

Le temps, un challenge à prendre en considération dans l'attrition des employés

Youssef Oubelmouh^{*,**}, Frédéric Fargon^{*}, Cyril de Runz^{**}, Arnaud Soulet^{**}, Cyril Veillon^{*}

^{*}Devoteam, Levallois-Perret, prenom.nom@devoteam.com

^{**}BDTLN, LIFAT, Université de Tours, Blois, prenom.nom@univ-tours.fr

Résumé. Réduire le taux d'attrition des employés est devenu un objectif majeur au sein des entreprises. Il existe une grande quantité de travaux sur la prédiction de l'attrition des clients mais beaucoup moins sur l'attrition des employés bien que ce phénomène soit de plus en plus important. Nous montrons dans cet article que le temps a un impact sur l'attrition alors que la majorité des variables utilisées pour prédire les démissions dans la littérature ne prennent pas en compte l'aspect temporel. Pour cela, nous proposons d'identifier si une variable temporelle est significativement différente entre les employés partis et ceux restés en nous appuyant sur les tests statistiques de Mann-Whitney U. Nous appliquons notre approche à des données internes à une entreprise de services du numérique permettant d'isoler des variables comme par exemple la durée des différentes missions effectuées. Notre étude conclut qu'il est nécessaire d'avoir la trajectoire temporelle des employés pour mieux appréhender le phénomène de l'attrition.

1 Introduction

L'attrition des employés est un problème qui s'accroît de plus en plus dans les entreprises technologiques à travers le monde et plus particulièrement dans les entreprises de services du numérique (ESN). Selon la DARES¹ (Adrien Lagouge, 2022), la France a enregistré près de 520 000 démissions par trimestre, à la fin de l'année 2021 et au début de l'année 2022, dont 470 000 démissions de contrats de travail à durée indéterminée (CDI), ce qui représente un taux de démissions de 2.7%. Ce taux est nettement plus élevé dans les ESN, il atteint par exemple 19% chez Accenture, 16% chez Atos et même 25% dans le secteur de la technologie en Inde (Standard, 2022). En effet, les consultants restent en moyenne entre 3 et 8 ans dans leur ESN (Morrison, 2021). Les démissions spontanées impliquent une baisse de productivité, d'autant plus si l'employé était spécialisé dans un domaine ou s'il avait de l'ancienneté. L'acquisition de nouveaux employés engendre de nombreux coûts et de la perte de temps que ce soit pour la recherche lors de la phase de recrutement ou la formation et l'adaptation

1. DARES : Direction de l'Animation de la Recherche, des Études et des Statistiques

Importance du temps dans l'attrition des employés

à un nouvel environnement après l'embauche. Le coût moyen du désengagement et de la non-disponibilité en France est de 14 580 euros par an et par salarié dont 9 185 euros de maîtrisables selon les conclusions de l'IBET² présenté par le groupe APICIL (2019). Il est donc préférable pour les entreprises d'identifier et d'activer les leviers leur permettant de retenir ses employés qui souhaitent partir.

Beaucoup d'études basées sur des approches issues des sciences humaines (Mobley, 1977) ont été effectuées depuis plus de 30 ans pour essayer de déterminer quels sont les facteurs qui poussent les employés à démissionner. Plus récemment, surtout depuis la diffusion par IBM d'un jeu de données fictives en 2016³, des travaux se sont penchés sur l'utilisation d'approches de fouille de données et d'apprentissage automatique dans un but principal de prédiction, car ce jeu représente un snapshot donnant très peu d'informations sur les dynamiques latentes. Par exemple, Zhao et al. (2018) sont partis de deux bases de données de ressources humaines, dont celle d'IBM, de tailles et de nombre d'attributs différents pour en déduire 1) que la taille des jeux d'apprentissage a un effet important sur la qualité des modèles et 2) que les algorithmes d'ensemble eXtreme Gradient Boosting, Gradient Boosting Tree et Random Forest semblent donner les meilleurs résultats. De leur côté, Yiğit et Shourabizadeh (2017) concluent que les machines à vecteur de support seraient le meilleur type de modèle pour les données d'IBM. Kane-Sellers (2007) indique que la régression logistique fonctionnerait le mieux si l'on se concentre principalement sur trois types de variables : caractéristiques personnelles, caractéristiques du travail, développement des ressources humaines. Brockett et al. (2019) sont les seuls qui ont introduit artificiellement de la temporalité dans leurs données. Pour cela, ils ont décidé aléatoirement pour chaque personne partie, s'il était parti en milieu d'année ou alors en fin d'année sur les données IBM. Ainsi, malgré le nombre grandissant de recherches menées sur la prédiction de l'attrition des employés, nous observons que les dynamiques temporelles dans les trajectoires des employés ne sont que peu ou pas considérées, et cela du fait même de la nature des jeux de données (snapshot). Par exemple, les rares variables liées au temps dans les données IBM ne sont pas assez exhaustives et/ou précises, e.g. « Years In Current Role », « Years With Current Manager ». Il est à noter que si les données fictives d'IBM se sont imposées dans ces recherches, c'est parce que la sensibilité des informations personnelles pouvant y être contenues ne permet pas de proposer des jeux de données ouverts à la communauté (contraintes du RGPD).

