



Analyse des Salaires dans la Tech Française

Comprendre les facteurs d'inégalité et rendre visible la « boîte noire » des rémunérations

Les salaire du secteur tech sont-il influencées ?

Le Défi

Le secteur tech présente des variations salariales importantes et opaques

Notre mission : transformer des données officielles en insights compréhensibles

Problématique

Quels sont les facteurs qui créent les écarts de salaires dans la tech française ?



Âge & Ancienneté



Formation & Diplôme



Profession & Hiérarchie



Métiers & Carrières




Géographie



Genre

Notre Approche : Données et Outils

Source des Données

 **Dataset :**
combined_dataset.csv

Origine : INSEE DADS
2021

Secteur : Information
et Communication (NAF
11_J)

Fiabilité : Source
officielle garantie

Justification Technique

Sources

**Rémunération brute par
secteur d'activité – La structure
des salaires en 2022 | Insee**

**Élaboration des Estimations
d'emploi | Insee**

La PCS 2020 | Insee



PapaParse.js

Lecture et transformation CSV → JSON pour le navigateur



ApexCharts.js

Graphiques modernes, variés et interactifs (Treemaps, Radials)



Leaflet.js

Référence pour les cartes interactives géographiques

Nettoyage des données

Pour améliorer l'interface de nos graphes on a décidé de rassembler certaines énormément de catégorie à la base

```
df = pd.read_csv("salaries_pays_complet_mini33.csv")
print(sorted(df['job_title'].dropna().unique()))
```

```
['AI Architect', 'AI Data Engineer', 'AI Data Scientist', 'AI Developer',
```

Classification de nos 312 intitulé de métier en 20 catégorie

Pour avoir une cohérence dans nos données on vas garder seulement les données avec un certain nombre d'entrée

```
pd.set_option('display.max_rows', None)
print(df['company_location'].value_counts())
```

```
company_location
USA      79762
CAN       3204
GBR       2584
AUS        303
DEU        272
```

Normalisation de nos données au format **ISO-3** pour la carte, ainsi que **suppression des colonnes ne contenant pas suffisamment d'informations.**

```
# Liste des fichiers CSV
csv_files = [
    'T2_CHP1_AGE.csv',
    'T2_CHP1_ANC.csv',
    'T2_CHP1_DIPLOME.csv',
    'T2_CHP1_NUTS.csv',
    'T2_CHP1_PROF.csv',
    'T2_CHP1_TAILLENT.csv'
]
```

```
# Ajouter une colonne pour identifier le type de données
if 'TR_AGE' in df.columns:
    df['data_type'] = 'age'
    df['category'] = df['TR_AGE']
elif 'TR_ANC' in df.columns:
    df['data_type'] = 'anciennete'
    df['category'] = df['TR_ANC']
elif 'B25' in df.columns:
    df['data_type'] = 'diplome'
    df['category'] = df['B25']
elif 'REG' in df.columns:
    df['data_type'] = 'region'
    df['category'] = df['REG']
elif 'CITP' in df.columns:
    df['data_type'] = 'profession'
    df['category'] = df['CITP']
elif 'A12' in df.columns:
    df['data_type'] = 'taille_entreprise'
    df['category'] = df['A12']
```

