**1. Introduction générale et définition du problème**

* **Contexte et justification**

Le problème majeur auquel chaque entreprise est confrontée chaque jour est celui d’avoir un profit maximal, autant que possible. C’est d’ailleurs légitime puisque c’est un principe de rationalité. Pour ce faire, l’entreprise oriente la majorité de ses actions de sorte à fidéliser ses clients, en leur offrant des services de qualité. Cependant, un aspect très important est négligé par ces entreprises : la main d’œuvre. En effet, en économie, la main d’œuvre représente l’un des quatre facteurs de production et occupe en général une place importante.

* **Problématique**

Dans un contexte économique de plus en plus compétitif, la fidélisation des employés est devenue un enjeu crucial pour les entreprises. Un personnel engagé et loyal contribue non seulement à la productivité, mais aussi à la culture organisationnelle. L'attrition des employés, ou le taux de rotation du personnel, représente un défi majeur pour de nombreuses entreprises, impactant non seulement la continuité des opérations, mais également la culture organisationnelle et la performance globale. Comprendre les raisons derrière cette attrition est essentiel pour développer des stratégies de rétention efficaces et réduire les coûts associés à la perte de talents. Ainsi, la question se pose sur l’Attrition des employés de « strategics consulting ». Pour un suivi cohérent de cet aspect de la réflexion, il convient de répondre aux interrogations spécifiques suivantes :

* Quels facteurs contribuent le plus à l'attrition des employés ?
* Quel type de mesures l'entreprise doit-elle prendre pour fidéliser ses employés ?
* **Méthodologie**

Grâce à l'analyse de données et à l'utilisation d'algorithmes prédictifs, nous allons identifier des signaux précurseurs de départ, tels que la baisse d'engagement ou des changements dans la performance.  Grâce à notre analyse, nous allons examiner toutes les variables à travers quelques graphiques et en déduire dans notre analyse exploratoire. Après notre exploration, nous allons construire quelques modèles basés sur les variables à portée de main et d'exclure de la base quelques variables non pertinentes. Et enfin, nous avons l'intention de construire un modèle qui peut prédire si un employé est sur le point de quitter. Ainsi, en anticipant ces risques, les entreprises peuvent mettre en place des stratégies ciblées pour améliorer la satisfaction et la rétention des employés, créant ainsi un environnement de travail plus stable et productif.

**2. Résultat de l’exploration des données**

Afin de mieux comprendre la structure et les caractéristiques des données, et d'identifier d'éventuelles anomalies ou incohérences avant de faire des prédictions nous avons fait des statistiques exploratoires sur nos données.

On peut retenir les principales caractéristiques suivantes :

* **Structure du jeu de données :** 1470 observations (lignes), 35 caractéristiques (variables)
* **Données manquantes :**Il existe de valeurs manquantes pour la modalité « DistanceFromHome ».
* **Données non pertinentes :** certaines variables sont non informatives dans notre jeu de données, il est importatnt de les exclurent de la base avant toute prédiction. Il s’agit notamment de : «EmployeeCount», «StandardHours», «Over18\_Index» et « EmployeeNumber ».
* **Type de données :** Nous avons seulement deux types de données : objet et entiers.
* **Étiquette :** L’étiquette dans notre jeu de données est "Attrition", et nous souhaitons comprendre pourquoi les employés quittent l’organisation.
* **Jeu de données déséquilibré :**1237 (84 % des cas) employés ne quittent pas l’organisation, alors que 237 (16 % des cas) la quittent. Cela rend notre jeu de données déséquilibré, car plus de personnes restent que celles qui partent.

Ce jeu de données étant déséquilibres, il est important d’etre prudent dans les tirages à faire pour la prédiction.

* **Statistiques descriptives univariées**

Les statistiques descriptives univariées montrent que :

* Analyse par sexe :
  + Les distributions sont **similaires**. L'âge moyen des femmes est de 37,33 ans, tandis que celui des hommes est de 36,65 ans.
  + Pour les personnes qui sont restées dans l'organisation, les niveaux de satisfaction ( dans le travail personnel) sont quasiment identiques. Cependant, parmi celles ayant **quitté l'organisation**, les femmes ont un niveau de satisfaction inférieur à celui des hommes. Bien plus dans les deux groupes, on note des individus insatisfait de leur environnement de travail ; ce qui peut etre un facteur d’attrition.
  + Les salaires moyens sont presque équivalents, avec une moyenne de 6 380,51 pour les hommes et 6 686,57 pour les femmes. On ne soupconne pas a ce niveau que le salaire soit un facteur d’attrition.
  + On observe un nombre plus élevé d'hommes dans trois départements, tandis que les femmes sont majoritaires dans le département **Recherche et Développement**.
* Impact du revenu sur l’attrition
  + Nous pouvons voir d'énormes différences dans chaque département selon le **statut d'attrition.** (explication)
  + semble que plus la satisfaction au travail est faible, plus l'**écart** selon le statut d'attrition est large en termes de niveaux de revenu.
  + Plus de 54 % des travailleurs qui ont quitté l'organisation ont travaillé **des heures supplémentaires** ! Cela pourrait-il être une raison pour laquelle les employés partent ?
* **Statistiques descriptives bivariées**
* Variables quantitatives

Les analyses sur les variables qualitatives ont etés faits a partir de la matrice de corrélation. On remarque que :

* + Plus le **nombre total d'années travaillées** est élevé, plus le **revenu mensuel** d'un employé est élevé.
  + Plus le **pourcentage d'augmentation de salaire** est élevé, plus la **note de performance** est élevée.
  + Plus les **années passées avec le manager actuel** sont élevées, plus les **années depuis la dernière promotion** sont élevées.
  + Plus l'**âge** est élevé, plus le **revenu mensuel** est élevé.
* Variables qualitatives

(Il faut calculer le V de Cramer)