# Projet

# Statistiques de la démographie des professionnels de santé

# Spécifications

SCN SIM des ARS – DNUM

## Alimentation des données

Le repository du script d’ETL est accessible sur Github : <https://github.com/dnum-sante-ars-data/demographie_professionnels_sante>

### Fournisseur et fréquence de rafraîchissement

|  |  |
| --- | --- |
| **Fournisseur** | Agence du Numérique en Santé via le projet DIAMANT |
| **Fréquence de rafraichissement** | Hebdomadaire, cible : quotidienne |

### Modèle des données sources

*A renseigner pour chaque fichier (dossier « donnees\_sources » :*

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Format | Longueur | Contraintes |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Les données sont mises à disposition sur le **serveur FTP Atlasanté** :

"username" : "sftp\_rdvvaccins\_ars",

"password" : "\*\*\*\*\*",

"host" : [www.atlasante.fr](http://www.atlasante.fr)

Les fichiers sont au format **.csv** et sont situés dans le répertoire **/demographie\_ps**.

### Stack

* Python 3.9
* argparse
* logging
* datetime
* pysqlite3
* json
* wget
* ftplib
* gnupg
* gzip

### Architecture du script

\_documentation

\_etl

|\_\_ \_utilitaire

|\_\_ \_data

|\_\_ \_database

|\_\_ \_input

|\_\_ \_output

|\_\_ \_log

|\_\_ main.py

|\_\_ \_modules

|\_\_ \_private\_transform

|\_\_ \_public\_transform

|\_\_ \_route\_sftp

|\_\_ \_route\_sqlite

|\_\_ \_\_init\_\_.py

|\_\_ sqlite\_config.py

|\_\_ \_route\_webservice

|\_\_ \_sql\_queries

|\_\_ \_settings

|\_\_ settings.json

|\_\_ \_utils

### Téléchargement des fichiers sources

|  |  |
| --- | --- |
| **Fonctionnalité** | Téléchargement des fichiers sources du serveur FTP [www.atlasanté.fr](http://www.atlasanté.fr) /demographie\_ps en local dans /data/input |
| **Commande** | import |
| **Fonction(s) dans main.py** | Pas en encore développé |
| **Fonction(s) dans /route\_sftp/route\_sftp.py** | Pas en encore développé |
| **Module(s) spécifiques** | wget |
| **Remarque(s)** |  |

### Déchiffrement des fichiers sources

Cf <https://medium.com/@almirx101/pgp-key-pair-generation-and-encryption-and-decryption-examples-in-python-3-a72f56477c22>

|  |  |
| --- | --- |
| **Fonctionnalité** | * Déchiffrement de l’archive * Décompression de l’archive |
| **Commande** | decrypt |
| **Fonction(s) dans main.py** | Pas en encore développé |
| **Fonction(s) dans /pgp\_decrypt/pgp\_decrypt.py** | Pas en encore développé |
| **Module(s) spécifiques** | Génération des clés de chiffrement / déchiffrement :   * Sur windows, Kleopatra * Sur Mac / linux : GnuPG et GPA   Modules python :   * Déchiffrement : python-gnupg * Décompression : gzip |
| **Remarque(s)** |  |

### Dump des données

|  |  |
| --- | --- |
| **Fonctionnalité** | Sauvegarde des données des fichiers sources de /data/input et /utils dans une base de données locale :   * création de la BDD * création du schéma des tables de dump : 1 table par fichier source * import des données des fichiers sources dans les tables de la BDD |
| **Commande** | dump |
| **Fonction(s) dans main.py** | Pas en encore développé |
| **Fonction(s) dans /route\_sqlite/route\_sqlite.py** | Pas en encore développé |
| **Module(s) spécifiques** | sqlite |
| **Remarque(s)** |  |

### Modèle des données cibles pour les visualisations

*A renseigner pour chaque fichier (dossier « donnees\_cibles » :*

*A remplir via demographie\_ps\_bascule\_v203\_V300\_mcd\_20210111*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Format | Longueur | Contraintes | Table source d’origine |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

### Transformation des données

|  |  |
| --- | --- |
| **Fonctionnalité** | * Construction de la table activites : croisements des tables :   A lister *via demographie\_ps\_bascule\_v203\_V300\_mcd\_20210111*  Avec les tables INSEE communes / départements / regions   * Construction de la table personnes :   A lister *via demographie\_ps\_bascule\_v203\_V300\_mcd\_20210111*  Avec les tables INSEE communes / départements / regions   * Extractions des fichiers cibles dans /output :   + activites.csv   + personnes.csv |
| **Commande** | transform |
| **Fonction(s) dans main.py** | Pas en encore développé |
| **Fonction(s) dans /transform\_private /transform\_private.py** | Pas en encore développé |
| **Module(s) spécifiques** |  |
| **Remarque(s)** |  |

### Publication des fichiers cibles

|  |  |
| --- | --- |
| **Fonctionnalité** | Publication des fichiers activites.csv et personnes .csv depuis /output vers le serveur FTP www.atlasante.fr |
| **Commande** | publish |
| **Fonction(s) dans main.py** | Pas en encore développé |
| **Fonction(s) dans /route\_sftp /route\_sftp.py** | Pas en encore développé |
| **Module(s) spécifiques** | ftplib |
| **Remarque(s)** |  |

## Visualisation des données

L’application OpenDataSoft (ODS) est utilisée pour mettre à disposition des utilisateurs les visualisations de données : <https://bce-sgsocialgouv.opendatasoft.com/login/?next=/backoffice/catalog/datasets/>

Cible : domaine dédié

### Accès libre et restreint

Certains champs des jeux de données sont en accès restreint et d’autres en accès public.

### Jeux de données

2 jeux de données

#### Activités :

|  |  |  |  |
| --- | --- | --- | --- |
| Champ (ODS) | Facette (ODS) | Variable correspondante (fichier .csv) | Type d’acès |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

#### Personnes :

|  |  |  |  |
| --- | --- | --- | --- |
| Champ (ODS) | Facette (ODS) | Variable correspondante (fichier .csv) | Type d’acès |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

### Représentations graphiques

Construire les maquettes en fonction du contenu du dossier « fichiers\_produits\_actuellement »