



DatArt " L'Art de faire parler la DATA"

SPRINT 1



Sommaire

- Présentation de l'équipe
- Logiciels utilisés
- Notre problématique
- Nos données

Présentation de l'équipe DatArt



Juliette VUONG
juliette.vng@gmail.com
Développeuse



Marine GALANTH
marine.galanth@gmail.com
Scrum Master (Développeuse)



Elodie LANTZ
elodie120399@gmail.com
Product Owner (Développeuse)



Emma GADEA
emma.gadea78@gmail.com
Développeuse

Nos outils



Discord



Jira



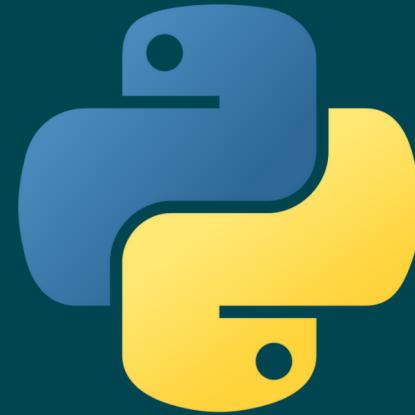
Google drive



Excel



SQL



Python

JIRA

Tableau Sprint 1



The Jira board displays the following status counts:

- A FAIRE: 0 tickets
- EN COURS 1 TICKET
- FINI 5 TICKETS ✓

Details for the EN COURS ticket:

Choix de nos indicateurs RSE
DS-21

Details for the FINI tickets:

- Attribution des rôles
ATTRIBUTION DES RÔLES
DS-15 ✓
- Slide équipe/rôles
CANVA
DS-16 ✓ E
- Slide outils
CANVA
DS-20 ✓ E
- Choix des données
DS-17 ✓
- Slide rétroplanning
DS-22 ✓

A search bar and filter buttons (Epic, E, +) are located at the top of the board.

Indicateurs et bdd

Méthode de recherche :



- Où et comment chercher une base de donnée : (mots clés, sites)

- Quels sont les différents indicateurs RSE, lequel veut on analyser?

bilan GES via ADEME

ID du Bilan GES	Type de structure	Type de collectivité	Raison sociale / Nom de l'entité	Année de reporting	Date de publication	Nombre de salariés / d'agents	Population
1	1 Entreprise		Mutuelle d'Ass	01/01/2011	04/03/2015	6848	
2	2 Entreprise		ADIDAS France	01/01/2011	05/03/2015	768	
3	3 Etablissement public		CROUS Nice-Tc	01/01/2011	05/03/2015	261	
4	4 Collectivité terri Communes		Fort de France	01/01/2010	25/03/2015	3000	94000
5	5 Collectivité terri Communes		concarneau Co	01/01/2011	05/03/2015	102	49000
6	6 Collectivité terri Communes		Communauté I	01/01/2012	05/03/2015	733	57490
7	7 Etablissement public		OPAC Saône et	01/01/2011	05/03/2015	557	
8	8 Entreprise		Système U Oue	01/01/2012	06/03/2015	2711	
9	9 Entreprise		Caisse d'Eparg	01/01/2011	06/03/2015	2732	
10	10 Collectivité terri Communauté d'ag		Communauté d'	01/01/2012	06/03/2015	210	123476
11	11 Etablissement public		ADEME (Agenc	01/01/2011	09/03/2015	963	
12	12 Etablissement public		Chambre des r	01/01/2011	09/03/2015	357	

Les bases de données

https://vcs.management-datasience.org/Management_Datascience/DB_SIREN

<https://www.data.gouv.fr/fr/reuses/combien-coute-1eur-de-ca-en-gco2/>



DatArt " L'Art de faire parler la DATA"

SPRINT 2



Sommaire

- Zoom sur Jira
- Nos jeux de données
- Extraction des données
- Nettoyage des données

Tableau Sprint 2

| A FAIRE 2 TICKETS | EN COURS 4 TICKETS | FINI 3 TICKETS ✓ |
| --- | --- | --- |
| Organisation des jeux de données DS-37 | Controle de la qualité des données DS-28 | Choix des indicateurs RSE DS-25 |
| Stockage sur Gitea DS-38 | Trouver un nom/problématique pour notre rapport de données DS-33 | Recherche de la nouvelle base de donnée (quels critères) DS-23 |
| | Extraction des données DS-35 | COLLECTE DE DONNÉES DS-27 |
| | Nettoyage des données DS-36 | COLLECTE DE DONNÉES DS-27 |

Sprint 2

Ce qui a été fait

Risque environnemental: L'entreprise se situe t-elle dans un secteur à risque environnemental ?

