

RAPPORT DE STAGE 2^e ANNEE

Etude de l'évolution des salaires sur le marché du travail en France et des déterminants du salaire : Cas des offres d'emploi en ligne

Année académique 2022 - 2023

Auteur :
Isaac GANIYU

Maître de stage :
Jean-Charles BRICONGNE
Reférent pédagogique :
Yamane IKKO

Septembre 2023

Table des matières

Remerciements	4
Résumé - Abstract	5
Introduction	6
1 Environnement du stage	7
2 La problématique et les données	8
2.1 Présentation de la problématique et source des données	8
2.2 Revue de littérature	8
2.3 Processus d'acquisition des données	10
2.4 Description et structure des données	11
2.4.1 Description des données recoltées	11
2.4.2 Traitement des données	12
2.5 Statistiques descriptives	13
2.5.1 Structure des données	13
2.5.2 Qualité des données	16
2.5.3 Quelques statistiques générales sur les salaires	17
2.5.4 Une hétérogénéité dans les niveaux de salaire	18
3 L'évolution des salaires en France en 2023	21
3.1 Représentation graphique	21
3.2 Stationnarité des séries temporelles	22
3.3 Modélisation de la série temporelle	23
3.3.1 La série des salaires moyens	23
3.3.2 La série des salaires médians	25
3.4 Prévision du niveau de salaire	27
4 Les déterminants du salaire : un modèle économétrique	28
4.1 Un premier modèle : estimation par moindres carrés ordinaires	28
4.1.1 Ecriture et analyse du modèle	28
4.1.2 Validation du modèle	29
4.2 Le modèle final : estimation par moindres carrés généralisés	30
4.2.1 Ecriture et estimation du modèle	30
4.2.2 Interprétation des résultats	30
Conclusion	32
Bibliographie	33
Annexes	34
Statistiques descriptives	34
Evolution du niveau de salaire	38
Les déterminants du salaire	39

Table des figures

1	Nombre d'offres d'emploi disponibles par site d'emploi	8
2	Nombre d'offres d'emploi par mois	13
3	Nombre d'offres d'emploi par département	14
4	Nombre d'offres d'emploi par departement normalisé par la population active	14
5	Nombre d'offres d'emploi par tranches de salaire	15
6	Nombre d'offres par catégorie socio-professionnelle	15
7	Nombre d'offres par qualification	15
8	Nombre d'offres d'emploi par groupe de métiers	16
9	Comparaison avec les données de l'Insee	16
10	Distribution des salaires mensuels selon l'Insee. Source : Insee, Base Tous salariés 2021	17
11	Boxplot de la distribution des salaires	17
12	Percentiles de la distribution des salaires	17
13	Distribution des salaires	18
14	Salaire moyen par département	18
15	Salaire moyen par secteur d'activité	19
16	salaire moyen par catégorie socio-professionnelle	19
17	Salaire moyen par qualification	19
18	Salaire moyen par durée du travail	20
19	Salaire moyen selon alternance	20
20	Salaire moyen par expérience demandée	20
21	Evolution journalière salaire	21
22	Evolution journalière du salaire moyen	22
23	Evolution journalière du salaire médian	22
24	Evolution journalière du salaire moyen différencié	23
25	Evolution journalière du salaire médian différencié	23
26	Graphe des autocorrélations de la série différenciée	24
27	Graphe des autocorrélations partielles de la série différenciée	24
28	Graphe d'autocorrélation des résidus du modèle ARIMA(1,1,0)	25
29	Graphe d'autocorrélation des résidus du modèle ARIMA(6,1,0)	25
30	Graphe des autocorrélations de la série différenciée du salaire médian . . .	26
31	Graphe des autocorrélations partielles de la série différenciée du salaire médian	26
32	Graphe des autocorrélations des résidus de la série	26
33	Prévision de la série du salaire moyen	27
34	Prévision de la série du salaire moyen	27
35	Estimation de la densité des résidus	29
36	Q-Q plot des résidus	29
37	Nombre d'offres par type de contrat	34
38	Nombre d'offres par secteur	34
39	Nombre d'offres par domaine professionnel	35
40	Salaire médian par département	35
41	Salaire médian par secteur d'activité	35
42	Salaire moyen par groupe de métiers	36
43	Salaire moyen par type de contrat	36

44	test de la présence d'une tendance linéaire dans la série des moyennes . . .	38
45	test de la présence d'une tendance linéaire dans la série des médianes . . .	38
46	test de la présence d'une tendance linéaire dans la série différenciée des moyennes	38
47	test de la présence d'une tendance linéaire dans la série différenciée des médianes	38
48	Test de stationnarité de la série des moyennes	39
49	Test de stationnarité de la série des médianes	39
50	Estimation du modèle ARIMA(1,1,0) de la série du salaire moyen	39
51	Estimation du modèle ARIMA(6,1,0) de la série du salaire moyen	39
52	Estimation du modèle ARIMA(1,1,0) de la série du salaire médian	39
53	Test d'homoscédasticité des résidus de la régression par moindres carrés ordinaires	39
54	Test d'autocorrélation des résidus de la régression par moindres carrés or- dinares	39
55	Résultats de la régression avec moindres carrés généralisés	40
56	Résultats de la régression avec moindres carrés généralisés	40
57	Résultats de la régression avec moindres carrés généralisés	41
58	Résultats de la régression avec moindres carrés généralisés	41
59	Résultats de la régression avec moindres carrés généralisés	42
60	Résultats de la régression avec moindres carrés généralisés	42
61	Résultats de la régression avec moindres carrés généralisés	43
62	Résultats de la régression avec moindres carrés généralisés	43

Liste des tableaux

1	P-valeurs obtenus à travers les régressions sur la variable temporelle	22
2	Estimation des paramètres du modèle	24
3	Estimation des paramètres du modèle	25
4	20 emplois les mieux rémunérés selon nos données	37
5	Salaires mensuels moyens en EQTP en 2021 selon l'Insee	37

Remerciements

Avant tout propos, je tiens à remercier les personnes ayant contribué, de près ou de loin, à la réalisation de cette étude et au bon déroulement de mon stage.

Je tiens tout d'abord à remercier mon maître de stage M. BRICONGNE Jean-Charles pour son assistance et sa disponibilité tout au long de ce stage. Je tiens également à remercier M. MONTORNES Jérémie, chercheur sénior au service des analyses microéconomiques de la Banque de France et M. GAUTIER Erwan, chef du service des analyses microéconomiques de la Banque de France pour leur avis métier sur les salaires et leur assistance technique dans la réalisation de l'étude. Aussi, je tiens à remercier mes collègues stagiaires pour leur accueil chaleureux et leur bienveillance tout au long de mon stage. Je remercie également M. LEBRUN Alexis grâce à qui j'ai pu m'acclimater à l'environnement informatique de la Banque de France et qui a été d'un grand secours concernant les problèmes techniques liés à l'informatique que j'ai rencontrés.

Enfin, je ne saurais finir sans remercier toute l'administration et le corps enseignant de l'ENSAI.

Résumé - Abstract

Les salaires représentent un aspect central de la décision publique. Il est important, à la fois pour les décideurs publiques et pour les simples acteurs du marché du travail, de comprendre les mécanismes qui régissent sa fixation et son évolution. Cependant, les données administratives disponibles sur les salaires peuvent être qualifiées d'ex-post c'est à dire qu'elles sont relatives à des faits antérieurs et donc elles reflètent une situation antérieure. Cela est dû au retard avec lequel elles sont postées ou mises à jour. De plus ces données ne reflètent généralement que la situation des travailleurs en poste. Pour pallier cette limite, une solution est de considérer les offres d'emploi postées en ligne. C'est donc sur cette base qu'est construite notre étude.

En nous servant de données en ligne postées sur l'année 2023 par le plus important site d'emploi en France, Pôle emploi, nous parvenons à des conclusions importantes. Nous remarquons une inégalité assez marquée des salaires. Cette inégalité semble être alimentée par certains facteurs. Dans notre analyse, il apparaît clairement que le niveau du salaire est fortement impacté par le métier exercé, la qualification, le département ou encore la catégorie socio-professionnelle. Certains domaines professionnels tels que la santé, la construction, l'art ou encore la logistique et transport semblent avoir des salaires assez élevés. Les emplois rémunèrent plus de 1.21 fois mieux les personnes hautement qualifiées que les personnes sans qualification. De même, ils rémunèrent plus 1.23 fois mieux les cadres par rapport à des manoeuvres. Aussi, la région Île-de-France et la région Corse disposent de salaires relativement élevés par rapport aux autres régions.

Wages are a central aspect of public decision-making. It is important, both for public decision-makers and for simple actors in the labor market, to understand the mechanisms that govern its setting and evolution. However, the administrative data available on wages can be described as ex-post, i.e. they relate to past events and therefore reflect a previous situation. This is due to the delay with which they are posted or updated. What's more, these data generally only reflect the situation of current workers. To overcome this limitation, one solution is to consider job offers posted online. This is the basis of our study.

Using online data posted for the year 2023 by the largest French job site, Pôle emploi, we arrive at some important conclusions. We note a fairly marked inequality in salaries. This inequality seems to be fuelled by a number of factors. In our analysis, it is clear that salary levels are strongly impacted by occupation, qualification, area and socio-professional category. Certain professional fields, such as health, construction, art or logistics and transport, seem to have fairly high salaries. Jobs pay more than 1.21 times better for highly qualified people than for those with no qualifications. Similarly, they pay executives more than 1.23 times better than labourers. Also, the Île-de-France and Corsica regions have relatively high salaries compared to other regions.

Introduction

La fixation des salaires est une question au centre des préoccupations des acteurs du marché du travail. Cela est encore plus vrai en cette période inflationniste ou éventuellement post-inflationniste. Appréhender le mécanisme de formation des salaires permettrait non seulement de comprendre les dynamiques économiques et sociales de notre société dans son état actuel, mais aussi d'aider à la décision. Les plateformes en ligne dédiées à l'emploi donnent une nouvelle dimension au marché du travail et fournissent des données riches, en temps réel sur ce marché. Elles permettent aussi, en particulier de capter des informations sur le flux entrant d'emplois.

Cette étude vise d'une part à analyser de manière détaillée l'évolution du salaire en France sur l'année 2023. Il sera alors éventuellement possible de faire des prévisions. Et d'autre part à détecter les facteurs qui influencent significativement les niveaux de salaire en France sur l'année 2023. Nous allons donc nous servir de données récentes afin de mettre en évidence les disparités géographiques, sectorielles ou encore professionnelles qui sont d'actualité sur le marché du travail français.

Afin de caractériser l'évolution des salaires en France, nous allons construire un indicateur du niveau de salaires en France et l'étudier comme une série temporelle. En ce qui concerne les déterminants du salaire, ils seront analysés à partir d'un modèle économétrique que nous expliciterons.

Toutefois, l'étape préalable et indispensable à cette étude est l'étape de récolte des données en ligne. Il nous a en effet fallu constituer la base de données en nous servant de ressources disponibles en ligne sur les plateformes dédiées à l'emploi en France. Cette base de données nous a ensuite servi à donner des réponses aux questions recherchées. Dans la suite de ce document, nous rappellerons des études semblables à la nôtre, nous présenterons ensuite les données récoltées, ainsi que le processus d'acquisition de ces données. Après quoi, nous nous attaquerons à l'évolution des salaires en France en 2023 ainsi qu'aux déterminants de ces salaires.

1 Environnement du stage

La Banque de France est la banque centrale de France. Elle a été créée le 18 janvier 1800 avant de devenir propriété de l'Etat Français le 1er janvier 1946 sous Charles De Gaulle. La Banque de France fait partie, depuis le 1er janvier 1999, du Système européen de banques centrales (SEBC), mais aussi de l'Eurosystème.

La Banque de France assure 3 missions principales :

- La stratégie monétaire
- La stabilité financière
- Les services à l'économie

Elle est constituée de plusieurs directions générales dont la Direction Générale des Statistiques, des Etudes et de l'International (DGSEI). C'est dans cette dernière que s'est déroulé ce stage. En particulier, il a eu lieu dans la Direction des Enquêtes de Conjoncture et des Analyses Micro-économiques et Structurelles (DECAMS), qui fait partie intégrante de la DGSEI.

Le maître de stage, Monsieur Jean-Charles BRICONGNE est le directeur adjoint de la DECAMS. Dans le cadre de sa fonction et du rôle de sa direction à l'intérieur de la Banque de France, il a jugé opportun de mener une étude sur le marché du travail en France. L'idée est de s'intéresser à l'évolution des salaires et à leurs déterminants sur l'année 2023.

2 La problématique et les données

2.1 Présentation de la problématique et source des données

Depuis l'année dernière, la France subit une crise inflationniste qui touche tous les secteurs d'activité. Les dépenses de consommation usuelles des ménages ont brutalement augmenté. Pour faire face à ces dépenses, les salaires ont très souvent été négociés à la hausse. Entre temps, les employeurs, quant à eux, ont au contraire intérêt à réduire les salaires.

L'objectif de notre étude est justement de comprendre ce qui se passe effectivement sur le marché du travail en France dans cette période. Il nous faut dans un premier temps mesurer l'évolution des salaires au cours des derniers mois, ensuite nous caractériserons le mécanisme de fixation de ces salaires.

Dans le cadre de notre étude, nous ne nous servons que des annonces d'offres d'emploi en ligne qui constituent de nos jours une source de données inestimable sur le marché de l'emploi. En effet, bien que des sources de données officielles existent sur les salaires, ces dernières ont des limites, elles sont souvent difficilement accessibles et lorsqu'elles le sont, il s'agit typiquement de données à basse fréquence et relatives à des périodes relativement lointaines. Les données que nous utiliserons pour notre étude ne présentent pas ces limites.

Vu que notre analyse ne concerne que le marché de l'emploi en France, nos données proviennent du site majeur de recherche d'emploi en France : Pôle emploi. Ce choix s'explique par le fait que Pôle emploi couvre à lui seul environ 30% des offres d'emploi publiées par les sites majeurs d'offres d'emploi en France, en ne considérant que les sites consacrés à la France.

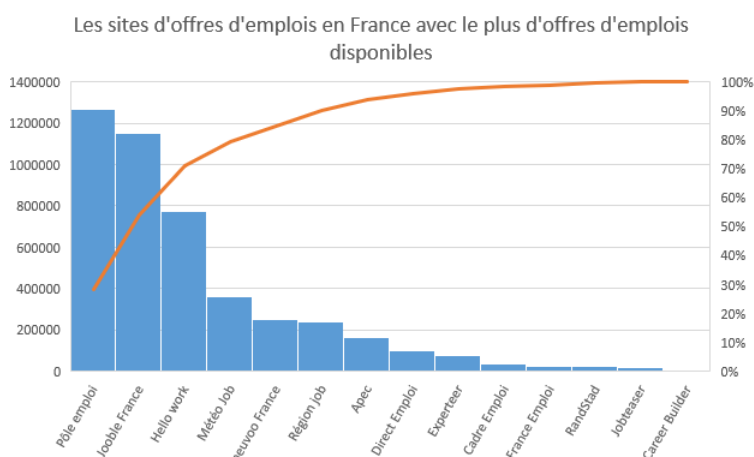


FIGURE 1 – Nombre d'offres d'emploi disponibles par site d'emploi

Notre base de données est donc entièrement tirée de Pôle emploi.

2.2 Revue de littérature

Dans cette étude, notre premier objectif est de collecter des données sur les offres d'emploi en France disponibles en ligne. Nous analyserons ensuite ces données et mettrons en évidence l'évolution des salaires ainsi que les facteurs qui déterminent ces salaires.

L'apparition des offres d'emploi en ligne est récente. Les études basées sur celles-ci sont donc relativement récentes. Le but de cette revue de littérature est de repérer des pistes concernant la méthodologie à adopter dans notre étude, le traitement de nos données ou encore les résultats auxquels nous pourrions nous attendre.

Tout d'abord, pour notre étude, nous avons l'intention de collecter des données sur les offres d'emploi présentes en ligne. La méthode de collecte de données que nous utilisons est le requêtage de l'API¹ de Pôle Emploi. Dans ce contexte, nous avons trouvé dans la littérature plusieurs études similaires à la nôtre.

En France, De Maricourt et Niang (2022) ont réalisé une étude pour la Direction de l'Animation de la Recherche, des Études et des Statistiques (DARES). La DARES a réussi à mettre en place une nouvelle base de données sur les offres d'emploi en France. Cette base est constituée uniquement d'offres d'emploi publiées en ligne sur 14 sites web sélectionnés. La base de données s'appelle Jocas (job offers collection and analysis system). La DARES a collecté les données en utilisant différentes méthodes : partenariat avec le site web, web-scraping et utilisation de l'API du site web. Les auteurs ont ensuite traité les données collectées. Ils ont harmonisé les intitulés de postes, ceux-ci pouvant différer d'un site à l'autre pour un même poste. Ils ont également déduplicé les données. Cependant, une des limites de la base de données Jocas est sa couverture inéquitable des différentes professions, ce qui pose un problème de représentativité. En effet, ils estiment que 66% des recrutements dans l'informatique et les télécommunications font l'objet d'une annonce en ligne. En revanche, seuls 29% des recrutements dans le secteur de la construction sont associés à une annonce en ligne.

