

Mini projet Data Frames Statistiques

Code pour inclure notre fichier valpineta_log.csv au projet python :

```
valpineta=pd.read_table('F:\\BUT-Informatique\\S2\\R2.08 - Stats\\dataFrames\\valpineta_log.csv',
                        sep=";",
                        index_col=(0),
                        encoding='ANSI')
```

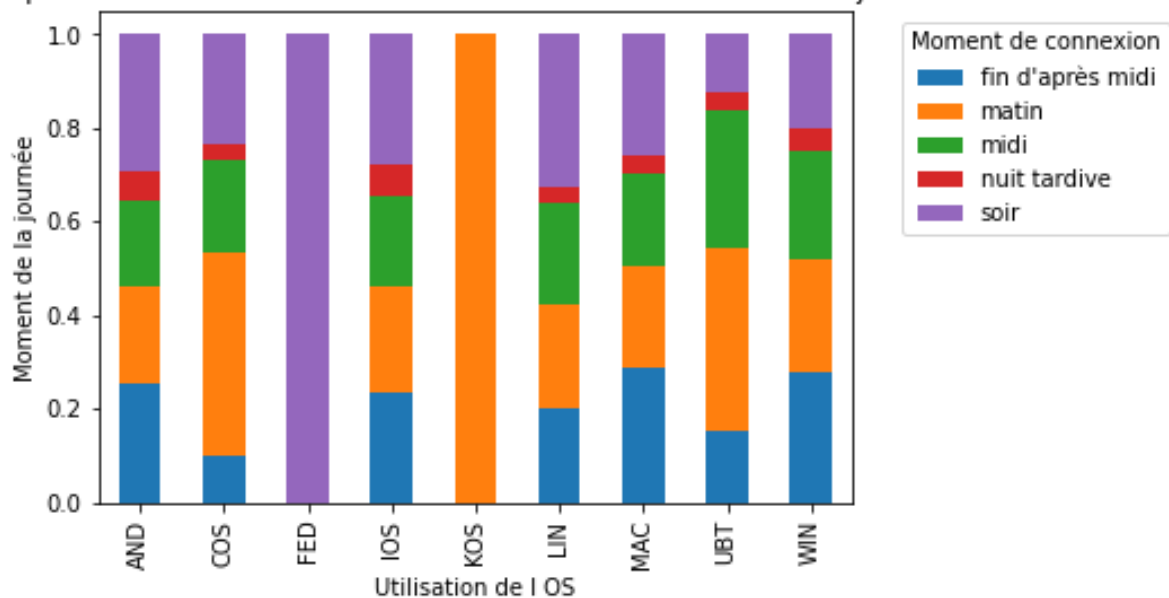
Analyse bivariee :2 Variables qualitatives

Code associé :

```
croise=pd.crosstab(valpineta['config_os'],valpineta['moment de connexion'],normalize='index')
croise.plot(kind="bar",stacked=True)
#croise.index
#croise=croise["matin","midi","fin d'après midi","soir","nuit tardive"]
#plt.legend(title="Moment de connexion",bbox_to_anchor=(1.05,1.0,loc='upper left'))
plt.title('Répartition de l utilisation des OS en fonction du moment de la journée')
plt.xlabel('Utilisation de l OS')
plt.ylabel('Moment de la journée')
plt.show()
```

Graphique :

Répartition de l utilisation des OS en fonction du moment de la journée

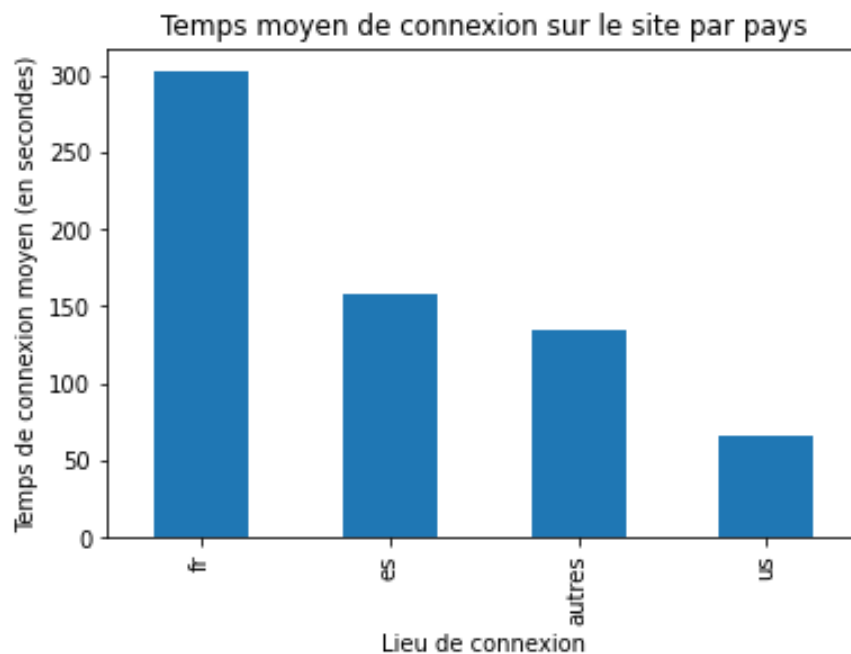


Variable qualitative et une quantitative 1 : diagramme en barres

Code associé :

```
croise2=valpineta['visit_total_time'].groupby(valpineta['countrié']).mean()
croise2 = croise2.sort_values(ascending = False)
croise2.plot(kind='bar')
plt.title('Temps moyen de connexion sur le site par pays')
plt.xlabel('Lieu de connexion')
plt.ylabel('Temps de connexion moyen (en secondes)')
plt.show()
```

Graphique :



Variable qualitative et une quantitative 2 : diagramme en moustaches

Code associé :

```
valpineta.boxplot(by='countrié', column=['visit_total_time'], grid=False,
                  showmeans=True,
                  showfliers=False)
plt.title('Temps moyen de connexion sur le site par pays')
plt.xlabel('Lieu de connexion')
plt.ylabel('Temps de connexion moyen (en secondes)')
plt.ylim(-100,600)
plt.show()
```

Graphique :