Dans cet article, contrairement à la littérature actuelle, nous cherchons à étudier les liens entre l'attrition et les dynamiques temporelles des trajectoires des employés. Nous cherchons à montrer que le nombre de changements entre états, et les durées des états dans ces trajectoires sont liées significativement à l'attrition sur un cas d'étude réel. Ce premier travail semble montrer que les jeux de données exploités dans la littérature sont peu adéquats car ils ne permettent pas de capturer des informations fortement significatives.

Nous commencerons par exposer notre cas d'étude dans la section 2, puis notre méthodologie présentant les variables que nous construisons et les tests statistiques

2. IBET : Indice du Bien-Être au Travail

3. <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>

(section 3), nous continuerons par l’analyse des résultats obtenus (section 4), puis nous conclurons (section 5).

2 Présentation du cas d’étude

Dans ce travail, nous nous sommes restreints aux employés qui sont ou qui ont été en CDI. Afin d’étudier leur attrition, nous nous intéressons à 4 tables obtenues à travers 3 sources de données différentes issues d’une ESN dont deux logiciels de gestion de ressources humaines. La constitution de notre entrepôt de données a un coût non négligeable en raison des différents processus *Extract-Transform-Load* à gérer tout en respectant le règlement général sur la protection des données (RGPD) particulièrement sensible à ce type de données personnelles.

Les 4 tables utilisées dans cet article sont **Job History**, **Mission History**, **Pricing Profil History** et **Compensation History**. Job History contient les changements liés au poste c’est-à-dire la date d’embauche, et/ou le nom des nouveaux postes (e.g. « Business Analyst », « Contrôleur de Gestion », « Product Owner »). Mission History contient les différentes missions effectuées avec la date de début et la date de fin. Pricing Profil History possède les différents changements liés aux types de profils vendus au client (« Consultant Junior », « Consultant Senior », « Expert Director », etc.). Enfin, Compensation History contient l’historique des salaires. Le tableau 1 décrit succinctement ces différentes tables en indiquant le nombre d’enregistrements, le nombre d’employés renseignés et le nombre d’employés ayant démissionné. Dans la suite de l’article, la variable « Active » vaut 0 pour les démissionnaires, 1 sinon.

Nom	Contenu	Nbr. de lignes	Nbr. d’employés <i>Active</i> = 1	Nbr. d’employés <i>Active</i> = 0
Job History	Changements liés au poste (embauche, nom du poste)	34 111	8 899	3 669
Job History (2020)	Changements liés au poste (embauche après 2020, nom du poste)	5 491	2 941	388
Mission History	Différentes missions effectuées	36 392	1 743	7 603
Pricing Profil History	Changements du type de profil de l’employé vendu aux clients	20 790	3 764	2 185
Compensation History	Historique des changements de salaire	15 549	8 025	3 214

TAB. 1 – Description des jeux de données (employé actif dénoté par *Active* = 1 / employé démissionnaire dénoté par *Active* = 0).

De manière générale, nous observons dans le tableau 1 que le nombre d’enregistrements par employé n’est pas très élevé avec une médiane au plus égale à 3 pour la table Mission History. Notons que la table Mission History contient des données subjectives puisque ce sont les employés qui ajoutent eux-mêmes ces données dans leur profil. Par conséquent, les données ne concernent pas exclusivement des missions chez les clients,

Importance du temps dans l'attrition des employés

mais également des formations, des séminaires ou encore des sous-missions ponctuelles. Les durées peuvent donc être très brèves de l'ordre de quelques jours. L'employé en faisant l'effort de renseigner ces informations dans son cursus traduit une forme d'engagement. Cette table capte donc des signaux informationnels sur la perception de sa situation dans l'entreprise. À l'inverse, les autres tables contiennent des données plus objectives puisqu'elles ne sont pas renseignées par les employés eux-mêmes.