Dans d'autres pays européens, Faryna et al. (2021) examinent la relation entre le marché du travail et la dynamique des salaires en Ukraine. Les auteurs utilisent une base de données de plus d'un million d'offres d'emploi en ligne. Cette base de données provient d'OLX.ua, l'une des principales plateformes d'annonces en ligne d'Ukraine. L'ensemble des données contient des informations sur le lieu d'emploi, la description, le salaire et le type d'emploi. Les chercheurs contrôlent l'hétérogénéité régionale, sectorielle, professionnelle et celle des segments de compétences lorsqu'ils étudient le lien entre la fixation des salaires et le chômage. À l'aide d'une régression linéaire (MCO), l'étude constate un faible arbitrage entre l'inflation salariale globale et le chômage.

Ziegler (2020) analyse les compétences requises mentionnées dans les offres d'emploi sur le principal site web autrichien consacré à l'emploi, et leur relation avec les offres salariales. L'auteur utilise donc des données provenant de plus de 380 000 offres d'emploi publiées sur le principal site web autrichien consacré à l'emploi. Il utilise également les offres d'emploi du plus grand concurrent sur le marché autrichien de l'emploi pour tester la robustesse des estimations. L'étude révèle un lien solide entre le nombre de compétences requises et les salaires proposés dans les offres d'emploi. Les offres d'emploi comportant de nombreuses compétences requises offrent des salaires nettement plus élevés.

En Amérique latine, Campos-Vázquez et al. (2021) ont effectué un balayage quotidien du web, de janvier à décembre 2020, sur un grand site d'offres d'emploi mexicain. Ils ont ainsi constitué la base de données nécessaire pour étudier l'influence de la pandémie de COVID-19 sur la demande de main-d'œuvre au Mexique. Les auteurs ont également utilisé l'analyse de texte pour rechercher les compétences et les caractéristiques personnelles recherchées par les employeurs. Ils constatent qu'il n'y a pas eu de changement fondamen-

1. Application Programming Interface : C'est une interface qui permet à des programmes informatiques de communiquer entre eux et éventuellement de s'échanger des données.

tal dans la nature des offres d'emploi au Mexique. Toutefois, le nombre d'offres d'emploi disponibles (en ligne) a diminué. En outre, ils n'ont trouvé aucune preuve permettant de soutenir l'idée que la crise du COVID-19 avait entraîné un changement permanent et significatif de la nature du travail au Mexique. Bien qu'ils constatent que la structure de la demande de main-d'œuvre a été modifiée, ce changement n'a été que temporaire.

Par ailleurs, nous avons trouvé des études qui analysent le marché du travail à partir de données d'offres d'emploi en ligne à disposition sans nécessairement passer par l'étape fastidieuse de récolte de données à travers le web-scraping ou le requêtage d'API.

Adrian et Lydon (2022) utilisent les données temporelles d'Indeed pour construire un indicateur mensuel de la croissance des salaires dans 7 pays : France, Allemagne, Irlande, Pays-Bas, Italie, Espagne et Royaume-Uni. Indeed publie des données tous les mois, ce qui donne plusieurs millions de données d'offres d'emploi publiées entre le 1er janvier 2018 et octobre 2022. Les données collectées concernent le type de salaire (horaire, mensuel ou annuel), la date à laquelle l'annonce a été publiée pour la première fois, la localisation de l'emploi et l'intitulé du poste. Les auteurs mènent leur étude en considérant des cellules regroupant les salaires correspondant à un même emploi, un même département et un même type de salaire. Pour chaque mois et chaque cellule, la croissance des salaires est calculée par rapport au niveau de salaire du même mois mais de l'année précédente. En utilisant cette méthodologie, ils constatent que la croissance des salaires dans les différents pays étudiés s'est fortement accélérée à la fin de la pandémie de COVID-19.

Shuai et al. (2020) utilisent des données à haute fréquence sur les offres d'emploi pour évaluer l'effet du COVID-19 sur les offres d'emploi aux États-Unis. Ils utilisent la base de données RTI (Real-Time-intelligence), qui est mise à jour quotidiennement. Elle recueille des offres d'emploi en ligne provenant de 15 000 sources. Les offres d'emploi sont classées et dédoublées. Ils ont examiné le nombre d'offres d'emploi et les salaires publiés. Les auteurs ont constaté que le COVID-19 a provoqué une baisse significative de la demande de travail, de l'ordre de 30%, la demande de travail étant mesurée par le nombre d'offres d'emploi. Toutefois, en analysant l'évolution des salaires publiés, ils constatent que le COVID-19 n'a pas eu d'impact négatif sur les salaires.

2.3 Processus d'acquisition des données

La difficulté majeure de cette étude a été l'acquisition des données et leur traitement. Comme annoncé, nous nous sommes servis de la base de données de Pôle emploi. Afin d'acquérir nos données nous avons utilisé l'API de Pôle Emploi. Grâce à cette API, nous pouvons récupérer toutes les offres d'emploi publiées dans un intervalle de temps à spécifier et présentes sur le site de Pôle emploi au moment du requêtage. Il faut donc que l'offre n'ait pas encore été pourvue au moment du requêtage pour pouvoir la récupérer.

L'étape de récolte a été très fastidieuse car l'API de Pôle emploi ne nous permet de récolter que 150 offres d'emploi par requête. Il nous a donc fallu être astucieux afin de récolter toutes les données désirées.

Notre base de données est construite en requêtant cette API chaque semaine du 21 juillet au 18 août 2023. À chaque fois, nous récupérons toutes les offres d'emploi présentes sur le site de Pôle emploi au moment du requêtage et postées entre le 1^{er} janvier 2023 et la date courante. Ainsi, l'offre d'emploi est récupérée seulement si elle n'a pas encore été pourvue au moment du requêtage de l'API. Notre base de données a été construite selon ce processus pour pouvoir également mesurer le flux sortant d'emplois. Toutefois ce sujet

n'est pas l'objet de la présente étude.

Nous procédons ensuite à une déduplication des offres d'emploi, en d'autres termes, nous supprimons les occurrences en trop des offres qui apparaissent plus d'une fois. Pour ce faire, nous nous basons sur l'identifiant unique accordé à chaque offre d'emploi en ligne par Pôle emploi.

En procédant ainsi, nous obtenons finalement une base de données de 1 481 393 offres d'emploi publiées entre le 1^{er} janvier 2023 et le 18 août 2023.

2.4 Description et structure des données

2.4.1 Description des données recoltées

Les données sont constituées de 1 481 393 lignes correspondant chacune à une offre d'emploi. Elles correspondent à une quarantaine de colonnes. Toutefois, celles qui nous intéressent sont les suivantes :

- **l'identifiant** : unique à chaque offre ;
- **l'intitulé du poste** : tel que donné par l'entreprise qui emploie ;
- **la description du poste** ;
- **la date de création du poste** ;
- **le lieu de travail** ;
- **le code ROME** : le ROME est un référentiel conçu par Pôle emploi et actualisé régulièrement, tenant compte des évolutions du marché du travail. Il présente l'ensemble des métiers regroupés par fiches, organisées par domaines professionnels.² Il est composé de 5 caractères : une lettre et quatre chiffres ;
- **le libellé ROME** : chaque fiche métier est identifiée par un code ROME et par ce libellé ;
- **l'appellation du poste** : il s'agit de l'intitulé du poste standardisé par pôle emploi. C'est cet intitulé que nous utiliserons pour la suite ;
- **le type de contrat** : il peut s'agir d'un contrat à durée déterminée (CDD), d'un contrat à durée indéterminée (CDI), d'un contrat d'intérimaire, d'un contrat libéral, etc. ;
- **l'expérience exigée** : cette colonne indique si pour l'offre considérée, aucune expérience n'est demandée, si une première expérience est souhaitée ou si une première expérience est exigée ;
- **le salaire proposé** : lorsqu'il est présent, il est sous forme d'un intervalle de salaire ou d'un salaire exact ;
- **les avantages extra-salariaux** ;
- **la qualification** : celle ci correspond au niveau d'études requis ;
- **les formations demandées par l'employeur** ;
- **la catégorie socio-professionnelle** : par exemple, cadre, technicien, employé qualifié, employé non qualifié, etc ;
- **les compétences demandées par l'employeur** ;

2. définition de Pôle emploi

- **une indicatrice indiquant si l'emploi proposé est une alternance ;**
- **une variable indiquant si l'emploi est à temps plein ou à temps partiel ;**
- **le code NAF :** nomenclature d'Activité Française. C'est un code à 4 chiffres et une lettre délivré par l'Insee. À chaque activité professionnelle est attribué un code NAF. Ce code correspond à une classification fine des secteurs d'activité ;
- **le secteur d'activité :** basé sur le code NAF. C'est une classification moins fine des secteurs d'activité.

2.4.2 Traitement des données

Nous avons commencé par procéder à une déduplication de nos données. Ainsi, une seule ligne a été gardée lorsque nous avons rencontré des lignes avec le même identifiant.

Ensuite, sur la base des colonnes présentées précédemment, nous avons créé de nouvelles colonnes qui sont plus adaptées à notre étude ou qui faciliteront nos analyses futures. Dans ce sens, les colonnes suivantes apparaissent :

- **le mois de création de l'offre ;**
- **le département de travail ;**
- **une variable indiquant le groupe de métiers auquel l'offre appartient :** il s'agit d'une forme agrégée de l'intitulé du poste basée sur le code ROME ;
- **une variable indiquant le domaine professionnel auquel l'offre appartient :** il s'agit d'une forme agrégée de la variable précédente ;
- **le nombre de formations demandées par l'employeur**
- **une nouvelle colonne pour les avantages extra-salariaux :** cette colonne a été délicate à créer. Les données initiales fournies par Pôle emploi contiennent des avantages extra-salariaux. Nous avons toutefois eu l'idée de rechercher, à partir d'une analyse de texte, les avantages extra-salariaux contenus dans le descriptif du poste. Nous avons ensuite comparé les avantages extra-salariaux contenus dans le descriptif de l'emploi aux avantages extra-salariaux contenus dans la colonne correspondante. Nous nous sommes aperçus qu'il n'y avait pas toujours concordance, il y avait donc de l'information en plus à récupérer dans le descriptif de l'emploi. La nouvelle colonne pour les avantages extra-salariaux combine donc les avantages extra-salariaux fournis avec ceux contenus dans le descriptif de l'emploi ;
- **une indicatrice indiquant si l'offre accorde des avantages extra-salariaux.**

L'objet de notre étude étant principalement les salaires, il a fallu effectuer certains traitements sur les salaires récupérés. Nous remarquons d'abord que les salaires fournis se présentent sous deux formes : sous la forme d'un intervalle ou sur la forme d'un salaire exact. Dans le premier cas, nous considérons la valeur au milieu de l'intervalle et dans le second cas, nous récupérons juste la valeur fournie. Nous avons ensuite remarqué que le salaire pouvait être de plusieurs types : horaire, mensuel ou annuel. Dans une optique de comparaison, nous avons ramené tous les salaires sur une base annuelle en faisant une simple interpolation. Par exemple, un salaire mensuel, sans 13^e mois sera multiplié par 12. Dans le cas d'un salaire horaire, nous faisons l'hypothèse que 7 heures sont travaillées par jour et considérons 250 jours ouvrés dans l'année. Avec ces considérations, pour un salaire horaire, en absence de 13^e mois, le salaire sera multiplié par 7 puis par 250.

Finalement, de nouvelles colonnes apparaissent concernant le salaire :

- **le salaire minimum** : lorsque le salaire est sous forme d'un intervalle, il s'agit de la borne inférieure et lorsque le salaire est sous forme d'un montant exact, il s'agit de ce montant ;
- **le salaire maximum** : lorsque le salaire est sous forme d'un intervalle, il s'agit de la borne supérieure et lorsque le salaire est sous forme d'un montant exact, il s'agit de ce montant ;
- **le salaire moyen** : la moyenne arithmétique du salaire minimum et du salaire maximum ;
- **le type de salaire** : horaire, mensuel ou annuel ;
- **la période sur laquelle le salaire est pourvu** : cette colonne indique la présence d'avantages comme le 13^e ou le 14^e mois de paie par exemple ;
- **le salaire corrigé** : nous avons ramené le salaire moyen sur une base annuelle en tenant compte d'un éventuel 13^e/14^e mois. Plusieurs contrôles ont été faits sur cette variable afin d'éviter les erreurs dans les données ou les valeurs aberrantes. Elle ne considère que les salaires contenus dans une certaine plage de salaire. la plage considérée est entre 5500€ et 400000€ par an. C'est cette colonne que nous considérerons pour l'étude des salaires ;
- **les tranches de salaire** : cette colonne indique la tranche de salaire correspondant à l'offre.

2.5 Statistiques descriptives

Les données que nous utilisons se composent de 1 481 393 offres d'emploi publiées entre le 1^{er} janvier 2023 et le 18 août 2023. Ces offres d'emploi couvrent les 96 départements français et 86 secteurs différents sur les 88 classés par l'Insee avec le code NAF. Elles concernent aussi 8916 emplois différents.

Ces informations soulèvent plusieurs questions concernant la structure de nos données ainsi que leur représentativité vis-à-vis du marché de l'emploi français.

2.5.1 Structure des données

Notre méthode de récolte de données nous conduit à avoir un nombre très différent de données d'un mois à l'autre comme nous pouvons le voir sur le graphique.

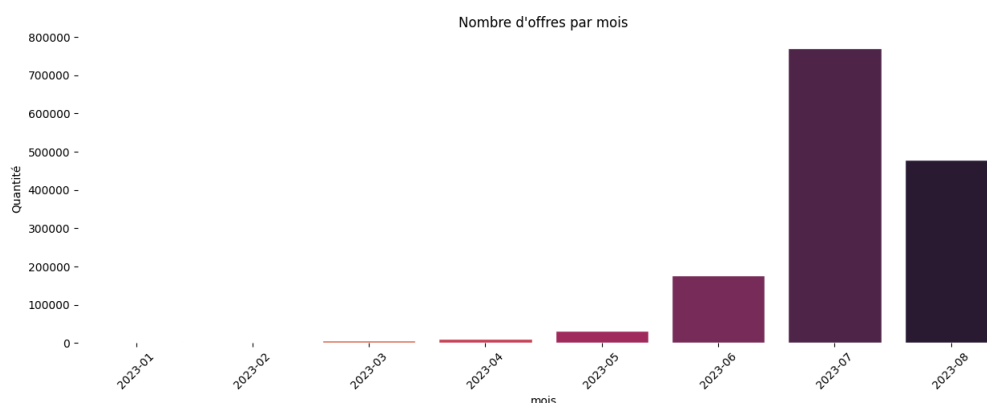


FIGURE 2 – Nombre d'offres d'emploi par mois

En effet, nous ne récupérons une offre que si elle est encore présente sur le site de Pôle emploi au moment du requêtage. De ce fait, sur l'ensemble de nos 5 semaines de requêtage, nous avons un nombre peu important d'offres d'emploi publiées en janvier ou en février. De plus du fait que la dernière date que nous requêtons est le 18 août, nous avons moins d'offres d'emploi datant d'août comparativement au mois de juillet.

Nous nous posons ensuite la question de la représentativité départementale de nos données. Nous illustrons donc cette représentativité départementale

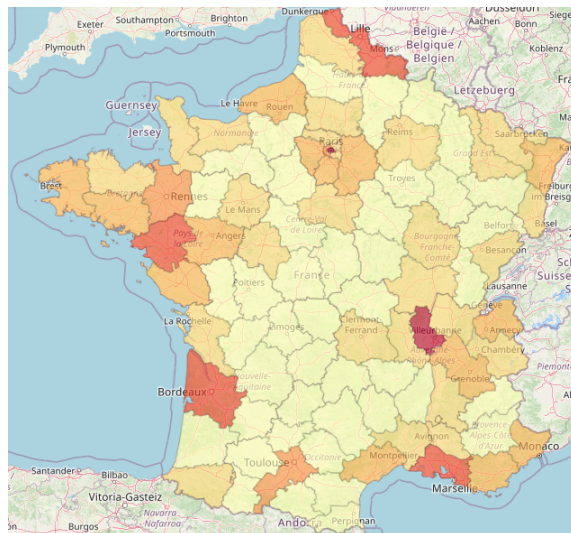


FIGURE 3 – Nombre d'offres d'emploi par département

Dans ce graphique, l'intensité de la couleur illustre l'importance du nombre d'offres d'emploi dans le département. On se rend alors compte que le département le plus représenté est le département Rhône (69) suivi par le département Paris (75). Toutefois, cet ordre change lorsqu'on normalise ce nombre d'offres d'emploi par la population active.

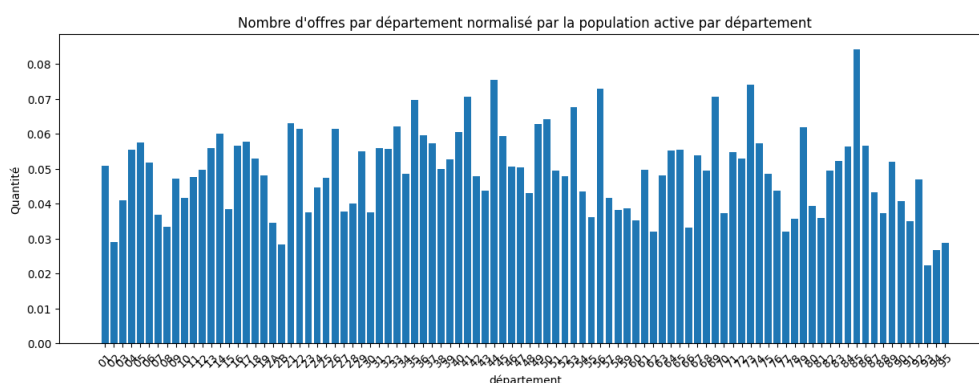


FIGURE 4 – Nombre d'offres d'emploi par département normalisé par la population active

Cette fois, les département 85 (Vendée) et 44 (Pays de la Loire) sont les mieux représentés. De plus, les écarts entre départements sont beaucoup moins importants.

