



---

# RAPPORT PROJET 4

---

Sylvain VASSEUR



16 AVRIL 2025

Projet personnel à but formateur et exploratoire

# Table des matières

I.	Résumé exécutif.....	2
II.	Introduction .....	2
III.	Méthodologie .....	3
	A. Analyse des départements : .....	3
	B. Analyse des facteurs explicatifs : .....	3
	C. Identification des profils de départements : .....	3
IV.	Résultats .....	4
	A. Analyse des départements .....	4
	a. Analyse des statistiques descriptives .....	4
	b. Tests d'hypothèses sur les différences .....	5
	c. Analyse de distribution des performances .....	6
	d. Synthèse des observations .....	7
	B. Analyse des facteurs explicatifs .....	8
	a. Analyse des variables quantitatives .....	8
	b. Analyse des variables qualitatives .....	11
	c. Résultats d'analyse .....	14
	C. Identification des profils par départements .....	15
	D. Synthèse et recommandations.....	17
V.	Analyse approfondie .....	18
	A. Une distribution étonnamment homogène .....	18
	B. Une variable dominante : le salaire mensuel .....	18
	C. Des profils éducatifs peu différenciés selon les fonctions.....	20
VI.	Recommandations.....	21
	A. Revoir le système de notation.....	21
	B. Améliorer la qualité et la granularité des données .....	22
	C. Réviser la grille salariale à partir de critères plus fins .....	22
	D. Réévaluer les niveaux de qualification attendus par département .....	22
	E. Mettre en place un système de satisfaction employé plus nuancé et personnalisé.....	22
VII.	Conclusion .....	22

# I. Résumé exécutif

- **Contenu:** Contexte, objectif, méthodes, résultats clés, recommandations principales
- **Bonnes pratiques:** Limiter à 1 page, être percutant, centrer sur la valeur business

## II. Introduction

### **Contexte :**

Dans un souci d'amélioration continue, une entreprise a exprimé le besoin d'évaluer les écarts de performance entre ses différents départements. Cette initiative vise à mieux comprendre les facteurs internes qui influencent la productivité et la satisfaction des collaborateurs.

### **Objectifs :**

L'objectif principal de cette analyse est d'identifier les différences de performance entre départements, à partir de variables à la fois qualitatives (comme le type de département ou le niveau de satisfaction) et quantitatives (comme les scores de performance ou les heures travaillées). Il s'agit également de détecter d'éventuelles corrélations ou tendances permettant d'orienter des décisions de gestion ou d'optimisation des ressources.

### **Questions de recherche**

- Existe-t-il des écarts de performance significatifs entre les départements ?
- Quels facteurs (satisfaction des employés, heures travaillées) expliquent les écarts de performance observés entre les départements ?
- Quelles relations spécifiques peut-on identifier entre les variables qualitatives (départements, satisfaction) et quantitatives (scores de performance, heures travaillées) ?
- Peut-on identifier des profils de départements plus performants ou moins performants que la moyenne ?

### III. Méthodologie

#### A. Analyse des départements :

**Méthodes :** Statistiques descriptives comparatives, tests d'hypothèses sur les différences (ANOVA, tests t), analyse de distribution des performances

**Objectif :** Identifier et quantifier les écarts de performance entre les départements, en répondant à la question : "Existe-t-il des écarts significatifs de performance entre les départements ?"

#### B. Analyse des facteurs explicatifs :

**Méthodes :** Analyse de corrélation bivariée et multivariée, modélisation des relations entre variables, analyse de régression

**Objectif :** Déterminer l'influence de la satisfaction des employés et des heures travaillées sur la performance, pour répondre à : "*Quels facteurs expliquent les écarts de performance observés entre les départements ?*"

#### C. Identification des profils de départements :

**Méthodes :** Techniques de segmentation, analyse comparative des groupes de performance, caractérisation multidimensionnelle des profils

**Objectif :** Caractériser les départements sur-performants et sous-performants et identifier leurs attributs distinctifs, répondant à : "*Peut-on identifier des profils de départements plus performants ou moins performants que la moyenne ?*"

#### D. Synthèse et recommandations :

**Méthodes :** Analyse croisée des résultats, hiérarchisation des facteurs d'influence, modélisation d'impact des interventions potentielles

**Objectif :** Formuler des recommandations concrètes et hiérarchisées pour réduire les écarts de performance entre départements.

## IV. Résultats

### A. Analyse des départements

a. Analyse des statistiques descriptives

Cette section présente les résultats de l'analyse quantitative des écarts de performance entre départements. Les données ont été analysées à l'aide de statistiques descriptives et de tests d'inférence statistique conformément à la méthodologie décrite précédemment.

Dans un premier temps, les indicateurs de tendance centrale et de dispersion sont examinés pour chaque département. Le Tableau 1 synthétise ces statistiques fondamentales, permettant une première comparaison factuelle des performances observées.

	Moyenne	Écart_type	Minimum	Maximum	Effectif
Department					
Engineering	3.02	1.42	1	5	10956
Operations	3.01	1.41	1	5	11181
Customer Support	3.00	1.41	1	5	11116
IT	3.00	1.42	1	5	11131
HR	3.00	1.41	1	5	10960
Sales	2.99	1.41	1	5	11122
Finance	2.98	1.41	1	5	11200
Marketing	2.98	1.41	1	5	11216
Legal	2.98	1.42	1	5	11118

Tableau 1 : Statistiques descriptives par département

Une remarquable **homogénéité** ressort des données : les moyennes varient dans un intervalle restreint, de 2,98 à 3,02, soit une amplitude de 0,04 points. De même, les écarts-types présentent une variabilité minimale (1,41 à 1,42). Les valeurs minimales et maximales sont identiques pour toutes les catégories, et le nombre d'observations par catégorie s'élève à environ 11 000 unités.

Ces résultats indiquent une **distribution** des performances **très uniforme** entre les neuf catégories, sans écart notable entre elles.

#### b. Tests d'hypothèses sur les différences

Afin d'identifier d'éventuelles disparités entre les départements, un test d'ANOVA est réalisé.

Ce test permet de détecter si, pour une variable donnée, **au moins deux départements** présentent une **différence significative de moyenne**.

Les variables sélectionnées pour cette analyse sont les suivantes :

- `Work_Hours_Per_Week` : volume horaire de travail
- `Overtime_Hours` : charge ou pression supplémentaire
- `Projects_Handled` : nombre de missions gérées
- `Performance_Score` : niveau de performance évalué
- `Remote_Work_Frequency` : fréquence de télétravail
- `Team_Size` : effectif moyen par équipe
- `Monthly_Salary` : niveau de rémunération mensuelle

Ces variables ont été choisies car elles permettent de **caractériser de façon concrète l'organisation du travail dans chaque département**. Un résultat significatif à l'ANOVA indiquerait qu'au moins deux départements présentent des **conditions ou résultats**

**sensiblement différents** sur l'un de ces aspects — ce qui pourrait orienter des décisions de pilotage ou d'équilibrage interne.

