

**Université Sorbonne Paris-Nord
IUT VILLETTANEUSE**

PORTFOLIO

ENZO SELVARATNAM

E-MAIL:

enzoselvaratnam@gmail.com

NUMÉRO TELEPHONE:

06 99 93 17 28

LINKEDIN:

Enzo SELVARATNAM

Sommaire

03	A propos de moi
04	Mes compétences
05	Mes qualités
06	Pourquoi les sciences de données ?
07	Une spécialité : les Statistiques
08	Une technique: Modèle de Regression
09	Bilan des compétences avec mes projets réalisés
10	Conseils de professionnels
11	Mes objectifs

A propos de moi

Bonjour! **Je suis** actuellement en 1ère année de **BUT Science des données** à l'IUT de Villetaneuse.

Je suis passionné par la science des données notamment dans l'analyse mathématiques des données.

Je souhaite devenir **Data Scientist** afin de développer des modèles prédictifs innovants pour résoudre des problématiques réelles.

Pour cela, j'ai commencé des études en **Science des données**, tout en développant mes compétences en **analyse et manipulation des données**.



Mes compétences

Au cours de cette première année de formation, j'ai consolidé mes bases en **mathématiques acquises au lycée**, et en analyse décisionnelle notamment en **statistiques descriptives et inférentielles**, en **probabilités** et en algèbre linéaire.

J'ai également appris à maîtriser plusieurs langages de programmation, tels que **Python**, **SQL** et **R**, qui sont essentiels pour **l'analyse et la manipulation des données**.

En parallèle, j'ai développé des compétences dans l'utilisation de **logiciels de datavisualisation**, comme **Excel** et **Tableau**, afin de rendre les résultats de mes analyses plus clairs et accessibles.

Mes qualités

Faculté d'adaptation ★★☆

Au fil des différents **projets** menés cette année, j'ai su **m'adapter** à **diverses contraintes** et intégrer **rapidement** les retours reçus, tout en veillant à fournir un **travail de qualité**.

Curieux ★★☆

Depuis **mon enfance**, j'ai toujours été animé par une **grande curiosité** et le **désir** d'apprendre de nouvelles choses pour enrichir mes **connaissances**.

Pensée analytique ★★☆

Au cours de cette première année de formation en sciences des données, j'ai affiné mon analyse des données. Bien qu'elle ne soit pas encore optimale, je progresse pour garantir un travail toujours plus rigoureux et de qualité.



Autonome ★★☆

Dans notre **formation**, l'autonomie joue un rôle **central**. La plupart des **projets** sont menés de façon autonome, développant notre **rigueur**, à prendre des **initiatives** et à mieux comprendre les **réalités** du **monde professionnel**.

Rapidité d'apprentissage ★★☆

Au cours de ma scolarité, j'ai acquis une capacité d'apprentissage qui me permet d'assimiler des concepts complexes, et de les appliquer efficacement notamment en sciences des données.

Travail en équipe ★★☆

À travers ma formation, j'ai développé ma capacité à travailler en équipe à travers les différents travaux de groupe, en partageant des idées et en adaptant ma méthode de travail aux exigences collectives. Cette compétence me permet d'échanger efficacement,

Pourquoi les Sciences des données ?



La **science des données** est un domaine qui m'attire **profondément**, notamment en raison de la **richesse** des **compétences** qu'elle requiert et de l'**impact** qu'elle peut avoir sur notre **société**.

Ce qui me **passionne** particulièrement, c'est la manière dont elle combine plusieurs **disciplines essentielles** : les **mathématiques, les statistiques, la programmation**, des matières que j'ai toujours aimés mais aussi la capacité à réfléchir de façon analytique.

Cette **Combinaison de plusieurs domaines** en fait un domaine à la fois exigeant et stimulant, où l'on ne cesse jamais d'apprendre à une époque où tous évoluent rapidement.

Dans un monde de **plus en plus numérique**, les données sont **omniprésentes** : elles se trouvent **dans nos téléphones, nos applications, nos déplacements, bref partout ...**

Savoir les **collecter, les traiter, les analyser et, surtout, les interpréter**, permet de donner du sens à ce flux massif d'informations. Cela ouvre la voie à des **décisions plus éclairées, plus stratégiques**, tant dans le secteur privé que dans les secteurs publics.