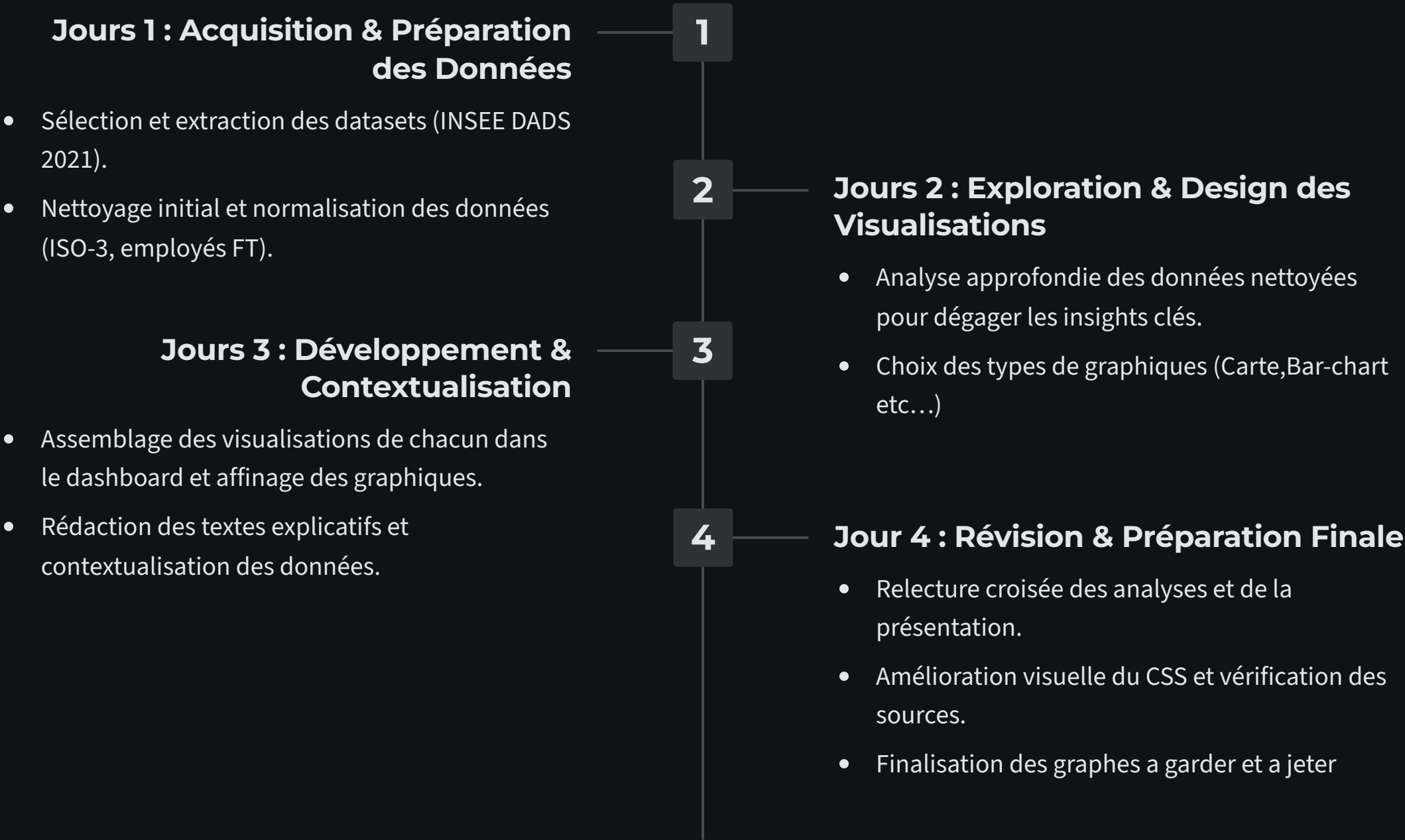
Nos datas combinés son issus de ces csv téléchargé sur le site de l'INSEE.

On a restructuré pour avoir les colonnes dont on avait besoin


Utilisation de lib comme Pandas pour manipuler les dataframes et de PyCountry pour le ISO 3 et d'un code github pour les coordonnées des régions. raw.githubusercontent.com/gregoiredavid/france-geojson/master/regions.geojson

Organisation et Répartition des Tâches

Notre semaine de travail a été structurée en phases distinctes pour optimiser l'analyse et la présentation des données, avec une répartition claire des responsabilités au sein de l'équipe.




