

30 janvier 2022

RAPPORT STATISTIQUE : RH & DATA

SOMMAIRE

RESUME	2
INTRRODUCTION	3
PROBLEMATIQUE	5
JEU DE DONNEE	7
OBJECTIFS	10
PRE-TRAITEMENTS	11
STATISITIQUE DESCRIPTIVE	12
MODELISATION	22
CONCLUSION	26

Parcours Data analyste -Openclassroom

PROJET 8 : COMMUNIQUEZ VOS RESULTATS

RESUME

PREVENIR L'ATTRITION DES TALENTS

La clé du succès dans toute organisation est d'attirer et de retenir les meilleurs talents. En tant que futur analyste RH dans mon entreprise , l'une de mes tâches consistera à déterminer les facteurs qui retiennent les employés dans mon entreprise et ceux qui par contre inciteraient certains à partir.

J'ai besoin de savoir quels facteurs prendre en compte pour éviter la perte de bons talents. En d'autres mots comment gérer au mieux les différentes sources de performance pour retenir au mieux mes collaborateurs ?

Concepts clés: Attrition, performance, motivations , salaire ,qualité de vie au travail

INTRODUCTION

Selon le dictionnaire en ligne , [www. dictionary.com](https://www.dictionary.com/) ¹, l'attrition fait référence à une réduction de la force ou à un affaiblissement d'un système quelconque.

En ressources humaines , le terme renvoie à une réduction de l'effectif causée par un départ à la retraite ou une démission.

Ces deux actions ne relèvent pas de la même nature car l'une est involontaire et l'autre initiée par le travailleur lui-même.

C'est le cas du départ volontaire du travailleur qui va nous intéresser dans ce rapport statistique.

Dans le domaine des ressources humaines , l'attrition fait partie intégrante de la gestion prévisionnelle des emplois et des compétences (GPEC) qui consiste à planifier la gestion de son effectif d'employés en initiant des actions pour une meilleure performance de l'entreprise.

En règle générale , l'attrition induit une perte de talents et peut causer à long terme un manque de compétitivité de l'entreprise si des actions idoines ne sont pas réalisées.

Si tant est que le départ du travailleur peut être soit volontaire soit involontaire notamment un départ à la retraite ou un licenciement , il n'en demeure pas moins que c'est à l'organisation de conjuguer les efforts pour retenir ses talents .

¹ <https://www.dictionary.com/>

L'entreprise à travers ses différentes politiques doit mettre en oeuvre un plan d'action efficace afin de réduire l'attrition qui est volontaire en favorisant une qualité de vie au travail² (QVT).

Depuis maintenant une décennie , la qualité de vie au travail est présente et prend une place importante dans la gestion des ressources humaines du fait de plusieurs evenements qui ont des impacts sur le travail , le travailleur et le sens donné au travail.

Fort de ces impacts , l'attrition si elle n'est pas jugulée peut influencer la productivité de l'entreprise car manquer de talents peut compromettre sérieusement les missions d'une entreprise .

Notre rapport de projet du parcours data analyste a choisi cette thématique pour étudier des données de ressources humaines afin de prédire le taux d'attrition et trouver des facteurs pouvant réduire celui-ci.

Il sera donc question dans ce rapport de mettre en rapport les données internes de l'entreprise pour favoriser la qualité de vie au travail et réduire de facto l'attrition chez le salarié.

² <https://www.anact.fr/themes/qualite-de-vie-au-travail>

PROBLEMATIQUE

Avec la généralisation du numérique dans toutes les sphères de notre société et plus particulièrement dans les entreprises , gérer , diriger ou conduire le management d'équipe peut s'appuyer sur des données foisonnantes.

On peut distinguer deux types de données sur lesquelles les décideurs ou opérationnels d'entreprise peuvent s'appuyer pour développer et mettre en oeuvre leurs stratégies de recrutement.

D'une part les données internes de l'entreprise sont celles produites par le travailleur , son environnement et l'entreprise elle-même. Il s'agit notamment du smartphone , des données de l'e-réputation et des performances de l'organisation .

D'autre part les données externes sont celles sur l'état du marché de l'entreprise , ses concurrents et les données d'autres parties prenantes.

Ce sont entre autre les CRM ,les législations , les études de marché.

Pour notre rapport , nous nous intéresserions aux données internes notamment les différents indicateurs de pilotage des ressources humaines qui sont autant de leviers pour bien manager l'effectif de son entreprise.

Face aux performances exigées au travailleur celui-ci se doit d'être écouté pour bien se sentir au travail notamment selon le concept de bonheur au travail développé récemment .

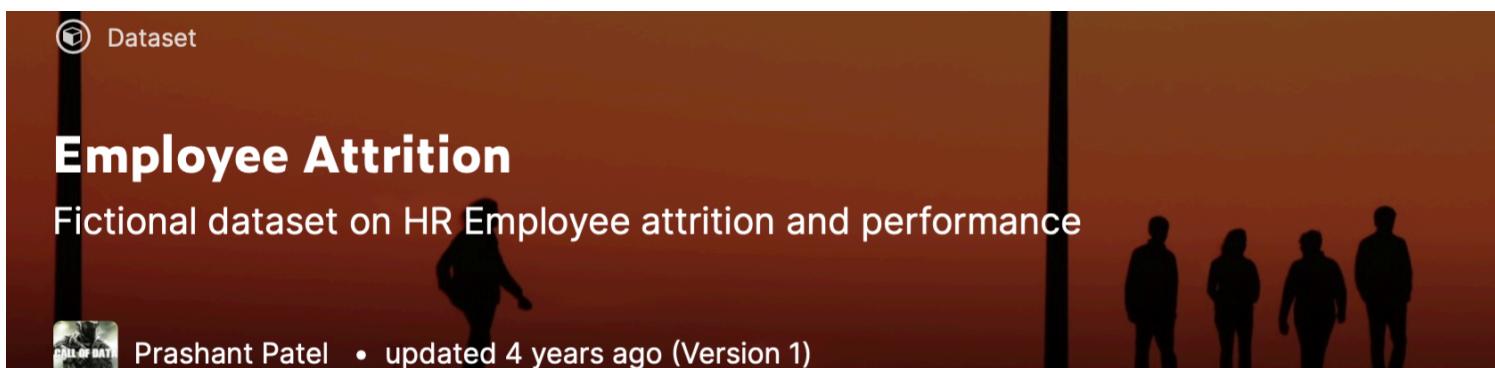
Notre problématique veut adresser le problème de l'attrition eu égard aux données internes de l'entreprise.

