# COMMENT MENER A BIEN MON PROJET

## **COLLECTION DES DONNEES**

**Etape 1** : **quelles données vais-je pouvoir collecter pour répondre à la demande de mon client ?**

Je liste les onglets qui pourraient me servir à répondre aux questions et ceux qui ne sont pas pertinents.  Dans mon cas tous peuvent m’apporter des informations pertinentes hormis les différentes options de l’onglet recherche sur le web.

Je choisis spécifiquement la France car le reste du monde n’entre pas en ligne de compte. Je commence donc à traiter mes problématiques une à une.

**Pblmatique n°1 : l’intérêt des français sur le thème Végan**

J’investigue sur l’intérêt des français sur l’alimentation végan par rapport aux autres centres d’intérêts végan qui sont pertinents :

* végan dans le domaine de la santé
* végan dans le domaine du bien-être
* végan dans hobbies et loisirs
* vegan dans shopping
* vegan dans marché commerciaux

En comparaison avec le terme végan dans l’alimentation

* Ceci sera l’un de mes indicateurs qui d’une façon simple me donnera une première idée du marché
* Je remonte à 6 années en arrière car avant cette période ce n’était pas le sujet en vogue

**Pblmatique n°2 : les plats préférés des français**

Vu que je souhaite être à la pointe de la ‘mode’ en termes de dégustation, je regarde uniquement les tendances de l’année. (Aux Vues du contexte je prends les 2 années précédentes, le Covid a impacté les habitudes alimentaires notamment les restaurants). Je regarde donc les requêtes et les sujets associés aux mots suivants : ‘recette vegan’, ‘alimentation végan’ et ‘restaurant vegan ’.

**Pblmatique n°3 : localiser les régions qui pourraient être les plus intéressantes pour s’implanter**

Je vais donc regarder l’évolution temporelle de l’intérêt pour la nourriture végan par région et par ville. (Avec le terme vegan et restaurant vegan). Je collecte alors les données sur les 5 dernières années.

**Etape 2 : Je caractérise mes données**

* Quels types de données j’obtiens ?
* Quelle est l’unité de mesure ?
* Y a-t-il des informations manquantes qui m’empêcheront de visualiser ce que je souhaite ? Si oui, comment vais-je y parer ?

**Etape 3 : Aux vues du jeu de données que j’obtiens, qu’est ce que je vais pouvoir et vouloir montrer ?**

* Quels types de graphes vais-je pouvoir utiliser ? mentalement je me représente déjà une maquette de mon dashboard
* Quelles difficultés je vais potentiellement rencontrer ? Comment pourrais-je y parer ?

1ère embuche : mes fichiers de localisation ne contiennent pas les coordonnées spatiales des régions. Je ne pourrais donc pas visualiser une carte de France, je dois donc ajouter un jeu de données avec les coordonnées de la France

En voici un avec les régions :

<https://raw.githubusercontent.com/gregoiredavid/france-geojson/master/regions.geojson>

2ième embuche : concernant les plats préférés, je possède des données textuelles qu’il faudrait que je classifie. 2 solutions envisageables : je crée un nuage de mots, ou bien je classifie dans un premier temps manuellement en définissant des catégories (nous verrons plus tard comment résoudre ce problème avec la bibliothèque nltk spécifique à la NLP)

* A la fin de cette étape, j’ai téléchargé l’ensemble de mon jeu de données dans mon fichier données\_originales
* Je copie colle l’intégralité dans mon fichier données\_preparees. C’est avec celles-ci que je vais pouvoir commencer à travailler.
* J’ai une idée précise de la direction de mon projet, je sais où et avec quoi je vais pouvoir explorer mes données

## **NETTOYAGE DES DONNEES**

J’ai beaucoup de fichier CSV je souhaite organiser mes informations dans un seul fichier de façon ordonnée. Pour ce faire j’utilise Pandas pour me créer un dataframe adapté. Dans notre cas il serait plutôt judicieux de créer un dataframe pour chaque problématique.

La gestion des données manquantes et l’encodage des données catégoriques (s’il y en a) se fera avec sklearn (voir notebook)

## **EXPLORATION ET ANALYSE**

Plusieurs options s’offrent à moi, j’utilise la bibliothèque dans laquelle je suis le plus confortable.

Pour ceux qui sont à l’aise avec python, essayer de le faire de différentes manières ! (matplotlib, seaborn, plotly, dash …)

* J’ai réalisé diverses analyses je vais donc pouvoir synthétiser toutes mes trouvailles et interpréter mes graphes pour en tirer des conclusions.
* Je mets mes scripts dans le dossier analyse

## **CONCLUSION**

Je sélectionne ceux qui me paraissent les plus pertinents

* Je crée un notebook clair avec commentaire que je mets dans le fichier insights

Je prépare ma restitution : j’élabore ma présentation visuelle et mon argumentaire. Je n’ai que 5 minutes donc 5 slides maximum

* Je mets ma présentation dans le fichier final