

**Etude sur les résultats des**

**Élections Présidentielles Françaises**

**1965-2017**

**Daria Antipova - Simplon**

**Formation Développeur Data 2021**

Développement d’une base de données, exploitation des

informations, partage des résultats

**Formateurs**: Azria David

Boumaiza Manel

Tobelem Josselin

**Simplon.co - 26 Rue des Amandiers, 92000 Nanterre**

# Table des matières

[**Table des matières**](#_qyx3odr8nwir)2

[**1.Présentation du projet & Contexte**](#_d3pz23v409kp)3

[Objectif](#_oyn664kz9rav) 4

[Stack technique](#_x298waj5c45n) 5

[**2. Construction de la base de données**](#_p6f9h5vif3p)6

[Sources des données](#_9cw0d87jik8b) 6

[Description des données](#_lndxhddf2gdp) 7

[Création du modèle](#_15qvsrpfygt3) 8

[Modèle physique des données](#_a9k27amcdevh) 11

[Nettoyer et valider les données](#_wnqk72moydc7) 11

[Insertion de données](#_szjkmp89uvkd) 12

[Sécuriser la base de données](#_z7t2qe8g2rc3) 13

[**3. Analyse de données et visualisations**](#_n3zo4e9shob4)13

[Exemple d’une requête SQL](#_mvs13025ea5v) 14

[La liste totale de tous les candidats](#_uxnzqa2oagh0) 14

[Optimisation de la BDD](#_qwoce5k8fiv7) 14

[Visualisation des données](#_ftshdsp5z6e6) 15

[La liste des gagnants (Président) de chaque élection](#_si7shk85ke77) 16

[Quels départements soutiennent le président le plus fort](#_377ygihu94p6) 17

[Les résultats des dernières élections en 2017 dans les 20 départements avec le plus haut niveau de participation - suffrages par candidats / le 1er tour, le 2ème tour](#_xnntklm1la81) 19

[**4. Partage des résultats**](#_nbtx95orbdge)20

[Power BI](#_bf35krgtev47) 20

[**5. Planification et Gestion du projet**](#_gxah14x6ho75)21

[Trello](#_4ioybv4x5i78) 21

[**6. Remerciements**](#_v7pkzwhy8gy3)22

[**7. Bibliographie**](#_l978228ljxjf)23

# 1.Présentation du projet & Contexte

En France, l’élection présidentielle détermine la personne qui sera élue président de la République pour un mandat de cinq ans. Depuis 1965, l’élection du président se fait au suffrage universel direct.

La Constitution du 4 octobre 1958 prévoit initialement une élection pour une durée de sept ans par un collège électoral comprenant les membres du Parlement, des conseils généraux et des assemblées des territoires d’outre-mer, ainsi que les représentants élus des conseils municipaux. Cette disposition est modifiée en 1962 après référendum, ainsi l’élection de 1965 est la première au suffrage universel direct requérant la majorité absolue des suffrages exprimés.

La durée du mandat est réduite en 2000 à cinq ans lors du référendum sur le quinquennat présidentiel. Jusqu’à la révision constitutionnelle du 23 juillet 2008, il n’y avait pas de limite au nombre de mandats que pouvait effectuer un président de la République. L’article 6 de la Constitution stipule désormais que le mandat ne peut être renouvelé qu’une fois consécutivement.

En 2011, l’âge minimum pour se présenter est abaissé de 23 à 18 ans pour les élections législatives et l’élection présidentielle.

Onze élections ont eu lieu depuis 1958, dont dix au suffrage universel, et huit présidents différents ont été élus.

Pour moi comme pour les étrangers qui habitent en France c’est important d’ avoir une meilleure compréhension du système des élections françaises et la construction de la société, l’activité de la population dans la vie politique, la popularité des présidents français, la fréquence des changements du chef de la République.

Le domaine d’application de ces enjeux dans mon projet va être la France par départements et en total.

La récupération des données électorales est mise en œuvre pour la période de 1965 à 2017

## Objectif

Mon objectif via ce projet est de travailler sur un cas concret avec des données réelles.

L’objectif du projet c'est la création d'un outil interactif (dashboard) très facile en pratique pour voir et mieux comprendre l'activité sociale au domaine politique. Ce dashboard ou page d'internet doit partager des données des suffrages sur toutes les élections présidentielles depuis 1965.

C’est une volonté de ma part d’appliquer mes connaissances vues durant ma formation à un cas pratique réel.

Mes missions seront donc les suivantes:

- regrouper les données sociopolitiques

- trouver les données manquantes

- nettoyer les données pour assurer une bonne cohérence

- centraliser ces informations dans une base de données créée à cet effet

- développer une partie dashboard affichant les résultats de toutes les élections présidentielles depuis 1965

- développer une interface web pour l’utilisateur

Domaine d’application: La France par départements.