On remarque également que les tranches de salaire ne sont pas représentées de manière équivalente dans nos données.

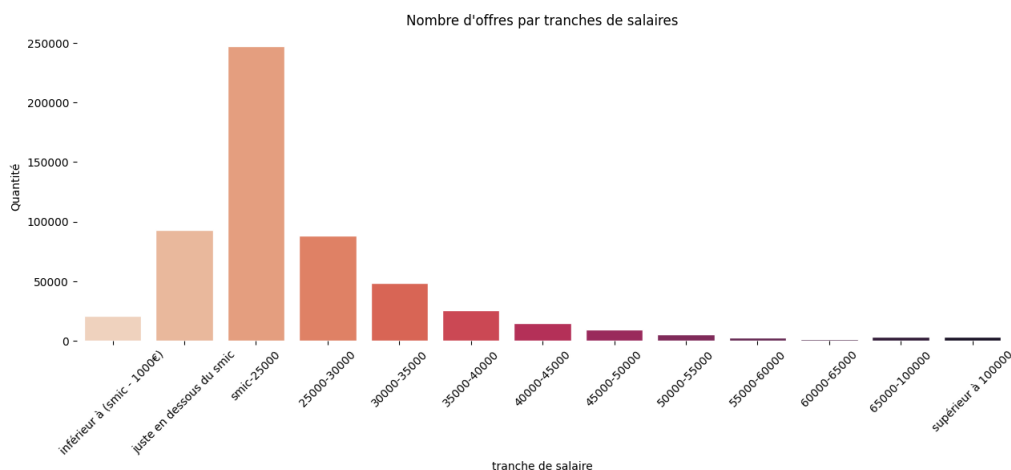


FIGURE 5 – Nombre d’offres d’emploi par tranches de salaire

On remarque dans un premier temps qu’un nombre important d’offres d’emploi concernent des salaires en dessous du smic. Cela s’explique en partie par le fait que nos données concernent également des alternances, sans concerner des stages. La proportion d’alternance dans nos données est de **0.036**, ce qui représente **53 359** offres d’emploi. Et, comme nous le verrons par la suite, les alternances sont en moyenne rémunérées en dessous du smic. Cela explique en partie le nombre d’offres d’emploi en dessous du smic. D’autre part, le Smic est la rémunération minimale en tenant compte non seulement de la rémunération affichée mais aussi des avantages extra-salariaux telles que les repas par exemple. Cela explique aussi le nombre d’offres à moins de 1000€ en dessous du smic.

La distribution du nombre d’offres par tranches de salaires est peut être également liée à la catégorie socio-professionnelle ou à la qualification demandée par les employeurs.

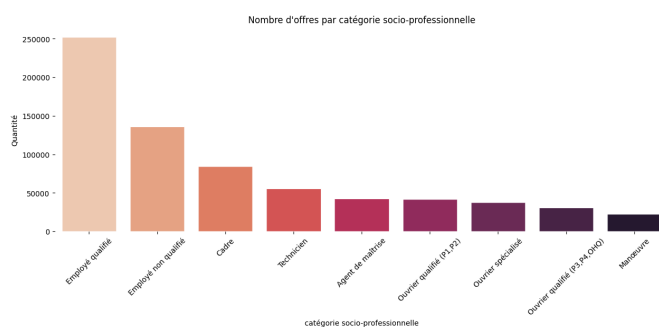


FIGURE 6 – Nombre d’offres par catégorie socio-professionnelle

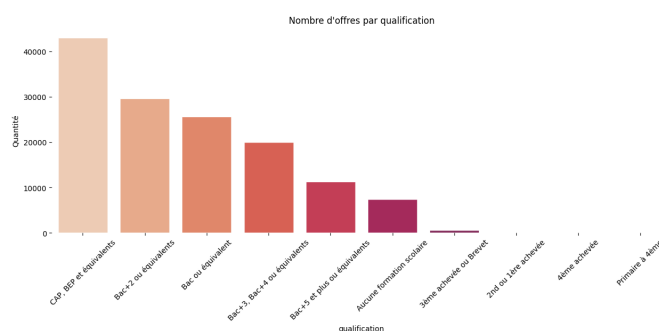


FIGURE 7 – Nombre d’offres par qualification

Nous remarquons que nos données contiennent majoritairement des offres correspondant à des niveaux d’étude en dessous du BAC ou équivalents au BAC. Ces offres correspondent généralement à des salaires dans les tranches inférieures. Aussi, nous constatons que les offres ne sont pas équitablement réparties entre les catégories socio-professionnelles, les employés étant les plus représentés.

Enfin, nous nous intéressons à la représentativité de nos données selon les groupes de métiers à travers le graphique suivant.

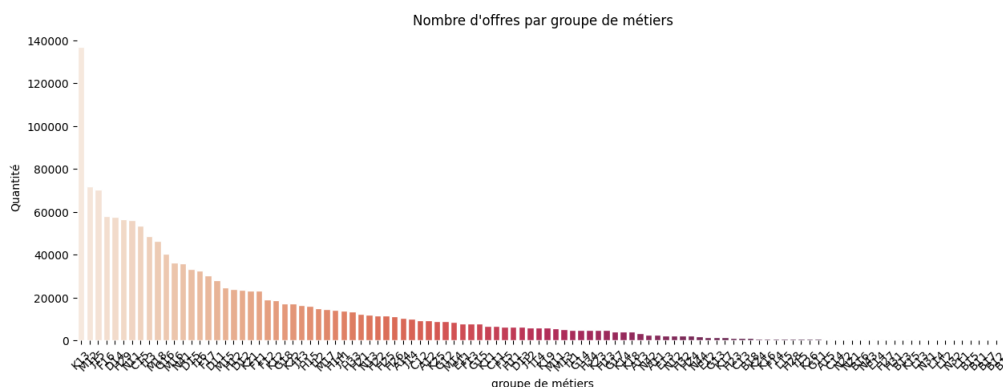


FIGURE 8 – Nombre d'offres d'emploi par groupe de métiers

Nous remarquons qu'un groupe de métiers est sur-représenté par rapport aux autres. Il s'agit du groupe K13 (Aide à la vie quotidienne) regroupant des métiers d'accompagnement médicosocial, auprès d'adultes, d'enfants, etc. Ce groupe regroupe près du dixième des offres d'emploi de notre base de données. Ensuite viennent les groupes M12 (comptabilité et gestion) et J15 (soins paramédicaux). De l'autre côté, certains groupes de métiers sont beaucoup moins représentés. Il s'agit notamment des métiers du domaine professionnel B (Arts et façonnage d'ouvrages d'art).

2.5.2 Qualité des données

On se pose maintenant la question de la qualité de nos données. Pour répondre à cette préoccupation, nous nous servons de statistiques officielles de l'INSEE sur le marché du travail français.

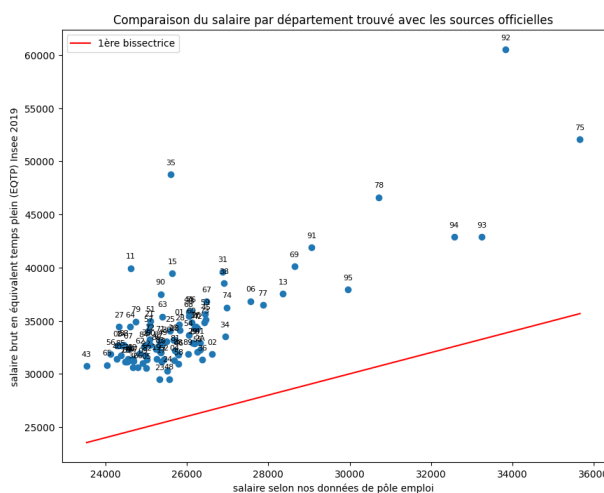


FIGURE 9 – Comparaison avec les données de l'Insee

Nous remarquons que les moyennes par département que nous produisons sont toujours inférieures aux données en équivalent temps plein de l'Insee. Cet écart est en moyenne de **8 434.64 €**. Cet écart suggère que nos données concernent des emplois moins bien rémunérés. En effet, comme nous l'avons vu à la figure 4, les tranches de salaires les plus présentes dans nos données sont les tranches inférieures. D'autre part, le salaire en

équivalent temps plein est un salaire converti à un temps plein pendant toute l'année, quelque soit le volume de travail effectif. Cette méthode de calcul n'est pas compatible avec nos données de salaire, elle aura tendance à surévaluer le niveau effectif de salaire annuel.

Une autre comparaison que nous faisons, consiste à regarder la distribution des salaires fournie par l'Insee. Cette dernière se présente comme suit :

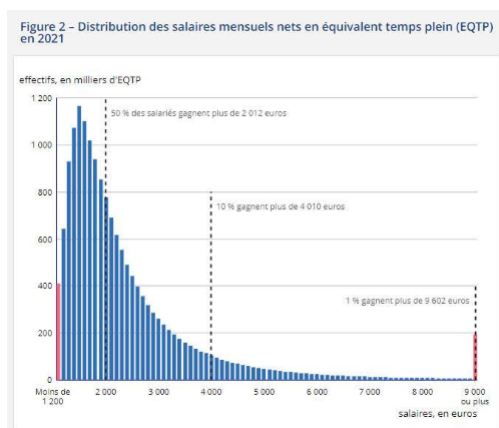


FIGURE 10 – Distribution des salaires mensuels selon l'Insee. Source : Insee, Base Tous salariés 2021

Nous notons une ressemblance avec le graphique du nombre d'offres par tranches de salaires présenté à la figure 5. En effet, nous remarquons dans un premier temps, un nombre assez important de salaires en dessous du smic. De plus les deux graphiques ont la même forme, nous constatons une croissance de l'histogramme puis un pic autour du smic puis une décroissance exponentielle. En ce sens, nos données sont assez similaires aux statistiques officielles. D'autres éléments de comparaison sont présentés en annexes.

2.5.3 Quelques statistiques générales sur les salaires

Comme annoncé, nous nous servons de la colonne des salaires ramenés sur base annuelle et contrôlés. La moyenne des salaires de notre base de données est **26 719.36€** et la médiane est de **22 800€**, avec un salaire maximum de **399 000€** et un salaire minimum de **5 783.75€**. Nous soupçonnons très rapidement une forte dispersion dans les niveaux de salaire. Nous représentons la distribution des salaires.

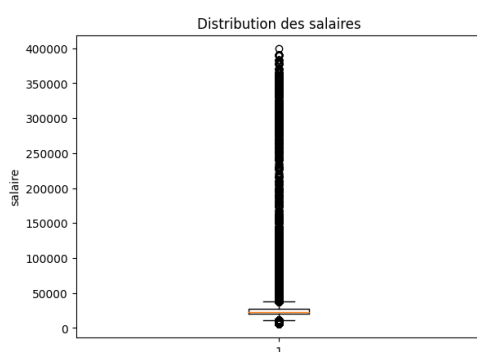


FIGURE 11 – Boxplot de la distribution des salaires

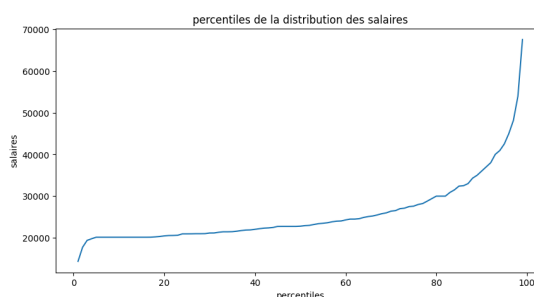


FIGURE 12 – Percentiles de la distribution des salaires

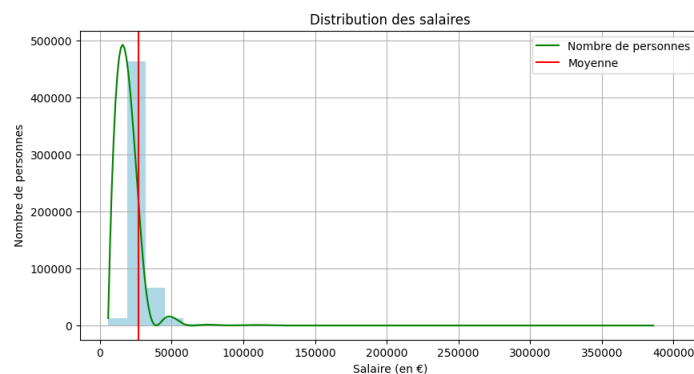


FIGURE 13 – Distribution des salaires

A travers ces graphiques, le soupçon d'une forte dispersion des salaires se confirme. Nous avons donc majoritairement des offres d'emploi avec des bas salaires, avec une médiane très proche du premier décile. Toutefois, il reste certaines offres avec des salaires très haut, plus de 10 fois supérieurs à la médiane. Il en existe exactement **1303** dans nos données soit moins de **0.1%** des données.

2.5.4 Une hétérogénéité dans les niveaux de salaire

Nous remarquons maintenant que les salaires dépendent de certaines caractéristiques observées dans nos données. Cette dépendance sera analysée plus en profondeur dans la suite de notre étude. Pour donner l'intuition derrière ces analyses, nous présentons dans ce paragraphe, les statistiques descriptives associées.

- Une hétérogénéité départementale

Tout d'abord, on anticipe une hétérogénéité départementale, comme le laisse présager la figure 9. Cette hétérogénéité se confirme avec le graphique suivant :

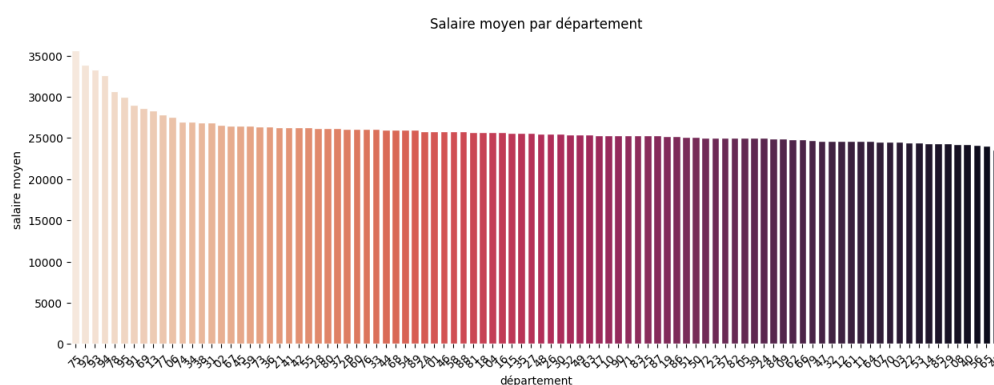


FIGURE 14 – Salaire moyen par département

Nous pouvons remarquer que les départements avec les plus hauts salaires moyens sont des départements d'Île-de-France, à savoir le 75 (Paris), le 92 (Hauts-de-Seine), le 93 (Seine-Saint-Denis), le 94 (Val-de-Marne), le 78 (Yvelines), le 95 (Val-d'Oise) et le 91 (Essonne). Ces départements ont un salaire moyen supérieur d'au moins 5000€ à celui du département avec le plus faible salaire moyen (Haute-Loire). Ces résultats sont assez intuitifs.

De plus, nous remarquons une homogénéité assez marquée des niveaux de salaires dans les autres départements.

- **Une hétérogénéité sectorielle**

Cette hétérogénéité peut être illustrée par le graphique suivant. Les secteurs sont constitués avec la classification sectorielle NAF.

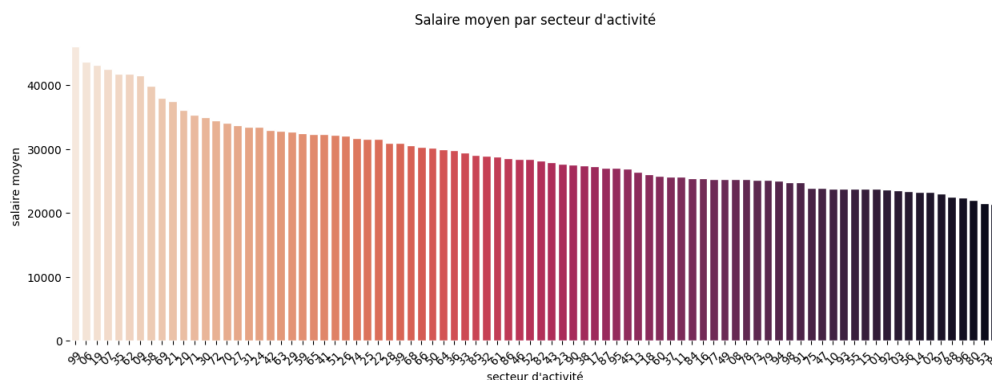


FIGURE 15 – Salaire moyen par secteur d'activité

Il existe des groupes de secteurs avec des niveaux de salaires plus ou moins similaires. Le secteur 99 (Activités des organisations et organismes extraterritoriaux) est le secteur le mieux payé. Il est suivi par les secteurs 06 (extraction d'hydrocarbures) et 19 (cockéfaction et raffinage).

- **Une hétérogénéité professionnelle**

Nous remarquons pour finir des différences dans les niveaux de salaires en fonction de certains caractéristiques qu'on qualifie de "professionnelles".

