

Processus Techniques

Tâches	progression	début	durée	fin
sprint 1 :				
Mise en place de l'environnement de travail		11.05.2023	1 Jours	12.05.2023
création du repository "trello"		11.05.2023	1 Jours	12.05.2023
Creation de l'env. de travail sur jupiter		11.05.2023	1 Jours	12.05.2023
Création du repository sur Github		11.05.2023	1 Jours	12.05.2023
Import du Notebook sur Github		11.05.2023	1 Jours	12.05.2023
Phase de tests et debugages		11.05.2023	1 Jours	12.05.2023
sprint 2 :				
Résolution du TP				
Résolution Techniques du TP (Sprint 2)		11.05.2023	1 Jours	12.05.2023
sprint 3 :				
Presentation du Projet		11.05.2023	1 Jours	12.05.2023
Elaboration du dossier Words		11.05.2023	1 Jours	12.05.2023

Etape 1 Mise en place de l'environnement de travail

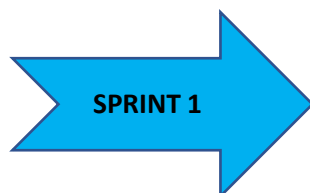
- Création de l'équipe
- Création d'un Tableau Trello
- Elaboration et répartition des tâches à faire
- Brainstorming général à propos des Sprints

Etape 2 Mise en place de l'environnement Anaconda

- Création d'un environnement de test en ligne de commande "TP_DataOps_BabyNames"
- Importation des modules **Jupyter, Numpy, Matplotlib.pyplot, et Pandas.**
- Création du notebook "TP_BabyNames"

Etape 3 Mise en place du repository Github.

- Création du repository
https://github.com/laylaTAH/Data_BabyNames.git
- Invitation des membres de l'équipe
- **Importation du Notebook** Jupyter BabyNames vers Github



SPRINT 2

Etape 1 Résolution du point Numéro 1 du TP

- Récupération du fichier ZIP demandé via le lien web <https://www.ssa.gov/oact/babynames/limits.html>
- Importation du dossier Names vers notre environnement de travail
- Conversion du fichier Yob1880.txt en fichier csv
- Importation des modules pandas, Numpy et Matplotlib sur le Notebook
- Import du dataframe `pd.read_csv("./names/yob1880.csv")`
- Renommage des colonnes `names1880 = pd.read_csv("./names/yob1880.csv", header=None, names=["name", "sex", "birth"])`

```
Entrée [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

Entrée [3]: pd.read_csv("./names/yob1880.csv")

Out[3]:
```

	name	sex	birth
	Mary	F	7065
0	Anna	F	2604
1	Emma	F	2003
2	Elizabeth	F	1939
3	Minnie	F	1746
4	Margaret	F	1578
...
1994	Woodie	M	5
1995	Worthy	M	5
1996	Wright	M	5
1997	York	M	5
1998	Zachariah	M	5

```
Entrée [6]: names1880 = pd.read_csv("./names/yob1880.csv", header=None, names=["name", "sex", "birth"])
names1880

Out[6]:
```

	name	sex	birth
0	Mary	F	7065
1	Anna	F	2604
2	Emma	F	2003
3	Elizabeth	F	1939
4	Minnie	F	1746
...
1995	Woodie	M	5
1996	Worthy	M	5
1997	Wright	M	5
1998	York	M	5
1999	Zachariah	M	5

2000 rows x 3 columns

Etape 2 Résolution du point Numéro 2 du TP

- Extraction du nombre de naissances par sexes : `births_by_sex = names1880.groupby('sex')['birth'].sum()` `print (births_by_sex)`

```
Entrée [7]: births_by_sex = names1880.groupby('sex')['birth'].sum()
print (births_by_sex)

sex
F    90994
M   110490
Name: birth, dtype: int64
```

Etape 3 Résolution des points Numéro 3 à 11

- nous nous sommes référés à la correction du professeur
-

Etape 1 Elaboration de la Conclusion du TP

- brainstorming sur l'ensemble des résultats



SPRINT 3

Etape 2 Elaboration du document Word de présentation du Projet

- brainstorming sur l'ensemble des points demandés
- capture d'écran des éléments importants
- vérification de la conformité de notre projet (tests, lien github ok, etc..)

Notre Conclusion

A la vue des résultats du dernier point du TP, on peut considérer que plus on avance dans le temps et plus les prénoms des Bébés sont diversifiés.