

ORGANISATION :

Pour démarrer ce projet, nous avons sélectionné le sujet des salaires. Nous avons entamé un brainstorming afin de déterminer notre problématique en fonction des données qui nous étaient disponibles. Une fois la problématique identifiée, nous l'avons divisée en quatre axes : temporalité, inégalité, entreprise et formation. Chacun de nous a pris en charge un axe spécifique pour mener des analyses univariées et croisées.

DONNEES :

Pour traiter nos données, nous avons choisi d'utiliser la bibliothèque Pandas de Python. Nous avons sélectionné environ un ou deux fichiers par personne afin de supprimer les colonnes dont nous n'avons pas besoin et de renommer l'ensemble des colonnes sur chaque fichier, car sur chacun d'entre eux, les colonnes n'avaient jamais le même nom. Ceci a demandé beaucoup de temps pour homogénéiser l'ensemble des données. Enfin, pour chaque analyse, nous avons éliminé les valeurs manquantes dans les colonnes utilisées, et même les données vides. Nous avons sauvegardé toutes les données traitées sur Git. La décision quant aux colonnes à conserver dépendait de nos axes d'analyses et également de la quantité de données à l'intérieur d'une variable. Nous avons estimé qu'une variable avec trop peu de données n'était pas intéressante à traiter.

AXE TEMPORALITE :

Analyse Univariée :

Nous avons effectué une analyse univariée portant sur la temporalité en relation avec la rémunération brute annuelle de l'emploi actuel. Les résultats indiquent que la rémunération du premier emploi et le nombre de mois occupés sur le poste actuel présentent des valeurs de p très élevées. Il semble y avoir une corrélation positive entre le nombre de mois occupés sur le poste actuel et le niveau de rémunération, où une augmentation du nombre de mois semble être associée à une rémunération plus élevée.

Nous avons entrepris une étude sur les diplômés qui ont fourni des réponses à trois reprises. L'objectif est d'analyser le pourcentage d'évolution des salaires à chaque occasion de réponse, en distinguant les réponses fournies après 1 an et 2 ans. Cependant, un problème se pose pour l'année 2020, où nous observons une augmentation moyenne de 50%, ce qui semble improbable. Cette situation pourrait être expliquée par un faible nombre de réponses ou l'impact de la période de la COVID-19. Il est possible qu'une erreur se soit glissée dans mes calculs. Voir Figure 1.

Nous avons ainsi choisi de diriger notre analyse vers le salaire du premier emploi, dans le but de comparer celui-ci avec le salaire de l'emploi actuel pour chaque année.

L'analyse des données de salaires pour les années 2018 à 2023 révèle plusieurs tendances significatives. Globalement, le salaire actuel a connu une augmentation progressive, passant de 34868.12 en 2018 à 39399.53 en 2023, avec des écarts types variant de 8709.454 à 10279.12. En parallèle, les salaires au premier emploi ont également augmenté au fil des ans, avec des moyennes passant de 29973.48 en 2018 à 33024.18 en 2023. Les variations observées soulignent l'évolution des conditions du marché du travail, tandis que les écarts

types signalent la dispersion des salaires. Notons également des valeurs minimales et maximales qui reflètent la diversité des situations salariales, avec des salaires actuels atteignant jusqu'à 141999 en 2022. Ces résultats suggèrent des dynamiques complexes dans le domaine de l'emploi au cours de cette période. Voir Figure 2.

Analyse Croisée :

Ainsi, nous avons opté pour cet angle d'approche afin de centrer notre étude sur le salaire du premier emploi.

Par la suite, dans la phase d'analyse croisée, nous avons entrepris de déterminer quelles variables étaient cruciales à examiner pour le salaire du premier emploi. Nous avons pris la décision de sélectionner des variables permettant une comparaison avec d'autres études sur les salaires actuels, notamment la localisation, le pays, la filière, la date de diplôme, ainsi que le sexe et le secteur d'activité.

1- Analyse Préliminaire

Dans un premier temps nous avons fait une analyse préliminaire avec un test de Tukey pour la rémunération du premier emploi avec la localisation et le pays.

En voici l'analyse :

- Différences de rémunération entre la France et l'étranger : La différence moyenne de rémunération entre la France et l'étranger est de 467,07, mais elle n'est pas statistiquement significative (p-value = 0,816). Cela suggère qu'il n'y a pas de différence significative dans la rémunération entre ces deux groupes.
- Différences de rémunération entre les pays : Il y a des différences significatives de rémunération entre certains pays : Danemark par rapport à la Belgique : +29 000, p-value = 0,035 Suisse par rapport à la Belgique : +56 850, p-value = 0 (très significatif) France par rapport à la Belgique : +20 561, p-value = 0,117 (légèrement significatif)
- Différences de rémunération entre les pays spécifiques : Par exemple, entre la France et le Brésil, la différence moyenne est de 24 461 avec une p-value de 0,026, ce qui suggère une différence significative. Différences de rémunération entre les localisations spécifiques : Par exemple, entre la Suisse et le Danemark, la différence moyenne est de 27 850 avec une p-value de 0,005, ce qui suggère une différence significative.