C omment partant des données et indicateurs de l'entreprise , le manager des ressources humaines peut motiver les travailleurs à rester durablement dans l'entreprise?

Pour répondre à une telle difficulté , nous nous appuierons sur les facteurs motivationnels du travailleur notamment les conditions de travail générales et spécifiquement sur la rémunération.

JEU DE DONNEE

Pour ce rapport , nous avons choisi l'option C qui demande de faire l'interface entre un métier et des données afférentes à celui-ci .



Capture d'écran du data set sur Kaggle

Nous avons fait ce choix en rapport avec notre projet professionnel dont les jalons sont visibles dans le métier de gestion des ressources humaines.

En effet , nous sommes ingénieurs pédagogiques et enseignants en gestion des ressources humaines depuis 6 ans .

Nous avons donc choisi de travailler sur un jeu de données trouvé sur le site Kaggle³ , un répertoire de jeux de données dans le domaines des data sciences.

Notre jeu de donnée est un échantillon fictif produit par un data analyste du domaine des GRH.

Il est constitué d'un fichier csv comportant 1470 individus et 35 colonnes.

³ <https://www.kaggle.com/>

Les variables définissent plusieurs indicateurs du domaine de pilotage des ressources humaines dont les champs principaux sont :

- * La qualité de vie au travail (QVT)
- * La performance de l'entreprise et du travailleur
- * La productivité

LES VARIABLES

Fort d'une trentaine de variables , nous avons choisi de ne pas exploiter toutes les varioles du jeu de donnée.

En effet , nous avons pris le soin de catégoriser les variables en trois grands indicateurs et retenir celles qui étaient afférentes aux motivations du travailleur et en lien directe avec ces conditions de travail.

Nous avons donc retenu cet ensemble d'attributs :

- * **L'âge** : Variable qui indique les âges des travailleurs
- * **Attrition** : C'est la variable cible qui indique si oui ou non le travailleur compte partir ou rester.
- * **EnvironnementSatisfaction** : variable quantitative qui mesure le taux de satisfaction du cadre de travail
- * **Gender** : Variable catégorielle qui informe sur le genre du travailleur

- * **JobInvolvement** : variable quantitative indiquant l'engagement du travailleur au sein de l'entreprise et dans ses missions
- * **MaritalStatus**: variable qualitative indiquant le statut marital du travailleur
- * **MonthlyIncome** : variable quantitative qui donne le salaire mensuel de l'individu
- * **PerformanceRating**: variable quantitative qui indique un score de performance du travailleur noté sur échelle
- * **RelationshipSatisfaction**: variable quantitative indiquant la coopération au sein de l'entreprise et les liens sociaux à l'intérieur de celle-ci
- * **TrainingTimelastYear** : variable quantitative qui marque le nombre de sessions de formation suivi par le salarié durant l'année passée
- * **WorkLifeBalance** : variable qui informe sur l'équilibre de vie du travailleur noté aussi sur une échelle.

Cet ensemble d'attributs forme notre échantillon qui se compose de variables qualitatives et quantitatives en plus de la variable cible de type booléenne qu'est l'Attrition.



Image source google image

OBJECTIFS

L'objectif de notre étude est de prédire le départ d'un de nos collaborateurs en nous basant sur différents facteurs .

Cette prédiction doit participer à réduire le taux d'attrition

Mais avant la phase de prédiction , nous ferons une analyse descriptive univariée et bivariée .

Ensuite ,nous réduirons les dimensions de notre échantillon en y appliquant une analyse en composante principale (ACP) afin de préparer la modélisation adéquate

Enfin , nous ferons une **régression logistique** avec comme variable cible (attrition) pour prédire le taux d'attrition , action permettant de révéler des pistes de solution de la problématique métier .

PRE-TRAITEMENTS

Notre échantillon téléchargé sur kaggle ne comportait pas de valeurs irrégulières ou de types NaN.

J'ai donc commencé l'analyse de l'échantillon en faisant de la statistique descriptive afin de comprendre la table.

Pour décrire au mieux notre échantillon et pouvoir capter des insights , nous avons eu recours au logiciel de data visualisation tableau software et du langage python et ses librairies dédiés aux data science

