Structure de la base de données Datan

Le document pour voir la structure de la base en format SQL se trouve dans le dossier : <https://github.com/datanfr/datan/tree/master/scripts/update_dataset/infos>

# amendements

Cette table récupère de l’open data de l’Assemblée nationale tous les amendements. Toutes les variables sont celles présentes dans l’open data.

# amendements\_auteurs

Cette table récupère de l’open data de l’Assemblée nationale tous les auteurs des amendements. Le type d’acteur peut soit être un député ou un gouvernement.

# candidate\_full

Cette table est une vue. Il s’agit d’avoir une table regroupant toutes les informations concernant chaque député qui est candidat à une élection. Ces données sont mises à jour à la main par l’équipe de Datan.

# catégories

Il s’agit de la liste des catégories des articles de blog (actualité politique, datan, et rapports).

# circos

Cette table recense toutes les villes et la circonscription à laquelle elles appartiennent. Les villes découpées en plusieurs circonscriptions (grande villes) apparaissent plusieurs fois dans la table. La table note également le code insee de la commune.

# cities\_adjacentes

Cette table recense toutes les villes adjacentes à d’autres villes. Dans cette table, chaque ville compte plusieurs lignes : une par ville qui la touche.

# cities\_infos

Cette table recense les infos de population des villes. Chaque ville a une ligne. Dans les colonnes se trouve le nombre d’habitants pour plusieurs années.

# cities\_mayors

Dans cette table se trouve le maire de chaque commune française.

# class\_groups

Dans cette groupe se trouvent plusieurs statistiques pour tous les groupes politiques de l’Assemblée nationale. Ces statistiques sont développées par datan. On y retrouve le taux de cohésion au groupe, le taux de proximité avec la majorité, le taux de participation à tous les votes, le taux de participation aux votes solennels, le taux de participation aux votes par spécialisation. Dans cette table, chaque groupe a plusieurs lignes, une par statistique. La statistique se trouve dans la colonne ‘value’. La colonne ‘votes’ donne le nombre de scrutins sur lesquels la statistique est calculée. La colonne ‘active’ prend la valeur 1 si le groupe est encore en activité.

# class\_groups\_proximite

Cette table donne la proximité (de vote) entre chaque groupe politique. Chaque groupe a plusieurs lignes, une avec tous les groupes pour lesquels nous calculons un score de proximité. Le score se trouve dans la colonne ‘score’. La variable ‘votesN’ est le nombre de scrutins sur lesquels la statistique est calculée. Le score va de 0 à 1. 1 signifie que deux groupes ont eu la même position sur tous les scrutins.

# class\_loyaute

Cette table donne le score de loyauté pour chaque député. Le score de loyauté est la proximité entre un député et son propre groupe politique. Si un député a été présent à plusieurs législatures, il sera présent plusieurs fois dans cette table, il a un score de loyauté par législature.

# class\_loyaute\_six

Cette table donne le score de loyauté pour chaque député. Ce score est basé sur les scrutins des 12 derniers mois.

# class\_majorite

Cette table donne le score de proximité entre chaque député et le groupe majoritaire à l’Assemblée nationale (groupe de la majorité présidentielle).

# class\_participation

Cette table donne le score de participation pour chaque député et chaque législature. Il s’agit ici du score de participation à tous les scrutins de l’Assemblée nationale qui ont eu lieu pendant la durée du mandat du député. La variable ‘votesN’ donne le nombre de scrutins sur lesquels le score est calculé. Le score de participation va de 0 à 1. Si le score est 1, cela veut dire que le député a participé à tous les scrutins.

# class\_participation\_commission

Cette table donne le score de participation par spécialité pour chaque député et chaque législature. Afin de mieux prendre en compte les contraintes liées à l’organisation de l’agenda à l’Assemblée nationale, nous avons développé cette statistique, qui ne prend en compte, pour chaque député, que les votes sur les textes liés à son domaine de spécialisation. Le domaine de spécialisation est opérationnalisée comme l’appartenance à une commission parlementaire (économie, affaires étrangères, etc).

# class\_participation\_six

Il s’agit du score de participation (tous les scrutins) mais uniquement basé sur les scrutins qui ont eu lieu ces 12 derniers mois.

# class\_participation\_solennels

Cette table donne le score de participation aux scrutins solennels pour chaque député et chaque législature.

# departement

Cette table donne des informations (code du département, slug utilisé sur Datan, libellé pour les textes automatique sur Datan) concernant les départements. Il y a une ligne par département

# deputes

Cette table est le résultat de l’export des fichiers open data de l’Assemblée nationale. Il y a dans cette table des informations pour chaque député (nom, prénom, genre, date de naissance, pays de naissance, ancien métier, catégorie socio-professionnelle).

