

Facteurs influençants la prise des transports en commun

RACHIDI Mustapha & SAUNIER Florent & SAADALLAH Malek

Janvier 2023

Introduction

Ce projet se base sur des données récoltées en 2010 dans la région Grenobloise. L'étude a pour but de déterminer les facteurs influençant la prise des transports en commun. Pour cela nous nous sommes pris comme limites : le réseau Mtag qui comprend les bus qualifiés de "ville" (Nous n'avons pas pris en compte les bus régionaux comme par exemple le bus Grenoble - Chamrousse) et le réseau du tramway dont les lignes depuis 2010 ont été augmentées.

Articles de la littérature

Familiarisation avec la base de données

La base de données contient 30 702 lignes et 116 colonnes ce qui correspond à nos variables, on peut la qualifier de base de données "moyenne" mais qui saura nous occuper. Concernant le nombre de valeurs manquantes, toutes variables confondues nous avons 971 658 valeurs manquantes soit 27.3% de notre base de données. De plus, 0% des lignes ont toutes leurs valeurs et c'est 21% des colonnes qui n'ont pas de valeurs manquantes. Il peut être intéressant de voir où sont les valeurs manquantes.

Visualisation valeurs manquantes titre à changer peut être

En annexe, quelques graphiques permettant de visualiser quelles variables ont le plus de valeurs manquantes. Ces graphiques nous permettront d'adopter un regard critique sur les variables que nous choisirons par la suite. Cependant, on peut établir quelques critères avec r : ration de valeurs manquantes dans la colonne.

Bon : $r \leq 5\%$ Moyen : $5\% < r \leq 20\%$ Mauvais $20\% < r \leq 45\%$ Très mauvais : $r > 45\%$

Plusieurs variables ont entre 80% 99% de valeurs manquantes

Préparation de la base de données

```
sample <- sample(c(TRUE,FALSE),size=nrow(DB_projet_full),replace=TRUE,prob=c(0.7,0.3))
DB_reduced <-DB_projet_full[!sample,] #base de donnée test , plus petite pour faire des opérations dessus

DB_reduced<-DB_reduced%>%mutate(frecqtcu=ifelse(frecqtcu<=3,1,0)) #On simplifie notre variable explicative
```

Variables du projet

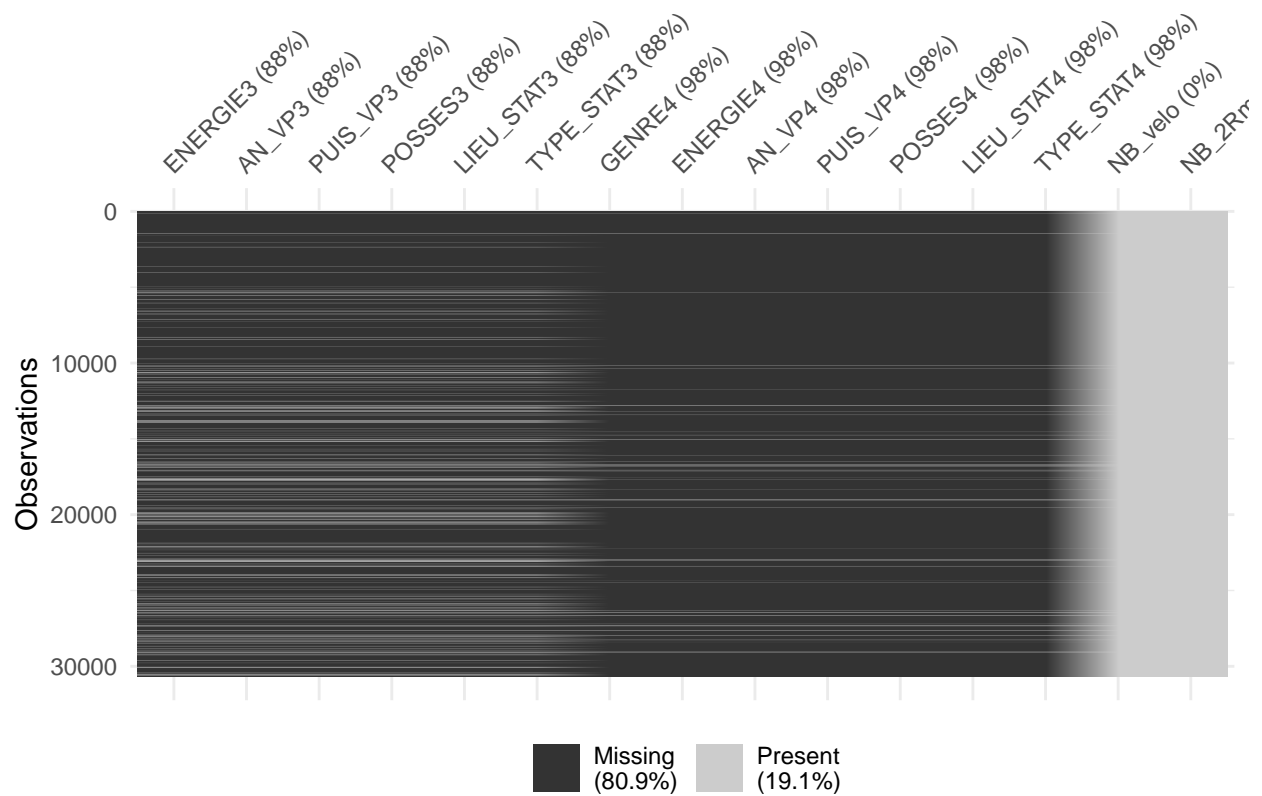
frecqtcu : Variable d'intérêt (Y) catégorielle qui indique la fréquence d'utilisation des transports en communs chez une personne. Elle prend les valeurs :

- 1 : Utilisation des transports en commun tous les jours
- 2 : Utilisation des transports en commun au moins deux fois par semaine
- 3 : Utilisation des transports en commun au moins deux fois par mois
- 4: Utilisation des transports en commun très rare
- 5 : Utilisation des transports en commun inexistante

```
#code de création de la variable frecqtcu , juste reprendre ce qui a été dit en haut
```

Annexes

```
data_2<-DB_projet_full[,c(30:44)]
vis_miss(
  data_2,
  cluster = FALSE,
  sort_miss = FALSE,
  show_perc = TRUE,
  show_perc_col = TRUE,
  large_data_size = 9e+06,
  warn_large_data = TRUE
)
```



Listes variables à plus de 80% de valeurs manquantes

-motoracc -situveil -STAT_TRAV -TYPE_STAT4 -LIEU_STAT4 -POSSES4 -PUIS_VP4 -AN_VP4 -
 ENERGIE4 -GENRE4 -TYPE_STAT3 -LIEU_STAT3 -POSSES3 -PUIS_VP3 -AN_VP3 -ENERGIE3 -
 motdeacc -nbarret -abonpage