Mes principaux indicateurs dans mon travail sont les suffrages: inscrits, votants, exprimés, blancs et nuls.

## Stack technique

**Tableau: Les outils techniques utilisées**

|  |  |
| --- | --- |
| **Outil** | **Utilisation / Raisons** |
| Python | Language orienté objet pour traiter la Data et développer des applications |
| Jupyter Notebook | Application web pour exécuter du code : sert aux chargement, exploration, traitement et nettoyage des données  . faciliter l'exécution de petit script  . pratique pour réaliser de l’exploration de données |
| MySQL | SGBDR : Système de Gestion de Bases de Données Relationnelles, qui stocke des  données de façon organisée et cohérente |
| Looping | un logiciel de modélisation conceptuelle de données entièrement gratuit et libre d'utilisation |
| MySQL Workbench 8.0 | IDE SQL pour la conception de la base de donnée, la visualisation et administration de la base de données |
| Power BI | Visualisation des données (Data Visualisation) |
| Trello | Gestion de projet, permet de visualiser les tâches selon la méthode Agile  . gratuit  . user friendly |
| Github | Service web d'hébergement et de gestion de  développement de logiciels  . un commit avec description par nouvelle fonctionnalité développées  . sauvegarde du code |
| Read.me | Fichier contenant des informations sur les autres fichiers du même projet  . facilite la compréhension dans la manière qu’a été développé le projet  . préciser comment exécuter les programmes |
| Flask | Framework open-source de développement web en Python |

# 2. Construction de la base de données

## Sources des données

Le “Centre de données socio-politiques SciencesPo” propose à la communauté scientifique des services autour des données des sciences sociales. Il coordonne et participe à des projets majeurs dans ce domaine. CDSP est unité mixte de services de Sciences Po et du Centre national de la recherche scientifique. J’ai trouvé ici tous les résultats des élections présidentielles entre 1965 et 2012: 1er et 2ème tours par circonscriptions en format csv. <https://cdsp.sciences-po.fr/fr/ressources-en-ligne/ressource/cdsp_presi2012t1_circ/>

J’ai récupéré les résultats des dernières élections de 2017 sur le site Kaggle [www.kaggle.com](http://www.kaggle.com) en format excel.

## Description des données

Au total il y a 20 fichiers de données avec la même structure :

Exemple de fichier brut:

**Tableau: Exemple de fichier brute**

|  |  |  |
| --- | --- | --- |
| **Nom de la colonne** | **Format de données** | **Description** |
| Code département | Texte | Code département français dont le territoire d’outre-mer et français établi dehors de France |
| département | Texte | Nom du département |
| circonscription | Nombre entier | Nombre de circonscription |
| Inscrits | Nombre entier | Nombre de gens français qui ont le droit de voter aux élections |
| Votants | Nombre entier | Nombre de gens français qui ont voté aux élections |
| Exprimés | Nombre entier | Les suffrages exprimés pour quelques candidats |
| Blancs et nuls | Nombre entier | Nombre de bulletins vides ou avec les erreurs |
| MITTERRAND (CIR) | Nombre entier | Les suffrages exprimés pour ce candidat |
| …... | Nombre entier | Une colonne pour chaque candidats |

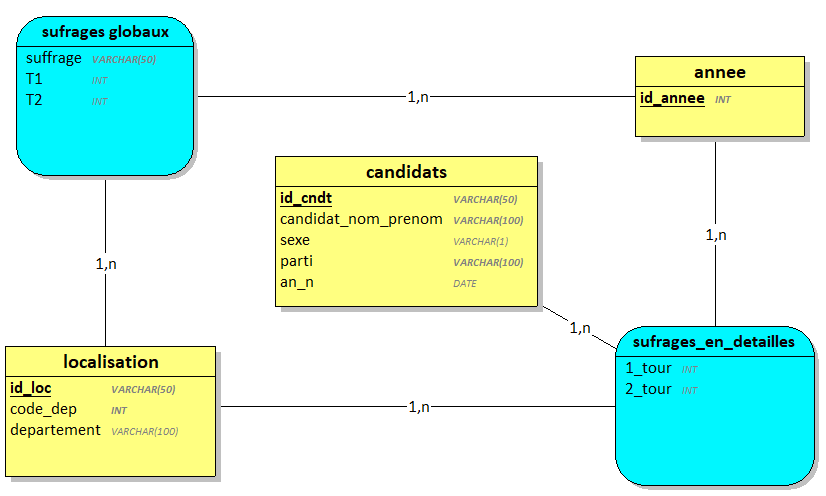
Je veux accumuler dans ma base de données les résultats de dix élections. Donc je vais concaténer les données de 20 fichiers ensemble avec des colonnes qui portent des information de même nature:

* les suffrages globaux
* les suffrages exprimés par chaque candidat.

