

09:00 / visio

## NOTES

- **QUESTIONS SUR LES DONNÉES** : Doit-on dissocier les réponses 1 et 2 ou pouvons-nous les regrouper ?

Les rubriques doivent être mises en valeur. Garder les questions car ce ne sont pas les mêmes rubriques. Ainsi, on peut contextualiser. Ensuite, il faudra créer des chapitres séparés, par exemple avec un markdown, en ajoutant un astérisque pour les différencier.

- **BERTopic** : BERTopic est normalement un modèle déjà prêt. Il faut essayer de faire tourner les exemples sur les textes, puis procéder à une extraction des données. Tester l'ajout des textes dans le modèle.
- Distribution multi-sujets : faire des visualisations, par exemple des aires des graphes. Une fois que nous serons plus à l'aise, on verra ce qu'on peut faire.
- Pour le moment, nous testons sans savoir exactement où cela nous mènera.
- Il est important de comprendre ce que sont les embeddings (vecteurs). Il faut vérifier les multiples représentations, cela permet de varier les présentations (par exemple : "je suis content" vs "content je suis", le modèle reconnaît ce genre de variations comme dans Star Wars).
- **SQLite** : Le code a été vérifié et des conseils ont été donnés. Attention à ne pas utiliser les notebooks pour l'exécution finale, mais plutôt du code propre en Python avec des fonctions. utiliser les notebooks seulement pour l'exploration.
- La structuration du code est aussi importante que le code lui-même.
- Objectifs de sqlite : Nous créons des tables pour structurer les relations, refaire la visualisation, et réinjecter ce que nous avons dans les tables relationnelles
- **Questions / Réponses** : Dans le cadre des questions/réponses, il y a plusieurs lieux et personnes à prendre en compte, donc il faut conserver ces informations.

- Dans un premier temps, répartir les relations par lieu, et mettre en évidence la personne qui en parle dans un champ dédié (par exemple avec un markdown qui l'extrait du reste).
- Est-ce que nous avons un répondant qui communique avec plusieurs lieux, ou un seul qui gère plusieurs lieux et répond pour les quatre ? Il faut attribuer les informations correctement. Sabrina devra être consultée pour cela (garder uniquement celui qui communique dans le texte).
- **Excel** : pour la colonne Responsable : il n'y a pas de formatage défini, ce qui complique les choses. Ajouter un champ "responsable :" si le modèle le trouve, tant mieux, sinon tant pis (vérifier cela avec Sabrina).
- **Structuration** : bien délimiter les différentes tâches, par exemple : nettoyage des données, puis définir des sous-tâches dans le nettoyage, etc. Cela doit se refléter dans la structure du code.
- Faire des petits tests unitaires simples (nous sommes libres de choisir si nous les utilisons ou non).
- **Décembre** : En décembre, il nous faut une visualisation. Un modèle BERTopic qui fonctionne, il faudra définir les classes dans les textes (classification totalement non supervisée).

## TÂCHES

1. Mettre le code au propre, et préfixer les fonctions avec un nom de répertoire bien choisi, comme "ars".