Dans un premiers temps, les salaires proposés sont différents selon la catégorie socio-professionnelle ou encore la qualification requise.

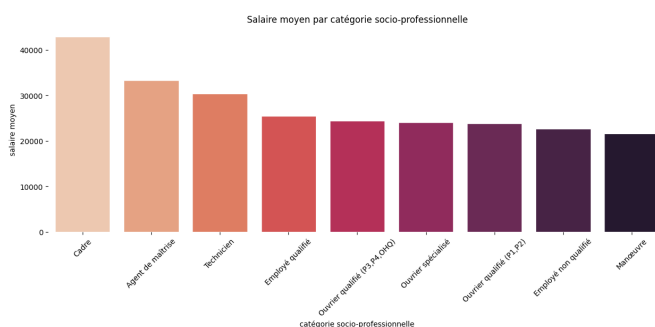


FIGURE 16 – salaire moyen par catégorie socio-professionnelle

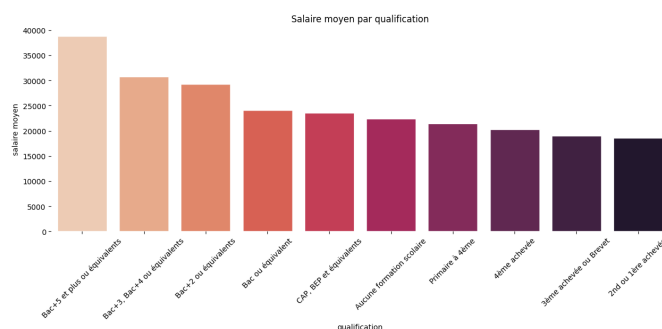


FIGURE 17 – Salaire moyen par qualification

De plus, le salaire diffère selon que l'emploi proposé soit à temps plein ou à temps partiel ou selon qu'il s'agisse d'une alternance ou pas.

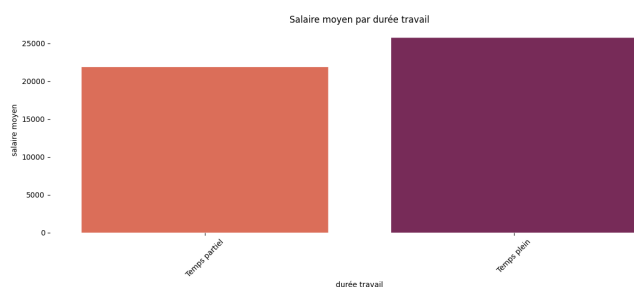


FIGURE 18 – Salaire moyen par durée du travail

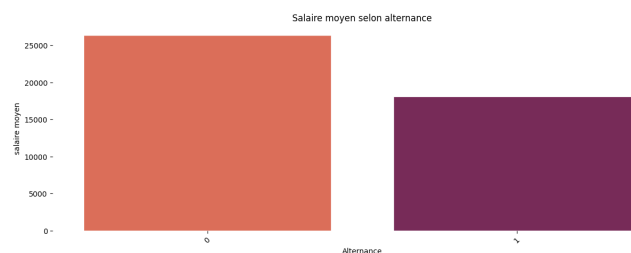


FIGURE 19 – Salaire moyen selon alternance

Ces résultats sont assez intuitifs. Un travailleur à temps plein gagne en moyenne à l'année environ 5 000 € de plus qu'un travailleur à temps partiel. Aussi, un alternant gagne à l'année environ 9 000 € de moins qu'un non alternant. On remarque en particulier que le salaire moyen pour un alternant est de **18 067.23 €**, ce qui est inférieur au smic. Cela explique en partie, le nombre élevé de travailleurs en dessous du smic dans nos données.

Enfin, nous nous sommes intéressés à l'expérience demandée par les employeurs. À un niveau très agrégée, nous avons trois modalités dans nos données :

- **D** : Débutant accepté. Dans ce cas, aucune expérience n'est demandé par l'employeur à l'employé ;
- **S** : Expérience souhaitée. Dans ce cas, l'employeur favorise les candidats avec une expérience ultérieure ;
- **E** : Expérience exigée. Dans ce cas, l'employeur estime qu'une expérience ultérieure du candidat est requise.

Nous remarquons que les niveaux de salaire sont différents selon cette caractéristique.

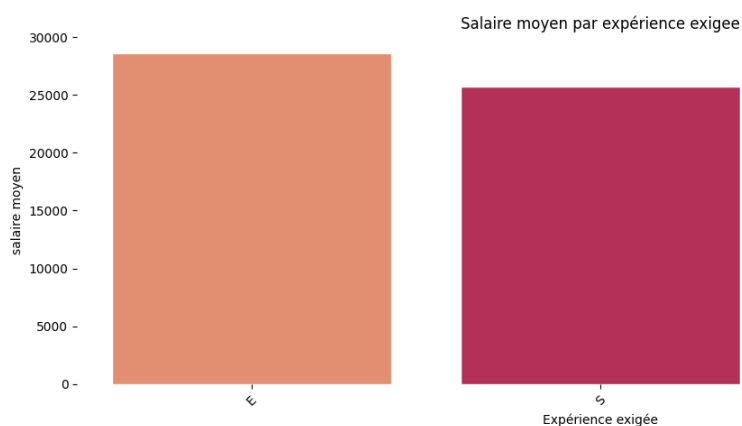


FIGURE 20 – Salaire moyen par expérience demandée

Comme nous nous y attendions, les salaires sont plus élevés lorsque l'employeur exige une expérience ultérieure à l'employé.

3 L'évolution des salaires en France en 2023

Nous nous intéressons dans ce paragraphe à l'évolution journalière du niveau des salaires en France sur l'année 2023.

Il nous a d'abord fallu construire un indicateur journalier du niveau de salaire. Pour ce faire, notre méthode a été inspirée de la méthodologie de Adrjan et Lydon (2022). Sur la base de cette étude, nous avons construit deux indicateurs du niveau de salaire sur l'ensemble du territoire français à une date donnée. L'un des indicateurs se sert de la moyenne, tandis que l'autre se sert de la médiane. Notre méthode se décline en 3 étapes :

- Pour chaque date, regrouper les salaires par intitulé de poste, type de contrat et département ;
- Pour chaque groupe généré, calculer le salaire moyen (respectivement médian) ;
- Nous obtenons une distribution de salaires pour le jour considéré. Nous prenons la moyenne (respectivement la médiane) de cette distribution. Nous obtenons finalement l'indicateur du niveau de salaire pour le jour considéré

Une représentation graphique nous permet de visualiser l'évolution dans le temps des indicateurs construits selon nos données.

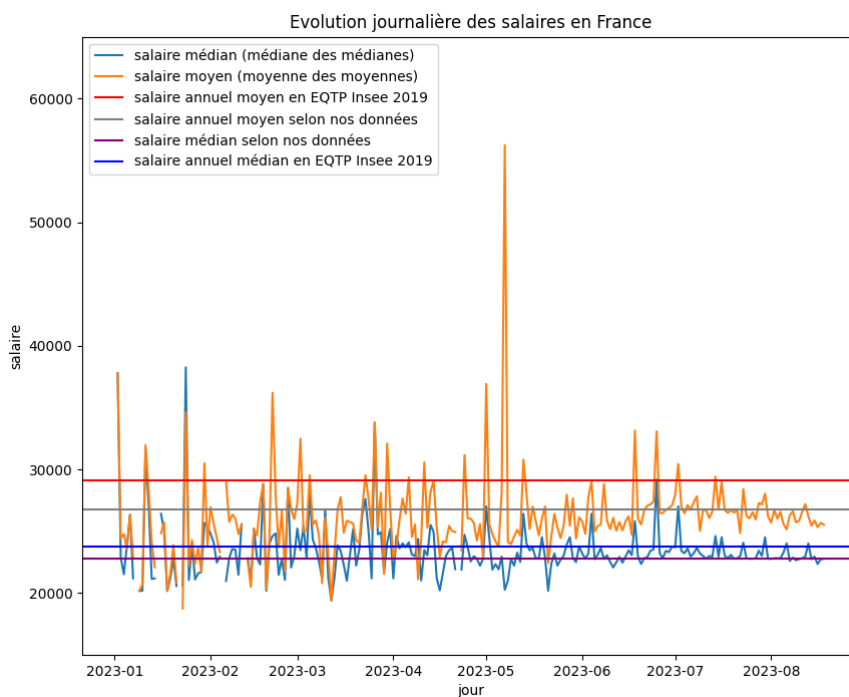


FIGURE 21 – Evolution journalière salaire

3.1 Représentation graphique

Le niveau de salaire est représenté par deux indicateurs différents : **la moyenne des salaires** et **la médiane des salaires**. Nous allons étudier ces deux séries temporelles.

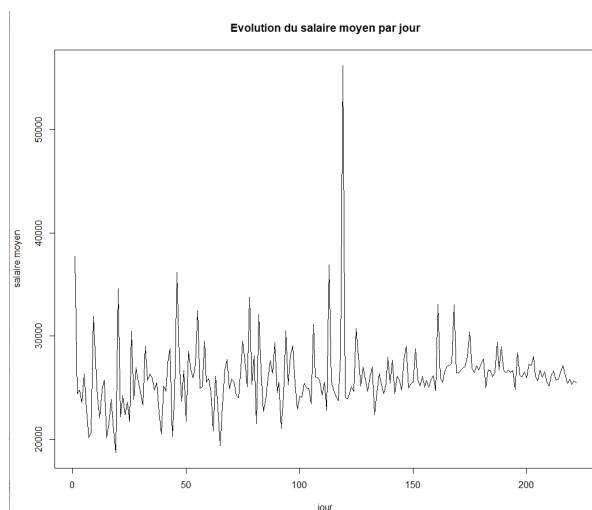


FIGURE 22 – Evolution journalière du salaire moyen

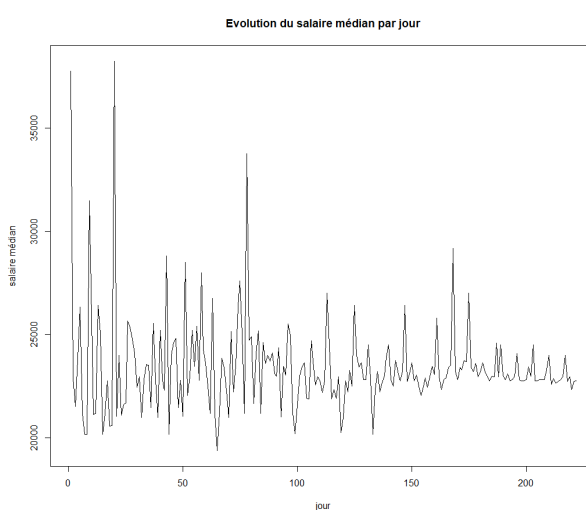


FIGURE 23 – Evolution journalière du salaire médian

Ces graphiques ne nous rassurent pas en ce qui concerne la stationnarité de nos deux séries. Nous allons donc nous pencher davantage sur cette stationnarité.

3.2 Stationnarité des séries temporelles

Nous suspectons une non stationnarité de nos séries temporelles. En particulier, nous suspectons l'existence d'une tendance temporelle. De ce fait, nous testons tout d'abord le cas le plus simple : la présence d'une tendance linéaire.

Nous supposons donc que la tendance est de la forme : $f(t) = a + b * t$ avec t le temps.

Afin de tester sa présence dans nos données, nous régressons la série temporelle sur notre variable temporelle. Le modèle de régression linéaire que nous utilisons est :

$Y_t = \beta_0 + \beta_1 * t$ Avec Y_t qui représentera alors la valeur du niveau de salaire (moyenne ou médiane) à l'instant t .

La significativité du coefficient β_1 nous indiquera alors la présence de la tendance suspecte. Les résultats de la régression sont détaillés en annexes. Nous présentons ici un extrait.

Série	β_0	β_1	modèle complet
Moyennes	$< 2 * 10^{-16}$	0.0215	0.02149
Médianes	$< 2 * 10^{-16}$	0.0955	0.09549

TABLE 1 – P-valeurs obtenus à travers les régressions sur la variable temporelle

Nous constatons que β_1 et le modèle sont significatifs au seuil de 5% en ce qui concerne la régression de la série des moyennes. Cela implique que nous pouvons à raison supposer l'existence d'une tendance linéaire dans cette série et donc une non stationnarité de la série journalière du salaire moyen.

En ce qui concerne la série journalière du salaire médian, la régression nous indique que β_1 et le modèle sont significatifs seulement au seuil de 10%. Nous allons donc considérer, avec moins d'évidences, que cette série est également non stationnaire.

Nous différencions donc nos séries temporelles afin de les stationnariser. les séries différenciées se présentent graphiquement comme suit :

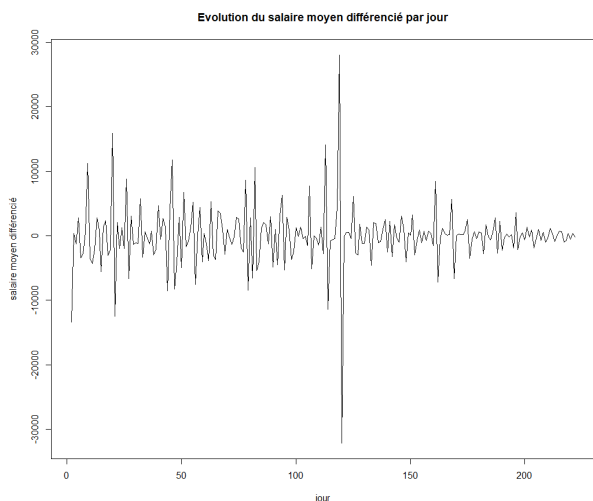


FIGURE 24 – Evolution journalière du salaire moyen différencié

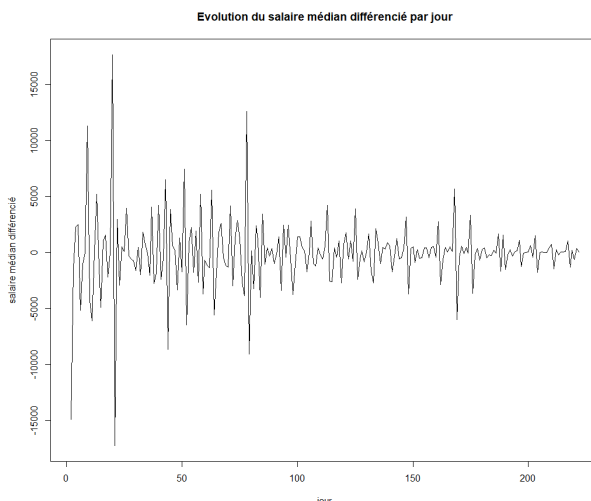


FIGURE 25 – Evolution journalière du salaire médian différencié

Ces séries semblent stationnaires. Toutefois, pour nous en assurer, nous effectuons le test de Dickey-Fuller augmenté. Ce test permet de tester la présence d'une racine unitaire dans la série temporelle, ce qui indiquerait une non stationnarité de la série. Il a pour hypothèse nulle H_0 : la série est non stationnaire. Les résultats du test sont présentés en annexes.

Nous constatons que l'hypothèse nulle est rejetée, avec une p-value de l'ordre de 0.01 pour chacune des deux séries temporelles.

Nous avons également effectuée la régression permettant de tester l'existence d'une tendance linéaire, et cette fois les coefficients de la régression et le modèle global sont clairement non significatifs. Les résultats sont présentés en annexes.

Nous avons donc assez d'éléments pour considérer la série différenciée comme étant stationnaire.

3.3 Modélisation de la série temporelle

Nous souhaitons dans ce paragraphe, faire correspondre la série temporelle que nous étudions à un modèle de série temporelle. Nous utilisons la méthode de Box-Jenkins qui se résume en 4 étapes :

- **Identification** : choix des ordres p (ordre de la partie autorégressive), d (ordre de différenciation) et q (ordre de la partie moyennes mobile) ;
- **Estimation** : estimation des paramètres pour le(s) modèle(s) retenu(s) ;
- **Validation du modèle** : vérification de la blancheur des résidus ;
- **Choix du modèle** : au cas où plusieurs modèles passent les étapes précédentes.

3.3.1 La série des salaires moyens

- **Identification du modèle**

Nous représentons le graphe des autocorrélations et celui des autocorrélations partielles de la série différenciée du salaire moyen.

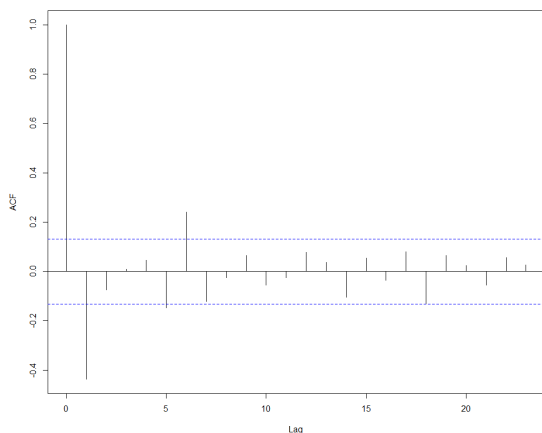


FIGURE 26 – Graphe des autocorrélations de la série différenciée

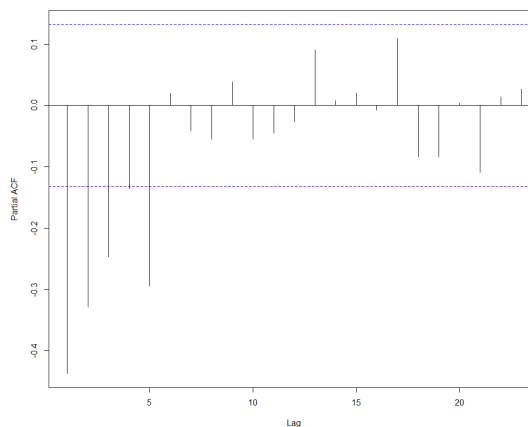


FIGURE 27 – Graphe des autocorrélations partielles de la série différenciée

A partir de ces graphiques, nous sommes tentés de représenter la série du salaire moyen par deux modèles : le modèle **ARIMA(1, 1, 0)**³ et le modèle **ARIMA(6, 1, 0)**⁴.