## Création du modèle

Pour la création d'un modèle conceptuel, j’ai utilisé une application Looping. La solution Looping pour la modélisation conceptuelle de données a été éditée par l'université de Toulouse III. Looping est un logiciel permettant la mise en œuvre intuitive et très rapide de modèles conceptuels de données (MCD) et la génération automatique temps-réel des modèles logiques (MLD) et des requêtes SQL de création des tables BD correspondantes.

La première étape de la conception repose sur l'analyse de l'existant et des besoins. De la qualité de la réalisation de cette première étape dépendra ensuite la pertinence de la base de données par rapport aux usages.



**Figure: Modèle conceptuel de données**

Choix du modèle de la base de données relationnelle:

* Données structurées avec un schéma établi et des valeurs normalisées
* La structure de données est identifiable
* L’intégrité des données peut être contrôlée

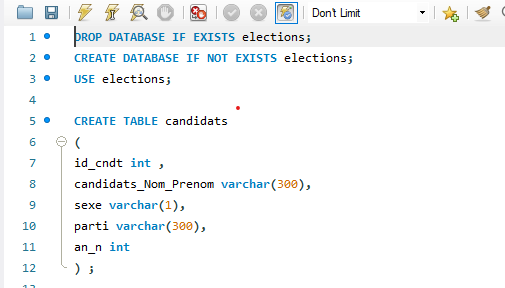
Suite à l’exploration de la donnée, j’ai pu déterminer les tables nécessaires à la création de la base de données relationnelles.

J’ai créé 5 tables: 3 entités indépendantes (candidats, année (élections), localisation (code departement, departement)) et 2 tables intermédiaires (suffrages\_globaux, details\_des\_suffrages (suffrages exprimés par candidats))

* Relations: “many-to-many”

Nous continuons en créant d’un modèle physique de la base de données via MySQL Workbench en utilisant un script SQL. Notre base de données s'appelle ‘elections’.

Voici un exemple de création de la table “candidats” utilisant SQL. Le nom de la table est ‘candidats’ et elle contient 5 attributs. Une classe représente une table, ainsi chaque attribut de la classe représente une colonne de la table. Chaque attribut va être créé avec son datatype et ses contraintes si besoin.



**Figure: Exemple du script créant une table**

Une fois que l’ensemble des tables a été créé avec le même processus que celui ci-dessus mentionné, j’ai installé les clés primaires.

La clé primaire permet d’identifier de manière unique chaque enregistrement dans une table. Elle doit être UNIQUE et ne doit pas contenir de valeur NULL. Elle sert d’index.

Pour identifier de manière unique chaque enregistrement, les clés primaires sont les suivantes:

**Tableau: Les clés primaires**

|  |  |  |
| --- | --- | --- |
| **Nom de la table** | **Nom des colonnes** | **Fonction** |
| candidats | id\_cndt | Clé primaire |
| annee | id\_annee | Clé primaire |
| localisation | id\_loc | Clé primaire |
| suffrages\_globaux | id\_loc, id\_annee, suffrages | Clé composite (composée de plusieurs colonnes) |
| details\_des\_suffrages | id\_loc, id\_annee, id\_cndt | Clé composite (composée de plusieurs colonnes) |

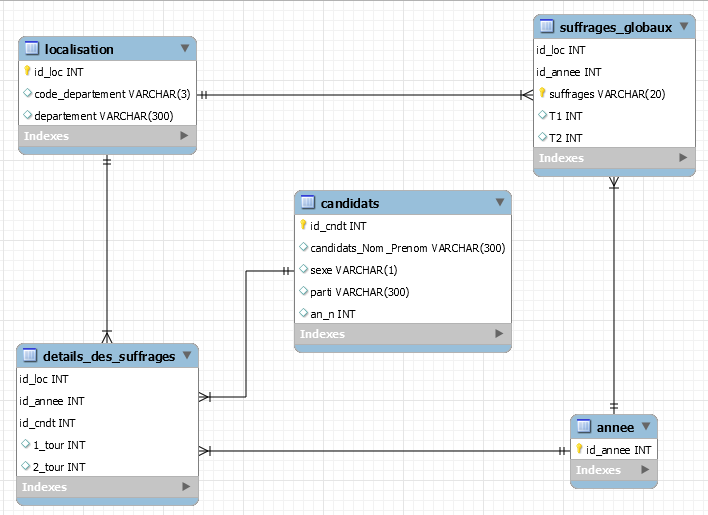
Les clés étrangères ont pour fonction principale la vérification de l'intégrité de la base. Elles permettent de gérer des relations entre plusieurs tables, et garantissent la cohérence des données.

