

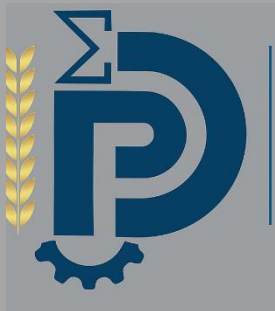
REPUBLIQUE DU CAMEROUN
Paix – Travail – Patrie
Ministère de l'Enseignement
Supérieur
UNIVERSITE DE DOUALA

Ecole Nationale Supérieure
Polytechnique de Douala



REPUBLIC OF CAMEROON
Peace – Work – Fatherland
Ministry of Higher Education
UNIVERSITY OF DOUALA

National Higher Polytechnic
School of Douala



RAPPORT DE STAGE ACADEMIQUE

**FILIERE : Génie Informatique et Télécommunication (GIT) – Génie
Logiciel**



Période : 10 juillet 2025 au 10 Septembre 2025

Rédigé par : MBOUMELA TIFFA JAURES WILSON

Matricule : **22G00249**

NIVEAU 3 - GENIE LOGICIEL

Sous l'encadrement professionnel :

M. Kevin Mbieleu

ANNEE ACADEMIQUE 2024 / 2025

DEDICACES



A mes parents

REMERCIEMENT

Pour commencer, je voudrais exprimer ma profonde gratitude et ma profonde reconnaissance à mon superviseur académique, **M. Kevin Mbieleu**, pour ses précieux conseils, ses encouragements et ses commentaires constructifs tout au long de mon stage. Je tiens également à remercier sincèrement mon superviseur technique, **M. Eta Cyril Nelson**, pour son soutien continu, ses critiques éclairées et son dévouement pour garantir que mes tâches soient accomplies selon les normes les plus élevées.

Je suis également reconnaissant au chef du département d'ingénierie et à tous mes professeurs de m'avoir fourni les connaissances et les ressources fondamentales qui se sont avérées utiles pendant mon stage.

Je tiens à remercier tout particulièrement mon grand frère **Mandela MBOUMELA**, les membres de ma famille et mes amis proches, pour leur soutien indéfectible, leurs encouragements et pour m'avoir soutenu dans les moments difficiles comme dans les moments joyeux. Vous êtes mes trésors.

Enfin, j'exprime ma profonde gratitude à mes remarquables collègues et membres de l'équipe du **CODEC** pour leur collaboration et leur camaraderie, qui ont rendu mon expérience de stage à la fois enrichissante et agréable.

INTRODUCTION

Dans le cadre de ma formation académique, j'ai eu l'opportunité d'intégrer l'entreprise **CODEC** pour un stage enrichissant, orienté vers l'analyse de données et l'apprentissage automatique. Cette expérience s'inscrit pleinement dans mon objectif professionnel de devenir **ingénieur en Data Science**, afin de pouvoir développer des modèles de **Machine Learning** et de **prévision** performants, capables d'apporter des solutions concrètes aux problématiques des entreprises.

L'entreprise **CODEC**, reconnue pour son dynamisme et son ouverture à l'innovation, m'a offert un cadre idéal pour renforcer mes connaissances et approfondir mes compétences. J'y ai eu l'occasion de mettre en pratique les notions apprises en cours, mais aussi de découvrir de nouvelles technologies clés pour mon futur métier, notamment le **nettoyage de données en Python**, l'utilisation de bibliothèques de **Machine Learning** et l'implémentation de modèles de prédiction.

Ce stage a été pour moi une véritable immersion dans le monde professionnel, me permettant de mieux comprendre les attentes du secteur et de développer une méthodologie de travail rigoureuse. Il a également renforcé ma passion pour l'intelligence artificielle et la data science, confirmant ainsi mon choix de carrière.

Ce rapport a pour objectif de retracer les étapes marquantes de mon parcours au sein de **CODEC**, tout en mettant en évidence les apprentissages significatifs réalisés, les outils utilisés, et les compétences pratiques acquises qui contribuent à mon développement vers le métier d'ingénieur en Data Science.

I – PRESENTATION DE L'ENTREPRISE

I – 1 - Situation de l'entreprise

CODEC est une entreprise située au Cameroun, dans la ville de Douala. Pour une situation exacte, elle est située à LOGBESSOU à environ 200 mètres du carrefour « marché pk14 ». (Voir le point très rouge)