Les tables Job History et Compensation History sont issues du même logiciel qui a été mis en place entre fin 2019 et début 2020. Par conséquent, les employés ayant quitté l'entreprise avant 2020 ne sont pas inclus dans ces tables. De même, les changements intervenus entre la date d'embauche et début 2020 (pour ceux toujours actifs ou ceux ayant quitté l'entreprise après 2020) ne sont pas inscrits dans ces deux tables. Cela a pour conséquence de générer un biais dans nos données. Nous pouvons remarquer que ces données correspondent à la période de la pandémie de la COVID.

Pour chaque mission de la table Mission History, nous disposons de la date de début et la date de fin permettant de calculer la durée de chaque mission pour tous les employés. Cependant, pour la table Pricing Profil History, nous avons seulement la date de changement effective. Sachant que nous n'avons pas à notre disposition les dates de fin de contrat des employés ayant quitté l'entreprise avant 2020, nous ne pouvons pas calculer la durée du dernier état qui serait pourtant une valeur importante. Nous avons donc décidé de nous restreindre à la même catégorie de personnes contenue dans les données de Job History et Compensation History sauf que cette fois nous avons les changements de profil entre la date d'embauche et l'année 2020.

Enfin, la table Compensation History, nous avons également omis les primes pour se concentrer uniquement sur les changements de salaire annuel fixe.

3 Méthodologie

Cette section présente notre méthode pour déterminer si le temps a un impact sur l'attrition. Nous cherchons dans un premier temps à construire des variables traduisant les dynamiques temporelles des trajectoires des employés, appelées *variables temporelles*. Ces variables ont pour objectif de fournir des informations sur le nombre de changements d'états et la durée des états de ces trajectoires. Nous cherchons ensuite à identifier les possibles dépendances significatives, au sens statistique, entre ces variables temporelles et l'attrition des employés.

Construction des variables temporelles Pour chacune des 4 tables de notre cas d'étude, nous créons 1 variable comptant le nombre de changements ayant eu lieu (après suppression ou correction des anomalies trouvées) et 4 variables de durées à partir des durées entre chaque changement qui sont *First duration*, *Last duration*, *Min duration*, *Max duration* qui correspondent respectivement à la durée du premier état, la durée du dernier état, la durée minimale et celle maximale parmi tous les états. Par exemple, pour la table Job History où chaque enregistrement concerne un changement de poste pour un employé, le nombre d'états indique le nombre de changements de poste ; la variable *First duration* indique la durée du premier poste ; etc.

Identification de dépendances temporelles significatives Afin d'identifier des possibles dépendances significatives entre les variables temporelles précédemment construites et l'attrition, nous séparons en deux populations les employés selon l'attrition (partis ou encore actif). Dans l'ensemble de notre étude, nous faisons l'hypothèse de l'indépendance entre les données de chaque employé. Nous cherchons à savoir si ces deux populations sont significativement différentes à la vue de nos variables temporelles.

Pour cela, en statistique, on utilise des tests paramétriques si la distribution de nos données suit une loi (le plus souvent la loi normale), non-paramétrique sinon. La loi normale ou la distribution gaussienne définit une représentation de données selon laquelle la plupart des valeurs sont regroupées autour de la moyenne et les autres s'en écartent symétriquement des deux côtés. Dans notre cas, nous commençons par déterminer si la distribution de nos données est gaussienne avec un test de normalité. Le test de normalité est un test statistique qui permet de déterminer si un échantillon de données suit une distribution normale. Pour cela, nous utilisons à la fois une méthode visuelle avec les boîtes à moustaches (box-plot) et une méthode statistique. Les deux méthodes statistiques les plus utilisées pour tester la normalité des données sont le test de Kolmogorov-Smirnov (Massey Jr, 1951) et le test de Shapiro-Wilk (Shapiro et Wilk, 1965). Le test de Shapiro-Wilk est considéré comme plus approprié pour les échantillons de petite taille (Razali et al., 2011) (<50 échantillons) alors que le test de Kolmogorov-Smirnov est utilisé pour un nombre d'échantillons plus grand que 50, ce qui est notre cas. Puisque nos données sont de grandes tailles nous avons choisi d'utiliser le test de Kolmogorov-Smirnov. Sur nos variables, les résultats des tests nous montrent que les distributions des données ne suivent pas une loi normale. Nous devons alors recourir à un test non-paramétrique. Pour comparer deux populations, l'une des solutions lorsque la distribution des données ne suit pas une loi normale est le test non-paramétrique de Mann-Whitney U (MWU) aussi connu sous le nom de Wilcoxon-Mann-Whitney test ou encore Rank Sum Test (Mann et Whitney, 1947). Nous avons donc choisi d'utiliser ce test pour comparer nos deux groupes.