Les clés étrangères sont les suivantes:

**Tableau: Les clés étrangères**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nom de la table** | **Nom de la colonne** | **Colonne de référence** | **Cardinalité** |
| suffrages\_globaux | id\_loc | localisation(id\_loc) | plusieurs-plusieurs |
| suffrages\_globaux | id\_annee | annee(id\_annee) | plusieurs-plusieurs |
| details\_des\_suffrages | id\_loc | localisation(id\_loc) | plusieurs-plusieurs |
| details\_des\_suffrages | id\_annee | annee(id\_annee) | plusieurs-plusieurs |
| details\_des\_suffrages | id\_cndt | candidats(id\_cndt) | plusieurs-plusieurs |

## Modèle physique des données



Ce schéma permet une bonne lecture des relations entre les tables ainsi qu’une

visualisation des composants de chaque table.

## Nettoyer et valider les données

Avant de procéder à l'insertion des données dans la base de données, une grosse partie du projet est consacrée au nettoyage de la donnée.

Pour l’étape de nettoyage j’ai importé les fichiers CSV et Excel sur un Jupyter Notebook et utilisé la librairie Pandas. Cette opération est cruciale pour éliminer les données erronées et combler d'éventuelles lacunes.

Voici les différentes étapes de nettoyage effectuées:

* Supprimer les données superflues (doublon, NaN) et les valeurs aberrantes
* Homogénéiser les informations: même nom des départements coordonnés par code département (j’ai gardé les deux noms du département s'il existe par exemple une version du nom avec les symbols particuliers mais avec le même code département)
* Ajouter les valeurs manquantes: compléter les informations en ajoutant des informations générales pour la description les candidats: prénom, sexe, année de naissance, nom entier du parti politique présenté.

Adapter les données à une structure standard :

- Harmoniser le nom des colonnes

- Normalisation des types de la colonne

* Réorganiser les colonnes
* Sauvegarder une copie (checkpoint) puis exporter les données en CSV

Lorsque les données ont été nettoyées, elles doivent être validées, à savoir déterminer si des erreurs se sont produites dans le processus de préparation des données (il peut arriver qu'une erreur apparaisse pendant cette étape, et il est alors nécessaire de la corriger avant de poursuivre).

On obtient à la fin 5 fichiers de données prêtes à être injectées dans les tables.

## Insertion de données

Après je passe l’ étape d’insertion de données en utilisant script SQL avec une commande LOAD DATA LOCAL INFILE. Cela me permet d'insérer des données par fichiers entiers en format csv. Ces fichiers contiennent un certain nombre de données et sont organisés en tables. Chaque ligne correspond à une entrée, et les colonnes de la table sont séparées par un caractère défini (souvent une virgule ou un point-virgule). Ce type de fichier est facile à produire (et à lire) avec un logiciel de type tableur (Microsoft Excel, ExcelViewer, Numbers…). Il est aussi possible de lire ce type de fichier avec MySQL, afin de remplir une table avec les données contenues dans le fichier.



**Figure: Exemple du script d’alimentation de la table de BDD**

Ensuite j’effectue des tests via des requêtes pour vérifier l’exactitude de la donnée intégrée.

## Sécuriser la base de données

Concernant la sécurité un compte admin gère la base puis un compte utilisateur a été créé avec des accès limités. Chaque utilisateur a bien entendu un mot de passe personnel.

* Création de compte utilisateurs via Workbench Server > User and Privilège ou via ligne de commande
* Sauvegarde de la base de donnée via Workbench ou via mysqldump en ligne de commande
* Sauvegarde des fichiers intermédiaires sous format pickle

# 3. Analyse de données et visualisations

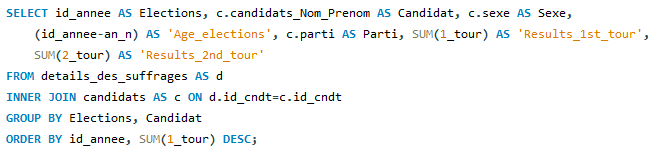
Selon l’objectif fixé au départ, voici la liste des questions à requêter dans la base de données:

1. La liste totale de tous les candidats à chaque élection de 1965 à 2017 avec les résultats des 1er et 2ème tours (nombre de voix), nom des partis politiques représentés, âge en année électorale
2. La liste des gagnants (Présidents) de chaque élection
3. Age moyen des présidents français / Âge minimal des présidents français
4. Le tableau avec le pourcentage de voix pour deux candidats au second tour, % de tous les votes, % de blancs et nuls, % d'abstention
5. Quels départements soutiennent le président le plus fort (2000-2017) en % de tous les électeurs inscrits
6. L'analyse du changement de niveau de l'abstention et niveau de votes par 'Blancs et nuls' aux élections présidentielles en France, 1965-2017 (2ème tour)
7. Participation aux élections par département (en % de Exprimés des Inscrits)

## Exemple d’une requête SQL

## La liste totale de tous les candidats

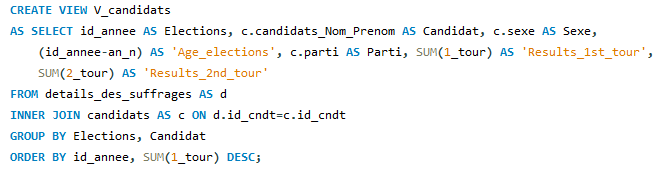
La liste totale de tous les candidats à chaque élection de 1965 à 2017 avec les résultats des 1er et 2ème tours (nombre de voix), nom des partis politiques représentés, âge en année électorale:



**Figure: La requête SQL avec jointure de deux tables “candidats” et “details\_des\_suffrages”.**

## Optimisation de la BDD

Comme je veux utiliser cette requête comme base pour l'analyse de données des candidats pour chaque élection, j'ai créé une vue “V\_candidats” - ça va me faciliter les tâches. Les vues permettent de simplifier les requêtes, de créer une interface entre l'application et la base de données, et/ou de restreindre finement l'accès en lecture des données aux utilisateurs.



**Figure: Creation de la vue ‘V\_candidats’**

Et maintenant au lieu d’utiliser la requête ci-dessous je peux faire une requête en une ligne:



**Figure: Exemple d’une requête en utilisant une vue**

## Visualisation des données

Pourquoi la visualisation est-elle si importante?

L'objectif de la visualisation des données est de simplifier les valeurs des données, de les comprendre et de communiquer des concepts et des notions importantes au public. Nos cerveaux sont programmés pour un traitement visuel rapide. Le cerveau humain peut traiter une image en seulement 13 millisecondes, 90% des informations transmises au cerveau sont visuelles.

Pour effectuer la partie sur l’analyse de données sur Jupiter Notebook, il faut tout d’abord connecter la base de données afin d’exporter la donnée de la base de données en dataframe.

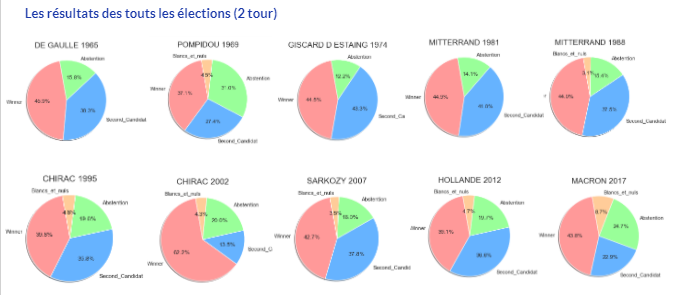


**Figure: Script de connection avec BDD et définition de la function sql\_to\_pandas**

Puis à l’aide des librairies pandas, geopandas, folium et matplotlib je réalise les différentes requêtes afin de modéliser la datavisualisation qui répondent aux problématiques.

Voici les exemples des codes correspondants aux questions:

## La liste des gagnants (Président) de chaque élection



**Figure: Les résultats des 2 tours de toutes les élections présidentielles françaises**

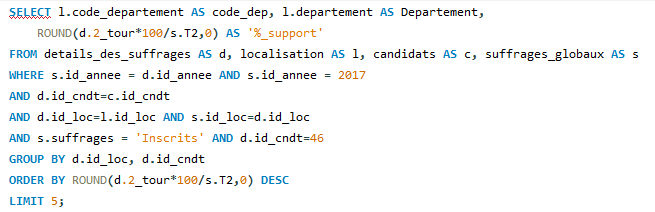
Ce set de diagrammes nous permet de comprendre le changement d'activité et la participation des Français dans la vie politique.

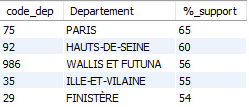
Le Président de la République qui été le plus soutenu par les Français c'était Jacques CHIRAC en 2002 (avec 62,22%) et par contre Le Président avec la plus faible côte de popularité était Georges POMPIDOU en 1969 (avec 37,14%).