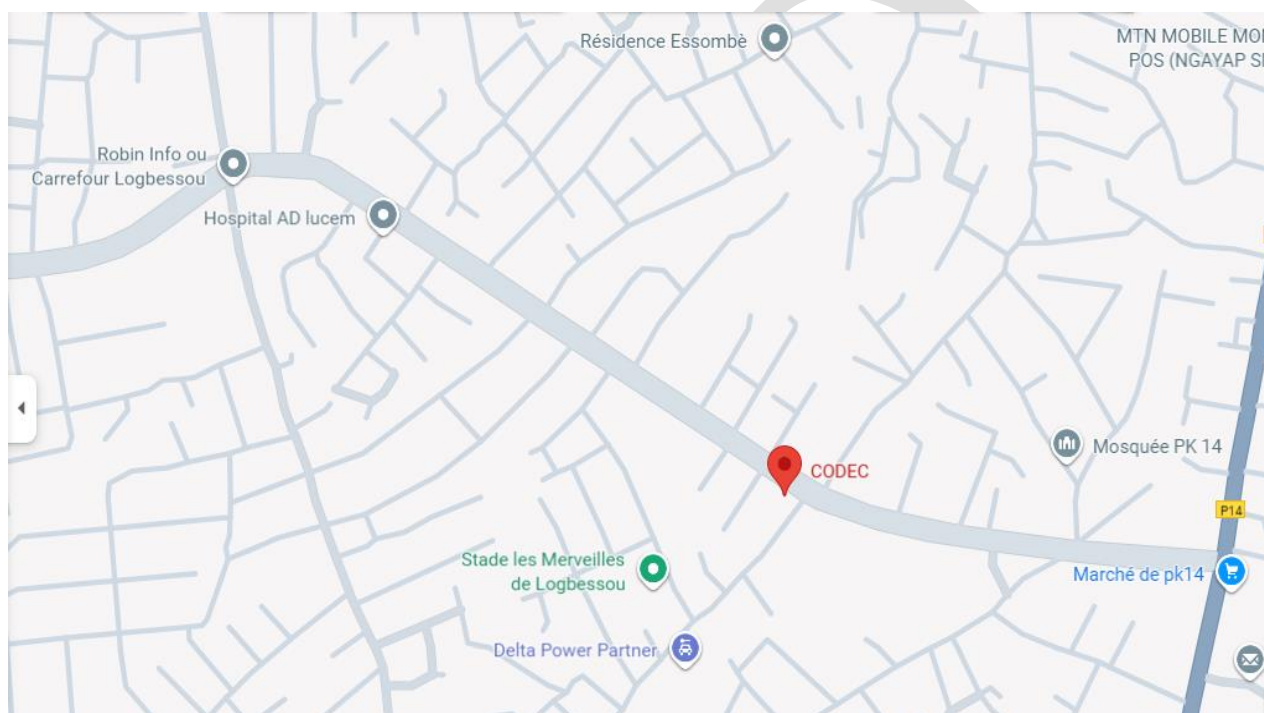


FIGURE 1 : Localisation de l'entreprise

I - 2 - Adresse complète de l'entreprise

CODEC	Logbessou - Douala	www.codec.cm
Contact :	Tel: +237 678 538 351	contact@codec.cm

I – 3 - Historique de l'entreprise

CODEC est une jeune startup technologique créée en **novembre 2022**. Son siège se situe à **Logbessou, Immeuble BAO, Douala, Cameroun**. Depuis sa création, l'entreprise s'est donné pour mission de relever les défis économiques du Cameroun à travers la **conception de solutions logicielles innovantes**, adaptées aux besoins

spécifiques des entreprises et des particuliers.

I-4- Indication du secteur d'activité

CODEC évolue dans le **secteur des technologies de l'information et de la communication (TIC)**, avec un positionnement sur plusieurs pôles complémentaires :

- **Hébergement web** (cloud, mutualisé, gestion de noms de domaine, sécurité SSL) ;
- **Développement d'applications web et mobiles sur mesure**, destinées à optimiser les processus métiers ;
- **Formation technologique** via l'Académie CODEC, qui dote les jeunes de compétences numériques de base et avancées ;
- **Espaces de coworking** et services associés pour les entrepreneurs et freelances.

Cette diversité d'activités permet à CODEC de proposer un **écosystème complet** favorisant l'innovation et la transformation digitale des entreprises.

I-5- Nombre du personnel

CODEC compte une **équipe jeune et dynamique d'une vingtaine de membres**, composée de plusieurs pôles :

- Une **Direction Générale** pilotée par le PDG et son assistant exécutif.
- Une **équipe technique** regroupant CTO, ingénieurs logiciels seniors et juniors, ainsi que des stagiaires.
- Une **division financière et académique** comprenant le CFO, un responsable financier, des coordinateurs de formation et des stagiaires.
- Des **spécialistes en infrastructure et en hébergement**, dont un gestionnaire d'infrastructure, des administrateurs système, des ingénieurs support et des stagiaires.
- Un **pôle opérationnel et coworking** avec responsables et agents dédiés.
- Des **services de support** incluant RH, marketing/ventes et support client 24/7.

Cela représente une organisation en **croissance**, orientée vers la collaboration et l'apprentissage continu, et intégrant régulièrement des stagiaires pour favoriser la montée en compétences.

I-6- Organigramme hiérarchique

Cadre organisationnel du CODEC

1. Direction Générale (Dirigée par le PDG)

- Veille à ce que les objectifs stratégiques soient atteints et à ce que les ressources soient utilisées efficacement.

Assistant exécutif – Assiste le PDG dans les tâches administratives et de coordination.

2. CODEC DEV (Géré par le CTO)

Supervise la technologie, le développement de logiciels et l'innovation.