Être capable de transformer des données brutes en informations pertinentes, puis en actions concrètes, est aujourd'hui une compétence précieuse et recherchée.

Être dans un métier de la data fais de nous un élément primordial dans une entreprise ou même dans une branche de l'Etat.

C'est cette capacité à créer de la valeur à partir de l'information, à résoudre des problèmes complexes et à anticiper les besoins futurs qui me motive à m'investir dans cette voie. J'aspire à devenir un professionnel capable de naviguer dans cette complexité, avec rigueur, curiosité et créativité.

Dans un monde où les données se comptent par milliard, la vraie valeur ne réside plus dans leur quantité, mais dans la capacité à en extraire du sens.

“You can have data without information, but you cannot have information without data.”

Daniel Keys Moran, écrivain de science-fiction et programmeur

Les statistiques dans la Science des données

Les statistiques sont l'une des **spécialités** fondamentales de la **science des données**. Elles regroupent des méthodes **mathématiques** permettant de collecter, organiser, **analyser et interpréter les données pour en tirer des conclusions fiables**.

Parmi elles, **la statistique descriptive** est celle que nous avons apprise en premier durant la formation. Elle permet de **résumer de grandes quantités de données à l'aide d'indicateurs simples** comme la moyenne, la médiane, l'écart-type, **offrant ainsi une vision claire et synthétique des données**.



Au-delà de cette première approche, les statistiques incluent aussi **des méthodes plus avancées** comme les **modèles de régression** et les **statistiques inférentielles**.

Celles-ci permettent de mesurer **l'incertitude (intervalles de confiance, fluctuations)**, d'évaluer la **validité des résultats** via des tests d'hypothèses, et de faire des **prédictions fiables**.

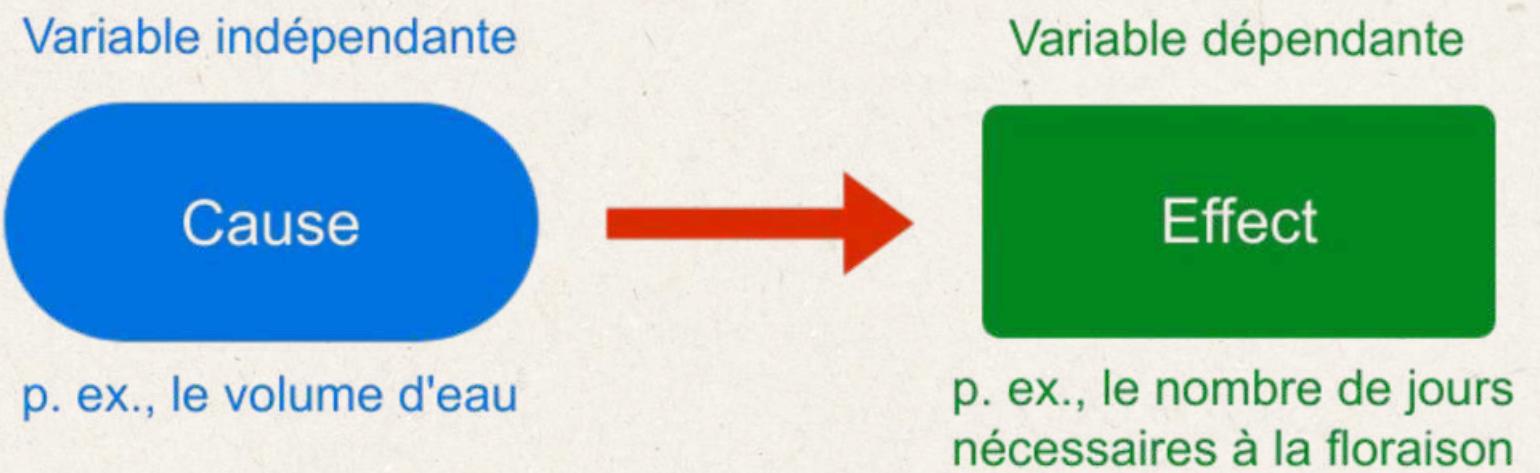
Ce qui m'a **particulièrement intéressé**, c'est le **rôle central** que jouent les **maths et les chiffres dans la prise de décision** : ils offrent un cadre rigoureux pour **comprendre les phénomènes**, identifier des **tendances**, **tester** des hypothèses et **anticiper** des **évolutions**.