# deputes\_accord

Cette table est la table qui sert à calculer ensuite le taux de proximité pour chaque député. Chaque député, sur chaque vote, a un score d’accord avec tous les groupes (1 si accord, sinon 0).

# deputes\_accord\_cleaned

Cette table donne le taux de proximité entre chaque député et chaque groupe de l’Assemblée nationale. Chaque député a plusieurs lignes : une par groupe politique. Le score va de 0 à 100. 100 veut dire que le député et le groupe ont voté sur la même ligne dans 100 pourcent des cas.

# deputes\_all

Quand sur Datan nous voulons des informations biographies pour les députés (nom, prénom, genre, département d’élection, etc.), nous les récupérons sur deux tables, deputes\_all et deputes\_last. Ces deux tables sont un condensé de toute l’information personnelle sur chaque député, afin d’éviter de faire trop de jointure de tables. La table deputes\_all a plusieurs lignes par députés : une par législature.

# deputes\_last

La table deputes\_last est similaire à la précédente mais elle ne compte qu’une ligne par député : son dernier mandat.

# deputes\_contacts

Cette table est un export de l’open data de l’Assemblée nationale. On y trouve les infos de contact des parlementaires (site internet, mail, twitter, facebook).

# deputes\_loyaute

Cette table donne le score de loyauté d’un député à son groupe politique. Contrairement à la table class\_loyaute, cette table a plusieurs lignes par député : une par groupe. Cela permet d’avoir tous les scores de loyauté des députés ayant appartenu à plusieurs groupes politiques.

# documents\_legislatifs

Cette table est un export de l’open data de l’AN. Il s’agit de tous les documents législatifs.

# dossiers

Cette table est un export de l’open data de l’AN. Il s’agit de tous les dossiers législatifs.

# dossiers\_acteurs

Cette table est un export de l’open data. On retrouve dedans les initiateurs ou rapporteurs des projets et proposition de loi.

# elect\_2019\_europe

Cette table donne les résultats aux élections européennes de 2019 par commune. Dans les colonnes se trouvent les informations concernant les différentes listes.

# elect\_2019\_europe\_clean

Cette table est la même que celle du dessus mais en format long (la précédente est en format wide). On retrouve pour chaque ligne le pourcentage de voix (‘value’) pour chaque liste dans chaque commune.

# elect\_2019\_europe\_listes

Cette table donne la liste des 34 listes pour l’élection européenne en France.

# elect\_deputes\_candidats

Cette table donne les députés candidats à une élection. Ces informations sont mises à jour manuellement par l’équipe de Datan. Dans cette table, il peut y avoir plusieurs lignes par député si celui-ci a été candidat à plusieurs élections lors de son mandat de député.

# elect\_legislatives\_cities

Cette table donne les résultats aux élections législatives par commune. Il y a plusieurs lignes par commune : une élection (2017 et 2022). Nous avons ici uniquement les résultats pour le second tour.

# elect\_legislatives\_infos

Cette table est identique à la table précédente, mais elle ne donne pas les résultats par candidat mais des informations générales (nombre de votants, abstention).

# elect\_legislatives\_results

Cette table donne les résultats globaux par circonscription. Chaque circo a plusieurs lignes : une par élection et ensuite une par candidat. Il y a ici les résultats pour le 1er tour et pour le 2nd tour.

# elect\_libelle

Cette table donne les élections qui sont traitées sur datan : <https://datan.fr/elections>

# elect\_pres\_2

Cette table donne les résultats au second tour des élections présidentielles de 2017 et 2022. Il y a plusieurs lignes par commune : une par élection et une par candidat.

# Famsocpro

Cette table donne les 8 familles socio-professionnelles de l’INSEE ainsi que le pourcentage de la population dans chaque catégorie.

# faq\_categories

Cette table donne les catégories de la FAQ de Datan.

# faq\_posts

Dans cette table se trouvent les entrées de la FAQ de Datan.

# fields

Cette table liste les différentes catégories (policy fields) utilisés sur Datan pour catégoriser les scrutins : culture, éducation, économie, environnement, etc.

# groupes\_accord

Cette table donne l’accord entre chaque groupe de l’Assemblée et pour chaque vote. Chaque groupe a plusieurs lignes : une par vote et une pour tous les autres groupes de l’Assemblée. La variable ‘accord’ prend la valeur 1 si les deux groupes ont eu la même position sur ce vote.

# groupes\_cohesion

Cette table donne le taux de cohésion pour chaque groupe et chaque scrutin. Il y a aussi d’autres informations : la position du groupe (1 si pour, 0 si abstention, et -1 si contre), le sort du vote (1 si adopté).