Variable	p-value
Projects_Handled	0.0864
Overtime_Hours	0.1370
Monthly_Salary	0.4500
Performance_Score	0.4673
Team_Size	0.5552
Remote_Work_Frequency	0.6141
Work_Hours_Per_Week	0.8018

Tableau 2 : Test ANOVA des variables quantitatives

Ces résultats suggèrent qu'il n'y a pas de disparité significative entre les départements concernant la charge de travail, la pression supplémentaire, le nombre de projets gérés, la performance, le télétravail, la taille des équipes, et les salaires. La p-value obtenue est supérieure à 0,05, ce qui signifie qu'aucune différence significative ne peut être affirmée. Même si certaines valeurs se rapprochent de ce seuil, elles ne permettent pas de tirer de conclusions définitives.

### c. Analyse de distribution des performances

Bien que certaines variables présentent des p-values proches du seuil de significativité (notamment `Projects_Handled` avec  $p = 0.086$ ), ces différences ne sont pas assez marquées pour être considérées comme statistiquement significatives au seuil classique de 5%. Par conséquent, ces résultats ne justifient pas une distinction notable entre les départements pour les aspects analysés.

Intéressons-nous à la répartition des scores de performance au sein des départements. L'analyse statistique suggère que les distributions des scores sont, à quelques nuances près, similaires entre les départements.

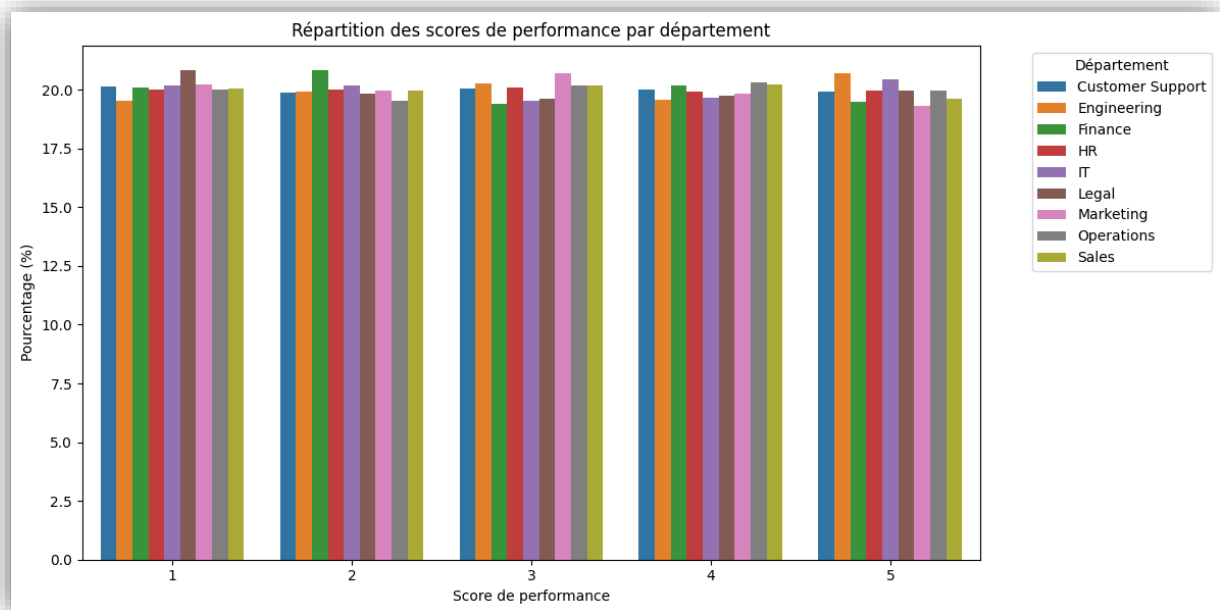


Figure 1 : Histogramme de la répartition des scores par département

Cette observation est confirmée par les données : la répartition des scores de performance ne montre pas de disparités significatives entre les départements. La proportion d'employés pour chaque niveau de score (de 1 à 5) reste globalement stable d'un secteur à l'autre.

#### d. Synthèse des observations

L'analyse des scores de performance selon les départements révèle une forte similarité entre les groupes. Les statistiques descriptives montrent des moyennes très proches et des écarts types quasi identiques, le tout avec un volume de données similaire entre départements. Les résultats du test ANOVA confirment cette stabilité, aucune des variables testées ne présentant de différence significative entre les groupes. Par ailleurs, la répartition des scores de performance est équilibrée dans l'ensemble des départements, chaque modalité représentant environ 20 % des effectifs. Ces constats ne mettent en évidence **aucune variation notable liée au département à ce stade de l'analyse**. Pour la suite, nous pourrions nous concentrer sur d'éventuels leviers internes aux départements ou sur des variables transverses susceptibles d'apporter un éclairage complémentaire.



## B. Analyse des facteurs explicatifs

### a. Analyse des variables quantitatives

#### 1. Corrélations avec le Performance\_Score

Afin de mieux comprendre quels facteurs peuvent influencer le *Performance\_Score*, nous avons réalisé une analyse de corrélation. L'objectif est d'identifier si certaines variables sont significativement liées à la performance des employés.