J'ai été marqué par leur **utilité concrète** dans des domaines variés comme la **médecine**, l'économie, ou le **marketing**.

Modèle de régression

Les modèles de régression sont des outils statistiques qui permettent de modéliser et quantifier la relation entre une variable à expliquer et une ou plusieurs variables explicatives.

La forme la plus connue est la régression linéaire simple, qui cherche à ajuster une droite aux données afin de prédire une valeur numérique en fonction d'une autre.

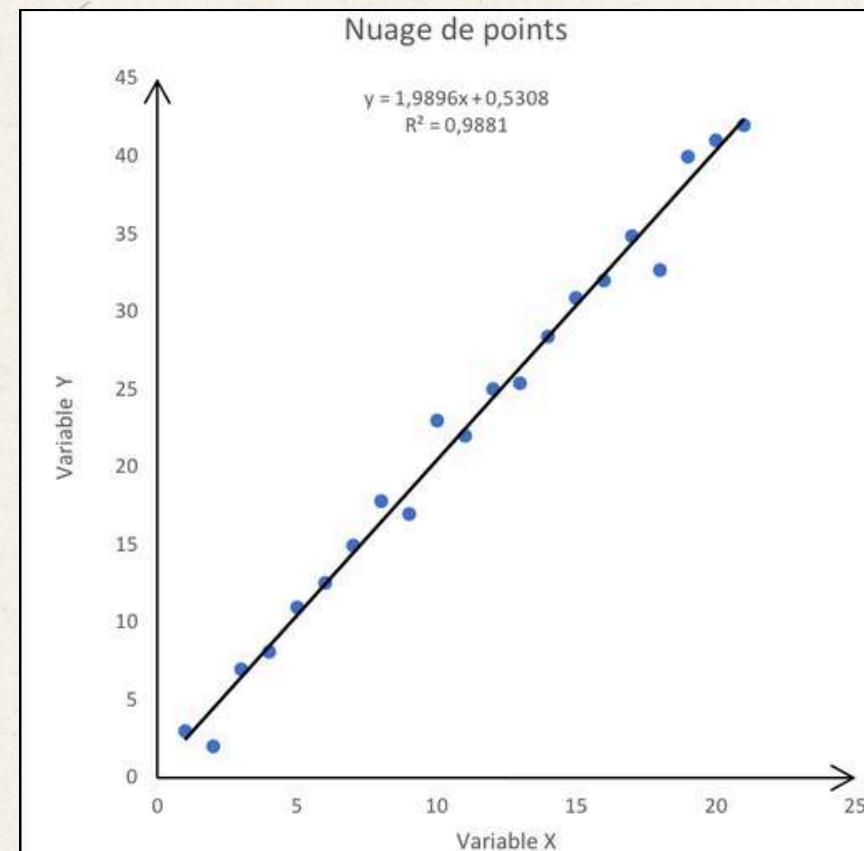


Par exemple, on peut utiliser la régression pour prédire le prix d'un appartement en fonction de sa surface, ou encore le revenu moyen en fonction du niveau d'études. Le modèle calcule alors une équation du type :