# groupes\_effectif

Cette table donne le nombre de députés membres des différents groupes politiques de l’Assemblée. Il s’agit uniquement des groupes toujours en activité, de la législature actuelle.

# groupes\_stats

Cette table donne, pour chaque groupe, quelques informations : le pourcentage de députées femmes, le nombre de députées femmes, l’âge moyen, et l’index ROSE (index de représentativité sociale).

# hatvp

Cette table récupère les données de la Haute autorité pour la transparence de la vie publique.

# history\_mps\_average

Cette table donne, pour chaque législature, la durée moyenne du temps passé à l’Assemblée par les députés en exercice.

# history\_per\_mps\_average

Cette table donne, pour chaque député, la durée moyenne de leur temps passé à l’Assemblée.

# insee

Cette table donne la correspondance, pour chaque commune, entre le code INSEE et le code postal.

# legislature

Cette table donne des informations sur les législatures de l’Assemblée nationale.

# mandat\_groupe

Cette table est un export de l’open data. Il s’agit des mandats de chaque député. Nous avons dans cette table uniquement les mandats liés au groupe politique.

# mandat\_principal

Cette table est un export de l’open data. Il s’agit des mandats de chaque député. Dans cette table nous avons uniquement les mandats principaux (mandat de député).

# mandat\_secondaire

Dans cette table nous avons les mandats secondaires des députés (appartenance à une commission parlementaire, à un groupe d’étude, etc).

# newsletter

Abonnés à la newsletter.

# organes

Cette table est un export de l’open data de l’Assemblée. Il s’agit de tous les organes (commission parlementaire, groupe politique, groupe d’étude, etc).

# parrainages

Cette table donne tous les parrainages des députés à un candidat à l’élection présidentielle. Ce sont les données pour 2022.

# parties

Cette table donne les partis politiques. La variable ‘effectif Total’ donne le nombre de députés actuellement rattaché financièrement au parti.

# password\_resets

Cette table est utilisée pour les mots de passe oubliés.

# posts

Dans cette table se trouvent les articles de blog sur le site DATAN.

# quizz

Dans cette table se trouvent les scrutins utilisés pour le quiz de Datan (voting aid application).

# readings

Cette table liste les différentes lectures possibles pour un texte : première lecture, deuxième lecture, lecture définitive, et texte de la commission mixte paritaire.

# regions

Cette table liste les régions françaises.

# sujets

Table non utilisée.

# table\_history

Cette table, pour le moment, liste les modifications apportées par les députés à certaines de leurs informations (information de candidature à une élection).

# users

Utilisateurs sur Datan.

# users\_mp

Utilisateurs députés sur Datan.

# users\_mp\_link

Table utilisée quand un député créé un compte sur Datan.

# votes

Cette table est un export de l’open data de l’Assemblée. Elle donne les informations concernant les positions de vote des députés. Chaque député a une ligne par scrutin quand il était présent. La position du député est dans la variable ‘vote’.

# votes\_datan

Cette table liste les votes qui sont décryptés et contextualisés par Datan.

# votes\_datan\_requested

Cette table liste les votes que les utilisateurs aimeraient voir décryptés.

# votes\_dossiers

Cette table fait la jointure entre les votes et les dossiers. Il s’agit du résultat d’un scrapping du site de l’Assemblée nationale.

# votes\_groupes

Cette table donne les positions de vote des groupes politique pour tous les scrutins de l’Assemblée. La position du groupe est dans la colonne ‘positionMajoritaire’.

# votes\_info

Cette table donne toutes les informations pour chaque scrutin. C’est un export de l’open data de l’Assemblée.

# votes\_participation

Cette table permet de calculer les taux de participation. Chaque député a une ligne dans cette table pour chaque scrutin qui a eu lieu au cours de son mandat. S’il a voté, la valeur de ‘participation’ est 1, sinon 0 s’il était absent.

# votes\_participation\_commission

Cette table est identique à la table précédente mais sert à calculer le score de participation par spécialisation (basée sur l’appartenance à une commission parlementaire). Chaque député a une ligne dans cette table pour chaque scrutin qui a eu lieu au cours de son mandat et qui a été préalablement discuté au sein de sa commission parlementaire. S’il a voté, la valeur de ‘participation’ est 1, sinon 0 s’il était absent.

# votes\_scores

Cette table sert à calculer les statistiques de Datan. Chaque député a une ligne par scrutin. Le vote du député est dans la variable ‘vote’. Le sort du code est dans ‘sortCode’ (si le vote est adopté ou rejeté). La position du groupe du député est dans ‘positionGroupe’. La position du gouvernement est dans ‘positionGvt’.