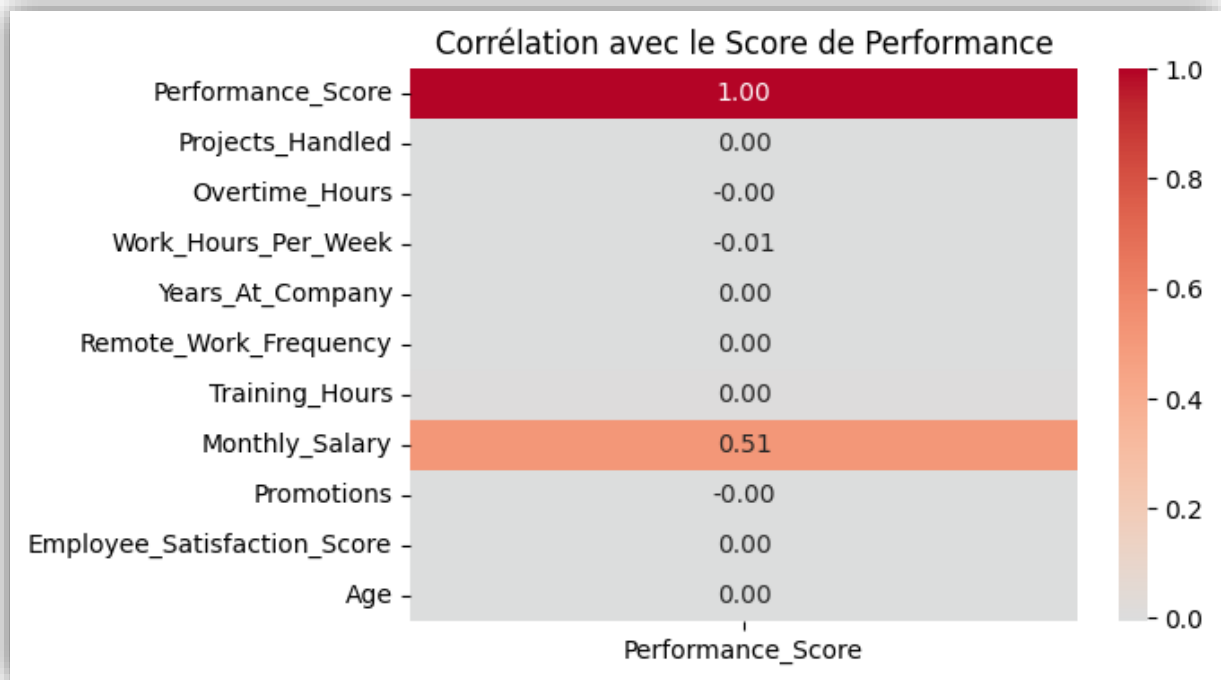


Figure 2 : Corrélation du score de performance sur les variables quantitatives

Les résultats sont globalement peu concluants : à l'exception d'une seule variable, toutes les corrélations obtenues sont proches de zéro. Cela signifie qu'aucun lien clair n'émerge entre le score de performance et la majorité des variables analysées. Ces résultats freinent notre capacité à dégager des pistes concrètes d'amélioration.

Seule exception notable : le *Monthly\_Salary*, qui présente une corrélation de **0.51** avec le *Performance\_Score*. Cette valeur suggère une relation modérée — autrement dit, plus un employé est rémunéré, plus son score de performance tend à être élevé. Ce lien reste cependant à nuancer : bien qu'une corrélation de 0.51 soit intéressante, elle ne permet pas, à elle seule, d'expliquer pleinement les différences de performance observées.

## 2. Régression multivariée

Afin d'avoir une vision globale de l'impact des variables quantitatives, nous réalisons une régression linéaire multiple, qui permet d'évaluer l'effet combiné de plusieurs variables sur une variable cible – ici, l'**indice de performance**.

	Coef.	Std.Err.	t	p-value	[0.025	0.975]
<b>Monthly_Salary</b>	0.000526	0.000003	187.507039	0.000000	0.000520	0.000531
<b>Work_Hours_Per_Week</b>	-0.000699	0.000430	-1.624142	0.104349	-0.001542	0.000145
<b>Projects_Handled</b>	0.000157	0.000266	0.591203	0.554386	-0.000364	0.000679
<b>Overtime_Hours</b>	0.000041	0.000444	0.091875	0.926797	-0.000830	0.000911
<b>Sick_Days</b>	0.000378	0.000888	0.425024	0.670820	-0.001364	0.002119
<b>Remote_Work_Frequency</b>	0.000077	0.000109	0.711313	0.476892	-0.000136	0.000291
<b>Team_Size</b>	-0.001726	0.000700	-2.465099	0.013699	-0.003099	-0.000354
<b>Training_Hours</b>	0.000143	0.000133	1.073611	0.283000	-0.000118	0.000404
<b>Promotions</b>	-0.004326	0.004717	-0.917086	0.359100	-0.013570	0.004919
<b>Employee_Satisfaction_Score</b>	0.001423	0.003344	0.425529	0.670452	-0.005132	0.007978
<b>Years_At_Company</b>	0.000963	0.001341	0.718044	0.472732	-0.001666	0.003592
<b>Age</b>	0.000018	0.000342	0.053433	0.957387	-0.000653	0.000689

Tableau 3 : Régression multivariée sur le Score de Performance

Nous nous intéressons uniquement aux variables dont la **valeur p** est inférieure à 0.05, ce qui signifie que leur influence est statistiquement significative. Dans notre cas, seules deux variables remplissent ce critère : **Monthly\_Salary** et **Team\_Size**.

Le coefficient associé au salaire mensuel peut paraître faible (0.0005), mais il s'applique à une variable aux valeurs élevées. Ainsi, une variation réaliste du salaire entraîne une variation notable du score de performance. Ce type de lecture rappelle que la signification d'un coefficient dépend autant de sa valeur que de l'échelle de la variable concernée.

De son côté, **Team\_Size** affiche un coefficient négatif (-0.0017), ce qui suggère qu'à mesure que la taille de l'équipe augmente, le score de performance individuel tend à diminuer légèrement.

Enfin, le **coefficient de détermination**  $R^2 = 0.26$  indique que l'ensemble de nos variables explicatives parvient à expliquer **26% de la variance observée** du score de

performance. Cela reste une proportion honorable dans un contexte RH, où de nombreux facteurs externes ou qualitatifs (non mesurés ici) peuvent influencer la performance.

### 3. Analyse par département

Nous poursuivons notre analyse en introduisant la dimension des départements. Pour cela, nous réalisons une régression linéaire (OLS) sur le score de performance, en isolant chaque département. Cette approche permet d'identifier les spécificités propres à chaque unité, qui pouvaient être noyées dans l'analyse globale.

Pour chaque département, deux résultats clés sont obtenus :

Le **R<sup>2</sup>**, qui indique la part de variance du score de performance expliquée par nos variables descriptives.

La liste des **variables significatives**, c'est-à-dire celles dont l'impact sur la performance est statistiquement prouvé (p-value < 0.05).

IT	→ R <sup>2</sup> : 0.267	Variables significatives : Monthly_Salary, Years_At_Company
Finance	→ R <sup>2</sup> : 0.256	Variables significatives : Monthly_Salary, Team_Size
Customer Support	→ R <sup>2</sup> : 0.268	Variables significatives : Monthly_Salary, Sick_Days
Engineering	→ R <sup>2</sup> : 0.256	Variables significatives : Monthly_Salary
Marketing	→ R <sup>2</sup> : 0.256	Variables significatives : Monthly_Salary
HR	→ R <sup>2</sup> : 0.257	Variables significatives : Monthly_Salary
Operations	→ R <sup>2</sup> : 0.259	Variables significatives : Monthly_Salary
Sales	→ R <sup>2</sup> : 0.261	Variables significatives : Monthly_Salary
Legal	→ R <sup>2</sup> : 0.267	Variables significatives : Monthly_Salary, Sick_Days

Tableau 4 : Régression linéaire du score de performance par département avec identification des variables significatives

Les résultats montrent que le R<sup>2</sup> reste relativement stable, autour de 0.26, ce qui suggère une cohérence globale : dans chaque département, nos variables expliquent environ 26 % du score de performance. En revanche, on observe l'apparition de nouvelles variables explicatives selon les départements. Par exemple, *Sick\_Days* devient significatif dans *Customer Support* et *Legal*, alors qu'il ne l'était pas au niveau global.