- **Estimation des paramètres du modèle**

Nous estimons maintenant les paramètres des deux modèles retenus. Le premier modèle ARIMA(1, 1, 0) s'écrit :

$$(I - B)Y_t - \phi_1(I - B)Y_{t-1} = \epsilon_t$$

Avec B l'opérateur retard, I l'opérateur identité et Y la série étudiée. L'estimation donne :

	ϕ_1
estimation	-0.4508
écart type	0.0612

TABLE 2 – Estimation des paramètres du modèle

La sortie du logiciel est présentée en annexes.

Quant au second modèle ARIMA(6, 1, 0), il s'écrit :

$$(I - B)Y_t - \sum_{i=1}^6 \phi_i(I - B)Y_{t-i} = \epsilon_t$$

L'estimation des différents paramètres est donnée dans le tableau suivant.

3. Modèle autorégressif moyennes mobiles intégré d'ordre (1, 1, 0)

4. Modèle autorégressif moyennes mobiles intégré d'ordre (6,1,0)

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
estimation	-0.8732	-0.8254	-0.7107	-0.5348	-0.4455	-0.0950
écart type	0.0699	0.0870	0.0972	0.0967	0.0863	0.0692

TABLE 3 – Estimation des paramètres du modèle

La sortie du logiciel est présentée en annexes.

• Validation du modèle

Nous commençons par représenter le graphe des autocorrélations des résidus.

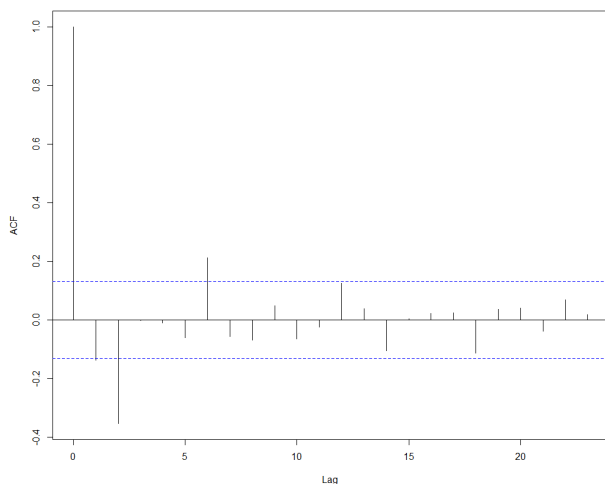


FIGURE 28 – Graphe d'autocorrélation des résidus du modèle ARIMA(1,1,0)

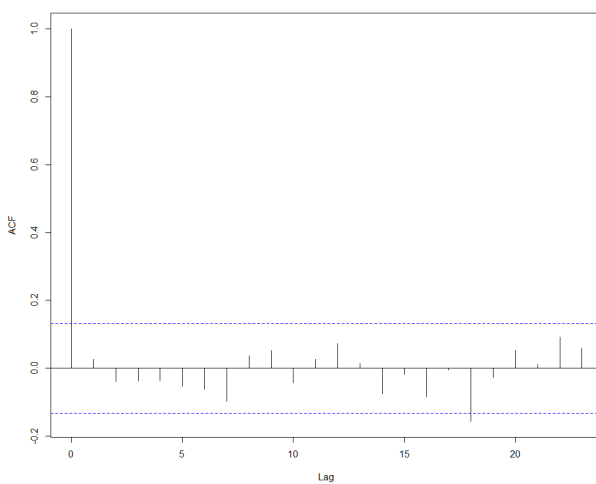


FIGURE 29 – Graphe d'autocorrélation des résidus du modèle ARIMA(6,1,0)

Le premier modèle semble avoir des résidus autocorrélés tandis que ce n'est pas le cas du second. Pour nous en assurer, nous effectuons le test de Box-Pierce sur la série des résidus. Ce test permettra de déterminer si la série des résidus présente une autocorrélation significative. L'hypothèse nulle du test est H_0 : la série n'est pas autocorrélée.

Nous obtenons que la p-valeur pour le premier modèle ARIMA(1,1,0) est de **0.04**. On rejette donc l'hypothèse nulle de non autocorrélation des résidus. Pour le second modèle ARIMA(6,1,0), le test affiche une p-valeur de **0.7**. On ne peut donc pas rejeter l'hypothèse nulle selon laquelle les résidus ne sont pas autocorrélés.

Finalement, dans notre étude, le salaire moyen sera représenté par le modèle **ARIMA(6,1,0)**.

3.3.2 La série des salaires médians

• Identification du modèle

Les graphiques des autocorrélations et des autocorrélations partielles de la série différenciée du salaire médian nous emmènent à considérer le modèle **ARIMA(1,1,0)**.

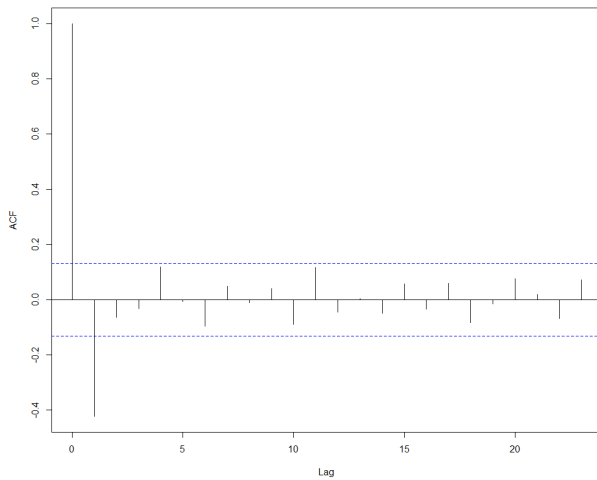


FIGURE 30 – Graphe des autocorrélations de la série différenciée du salaire médian

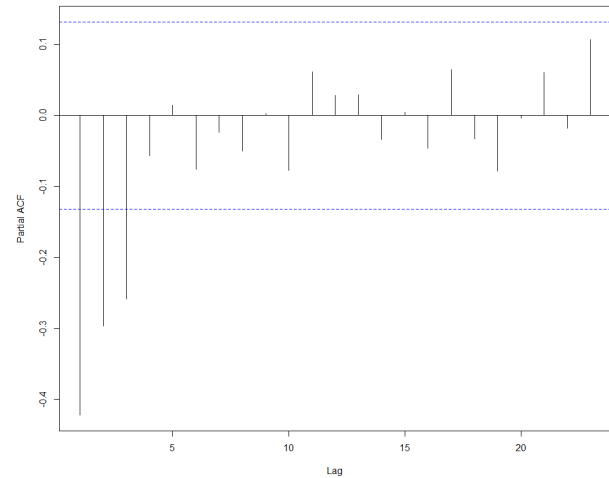


FIGURE 31 – Graphe des autocorrélations partielles de la série différenciée du salaire médian

- **Estimation des paramètres du modèle**

Il s'agit d'un modèle ARIMA(1,1,0). Comme précédemment, il s'écrit

$$(I - B)Y_t - \phi_1(I - B)Y_{t-1} = \epsilon_t$$

Avec Y_t représentant cette fois la série des médianes.

Nous obtenons un coefficient ϕ_1 estimé à **-0.4657** avec un écart type de **0.0632**. La sortie du logiciel est présentée en annexes.

- **Validation du modèle**

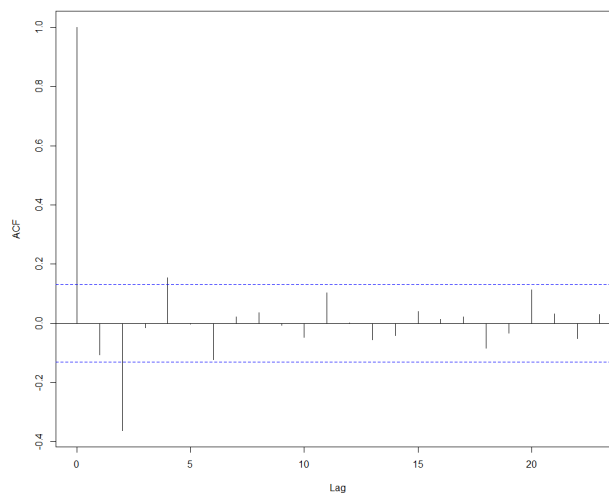


FIGURE 32 – Graphe des autocorrélations des résidus de la série

Ce graphe suggère une éventuelle présence d'autocorrélation des résidus. Nous effectuons le test de Box-Pierce sur la série des résidus. Ce dernier nous fournit une p-valeur de **0.11**. Nous ne pouvons donc pas rejeter l'hypothèse nulle selon laquelle les résidus ne sont pas autocorrélés.

Finalement, dans notre étude, le salaire médian sera représenté par le modèle **ARIMA(1,1,0)**.

3.4 Prédiction du niveau de salaire

Dans ce paragraphe, nous nous intéressons à la prédiction de la série du salaire moyen et à celle du salaire médian. Nous faisons une prédiction sur un mois. Celle-ci est basée sur la modélisation faite dans le paragraphe précédent. Ces prévisions ainsi qu'un intervalle de confiance à 80% et à 95% sont présentés sur les graphiques suivants.

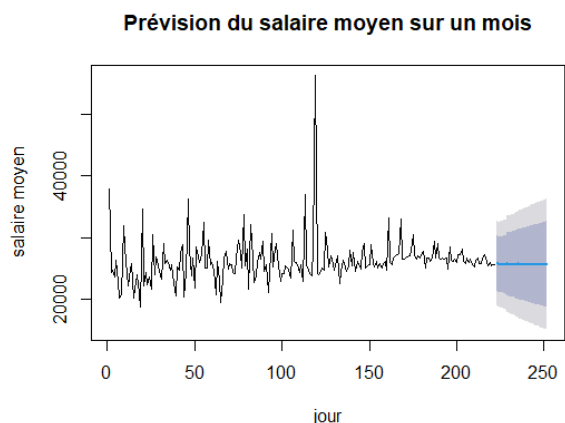


FIGURE 33 – Prédiction de la série du salaire moyen

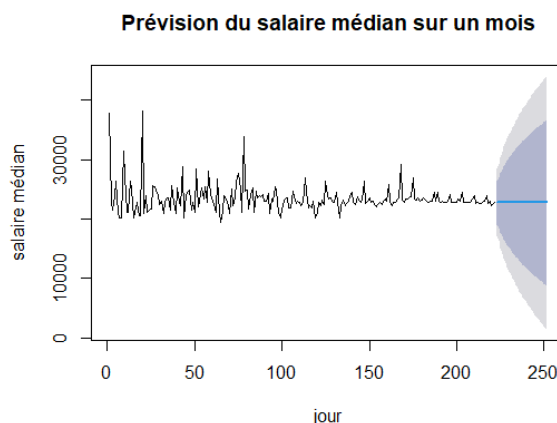


FIGURE 34 – Prédiction de la série du salaire médian

Les prévisions présentent une très faible volatilité. Cela s'explique par le fait que nous observons les salaires sur une période trop courte. La théorie économique nous enseigne que les salaires sont rigides de manière globale. Ainsi, malgré l'inflation enregistrée, il faudrait les observer sur un plus grand intervalle de temps pour observer des variations significatives dans les prévisions.

Nous remarquons également que l'intervalle de confiance est relativement large, particulièrement pour la série du salaire médian.

4 Les déterminants du salaire : un modèle économétrique

Dans le but de déterminer les variables de notre base de données qui influencent significativement le salaire en France, nous effectuons un modèle économétrique dans lequel nous régressons le logarithme du salaire sur différentes caractéristiques.

4.1 Un premier modèle : estimation par moindres carrés ordinaires

4.1.1 Ecriture et analyse du modèle

Les analyses descriptives que nous avons effectuées, l'expertise métier ainsi que la revue de littérature que nous avons effectuée nous inspirent un premier modèle incluant un certain nombre de variables. Ce modèle s'écrit :

$$\begin{aligned} \log(\text{salaire}_i) = & \alpha_0 + \alpha_1 * \text{metier}_i + \alpha_2 * \text{secteur}_i + \alpha_3 * \text{departement}_i + \\ & \alpha_4 * \text{contrat}_i + \alpha_5 * \text{qualification}_i + \alpha_6 * \text{indicAvantages}_i + \\ & \alpha_7 * \text{categorieSocioPro}_i + \alpha_8 * \text{DureeTravail}_i + \alpha_9 * \text{experienceExigee}_i + \\ & \alpha_{10} * \text{indicAlternance}_i + \alpha_{11} * \text{NombreFormations}_i + \alpha_{12} * \text{mois}_i + \epsilon_i \end{aligned}$$

Avec :

- *metier* : le groupe de métier ;
- *secteur* : le secteur d'activité ;
- *contrat* : le type de contrat ;
- *qualification* : le niveau de qualification demandé ;
- *indicAvantages* : une indicatrice indiquant si l'offre propose des avantages extra-salariaux ;
- *categorieSocioPro* : la catégorie socio-professionnelle ;
- *DureeTravail* : la durée du travail c'est à dire temps plein ou temps partiel ;
- *experienceExigee* : l'expérience demandée par l'employeur ;
- *indicAlternance* : l'indicatrice indiquant s'il s'agit d'une alternance ;
- *NombreFormations* : le nombre de formations demandées par l'employeur ;
- *mois* : le mois dans lequel l'offre a été créée.

Ces variables ont été expliquées plus en détail dans la description des données (partie 2.4)

Toutes les variables explicatives sont des variables qualitatives catégorielles sauf le nombre de formations. Ainsi, pour l'exécution du modèle, des indicatrices sont créées pour chaque modalité de chacune des variables qualitatives catégorielles.

Nous estimons le modèle considéré par moindres carrés ordinaires. En analysant les résultats, nous nous rendons compte qu'une variable n'est pas significative, car aucune de ses modalités (représentée chacune par une variable indicatrice) ne l'est au seuil de 5%. Il s'agit de la variable indiquant le mois de création de l'offre.

Nous retirons donc cette variable de notre modèle et étudions le nouveau modèle obtenu qui s'écrit :

$$\begin{aligned} \log(\text{salaire}_i) = & \alpha_0 + \alpha_1 * \text{metier}_i + \alpha_2 * \text{secteur}_i + \alpha_3 * \text{departement}_i + \\ & \alpha_4 * \text{contrat}_i + \alpha_5 * \text{qualification}_i + \alpha_6 * \text{indicAvantages}_i + \\ & \alpha_7 * \text{categorieSocioPro}_i + \alpha_8 * \text{DureeTravail}_i + \alpha_9 * \text{experienceExigee}_i + \\ & \alpha_{10} * \text{indicAlternance}_i + \alpha_{11} * \text{NombreFormations}_i + \epsilon_i \end{aligned}$$

Après estimation par moindres carrés ordinaires du modèle, toutes les variables sont significatives. Le modèle est globalement significatif et le coefficient de détermination R^2 du modèle vaut **0.4**.

4.1.2 Validation du modèle

Il s'agit maintenant de tester les hypothèses d'homoscédasticité, d'autocorrélation et de normalité des résidus de notre régression.

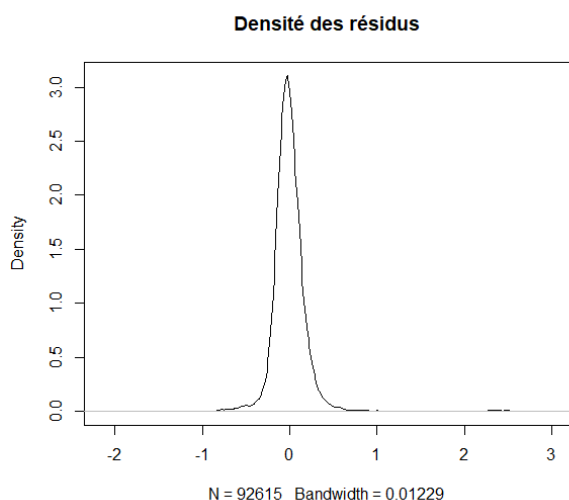


FIGURE 35 – Estimation de la densité des résidus

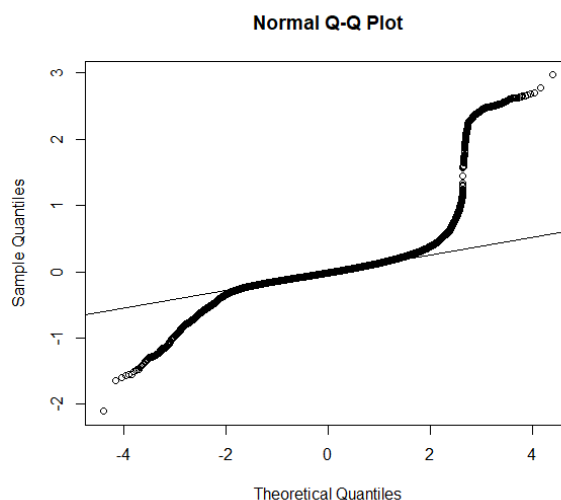


FIGURE 36 – Q-Q plot des résidus

Selon les deux graphiques ci-dessus, la normalité des résidus est une hypothèse plausible.