Le test de **Mann-Whitney U** (MWU-test) ne suppose aucune distribution spécifique (telle qu'une distribution normale des échantillons) pour calculer les statistiques du test et les p -valeurs. La p -valeur est une probabilité mesurant l'importance d'un résultat statistique. En revanche, le MWU-test suppose que les observations soient indépendantes. Basé sur les rangs, son principe est de déterminer s'il est probable qu'une observation dans un groupe soit supérieure à une observation dans l'autre groupe. En d'autres termes, il s'agit de savoir si un échantillon a une dominance stochastique par rapport à l'autre. L'hypothèse nulle et l'hypothèse alternative de ce test sont les suivantes :

- H0 : Les deux groupes sont échantillonnés à partir de populations ayant des distributions identiques. Généralement, les populations échantillonnées présentent une égalité stochastique.
- H1 : Les deux groupes sont échantillonnés à partir de populations ayant des distributions différentes. La plupart du temps, cela signifie que l'une des populations échantillonnées présente une dominance stochastique.

Nous détaillons maintenant les 4 étapes pour appliquer le MWU-test :

Importance du temps dans l'attrition des employés

1. Classer les valeurs des deux échantillons de la plus basse à la plus haute, quel que soit le groupe auquel chaque valeur appartient.
2. Additionner les rangs des deux échantillons séparément, R_1 et R_2 (expliquant la dénomination du test par *rank-sum*).
3. Calculer les statistiques du test : $U = \min(U_1, U_2)$ avec $U_1 = n_1 n_2 - n_1(n_1 + 1)/2 - R_1$ et $U_2 = n_1 n_2 - n_2(n_2 + 1)/2 - R_2$
4. Déterminer la p -valeur.

Si la p -valeur est inférieure à 5% le test est statistiquement significatif et l'hypothèse nulle est rejetée. Dans ce cas, nous pouvons conclure que les deux populations sont différentes.

Le test qui a été décrit dans cette section est un test bilatéral, il peut être légèrement modifier pour le transformer en test unilatéral. Le test bilatéral permet de rejeter ou non l'hypothèse nulle énoncée précédemment. Les tests unilatéraux permettent de savoir quelle population a une dominance stochastique sur l'autre. Nous allons donc utiliser le test bilatéral ainsi que les deux tests unilatéraux sur l'ensemble de nos tables.

4 Analyse des résultats

En suivant notre méthodologie, pour chaque table, nous analysons nos variables temporelles construites afin de déterminer leur importance dans la compréhension du phénomène d'attrition. Comme indiqué précédemment, tous les tests de Kolmogorov-Smirnov sur nos variables ont significativement montré qu'aucun des groupes des 4 tables ne suit une loi normale. Aussi, l'ensemble des résultats des tests statistiques donnés dans cette section sont issus des tests de Mann-Whitney U.

4.1 Analyse de la table Job History

Nous commençons par analyser les 5 variables temporelles (le nombre et les 4 durées) construites au regard de la table Job History en considérant dans un premier temps l'ensemble des employés, puis en se restreignant à ceux embauchés après 2020 (pour lesquels les données sont complètes). La figure 1 représente les box-plots pour les 5 variables en séparant les deux populations à savoir employés partis (en bleu) et ceux restés (en orange). Le tableau 2 présente les différents tests de Mann-Whitney U bilatéral puis unilatéraux. « $0 > 1$ » et « $0 < 1$ » signifient que la valeur de la variable considérée est respectivement supérieure et inférieure pour les employés partis par rapport à ceux qui sont restés.

