

Projet SAE : Analyse de Données, Reporting et Datavisualisation

Aimad Hamdaoui
hamdaoui.cc

22 juin 2025

Résumé

Ce document présente la synthèse d'un projet réalisé dans le cadre d'une Situation d'Apprentissage en Entreprise (SAE). L'objectif était d'analyser l'influence de l'âge sur la capacité à acquérir un logement, en utilisant des jeux de données fournis par l'établissement. L'analyse se concentre sur les difficultés rencontrées par les jeunes générations par rapport aux plus anciennes, en explorant des facteurs comme le type de contrat de travail et l'évolution du coût de l'immobilier.

1 Objectifs et Problématique

Le point de départ de ce projet était de répondre à la problématique suivante : « *Comment l'âge influence-t-il la possibilité d'acquérir un certain statut d'occupation ?* ». Pour ce faire, nous avons analysé des jeux de données portant sur le logement et l'emploi.

L'analyse visait spécifiquement à comprendre et quantifier les difficultés accrues que rencontre un jeune pour accéder à un logement, en comparaison avec une personne plus âgée. Cela impliquait une compréhension fine des données, une analyse ciblée et la production de visualisations pertinentes pour illustrer nos conclusions.

Une première analyse exploratoire a permis de visualiser la répartition des types de logement (maisons, appartements) en fonction des différentes classes d'âge dans le département du Vaucluse, comme l'illustre la figure 1.

2 Compétences Techniques et Méthodologie

Pour mener à bien ce projet, nous avons mobilisé plusieurs compétences techniques. Les données brutes ont d'abord été structurées et stockées dans une base de données SQL pour en faciliter la manipulation. L'ensemble de l'analyse et de la visualisation a ensuite été réalisé à l'aide du langage de programmation **R** et de ses bibliothèques dédiées, notamment **ggplot2**.

Une des hypothèses clés était que la précarité de l'emploi joue un rôle majeur. Nous avons donc utilisé un jeu de données de l'INSEE pour isoler les individus en Contrat à Durée Déterminée (CDD) de ceux en Contrat à Durée Indéterminée (CDI). Le code ci-dessous montre un extrait des étapes de traitement des données.

```
data <- read.csv("ID_ACT2B_2020.csv", sep = ";")
filtered_data <- subset(data, substr(CODEGE, 1, 2) == "15")
write.csv(filtered_data, "fait4.csv", row.names = FALSE)
filtered_data <- filtered_data %>% filter(EMPL %in% c(15, 16))
filtered_data <- filtered_data %>% mutate(EMPL = recode(EMPL, '15' = "COD"))
```