- Ingénieurs logiciels seniors – Dirigent des projets de développement et encadrent des ingénieurs juniors.
- Ingénieurs logiciels juniors – Participent au codage, au débogage et aux tests des applications.
- Stagiaires – Soutiennent les tâches de développement de logiciels afin d'acquérir une expérience pratique.

3. Académie CODEC (Supervisée par le CFO)

Gère les questions financières et supervise les programmes de formation.

- Responsable financier – Gère la comptabilité, la budgétisation et la paie.
- Coordonnateurs de formation – Élaborent et gèrent les programmes de formation.
- Stagiaires – Participent aux analyses financières et aux activités de formation

4. Hébergement CODEC (Division Infrastructure)

- Gestionnaire d'infrastructure – Gère les services d'hébergement, la sécurité et l'évolutivité.
- Administrateurs système – Maintiennent l'infrastructure d'hébergement et optimisent les performances.
- Ingénieurs de support – Résolvent les problèmes d'hébergement des clients et fournissent un support technique.
- Stagiaires – Participent à la surveillance des services d'hébergement et acquièrent de l'expérience en matière d'infrastructure informatique.

5. Espaces CODEC (axés sur les services de coworking)

- Responsable des opérations – Supervise les espaces de coworking et la logistique.

- Coordonnateurs d'installations – Gèrent les réservations et l'entretien des espaces.
- Agents des relations clients – Répondent aux demandes des clients et améliorent l'engagement client.
- Stagiaires – Participent à la gestion de l'espace de coworking et à la logistique opérationnelle.

6. Services de soutien

- Responsable des ressources humaines – Gère le recrutement, le bien-être des employés et la conformité.
- Responsable marketing et ventes – Élabore des stratégies de marque et stimule la croissance de l'entreprise.
- Équipe de support client – Fournit une assistance et un support client 24h/24 et 7j/7.

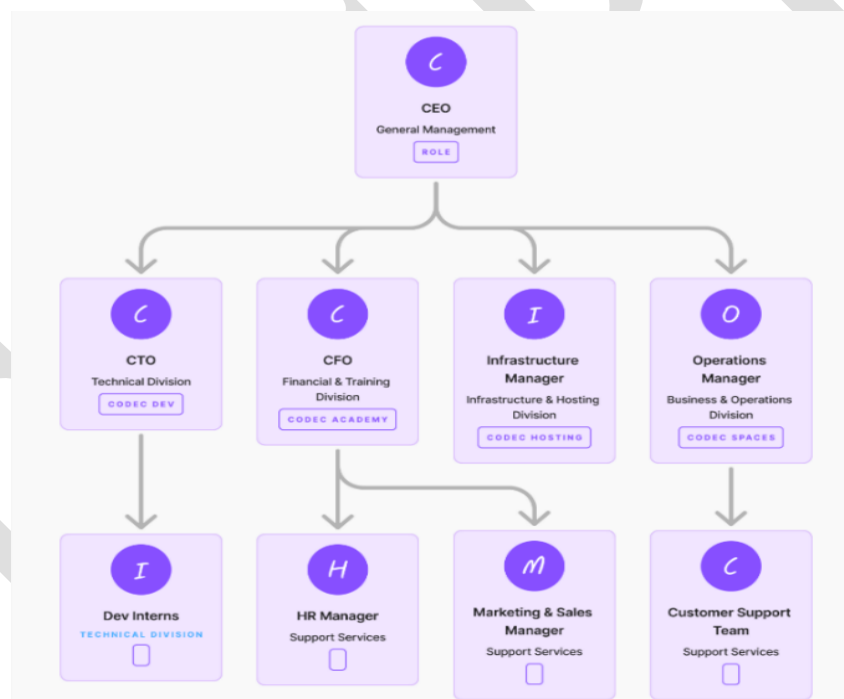


FIGURE 2 : Organigramme de CODEC

II – LES TRAVAUX EFFECTUES

Au cours de ce stage chez **CODEC**, j'ai eu l'opportunité de découvrir le métier de **développeur full-stack orienté Data Science** et de comprendre de manière concrète les défis rencontrés lors de la conception d'une application web intégrant des fonctionnalités d'analyse de données. Pour mieux présenter les travaux réalisés, il est pertinent de commencer par les outils et technologies appris, avant de détailler le projet qui m'a été confié.

1. Les outils et technologies mis à ma disposition

Durant la première phase de mon stage, l'accent a été mis sur l'apprentissage des technologies et des langages nécessaires à la réalisation du projet. Cette phase a été indispensable pour devenir pleinement opérationnel.

Ainsi, j'ai pu me former sur :

- **Front-end** : React.js et Tailwind CSS pour créer des interfaces web interactives et modernes.
- **Back-end** : Spring Boot pour construire des API REST robustes et gérer la communication avec la base de données.
- **Base de données** : MySQL pour le stockage et la gestion des datasets.
- **Data Science** : Python avec les bibliothèques Pandas, NumPy, Matplotlib, Seaborn et Scikit-learn pour le traitement des données, l'analyse statistique et la modélisation prédictive.
- **Outils complémentaires** : Git pour le versioning, Postman pour tester les API, et un IDE adapté pour Java (Spring Tool Suite).