Notre Équipe et Leurs Contributions




Japhet

Gestion du CSS et de l'affichage des graphes dans le Dash-Board




Lyes

Nettoyage des Datasets



Selim

Contextualisation des graphes



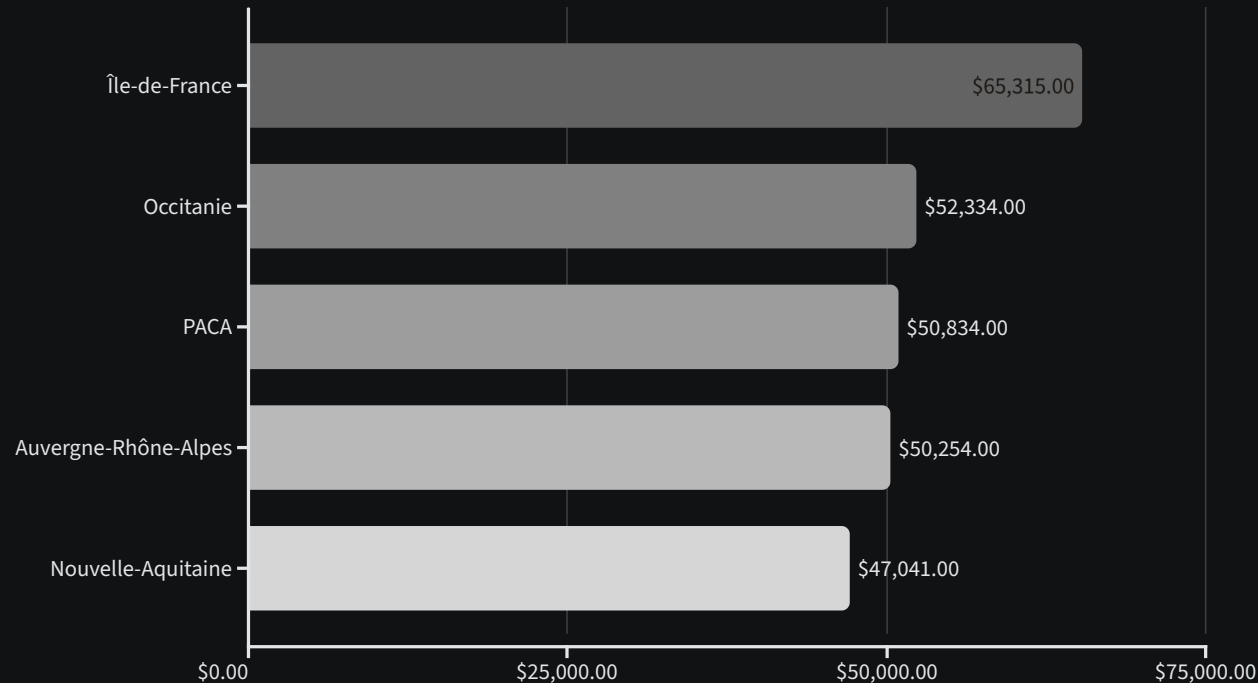
Derick

Recherche et exploitation des données INSEE pour les graphes

L'Impact de la Géographie

 L'Anomalie Française

Concentration extrême : L'Île-de-France domine sur tous les plans



+38%

Écart IdF vs Province

65%

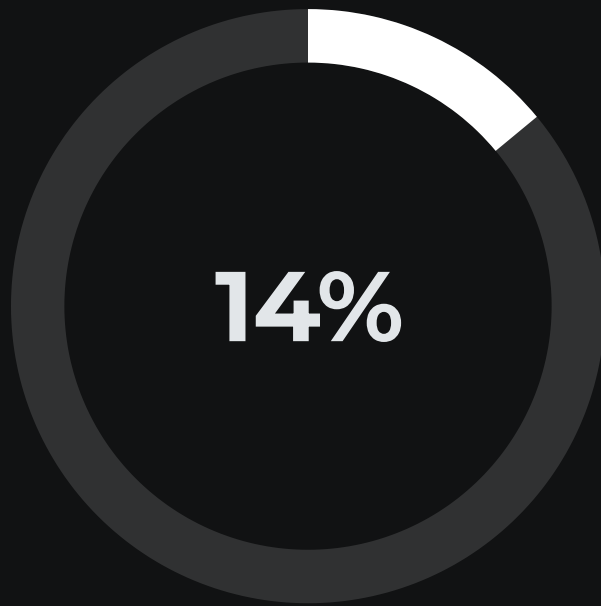
Part des postes tech

Concentrés en Île-de-France

Double effet : L'Île-de-France cumule les salaires les plus élevés *et* le plus grand nombre de postes — une anomalie structurelle du marché français