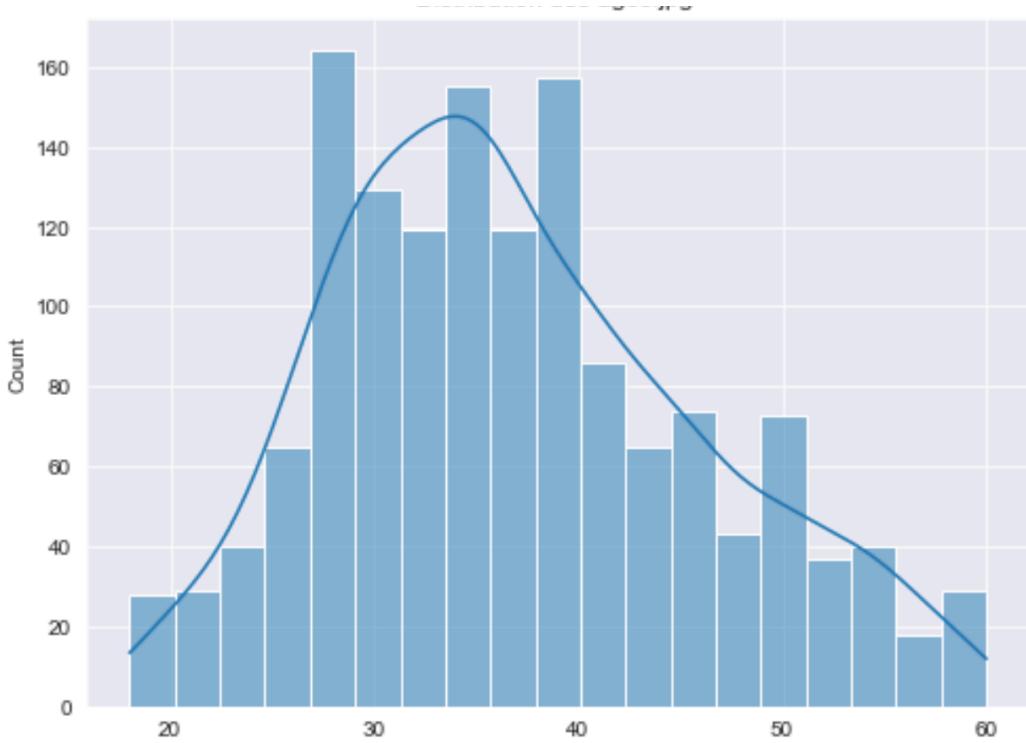
STATISTIQUE DESCRIPTIVE

Pour nous orienter dans notre exploration de l'échantillon , sous avons essayé de répondre à des questions selon notre connaissance métier.

Notre problématique abordant les concepts de l'attrition, de la qualité de vie au travail et des données de l'entreprise , nous avons analysé les variables de l'âge , du salaire mensuel et du statut marital.

Selon la distribution des âges de l'entreprise , la majorité des employés ont un âge moyen de 35 ans et un revenu moyen de 2340 € (Annexe) .

Cet âge constitue une forme de maturité qui se traduit par une stabilité du foyer.



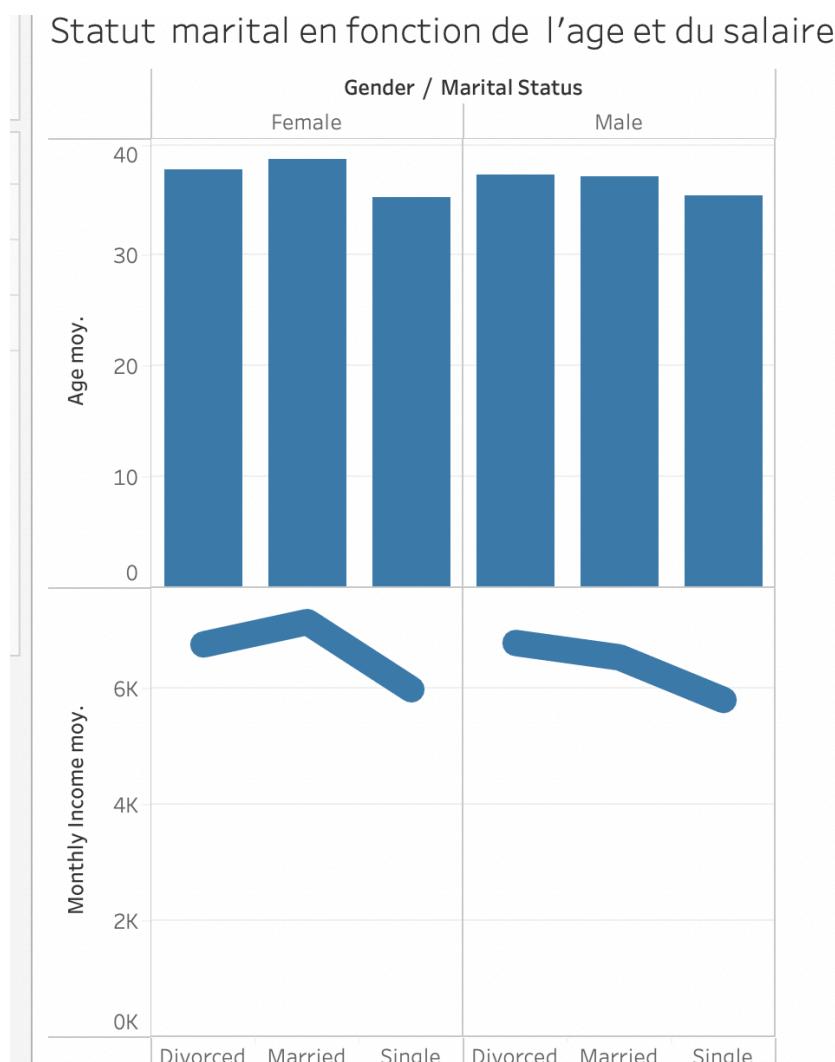
Pyramide des âges de l'échantillon

Avec cette allure centrée et symétrique , la distribution des âges montre une pluralité de classes d'âge constituant la force de travail de l'entreprise.

L'âge moyen de 35 ans est un élément important pour planifier une gestion optimale de l'attrition dans cette entreprise; En effet les travailleurs à cet âge visant la stabilité ou une considération seront plus réceptifs aux éléments motivationnels de l'entreprise .

Ces éléments motivationnels peuvent être constitué de motivations intrinsèques et des motivations extrinsèques.