En conclusion, nos résultats indiquent des différences significatives de rémunération entre certains pays et localisations, tandis que d'autres comparaisons ne montrent pas de différences significatives. Les p-values indiquent le degré de confiance dans les différences observées, où une valeur inférieure à 0,05 est généralement considérée comme statistiquement significative.

Ensuite, nous avons effectué une anova à deux facteurs pour la rémunération du premier emploi avec le sexe et le secteur d'activité.

En voici l'analyse :

- Effet du sexe sur la rémunération : La somme des carrés associée au facteur "sexe" est de 730,41, avec 2 degrés de liberté. Cela suggère qu'il y a une certaine variation de la rémunération liée au sexe.
- Effet du secteur d'activité sur la rémunération : La somme des carrés associée au facteur "secteur_premiere_entreprise" est de 3 262,34 millions, avec 27 degrés de

liberté. Cela indique une variation significative de la rémunération entre les différents secteurs d'activité.

- Interaction entre le sexe et le secteur d'activité : La somme des carrés associée à l'interaction entre le sexe et le secteur d'activité est de 1 114,54 millions, avec 20 degrés de liberté. Cela suggère qu'il y a une interaction significative entre le sexe et le secteur d'activité en termes d'effet sur la rémunération.
- Résidus : La somme des carrés des résidus est de 19 237,68 millions, avec 296 degrés de liberté. Cela représente la variation non expliquée par les facteurs examinés.
- Erreur standard résiduelle : L'erreur standard résiduelle est de 8 061,77. Cela représente l'écart-type moyen des résidus, mesurant la dispersion des observations autour de la moyenne prédite.

En conclusion, l'ANOVA suggère que le sexe, le secteur d'activité et leur interaction ont des effets significatifs sur la rémunération du premier emploi. La variation observée dans la rémunération peut être attribuée à ces facteurs, mais une partie reste inexpliquée (résidus). Les estimations non réalisables pour certains effets peuvent être dues à des déséquilibres dans les données.

Enfin, nous avons réalisé une analyse de régression linéaire pour la rémunération du premier emploi avec la filière et la date d'obtention du diplôme.

En voici l'analyse :

- Intercept (Constante) : L'intercept est estimé à -406 696,4. Cela représente la valeur attendue de la rémunération lorsque toutes les autres variables sont égales à zéro.
- Coefficients des variables explicatives (filière et date_diplome) : Pour chaque unité de changement dans la filière « Energétique – énergies Renouvelables (EnR) », la rémunération diminue d'environ 8 232,5. Pour chaque unité de changement dans la filière « Génie Biologique et Agroalimentaires », la rémunération diminue d'environ 5 052,8. Pour chaque unité de changement dans la filière « Informatique et Gestion (IG) », la rémunération diminue d'environ 1 283,2. ... (pour toutes les autres filières et la variable date_diplome)
- P-valeurs associées aux coefficients : Certains coefficients ont des p-valeurs significatives (inférieures à 0,05), ce qui suggère que ces filières ont un impact significatif sur la rémunération du premier emploi. Par exemple, les filières "Energétique - énergies Renouvelables (EnR)" et "Génie Biologique et Agroalimentaires" sont significatives.
- Residuals (Résidus) : Les résidus représentent les écarts entre les valeurs observées et les valeurs prédites par le modèle. La distribution des résidus semble centrée autour de zéro, indiquant que le modèle est raisonnablement approprié.
- Erreur standard résiduelle : L'erreur standard résiduelle est de 7 606. Cela représente l'écart-type moyen des résidus, mesurant la dispersion des observations autour de la ligne de régression.
- Test F-statistic : Le test F-statistic évalue globalement la significativité du modèle. Ici, la statistique F est de 7,579 avec un p-value très faible (1.616e-11), indiquant que le modèle est significatif dans son ensemble.

En conclusion, le modèle de régression suggère que la filière et la date d'obtention du diplôme ont un impact significatif sur la rémunération du premier emploi. Les filières spécifiques, telles que "Energétique - énergies Renouvelables (EnR)" et "Génie Biologique et Agroalimentaires", sont associées à des variations significatives dans la rémunération. Cependant, l'ensemble du modèle n'explique qu'une petite partie de la variance totale de la rémunération.

