TANG Kévin p1501263

**TP1 Nettoyage de données**

Connexion

psql -h bd-pedago.univ-lyon1.fr -U p1501263

mdp : fMxSpNIkjneR

**CREATE** **TABLE** s2016 (

"Programme 2016" DECIMAL **NOT** **NULL**,

"SIREN" VARCHAR **NOT** **NULL**,

"NIC" VARCHAR,

"Dénomination" VARCHAR **NOT** **NULL**,

"Montant" DECIMAL **NOT** **NULL**,

"Objet" VARCHAR **NOT** **NULL**,

"Parlementaire" VARCHAR,

"Réserve 2016" VARCHAR,

"Convention 2016" VARCHAR,

"COG : code département" VARCHAR,

"COG : code commune" DECIMAL,

"COG : ville ou pays" VARCHAR **NOT** **NULL**,

"Nomenclature juridique" DECIMAL **NOT** **NULL**,

"Code NAF" VARCHAR **NOT** **NULL**,

"Situation SIRENE" VARCHAR,

"RNA" VARCHAR

);

* SIREN :

Certains numéros SIREN sont manquants, ils sont donc indiqués comme ‘N/C’ ou ‘Indéterminé’. On va les modifier pour mettre NULL à la place.

ALTER TABLE s2016

ALTER COLUMN "SIREN" DROP NOT NULL;

UPDATE s2016

SET "SIREN" = NULL

WHERE NOT ("SIREN" ~ E'^\\d+$');

ALTER TABLE s2016

ALTER COLUMN "SIREN" TYPE NUMERIC USING "SIREN"::NUMERIC;

* NIC :

Une ligne contenait “00 00”.

UPDATE s2016

SET "NIC" = NULL

WHERE "NIC" ~ E'[^0-9]+';

ALTER TABLE s2016

ALTER COLUMN “NIC” TYPE NUMERIC USING “NIC”::NUMERIC;

* Réserve 2016 :

UPDATE s2016 SET "Réserve 2016" = True where "Réserve 2016" is not null;

UPDATE s2016 SET "Réserve 2016" = False where "Réserve 2016" is null;

ALTER TABLE s2016

ALTER COLUMN "Réserve 2016" TYPE BOOLEAN USING "Réserve 2016"::BOOLEAN;

***Expliquer pourquoi, du point de vue métier, le montant de la subvention n’est cependant pas un attribut qui devrait faire partie de la clé.***

Pour une même entreprise (même dénomination, SIREN, COG…), ce montant peut varier selon la subvention demandée.

***Vérifier la satisfaction de ces dépendances.***

Le numéro de SIREN devrait être propre à chaque entreprise.

select distinct "SIREN", "Dénomination" from s2016 ORDER BY "Dénomination";

Or, avec cette commande, on voit que ce n’est pas le cas. Plusieurs entreprises ont le même numéro SIREN ou plusieurs numéros SIREN sont affectés à la même entreprise.

On peut dire que les colonnes clés devraient être SIREN et NIC, l’association des deux éléments forme le SIRET qui est l’identifiant unique d’une entreprise.

Entreprise : idEntreprise, SIREN, NIC, Dénomination, codeNAF, Situation SIRENE, RNA

Subvention : idEntreprise, Montant, Objet,

Lieu : idEntreprise, codeCommune, codeDépartement, VilleOuPays

Parlementaire : nom

Réserve 2016 ? Convention 2016 ? Nomenclature juridique ?

-- SIREN, NIC -> denomination, dep, com, ville\_ou\_pays, nom\_jur, naf, situation\_SIRENE, RNA, NAF, nomenclature\_juridique

-- df précedente pas tout à fait vraie, id à rajouter ? corriger naf et nomenclature.

-- dep, com -> ville\_ou\_pays

-- SIREN, NIC -> liste de parlementaires (pas en 1FN) id à rajouter ?

-- SIREN, NIC, annee, programme, numero? -> objet, montant, reserve, convention

