainsi que le nettoyage de la base, nous avons choisis de supprimer les colonnes de type « précisez … » et de rassembler les réponses

dans une seule pour faciliter la lecture et l'analyse. Nous n'avons pas rencontré de problème pour cette partie.

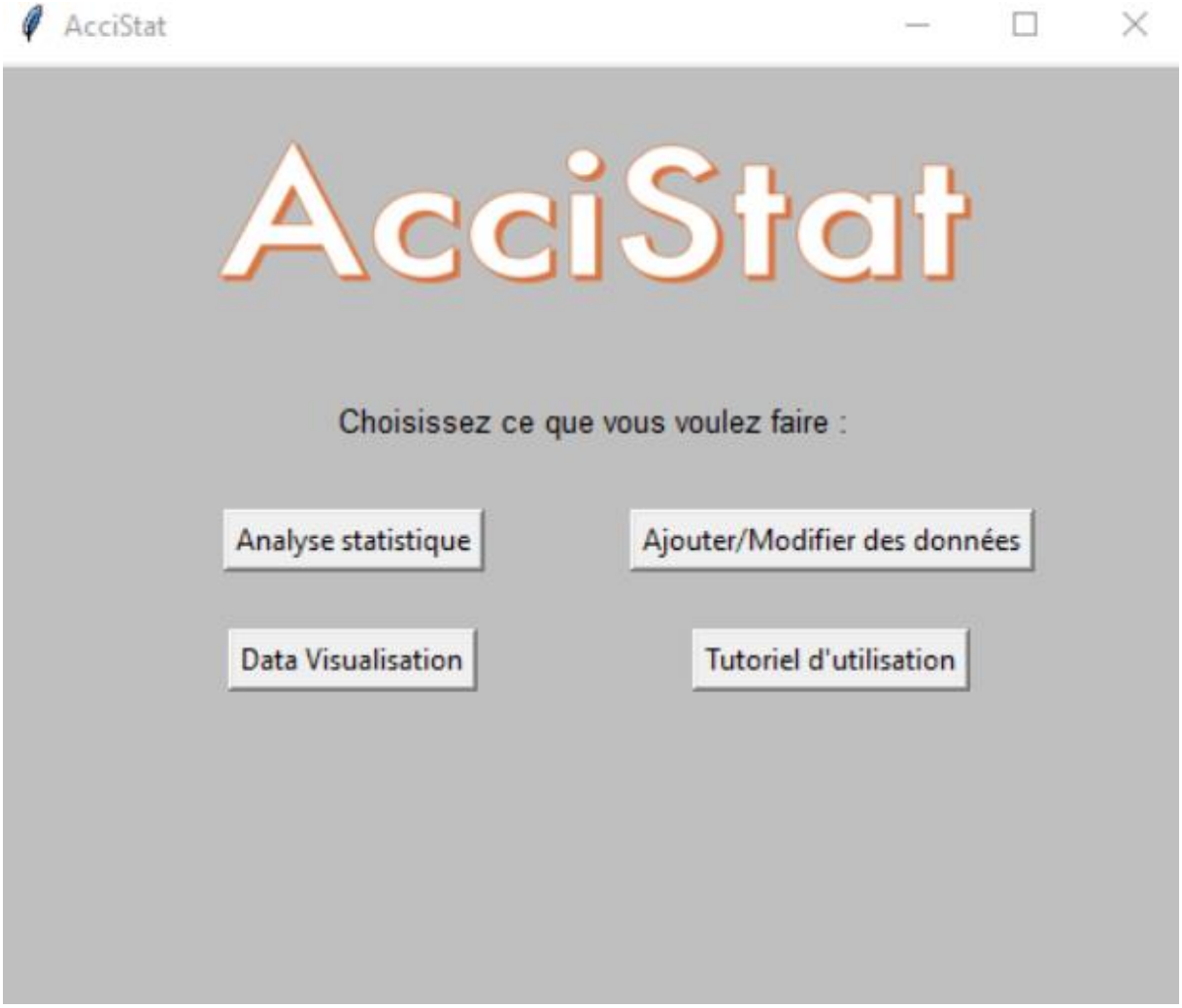
II. L'analyse statistique :

Pour effectuer notre analyse statistique en python, nous avons décidé de faire une jointure entre nos deux tables, « indiv.csv » et « accident_traite.csv » pour pouvoir mettre en relation les informations présentent dans les deux. Pour ce faire, nous avons écrit une procédure : « jointure ».

Après cette concaténation, pour chaque calcul statistique, nous avons réalisé une fonction qui retourne le résultat pour pouvoir l'afficher dans le Tkinter, par la suite. Dans un premier temps, nous avons calculé la part des individus accidentés en fonction du sexe puis en fonction de leur âge après avoir regroupé les individus par tranche d'âge de 10 ans. Nous avons aussi calculé la proportion de personnes qui ont un escalier chez eux parmi les volontaires accidentés puis parmi tous les volontaires. Ensuite, nous avons cherché à répartir le nombre d'accidents par type de logement, par département et enfin par type de situation familiale. Pour finir notre analyse, nous avons réalisé un test du khi-deux d'indépendance pour voir s'il y a un lien entre le fait d'avoir un accident et le sexe de la personne concernée. Cependant, après réflexion, nous aurions souhaité laisser le choix à l'utilisateur de choisir ses variables pour effectuer ce test, pour qu'il puisse faire le test qu'il aurait voulu et le plus intéressant pour lui.