Les motivations intrinsèques sont inhérentes aux plaisirs du travailleur sans aucune contrainte quand la motivation extrinsèque fait référence aux contraintes .



Graphique de l'âge, statut marital et du salaire

Pour tester la question de la motivation , nous avons analysé le salaire de l'entreprise pour comprendre son influence sur l'attrition.

Il apparait que le genre n'influence pas suffisamment sur le revenu et que le statut marital notamment les mariés ont un meilleur traitement salarial.

Ce qui est compréhensible du fait que le mariage confère une sécurité aux employeurs et les avantages sociaux se greffent à ce statut pour amoindrir les taxes ou impositions dont le fait de vivre seul peut engendrer.

Cependant les femmes connaissent une forte tendance à la baisse du revenu quelques soient leur statut marital.

Ce fait est important à signifier car il est notoirement connu que les femmes subissent les coups de la vie maritale dans leur foyer plus que les hommes.

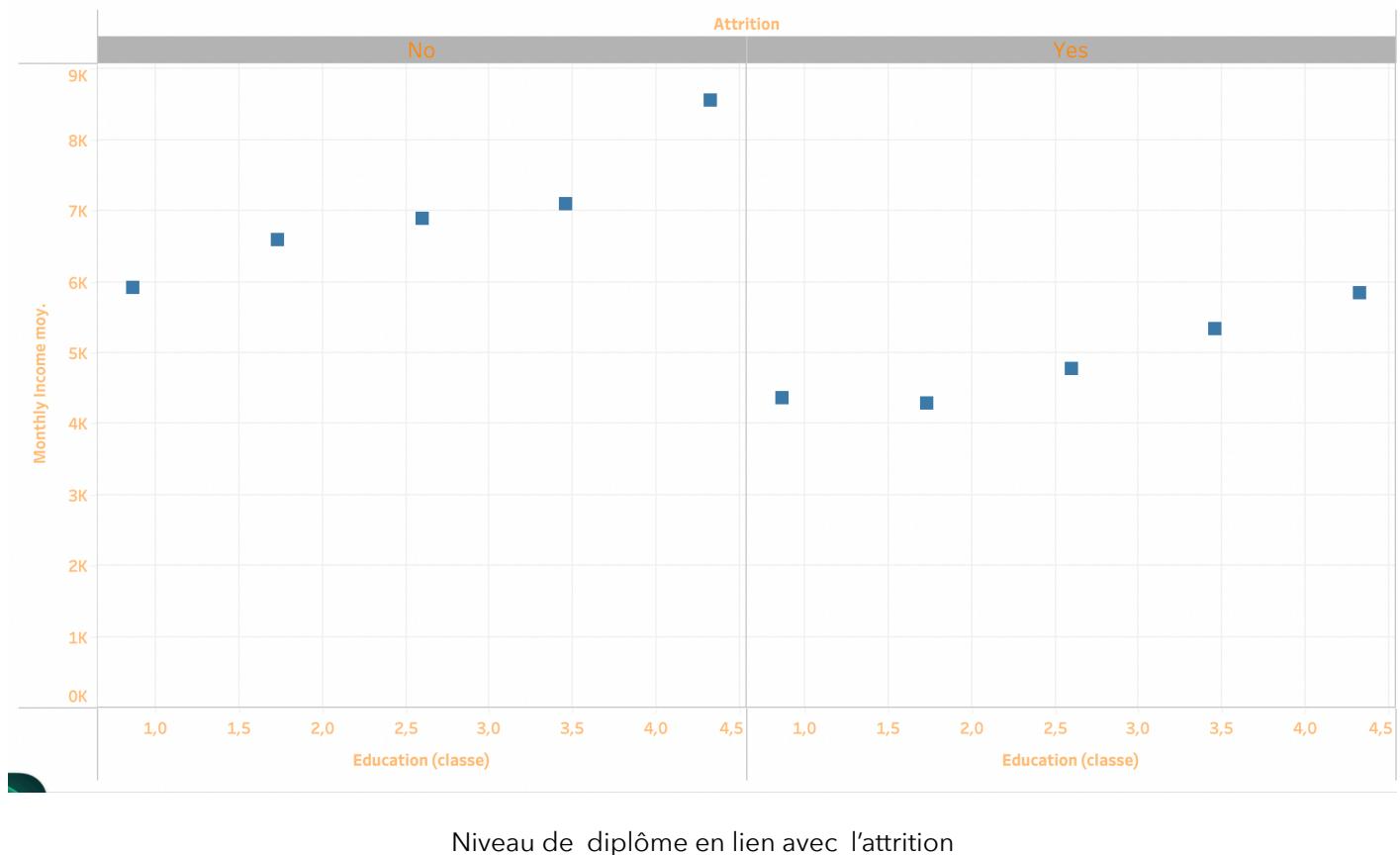
En effet du fait des grossesses et autres événements dans la vie de couple les femmes connaissent un moindre succès que les hommes dans leur carrière.

Une des raisons à ce déséquilibre c'est le diplôme d'accès à certaines fonctions ou aux premières expériences.

Sur ce graphe , l'abscisse représente le niveau de diplôme et se détaille comme suit : 1= niveau avant Bac , 2 = Niveau Baccalauréat, 3 = Niveau Bachelor , 4 = niveau master et 5= niveau Doctorat(PhD) .

On peut voir que dans les deux clauses de notre étude qu'il y'a la même tendance d'un meilleur salaire selon le niveau de diplôme. Plus on est diplômé plus on un plus grand salaire qu' ce soit pour les femmes ou pour les hommes.

le diplôme est il important pour l'attrition?



Niveau de diplôme en lien avec l'attrition

Cependant le niveau d'augmentation est plus grand chez les hommes que chez les femmes.