$$y = a + bx$$

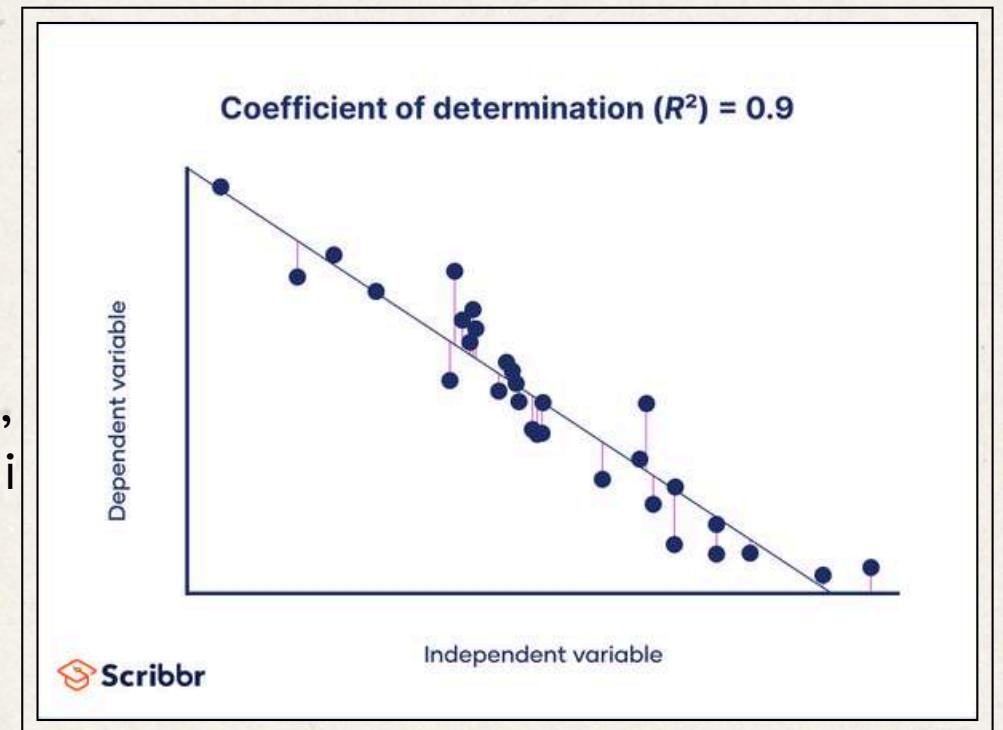
où a est l'ordonnée à l'origine et b le coefficient directeur qui mesure l'impact de x sur y .

Plus la **droite de régression** passe près des points, plus le modèle est ajusté aux données, et donc plus les prédictions ont de chances d'être fiables.



A QUOI ÇA SERT ?

Le coefficient R^2 mesure la qualité de l'ajustement du modèle : plus les résidus c'est-à-dire les écarts entre les valeurs observées et les valeurs prédites sont faibles, plus R^2 est proche de 1, ce qui indique que le modèle explique bien la variabilité des données.

