# Projet NF26

Données : France de 2005 à 2014

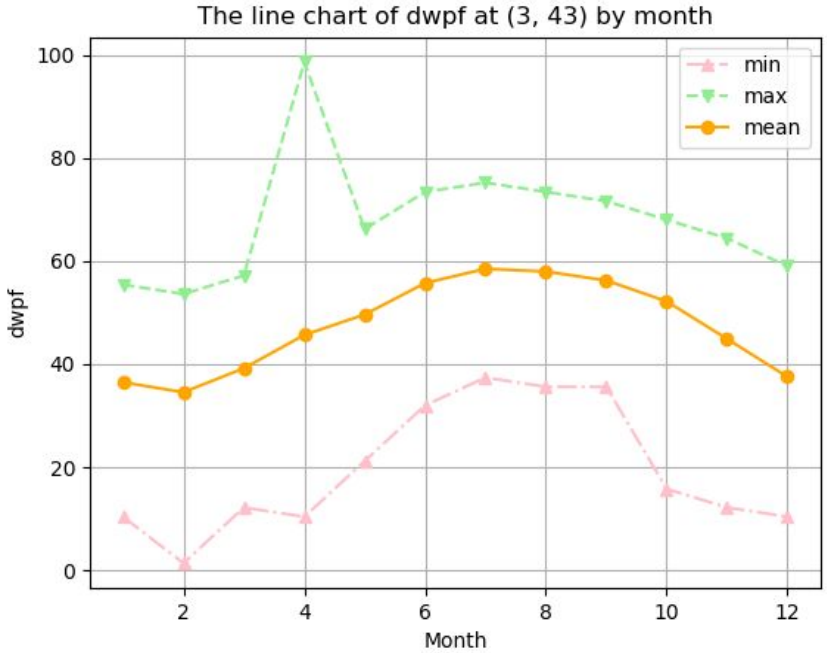
**Hugo DURET**
**Haojie LU**

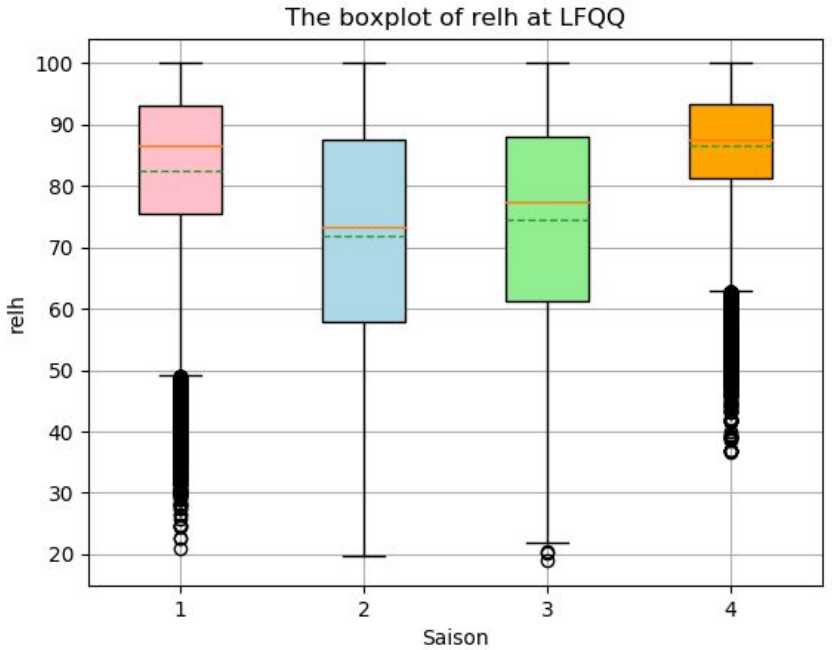
UTC - PRÉSENTATION
PROJET NF26
LU Haojie
DURET Hugo

