Rapport d’étude

Ilan FUNDENBERGER et Antoine HOUY

2024-10-13

Dans ce rapport, nous allons voir l’importance d’avoir une bonne étiquette DPE (notamment sur l’aspect financier). Nous présenteront d’abord les communes possèdant les meilleures logement (selon l’étiquette DPE) et celles qui ont le plus de passoire thermique. Ensuite, nous verrons l’importance d’avoir une bonne étiquette DPE et pour finir ce qui impactent sur la consommation d’energie et comment consommer moins.

df = read.csv("C:/Etudes/BUT SD/Année 2/R/Projet\_DPE\_logement/test.csv", sep = ",", dec = ".")

Voici donc les communes possèdant le plus de passoires thermiques :

### Histogramme de la distribution des étiquettes DPE G par Code postal

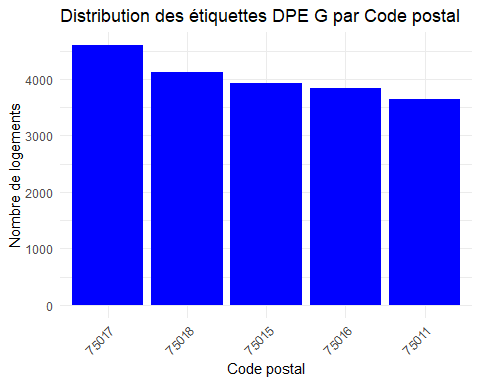
# Chargement des bibliothèques nécessaires  
library(ggplot2)  
library(dplyr)

##   
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:stats':  
##   
## filter, lag

## Les objets suivants sont masqués depuis 'package:base':  
##   
## intersect, setdiff, setequal, union

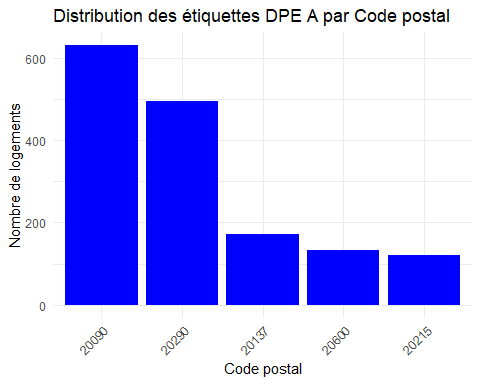
etiquette <- "G"  
  
# Filtrage des données pour l'étiquette DPE G  
filtered\_data\_G <- df %>%  
 dplyr::filter(Etiquette\_DPE == etiquette,) %>%  
 group\_by(Code\_postal\_.BAN.) %>%  
 summarise(count = n()) %>%  
 arrange(desc(count))  
  
top\_5\_G <- filtered\_data\_G[1:5, ]  
top\_5\_G <- na.omit(top\_5\_G)  
  
  
# Si les données sont disponibles, généreration du graphique  
if (nrow(top\_5\_G) > 0) {  
 ggplot(top\_5\_G, aes(x = reorder(Code\_postal\_.BAN., -count), y = count)) +  
 geom\_bar(stat = "identity", fill = "blue") +  
 labs(title = "Distribution des étiquettes DPE G par Code postal",  
 x = "Code postal",   
 y = "Nombre de logements") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))  
} else {  
 cat("Aucune donnée disponible pour l'étiquette DPE G.")}



Et voici, les communes possèdant le plus de logements avec la meilleure étiquette DPE :

### Histogramme de la distribution des étiquettes DPE A par Code postal

# Chargement des bibliothèques nécessaires  
library(ggplot2)  
library(dplyr)  
  
etiquette <- "A"  
  
# Filtrage des données pour l'étiquette DPE A  
filtered\_data\_A <- df %>%  
 dplyr::filter(Etiquette\_DPE == etiquette,) %>%  
 group\_by(Code\_postal\_.BAN.) %>%  
 summarise(count = n()) %>%  
 arrange(desc(count))  
  
top\_5\_A <- filtered\_data\_A[1:5, ]  
top\_5\_A <- na.omit(top\_5\_A)  
  
  
# Si les données sont disponibles, généreration du graphique  
if (nrow(top\_5\_A) > 0) {  
 ggplot(top\_5\_A, aes(x = reorder(Code\_postal\_.BAN., -count), y = count)) +  
 geom\_bar(stat = "identity", fill = "blue") +  
 labs(title = "Distribution des étiquettes DPE A par Code postal",  
 x = "Code postal",   
 y = "Nombre de logements") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))  
} else {  
 cat("Aucune donnée disponible pour l'étiquette DPE A.")}



Ici, nous allons voir en moyenne la différence de coût en enérgie par ans qu’il peut y avoir entre les différentes étiquettes DPE.

library(dplyr)  
  
# Calcul de la moyenne des coûts par type de logement  
moyennes <- df %>%  
 group\_by(Etiquette\_DPE) %>% # Regrouper par la variable 'Logement'  
 summarise(Moyenne\_Cout = mean(Coût\_total\_5\_usages, na.rm = TRUE)) # Calculer la moyenne  
  
print(moyennes)

## # A tibble: 7 × 2  
## Etiquette\_DPE Moyenne\_Cout  
## <chr> <dbl>  
## 1 A 511.  
## 2 B 874.  
## 3 C 1236.  
## 4 D 1647.  
## 5 E 1820.  
## 6 F 1809.  
## 7 G 2113.