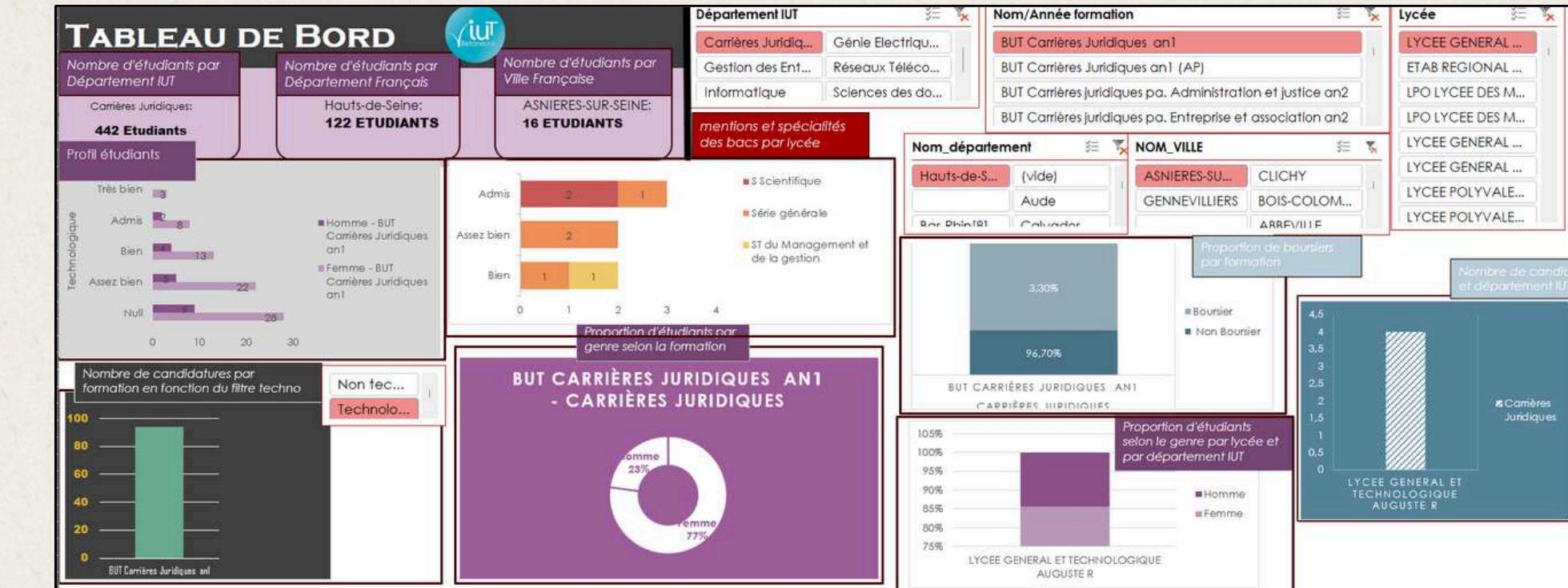


“Ce que je trouve particulièrement intéressant, c'est que la régression ne se limite pas à prédire, elle permet aussi d'interpréter les relations entre les variables. Elle peut mettre en lumière des tendances cachées ou des corrélations inattendues. Dans la version multiple du modèle, on peut intégrer plusieurs variables explicatives à la fois, ce qui se rapproche de situations réelles plus complexes”.

Mes projets

CRÉATION D'UN TABLEAU DE BORD PROFESSIONNEL

Nous devions crée un tableau de bord sur Excel à partir de données sur des étudiants de notre Université



CE QUE J'AI APPRIS

- Comprendre le jeux de données pour en extraire des informations fiables
- Le nettoyage de données est primordiale pour fournir de bonnes analyses fiables
- Utiliser des bons graphiques pour que tous le monde puisse les comprendre
- Savoir comprendre le besoin du client

LES DIFFICULTÉS RENCONTRÉS

- Difficultés a utiliser des graphiques pertinents au début mais j'ai su m'améliorer et je sais que je vais progresser
- Utilisation avancés d'Excel donc j'ai eu besoin de me former et m'améliorer sur ce projet
- Trouvé une bonne charte graphique pour capter l'attention du client

Mes projets

MISE EN ŒUVRE D'UNE ENQUÊTE

Nous avons du mettre en oeuvre un étude statistique entière passant de la récolte des données avec un sondage que nous avions crée à l'analyse de ses données récoltés et de leurs valorisation

le thème était :

Les pratiques alimentaires des usagers sur le campus de Villetaneuse

CE QUE J'AI APPRIS

- La gestion de projet en groupe est essentiel pour fournir un travail de qualité
- Savoir toutes les étapes de la mise en oeuvre d'une enquête
- On a su apprendre à travailler en groupe et rendre des travaux de qualités en suivant les délais

Extrait du questionnaire

Les pratiques alimentaires des usagers sur le campus de Villetaneuse

Nous avons besoin de votre participation et de vos retours précieux !

Étudiants de première année du BUT Science des données de l'Université Sorbonne Paris Nord, nous menons une enquête sociologique sur les pratiques alimentaires des usagers sur le campus de Villetaneuse.

Pour ce faire, nous avons élaboré un questionnaire d'une quarantaine de questions afin de mieux comprendre vos habitudes et vos choix en matière d'alimentation.

Nous serions très reconnaissants de bien vouloir y répondre le plus sincèrement possible. Vos réponses sont entièrement anonymes et pourront ainsi contribuer à l'amélioration des services alimentaires sur le campus. Cela ne vous prendra pas plus d'une **dizaine de minutes**.

Tout le groupe du BUT vous remercie sincèrement à l'avance pour votre participation, essentielle à la réussite de cette étude.

Bon appétit !

VOTRE IDENTITÉ

1. Êtes-vous :

- Étudiant.e
- Personnels travaillant à l'université (enseignant, administratif, etc.)
- Je ne suis que de passage sur le campus

Voici un extrait du sondage que nous avions fait sur Sphinx puis partagés aux individus de notre université.

Puis nous avons fais des analyses sur Excel

Extrait des analyses

Pour analyser cette première hypothèse voyons avec un graphique la répartitions des revenus sur la population interrogée.

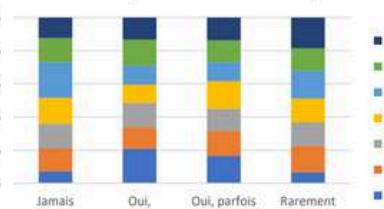


1. Croisement pour la première hypothèse :

Variables :

Variable 1 : Revenu mensuel net de votre foyer (variable qualitative ordinale qui regroupe les participants en différentes catégories de revenu).
Variable 2 : Vous est-il déjà arrivé de sauter un repas ?

Diagramme montrant le revenu selon la fréquence de saut de repas



Voici un extrait de nos analyses à partir des données récoltés que nous avions du récoltés puis valorisés dans un oral.