# I. - Résultats du projet : Q1

➔ *linechartparmois(point, indicateur) ⇒ linechartparmois((3,43), 'dwpf')*

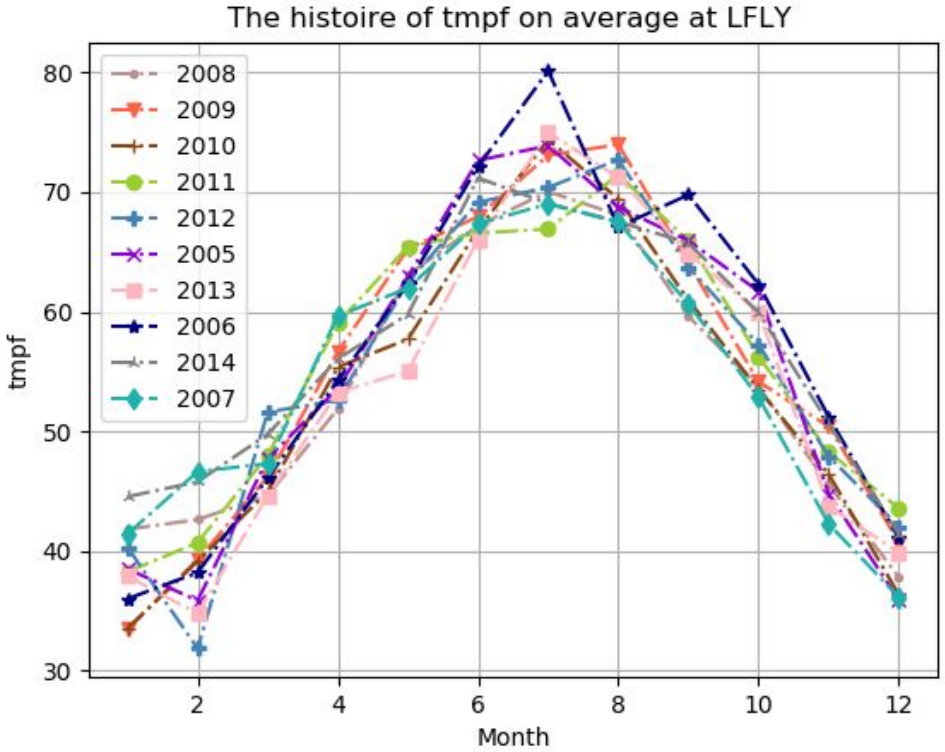
➔ *boxplot(point, indicateur) ⇒ boxplot('LFQQ', 'relh')*