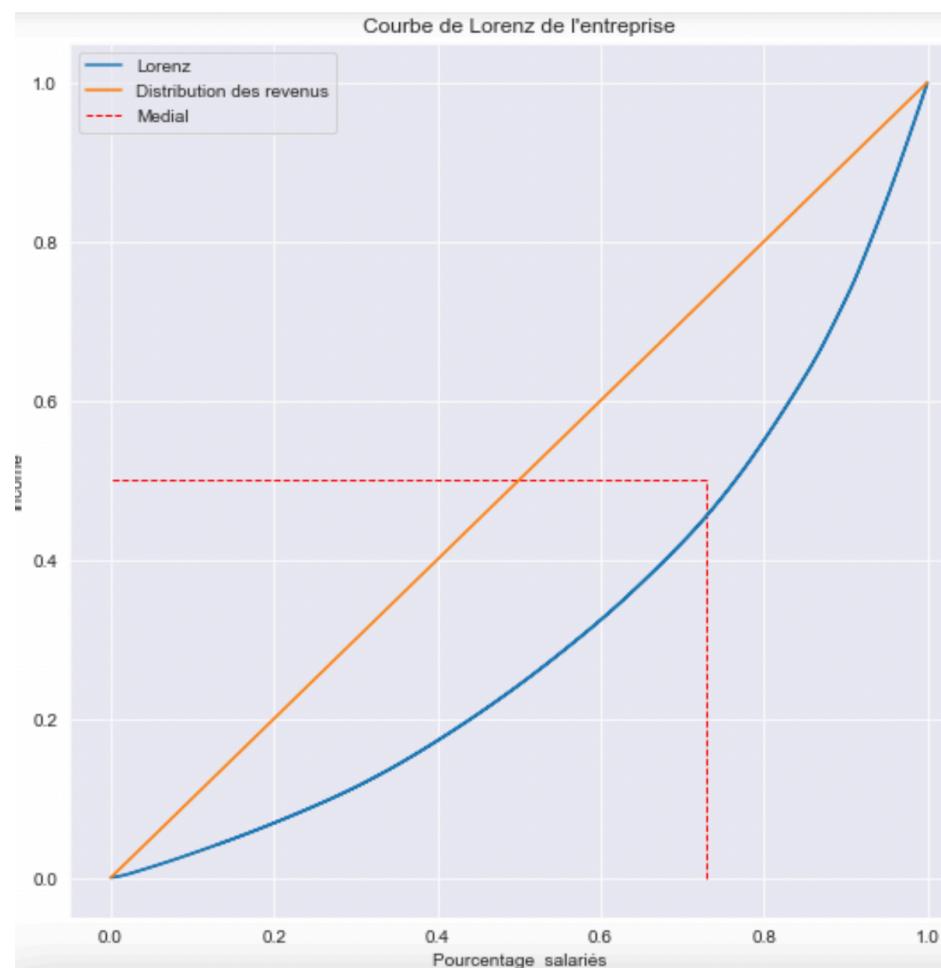
En outre l'évolution du revenu est nettement moins marqué chez les femmes de sorte qu'il est difficile de voir une différence entre celles qui n'ont pas eu le baccalauréat et celle en possédant.

Il est impérieux de tenir compte de ces indicateurs si l'on veut comprendre les facteurs entraînant l'attrition dans l'entreprise.

Fort de ces indicateurs , nous avons réalisé un tableau de bord capable d'aider à surveiller les indicateurs et à pouvoir trouver les leviers pour un meilleur pilotage des ressources humaines.

Nous avons ajouter d'autres indicateurs comme les fonctions occupées et les sessions de formations suivies par les employés afin de comprendre leurs motivations .

L'analyse descriptive montre que les sessions de formation confèrent un sentiment de considération au travailleur qui le voit comme l'intérêt que lui porte son organisation . Ainsi , il faut veiller en plus de l'équité salariale que les travailleurs soient souvent formés et leurs compétences mises à jour .



Courbe de Lorenz de l'échantillon

Le tableau de bord indique en plus des revenus , des diplômes , du temps de sessions de formation suivies , les fonctions qui se sentent à l'aise avec le cadre du travail.

La fonction de sales tient le peloton de tête suivi des scientifiques. mais il est à noter que c'est la fonction des ressources humaines qui occupe le niveau le plus bas quand au poste occupé dans l'entreprise.

Mais ce métrique est à relativiser car elle se base sur l'effectif des différentes services et départements.Les commerciaux occupent le premier rang parce qu'ils sont les plus nombreux et vice versa pour le département des ressources humaines.

Ce tableau dd bord nous sert de transition pour passer de la data visualisation de Tableau software à l'analyse d'autre indicateurs.

Par ailleurs , nous avons analysé par le biais des mesures de concentration la répartition égalitaire du salaire dans notre échantillon . Pour ce faire nous avons calculé l'indice de Gini mesurant la concentration des revenus proportionnellement aux catégories des salariés.

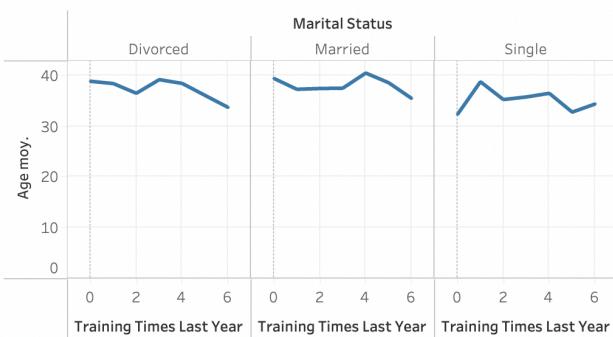
Avec un indice de Gini de 0,37 , il apparait que notre échantillon traduit une répartition inégalitaire , ce qui est e nature à crisper le personnel ou à démotiver.

