

PANDAS : LE B.A.-BA

- ✓ Définir un mask lié à un ou des critères à appliquer pour filtrer des données
 - `mon_mask=(mon_df['col1']==valeur_voulue)`
 - pour une conjonction de critères : `mon_mask=(mon_df['col1']==valeur_voulue)&...`
- ✓ Appliquer un mask pour extraire la sélection voulue
 - `Selec_donnees=mon_df[mon_mask]`
- ✓ Utiliser `[...]` pour renvoyer un dataframe à partir d'un autre dont on ne souhaite garder qu'une ou plusieurs colonnes
 - `Recup_dataf=mon_df[['col1','col2']]`
- ✓ Utiliser `[...]` pour renvoyer une series (à une dimension donc) qu'on peut tout particulièrement exploiter pour une représentation avec pyplot
 - `Recup_series=mon_df['col1']`
- ✓ Définir une nouvelle colonne dans un dataframe avec des opérations réalisées à partir des valeurs d'autres colonnes
 - `mon_df['new_col']=mon_df['col1']+mon_df['col2']`
- ✓ Réaliser une jointure entre deux dataframe
 - `mon_df_join=pd.merge(...)`
- ✓ Effectuer des calculs synthétisant des données sur des segments de population définis selon les différentes modalités d'une ou plusieurs colonnes
 - `mon_df_calc=mon_df(['col_criterie']).sum()` (ou `mean()`, ou `median()`...)

REALISATION

- ✓ Savoir réaliser des graphiques divers avec pyplot (pie, bar, hist, boxplot, scatter, plot) et lire les données qu'on peut y trouver (savoir interpréter : expliquer, donner le sens)
- ✓ Savoir représenter une droite de régression en ayant calculé les caractéristiques d'un modèle linéaire avec `scipy.stats.linregress`

INTERPRETATION

- ✓ Savoir interpréter la valeur d'une médiane, de quartiles et déciles
- ✓ Savoir analyser les différences éventuelles entre moyenne et médiane (ou l'absence de différence : répartition semblable des valeurs de part et d'autre de la médiane)
- ✓ Savoir identifier des indicateurs permettant de mesurer la dispersion des données
- ✓ Savoir interpréter la valeur d'un coefficient de corrélation et du coefficient directeur de la droite de régression