# I. - Résultats du projet : Q1

➔ *linechart_histoire(point, indicateur) ⇒ linechart_histoire('LFLY', 'tmpf')*

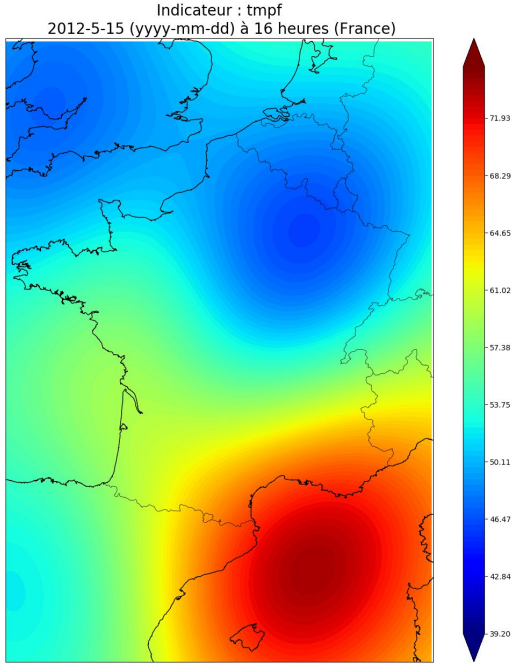


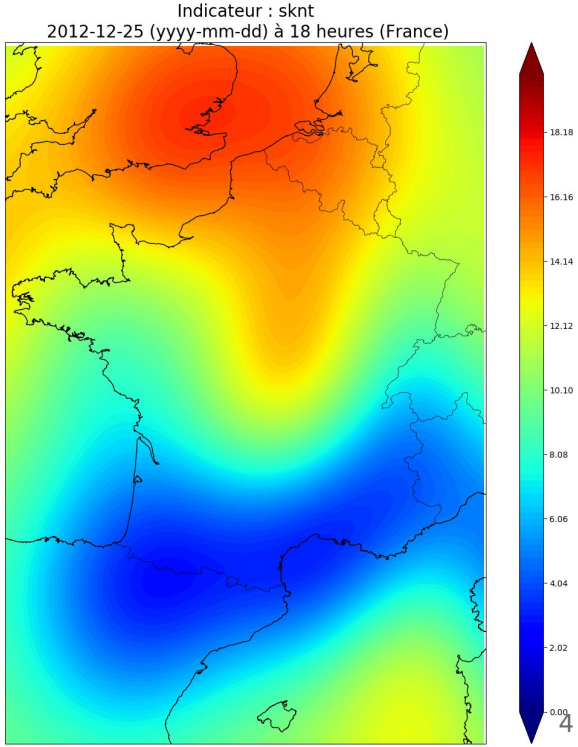
# I. - Résultats du projet : Q2

➔ *get_plot_per_hour_and_indicator(year, month, day, hour, indicator)*

⇒ *get_plot_per_hour_and_indicator(2012, 5, 15, 16, 'tmpf')*

⇒ *get_plot_per_hour_and_indicator(2012, 12, 25, 18, 'sknt')*



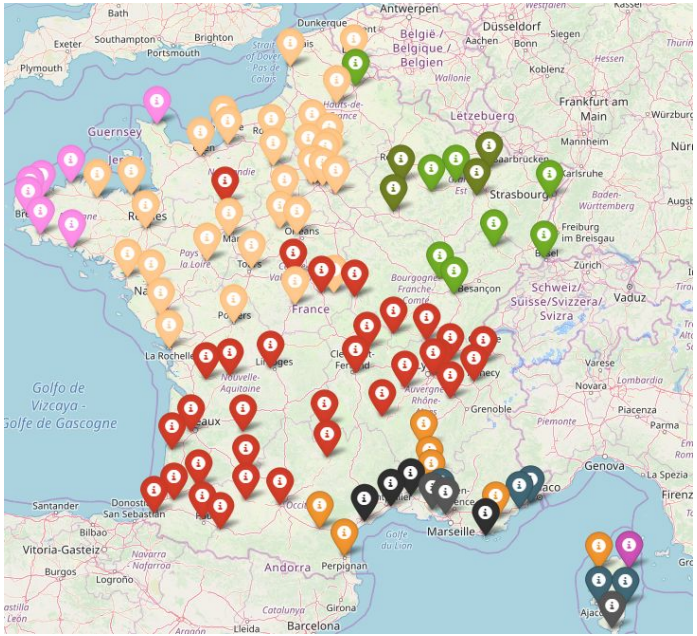
UTC - PRÉSENTATION
PROJET NF26
LU Haojie
DURET Hugo