Nous analysons l'homoscédasticité des résidus à l'aide du test de Breusch-Pagan. L'hypothèse nulle du test est l'homoscédasticité des résidus. La sortie du logiciel est présentée en annexes. Le test nous fournit une p-valeur inférieure au seuil de 1%. Nous rejetons donc l'hypothèse d'homoscédasticité des résidus.

De même, nous analysons l'autocorrélation des résidus à l'aide du test de Ljung-Box, dont l'hypothèse nulle est l'absence d'autocorrélation des résidus. Le test nous fournit une p-valeur inférieure au seuil de 1%. Nous rejetons donc l'hypothèse de non autocorrélation des résidus.

Finalement, le modèle que nous avons réalisé n'est pas valide et nous mettons en place une alternative.

4.2 Le modèle final : estimation par moindres carrés généralisés

4.2.1 Ecriture et estimation du modèle

Les hypothèses d'homoscédasticité et de non autocorrélation des résidus n'étant pas respectées, nous implémentons la même régression en utilisant cette fois comme estimateur l'estimateur par moindres carrés généralisés. Dans ce contexte, l'hétéroscédasticité et la non autocorrélation des résidus sont autorisées. Nous rappelons le modèle :

$$\begin{aligned} \log(\text{salaire}_i) = & \alpha_0 + \alpha_1 * \text{metier}_i + \alpha_2 * \text{secteur}_i + \alpha_3 * \text{departement}_i + \\ & \alpha_4 * \text{contrat}_i + \alpha_5 * \text{qualification}_i + \alpha_6 * \text{indicAvantages}_i + \\ & \alpha_7 * \text{categorieSocioPro}_i + \alpha_8 * \text{DureeTravail}_i + \alpha_9 * \text{experienceExigee}_i + \\ & \alpha_{10} * \text{indicAlternance}_i + \alpha_{11} * \text{NombreFormations}_i + \epsilon_i \end{aligned}$$

Les graphiques Q-Q plot et de densité des résidus sont semblables à ceux obtenus précédemment. Les résultats de la régression sont maintenant interprétables.

4.2.2 Interprétation des résultats

La liste complète des coefficients de la régression est présentée en annexes.

Au niveau des métiers, certains groupes de métiers ont une très forte influence sur le salaire. C'est le cas des groupes N31 (personnel navigant du transport maritime et fluvial), J11 (practiciens médicaux), N32 (Personnel sédentaire du transport maritime et fluvial) et J12 (practiciens médico-techniques). Ces groupes de métiers ont un salaire supérieur en moyenne, respectivement de 58.02%, 56,50%, 45.82% et 31.97% à celui d'un employé du groupe de référence A11 (engins agricoles et forestiers), toute chose égale par ailleurs. Toutefois, ce résultat est à considérer avec précaution pour les secteurs N31 et N32 car ces derniers sont peu représentés dans nos données.

Nous remarquons également que les métiers du domaine professionnel G (hôtellerie-restauration, tourisme, loisirs et animation) sont légèrement moins bien rémunérés par rapport au groupe de référence A11, sauf le groupe de métiers G16 (production culinaire). Toutefois, ce domaine professionnel est assez bien représenté dans nos données, le graphe du nombre d'offres par domaine professionnel est présenté en annexes. Ainsi, on en déduit que ce domaine propose beaucoup d'emplois avec des bas salaires. Dans le même temps, certains domaines professionnels ont tous leurs groupes de métiers mieux rémunérés que le groupe de référence, toute chose égale par ailleurs. C'est le cas des domaines B (Arts et façonnage d'ouvrages d'art) et F (Construction, Bâtiment et Travaux Publics).

Au niveau sectoriel, l'influence la plus importante sur le salaire est celle du secteur 9 (services de soutien aux industries extractives). Ce dernier augmenterait, toute chose égale par ailleurs, le salaire de 79.85% par rapport au secteur 1 (Culture et production animale, chasse et services annexes). Toutefois, nous ne disposons que de 207 données de salaires relatives au secteur 9, ce qui peut paraître peu pour justifier de la robustesse de ce résultat. Hormis ce secteur, les secteurs qui influencent le plus le salaire sont dans l'ordre les secteurs d'activité 99 (activités des organisations et organismes extraterritoriaux), 35 (production et distribution d'électricité, de gaz, de vapeur et d'air conditionné) et 21 (industrie pharmaceutique). Ces secteurs augmentent en moyenne le salaire respectivement de 35.5%, de 19.22% et de 13.15% par rapport au secteur 1, toute chose égale par ailleurs.

Au niveau professionnel, les emplois avec la qualification "Bac+5 et plus ou équivalents" ont un salaire supérieur en moyenne de 21.12% aux salaires dans les emplois

n'exigeant aucune formation requise, toute chose égale par ailleurs. Ensuite viennent les emplois avec la qualification "Bac+3, Bac+4 ou équivalents" qui ont un salaire supérieur en moyenne de 13.79%, toute chose égale par ailleurs.

De même, la catégorie socio-professionnelle "cadre" est celle qui influence le plus le salaire. Les cadres ont en moyenne un salaire supérieur de 23.80% à celui des manoeuvres, toute chose égale par ailleurs.

Au niveau départemental, les emplois des départements 75(Paris), 2B(Haute-Corse), 93(Seine-Saint-Denis), 92(Hauts-de-Seine), 2A(Corse-du-Sud) et 94(Val-de-Marne) ont les plus hauts salaires avec un salaire supérieur en moyenne, respectivement de 9.58%, 9.25%, 8.24%, 7.05%, 6.90% et 6.88% à ceux du département 01 (Ain), toute chose égale par ailleurs. Il s'agit de départements des régions Île-de-France et Corse.

Nous constatons aussi qu'un CDI a un salaire supérieur en moyenne de 3.70% à celui d'un CDD, toute chose égale par ailleurs. Un alternant a un salaire inférieur en moyenne de 37.05% par rapport à un non alternant, toute chose égale par ailleurs. Les emplois exigeant une première expérience professionnelle à l'employé ont un salaire supérieur en moyenne de 3.62% à celui des emplois acceptant des débutants, toute chose égale par ailleurs. Aussi, de manière controversée, le fait d'avoir des avantages extra-salariaux augmente le salaire en moyenne de 0.55%, toute chose égale par ailleurs. Cela sous-entend que c'est justement aux salariés les mieux payés que des avantages extra-salariaux sont le plus souvent accordés.

Conclusion

L'objectif dans cette étude a été dans un premier temps de construire une base de données sur les offres d'emploi en France en 2023 en se servant uniquement de ressources en ligne puis d'analyser à la fois l'évolution des salaires et les déterminants de ces salaires. Cette étude est utile à la prise de décision des acteurs du marché du travail, du directeur des ressources humaines à l'étudiant tout juste diplômé en passant par l'employé désirant changer de poste. Ces acteurs ont besoin de comprendre l'offre d'emploi, la fixation des salaires ainsi que les dynamiques du marché du travail.

A ce titre, notre étude nous renseigne sur l'évolution du niveau général de salaire au cours des derniers mois. Elle nous montre également la forte dispersion des salaires en France. De plus, ces salaires varient selon des caractéristiques géographiques, sectorielles et professionnelles. Dans ce contexte, les employés du domaine médical, du domaine de la construction, de l'art ou encore du domaine de la logistique et transport gagnent plus relativement aux autres domaines professionnels. De même, les cadres, les personnes hautement qualifiées ou encore les employés en Île-de-France ou en Corse ont des plus hauts salaires.

Toutefois, notre étude présente des limites. Ces limites sont principalement liées à la courte période observée et à la faible représentativité de nos données en ce qui concerne certaines caractéristiques. Par conséquent il existe un manque de robustesse de nos résultats relativement à ces caractéristiques. La solution pourrait être de reproduire le processus d'acquisition de données sur une période assez longue pour, d'une part, avoir plus de points temporels pour analyser l'évolution des salaires dans le temps et d'autre part, avoir des données plus représentatives pour analyser les déterminants du salaire.

De plus, dans le processus d'acquisition de nos données, nous avons commencé à récolter les offres d'emploi en juillet. Ainsi les offres d'emploi pour les mois de janvier ou février que nous avons sont des offres d'emploi qui n'ont pas été pourvues depuis janvier ou février. Ces offres ont donc potentiellement des caractéristiques particulières inobservées qui font qu'elles n'ont pas été pourvues durant tout ce temps. Une tentative de solution a été d'inclure la variable indiquant le mois dans le modèle économétrique. Cependant, cette variable s'est avérée non significative.

Enfin, il serait intéressant de combiner les données de Pôle emploi avec des données en provenance d'autres plateformes en ligne dédiées à l'emploi. En particulier, des plateformes qui représenteraient assez bien les hauts salaires.

Bibliographie

- Adrjan, P., & Lydon, R. (2022). Wage Growth in Europe : Evidence From Job Ads (No. 7/EL/22). Central Bank of Ireland.
- Campos-Vazquez, R. M., Esquivel, G., & Badillo, R. Y. (2021). How has labor demand been affected by the COVID-19 pandemic ? Evidence from job ads in Mexico. *Latin American Economic Review*, 30(1), 1-42.
- De Maricourt, C., & Niang, M. (2022). Les offres d'emploi en ligne, nouvelle source de données sur le marché du travail.
- Faryna, O., Pham, T., Talavera, O., & Tsapin, A. (2022). Wage and unemployment : Evidence from online job vacancy data. *Journal of Comparative Economics*, 50(1), 52-70.
- Shuai, X., Chmura, C., & Stinchcomb, J. (2021). COVID-19, labor demand, and government responses : Evidence from job posting data. *Business Economics*, 56(1), 29-42.
- Ziegler, L. (2020). Skill demand and posted wages : evidence from online job ads in Austria. Vienna : Department of Economics, University of Vienna.
- Lien : Liste des codes NAF
- Lien : Liste des codes ROME
- Lien : Données de l'Insee sur les salaires en 2019