Nous pouvons voir que pour le Job History les deux populations sont statistiquement différentes (rejet de l'hypothèse nulle) concernant toutes les variables (nombre de jobs et durées des jobs). Les tests unilatéraux nous montrent que le nombre de jobs ainsi que la durée du dernier job sont inférieurs chez ceux qui sont partis et que les autres durées sont supérieures chez les employés toujours actifs. Nous pouvons en déduire que l'attrition survient plus fréquemment chez les employés qui restent longtemps sur un même job et n'en changent pas assez. Ces derniers décident donc de partir au bout de quelques mois après le début de leur dernier job. Cependant, ces résultats

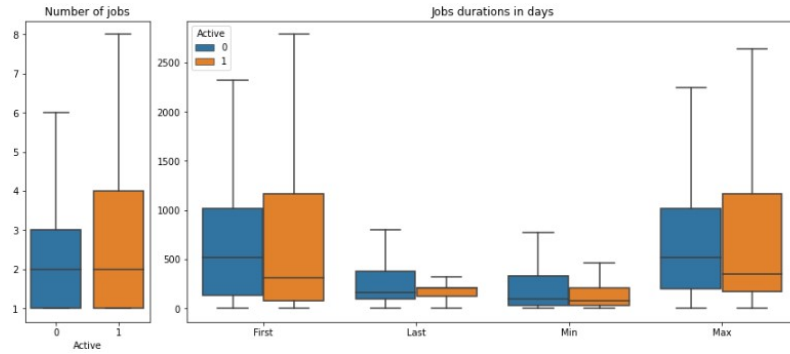


FIG. 1 – *Box-plots des variables pour la table job history.*

	bilatéral	unilatéral 0 > 1	unilatéral 0 < 1
Count	<0.001	1.0	<0.001
First duration	<0.001	<0.001	1.0
Last duration	0.014	0.992	0.007
Min duration	<0.001	<0.001	1.0
Max duration	<0.001	<0.001	1.0

TAB. 2 – *P-valeurs des variables pour la table Job History.*

sont à pondérer car, comme mentionné dans la description des données, cette table ne contient pas les changements survenus avant 2020. Par conséquent, cette absence biaise les données des personnes ayant été embauchées dans l’entreprise avant 2020. Nous allons donc maintenant nous focaliser uniquement sur les employés arrivés après 2020 afin de confirmer ou d’infirmier notre analyse précédente.

	bilatéral	unilatéral 0 > 1	unilatéral 0 < 1
Count	<0.001	1.0	<0.001
First duration	0.003	0.002	0.998
Last duration	<0.001	1.0	<0.001
Min duration	0.032	0.016	0.984
Max duration	0.042	0.979	0.021

TAB. 3 – *P-valeurs des variables pour la table Job History (depuis 2020).*

Si on se concentre uniquement sur les employés ayant intégré l’entreprise après 2020, nous retrouvons toujours des différences statistiquement significatives entre les deux populations pour toutes les variables temporelles (voir figure 2 et tableau 3). On remarque par contre que la variable max est maintenant inférieure chez les personnes ayant démissionné contrairement à notre analyse précédente. Cela peut être expliqué

Importance du temps dans l'attrition des employés

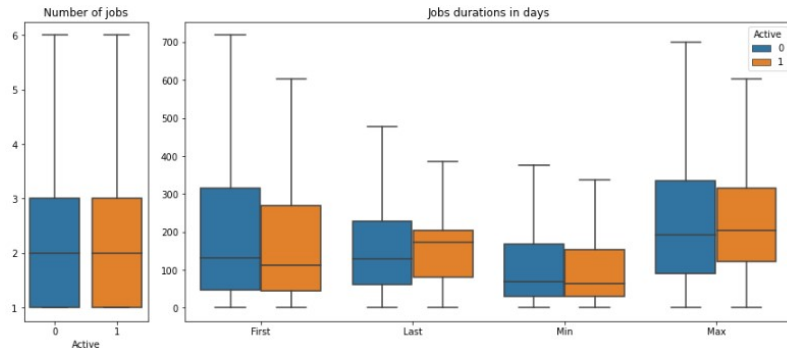


FIG. 2 – *Box-plots des variables pour la table Job History (depuis 2020).*

par le fait qu'avec moins de 3 ans de contrat, la durée maximale d'un job sera forcément plus petite que la durée de ceux qui sont toujours en poste à l'heure actuelle.

4.2 Analyse de la table Mission History

Nous appliquons la même méthode d'analyse sur les variables temporelles de la table des missions en nous appuyant sur la figure 3 et sur le tableau 4.

