

R Notebook

```
library(FactoMineR)
library(MASS)
```

On voit que les deux premières colonnes sont inutiles pour nos analyses, on va s'occuper de les supprimer

```
rm(list=ls())
data <- read.csv("data/train.csv")
head(data)
```

```
##      X      id Gender      Customer.Type Age Type.of.Travel      Class
## 1 0 70172   Male    Loyal Customer   13 Personal Travel Eco Plus
## 2 1  5047   Male disloyal Customer   25 Business travel Business
## 3 2 110028 Female    Loyal Customer   26 Business travel Business
## 4 3  24026 Female    Loyal Customer   25 Business travel Business
## 5 4 119299   Male    Loyal Customer   61 Business travel Business
## 6 5 111157 Female    Loyal Customer   26 Personal Travel      Eco
##      Flight.Distance Inflight.wifi.service Departure.Arrival.time.convenient
## 1                460                      3                             4
## 2                235                      3                             2
## 3                1142                     2                             2
## 4                562                      2                             5
## 5                214                      3                             3
## 6                1180                     3                             4
##      Ease.of.Online.booking Gate.location Food.and.drink Online.boarding
## 1                      3                1                5                3
## 2                      3                3                1                3
## 3                      2                2                5                5
## 4                      5                5                2                2
## 5                      3                3                4                5
## 6                      2                1                1                2
##      Seat.comfort Inflight.entertainment On.board.service Leg.room.service
## 1                5                      5                4                3
## 2                1                      1                1                5
## 3                5                      5                4                3
## 4                2                      2                2                5
## 5                5                      3                3                4
## 6                1                      1                3                4
##      Baggage.handling Checkin.service Inflight.service Cleanliness
## 1                4                4                5                5
## 2                3                1                4                1
## 3                4                4                4                5
## 4                3                1                4                2
## 5                4                3                3                3
## 6                4                4                4                1
##      Departure.Delay.in.Minutes Arrival.Delay.in.Minutes      satisfaction
```

```
## 1      25      18 neutral or dissatisfied
## 2      1      6 neutral or dissatisfied
## 3      0      0      satisfied
## 4     11      9 neutral or dissatisfied
## 5      0      0      satisfied
## 6      0      0 neutral or dissatisfied
```

```
data <- data[,c(-1,-2)] # On retire les colonnes inutiles
row.names(data) <- paste("id_", sep="", 1:dim(data)[1]) # On modifie le nom des colonnes pour le style
str(data)
```

```
## 'data.frame': 103904 obs. of 23 variables:
## $ Gender : chr "Male" "Male" "Female" "Female" ...
## $ Customer.Type : chr "Loyal Customer" "disloyal Customer" "Loyal Customer" "Lo
## $ Age : int 13 25 26 25 61 26 47 52 41 20 ...
## $ Type.of.Travel : chr "Personal Travel" "Business travel" "Business travel" "Bus
## $ Class : chr "Eco Plus" "Business" "Business" "Business" ...
## $ Flight.Distance : int 460 235 1142 562 214 1180 1276 2035 853 1061 ...
## $ Inflight.wifi.service : int 3 3 2 2 3 3 2 4 1 3 ...
## $ Departure.Arrival.time.convenient: int 4 2 2 5 3 4 4 3 2 3 ...
## $ Ease.of.Online.booking : int 3 3 2 5 3 2 2 4 2 3 ...
## $ Gate.location : int 1 3 2 5 3 1 3 4 2 4 ...
## $ Food.and.drink : int 5 1 5 2 4 1 2 5 4 2 ...
## $ Online.boarding : int 3 3 5 2 5 2 2 5 3 3 ...
## $ Seat.comfort : int 5 1 5 2 5 1 2 5 3 3 ...
## $ Inflight.entertainment : int 5 1 5 2 3 1 2 5 1 2 ...
## $ On.board.service : int 4 1 4 2 3 3 3 5 1 2 ...
## $ Leg.room.service : int 3 5 3 5 4 4 3 5 2 3 ...
## $ Baggage.handling : int 4 3 4 3 4 4 4 5 1 4 ...
## $ Checkin.service : int 4 1 4 1 3 4 3 4 4 4 ...
## $ Inflight.service : int 5 4 4 4 3 4 5 5 1 3 ...
## $ Cleanliness : int 5 1 5 2 3 1 2 4 2 2 ...
## $ Departure.Delay.in.Minutes : int 25 1 0 11 0 0 9 4 0 0 ...
## $ Arrival.Delay.in.Minutes : num 18 6 0 9 0 0 23 0 0 0 ...
## $ satisfaction : chr "neutral or dissatisfied" "neutral or dissatisfied" "sati
```

```
head(data)
```

```
##      Gender      Customer.Type Age Type.of.Travel      Class Flight.Distance
## id_1   Male      Loyal Customer   13 Personal Travel Eco Plus           460
## id_2   Male disloyal Customer   25 Business travel Business           235
## id_3 Female      Loyal Customer   26 Business travel Business           1142
## id_4 Female      Loyal Customer   25 Business travel Business           562
## id_5   Male      Loyal Customer   61 Business travel Business           214
## id_6 Female      Loyal Customer   26 Personal Travel      Eco           1180
##      Inflight.wifi.service Departure.Arrival.time.convenient
## id_1                      3                               4
## id_2                      3                               2
## id_3                      2                               2
## id_4                      2                               5
## id_5                      3                               3
## id_6                      3                               4
##      Ease.of.Online.booking Gate.location Food.and.drink Online.boarding
```

| | | | | |
|---------|----------------------------|--------------------------|------------------|------------------|
| ## id_1 | 3 | 1 | 5 | 3 |
| ## id_2 | 3 | 3 | 1 | 3 |
| ## id_3 | 2 | 2 | 5 | 5 |
| ## id_4 | 5 | 5 | 2 | 2 |
| ## id_5 | 3 | 3 | 4 | 5 |
| ## id_6 | 2 | 1 | 1 | 2 |
| ## | Seat.comfort | Inflight.entertainment | On.board.service | Leg.room.service |
| ## id_1 | 5 | 5 | 4 | 3 |
| ## id_2 | 1 | 1 | 1 | 5 |
| ## id_3 | 5 | 5 | 4 | 3 |
| ## id_4 | 2 | 2 | 2 | 5 |
| ## id_5 | 5 | 3 | 3 | 4 |
| ## id_6 | 1 | 1 | 3 | 4 |
| ## | Baggage.handling | Checkin.service | Inflight.service | Cleanliness |
| ## id_1 | 4 | 4 | 5 | 5 |
| ## id_2 | 3 | 1 | 4 | 1 |
| ## id_3 | 4 | 4 | 4 | 5 |
| ## id_4 | 3 | 1 | 4 | 2 |
| ## id_5 | 4 | 3 | 3 | 3 |
| ## id_6 | 4 | 4 | 4 | 1 |
| ## | Departure.Delay.in.Minutes | Arrival.Delay.in.Minutes | | |
| ## id_1 | 25 | 18 | | |
| ## id_2 | 1 | 6 | | |
| ## id_3 | 0 | 0 | | |
| ## id_4 | 11 | 9 | | |
| ## id_5 | 0 | 0 | | |
| ## id_6 | 0 | 0 | | |
| ## | satisfaction | | | |
| ## id_1 | neutral or dissatisfied | | | |
| ## id_2 | neutral or dissatisfied | | | |
| ## id_3 | satisfied | | | |
| ## id_4 | neutral or dissatisfied | | | |
| ## id_5 | satisfied | | | |
| ## id_6 | neutral or dissatisfied | | | |