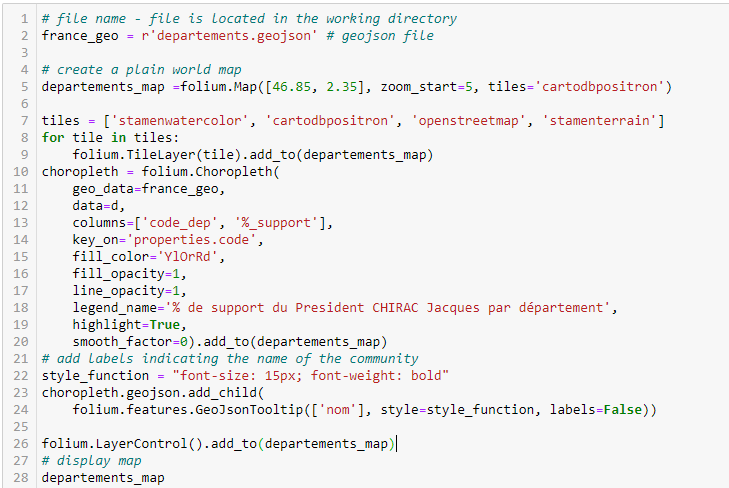
Il était intéressant de voir que lors de certaines élections la gagnant l’ a emporté avec une très faible différence. Par exemple: en 1974 Giscard d'Estaing l'emporte avec une très courte avance - la différence entre Valéry GISCARD D'ESTAING et François MITTERRAND a égalé 1%. La campagne a été très intense, accompagnée d'un débat sérieux. Et la société était fortement divisée.

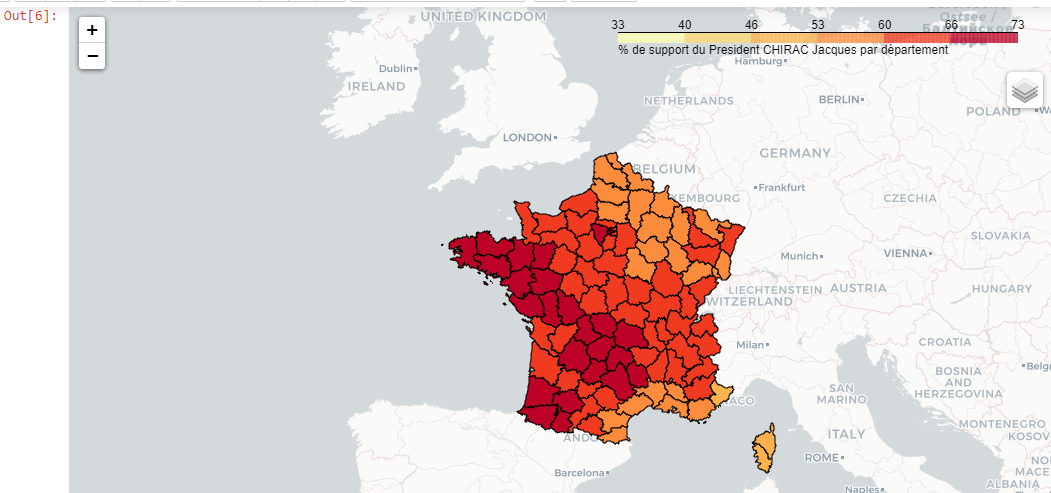
## 

## Quels départements soutiennent le président le plus fort

(élections de 2000 à 2017) en % de tous les électeurs inscrits avec visualisation de réponse 



