

Predictus Olympiae:

Prédiction des  
performances des nations  
aux J.O. d'été.

JULIEN MUHLEMANN

CRISTHIAN RONQUILLO

DR. ING JULIEN BILLETER





---

# Avancement du projet

- Recherche préalable de données
- Etablissement du cahier des charges
- Nettoyage général des données
- Recherche complémentaire
- Visualisation des données

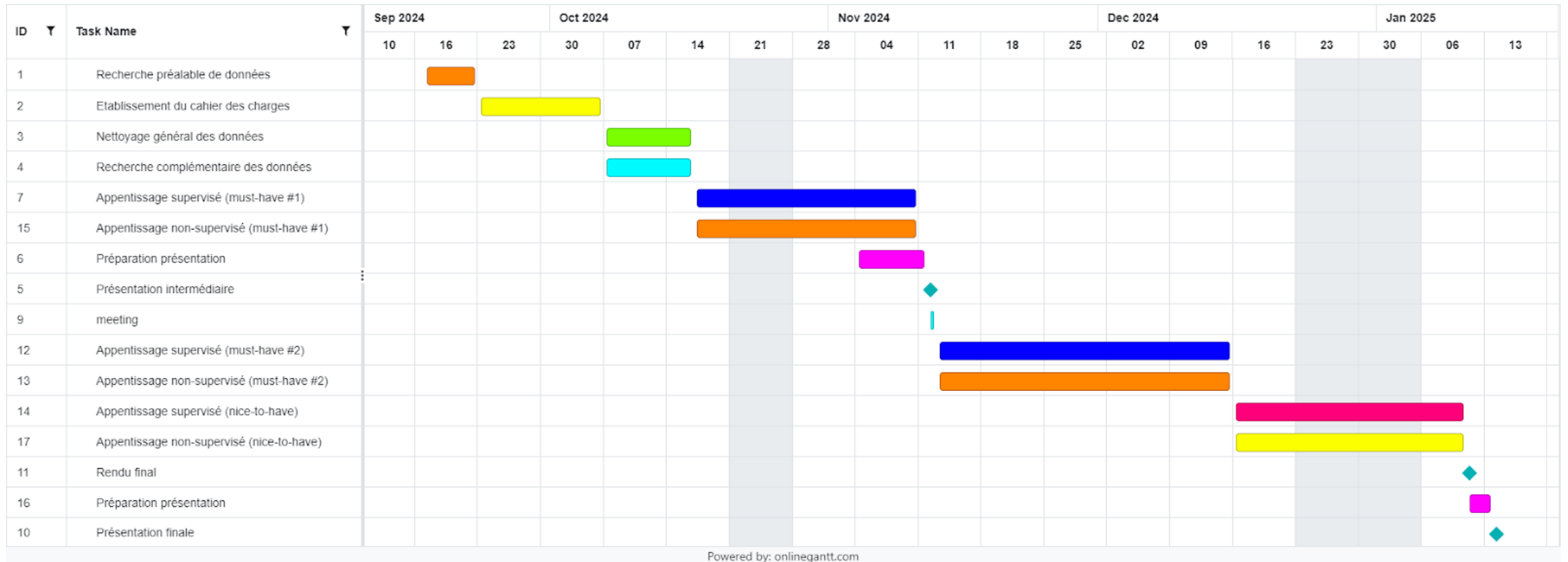
# Recherche de datasets

Type ▾	People ▾	Modified ▾
Name	↑	
archive.zip		
archive2.zip		
athlete_events.csv.zip		
Cahier des charges-exemple-GML.docx		
Countries GDP 1960-2020.zip		
Countries population by year 2020.zip		
data socio economique.zip		
Global Economy Indicators.zip		
Journal de travail-exemple-GML.xlsx		
Notes Projet JO.pdf		
Summer-Olympic-medals-1976-to-2008.zip		