Annexes

Statistiques descriptives

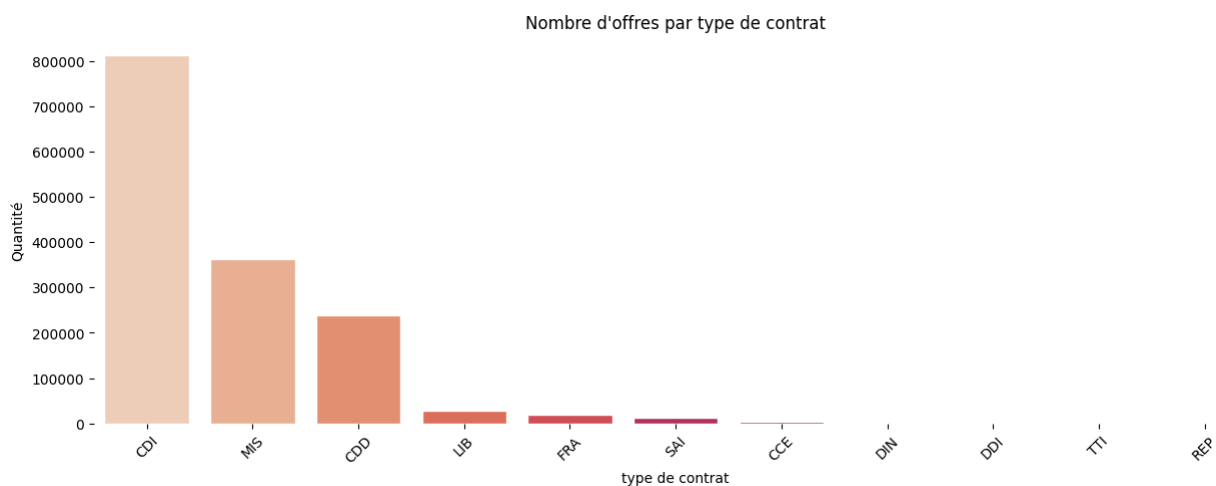


FIGURE 37 – Nombre d'offres par type de contrat



FIGURE 38 – Nombre d'offres par secteur

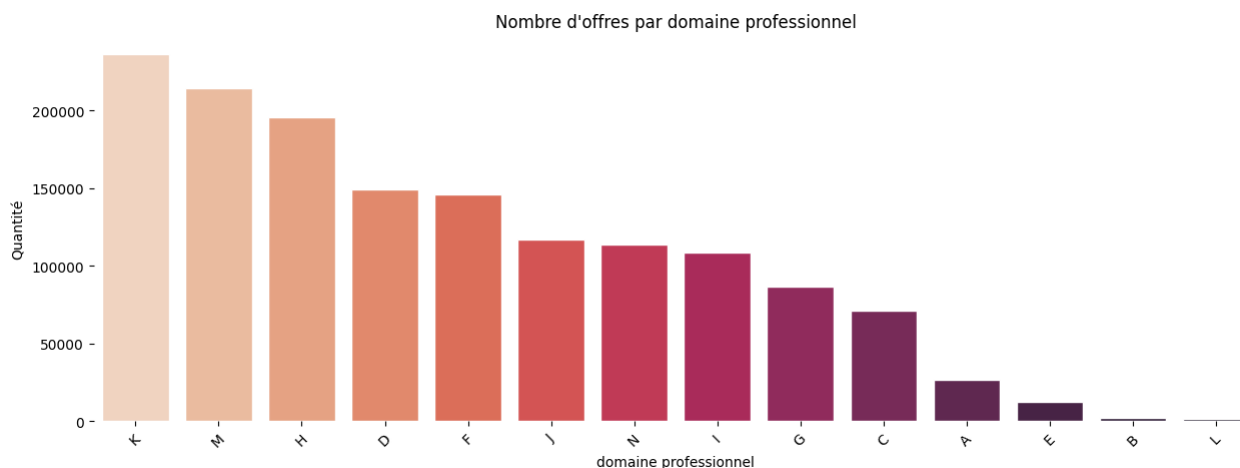


FIGURE 39 – Nombre d'offres par domaine professionnel

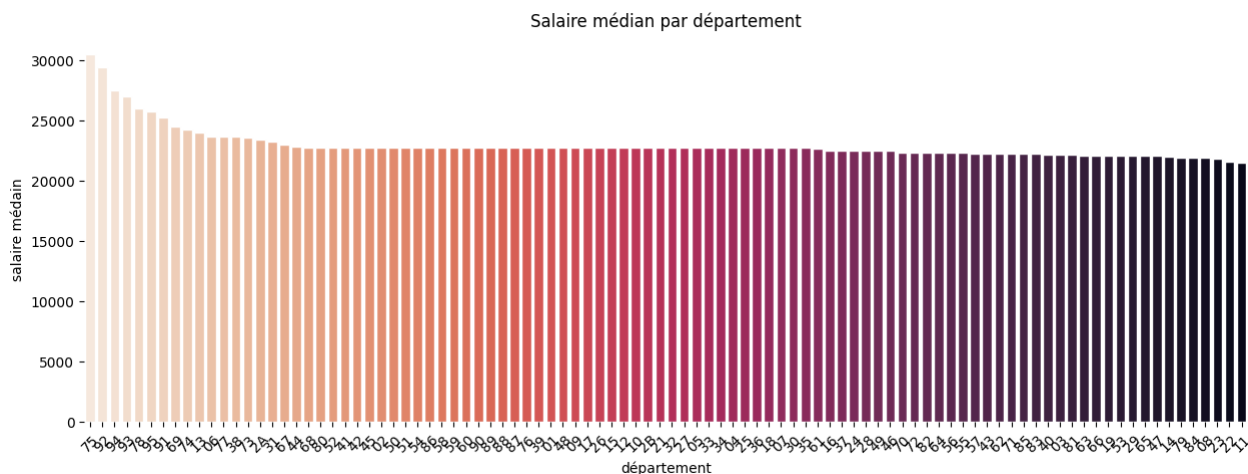


FIGURE 40 – Salaire médian par département

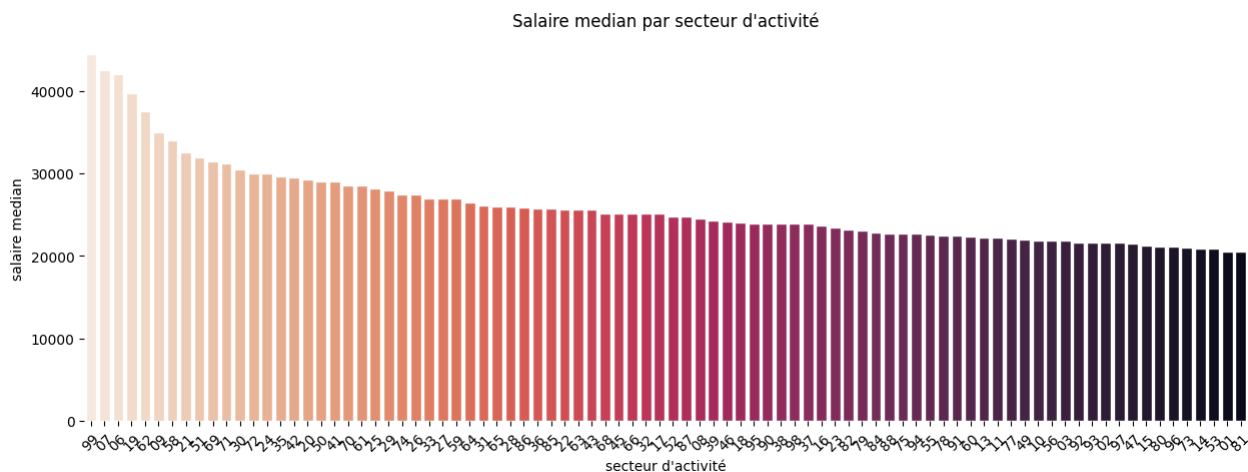


FIGURE 41 – Salaire médian par secteur d'activité

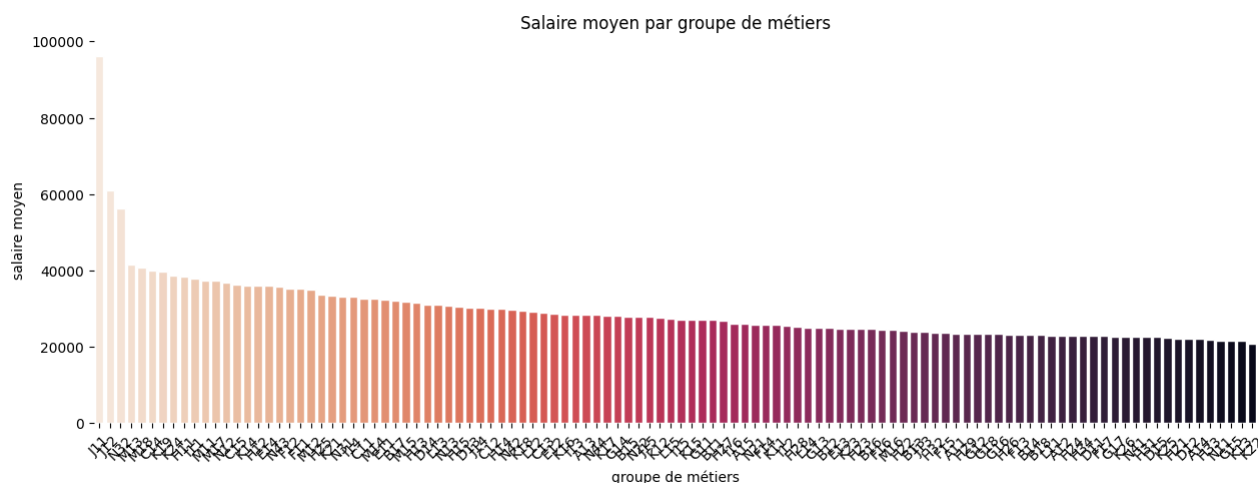


FIGURE 42 – Salaire moyen par groupe de métiers

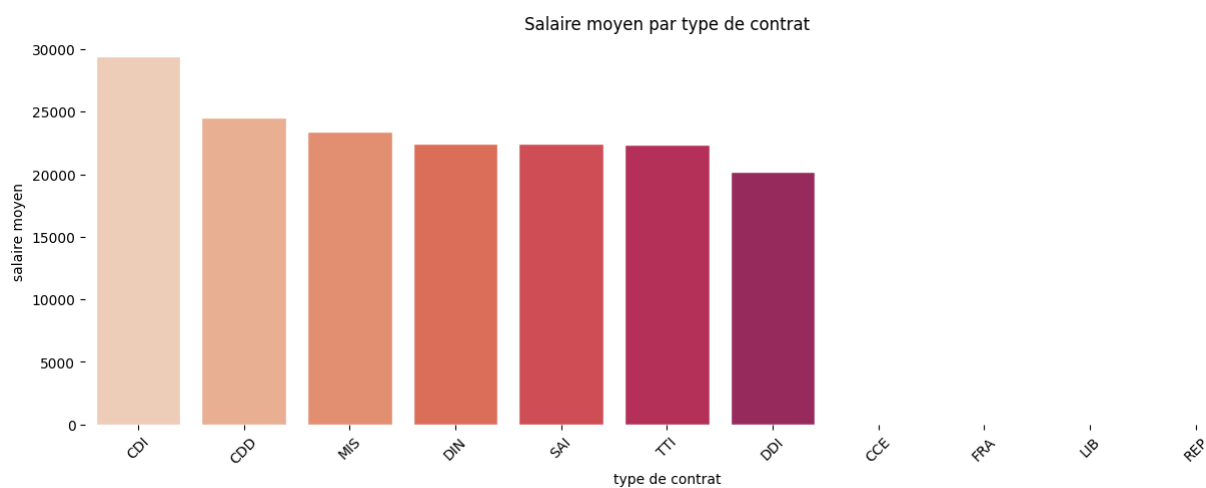


FIGURE 43 – Salaire moyen par type de contrat

Intitulé de poste	salaire moyen	salaire médian	nb salaires
Responsable des relations publiques digitales	330000.0	330000.0	7
Chirurgien	173764.7	160000.0	17
Gastro-entérologue	167516.0	136480.0	10
Diagnostiqueur construction amiante	160060.0	27500.0	5
Ophtalmologue	155333.3	145000.0	8
Câbleur haute et basse tension	140396.8	27950.0	15
Agent de transit import	120728.2	30300.0	20
Chef de produit assurance vie	118200.0	75000.0	5
Chirurgien-dentiste	117938.3	120000.0	801
Médecin biologiste	117314.8	120000.0	27
Médecin urgentiste	111720.2	120834.0	98
Biologiste médical	109075.5	120000.5	85
Médecin du travail et de prévention	108339.3	113500.0	28
Médecin Anesthésiste Réanimateur -MAR-	105660.6	99750.0	53
Radiologue	105443.4	90000.0	28
Responsable de projet de développement local	105354.0	51000.0	5
Gynécologue-obstétricien	105061.7	95000.0	67
Agent d'environnement de proximité	104336.4	28770.0	5
Médecin spécialiste	103730.7	105000.0	39
Pneumologue	102546.1	100000.0	15

TABLE 4 – 20 emplois les mieux rémunérés selon nos données

Un élément de comparaison concernant ce tableau pourrait être les estimations de salaires par intitulé de poste faites par Indeed avec les données de Pôle emploi : <https://fr.indeed.com/cmp/P%C3%B4le-Emploi-5/salaries>. Toutefois les emplois ne sont pas tout à fait comparables car nous avons, de toute évidence, plus d'intitulés de poste que ceux présentés par Indeed.

Catégorie socio-professionnelle	salaire mensuel brut 2021
Cadres	5773
Professions intermédiaires	3261
Employés	2339
Ouvriers	2411

TABLE 5 – Salaires mensuels moyens en EQTP en 2021 selon l'Insee

Evolution du niveau de salaire

```
Call:
lm(formula = moyennes ~ seq(1, length(moyennes)))

Residuals:
    Min       1Q   Median       3Q      Max
-6638.9 -1519.2  -492.3   677.7  30038.2

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  25206.942    458.673   54.956  <2e-16 ***
seq(1, length(moyennes))    8.259      3.567    2.316  0.0215 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3405 on 220 degrees of freedom
Multiple R-squared:  0.0238,    Adjusted R-squared:  0.01936
F-statistic: 5.363 on 1 and 220 DF,  p-value: 0.02149
```

FIGURE 44 – test de la présence d’une tendance linéaire dans la série des moyennes

```
Call:
lm(formula = medianes ~ seq(1, length(medianes)))

Residuals:
    Min       1Q   Median       3Q      Max
-4227.5  -894.9  -221.4   491.6  14468.6

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  23862.018    309.619   77.069  <2e-16 ***
seq(1, length(medianes))   -4.031      2.408   -1.674  0.0955 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2299 on 220 degrees of freedom
Multiple R-squared:  0.01258,    Adjusted R-squared:  0.008094
F-statistic: 2.803 on 1 and 220 DF,  p-value: 0.09549
```

FIGURE 45 – test de la présence d’une tendance linéaire dans la série des médianes

```
Call:
lm(formula = diff_moy ~ seq(1, length(diff_moy)))

Residuals:
    Min       1Q   Median       3Q      Max
-32088  -1764     -4    1399   28082

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -206.525    644.545   -0.32   0.749
seq(1, length(diff_moy))    1.361      5.034    0.27   0.787

Residual standard error: 4775 on 219 degrees of freedom
Multiple R-squared:  0.0003334, Adjusted R-squared:  -0.004231
F-statistic: 0.07304 on 1 and 219 DF,  p-value: 0.7872
```

FIGURE 46 – test de la présence d’une tendance linéaire dans la série différenciée des moyennes

```
Call:
lm(formula = diff_med ~ seq(1, length(diff_med)))

Residuals:
    Min       1Q   Median       3Q      Max
-17001.9  -970.8   -21.5   929.3  17893.5

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -255.612    434.272   -0.589   0.557
seq(1, length(diff_med))    1.690      3.392    0.498   0.619

Residual standard error: 3217 on 219 degrees of freedom
Multiple R-squared:  0.001133, Adjusted R-squared:  -0.003428
F-statistic: 0.2483 on 1 and 219 DF,  p-value: 0.6187
```

FIGURE 47 – test de la présence d’une tendance linéaire dans la série différenciée des médianes

Augmented Dickey-Fuller Test

data: moyennes
Dickey-Fuller = -5.0268, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

FIGURE 48 – Test de stationnarité de la série des moyennes

Augmented Dickey-Fuller Test

data: medianes
Dickey-Fuller = -5.9664, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

FIGURE 49 – Test de stationnarité de la série des médianes

```
Call:
arima(x = moyennes, order = c(1, 1, 0))

Coefficients:
    ar1
-0.4508
s.e.   0.0612

sigma^2 estimated as 18130470: log likelihood = -2160.5, aic = 4324.99
```

FIGURE 50 – Estimation du modèle ARIMA(1,1,0) de la série du salaire moyen

```
Call:
arima(x = moyennes, order = c(6, 1, 0))

Coefficients:
    ar1      ar2      ar3      ar4      ar5      ar6
-0.8732 -0.8254 -0.7107 -0.5348 -0.4455 -0.0950
s.e.    0.0699  0.0870  0.0972  0.0967  0.0863  0.0692

sigma^2 estimated as 12239137: log likelihood = -2117.8, aic = 4249.59
```

FIGURE 51 – Estimation du modèle ARIMA(6,1,0) de la série du salaire moyen

```
Call:
arima(x = medianes, order = c(1, 1, 0))

Coefficients:
    ar1
-0.4657
s.e.   0.0632

sigma^2 estimated as 8241312: log likelihood = -2073.38, aic = 4150.77
```

FIGURE 52 – Estimation du modèle ARIMA(1,1,0) de la série du salaire médian

Les déterminants du salaire

studentized Breusch-Pagan test

data: reglin2
BP = 2352.4, df = 316, p-value < 2.2e-16

FIGURE 53 – Test d'homoscédasticité des résidus de la régression par moindres carrés ordinaires

Box-Ljung test

data: residus
X-squared = 417.61, df = 1, p-value < 2.2e-16

FIGURE 54 – Test d'autocorrélation des résidus de la régression par moindres carrés ordinaires


```
Generalized least squares fit by REML
Model: lsalaire ~ metier_agg + secteurActivite + departement + typecontrat + indic_compl + experience
x1ge + indic_alter + nb_formation + dureeTravailLibelleConverti + csp + qualification
Data: data_pe
AIC      BIC      logLik
-6260.442 -3273.187 3447.221

Coefficients:
(Intercept)          10.043349 0.03861849 260.06583 0.0000
metier_aggA12        -0.007835 0.03490379  -0.22448 0.8224
metier_aggA13          0.024057 0.03664382  0.65651 0.5115
metier_aggA14          0.015083 0.03554000  0.42438 0.6713
metier_aggA15          0.008653 0.04491562  0.19264 0.8472
metier_aggB11          0.123910 0.23423168  0.52901 0.5968
metier_aggB12          0.203379 0.23414812  0.86859 0.3851
metier_aggB13          0.112783 0.10939626  1.03096 0.3026
metier_aggB15          0.140017 0.16700052  0.83842 0.4018
metier_aggB16          0.025338 0.04863524  0.52098 0.6024
metier_aggB18        -0.000807 0.04234106  -0.01907 0.9848
metier_aggC11          0.003720 0.03595932  0.15907 0.8736
metier_aggC12          0.018837 0.03631054  0.51879 0.6039
metier_aggC13        -0.025543 0.04649729  -0.54935 0.5828
metier_aggC14        -0.008992 0.04980038  -0.18055 0.8567
metier_aggC15          0.060496 0.03604149  1.67851 0.0933
metier_aggD11        -0.001153 0.03466240  -0.03325 0.9735
metier_aggD12        -0.083797 0.03446757  -2.43117 0.0151
metier_aggD13          0.017301 0.03760486  0.46008 0.6455
metier_aggD14          0.030806 0.03441539  0.89513 0.3707
metier_aggD15        -0.021167 0.03510769  -0.60292 0.5466
metier_aggE11        -0.025892 0.03669940  -0.70550 0.4805
metier_aggE12        -0.045276 0.04741583  -0.95487 0.3396
metier_aggE13        -0.006117 0.04612986  -0.13260 0.8945
metier_aggE14          0.009007 0.08865603  0.10159 0.9191
metier_aggF11          0.079074 0.03494350  2.26291 0.0236
metier_aggF12          0.129600 0.03537513  3.66360 0.0002
metier_aggF13          0.001810 0.04040873  0.04480 0.9643
metier_aggF14          0.038802 0.08954329  0.43333 0.6648
metier_aggF15          0.025160 0.03820490  0.65856 0.5102
metier_aggF16          0.033263 0.03451444  0.96374 0.3352
metier_aggF17          0.020350 0.03527659  0.57687 0.5640
```

FIGURE 55 – Résultats de la régression avec moindres carrés généralisés

```
metier_aggF17          0.020350 0.03527659  0.57687 0.5640
metier_aggG11        -0.014304 0.04493014  -0.31835 0.7502
metier_aggG12        -0.006352 0.03456981  -0.18375 0.8542
metier_aggG13        -0.054002 0.04351922  -1.24087 0.2147
metier_aggG14        -0.003432 0.03646579  -0.09411 0.9250
metier_aggG15        -0.046282 0.03846385  -1.20325 0.2289
metier_aggG16          0.024559 0.03450224  0.71182 0.4766
metier_aggG17        -0.075784 0.03679042  -2.05990 0.0394
metier_aggG18        -0.006732 0.03509305  -0.19184 0.8479
metier_aggH11          0.144129 0.03731925  3.86204 0.0001
metier_aggH12          0.067313 0.03481452  1.93348 0.0532
metier_aggH13          0.071950 0.03654340  1.96889 0.0490
metier_aggH14          0.067057 0.03576570  1.87489 0.0608
metier_aggH15          0.023145 0.03483545  0.66441 0.5064
metier_aggH21        -0.036347 0.03662087  -0.99253 0.3209
metier_aggH22          0.039480 0.03589675  1.09983 0.2714
metier_aggH23          0.034535 0.04015616  0.86002 0.3898
metier_aggH24        -0.041735 0.04375529  -0.95383 0.3402
metier_aggH25          0.130662 0.03590235  3.63937 0.0003
metier_aggH26        -0.009812 0.03625232  -0.27065 0.7867
metier_aggH27          0.053739 0.05013697  1.07185 0.2838
metier_aggH28          0.049702 0.05408127  0.91903 0.3581
metier_aggH29          0.010840 0.03453005  0.31393 0.7536
metier_aggH31        -0.042499 0.06712055  -0.63317 0.5266
metier_aggH32        -0.015961 0.04100194  -0.38928 0.6971
metier_aggH33        -0.038836 0.03687076  -1.05330 0.2922
metier_aggH34          0.020101 0.03952792  0.50852 0.6111
metier_aggI11          0.154716 0.03727647  4.15049 0.0000
metier_aggI12        -0.015252 0.03536024  -0.43133 0.6662
metier_aggI13          0.043358 0.03440497  1.26023 0.2076
metier_aggI14          0.022503 0.03635687  0.61895 0.5360
metier_aggI15          0.193468 0.04214218  4.59083 0.0000
metier_aggI16          0.070504 0.03457568  2.03911 0.0414
metier_aggJ11          0.561533 0.03766008 14.91057 0.0000
metier_aggJ12          0.316776 0.03667429  8.63755 0.0000
metier_aggJ13        -0.016408 0.03447439  -0.47594 0.6341
metier_aggJ14        -0.033883 0.03496612  -0.96902 0.3325
metier_aggJ15          0.063718 0.03436816  1.85399 0.0637
metier_aggK11        -0.189487 0.03514444  -5.39166 0.0000
metier_aggK12        -0.016828 0.03442787  -0.48880 0.6250
metier_aggK13          0.019069 0.03435688  0.55503 0.5789
metier_aggK14          0.089546 0.03564920  2.51186 0.0120
metier_aggK15          0.065190 0.04116260  1.58372 0.1133
metier_aggK16        -0.081280 0.04225099  -1.92375 0.0544
metier_aggK17          0.029765 0.06195684  0.48041 0.6309
metier_aggK18        -0.029307 0.03514269  -0.83395 0.4043
metier_aggK19          0.010607 0.03627568  0.29241 0.7700
metier_aggK21          0.071824 0.03458820  2.07655 0.0378
metier_aggK22        -0.020420 0.03647636  -0.55982 0.5756
metier_aggK23          0.014002 0.03756058  0.37278 0.7093
metier_aggK24          0.054087 0.03899389  1.38706 0.1654
metier_aggK25        -0.055442 0.03657392  -1.51589 0.1296
metier_aggK26        -0.031208 0.04481847  -0.69631 0.4862
metier_aggL11        -0.108395 0.24277615  -0.44648 0.6553
metier_aggL12          0.063640 0.23346270  0.27259 0.7852
```