Cette phase de formation m'a permis de comprendre non seulement la syntaxe des langages, mais également **les bonnes pratiques de développement**, la structuration d'un projet full-stack et les méthodes de traitement des données. C'est seulement après cette période d'apprentissage que j'ai pu passer à la réalisation concrète du projet.

2. Les missions du poste occupé

Une fois les bases techniques acquises, j'ai été impliqué dans la réalisation d'un projet pratique et complet : **DataLab – Plateforme d'analyse et de prédiction des salaires**. Ce projet avait pour objectif de mettre en application les connaissances acquises, et de maîtriser les étapes de développement d'une application full-stack intégrant la data science.

Présentation du projet

Le projet consistait à développer une application web capable de :

- Explorer un jeu de données socio-professionnelles (âge, sexe, diplôme, secteur, heures travaillées, salaire).
- Réaliser des **statistiques descriptives** et des visualisations graphiques (histogrammes, boxplots, barplots).
- Analyser les **corrélations et relations entre variables** (ANOVA, heatmaps, scatterplots).
- Mettre en œuvre des **modèles prédictifs** (régression linéaire, régression logistique, Random Forest, Gradient Boosting) et évaluer leurs performances.
- Permettre à un utilisateur de saisir son profil et d'obtenir une prédiction de salaire ou de probabilité de dépasser un certain seuil.

Implémentation du projet

a) Création de la base de donnée et des tables à l'aide de XAMPP

```
-- 1. Création de la base
CREATE DATABASE IF NOT EXISTS datalab CHARACTER SET utf8mb4 COLLATE utf8mb4_general_ci;
USE datalab;

-- 2. Table des utilisateurs
CREATE TABLE users (
  id INT AUTO_INCREMENT PRIMARY KEY,
  username VARCHAR(50) NOT NULL UNIQUE,
  email VARCHAR(100) NOT NULL UNIQUE,
  password VARCHAR(255) NOT NULL,
  role ENUM('admin', 'analyst', 'user') DEFAULT 'user',
  created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP
);

-- 3. Table socio-professionnelle (salary est maintenant CONTINU)
CREATE TABLE socio_data (
  id INT AUTO_INCREMENT PRIMARY KEY,
  age INT NOT NULL,
  sexe ENUM('Male', 'Female') NOT NULL,
  education_level VARCHAR(50) NOT NULL,
  sector VARCHAR(100) NOT NULL,
  hours_per_week INT NOT NULL,
  salary FLOAT NOT NULL,
  inserted_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP
);

-- 4. Table résultats d'analyses (inchangé)
CREATE TABLE analysis_results (
  id INT AUTO_INCREMENT PRIMARY KEY,
  analysis_type VARCHAR(100) NOT NULL,
  result_json JSON NOT NULL,
  created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
  created_by INT,
  FOREIGN KEY (created_by) REFERENCES users(id) ON DELETE SET NULL
);

-- 5. Table des prédictions (salary continu)
CREATE TABLE predictions (
  id INT AUTO_INCREMENT PRIMARY KEY,
  user_id INT,
  age INT NOT NULL,
  sexe ENUM('Male', 'Female') NOT NULL,
  education_level VARCHAR(50) NOT NULL,
  sector VARCHAR(100) NOT NULL,
  hours_per_week INT NOT NULL,
  predicted_salary FLOAT NOT NULL,
  model_used VARCHAR(50) NOT NULL,
  predicted_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
  FOREIGN KEY (user_id) REFERENCES users(id) ON DELETE CASCADE
);
```

Figure 3 : Création de la base de donnée et des tables à l'aide de XAMPP

Users : Gère les comptes des personnes qui utilisent la plateforme. Rôle : permet de distinguer un administrateur d'un simple utilisateur.

socio_data : Contient les données brutes importées depuis le dataset Kaggle (âge, sexe, diplôme, secteur, heures/semaine, salaire). C'est la table principale pour la data science.

analysis_results : Permet de stocker les résultats de statistiques ou corrélations en JSON. Exemple : la moyenne des salaires par secteur, la matrice de corrélation. Facilite la réutilisation sans recalculer à chaque fois.

