

# Projet NF26

## *ASOS - Italie 2011 à 2014*

Claire Guyot - Haifei Zhang  
TD2-5-6



# Présentation des données



## asos.csv

- 3175968 données
- 31 variables
- 105 stations

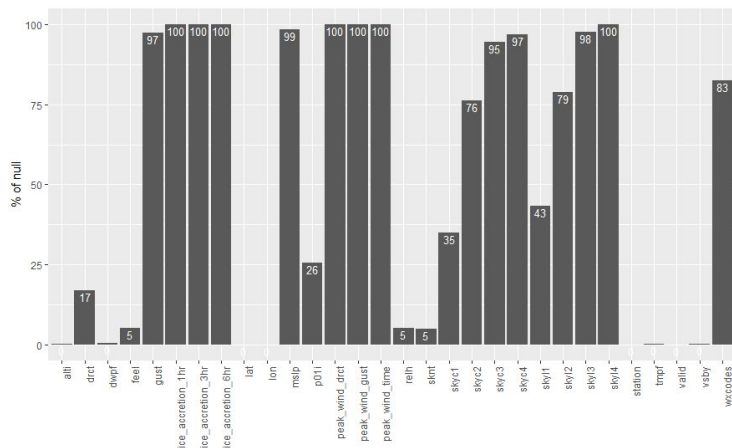


TABLE 1 – Variables du jeu de données ASOS

<i>station</i>	id de la station
<i>valid</i>	timestamp de l'observation
<i>tmpf</i>	température de l'air (F)
<i>dwpf</i>	température au point de rosée (F)
<i>relh</i>	humidité relative (%)
<i>drct</i>	direction du vent (deg. du Nord)
<i>sknt</i>	vitesse du vent (noeud)
<i>p01i</i>	heure de précipitation
<i>alti</i>	altimètre de pression (pouce)
<i>mslp</i>	pression au niveau de la mer (mbar)
<i>vsby</i>	visibilité (mile)
<i>gust</i>	rafales de vent (noeud)
<i>skycx</i>	couverture du niveau $x$ du ciel ( $x = 1,2,3,4$ )
<i>skylx</i>	altitude du niveau $x$ du ciel (pied) ( $x = 1,2,3,4$ )
<i>wxcodes</i>	codes du temps actuel
<i>feel</i>	température ressentie (F)
<i>ice_accretion_zhr</i>	accumulation de glace en $x$ heures (pouce) ( $x = 1,3,6$ )
<i>peak_wind_gust</i>	rafales de vent maximale (noeud)
<i>peak_wind_drct</i>	direction de la rafale de vent maximale (deg. du Nord)
<i>peak_wind_time</i>	instant de la rafale de vent maximale
<i>metar</i>	autres observations non traitées (format metar)

# Stockage des données

## Stockages géographiques :

- ❖ stations
  - latitude, longitude
  - stations
- ❖ asos1
  - latitude, longitude
  - année, mois, jour, heure, minute

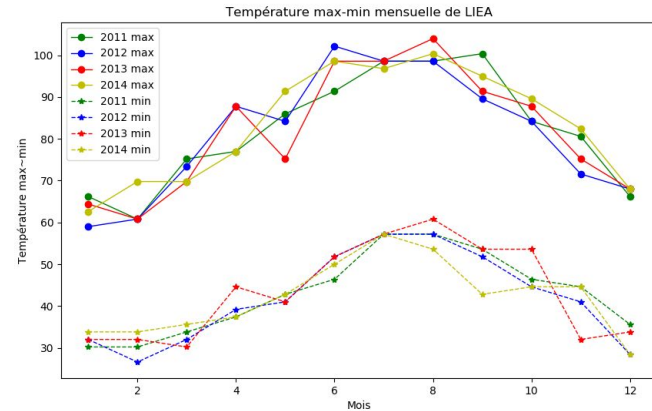
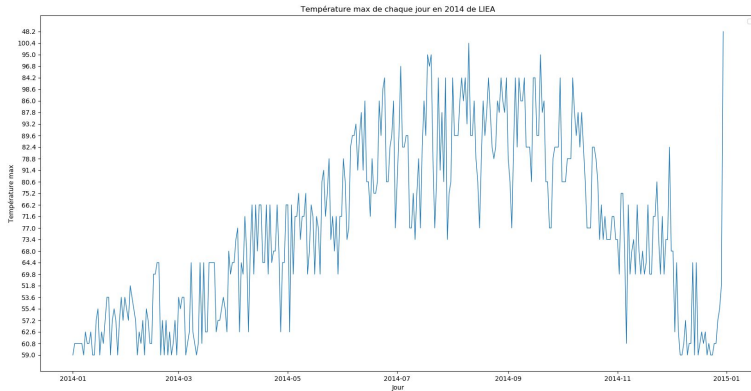
## Stockage temporel :