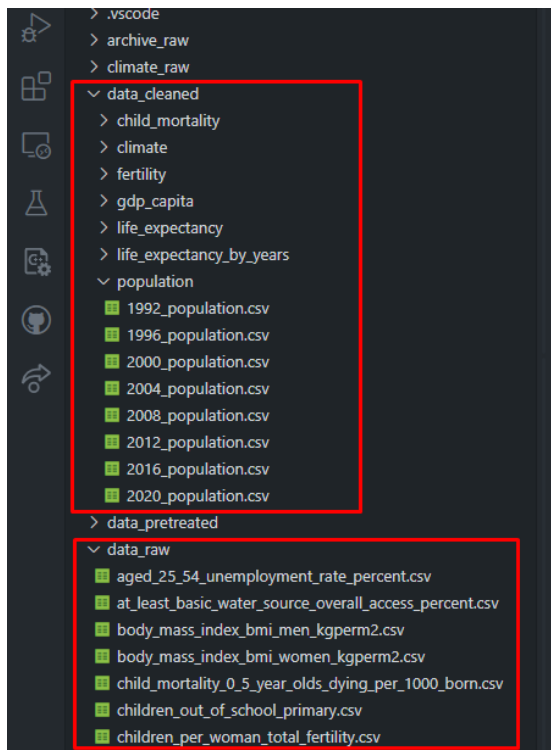
- Création d'un drive chez google
- Centralisation des données pour toute l'équipe
- Stockage au format .zip avec ULR du site web source (Kaggle...)

# Rédaction du cahier des charges

## Etablissement d'un diagramme de Gantt



# Polissage des données brutes



- Organisation du dépôt Git selon arborescence de fichiers
- Nous avons essentiellement 3 NB qui traitent les données:
  - 1 Création des fichiers .csv pour classements olympiques (top 10)
  - 2 Préparation des fichiers .csv des données d'entrées (gapminder, etc.)
    - Normalisation, imputations, etc.
  - 3 Visualisation des données
    - Recherche de tendances exploitables, corrélations, clustering, etc.
- Nous générons un fichier .csv par événement olympique

Köppen climate classification scheme symbols description

table<sup>[8][7][10]</sup>

1st	2nd	3rd
<b>A (Tropical)</b>	f (Rainforest)	
	m (Monsoon)	
	w (Savanna, dry winter)	
	s (Savanna, dry summer)	
<b>B (Dry)</b>	W (Arid desert)	h (Hot)
	S (Semi-arid steppe)	k (Cold)
<b>C (Temperate)</b>	w (Dry winter)	a (Hot summer)
	f (No dry season)	b (Warm summer)
	s (Dry summer)	c (Cold summer)
<b>D (Continental)</b>	w (Dry winter)	a (Hot summer)
	f (No dry season)	b (Warm summer)
	s (Dry summer)	c (Cold summer)
<b>E (Polar)</b>	T (Tundra)	
	F (Ice cap)	

## Recherche complémentaire: données climatique

- Agrégation de données climatiques (Köppen Geiger)
- *Hypothèse*: Possibilité de partir des données discrètes -> obtenir des données continues.
- Ajout pour les pays d'une contribution à chaque composante: temperature, précipitation, saisonnalité, aridité

	country	Avg_Temperature	Avg_Precipitation	Avg_Seasonality	Avg_Aridity
0	Afghanistan	0.446153846	0.407692308	0.338461538	0.49231
1	Albania	0.55	0.6	0.475	0.45
2	Algeria	0.5	0.36	0.24	0.64
3	Angola	0.516666667	0.483333333	0.333333333	0.5
4	Antarctica	0.1	0.1	0.1	0.1
5	Argentina	0.423076923	0.469230769	0.338461538	0.45385
6	Armenia	0.48	0.46	0.42	0.44
7	Australia	0.6	0.583333333	0.475	0.46667
8	Austria	0.375	0.35	0.325	0.35
9	Azerbaijan	0.5	0.5	0.425	0.4375
10	Bahamas	0.7	0.7	0.6	0.3



# Difficultés rencontrées

- Concordance: *délégations olympiques VS nations présentes dans les données d'entrées (gapminder) -> Edition manuelle.*
- Harmonisation de l'état des données et des versions des NB au sein de l'équipe.  
*-> création d'un notebook avec fonctions utilitaires à importer.*
- *Prévoir la quantité de temps à allouer à chaque étape. -> expérience...*



# Graphiques interactifs: visualisation des données







# Questions & Réponses