Predictions : Stocke les prédictions faites par les utilisateurs via l'interface. Exemple : un utilisateur entre son profil → on enregistre sa demande et la sortie du modèle ML (salaire estimé).

b) Nettoyage des données (dataset) et sauvegarde dans un fichier csv

```
import pandas as pd
import numpy as np

# Charger le dataset depuis le fichier CSV
data = pd.read_csv("adult.csv")

# Remplacer les "?" (valeurs manquantes codées sous forme de texte) par de vrais NaN de NumPy
data = data.replace('?', np.nan)

# Vérifier combien de valeurs manquantes il y a par colonne
data.isnull().sum()

# Remplir les valeurs manquantes dans certaines colonnes avec la valeur la plus fréquente (mode)
for col in ['workclass', 'occupation', 'native.country']:
    data[col].fillna(data[col].mode()[0], inplace=True)

# Vérifier le nombre de doublons dans le dataset
print(data.duplicated().sum())

# Supprimer toutes les lignes dupliquées pour ne garder qu'une occurrence unique
data = data.drop_duplicates()
print(data.shape) # Afficher les dimensions du dataframe après suppression

# Supprimer les espaces inutiles au début et à la fin des chaînes de caractères
data = data.applymap(lambda x: x.strip() if isinstance(x, str) else x)

# Afficher les valeurs uniques pour chaque colonne de type "object" (catégorielle)
for col in data.select_dtypes(include='object').columns:
    print(col, data[col].unique())

# Définir une fonction pour détecter les valeurs aberrantes (outliers) dans une colonne numérique
def look_outliers(df, column):
    Q1 = df[column].quantile(0.25) # Premier quartile (25%)
    Q3 = df[column].quantile(0.75) # Troisième quartile (75%)
    IQR = Q3 - Q1 # Intervalle interquartile (IQR)
    lower = Q1 - IQR*1.5 # Limite inférieure
    upper = Q1 + IQR*1.5 # Limite supérieure
    # Retourner les lignes où la valeur est en dehors de ces bornes (potentiels outliers)
    return df[(df[column] < lower) & (df[column] > upper)]

# Afficher les 5 premières lignes pour vérifier l'état du dataframe
data.head()

# Vérifier les outliers dans certaines colonnes clés
columnc = ['age', 'fnlwgt', 'education.num', 'capital.gain', 'capital.loss', 'hours.per.week', 'salary']
for col in columnc:
    da = look_outliers(data, col)
    print(col, da.shape) # Afficher combien d'outliers ont été détectés par colonne

# Supprimer la colonne "fnlwgt" jugée peu utile pour l'analyse
data = data.drop(columns=['fnlwgt'])

# Sauvegarder le dataset nettoyé dans un nouveau fichier CSV
data.to_csv('adult_clean.csv', index=False)

# Charger les données nettoyées dans une base MySQL (table "adult_clean")
from sqlalchemy import create_engine
engine = create_engine("mysql+pymysql://root:localhost@localhost/datalab")
data.to_sql("adult_clean", con=engine, if_exists="replace", index=False)
```

Figure 4 : Lignes de code pour le nettoyer des données (un dataset).

c) Connexion entre l'api de Spring Boot et la base de données SQL.

```
package com.backendAnalyseSalaire.api.controller;

import org.springframework.web.bind.annotation.*;
import org.springframework.web.client.RestTemplate;
import java.util.*;

@RestController
@RequestMapping("/api/ml")
@CrossOrigin(origins = "")
public class MLController {

    @PostMapping("/evaluate")
    public Object evaluateModels(@RequestBody Map<String, Object> payload) {
        RestTemplate restTemplate = new RestTemplate();
        String flaskUrl = "http://127.0.0.1:5000/evaluate";
        return restTemplate.postForObject(flaskUrl, payload, Object.class);
    }

    @PostMapping("/predict")
    public Object predictSalary(@RequestBody Map<String, Object> payload) {
        RestTemplate restTemplate = new RestTemplate();
        String flaskUrl = "http://127.0.0.1:5000/predict";
        return restTemplate.postForObject(flaskUrl, payload, Object.class);
    }
}
```

Figure 5 : Lignes de codes pour sauvegarder les prédictions dans la base de données via API.

d) Représentation des statistiques descriptives avec Flask et Réact.js.



Figure 6 : représentation de la page des statistiques descriptives du dataset.

e) Analyse bivariée par la corrélation et la méthode ANOVA.

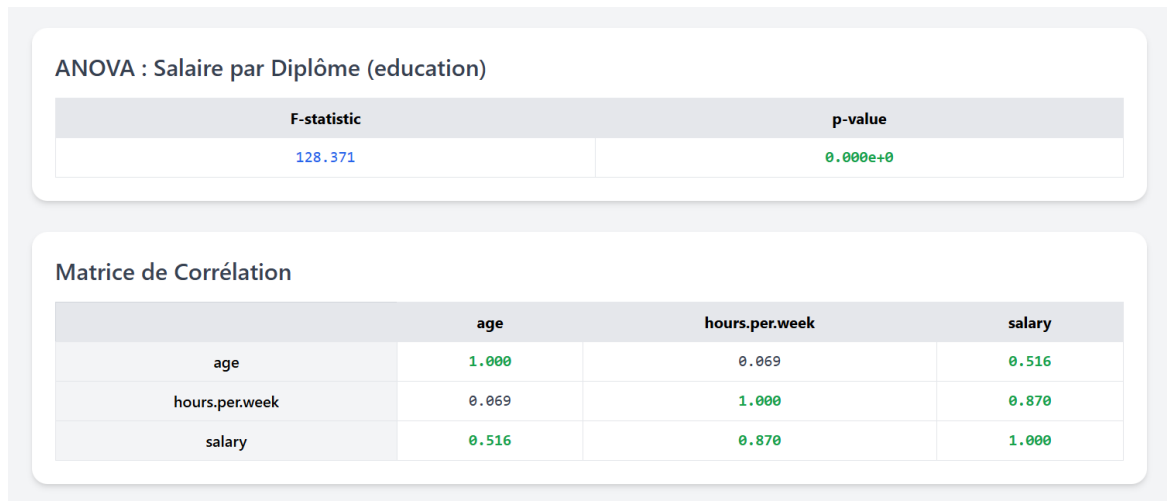


Figure 7 : Représentation des métriques d'analyse bivariée (Corrélation et ANOVA)