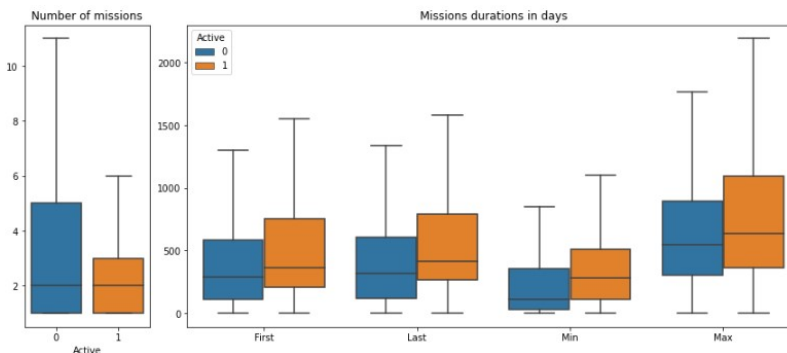


FIG. 3 – *Box-plots des variables pour la table Mission History.*

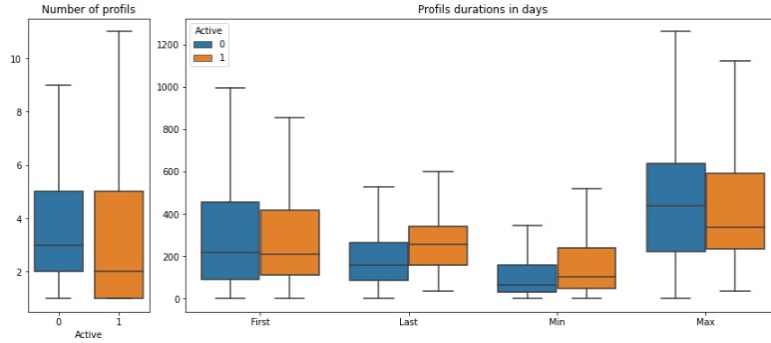
Concernant les missions, nous observons également que toutes les variables sont significativement différentes entre les deux populations. Le nombre de missions effectuées par les employés qui démissionnent est plus important que le nombre de missions de ceux qui restent. Cependant, les valeurs de toutes les variables de durées sont inférieures. Cela signifie donc que les employés ayant eu beaucoup de missions et ce, d'autant plus avec des courtes durées, ont tendance à partir.

	bilatéral	unilatéral 0 > 1	unilatéral 0 < 1
Count	<0.001	<0.001	1.0
First duration	<0.001	1.0	<0.001
Last duration	<0.001	1.0	<0.001
Min duration	<0.001	1.0	<0.001
Max duration	<0.001	1.0	<0.001

TAB. 4 – *P-valeurs des variables pour la table Mission History.*

4.3 Analyse de la table Pricing Profile History

Nous considérons maintenant les variables temporelles de la table du profil de coût des employés en nous appuyant sur la figure 4 et sur le tableau 5.

FIG. 4 – *Box-plots des variables pour la table Pricing Profile History.*

	bilatéral	unilatéral 0 > 1	unilatéral 0 < 1
Count	<0.001	<0.001	1.0
First duration	0.388	0.805	0.194
Last duration	<0.001	1.0	<0.001
Min duration	<0.001	1.0	<0.001
Max duration	0.006	0.003	0.997

TAB. 5 – *P-valeurs des variables pour la table Pricing Profil History.*

Pour la table Pricing Profil History, le comportement des deux populations ne se différencie pas significativement suivant toutes les variables temporelles. En effet, d'après les p -valeurs du test de Mann-Whitney U nous avons la première durée qui ne rejette pas l'hypothèse nulle selon laquelle les deux populations sont similaires. Nous avons donc la première durée de profil qui est similaire pour les deux populations. Les tests unilatéraux nous montrent que ceux qui ne sont plus en poste ont subi plus de modifications de profil et que la durée max d'un de leur profil est plus élevé par rapport

Importance du temps dans l'attrition des employés

à ceux qui sont toujours sous contrat. Tandis que la durée minimale parmi toutes les durées ainsi que la dernière durée sont plus élevées pour les personnes encore actives.

4.4 Analyse de la table Compensation History

Cette section porte sur les variables temporelles de la table des salaires des employés en nous appuyant sur la figure 5 et sur le tableau 6.

