

# Données

Pivot et jointure

---

Véronique Tremblay

# Retour sur les étapes d'un projet d'analyse de données

---

1. Définition des objectifs
2. Données
  - Inventaire et qualité
  - Constitution et nettoyage
  - Exploration et traitement préliminaire
3. Élaboration et validation des modèles
4. Mise en oeuvre
5. Suivi de la performance et amélioration

## Pivot

---

# Pivoter

## Format Long

Client	Âge	Produit	Montant
A	32	Auto	350
A	32	Habitation	1200
B	55	Auto	200
C	55	Habitation	900

## Format Large

Client	Âge	Auto	Habitation
A	32	350	1200
B	55	200	
C	55		900

## Pivot: long vers large

Client	Âge	Produit	Montant
A	32	Auto	350
A	32	Habitation	1200
B	55	Auto	200
C	55	Habitation	900

```
pivot_wider(long,  
            names_from = Produit,  
            values_from = Montant) %>%  
kable()
```

## Pivot: large vers long

Client	Âge	Auto	Habitation
A	32	350	1200
B	55	200	
C	55		900

```
pivot_longer(large,  
             cols = c('Auto', 'Habitation'),  
             names_to = "Produit",  
             values_to = "Montant",  
             values_drop_na = TRUE)
```

## Les jointures

---

# Joindre deux tables

## Auto

Client	Âge	Montant
A	32	350
B	55	200

## Habitation

Client	Âge	Montant
A	32	1200
C	55	900



# Jointure externe complète (full\_join)

## Auto

Client	Âge	Montant
A	32	350
B	55	200

## Habitation

Client	Âge	Montant
A	32	1200
C	55	900

## Jointure complète

Client	Âge	Montant.Auto	Montant.Habitation
A	32	350	1200
B	55	200	
C	55		900

## Jointure à gauche (left\_join)

### Auto

Client	Âge	Montant
A	32	350
B	55	200

### Habitation

Client	Âge	Montant
A	32	1200
C	55	900

### Jointure à gauche

Client	Âge	Montant.Auto	Montant.Habitation
A	32	350	1200
B	55	200	

## Jointure à droite (right\_join)

### Auto

Client	Âge	Montant
A	32	350
B	55	200

### Habitation

Client	Âge	Montant
A	32	1200
C	55	900

### Jointure à droite

Client	Âge	Montant.Auto	Montant.Habitation
A	32	350	1200
C	55		900

## Jointure interne (inner\_join)

### Auto

Client	Âge	Montant
A	32	350
B	55	200

### Habitation

Client	Âge	Montant
A	32	1200
C	55	900

### Jointure interne

Client	Âge	Montant.Auto	Montant.Habitation
A	32	350	1200

```
inter <- inner_join(Auto, Habitation,  
  by = c('Client', 'Âge'),  
  suffix = c(".Auto", ".Habitation")  
)
```

Client	Âge	Montant.Auto	Montant.Habitation
A	32	350	1200

- Pivot

S'assurer qu'une ligne = une observation

- Jointure

Choisir le type de jointure en fonction de la population cible