f) Implémentation de l'apprentissage automatique (machine learning) par régression linéaire et sauvegardes des modèles avec joblib en fichier.pkl.

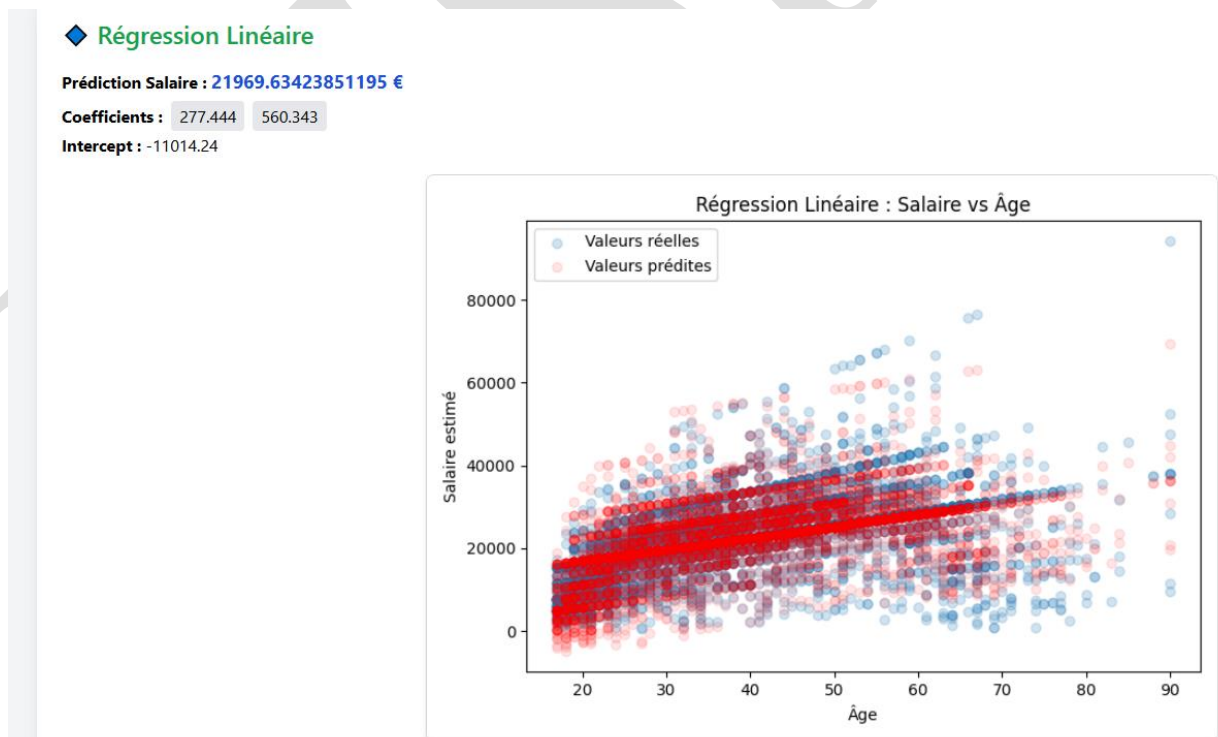


Figure 8 : Comparaison des valeurs réelles en bleu ou valeur prédites par régression linéaire en rouge.

- g) Implémentation des modèles d'apprentissage avancée (machine learning) tel que le Random-Forest et le GradientBoosting.

Modèles utilisés : RandomForest & GradientBoosting


Modèle	Hyperparamètres	R ²	RMSE	MAE
 Random Forest	<pre>{ "bootstrap": true, "ccp_alpha": 0, "criterion": "squared_error", "max_depth": null, "max_features": 1, "max_leaf_nodes": null, "max_samples": null, "min_impurity_decrease": 0, "min_samples_leaf": 1, "min_samples_split": 2, "min_weight_fraction_leaf": 0, "monotonic_cst": null, "n_estimators": 100, "n_jobs": null, "oob_score": false, "random_state": 42, "verbose": 0, "warm_start": false }</pre>	0.999	197.33	32.05
 Gradient Boosting	<pre>{ "alpha": 0.9, "ccp_alpha": 0, "criterion": "friedman_mse", "init": null, "learning_rate": 0.1, "loss": "squared_error", "max_depth": 5, "max_features": null, "max_leaf_nodes": null, "min_impurity_decrease": 0, "min_samples_leaf": 1, "min_samples_split": 2, "min_weight_fraction_leaf": }</pre>	1.000	142.24	55.07

Figure 9 : affichage des indicateurs de qualité du modèle de prédiction et des meilleurs paramètres à utiliser pour prédire.

h) Essaie de prédiction sur l'interface utilisateur.

