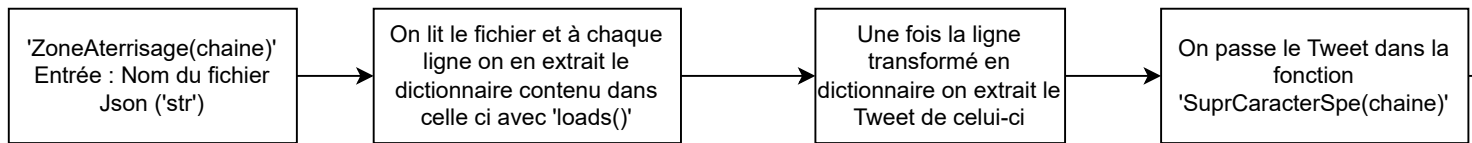
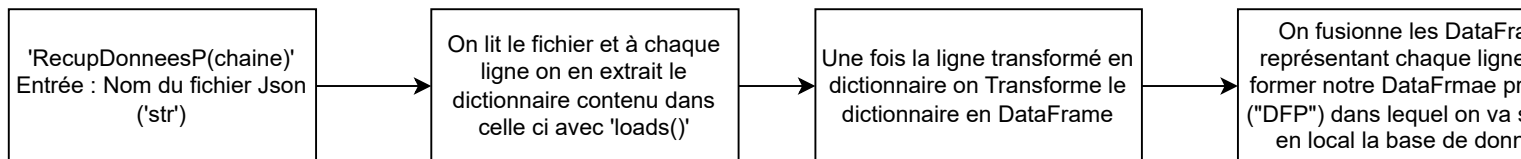


Étape n°1 : Zone d'Aterrisage

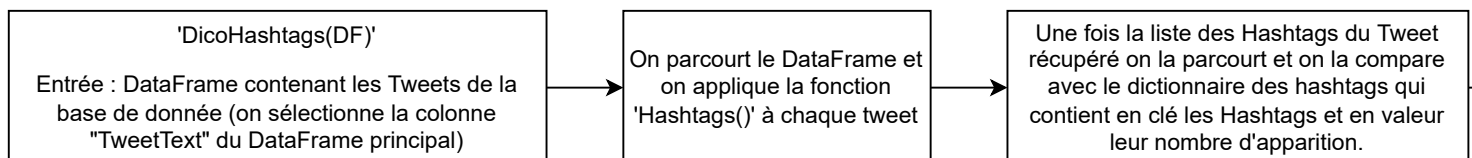
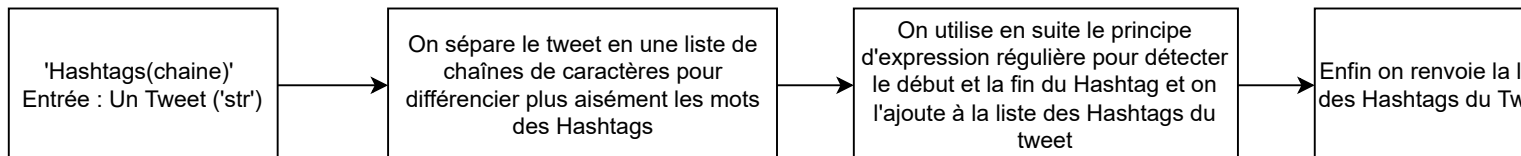


Étape n°2 : DataFrame Principal



Étape n°3 : Opérations d'Analyse

Méthode : Récupération des Tweets du **DataFrame**
(principal) → **Analyse** → **DataFrame** (d'analyse) Les Données d'Analyse seront liées
aux données du **DataFrame** principal par leur index (le numéro de leur ligne
tout simplement)



La même **méthode** d'analyse est en
suite utilisée pour analyser les @, les
sentiments et les topics.

Étape n°4 : Complément de la base de donnée (donc du DataFrame principal 'DFP')

L'enjeu est ici de compléter notre base de données incomplète. En effet notre base de données comporte pour chaque tweet un champ ID mais aucun "Pseudo" ou "Nom d'Utilisateur". C'est pourquoi nous allons ajouter à

→ On obtient le Tweet sans caractère spéciaux, on l'écrit donc dans le fichier Json 'Zone d'atterrissage'