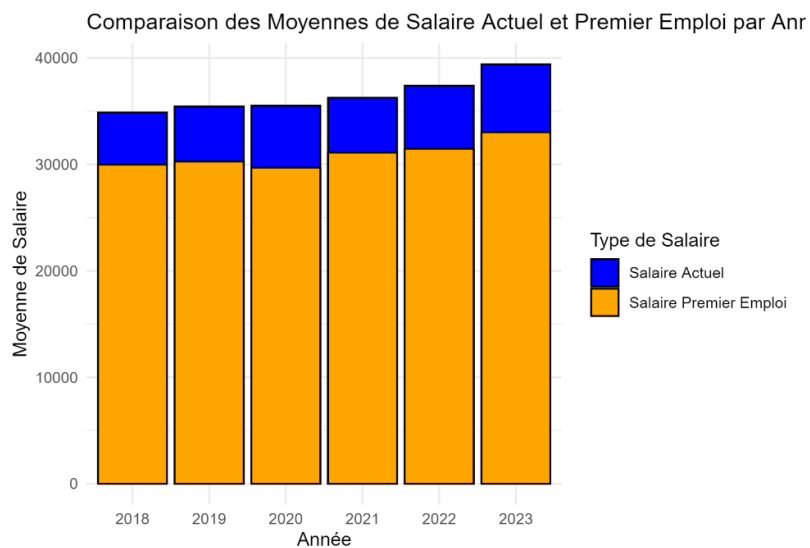
2- Diagramme

Au vue des résultats précédents, nous avons mis en place des diagrammes pour représenter les résultats.

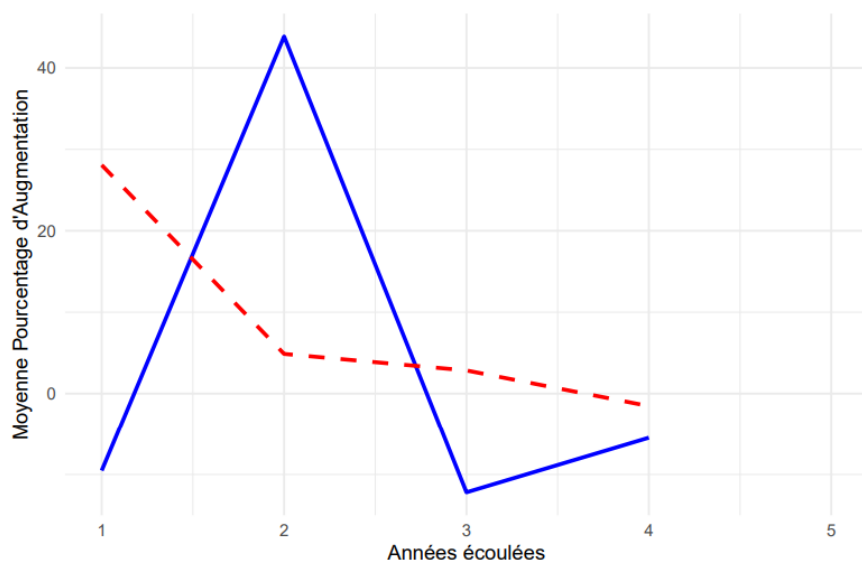
Il est observé dans le diagramme à barres (voir Figure3) que, globalement, dans l'ensemble des secteurs, la part des hommes est associée à un salaire nettement plus élevé. Pour certains secteurs tels que les services liés aux technologies de l'information et de la communication (TIC), la fabrication de produits métalliques, et l'industrie agroalimentaire, le salaire des hommes est presque deux fois supérieur. Dans certains secteurs, où seules des réponses masculines ou féminines ont été fournies, la distinction n'est pas possible. Par ailleurs, certains secteurs révèlent un salaire supérieur pour les femmes, comme dans l'agriculture et l'industrie pharmaceutique.

nous appuyons toujours sur une rémunération moyenne, et il est remarquable sur le diagramme à barres (voir Figure 4) que très peu de répondants ont effectué leur premier emploi à l'étranger, avec seulement 8 pays principalement européens identifiés. La Suisse ressort comme le pays offrant la rémunération moyenne la plus élevée, suivie de la France, bien que cette dernière puisse être classée en deuxième position potentiellement en raison d'un plus grand nombre de réponses. On observe également que le Brésil est le pays offrant la rémunération moyenne la moins élevée, mais cela peut également s'expliquer par les différences de PIB entre les pays.

