

Rapport d'analyse de la collecte de données : Projet Marché du Travail en Informatique

Groupe 2 :

- Youssef EL-OTMANI
- Noé Flandre
- Marc-Alexandre MICHELIN
- Arthur Rubio
- Ludovic Terrasson

Introduction

Ce rapport vise à analyser les données collectées à l'aide d'une méthode de **Web Scraping** sur le site d'emploi **Hello Work** concernant les postes d'**ingénieurs logiciel (Software Engineer)** en France. L'étude a pour but de recueillir des données pertinentes pour orienter les **étudiants en ingénierie** vers des spécialisations adéquates après leurs diplômes, mais aussi de connaître précisément les **compétences demandées** et **opportunités d'emploi** sur l'ensemble du territoire.

Objectifs de notre projet de Collecte de Données

1. Identifier les **secteurs clés** recrutant des ingénieurs logiciels.
2. Évaluer la **distribution géographique** des offres d'emploi.
3. Analyser les **types de contrats** et les niveaux d'études demandés.
4. Déterminer la prévalence des options de **télétravail**.
5. Comprendre les gammes de **salaires** en fonction du secteur et de la région.
6. Connaître les **compétences demandées** dans l'industrie.

Méthode de Collecte des Données

Nous avons utilisé la librairie Python **Selenium** afin d'effectuer le Web Scraping, en veillant à la conformité avec les termes légaux du site. L'accent a été mis sur l'extraction de détails essentiels des annonces d'emploi, malgré le manque de structure uniforme entre les listes. Les données extraites se présentent comme tel :

- | | |
|---------------------------|----------------|
| • Nom de l'emploi | String |
| • Entreprise | String |
| • Type de contrat | String |
| • Niveau d'études demandé | int |
| • Secteur de l'entreprise | String |
| • Télétravail partiel | boolean |
| • Années d'expériences | int |

- Salaire **double**
- Compétences requises **String**
- Ville et département de l'offre **String, int**

Outils d'Analyse de Données

Afin d'analyser correctement nos données, nous avons utilisé différents outils :

- Visualisation des données selon plusieurs critères
 - **Matrice de corrélation**
 - **Forêt aléatoire**

Analyse Exhaustive des Données

Catégorie	Détails
Nombre total d'annonces	1119
Départements avec le plus d'offres	Paris (75) : 286, Hauts de Seine (92) : 102, Lyon (69) : 91
Type de contrat	CDI : 1018 (91%)
Niveau d'études demandé	Bac+5 : 803 (72%)
Années d'expérience	Non spécifié : 672, 1 an : 213, 7 ans : 234
Télétravail	Partiel dans 42% des offres
Secteur d'activité avec le plus d'offres	Informatique/ESN : 390 (35%)
Salaires	Moyenne : 43826€, Echelle : 13,5k - 90k €
Compétences demandées	Python (53%), Java (39%), PHP (34%), Git (14%), Devops (13%)

Conclusion

En conclusion, notre projet offre des perspectives précieuses sur le marché de l'emploi en ingénierie logicielle en France, révélant l'**hégémonie régionale** de la région parisienne avec des salaires plus élevés et un plus grand nombre d'opportunités, ainsi qu'une demande variée de compétences, soulignant l'importance de la **polyvalence des développeurs** et de la maîtrise des **outils cloud**.

Cette étude confirme également une **corrélation négative** entre les salaires et les compétences spécifiques, avec une demande croissante pour des **experts en langage spécialisés** pour les postes de haut niveau.

De plus, elle consolide le **statut lucratif du marché de l'emploi technologique**. Les résultats soulignent une demande significative pour les ingénieurs logiciels en France, avec des opportunités variées et une flexibilité de travail, bien que des disparités existent entre les secteurs et les régions, avec une concentration d'emplois dans les métropoles.

La majorité des offres sont des CDI, souvent avec la possibilité de télétravail, et les années d'expérience sont valorisées, même pour les jeunes ingénieurs, avec une préférence pour un niveau d'études élevé (Bac+5).

Limites de l'Analyse

Il est important de noter que cette étude présente des limites.

- Tout d'abord, l'ensemble de données a été collecté via le web scraping sur Hello Work, un site légalement autorisé pour cette pratique, mais il peut ne pas être représentatif de l'ensemble du marché de l'emploi et peut ne pas être entièrement à jour. De plus, les données, en particulier les informations salariales, peuvent être approximatives et **sujettes à négociation**.
- Deuxièmement, le **marché caché de l'emploi** limite la visibilité des offres non publiées en ligne.
- Enfin, le secteur technologique est **dynamique**, avec des compétences recherchées qui évoluent rapidement en fonction des avancées et des besoins du marché.