Projet de Data Science

[**Description du Projet**](#_30j0zll) **2**

[**Des exemples de sujets**](#_1fob9te) **2**

[**Livrable**](#_3znysh7) **3**

[**Structure conseillée**](#_2et92p0) **4**

[**Deadlines**](#_n36zz9tea3xr) **4**

[**Ressources**](#_tyjcwt) **4**

[**Publication**](#_3w6nleidppwc) **5**

Le but de ce projet est de solidifier vos connaissances en Data Sciences. Cela va vous permettre de :

1. Développer vos compétences et de les pratiquer sur un problèmes que vous avez choisi
2. Vous donner une expérience réelle de ce que fait un Data Scientist au jour le jour
3. Vous faire travailler en équipe et apprendre de chacun

# Description du Projet

Vous allez pouvoir explorer de vraies bases de données que vous aurez choisi pour votre recherche

Voici les étapes :

1. Sélectionnez un sujet qui vous intéresse et trouvez des données pertinentes par rapport à ce sujet que vous souhaitez explorer. Vous pouvez aussi apporter des données de votre entreprise mais faîtes attention que vous ayez bien l’autorisation de les exploiter les montrer à des personnes extérieures à votre entreprise.
2. Trouvez un problème que vous souhaitez résoudre et montrez que ce problème est bien réel.
3. Téléchargez les bases de données sur votre ordinateur et utilisez les technologies vues en cours pour résoudre votre problème.
4. Explorez votre base de données et faites quelques statistiques descriptives, construisez des intervalles de confiance si nécessaire.
5. Appliquez un modèle de Machine Learning appris en cours qui pourrait vous aider à apporter une réponse à votre problème.
6. Mettez vos résultats en forme en utilisant les outils de Business Intelligence vus en cours.
7. Amusez vous!

# Des exemples de sujets

Voici quelques bases de données et problèmes assez populaires trouvés sur Kaggle :

* **Les salaires à San Francisco**
* Data:<https://www.kaggle.com/kaggle/sf-salaries>
* Question 1: Comment les salaires ont évolué dans le temps ?
* Question 2: Pouvez trouver des preuves de discrimination dans les salaires ?
* **Analyse Ressources Humaines**
  + Data:<https://www.kaggle.com/ludobenistant/hr-analytics>
  + Question 1: Pourquoi certains employés s’en vont prématurément ?
  + Question 2: Y’a t-il quelque chose qu’on puisse faire pour réduire le turnover ?
* **Prêt d’argent chez Lending Club**
  + Data:<https://www.kaggle.com/wendykan/lending-club-loan-data>
  + Question 1: Quelle est la cible de Lending Club ?
  + Question 2: Pouvez vous prédire les clients qui sont susceptibles de faire défaut ?
* **Changement Climatique** 
  + Data: <https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data>
  + Question 1: Quand est ce que le changement climatique a commencé ?
  + Question 2: Pouvez vous corréler le changement climatique avec l’activité humaine ?

# Livrable

1- Préparez **une présentation** que vous allez faire devant la classe. La présentation sera de 15 minutes incluant 5 minutes de questions. Votre présentation devra inclure :

* Une sélection rationnelle de vos données en concordance avec le problème que vous essayez de résoudre.
* Pourquoi les outils que vous avez utilisé étaient les mieux à même de vous aider dans votre analyse ?
* Avez vous beaucoup de nettoyage de données à faire ?
* Présentez vos résultats et n’oubliez pas d’y ajouter des preuves statistiques
* Faites une présentation pertinente et synthétique
* Faites-nous une partie recommandations qui ouvre à une discussion.

2- **Un article** d’environ 2 pages word (5 pages si vous incluez des images de votre code) qui reprendra et développera les différents points de votre présentation.

3- **Votre code,** si vous en avez.

Vous n’utiliserez que votre présentation comme support durant la présentation mais vous rendrez à votre professeur les trois items évoqués plus haut.

# Structure conseillée

**1) Etablissement du problème:** Expliquer le problème que vous voulez résoudre

**2) Comment:** Donnez des preuves que votre problème existe

**3) Acteurs:** Qui sont les acteurs de votre problème ? Essayez d’être précis (ex : Ingénieur commerciaux plutôt que département commercial)

**4) Clients:** Qui va utiliser la solution à votre problème ? Les clients de l’entreprise, les équipes en interne, les dirigeants ?

**5) Implémentation:**  Comment est ce que vous résolvez le problèmes ? Quelles sont les technologies que vous allez utiliser.

# Deadlines

**3A** - Introduction du projet

**4A** - Restitution des bases de données au professeur

**6A** - Restitution des bases de données nettoyées

**7B** - Analyses et interprétations prêtes

**8A** - Finition des projets et début des présentations

**8B** - Présentations

# 

# Ressources

<https://www.data.gov/>

<https://www.enigma.com/>

<https://snap.stanford.edu/data/index.html> (social data)

<https://opendata.cityofnewyork.us/>

<https://mattermark.com/>

<https://www.crunchbase.com/>

[https:///www.kaggle.com/](about:blank)

# Publication

Pour vous faire gagner en visibilité, nous proposons aux étudiants qui le souhaitent de filmer leur présentation et de publier leurs projets sur nos réseaux en plus des leurs. L’idée est de vous faire voir par le plus de monde possible pour augmenter vos opportunités d’emploi.

Voici les réseaux sur lesquels votre projet sera publié :

* Blog
* Facebook
* Linkedin
* Medium
* Quora