The image displays two versions of a web application interface titled "Estimation de Salaire".

The left version shows the input form with the following fields and values:

- Age: (empty)
- Sexe: Male
- Education Level: (empty)
- Sector: (empty)
- Hours Per_week: (empty)
- Prédire button

The right version shows the form with the following values filled in:

- Age: 25
- Sexe: Male
- Education Level: 15
- Sector: génie logiciel
- Hours Per_week: 45
- Prédire button
- Salaire estimé : 20 003 FCFA

Figure 10 : Exemple d'estimation (prédiction) de salaire en fonction des données remplies

Tâches réalisées :

Au cours de ce projet, différentes activités m'ont été confiées :

- **Intégration du dataset et nettoyage des données** : gestion des valeurs manquantes, doublons et outliers.
- **Développement back-end** : création d'API REST avec Spring Boot pour récupérer et stocker les données.
- **Implémentation de la partie Data Science** : traitement statistique, modélisation et prédiction via Python.
- **Développement front-end** : création des pages React.js interactives, intégration de graphiques dynamiques avec Recharts et liaison avec l'API.
- **Évaluation et comparaison des modèles** : calcul de métriques (RMSE, R^2 , précision, recall) et affichage des résultats pour l'utilisateur.

3. Les tâches périphériques

Au cours de mon stage, j'ai également réalisé des tâches complémentaires qui ont enrichi ma compréhension globale du projet et de l'environnement professionnel :

- **Documentation technique et rapport** : rédaction d'un guide d'utilisation et d'un rapport de stage incluant les étapes du projet, les choix techniques et les analyses réalisées.
- **Amélioration de l'interface utilisateur** : optimisation de l'UI avec Tailwind CSS pour rendre l'expérience utilisateur plus fluide et intuitive.

Ces activités ont permis de consolider mes compétences, tant sur le plan technique que méthodologique, et m'ont préparé à la réalisation autonome de projets de data science et développement web full-stack.

III - ACQUIS DU STAGE

Au cours de ce stage, j'ai beaucoup appris. Les apports tirés de cette expérience professionnelle peuvent être regroupés autour de trois axes principaux : **les compétences acquises, les difficultés rencontrées et les solutions apportées, et la vie en société.**

1. Compétences acquises

Ce stage m'a permis de consolider et de développer un ensemble de connaissances et de savoir-faire directement liés au métier de **data scientist**. Plus précisément, j'ai acquis :

- **Compétences techniques :**
 - Manipulation et nettoyage de données avec pandas et numpy.
 - Création de visualisations pertinentes avec matplotlib et seaborn.
 - Mise en place de modèles de machine learning simples avec scikit-learn (régression logistique, arbres de décision).
 - Utilisation de requêtes SQL pour extraire des données dans une base **MySQL**.
 - Déploiement d'un petit tableau de bord interactif avec **Plotly Dash** pour la restitution des résultats.
- **Compétences méthodologiques :**
 - Approche structurée de la résolution de problèmes (collecte → nettoyage → analyse → modélisation → interprétation).
 - Bonnes pratiques de versionnage avec **GitHub** pour assurer un travail reproductible.
 - Documentation des résultats et communication des analyses à l'équipe.
- **Compétences personnelles :**
 - Développement de l'autonomie dans la recherche de solutions (tutoriels, documentation, forums techniques).
 - Amélioration de l'organisation du travail et de la gestion du temps pour respecter les délais impartis.

2. Difficultés rencontrées et solutions apportées

Comme dans tout projet technique, certaines difficultés sont survenues :

- **Problèmes de qualité des données :** J'ai rencontré des datasets avec beaucoup de valeurs manquantes et des incohérences dans les formats.

Solution : mise en place de scripts de nettoyage automatisés et apprentissage des méthodes d'imputation pour combler les données manquantes.

- **Difficulté à choisir le bon modèle de prédiction :** Les premiers modèles donnaient de mauvaises performances.

Solution : apprentissage de la validation croisée, test de plusieurs modèles (baseline simple puis modèles plus complexes) et réglage des hyperparamètres.

- **Gestion du temps :** Les tâches prenaient parfois plus de temps que prévu.

Solution : utilisation d'un planning hebdomadaire et communication régulière avec le tuteur pour prioriser les tâches.

Ces défis ont renforcé ma capacité à garder mon calme, à m'adapter et à progresser rapidement dans un environnement exigeant.

3. La vie en société

Mon stage au sein de cette entreprise a été très formateur, non seulement sur le plan technique mais aussi humain. J'ai pu observer le fonctionnement global d'une organisation où les différents départements collaborent pour atteindre un objectif commun : la valorisation des données. J'ai compris que le data scientist ne travaille pas isolément, mais en interaction avec les équipes techniques et métiers (marketing, IT, direction).