LES DIFFICULTÉS RENCONTRÉS

- Être en parfaite connexion avec son groupe est difficile, chacun a ses idées, donc il faut faire des concessions, mais nous avons su nous adapter et trouver des solutions communes qui ont permis d'avancer ensemble.
- Il est difficile pour des nouveaux dans ce domaine de savoir gérer toutes les étapes d'une enquête mais on c'est améliorer au fil de ce projet

Mes projets

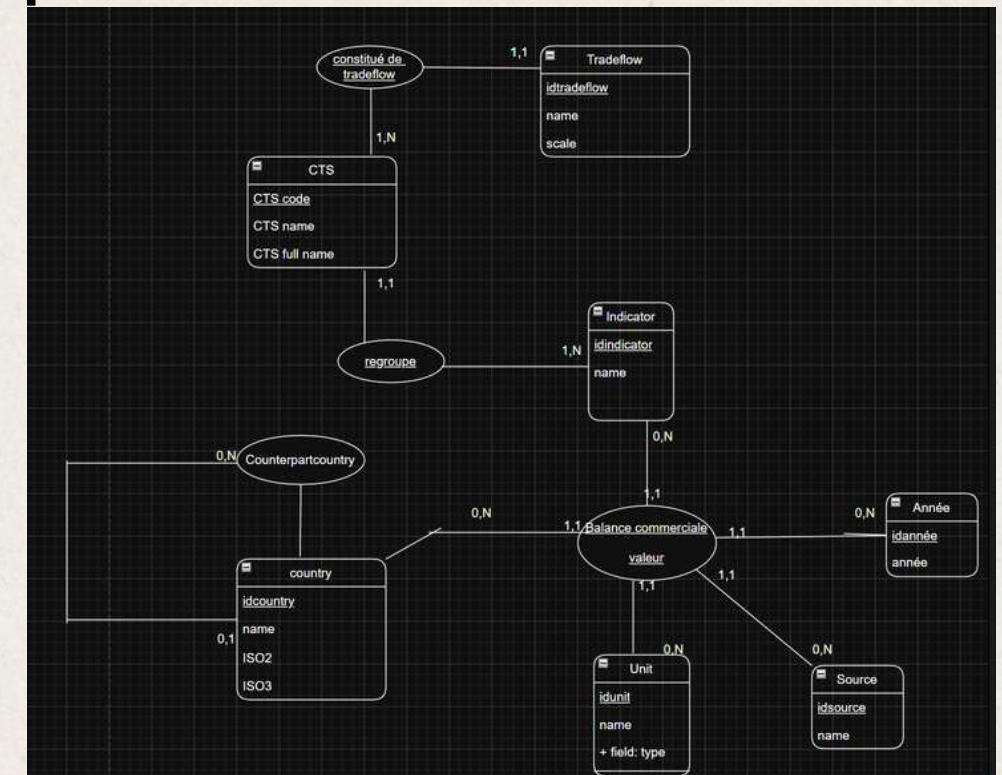
CONCEPTION ET IMPLÉMENTATION D'UNE BASE DE DONNÉES

Ce projet avait pour but de concevoir une base de données relationnelle, de l'alimenter en données via SQL, puis d'effectuer des analyses pour en extraire des informations pertinentes.

CE QUE J'AI APPRIS

- J'ai appris à faire des Modelisation de bases de données (Modèle EA et schéma relationnels)
- Utilisations des différents logiciels de datavisualisation comme Metabase et SQL

Modele EA



Exemple de requête

```
CREATE TABLE raw_unilateral_long AS
SELECT
    r.object_id,
    r.direct_country,
    r.direct_iso2,
    r.direct_iso3,
    r.counterpart_country,
    r.counterpart_iso2,
    r.counterpart_iso3,
    r.indicator,
    r.unit,
    r.source,
    r.cts_code,
    r.cts_name,
    r.cts_full_descriptor,
    r.trade_flow,
    r.scale,
    year::INT AS année,
    value::NUMERIC AS valeur
FROM raw_unilateral r,
LATERAL UNNEST (
    ARRAY[
        '1994','1995','1996','1997','1998','1999','2000','2001','2002','2003',
        '2004','2005','2006','2007','2008','2009','2010','2011','2012','2013',
        '2014','2015','2016','2017','2018','2019','2020','2021','2022','2023'
    ],
    ARRAY[
        f1994, f1995, f1996, f1997, f1998, f1999, f2000, f2001, f2002, f2003,
        f2004, f2005, f2006, f2007, f2008, f2009, f2010, f2011, f2012, f2013,
        f2014, f2015, f2016, f2017, f2018, f2019, f2020, f2021, f2022, f2023
    ]
) AS t(year, valeur);
```