- ❖ asos2
  - instant de l'observation
  - station, latitude, longitude

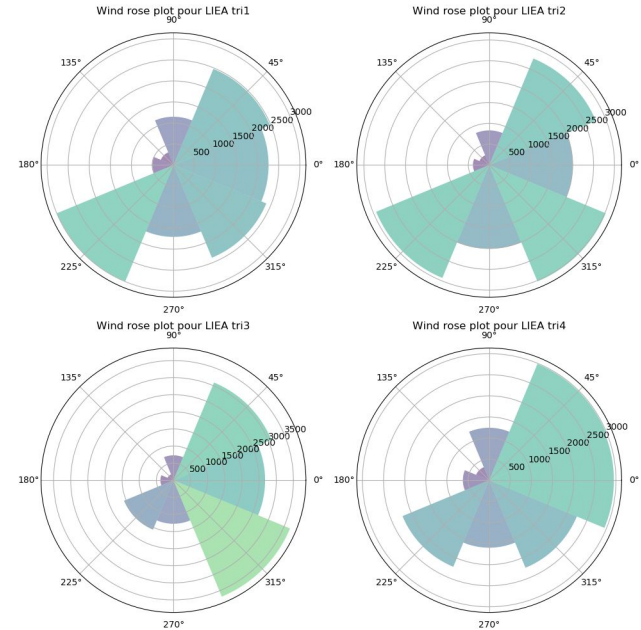
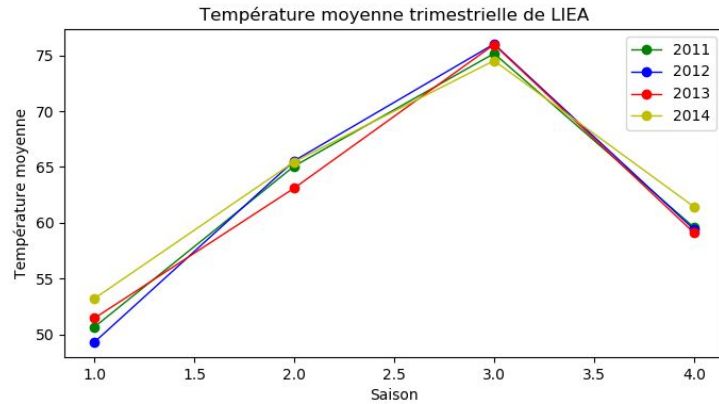
# Problématique 1

*“Pour un point de l'espace donné, je veux pouvoir avoir un historique du passé, avec des courbes adaptées. Je veux pouvoir mettre en évidence la saisonnalité et les écarts de saisonnalité.”*

➤ matplotlib



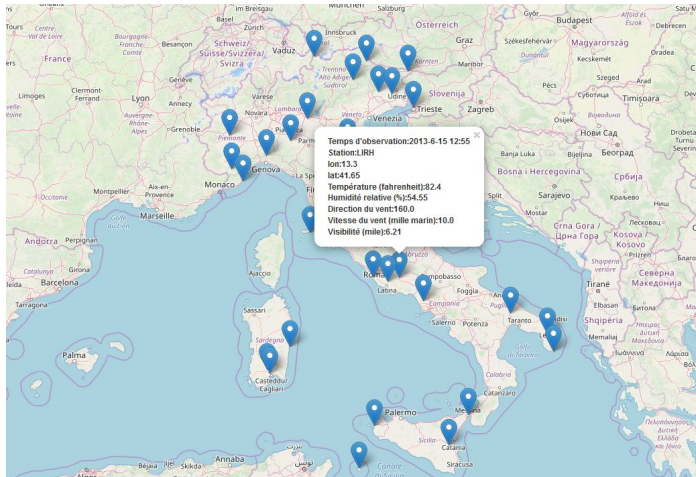
# Problématique 1



# Problématique 2

*“Pour un instant donné, je veux pouvoir obtenir une carte me représentant n'importe quel indicateur d'une station.”*

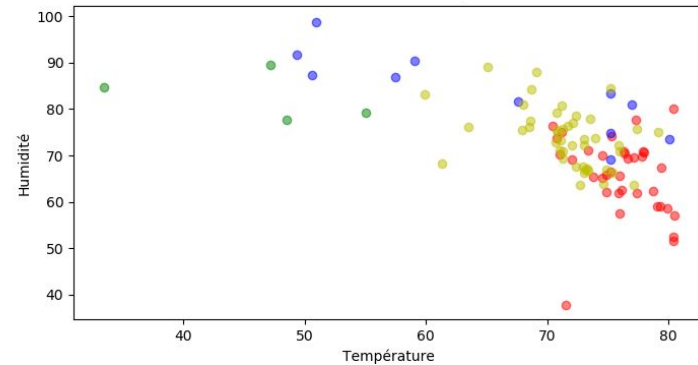
➤ folium



# Problématique 3

*“Pour une période donnée, je veux pouvoir clusteriser l'espace et représenter cette clusterisation.”*

- k-means
- (station, tmp\_moy, tmp\_std, tmp\_min, tmp\_max, relh\_moy, relh\_std, relh\_min, relh\_max)



# Conclusion

- ❖ Prise en main de Python, notamment de *matplotlib*
- ❖ Application de *Cassandra* et *Spark* en contexte réel et concret
- ❖ Donnée incohérente : la station QLQ
- ❖ *Clustering* non optimal