- **Création du Gitea**
- **Identifier les sources de données sur les indicateurs RSE:**
 - Jeu de donnée sur le zonage inondation en France 2020
 - Jeu de donnée adresse des entreprises avec le code siren
- **Sélectionner les données les plus pertinentes:**
 - *te_siren_adresse*: siren,complementAdresseEtablissement,codePostalEtablissement,codeCommuneEtablissement
 - *n_enjeu_rapport*: scenario (échelle 1-4), id(nom de la commune), hab_perm(surfaces inondable)
- **Nettoyer et transformer les données** (commencement)

Nos difficultés

Nettoyage des données avec python
Comprendre les indicateurs sans "traduction"

Ce qui reste à faire

Nettoyer et transformer les données (à poursuivre)
Se familiariser avec Gitea et Python

Extraction des données

Jeux de données des entreprises

https://vcs.management-datasience.org/Management_Datasience/DB_SIREN/src/branch/master/Tables_indicateurs

Code postal

Jeux de données Bilan GES

<https://www.georisques.gouv.fr/donnees/bases-de-donnees/inondations-par-remontee-de-nappes>

<https://www.georisques.gouv.fr/donnees/bases-de-donnees/base-de-donnees-mouvements-de-terrain>

Nettoyage des données avec Python

Apprentissage de Python :

https://www.youtube.com/watch?v=x8yu8sq8mdw&ab_channel=MachineLearnia

https://www.youtube.com/watch?v=efSjcrp870Y&ab_channel=PrettyPrinted



DatArt " L'Art de faire parler la DATA"

Merci pour votre attention

Next step :

- Organisation des données
- Stockage sur Gitea
- Elaboration de l'introduction/conceptualisation de notre rapport





DatArt " L'Art de faire parler la DATA"

SPRINT 3

Tableau Sprint 3



A FAIRE 2 TICKETS

Trouver un nom/problématique pour notre rapport de données

DS-33

Fusion de la base 1 et de la base 1.2

DS-42

EN COURS 2 TICKETS

Contrôle de la qualité des données

COLLECTE DE DONNÉES

DS-28

Extraction des données

DS-35

FINI 5 TICKETS ✓

Organisation des jeux de données

DS-37

Nettoyage du jeux 1

DS-39

Nettoyage du jeux 1.2

DS-40

Nettoyage jeux 2

DS-41

Stockage sur Gitea

DS-38



```
Entrée [15]: print(df.head())
```

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import csv
df = pd.read_csv("te_siren_adresse.csv")
print(df.head())
    siren  nic      siret complementAdresseEtablissement \
0  807626460  24  80762646000024                      LA TOUCHE
1  807639737  12  80763973700012                      NaN
2  807651666  16  80765166600016                      NaN
3  807658679  20  80765867900020          HAMEAU DE SERRIERES
4  807663893  38  80766389300038                      ZI BEL AIR
    numeroVoieEtablissement indiceRepetitionEtablissement \
0                NaN                    NaN
1                4.0                   B
2                3.0                    NaN
3               11.0                    NaN
4                9.0                    NaN
    typeVoieEtablissement libelleVoieEtablissement codePostalEtablissement \
0            RTE             DE LINERS           16200.0
1            NaN             GRANDE RUE           91600.0
2            RUE             ANDRE HUYS           59100.0
3            RUE             DU MOULIN           1230.0
4            RUE             DE L'ETANG           97450.0
    libelleCommuneEtablissement libelleCommuneEtrangerEtablissement \
0        BOURG-CHARENTE                      NaN
1      SAVIGNY-SUR-ORGE                      NaN
2        ROUBAIX                           NaN
3  SAINT-RAMBERT-EN-BUGEY                      NaN
4      SAINT-LOUIS                           NaN
    distributionSpecialeEtablissement codeCommuneEtablissement \
0                  NaN                 16056
1                  NaN                 91589
2                  NaN                 59512
3                  NaN                 01384
4                  NaN                 97414
    codeCedexEtablissement libelleCedexEtablissement \
0                  NaN                      NaN
```

```
         siren  nic      siret codePostalEtablissement \
0  807626460  24  80762646000024           16200.0
1  807639737  12  80763973700012           91600.0
2  807651666  16  80765166600016           59100.0
3  807658679  20  80765867900020           1230.0
4  807663893  38  80766389300038           97450.0
    libelleCommuneEtablissement codeCommuneEtablissement \
0        BOURG-CHARENTE                     16056
1      SAVIGNY-SUR-ORGE                     91589
2        ROUBAIX                         59512
3  SAINT-RAMBERT-EN-BUGEY                     01384
4      SAINT-LOUIS                        97414
```

```
Entrée [ ]:
```

```
Entrée [4]: print(df['distributionSpecialeEtablissement'])
```

```
0      NaN
1      NaN
2      NaN
3      NaN
4      NaN
...
1095676   NaN
1095677   NaN
1095678   NaN
1095679   NaN
1095680   NaN
Name: distributionSpecialeEtablissement, Length: 1095681, dtype: object
```

```
Entrée [5]: print(df['distributionSpecialeEtablissement'].isnull())
```

```
0      True
1      True
2      True
3      True
4      True
...
1095676   True
1095677   True
1095678   True
1095679   True
1095680   True
Name: distributionSpecialeEtablissement, Length: 1095681, dtype: bool
```

Sprint 3

Ce qui a été fait

- S'imprégner de notre sujet
- Vérification de la fiabilité des données
- Exportation des données
- Nettoyage des données
- Début d'architecture de la base de données

Nos difficultés

- Commande visualisation des données
- Déposer les données sur gitea
- Assurer la fiabilité des données



DatArt " L'Art de faire parler la DATA"

Merci pour votre attention

Next step :

- Fusion de deux jeux de données
- Conception du début du rapport
- Conception de graphiques
- Architecture des données





DatArt " L'Art de faire parler la DATA"

SPRINT FINAL - RETROSPECTIVE

LANTZ Elodie - GADEA Emma - GALANTH Marine - VUONG Juliette



DatArt " L'Art de faire parler la DATA"

Merci pour votre attention