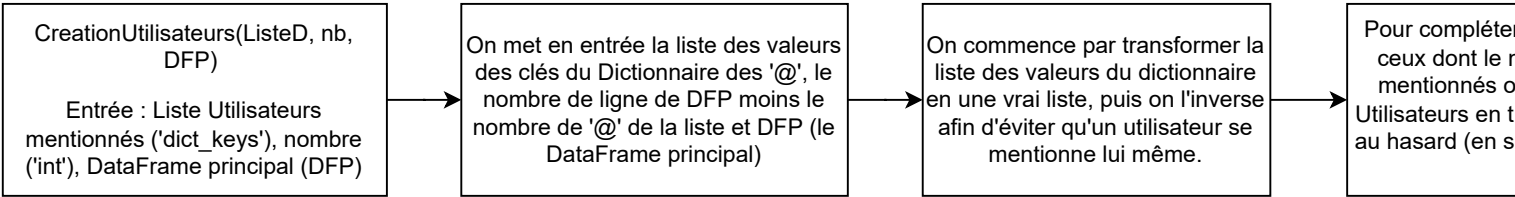
ame
e pour
incipal
stocker
ées

iste
weet

→ Si le Hashtag n'est pas dans le dictionnaire on l'ajoute.
Sinon on ajoute 1 à sa valeur (qui modélise le compteur du nombre de mention du Hashtag)

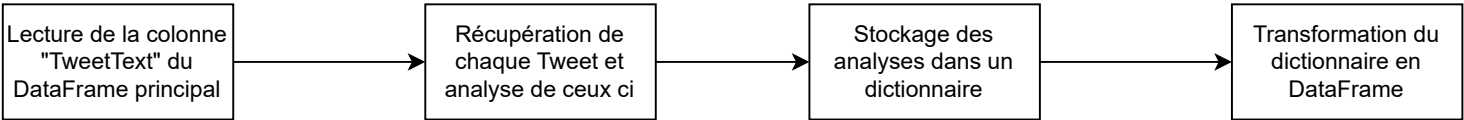
notre DataFrame principal ce nouveau champ "Utilisateurs".

La création d'Utilisateurs est essentielle au bon déroulement de certaines opérations d'Analyses, notamment pour 'Les utilisateurs mentionnés par un utilisateur spécifique'. Il est donc impératif que tous les utilisateurs mentionnés dans les tweets soient présent au sein de la base de données, donc dans le champ "Utilisateurs" du DataFrame DFP.



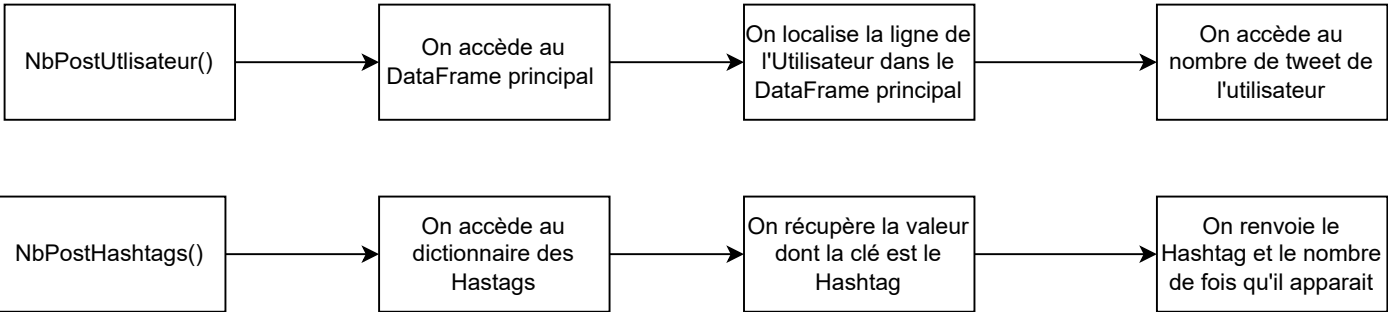
Étape n°5 : Stockage des données d'Analyse

Par soucis de rapidité d'exécution nous avons opter pour la méthode de parcourir le DataFrame contenant tous les tweets et d'effectuer au fur et à mesure les opérations d'analyses sur ceux-ci pour en même temps les réunir dans un même dictionnaire et faciliter la création d'un DataFrame réunissant toutes les analyses.



Étape n°6 : Fonctionnalités de Inpoda

Maintenant que l'on dispose de toutes les analyses de la base de données stockées dans des structures adaptées, on peut se préoccuper des fonctionnalités de Inpoda



er les Utilisateurs en plus de
nom est connu car ils sont
n crée de toute pièce des
irant des lettres et un chiffre
appuyant sur les intervalles
d'Unicode).

A chaque utilisateur créé on ajoute son nom
précédé d'un '@' dans la liste des Utilisateurs.
Puis on crée un DataFrame contenant comme
nom de colonne "Utilisateurs", comme valeur
pour chaque ligne le nom d'un Utilisateur.

Enfin on fusionne ce
DataFrame d'utilisateurs
d'une seule colonne avec
le DataFrame principal
('DFP')