FIGURE 56 – Résultats de la régression avec moindres carrés généralisés

metier_aggL12	0.063640	0.23346270	0.27259	0.7852
metier_aggL13	-0.020994	0.06511814	-0.32240	0.7471
metier_aggL15	-0.209912	0.23342044	-0.89929	0.3685
metier_aggL15	0.010268	0.04740531	0.21660	0.8285
metier_aggM11	0.114978	0.03672757	3.13058	0.0017
metier_aggM12	0.043819	0.03440905	1.27348	0.2029
metier_aggM13	0.124061	0.04014885	3.09003	0.0020
metier_aggM14	0.002791	0.03715537	0.07511	0.9401
metier_aggM15	-0.000310	0.03477895	-0.00891	0.9929
metier_aggM16	-0.046869	0.03437505	-1.36347	0.1727
metier_aggM17	0.039252	0.03605554	1.08866	0.2763
metier_aggM18	0.114998	0.03500423	3.28526	0.0010
metier_aggN11	-0.037642	0.03468141	-1.08535	0.2778
metier_aggN12	0.182508	0.04118348	4.43158	0.0000
metier_aggN13	0.042837	0.03581744	1.19598	0.2317
metier_aggN22	-0.028815	0.08872543	-0.32476	0.7454
metier_aggN31	0.576118	0.16019401	3.59637	0.0003
metier_aggN32	0.454626	0.08423368	5.39720	0.0000
metier_aggN41	-0.023639	0.03510397	-0.67340	0.5007
metier_aggN42	-0.001293	0.03968897	-0.03257	0.9740
metier_aggN43	0.110300	0.05392793	2.04533	0.0408
metier_aggN44	0.082268	0.05043584	1.63114	0.1029
secteurActivite2	0.066434	0.04784854	1.38843	0.1650
secteurActivite3	0.000928	0.05217260	0.01779	0.9858
secteurActivite6	0.024127	0.11685734	0.20646	0.8364
secteurActivite8	-0.072084	0.05737984	-1.25627	0.2090
secteurActivite9	0.798142	0.23153362	3.44720	0.0006
secteurActivite10	0.060973	0.01663809	3.66465	0.0002
secteurActivite11	0.072269	0.03383030	2.13623	0.0327
secteurActivite13	0.008489	0.04073673	0.20838	0.8349
secteurActivite14	-0.021640	0.03048692	-0.70981	0.4778
secteurActivite15	0.015667	0.05149130	0.30427	0.7609
secteurActivite16	0.052323	0.02749143	1.90326	0.0570
secteurActivite17	0.101386	0.03547898	2.85763	0.0043
secteurActivite18	0.081145	0.03248900	2.49763	0.0125
secteurActivite19	-0.177693	0.23171300	-0.76686	0.4432
secteurActivite20	0.114380	0.02513598	4.55043	0.0000
secteurActivite21	0.130943	0.03031684	4.31914	0.0000
secteurActivite22	0.087879	0.02419590	3.63198	0.0003
secteurActivite23	0.117038	0.02804490	4.17325	0.0000
secteurActivite24	0.127979	0.03251116	3.93646	0.0001
secteurActivite25	0.093344	0.01848084	5.05083	0.0000
secteurActivite26	0.042306	0.02306670	1.83409	0.0666
secteurActivite27	0.123934	0.02420065	5.12111	0.0000
secteurActivite28	0.015549	0.01753910	0.88651	0.3753
secteurActivite29	0.078484	0.02675216	2.93374	0.0033
secteurActivite30	0.079639	0.02673587	2.97872	0.0029
secteurActivite31	0.030530	0.03256358	0.93754	0.3485
secteurActivite32	0.092279	0.02355698	3.91726	0.0001
secteurActivite33	0.080441	0.01763114	4.56242	0.0000
secteurActivite35	0.192254	0.02131664	9.01896	0.0000
secteurActivite36	0.016441	0.02844777	0.57794	0.5633
secteurActivite37	0.064083	0.05161743	1.24150	0.2144

FIGURE 57 – Résultats de la régression avec moindres carrés généralisés

secteurActivite37	0.064083	0.05161743	1.24150	0.2144
secteurActivite38	0.042036	0.02483872	1.69234	0.0906
secteurActivite39	-0.063735	0.07499093	-0.84991	0.3954
secteurActivite41	0.087708	0.02130533	4.11671	0.0000
secteurActivite42	0.020456	0.02428802	0.84221	0.3997
secteurActivite43	0.055947	0.01606810	3.48187	0.0005
secteurActivite45	0.021303	0.01700081	1.25306	0.2102
secteurActivite46	0.060874	0.01604595	3.79370	0.0001
secteurActivite47	0.019218	0.01577349	1.21835	0.2231
secteurActivite49	0.103722	0.01784910	5.81104	0.0000
secteurActivite50	0.026041	0.08334062	0.31246	0.7547
secteurActivite51	0.098922	0.23136957	0.42755	0.6690
secteurActivite52	0.070737	0.01998650	3.53923	0.0004
secteurActivite53	-0.028939	0.04743655	-0.61005	0.5418
secteurActivite55	0.046768	0.01656629	2.82308	0.0048
secteurActivite56	0.035949	0.01625612	2.21141	0.0270
secteurActivite58	0.019666	0.02530168	0.77725	0.4370
secteurActivite59	-0.038133	0.03860437	-0.98778	0.3233
secteurActivite60	-0.019879	0.06869249	-0.28940	0.7723
secteurActivite61	0.010377	0.02862487	0.36250	0.7170
secteurActivite62	0.031501	0.01775952	1.77375	0.0761
secteurActivite63	0.003074	0.02926858	0.10502	0.9164
secteurActivite64	0.075701	0.01755887	4.31126	0.0000
secteurActivite65	0.033419	0.02068764	1.61539	0.1062
secteurActivite66	0.077038	0.01960339	3.92981	0.0001
secteurActivite68	0.069551	0.01765956	3.93844	0.0001
secteurActivite69	0.100609	0.01762735	5.70753	0.0000
secteurActivite70	0.102462	0.01560801	6.56469	0.0000
secteurActivite71	0.054787	0.01614614	3.39316	0.0007
secteurActivite72	0.029917	0.02053293	1.45704	0.1451
secteurActivite73	0.013813	0.02700657	0.51147	0.6090
secteurActivite74	0.053474	0.02180912	2.45191	0.0142
secteurActivite75	0.008923	0.06199813	0.14393	0.8856
secteurActivite77	0.074449	0.02126483	3.50103	0.0005
secteurActivite78	0.104021	0.01535565	6.77415	0.0000
secteurActivite79	-0.032812	0.02771141	-1.18407	0.2364
secteurActivite80	0.014321	0.02026510	0.70670	0.4798
secteurActivite81	0.022883	0.01751203	1.30672	0.1913
secteurActivite82	0.027341	0.01738848	1.57236	0.1159
secteurActivite84	-0.004419	0.01556366	-0.28392	0.7765
secteurActivite85	0.014542	0.01575650	0.92295	0.3560
secteurActivite86	0.047407	0.01567019	3.02528	0.0025
secteurActivite87	0.050848	0.01557480	3.26478	0.0011
secteurActivite88	0.004571	0.01551050	0.29469	0.7682
secteurActivite90	0.042867	0.02577052	1.66342	0.0962
secteurActivite91	-0.016677	0.03163800	-0.52712	0.5981
secteurActivite92	0.038394	0.03617853	1.06123	0.2886
secteurActivite93	0.023147	0.01719813	1.34591	0.1783
secteurActivite94	0.008564	0.01591469	0.53811	0.5905
secteurActivite95	-0.020576	0.02697454	-0.76279	0.4456
secteurActivite96	0.063033	0.01620390	3.89000	0.0001
secteurActivite97	0.082307	0.02410955	3.41388	0.0006
secteurActivite99	0.354333	0.07136900	4.96480	0.0000

FIGURE 58 – Résultats de la régression avec moindres carrés généralisés

secteurActivite99	0.354333	0.07136900	4.96480	0.0000
departement02	-0.014825	0.01212577	-1.22259	0.2215
departement03	0.006171	0.01335801	0.46195	0.6441
departement04	-0.015838	0.01503911	-1.03214	0.2923
departement05	-0.030633	0.01340687	-2.28490	0.0223
departement06	0.027972	0.00937587	2.98336	0.0029
departement07	-0.015875	0.01327496	-1.19587	0.2317
departement08	-0.027157	0.01389294	-1.95473	0.0506
departement09	-0.032083	0.01554281	-2.06419	0.0390
departement10	-0.036431	0.01283322	-2.83879	0.0045
departement11	-0.038765	0.01237042	-3.13367	0.0017
departement12	0.008262	0.01216211	0.67932	0.4969
departement13	0.015451	0.00831619	1.85799	0.0632
departement14	-0.009643	0.01017852	-0.94735	0.3435
departement15	0.026638	0.01599090	1.66585	0.0957
departement16	-0.017361	0.01228164	-1.41353	0.1575
departement17	-0.015339	0.00998613	-1.53605	0.1245
departement18	-0.026830	0.01247329	-2.15097	0.0315
departement19	-0.013793	0.01383421	-0.99705	0.3187
departement21	-0.002129	0.01061183	-0.20063	0.8410
departement22	-0.009522	0.01048644	-0.90804	0.3639
departement23	-0.018211	0.01882754	-0.96724	0.3334
departement24	0.009760	0.01248006	0.78209	0.4342
departement25	0.007558	0.01104379	0.68433	0.4938
departement26	-0.001465	0.01023130	-0.14317	0.8862
departement27	-0.032935	0.01109342	-2.96884	0.0030
departement28	-0.007116	0.01219020	-0.58378	0.5594
departement29	0.011439	0.01002624	1.14095	0.2539
departement2A	0.068071	0.01568087	4.34101	0.0000
departement2B	0.081989	0.01697002	4.83142	0.0000
departement30	-0.004754	0.01034369	-0.45962	0.6458
departement31	0.003124	0.00846231	0.36922	0.7120
departement32	0.006438	0.01497438	0.42993	0.6672
departement33	0.012982	0.00840702	1.54420	0.1225
departement34	0.008976	0.00898512	0.99895	0.3178
departement35	0.014251	0.00919325	1.55017	0.1211
departement36	-0.009406	0.01433171	-0.65629	0.5116
departement37	-0.006428	0.01033681	-0.62184	0.5340
departement38	0.023532	0.00886364	2.65492	0.0079
departement39	-0.006486	0.01336739	-0.48520	0.6275
departement40	0.010103	0.01180233	0.85601	0.3920
departement41	0.004560	0.01152851	0.39550	0.6925
departement42	0.014170	0.01071966	1.32186	0.1862
departement43	-0.011432	0.01525802	-0.74922	0.4537
departement44	0.007860	0.00863983	0.90974	0.3630
departement45	-0.003279	0.01019361	-0.32167	0.7477
departement46	0.009095	0.01507231	0.60345	0.5462
departement47	-0.017202	0.01238493	-1.38898	0.1648
departement48	0.044963	0.02172188	2.06994	0.0385
departement49	0.000701	0.01012022	0.06928	0.9448
departement50	-0.004872	0.01166037	-0.41782	0.6761
departement51	-0.034332	0.00990700	-3.46541	0.0005
departement52	-0.021443	0.01475240	-1.45353	0.1461

FIGURE 59 – Résultats de la régression avec moindres carrés généralisés

departement52	-0.021443	0.01475240	-1.45353	0.1461
departement53	0.009188	0.01345017	0.68310	0.4945
departement54	-0.002431	0.01038550	-0.23409	0.8149
departement55	-0.044913	0.01497194	-2.99978	0.0027
departement56	0.011811	0.01046527	1.12859	0.2591
departement57	-0.014064	0.00992432	-1.41714	0.1564
departement58	-0.005709	0.01526846	-0.37393	0.7085
departement59	-0.015767	0.00860953	-1.83132	0.0671
departement60	-0.007772	0.01086158	-0.71557	0.4743
departement61	-0.037221	0.01361831	-2.73314	0.0063
departement62	-0.032914	0.01029115	-3.19827	0.0014
departement63	-0.008322	0.01042740	-0.79808	0.4248
departement64	-0.011523	0.01010303	-1.14050	0.2541
departement65	-0.009494	0.01465734	-0.64775	0.5171
departement66	-0.030988	0.01186815	-2.61106	0.0090
departement67	0.020215	0.00930837	2.17167	0.0299
departement68	0.022426	0.01059780	2.11607	0.0343
departement69	0.005714	0.00822812	0.69441	0.4874
departement70	-0.010337	0.01644033	-0.62875	0.5295
departement71	-0.023407	0.01067296	-2.19312	0.0283
departement72	0.010359	0.01113323	0.93042	0.3522
departement73	0.043521	0.01106398	3.93356	0.0001
departement74	0.032862	0.00957592	3.43176	0.0006
departement75	0.095177	0.00875277	10.87398	0.0000
departement76	-0.018278	0.00935113	-1.95458	0.0506
departement77	0.035770	0.00938669	3.81069	0.0001
departement78	0.067368	0.00975717	6.90442	0.0000
departement79	-0.000647	0.01239101	-0.05222	0.9584
departement80	0.004964	0.01243503	0.39917	0.6898
departement81	-0.026665	0.01236487	-2.15654	0.0310
departement82	-0.019501	0.01382854	-1.41018	0.1585
departement83	0.003691	0.00908751	0.40615	0.6846
departement84	-0.020526	0.01012042	-2.02817	0.0425
departement85	0.005912	0.01096914	0.53900	0.5899
departement86	-0.038075	0.01133733	-3.35834	0.0008
departement87	-0.023344	0.01232987	-1.89328	0.0583
departement88	-0.025168	0.01269218	-1.98296	0.0474
departement89	0.005974	0.01302474	0.45864	0.6465
departement90	-0.011107	0.01558205	-0.71280	0.4760
departement91	0.043807	0.00985067	4.44715	0.0000
departement92	0.070257	0.00938827	7.48352	0.0000
departement93	0.081326	0.01019566	7.97657	0.0000
departement94	0.068669	0.01053648	6.51725	0.0000
departement95	0.033551	0.00978300	3.42955	0.0006
typeContratC0I	0.036975	0.00204805	18.05366	0.0000
typeContratDDI	0.027050	0.05065859	0.53396	0.5934
typeContratDIN	-0.068696	0.03054912	-2.24871	0.0245
typeContratMIS	-0.094548	0.00419873	-22.51830	0.0000
typeContratSAI	0.016524	0.00841537	1.96358	0.0496
typeContratTTI	-0.195184	0.09436675	-2.06835	0.0386
indic_comp11	0.005350	0.00181128	2.95367	0.0031
experienceExigee	0.036334	0.00175920	20.65381	0.0000
experienceExigees	-0.005051	0.00428779	-1.17792	0.2388

FIGURE 60 – Résultats de la régression avec moindres carrés généralisés

indic_compl1	0.005350	0.00181128	2.95367	0.0031
experienceExigeE	0.036334	0.00175920	20.65381	0.0000
experienceExigeS	-0.005051	0.00428779	-1.17792	0.2388
indic_alter1	-0.374899	0.00618959	-60.56932	0.0000
nb_formation	-0.009809	0.00267850	-3.66211	0.0003
dureeTravailLibelleConvertiTemps partiel	-0.215904	0.01504298	-14.35244	0.0000
dureeTravailLibelleConvertiTemps plein	-0.063342	0.01483895	-4.26862	0.0000
cspouvrier qualifié (P3,P4,OHQ)	0.025023	0.00645651	3.87566	0.0001
cspouvrier spécialisé	-0.006939	0.00526723	-1.31748	0.1877
cspEmployé non qualifié	-0.015238	0.00406280	-3.75068	0.0002
cspEmployé qualifié	0.010380	0.00356169	2.91427	0.0036
cspTechnicien	0.053780	0.00425043	12.65272	0.0000
cspAgent de maîtrise	0.087696	0.00474606	18.47763	0.0000
cspCadre	0.208387	0.00517151	40.29512	0.0000
csp	0.205602	0.01735697	11.84548	0.0000
qualificationPrimaire à 4ème	-0.044260	0.08734814	-0.50670	0.6124
qualification4ème achevée	-0.055607	0.06195178	-0.89758	0.3694
qualification3ème achevée ou Brevet	0.005871	0.01601374	0.36663	0.7139
qualification2nd ou 1ère achevée	0.052213	0.03896932	1.33986	0.1803
qualificationCAP, BEP et équivalents	0.004783	0.00387715	1.23363	0.2173
qualificationBac ou équivalent	0.014080	0.00408797	3.44433	0.0006
qualificationBac+2 ou équivalents	0.072251	0.00435416	16.59346	0.0000
qualificationBac+3, Bac+4 ou équivalents	0.137532	0.00474917	28.95923	0.0000
qualificationBac+5 et plus ou équivalents	0.210941	0.00585183	36.04703	0.0000
qualification	0.100004	0.01253807	7.97605	0.0000

FIGURE 61 – Résultats de la régression avec moindres carrés généralisés

Standardized residuals:
 Min Q1 Med Q3 Max
 -9.12807895 -0.43855414 -0.06933265 0.34536439 12.88886591
 Residual standard error: 0.2306971
 Degrees of freedom: 91763 total; 91447 residual

FIGURE 62 – Résultats de la régression avec moindres carrés généralisés