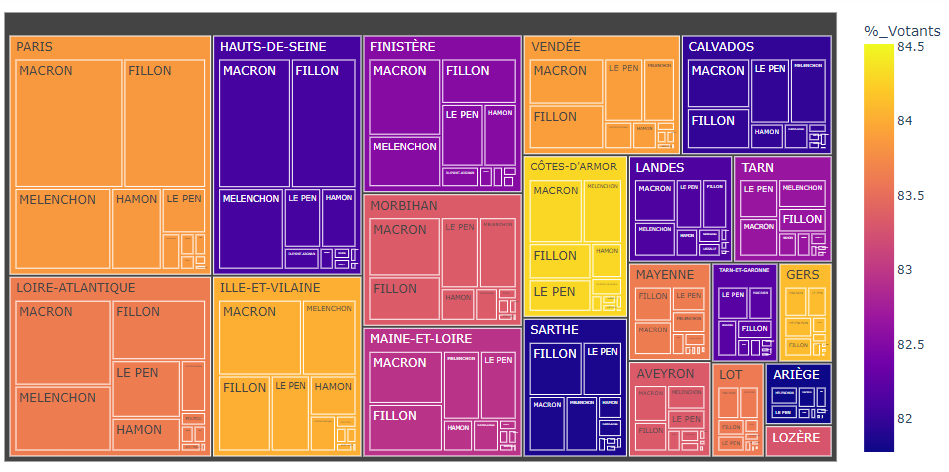




**Figure: La carte représentant le niveau de soutien du président CHIRAC à 2002**

Déjà en regardant les résultats de seulement 4 élections, nous pouvons être sûrs que le soutien le plus fort pour chaque président se situait dans des départements très différents de France. Pour comprendre cette corrélation, nous devrions faire une analyse plus approfondie et rechercher d'autres facteurs. Je peux aussi constater que les gens qui habitent en Corse nous montrent un faible niveau d'activité politique. Leur plus fort soutien n'a été fixé que pour le président Sarkozy en 2007.

## Les résultats des dernières élections en 2017 dans les 20 départements avec le plus haut niveau de participation - suffrages par candidats / le 1er tour, le 2ème tour



**Figure: Les résultats du 1er tour des élections en 2017: les 20 départements avec le plus haut niveau de participation aux élections**



**Figure: Les résultats du 2ème tour des élections en 2017: les 20 départements avec plus haut niveau de participation aux élections**

# 4. Partage des résultats

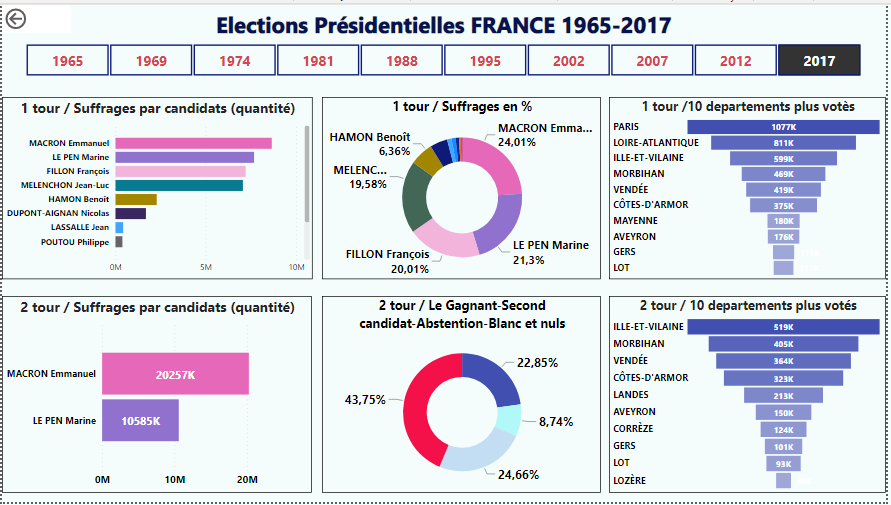
## Power BI

Power BI est un ensemble de services logiciels, d’applications et de connecteurs qui œuvrent ensemble pour transformer des sources de données disparates en informations visuelles immersives et interactives. Les données peuvent être sous forme de feuille de calcul Excel ou de collection d’entrepôts de données hybrides locaux ou sur le cloud. Power BI nous permet de nous connecter facilement à nos sources de données, de visualiser et de découvrir ce qui est important, et de partager ces informations.

Power BI permet de créer du contenu attrayant et interactif ainsi que des tableaux de bord améliorés.

Mon dashboard Power BI contient les réponses pour les questions suivantes:

* Les résultats par candidats au premier et deuxième tour des élections
* Parti politique et âge de chaque candidat au l’année des élections
* Les départements avec le plus haut niveau de participation aux élections au premier et deuxième tour et les chiffres des suffrages globaux