Cela s'explique par le fait qu'en segmentant par département, on réduit l'effet de dilution dû à la diversité des situations. Certaines variables, peu visibles à l'échelle globale, gagnent en importance une fois replacées dans leur contexte métier. Cela confirme la pertinence d'une analyse segmentée pour mieux comprendre les leviers de performance propres à chaque service.

Enfin, il est intéressant de noter que le nombre de variables significatives n'a pas de lien direct avec le  $R^2$  obtenu. Un département avec une seule variable explicative peut avoir un  $R^2$  comparable à un autre en ayant deux. Cela signifie que certaines variables, bien que significatives, n'apportent qu'une faible contribution à la variance expliquée.

## b. Analyse des variables qualitatives

### 1. Répartition des variables qualitatives

Nous poursuivons l'analyse en nous intéressant à la répartition des variables qualitatives au sein des départements. Cette étape permet de mieux comprendre la variation des caractéristiques et d'identifier d'éventuelles tendances ou disparités.

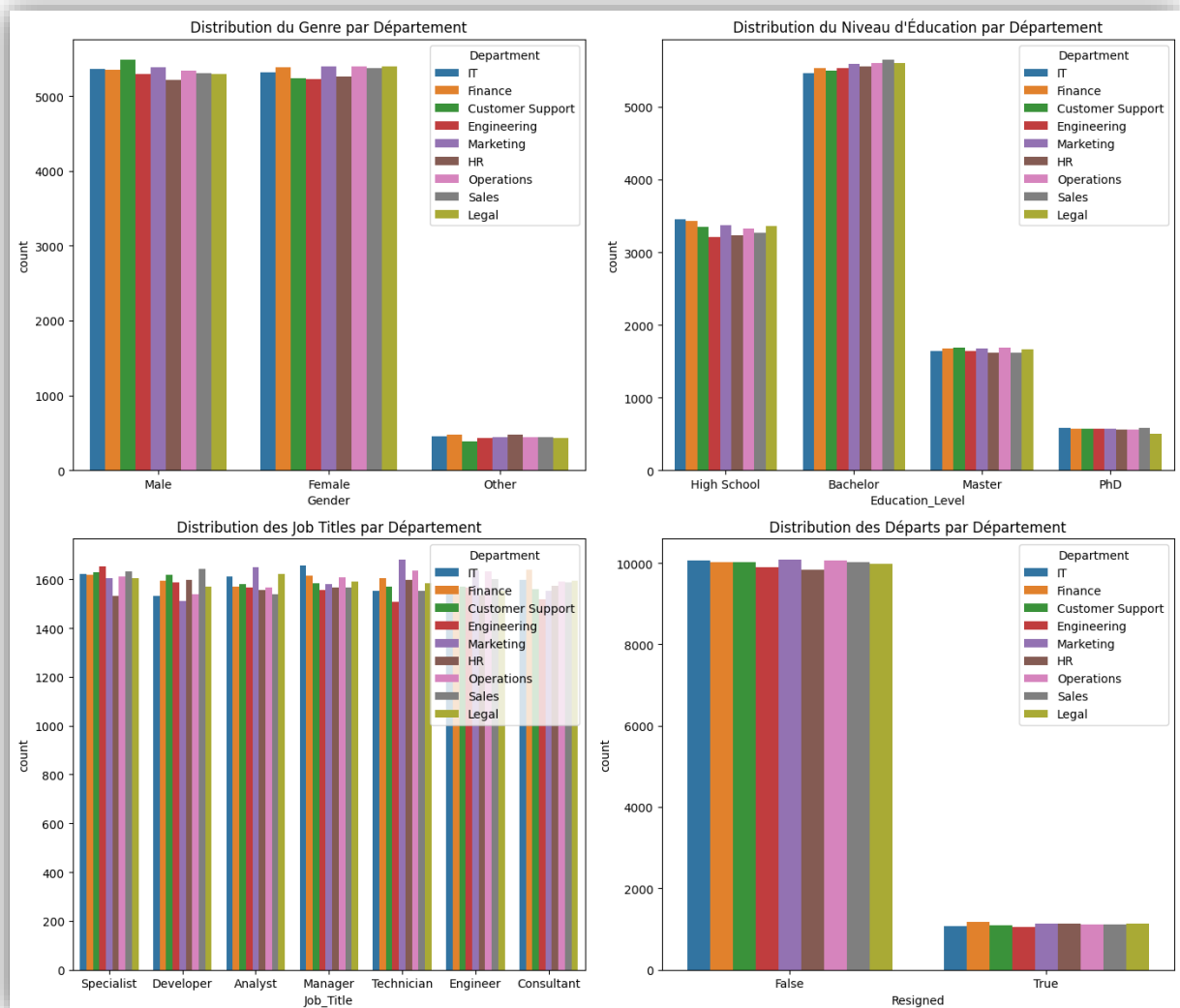


Figure 3 : Distribution des variables qualitatives par département

- **Genre** : La répartition par genre est équitable entre les départements, avec un pourcentage similaire d'hommes et de femmes. La catégorie "Other" reste faible, ce qui est un résultat attendu compte tenu de la faible représentation de cette catégorie dans l'ensemble.
- **Niveau d'Étude** : Le niveau d'étude "Bachelor" est majoritaire, représentant près de 50% des effectifs dans chaque département. La catégorie "High School" enregistre environ 3000 employés par département. La distribution des niveaux d'études présente une homogénéité remarquable entre les départements, quel que soit le niveau d'étude (PhD, Master, Bachelor, ou High School).
- **Postes** : Les intitulés de postes ("Job\_Title") montrent une répartition homogène à travers les départements, sans variation notable.
- **Résignations** : Les données sur les départs volontaires ("Resigned") indiquent une répartition stable, avec environ 1000 départs dans chaque département, sans variation significative entre eux.