L'équité étant une valeur de bien-être social , il est important d'ajuster les revenus et les avantages liés pour mieux motiver son personnel.

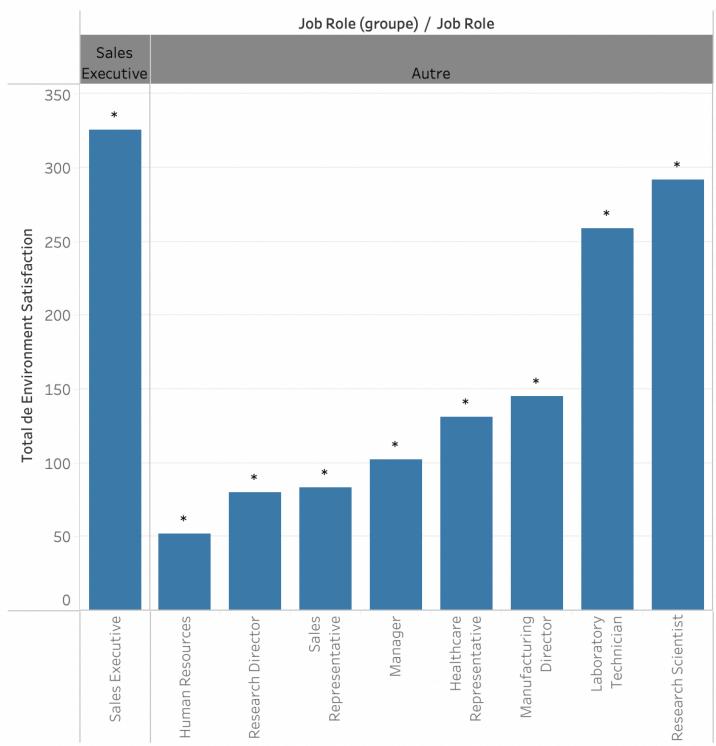
Au titre des analyses multi-variées , nous avons calculé la correlation entre les variables de notre échantillon. Il apparait que c'est la variable "MonthlyIncome " .qui est corrélée avec la variable "Age".

KPI (1) du manager des RH

L'action de formation implique t'elle davantage ?



fonctionnelles et statut marital et cadre de vie



le diplome est il important pour l'attrition ?



Tableau de bord du manager RH

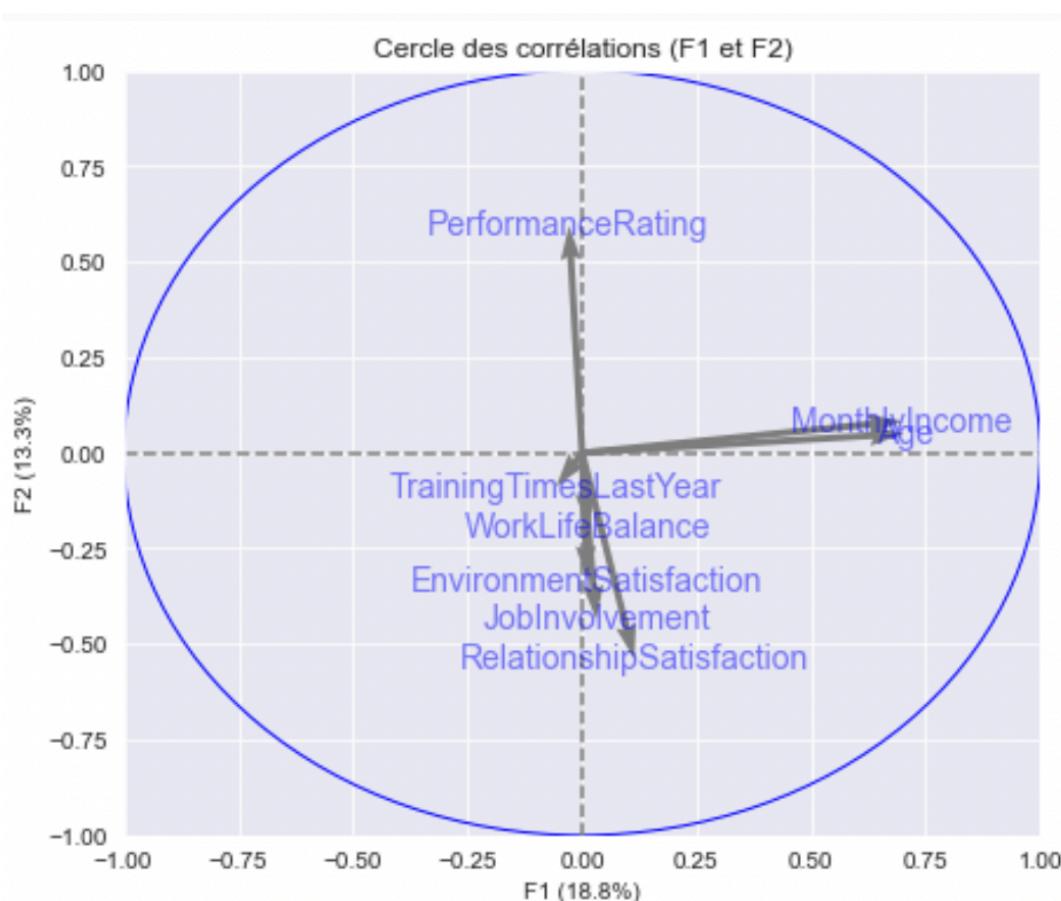
Nous obtenons un score de 0,5 de corrélation entre l'âge et le revenu mensuel. Nous pouvons donc en déduire qu'il y'a une corrélation linéaire entre ces deux variables .Il es patent que l'âge faisant un plus grand nombre d'expériences professionnelles dans une entreprise , le salaire s'en trouve être amélioré eu égard

Par conséquent , l'entreprise gagne à conserver ou retenir longtemps ses talents .