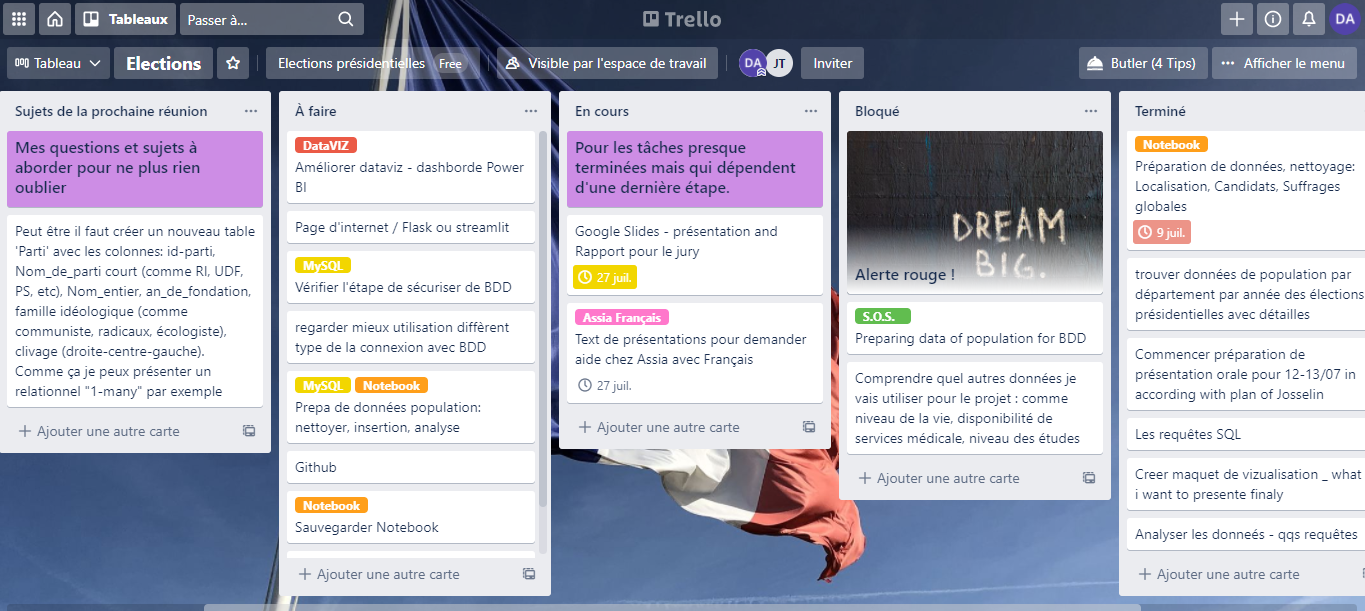


**Figure: Dashboard Power BI**

# 5. Planification et Gestion du projet

## Trello

Pour mener à bien ce projet et organiser les différentes tâches, je me suis appuyé sur le tableau Trello. En créant simplement quatre colonnes (à faire, en cours, bloqué, terminé), j’ai pu suivre l’avancement du projet et ainsi respecter les délais fixés.



**Figure: capture d’écran avec tableau Trelo**

# 6. Remerciements

En premier lieu, je souhaite remercier les personnes qui ont décidé de me choisir pour cette formation et qui ont tellement changé ma vie! Merci - c’était la meilleure chose qui s'est passée pour moi cette année.

Je veux dire merci à tous les formateurs qui ont accompagné la promotion 2021 de Développeur.ses Data de Nanterre - Mme Manel BOUMAIZA, M. Josselin Tobelem, M. David AZRIA et M. Nicolas Zanforlini, pour leur accompagnement et conseils dans la création de mon projet. Vous faites un grand travail, vous changez la vision et donnez un bon exemple.

Je veux personnellement remercier Mme Assia M’JALLED, la formatrice de français, pour son soutien pendant toute la durée de la formation.

Je remercie également, M. Joachim Zentici, pour ses idées et inspiration qui m’ont permis de prendre des décisions concernant le développement du projet. Enfin, un grand merci aux apprenants de la promotion, pour ces 7 mois de soutien dans notre reconversion vers un nouveau métier.

# 7. Bibliographie

**Ressources officielles:**

1. Stakoverflow: <https://stackoverflow.com>
2. Power BI: <https://powerbi.microsoft.com/fr-fr/>
3. Flask: <https://flask.palletsprojects.com/en/2.0.x/>
4. SQL: <https://sql.sh/>
5. Geopandas: <https://geopandas.org/gallery/polygon_plotting_with_folium.html>
6. Matplotlib: <https://matplotlib.org/>
7. IT proger: <https://itproger.com/course/flask>
8. Trello: <https://trello.com/fr>

**Ressources en ligne:**

1. Le “Centre de données socio-politiques SciencesPo” <https://cdsp.sciences-po.fr/fr/ressources-en-ligne/ressource/cdsp_presi2012t1_circ/>
2. Wikipedia <https://fr.wikipedia.org/wiki/%C3%89lection_pr%C3%A9sidentielle_fran%C3%A7aise_de_2022#:~:text=L'%C3%A9lection%20pr%C3%A9sidentielle%20fran%C3%A7aise%20de,d%C3%A9mission%20du%20pr%C3%A9sident%20en%20exercice%2C>
3. Kaggle [www.kaggle.com](http://www.kaggle.com)
4. France Culture <https://www.franceculture.fr/politique/ce-que-les-elections-de-2017-nous-ont-appris-des-francais>
5. Fondation Robert Schuman <https://www.robert-schuman.eu/fr/oee/0002-2002-une-crise-de-la-democratie-electorale-les-enseignements-de-l-election-presidentielle>