## 2. Test d'indépendance ( $\chi^2$ ) global

Nous poussons nos recherches et réalisons un test de  $\chi^2$  afin de déterminer si parmi ces variables certaines sont liées au Performance Score :

```
Test Chi-carré entre 'Job_Title' et 'Performance_Score':
Chi²: 22.5891, p-value: 0.5441
=====
Test Chi-carré entre 'Gender' et 'Performance_Score':
Chi²: 15.6734, p-value: 0.0473
=====
Test Chi-carré entre 'Education_Level' et 'Performance_Score':
Chi²: 20.0170, p-value: 0.0668
=====
Test Chi-carré entre 'Resigned' et 'Performance_Score':
Chi²: 3.1946, p-value: 0.5258
=====
Test Chi-carré entre 'Department' et 'Performance_Score':
Chi²: 29.8401, p-value: 0.5763
```

Tableau 5 : Test du  $\chi^2$  sur les variables qualitatives

Deux variables ressortent particulièrement de cette analyse : le genre et le niveau d'études. Les p-values associées à ces variables sont significatives, avec une p-value pour le genre inférieure à 0.05, respectant ainsi le seuil de significativité. En revanche,

le niveau d'études à une p-value de 0.06, ce qui ne respecte pas strictement le seuil de 0.05, mais reste suffisamment proche pour justifier son inclusion dans l'analyse.

Cela nous permet donc d'identifier le genre comme une variable significative à inclure dans l'évaluation de la performance des départements. Le niveau d'études, bien que légèrement au-dessus du seuil, peut également être pris en compte dans l'interprétation des résultats.

### 3. Test d'indépendance par département

Nous pouvons maintenant aller plus loin en réalisant un test du Chi<sup>2</sup> pour examiner l'indépendance de ces variables par rapport aux départements. Cela nous donnera une perspective supplémentaire sur la relation entre ces facteurs et la performance des départements.

Department	Variable	Chi2	p-value	Degrees_of_Freedom
HR	Resigned	10.561000	0.032000	4
HR	Job_Title	37.966200	0.034900	24
IT	Job_Title	34.237100	0.080500	24
Operations	Gender	12.940300	0.113900	8
Marketing	Gender	12.637400	0.125000	8

Tableau 6 : Cinq relations les plus significatives entre les variables qualitatives et le Score de Performance

Afin de rendre l'analyse plus lisible, nous nous limitons à l'affichage des cinq premières relations les plus significatives entre les variables qualitatives et le Score de Performance. Les résultats les plus intéressants étant inclus dans cette sélection, cela ne restreint pas l'analyse, tout en évitant un affichage trop volumineux.

L'analyse des relations entre les variables qualitatives (Job\_Title, Gender, Education\_Level, Resigned) et le Performance\_Score, menée séparément par département, indique que **seul le département HR présente des relations statistiquement significatives** (p-value < 0.05).

Dans ce département, les variables Job\_Title et Resigned sont liées au score de performance.

Il est à noter que la variable `Education_Level`, qui présentait une p-value proche du seuil de significativité (0.06) dans l'analyse globale, **n'a pas confirmé cette tendance lorsqu'elle a été examinée séparément par département**. À chaque segmentation, les p-values associées à cette variable sont restées **largement supérieures à 0.05**, ce qui ne permet pas de retenir une relation significative entre le niveau d'éducation et le score de performance dans un cadre départemental. Une situation similaire a été observée pour la variable `Gender`, dont la p-value sur le test du Chi<sup>2</sup> global était à 0.04. Même si par départements ses valeurs sont relativement proches du seuil de significativité (0.11 à 0.15) elles restent néanmoins **trop élevées** pour conclure à une relation statistiquement significative.

Ces résultats suggèrent que **ces variables n'ont pas un impact significatif** sur la performance, du moins dans le cadre de cette **analyse segmentée par département**.

**Aucune autre relation significative n'a été observée dans les autres départements.**

Cette analyse met en évidence une **disparité entre départements** quant à la présence de variables qualitatives associées à la performance.

### c. Résultats d'analyse

L'analyse des variables quantitatives met en évidence **des résultats contrastés**. À l'échelle globale, seule la variable `Monthly_Salary` présente une corrélation notable avec le `Performance_Score`, avec un coefficient de 0.51, tandis que les autres variables affichent des corrélations proches de zéro. Cette observation est confirmée par la régression multivariée, qui retient également `Team_Size` comme variable significative, bien que son effet reste modéré. Lorsqu'on isole les départements, le R<sup>2</sup> reste stable autour de 0.26, ce qui indique une capacité explicative comparable dans chaque unité. Toutefois, de **nouvelles variables deviennent significatives localement**, comme `Sick_Days` dans *Customer Support* et *Legal*, `Years_at_Company` dans *IT*, ou encore `Team_Size` dans *Finance*. Ces résultats révèlent que **certains effets se manifestent uniquement dans des contextes spécifiques**, masqués dans l'analyse globale.

Les variables qualitatives présentent une **distribution relativement homogène** entre départements. Le test du Chi<sup>2</sup> global révèle une association significative entre le genre et le score de performance, tandis que le niveau d'étude approche le seuil de significativité sans le franchir. En affinant l'analyse **au niveau départemental, seules deux relations apparaissent significatives**, toutes deux dans le département *RH* : les variables `Job_Title` et `Resigned` y présentent un lien statistique avec le `Performance_Score`. Aucune autre association notable n'est détectée dans les

autres départements, y compris pour `Education_Level`, dont l'effet ne se confirme pas localement.

Dans l'ensemble, les données quantitatives permettent d'identifier quelques facteurs explicatifs, dont certains conservent leur influence à travers les départements, tandis que d'autres **ne deviennent visibles qu'une fois les unités analysées séparément**. Les données qualitatives, plus constantes dans leur répartition, révèlent peu de relations significatives avec la performance, en dehors de cas très localisés. Ces résultats illustrent **les limites des variables observées** pour expliquer la performance individuelle, tout en soulignant l'intérêt d'analyses segmentées.

### C. Identification des profils par départements

Après avoir déterminé les variables significatives pour nos différents départements, l'étape suivante consiste à identifier des profils types. Pour chaque département, nous calculerons le  $R^2$  afin de mesurer l'impact des variables explicatives sur la variance du score de performance. Ce calcul nous permettra de quantifier la part de la performance expliquée par les variables sélectionnées, et d'évaluer si certains départements présentent des profils distincts de surperformance ou de sous-performance en fonction de leur capacité explicative.

#### a. Variables significatives par département

À partir des analyses précédentes, nous pouvons dresser un tableau récapitulatif associant à chaque département les variables explicatives du score de performance:

- IT : `Monthly_Salary`, `Years_At_Company`
- Finance : `Monthly_Salary`, `Team_Size`
- Customer Support : `Monthly_Salary`, `Sick_Days`
- Engineering : `Monthly_Salary`
- Marketing : `Monthly_Salary`
- HR : `Monthly_Salary`, `Job_Title`, `Resigned`
- Operations : `Monthly_Salary`
- Sales : `Monthly_Salary`
- Legal : `Monthly_Salary`, `Sick_Days`

Trois groupes se dégagent selon le nombre de variables explicatives identifiées.