Car plus ceux ci sont longtemps sensibilisés à la culture d'entreprise , plus ils s'y sentiront bien et seront de facto productifs.

Une bonne prise en compte des besoins des travailleurs à travers une bonne équité du traitement salarial peut réduire le taux de turnover du personnel et permettre une bonne compétitivité de l'entreprise.

Après cette statistique descriptive , nous allons continuer à analyser notre échantillon en réduisant ces dimensions en vue de préparer la régression logistique que nous envisageons de faire.



ACP/Cercle de correlation de l'échantillon

L'analyse de composante principale nous permettra de comprendre les liaisons entre variables de notre échantillon;

Sur le nuage des points les individus sont représentés selon les deux axes et étalés sur les deux axes principaux.

L'analyse en composantes principales a révélé l'importance de certaines variables a forger les composantes principales. Elle met en exergue les variables d'âge et monthly income comme importante pour la composante principale F1.

Mais aussi le taux de performance est aussi importante pour la composante principale F2.

Quant aux autres variables elles contribuent très peu à F2 car mal représentées à l'exception de relationshipsatisfaction qui est anti -corrélée à F2.

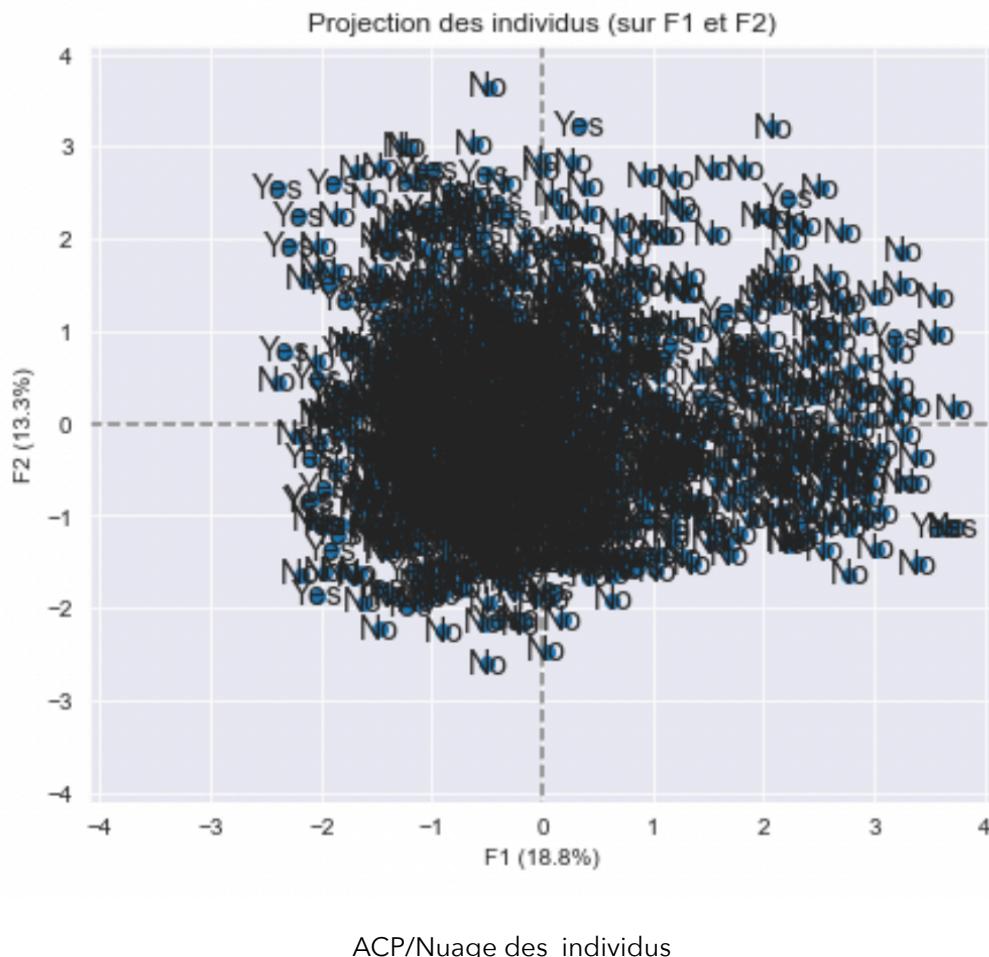
L'ACP nous permet de comprendre l'importance des variables en lien avec la variable cible de l'attrition.

Nous allons appliquer une régression logistique pour prédire le taux d'attrition ,

Cette variable étant catégorielle avec deux modalités de Yes et no;

La clause No représente notre variable positive puisqu'elle indique une attrition négative qui signifie que le travailleur ne part pas de l'entreprise. Ainsi la modalité Yes indique la classe négative qui indique le départ du salarié.

MODELISATION



LE SALAIRE REDUIT LE TAUX D'ATTRITION

Avant la régression logistique , nous scindons notre jeu de donnée en jeu d'entraînement et en jeu de test.

Ensuite nous entraînons le modèle pour ensuite tester sa performance.