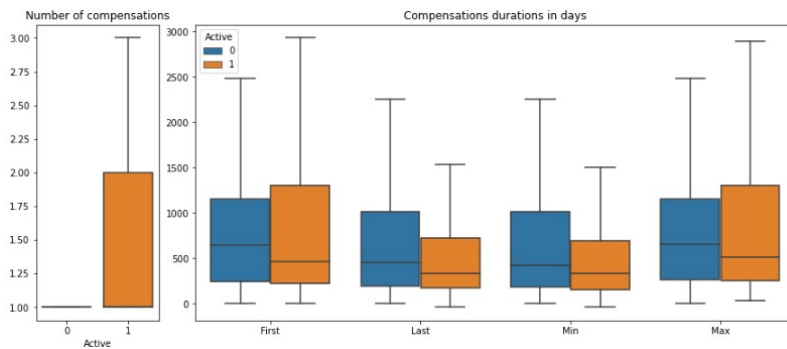


FIG. 5 – *Box-plots des variables pour la table Compensation History.*

	bilatéral	unilatéral 0 > 1	unilatéral 0 < 1
Count	<0.001	1.0	<0.01
First duration	0.055	0.027	0.973
Last duration	<0.001	<0.001	1.0
Min duration	<0.001	<0.001	1.0
Max duration	0.262	0.131	0.869

TAB. 6 – *P-valeurs des variables pour la table Compensation History.*

Dans la table Compensation History, le test bilatéral sur la donnée First duration ne montre pas de différence significative entre les deux populations. Pourtant, le test unilatéral nous montre la supériorité significative de cette variable pour les personnes ayant démissionné. Cela est confirmé visuellement par la boîte à moustache de cette variable qui présente une médiane bien plus élevée que celle des personnes qui sont restées et ce, malgré que les troisièmes quartiles soient inversés pour les deux groupes. La durée maximale avec le même salaire n'est pas différente. Quant aux autres variables temporelles, elles ont une valeur plus élevée pour les non actifs. Comme attendu, nous pouvons déduire de ces résultats que les employés qui démissionnent ont moins de changement de salaire et que leur salaire reste fixe durant une plus longue durée.

4.5 Comparaison des résultats

Afin d'identifier des informations intéressantes quant à l'impact du temps sur l'attrition, nous comparons les résultats des tests unilatéraux entre les différentes tables. Le tableau 7 permet de comparer à l'aide de flèches les employés ayant démissionné par rapport à ceux toujours en CDI. Une flèche dirigée vers le haut signifie que la valeur de la variable concernée est supérieure chez les employés inactifs, une flèche vers le bas signifie l'inverse et un tiret signifie qu'il n'y pas de différences entre les deux populations.

	Job	Mission	Pricing	Compensation
Count	↓	↑	↑	↓
First duration	↑	↓	-	-
Last duration	↓	↓	↓	↑
Min duration	↑	↓	↓	↑
Max duration	↑	↓	↑	-

TAB. 7 – *Supériorité et infériorité des employés ayant quitté l'entreprise.*

À partir de ce tableau, nous pouvons en déduire les observations et possibles explications présentées dans le tableau 8. Ces informations nous semblent utiles pour les ressources humaines et surtout indique clairement l'apport que peut représenter une analyse des dynamiques temporelles pour expliquer et prédire l'attrition des employés. Attention, du fait des biais de nos données, les comportements observés peuvent être différents de la réalité.

Tables d'origine	Observations groupe démission	Explications possibles
Mission	Nombre élevé de missions de courtes durées.	Les employés pourraient chercher une certaine stabilité dans leur travail.
Job	Nombre bas de jobs et en plus de longues durées.	Les employés semblent vouloir évoluer dans l'entreprise et ne pas rester au même niveau.
Compensation	Nombre d'augmentations de salaire faible et salaire fixe durant une plus longue période.	Les augmentations seraient perçues comme peu fréquentes.
Job/Mission	Beaucoup de missions effectuées avec des durées plus courtes mais résultats opposés pour job.	Les employés veulent sans doute de l'évolution dans leur job mais une stabilité sur les lieux de missions.
Pricing/Compensation	Nombre élevé de pricings de courte durée et opposé pour le salaire.	Le pricing profile évolue rapidement mais le salaire ne semble pas suivre.
Mission/Pricing	Nombre élevé de changements de pricing profil et de missions et durée plus courte (sauf pour le max).	Les changements de missions entraînent des changements de clients permettant à l'entreprise de modifier le pricing profile de l'employé.