Département: IT  
Monthly\_Salary:  $R^2 = 0.266$   
Years\_At\_Company:  $R^2 = 0.0$

Département: Finance  
Monthly\_Salary:  $R^2 = 0.255$   
Team\_Size:  $R^2 = 0.0$

Département: Customer Support  
Monthly\_Salary:  $R^2 = 0.267$   
Sick\_Days:  $R^2 = 0.0$

Département: Engineering  
Monthly\_Salary:  $R^2 = 0.256$

Département: Marketing  
Monthly\_Salary:  $R^2 = 0.255$

Département: HR  
Monthly\_Salary:  $R^2 = 0.256$   
JobTitle\_Analyst:  $R^2 = 0.0$   
JobTitle\_Consultant:  $R^2 = 0.001$   
JobTitle\_Developer:  $R^2 = 0.0$   
JobTitle\_Engineer:  $R^2 = 0.0$   
JobTitle\_Manager:  $R^2 = 0.0$   
JobTitle\_Specialist:  $R^2 = 0.0$   
JobTitle\_Technician:  $R^2 = 0.0$   
Resigned\_ID:  $R^2 = 0.0$

Département: Operations  
Monthly\_Salary:  $R^2 = 0.259$

Département: Sales  
Monthly\_Salary:  $R^2 = 0.26$

Département: Legal  
Monthly\_Salary:  $R^2 = 0.266$   
Sick\_Days:  $R^2 = 0.0$

Tableau 7 : Résultats du  $R^2$  sur les variables explicatives pas département

Dans un premier temps, nous avons cherché à analyser les  $R^2$  pour chaque département afin d'identifier l'impact des variables explicatives sur le score de performance. Pour obtenir les résultats les plus précis possibles, nous avons appliqué du **one-hot encoding** sur des variables telles que Gender (même si elle ne s'avère pas significative ici) et Job\_Title. Cette méthode a permis d'éviter d'attribuer une importance injustifiée à certaines catégories de ces variables, en leur attribuant des valeurs binaires équivalentes et en traitant chaque catégorie sur un pied d'égalité.

Bien que certaines variables, comme **le salaire** (ce qui était attendu), montrent un  **$R^2$  modéré**, l'ajout d'autres variables significatives n'a pas permis d'améliorer substantiellement la capacité explicative des modèles. Les résultats obtenus restent globalement faibles, ce qui indique qu'aucune relation marquée entre les variables et la performance n'émerge à ce stade. Par conséquent, il n'est pas possible d'établir des profils distincts entre les départements.

La suite logique, dans le cas où certaines variables auraient montré des relations plus intéressantes, aurait été d'analyser **le recouvrement** et les **VIF** pour évaluer si certains départements pouvaient être mieux expliqués. Toutefois, étant donné que ce n'est pas le cas, il n'y a pas lieu de pousser davantage l'analyse sur ce point, et nous passons donc à la suite.

Même avec les variables censées être explicatives de chaque département, il est impossible d'identifier des profils distincts, ce qui rend **difficile la distinction** entre départements **surperformants** ou **sous-performants**.

## b. Synthèse des profils

L'analyse par département n'a pas permis d'identifier des profils explicatifs distincts. Malgré la présence de variables significatives dans certains cas, les  $R^2$  obtenus restent faibles, ne permettant pas de dégager une structure claire de la performance. Aucune surperformance ni sous-performance n'a pu être mise en évidence.

## D. Synthèse et recommandations

L'ensemble des analyses menées ne permet pas d'identifier des profils types ou des logiques de performance différenciées entre les départements. La variable « Monthly\_Salary » présente un pouvoir explicatif modéré ( $R^2 \approx 0.26$ ) et relativement stable dans la majorité des départements. En revanche, l'ajout de variables pourtant identifiées comme significatives (telles que « Team\_Size », « Years\_At\_Company » ou « Sick\_Days ») n'a pas permis d'améliorer substantiellement la capacité explicative des modèles. Les  $R^2$  obtenus restent faibles, parfois nuls, traduisant une hétérogénéité des résultats et une absence de relation marquée entre les variables testées et la performance. Dans ces conditions, il n'est pas possible de mettre en évidence des profils distincts ni de conclure à une surperformance ou une sous-performance de certains départements.

Ces constats suggèrent que les relations mises en évidence jusqu'à présent ne suffisent pas à expliquer de manière satisfaisante les variations du score de performance, et qu'elles sont probablement influencées par d'autres dynamiques plus complexes, non captées à ce stade. Dans cette perspective, il serait pertinent de recentrer les analyses futures directement sur ce score, en s'autorisant des hypothèses explicatives plus ciblées et contextuelles. Cela pourrait inclure l'examen d'effets non linéaires, l'analyse de sous-groupes d'employés partageant certaines caractéristiques, ou encore la prise en compte de dynamiques temporelles comme l'évolution de la performance en fonction de l'ancienneté. Cette phase d'approfondissement, plus interprétative, permettra le cas échéant de dégager des leviers d'action concrets, mieux ancrés dans les réalités propres à l'organisation.

## V. Analyse approfondie

### A. Une distribution étonnamment homogène

La répartition des scores de performance est étonnamment homogène entre les départements, avec une distribution équilibrée entre les niveaux 1 à 5. Ce type de régularité laisse penser que le **processus d'évaluation suit potentiellement une logique normée ou contrainte**, comme le respect de quotas implicites ou explicites. Cette structuration des données **pourrait expliquer** en partie la **faiblesse des relations statistiques observées**, notamment le manque de variation dans les scores selon les facteurs analysés. Cela constitue une limite importante à la capacité explicative des modèles, et suggère que le score de performance, tel qu'il est attribué, pourrait ne pas refléter une variabilité « naturelle » de la performance individuelle.

Afin de valider cette hypothèse, un test de normalité a été effectué. Ce test nous fournit deux résultats clés :

- **Statistique de test (W)** : Plus cette valeur se rapproche de 1, plus la distribution des données est proche de la normale.
- **P-value** : Si elle est inférieure à 0.05, nous rejetons l'hypothèse nulle, ce qui signifie que les données ne suivent pas une distribution normale.

```
- Statistique de test (W) : 0.888  
- P-value : 0.000
```

Tableau 8 : Test de normalité avec Shapiro-Wilk

Les résultats du test confirment nos suspicions concernant la répartition biaisée des valeurs, ce qui pourrait expliquer les faibles relations statistiques observées précédemment. Il serait donc pertinent, dans d'éventuelles analyses futures, de prendre en compte cette structuration probable des données, afin de mieux évaluer dans quelle mesure elle limite l'interprétation des résultats

### B. Une variable dominante : le salaire mensuel

Dans un contexte où la performance des employés est mesurée à l'aide du `Performance Score`, il est essentiel de comprendre les facteurs qui influencent cette

évaluation. Bien que de nombreuses variables aient été explorées, une analyse plus approfondie s'avère nécessaire pour identifier des relations plus nuancées, notamment en ce qui concerne les variables quantitatives. Parmi celles-ci, le `Monthly Salary` semble être une donnée particulièrement pertinente, car sa corrélation avec le score de performance suggère qu'il pourrait jouer un rôle dans l'évaluation globale des employés. Cependant, étant donné les caractéristiques des données et la distribution homogène du score de performance, il est crucial de tester l'influence du salaire sur la performance en fonction des départements. Cette segmentation permettra de mieux comprendre comment cette variable interagit dans des contextes différents, et si des ajustements sont nécessaires dans l'interprétation des résultats obtenus. Cette analyse approfondie vise ainsi à explorer les corrélations entre le salaire et le `Performance Score` dans le but de mieux appréhender les dynamiques sous-jacentes au sein de chaque département.