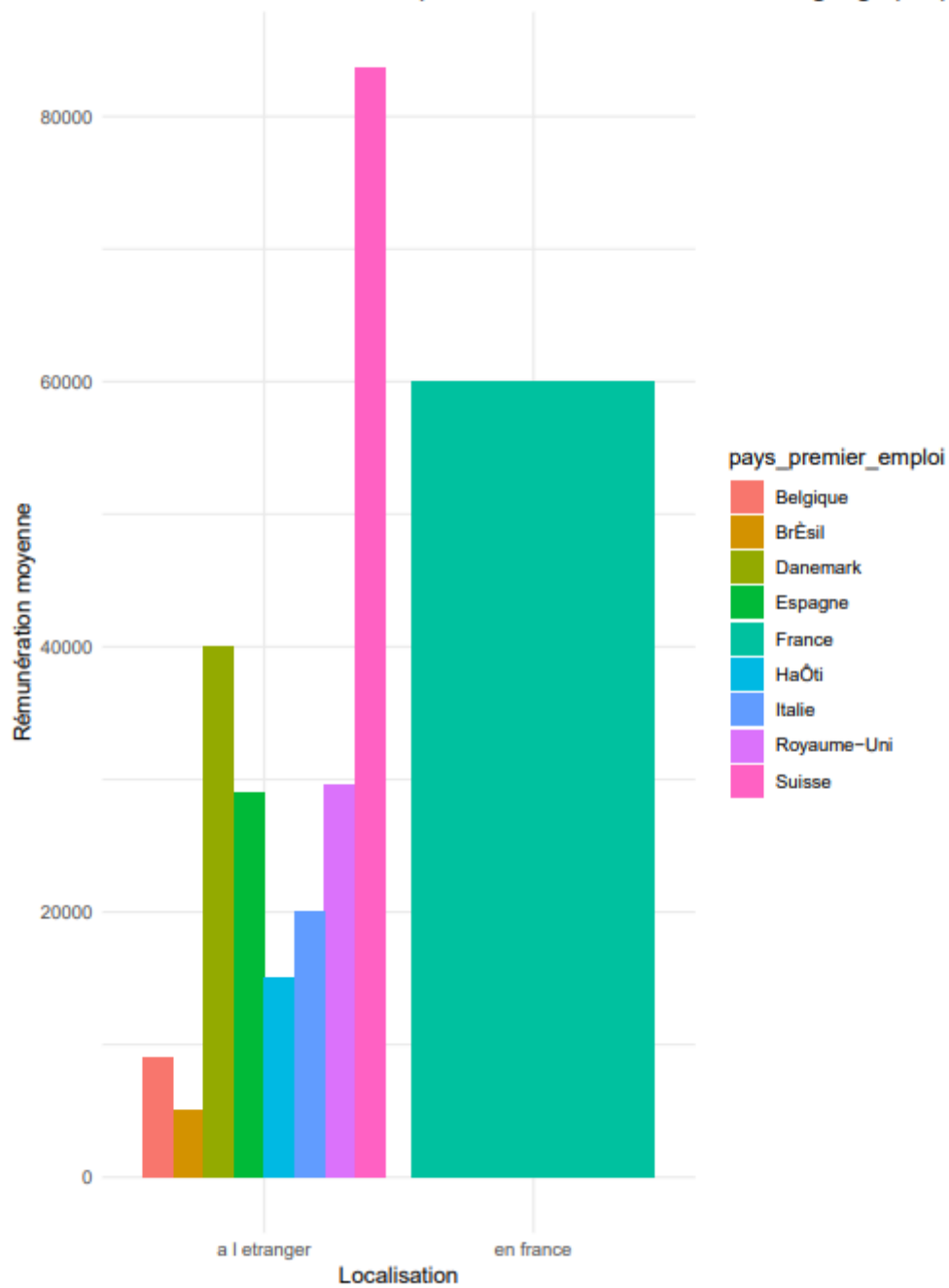
Enfin, il est notable sur le diagramme de dispersion (voir Figure 5) que les données sont assez concentrées pour chaque filière. Les valeurs les plus extrêmes sont généralement associées à des dates d'obtention de diplôme plus anciennes. De plus, on observe une augmentation significative du nombre de réponses aux questionnaires pour les personnes ayant obtenu leur diplôme à partir de 2020. Les valeurs extrêmes, qu'elles soient positives ou négatives, sont rares. Pour la filière Gestion et Business Administration (GBA), la fourchette générale de rémunération va de 19 000 à 38 000, avec la majorité des réponses situées entre 20 000 et 40 000. En ce qui concerne l'Informatique et la Gestion (IG), la rémunération se situe plutôt entre 29 000 et 36 000, avec une distribution plus resserrée.



Annexe :



Rémunération Premier Emploi en fonction de la localisation géographique



Rémunération Premier Emploi en fonction de la Filière et de la Date de D