Nous décidons d'utiliser la librairie Statsmodel afin de révéler les variables significatives ou non . Cela nous permettra de réaliser une bonne performance de notre modèle pour mieux prédire.

	precision	recall	f1-score	support		
=====	0	0.00	0.00	49		
Dep. \ Model:	1	0.83	1.00	245		
Model						
Link F Method	accuracy		0.83	294		
Date:	macro avg	0.42	0.50	294		
Time:	weighted avg	0.69	0.83	294		
No. It						
Covari						
=====						
	coef	Scores de la prediction	P> z	[0.025	0.975]	
Intercept	-1.3538	0.793	-1.707	0.088	-2.909	0.201
MonthlyIncome	9.381e-05	2.38e-05	3.946	0.000	4.72e-05	0.000
PerformanceRating	0.0078	0.200	0.039	0.969	-0.385	0.401
Age	0.0333	0.010	3.442	0.001	0.014	0.052
RelationshipSatisfaction	0.1002	0.067	1.504	0.133	-0.030	0.231
TrainingTimesLastYear	0.1398	0.058	2.414	0.016	0.026	0.253
WorkLifeBalance	0.2276	0.100	2.275	0.023	0.031	0.424
=====						

Modélisation de type régression logistique

La modélisation nous indique deux variables non significatives pour notre modèle de prédiction en préparation puisqu'elles sont supérieures à notre seuil de test de 0.05.

Ce sont les variables de taux de performance , de la qualité relationnelle entre travailleur.

Nous les supprimerons donc de la table du modèle de prédiction pour mieux permettre une bonne performance du modèle.

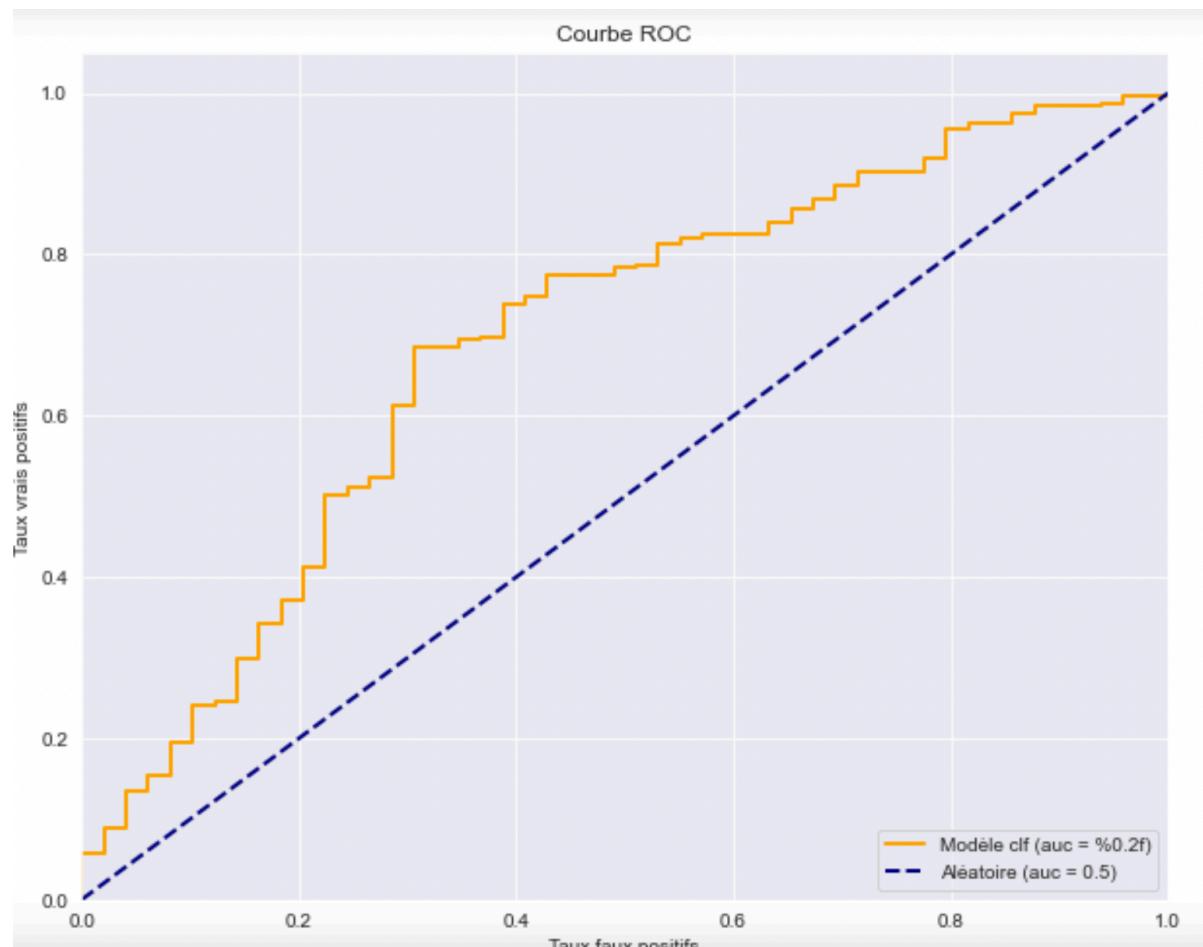
Avec un score de 0,83 , notre modèle est totalement satisfaisant .Nous avons un modèle précis à 83% en nous fiant au accuracy score .

Nous avons un score très proche de 1 donc nous pouvons dire que nous avons une bonne performance de notre algorithme .qui est la moyenne harmonique de precision et de recall.

Il est à rappeler que tout cela doit rester contextuel car les métriques dépendent de la problématique abordée pour le métier donné.

Après sélection des meilleures variables productrices qui aideront à bien modéliser, nous sélectionnons selon une significativité avec le seuil p-value avec stats model.

A ces différents scores de performance de notre model, nous allons y ajouter une courbe ROC pour comprendre son utilité quant à la thématique métier que peut revêtir la régression logistique.



Courbe ROC de la regression logistique

CONCLUSION

La régression logistique a permis de prédire et de classer les travailleurs selon un score .

Nous pouvons affirmer que le salaire est un facteur très décisionnel pour réduire l'attrition des travailleurs.

Cependant il faut contextualiser cette modélisation car d'autres variables doivent être prises en compte .