Tableau 9 : Corrélations du `Monthly_Salary` global et par département

<code>Performance_Score</code>	0.516683	0.505659	0.505038	0.506381	0.516064	0.515997	0.505427	0.508602	0.510131
<code>Projects_Handled</code>	0.007836	0.001688	-0.015103	0.002047	-0.021588	0.001744	0.013260	0.000200	-0.007309
<code>Overtime_Hours</code>	0.002146	-0.002848	-0.011801	0.003620	-0.005388	-0.007546	0.008847	-0.003907	-0.010735
<code>Work_Hours_Per_Week</code>	0.003366	-0.012524	-0.001103	0.013195	0.005146	-0.003069	-0.005576	-0.019843	-0.001006
<code>Years_At_Company</code>	0.027806	0.002532	-0.008983	-0.010319	-0.002261	-0.017724	-0.005261	0.002526	0.006285
<code>Remote_Work_Frequency</code>	0.007059	0.001577	-0.010381	-0.012124	0.007160	-0.021053	-0.009558	0.014756	0.018090
<code>Training_Hours</code>	-0.004394	-0.006033	-0.008346	-0.000852	-0.009689	0.002989	0.009237	0.003580	0.003513
<code>Promotions</code>	0.003585	0.005118	-0.001720	-0.015847	0.000017	-0.007904	-0.001033	0.008811	-0.008682
<code>Employee_Satisfaction_Score</code>	-0.004783	0.000995	-0.001874	0.005710	-0.013915	0.011525	-0.006405	0.016399	0.001769
<code>Age</code>	0.009023	0.008412	0.016367	-0.006548	0.007005	0.004631	0.003130	-0.007647	-0.009411

Les résultats sont à la fois attendus et surprenants : la corrélation entre le score de performance et le salaire était anticipée, mais il est surprenant de constater l'**absence totale de corrélation** avec les autres variables. Cela soulève des questions importantes sur la pertinence et la justesse des critères d'évaluation utilisés.

En particulier, malgré l'absence de corrélation notable avec de nombreuses variables, il est intéressant de noter que celles liées aux efforts fournis par les employés, telles que `Projects_Handled`, `Overtime_Hours`, et `Training_Hours`, ne montrent aucune influence sur le salaire et donc le score de performance. Cette situation suggère que, peu importe l'implication ou les efforts fournis, ceux-ci ne sont pas reconnus dans l'évaluation, ce qui amène à penser que le score de performance ne reflète pas nécessairement la contribution réelle des employés.

### C. Des profils éducatifs peu différenciés selon les fonctions

Dans cette analyse, nous cherchons à comprendre si les pratiques de recrutement et d'évaluation des performances reflètent réellement les spécificités des différents départements. Une question clé est de savoir si les niveaux d'études et les profils des employés correspondent aux exigences et aux attentes propres à chaque fonction. Les résultats que nous allons explorer mettent en lumière des anomalies et des **uniformités surprenantes** dans la répartition des niveaux d'études par département, suggérant un manque de différenciation qui mérite une réflexion plus approfondie.

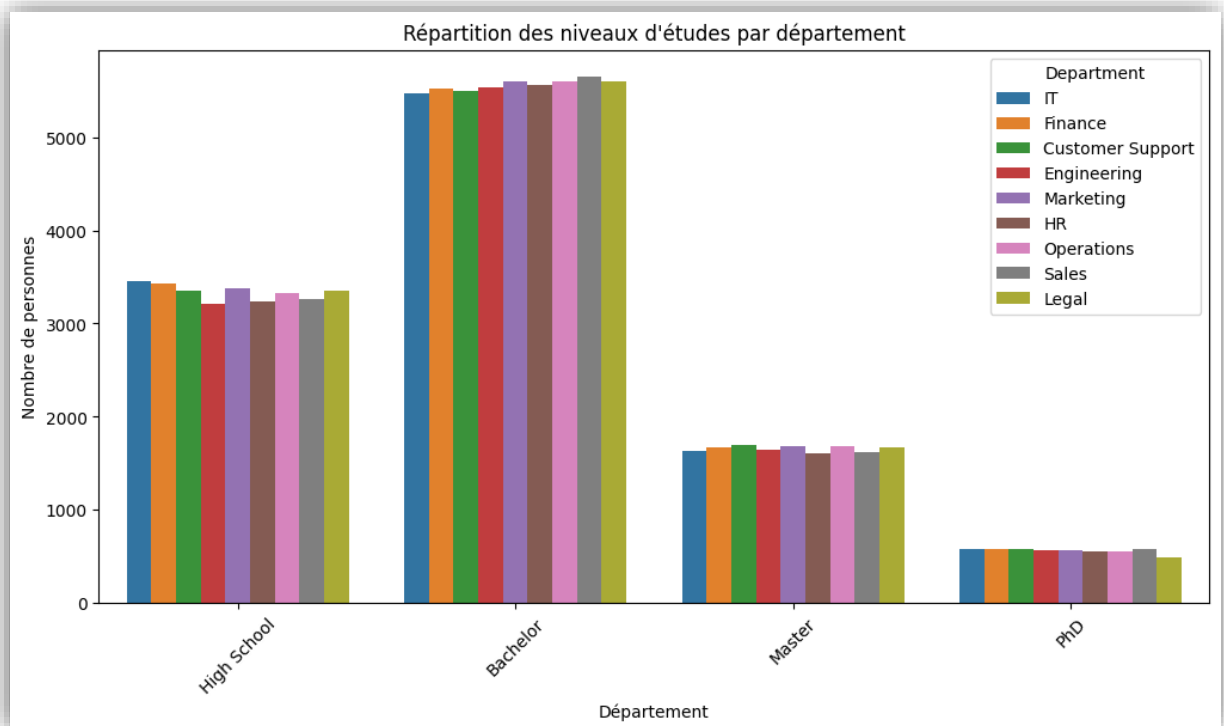


Figure 4 : Histogramme de la répartition des niveaux d'études par département

Une observation marquante dans la répartition des niveaux d'études par département est l'uniformité des données, qui révèle un manque de différenciation entre les différents corps de métiers. Par exemple, dans des départements comme le *Sales* ou le *Customer Support*, on pourrait s'attendre à ce que les niveaux d'études soient principalement centrés autour de diplômes de premier cycle ou de niveau licence, avec peu de présence de doctorats. Pourtant, la répartition des niveaux d'études montre une homogénéité surprenante entre ces départements et des secteurs comme *IT*, *Finance* ou *Engineering*, qui, eux, exigeraient logiquement des qualifications plus élevées, telles que des masters ou des doctorats.