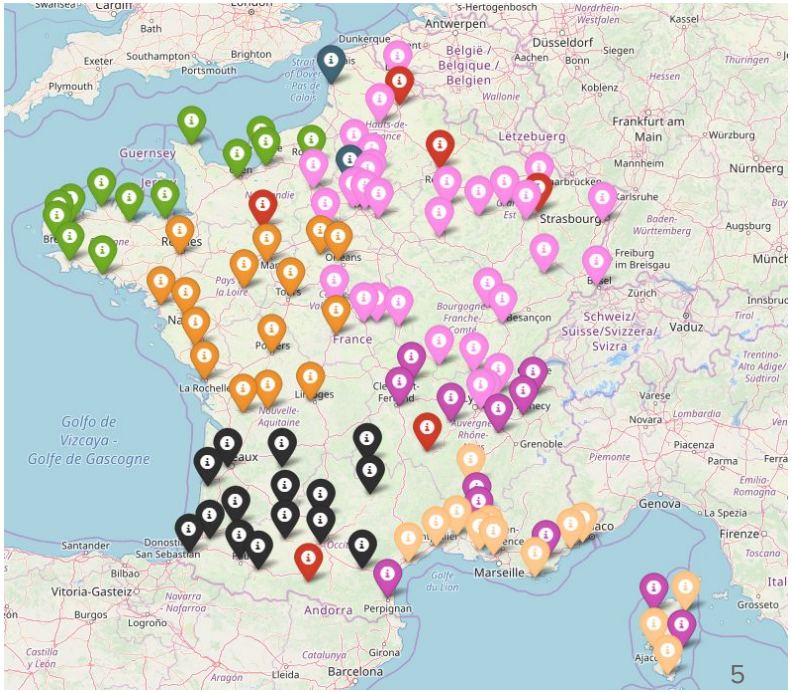
# I. - Résultats du projet : Q3

➔ *mapofclusterisation(start_date, end_date, nombre_de_gruope(10 par défaut))*

⇒ *mapofclusterisation('2012-06-09', '2012-08-09')*

⇒ *mapofclusterisation('2012-03-09', '2012-08-22',8)*



# II. - Choix de stockage

Q1

| Clé de partitionnement | Clés de tri |
|---|---|
| station | year, month, day, hour, minute |

Q2

| Clé de partitionnement | Clés de tri |
|---|---|
| year, month, day, hour | minute, station |

Q3

| Clé de partitionnement | Clés de tri |
|---|---|
| timestamp_day | hour, minute, station |

# III- Traitement : Q1

**Line Charts par mois :**

( station, année, mois, jour, ind ) ⇒ ( mois, (1, ind, ind, ind) )

r : ( a , b ) ⇒ ( a[0]+b[0], min(a[1],b[1]), max(a[2],b[2]), a[3]+b[3] )

( mois, (s0, s1, s2, s3) )⇒ ( mois, (s1, s2, s3/s0) )

**Boxplots pour la saisonnalité :**

( station, année, mois, jour, ind ) ⇒ ( saison, ind ) => ( saison, *[itérable]* )

On crée les boxplots à partir du mapping obtenu précédemment.

# III- Traitement : Q1

**Line Charts par mois pour toute la période :**

( station, année, mois, jour, ind) ⇒ ( année, mois, (1, ind )

r : ( a , b ) ⇒ ( a+b )

( (année, mois), (s0, s1) )⇒ ( (année, mois), (s1/s0) )

( (année, mois), s0 )⇒ ( année, (mois, s0) ) =>( année, [itérable] )

# III- Traitement : Q2

**Création d'une heatmap pour un indicateur :**

Données des stations pour une heure

Krigeage

Heatmap

UTC - PRÉSENTATION
PROJET NF26
LU Haojie
DURET Hugo

# III- Traitement : Q3

## Traitement de données

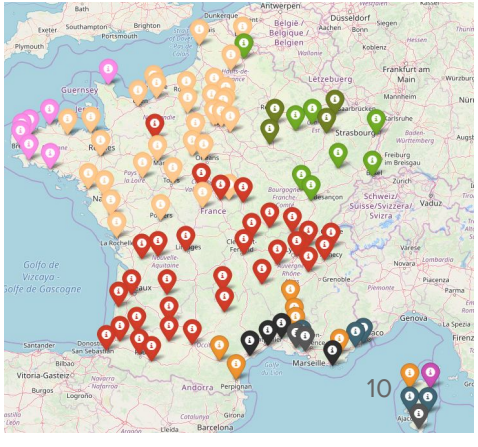
- *Principe : numérique, non nul  => (tmpf, dwpf, relh, drct, sknt, alti, vsby ,skyl1, feel）*
- *Moyenne*

## Clusterisation : Kmeans

- Initialiser au hasard K centres des catégories.
- Mapping :le centre *i* le plus proche, la distance d'Euler, une pair clé-value (i,(1,enregistrement)).
- Réduction et Mapping : mettre à jour le centre de chaque catégorie avec le moyen de tous les enregistrements actuels de cette catégorie.

## Représentation

- *Folium*
- *(latitude, longitude)*
- HTML



LU Haojie
DURET Hugo

# Conclusion

Gros volume de données

Problèmes dans les données (valeurs nulles, stations sans données avant 2010...)

Choix de stockage spécifique à une question

Cassandra pour le stockage en Orienté Colonnes

Spark pour des calculs en streaming

# MERCI