On peut voir qu’il y a une différence de coût énorme entre un logement avec une étiquette A et G (quasiment 1600 €).

### Le choix du type d’énergie principale de chauffage

library(dplyr)  
  
# Calcul de la moyenne des coûts par type d'e logement'énergie principale de chauffage  
moyennes <- df %>%  
 group\_by(Type\_énergie\_principale\_chauffage) %>%   
 summarise(Moyenne\_Cout = mean(Coût\_total\_5\_usages, na.rm = TRUE))   
  
print(moyennes)

## # A tibble: 15 × 2  
## Type\_énergie\_principale\_chauffage Moyenne\_Cout  
## <chr> <dbl>  
## 1 Bois – Bûches 1732.  
## 2 Bois – Granulés (pellets) ou briquettes 1469.  
## 3 Bois – Plaquettes d’industrie 3927.  
## 4 Bois – Plaquettes forestières 1001.  
## 5 Butane 3777.  
## 6 Charbon 2851.  
## 7 Fioul domestique 2693.  
## 8 GPL 4315.  
## 9 Gaz naturel 1655.  
## 10 Propane 2247.  
## 11 Réseau de Chauffage urbain 1992.  
## 12 Réseau de Froid Urbain 1306.  
## 13 Électricité 1295.  
## 14 Électricité d'origine renouvelable utilisée dans le bâtiment 2009.  
## 15 <NA> 2035.

Nous pouvons voir avoir le tableau ci-dessus l’importance de bien choisir sont type d’énergie principale pour le chauffage, le bois en plaquettes forestières étant le moins chères et le GPL le plus chère.

Nous allons maintenant vérifier que l’isolation est bien la clé pour avoir un cout à payer du à la consommation énergétique moins élevé. Pour cela, nous allons utiliser le test de krustal avec les hypothèses suivantes :

H0 : Il n’y a pas de différence dans la distribution de la variable continue (dans notre cas, le coût énergétique) entre les différents niveaux de la variable qualitative (dans notre cas, la qualité de l’isolation).

H1 : Il existe au moins une différence significative dans la distribution de la variable continue entre les groupes

# Test de Kruskal-Wallis  
kruskal\_test\_enveloppe <- kruskal.test(Coût\_total\_5\_usages ~ Qualité\_isolation\_enveloppe, data = df)  
kruskal\_test\_murs <- kruskal.test(Coût\_total\_5\_usages ~ Qualité\_isolation\_murs, data = df)  
kruskal\_test\_menuiseries <- kruskal.test(Coût\_total\_5\_usages ~ Qualité\_isolation\_menuiseries, data = df)  
kruskal\_test\_plancher\_bas <- kruskal.test(Coût\_total\_5\_usages ~ Qualité\_isolation\_plancher\_bas, data = df)  
  
  
# Affichage des résultats  
print(kruskal\_test\_enveloppe)

##   
## Kruskal-Wallis rank sum test  
##   
## data: Coût\_total\_5\_usages by Qualité\_isolation\_enveloppe  
## Kruskal-Wallis chi-squared = 58981, df = 3, p-value < 2.2e-16

print(kruskal\_test\_murs)

##   
## Kruskal-Wallis rank sum test  
##   
## data: Coût\_total\_5\_usages by Qualité\_isolation\_murs  
## Kruskal-Wallis chi-squared = 99485, df = 3, p-value < 2.2e-16

print(kruskal\_test\_menuiseries)

##   
## Kruskal-Wallis rank sum test  
##   
## data: Coût\_total\_5\_usages by Qualité\_isolation\_menuiseries  
## Kruskal-Wallis chi-squared = 37561, df = 3, p-value < 2.2e-16

print(kruskal\_test\_plancher\_bas)

##   
## Kruskal-Wallis rank sum test  
##   
## data: Coût\_total\_5\_usages by Qualité\_isolation\_plancher\_bas  
## Kruskal-Wallis chi-squared = 29391, df = 3, p-value < 2.2e-16

Étant donné que les différentes p-value sont beaucoup inférieure au seuil de signification habituel de 0.05, nous pouvez rejeter l’hypothèse nulle (H0), qui stipule qu’il n’y a pas de différence significative entre les groupes.

On conclue qu’il existe des différences significatives dans le coût énergétique entre au moins deux des niveaux des différentes qualité d’isolation que nous avez testés. Cela signifie que la qualité de l’isolation a un impact mesurable sur le coût énergétique.

En pratique, cela signifie que les logements avec une meilleure isolation (comme ceux classés “très bonne”) tendent à avoir des coûts énergétiques inférieurs par rapport à ceux avec une isolation “insuffisante”. Les résultats confirment donc bien que des améliorations de l’isolation pourraient conduire à des économies sur les coûts d’énergie.