TAB. 8 – *Observations et tentatives d'explications des démissions des employés.*

5 Conclusion

Nous avons vu à travers le test statistique de Mann-Whitney U appliqué aux 4 tables que nos variables temporelles sont souvent significativement différentes entre les employés toujours en CDI et ceux qui avaient démissionnés. Cela nous conforte dans l'idée que le temps a un impact sur l'attrition. Mais il s'avère dans la littérature existante que ce facteur temps n'est pratiquement jamais utilisé pour prédire l'attrition des employés. Le fait d'avoir des évolutions et des changements est a priori un facteur sur lequel les entreprises pourraient agir pour remédier à l'attrition, ce qui n'est pas traité dans la littérature orientée informatique, et représente donc une faille dans les outils numériques actuels à disposition des RH.

Pour renforcer notre étude, l'une des pistes est de faire une distinction entre jeunes et moins jeunes salariés car d'après l'INSEE⁴ (Picart, 2014), le taux de rotation chez les jeunes est beaucoup plus élevé que chez les seniors. Comme la période d'étude des données chevauche la période de la pandémie de la COVID, il faudra vérifier nos hypothèses sur les années antérieures. De plus, une perspective découlant de cette étude est de construire des séquences qui caractérisent les trajectoires des employés dans l'entreprise, depuis l'embauche jusqu'à la démission. Sur ces séquences d'événements, nous chercherons à la fois à prédire mais aussi à expliquer l'attrition à l'aide d'extractions de motifs séquentiels significatifs facilement interprétables.

Références

- Adrien Lagouge, Ismaël Ramaïo, V. B. (2022). La France vit-elle une "grande démission". Technical report, DARES : Direction de l'animation de la recherche, des études et des statistiques.
- APICIL, G. (2019). Ibet – désengagement des salariés : un coût de 14580€/an/salarié. Technical report, Groupe APICIL.
- Brockett, N., C. Clarke, M. Berlingerio, et S. Dutta (2019). A system for analysis and remediation of attrition. In *2019 IEEE International Conference on Big Data (Big Data)*, pp. 2016–2019.
- Kane-Sellers, M. L. (2007). *Predictive models of employee voluntary turnover in a North American professional sales force using data-mining analysis*. Ph. D. thesis, Texas University.
- Mann, H. B. et D. R. Whitney (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50–60.
- Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association* 46(253), 68–78.
- Mobley, W. H. (1977). Intermediate linkages in the relationship between job satisfaction and employee turnover. *Journal of applied psychology* 62(2), 237.
- Morrison, G. (2021). How long do people stay at their firms? Technical report, Consulting Point.
- Picart, C. (2014). Une rotation de la main d'œuvre presque quintuplée en 30 ans : plus qu'un essor des formes particulières d'emploi, un profond changement de leur usage. Technical report, INSEE : Institut National de la Statistique et des Études Économiques.

4. INSEE : Institut National de la Statistique et des Études Économiques

- Razali, N. M., Y. B. Wah, et al. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics* 2(1), 21–33.
- Shapiro, S. S. et M. B. Wilk (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52(3/4), 591–611.
- Standard, B. (2022). It industry logged 25% attrition in fy22, trend to continue : Report. Technical report, Business Standard.
- Yiğit, İ. O. et H. Shourabizadeh (2017). An approach for predicting employee churn by using data mining. In *2017 International IDAP*, pp. 1–4. IEEE.
- Zhao, Y., M. K. Hryniewicki, F. Cheng, B. Fu, et X. Zhu (2018). Employee turnover prediction with machine learning : A reliable approach. In *Proc. of SAI intelligent systems conference*, pp. 737–758. Springer.

Summary

Reducing employee attrition has become a major objective within companies. There is a large amount of scientific literature on the prediction of customer attrition but much less on employee attrition, although there is a growing body of literature. We want to show in our paper that time has an impact on attrition, since most of the variables used to predict resignations in the literature do not take into account the temporal aspect. To do so, we propose to identify whether a temporal variable is significantly different between employees who left and those who stayed using Mann-Whitney U statistical tests. We apply our approach to internal data of an information technology consulting company allowing us to isolate variables such as the duration of the different missions performed. Our study concludes that it is necessary to have the temporal trajectory of employees to better understand the attrition phenomenon.