L'ambiance au sein de l'équipe était conviviale et professionnelle à la fois. Les échanges étaient simples et directs, ce qui favorisait la bonne circulation de l'information et une meilleure efficacité collective. Cette expérience m'a appris l'importance de la communication et de l'esprit d'équipe pour mener à bien un projet de data science.

CONCLUSION

Ce stage a constitué une étape clé dans mon processus de formation en **Data Science**, en me permettant de passer de la théorie à la pratique et de mettre en application les notions apprises au cours de mon cursus. J'ai pu explorer l'ensemble du cycle de traitement des données : de leur collecte à leur modélisation, en passant par leur nettoyage et leur visualisation. Cette expérience m'a offert une vision plus concrète du rôle du data scientist dans un environnement professionnel et m'a permis de comprendre les défis liés à l'exploitation des données en entreprise.

Sur le plan des acquis, ce stage m'a permis de développer des **compétences techniques solides** (analyse de données, modélisation prédictive, visualisation) et des **compétences méthodologiques** (planification, documentation, validation de modèles), tout en renforçant des qualités personnelles essentielles comme l'autonomie et la rigueur. Il a également répondu à la problématique posée en introduction : comment transformer les données brutes en informations exploitables pour appuyer la prise de décision. Le projet final réalisé a démontré qu'il est possible, à partir d'un dataset initialement bruité, de produire un modèle de prédiction performant et utile pour l'entreprise.

En termes de bilan, ce stage a été extrêmement formateur et a confirmé mon choix d'orientation professionnelle vers l'ingénierie en Data Science. Il m'a également sensibilisé aux contraintes réelles du monde du travail : respect des délais, communication claire avec les parties prenantes, documentation rigoureuse des résultats.

Enfin, je formulerais quelques recommandations pour les futurs stagiaires : se former à l'avance aux outils de base (Python, SQL, bibliothèques de Machine Learning), apprendre à poser les bonnes questions au début du projet pour clarifier les objectifs, et documenter systématiquement les travaux pour faciliter la transmission. Pour l'entreprise, il pourrait être intéressant de mettre à disposition davantage de données historiques afin de permettre des analyses plus riches et plus robustes.

Ce stage marque donc pour moi un **tournant dans ma formation**, car il m'a non seulement permis d'acquérir des compétences concrètes, mais aussi de mieux cerner le rôle que je souhaite jouer en tant que futur ingénieur data scientist : un professionnel capable de transformer les données en valeur et de contribuer activement à la prise de décision dans un contexte organisationnel.

REFERENCES

- **React.js Documentation** : <https://react.dev/>
- **Tailwind CSS Documentation** : <https://tailwindcss.com/docs>
- **Spring Boot Documentation** : <https://spring.io/projects/spring-boot>
- **Spring Data JPA Reference** : <https://spring.io/projects/spring-data-jpa>
- **MySQL Documentation** : <https://dev.mysql.com/doc/>
- **Python Official Documentation** : <https://docs.python.org/3/>
- **Pandas Documentation** : <https://pandas.pydata.org/docs/>
- **NumPy Documentation** : <https://numpy.org/doc/>
- **Matplotlib Documentation** : <https://matplotlib.org/stable/contents.html>
- **Seaborn Documentation** : <https://seaborn.pydata.org/>
- **Scikit-learn Documentation** : <https://scikit-learn.org/stable/>
- **Flask Documentation** : <https://flask.palletsprojects.com/en/stable/>
- **Kaggle – Adult Census Income Dataset** : <https://www.kaggle.com/datasets/uciml/adult-census-income>
- **GitHub** (versionning & collaboration) : <https://github.com/>
- **Postman** (tests API REST) : <https://www.postman.com/>
- **FreeCodeCamp** (apprentissage en ligne) : <https://www.freecodecamp.org>
- **W3Schools** (référence rapide pour HTML/CSS/JS) : <https://www.w3schools.com/>
- **GeeksforGeeks** (tutoriels & exemples de code) : <https://www.geeksforgeeks.org/>
- **Figma** (prototypage UI) : <https://www.figma.com/>

Table des matières

DEDICACES	i
REMERCIEMENT	ii
INTRODUCTION	iii
I – PRESENTATION DE L’ENTREPRISE	1
I – 1 - Situation de l’entreprise	1
I - 2 - Adresse complète de l’entreprise	1
I – 3 - Historique de l’entreprise	1
I-4- Indication du secteur d’activité	2
I-5- Nombre du personnel	2
I-6- Organigramme hiérarchique	3
II – LES TRAVAUX EFFECTUES	6
1. Les outils et technologies mis à ma disposition.....	6
2. Les missions du poste occupé	6
3. Les tâches périphériques	13
III - ACQUIS DU STAGE	14
1. Compétences acquises	14
2. Difficultés rencontrées et solutions apportées	14
3. La vie en société.....	15
CONCLUSION	16
REFERENCES	17
Table des matières.....	18