III. L'interface utilisateur :

Nous avions pour mission de concevoir un outil dynamique répondant aux besoins du commanditaire. Nous avons donc décidé pour cette partie d'utiliser Tkinter via python. Voici à quoi ressemble notre page d'accueil.



Concernant le TKinter, nous n'avons pas rencontré de problème en particulier. Nous avons juste dû nous mettre d'accord sur la manière dont nous voulions disposer les différents boutons. Nous avons hésité pour la partie « Ajouter/Modifier » pour savoir si nous le faisions comme il est actuellement ou si, au lieu d'ouvrir une autre page pour le deuxième fichier nécessaire pour la modification, on faisait apparaître sur la même page ce dont nous avions besoin. Le bouton visualisation nous permet de visualiser les résultats de l'analyse statistiques. Le bouton ajouter/modifier des données nous permet d'ajouter un fichier et modifier de faire une concaténation. Les deux boutons d'en bas permettent d'accéder au Power Bi et au tutoriel d'utilisation sur YouTube. Nous n'avons rencontré aucune difficulté sur ces deux derniers boutons. La partie qui nous a pris le plus de temps était pour ajouter et modifier des fichiers car nous récupérons le chemin complet sous forme de string et nous devenions seulement récupérer le nom pour ensuite pour traiter les fichiers. Le fond du TKinter n'a été ajouté qu'à la fin, car nous n'avions pas encore réalisé le tutoriel d'utilisation ni terminé le

Power BI. Nous avons juste rencontré quelque difficulté lors de l'exécution de notre programme, nous avons dû relancer Spyder pour que le programme fonctionne. Autrement, nous n'avons pas rencontré de difficultés particulières.

IV. La visualisation :

Pour la représentation graphique, nous avons choisi Power Bi comme logiciel, car selon nous, il est plus simple d'utilisation qu'Excel par exemple. De plus, il nous permet une actualisation des tables de données.

La data visualisation permet de répondre à notre problématique en trois parties : le profil des volontaires et des victimes, les circonstances des accidents et enfin la représentativité de notre base. Nous n'avons pas rencontré de problèmes à part lors de l'insertion des tables avant la modification de la procédure convertisseur. Cependant, nous n'avons pas pu faire tout ce que l'on voulait du fait du faible jeu de données que nous avions à disposition. Nous n'arrivions pas à visualiser ce que nous pouvions mettre en relation entre les différentes variables. De plus, nous voulions faire une comparaison avec la population française mais nous avons trouvé que des bases de données sur le recensement de la population française par département et non globale. De plus, elles ne contenaient pas les caractéristiques que nous souhaitions, alors nous avons juste importé une table qui contient la répartition de la population française par tranche d'âge et par sexe.

V. Le bilan :

1. Ce qui ne fonctionne pas :

Dans l'idée, tout fonctionne, mais nous n'avons pas fait de contrôle au cas où les fichiers que l'on devrait rajouter ne seraient pas faits comme nos fichiers.

2. Ce qui fonctionne :

- Toute la partie code : fonction et procédure, collecte et traitement, analyse statistique
- L'interface utilisateur : Tkinter avec tous les liens (Power Bi, Tuto vidéo)
- La visualisation : Power Bi

3. Ce que l'on pourrait améliorer :

- Gestion du répertoire du fichier : pour l'instant, il faut obligatoirement que les codes et les fichiers soient dans le même dossier pour que cela fonctionne.
- Gestion des variables pour les analyses statistiques : c'est-à-dire, vérifier que les variables sont utilisables pour faire l'analyse.
- Gestion des erreurs : par exemple, qu'il y ait le même nombre de colonnes dans la table que l'on souhaite ajouter que dans notre fichier pour pouvoir l'insérer.
- Rendre l'interface utilisateur plus attractive.
- Faire des graphiques plus pertinents sur Power Bi (si nous avions un jeu de données plus conséquent).
- Régler les problèmes d'encodage pour certains caractères.
- Mettre des graphiques sur l'interface Tkinter.
- Faire des prédictions (si nous avions eu plus de personnes qui ont eu un accident).

Conclusion :

Pour conclure, ce projet ne nous a pas inspirées du fait de la faible quantité de données. Nous aurions aimé avoir plus de données afin de pouvoir consolider nos observations, faire des tests statistiques plus poussées et nous aurions aussi aimé avoir plus de données utiles, c'est-à-dire pas de réponse rédigée par les volontaires.

Vous pourrez retrouver ce que chacun a fait ainsi que son ressenti dans les rapports individuels.