LES DIFFICULTÉS RENCONTRÉS

- Avoir un Modele EA qui est en accord avec les données et création des tables
- Conception des requêtes performantes par rapport à la base de données
- Coordination avec le groupe pour se répartir les tâches

Mes projets

Au fil de cette première année j'ai appris et développé énormément de compétences informatiques économiques,mathématiques et bien d'autres ...

Catégorie	Compétences	Projets réalisés	Niveau
Informatique	Python, R, SQL, , Excel, Tableau	Programme R de régression,automatisations des tâches avec un algorithme ...)	★★★★☆
	Visualisation (ggplot2, matplotlib, pandas)	Infographie pour l'Insee,visualisation statistique sur vrais données)	★★★★☆
	Analyse de données, traitement de bases de données	Création d'une base de données,nettoyage de données avec Python ...	★★★★☆
Mathématiques	Statistiques descriptives, inférentielles	Projets sur des intervalles de confiances sur une population	★★★★☆
	Probabilités, Modélisation, Régressions,Algèbre linéaire	Travaux réalisés sur des modèles de régressions sur les cancers malignes ou bénines	★★★☆☆
	Analyse de données quantitatives / qualitatives	Projet réalisés sur des analyses statistiques bivariés	★★★★☆
Économie	Microéconomie, macroéconomie d'une entreprise	Analyse des indicateurs de performances d'une entreprise	★★★☆☆
ANGLAIS	Lecture et compréhension, communication écrite	TOEIC (890) Niveau C1 pour l'écoute B2 pour la rédaction	★★★★★
	Présentation orale, échanges professionnels	Création d'un poster à présenter à l'oral	★★★★☆

Conseil de professionnel

BIHI Alexandre

Data Scientist à l'Insee



Le premier échange que j'ai eu s'est déroulé avec **Alexandre Bihi, ingénieur et chercheur à l'INSEE**. Je l'ai recontacté après son **intervention** dans notre **formation**, où il était venu présenter **un projet** à réaliser pour l'INSEE. Son **retour d'expérience** m'a tout de suite **intéressé**, et j'ai souhaité en savoir plus sur **son parcours** et sur les **exigences du domaine**.

Lors de notre échange, il m'a **conseillé sur plusieurs aspects, techniques comme non techniques**.

Il a notamment insisté sur l'importance de maîtriser parfaitement des **outils** comme **R** ou **Python**, qu'il considère comme des compétences fondamentales pour tout data scientist.

Il m'a également apporté des conseils précieux sur l'**approche à adopter** pour les stages ou les alternances : commencer ses **démarches le plus tôt possible**, et surtout, se démarquer par un profil **atypique**, capable de sortir du lot. Cette discussion m'a permis de mieux comprendre **les réalités du métier**, mais aussi **l'attitude à adopter pour réussir dans ce domaine** : curiosité, rigueur, sens de l'initiative, et **capacité à se distinguer** dans un environnement **très concurrentiel**.

Mes objectifs

OBJECTIF À COURT-TERME

Dans un premier temps j'aimerais finir mon BUT avec de très bonnes notes pour avoir une alternance et continuer mes études. Je vais continuer de me former pour toujours continuer à apprendre.



OBJECTIF A MOYEN-TERME

Après l'obtention de mon BUT, j'aimerais intégrer une école d'ingénieur en alternance. Pour moi, c'est l'un des meilleurs moyens de devenir un data scientist compétent, en alliant l'acquisition de connaissances théoriques solides à une véritable expérience de terrain. L'alternance me permettra de monter en compétences tout en étant confronté aux enjeux concrets du métier.

OBJECTIF À LONG TERME

À long terme, je souhaite devenir un data scientist compétent, capable de relever des défis complexes et de mener des projets innovants. Mon ambition est de me faire un nom dans ce domaine en développant une expertise reconnue et en apportant une réelle valeur ajoutée grâce à mes compétences.