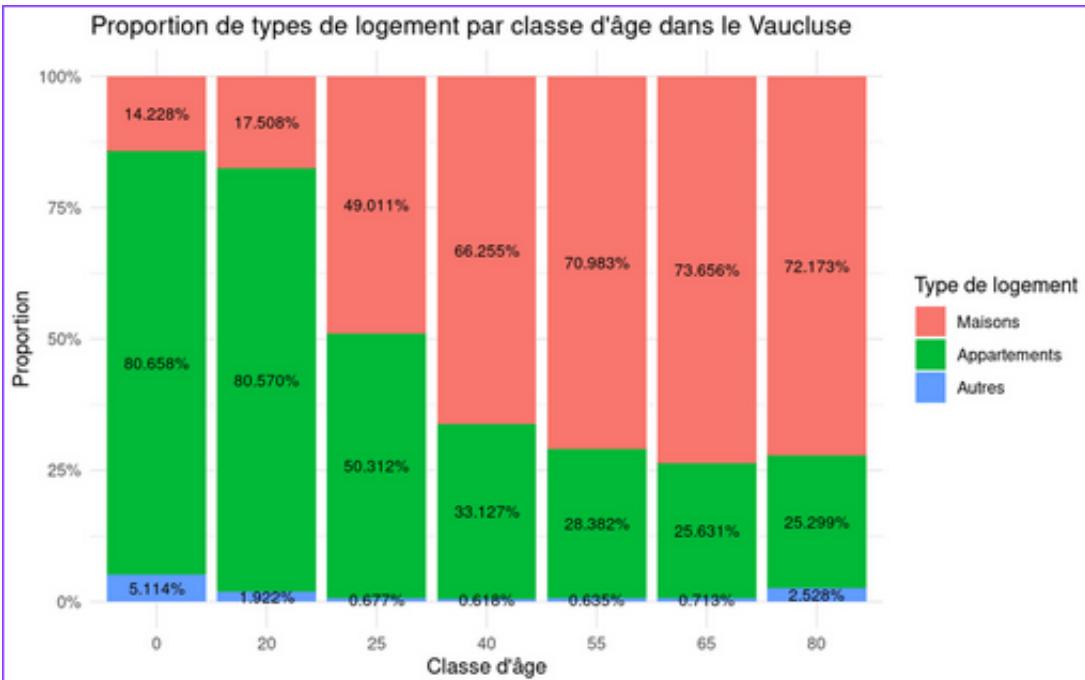


FIGURE 1 – Proportion des types de logement par classe d’âge dans le Vaucluse. On observe que la part des appartements (en vert) est majoritaire chez les plus jeunes, tandis que celle des maisons (en rouge) augmente significativement avec l’âge.

L’analyse a révélé une proportion significativement plus élevée de jeunes de moins de 40 ans en CDD (voir Figure 2). Cette instabilité professionnelle explique en partie pourquoi un jeune peut avoir plus de difficultés à obtenir un prêt ou à garantir un loyer sur le long terme.

3 Résultats et Conclusion

Notre analyse nous amène à conclure qu’un jeune rencontre effectivement de plus grandes difficultés pour entretenir un logement stable. Cette situation s’explique par une combinaison de facteurs :

- **La précarité de l’emploi** : Comme démontré, les jeunes générations (génération 2000 et suivantes) sont davantage concernées par les contrats courts (CDD).
- **Le contexte économique et historique** : Les générations précédentes (avant 1980) ont bénéficié de conditions différentes. Par exemple, la loi de 1948 a longtemps bloqué les loyers, protégeant les locataires en place.
- **L’explosion du coût de l’immobilier** : Pour appuyer nos propos, nous avons utilisé des ressources externes, comme les indices du coût de la construction.

Le tableau 1 met en perspective l’évolution drastique du coût de la construction entre 1970 et aujourd’hui. Une maison dont la construction coûtait environ 50 000 € en 1970 (valeur convertie) en coûterait près de 480 000 € en 2024, une augmentation sans commune mesure avec celle des salaires.

Ce projet nous a permis de développer notre capacité à utiliser des études scientifiques et des données externes pour contextualiser et renforcer une analyse. Il a également souligné l’importance de la gestion rigoureuse des données, de l’optimisation des requêtes et de la conformité des informations présentées.

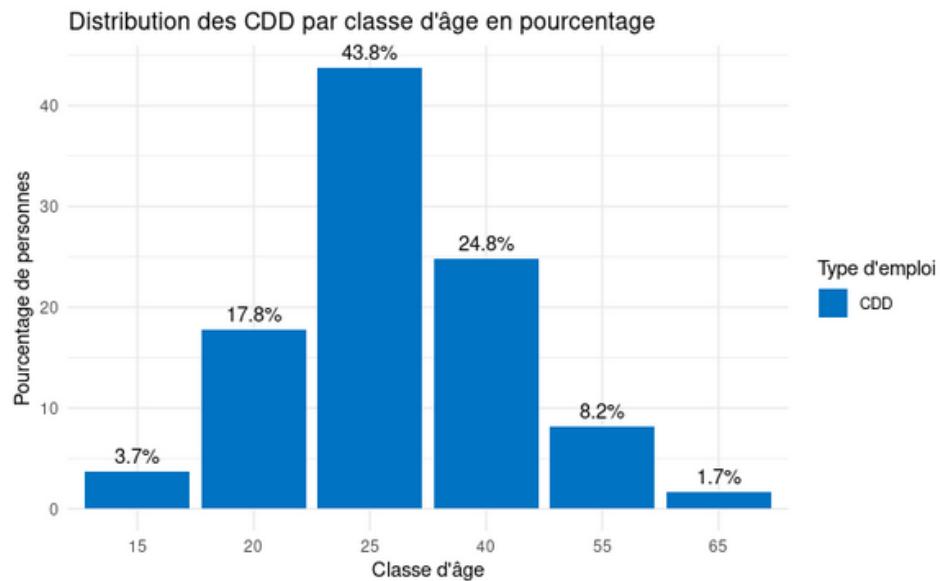


FIGURE 2 – Distribution en pourcentage des emplois en CDD par classe d'âge. Le pic chez les 25-40 ans (43.8%) met en évidence la précarité de l'emploi pour cette tranche de la population active.

TABLE 1 – Comparaison de l'indice du coût de la construction (ICC) entre 1970 et 2023.

Information	Année 2023	Année 1970
Trimestre	T4	T4
Valeur de l'indice	2 162	226
Parution au J.O.	04/04/2024	26/02/1971

Source : Données basées sur l'indice du coût de la construction.