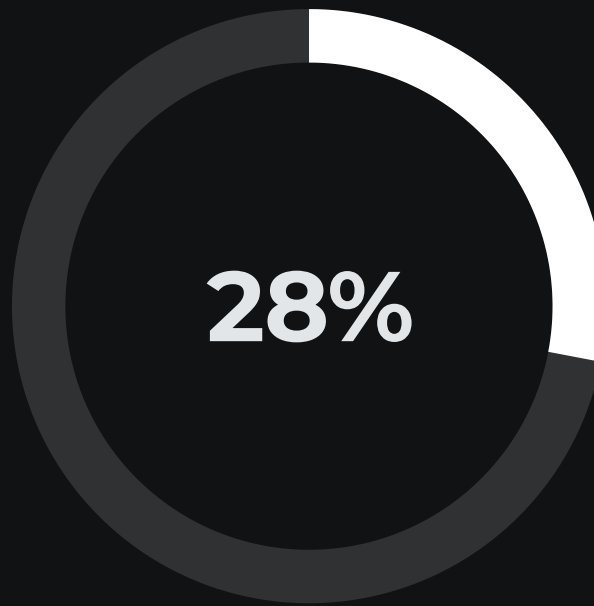
Un Écart Structurel et Persistant

Inégalités de Genre dans la Tech



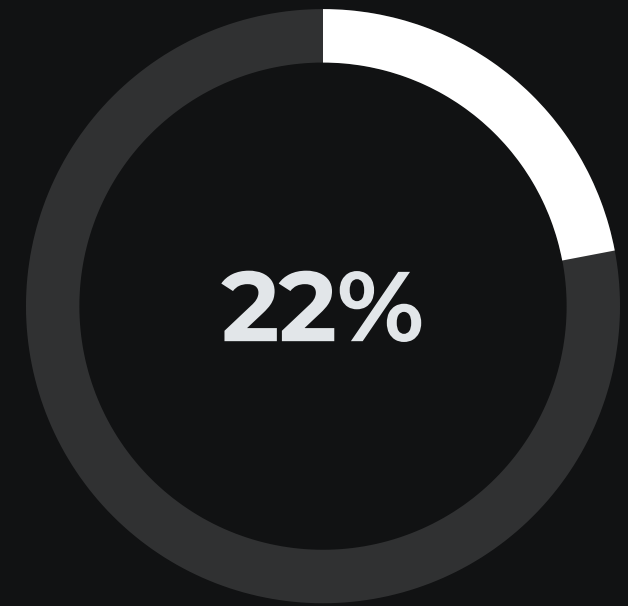
Écart salarial moyen

Entre hommes et femmes dans la tech



Part des femmes

Dans le secteur tech français



Écart chez les cadres

L'inégalité s'aggrave avec la hiérarchie

L'écart se creuse

Avec l'ancienneté :

- 0-5 ans : écart de 8%
- 5-15 ans : écart de 12%
- 15+ ans : écart de 18%

Où sont les femmes ?

Répartition hiérarchique :

- Cadres : 32% de femmes
- Intermédiaires : 28%
- Employés : 52%

Origine structurelle

Sous-représentation en haut de l'échelle + écart à poste égal

Hiérarchie des Facteurs

Réponse à la Problématique

Quels facteurs créent les écarts de salaires dans la tech française ?



La Géographie

L'anomalie Île-de-France : +38% de salaire, 65% des postes



La Hiérarchie

Cadres vs Autres : divergence progressive des trajectoires



Formation & Âge

Facteurs secondaires mais déterminants : 72% Bac+2 minimum



Le Genre

Un écart structurel de 14% qui s'aggrave avec l'ancienneté



Les Outils de Notre Équipe

Voici les logiciels que nous utilisons quotidiennement pour le développement, le partage de code et la communication.



VS Code

Notre environnement de développement intégré pour l'écriture de code propre et efficace.



GitHub

Plateforme indispensable pour le contrôle de version, la collaboration et la gestion des projets.



WhatsApp

Utilisé pour la communication rapide et informelle au sein de l'équipe.



Teams

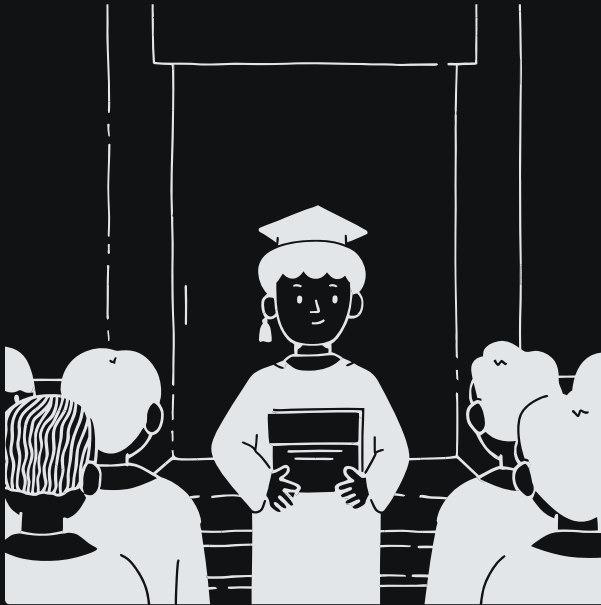
Outil centralisé pour la gestion des tâches, le suivi des projets et la coordination d'équipe.



Snapchat

Pour des échanges plus créatifs et des annonces rapides de l'équipe.

Mission Accomplie



✓ Nos Objectifs Atteints

Données rendues lisibles

6 dimensions analysées avec rigueur

Insights identifiés

Hiérarchie claire des facteurs d'inégalité

Visualisations adaptées

Chaque graphique justifié par son usage

📄 **Méthodologie validée** : Source officielle INSEE + stack technique moderne (PapaParse, ApexCharts, Leaflet, Plotly)

Questions ?

Merci de votre attention

Demo