Cette absence de différenciation pourrait signaler une **politique de recrutement trop standardisée**, ne prenant pas en compte les spécificités des compétences et

qualifications requises pour chaque département. D'une part, des postes dans des domaines techniques ou spécialisés devraient attirer des profils avec des niveaux d'études plus élevés. D'autre part, cette uniformité pourrait également pointer un **manque de reconnaissance des compétences spécifiques requises** par chaque fonction dans le cadre des évaluations. Si le score de performance ne tient pas compte des particularités de chaque département, il risque de sous-estimer les contributions réelles des employés, ce qui accentue la perception d'un manque de différenciation dans le traitement des profils.

Cette homogénéité des niveaux de qualification entre des départements aux exigences techniques très différentes soulève une forme d'incohérence vis-à-vis des scores de performance observés. En effet, des départements comme *Sales* ou *Customer Support*, où l'on s'attendrait à des niveaux d'études globalement plus bas, présentent une distribution similaire à celle de *l'IT* ou *l'Engineering*, sans pour autant afficher une performance supérieure ou différenciée. Cela pourrait indiquer que, d'une part, les départements les plus opérationnels présentent une **sous-performance au regard du niveau de qualification des profils recrutés**, et que, d'autre part, les départements plus techniques affichent une **sur-performance relative**, en valorisant plus efficacement ces qualifications.

En somme, bien que les données ne permettent pas d'établir de conclusions fermes, deux profils de départements semblent émerger à travers cette lecture : d'un côté, des départements à dominante opérationnelle, potentiellement en **sous-performance** au regard des profils recrutés ; de l'autre, des départements plus techniques, qui pourraient présenter une **sur-performance relative**, suggérant une meilleure adéquation entre niveau de qualification et performance mesurée. Ces constats restent à nuancer, mais ils offrent une grille d'interprétation intéressante pour de futures analyses plus ciblées.

## VI. Recommandations

### A. Revoir le système de notation

La distribution trop homogène des scores de performance, observée quel que soit le département, suggère l'existence d'un système d'évaluation normé, voire contraint. Il serait pertinent de réviser ce système pour y introduire davantage de variabilité réelle, permettant de mieux distinguer les niveaux de performance individuelle et de rendre les scores exploitables à des fins analytiques et managériales.

## B. Améliorer la qualité et la granularité des données

La faiblesse des corrélations entre les variables étudiées et le score de performance souligne un manque de données explicites et nuancées. L'enrichissement des bases de données, notamment en intégrant des variables plus représentatives de la réalité du travail (efforts fournis, contexte de mission, feedback qualitatif), permettrait de renforcer la capacité explicative des analyses futures.

## C. Réviser la grille salariale à partir de critères plus fins

Les analyses ont mis en évidence l'absence d'impact des efforts individuels (projets menés, heures supplémentaires, formations suivies) sur la rémunération. Il serait donc pertinent de repenser la grille salariale en y intégrant des critères valorisant l'engagement et les compétences développées, tout en tenant compte du niveau d'études requis par poste.

## D. Réévaluer les niveaux de qualification attendus par département

La répartition homogène des niveaux d'études entre départements très différents interroge sur la pertinence des exigences en matière de formation. Une meilleure adéquation entre le niveau de qualification requis et les compétences réellement nécessaires permettrait d'optimiser les recrutements, de limiter la surqualification et de renforcer la cohérence entre profil et performance attendue.

## E. Mettre en place un système de satisfaction employé plus nuancé et personnalisé

Au-delà de la performance mesurée, la compréhension fine des leviers de motivation et de satisfaction des employés est essentielle. Un dispositif d'écoute plus qualitatif, adapté aux spécificités de chaque département et centré sur les besoins individuels, permettrait de mieux piloter la fidélisation et d'anticiper les risques de désengagement.

# VII. Conclusion

Cette analyse s'inscrit dans une démarche d'amélioration continue engagée par l'entreprise, soucieuse de mieux comprendre les écarts de performance entre ses départements. En s'appuyant sur un ensemble de variables qualitatives et

quantitatives, l'étude avait pour objectif d'identifier les leviers internes susceptibles d'expliquer les différences observées, tant en termes de productivité que de satisfaction des collaborateurs.

Notre analyse n'a pas mis en évidence de variation notable des scores de performance entre les départements. Quelle que soit la variable considérée, les résultats restent particulièrement homogènes d'un service à l'autre. Cette absence de différenciation empêche d'identifier des facteurs explicatifs clairs, et limite la portée des analyses statistiques réalisées.

Au-delà des performances elles-mêmes, c'est la structure globale des données qui interpelle. Qu'il s'agisse des niveaux d'études, des salaires, des effectifs ou des indicateurs d'implication individuelle, toutes les variables présentent une répartition étonnamment stable entre les départements, indépendamment des fonctions exercées. Cette homogénéité généralisée suggère que les pratiques de gestion et d'évaluation appliquent une logique uniforme, qui ne semble pas tenir compte des particularités propres à chaque métier ou département.

Bien que nos analyses ne permettent pas d'identifier des profils types de départements en surperformance ou sous-performance, certains éléments de réflexion émergent. D'un côté, les départements aux fonctions plus opérationnelles, où l'on pourrait s'attendre à des niveaux de formation plus modestes, affichent une performance comparable à celle de départements plus techniques, malgré un niveau de qualification équivalent. De l'autre, cette situation peut également indiquer que les départements techniques parviennent à mieux valoriser le capital de compétences mobilisé. Ces interprétations restent cependant spéculatives, en l'absence de données réellement discriminantes.

En définitive, si l'analyse ne permet pas de dresser des profils explicites à partir des données disponibles, elle met clairement en lumière un enjeu structurel : l'outil de mesure de la performance, dans sa forme actuelle, limite fortement la capacité à produire des analyses pertinentes, nuancées et exploitables. Cela appelle à une révision des indicateurs mobilisés et à une réflexion plus large sur les critères d'évaluation de la performance au sein de l'organisation.