

Année Universitaire 2022 – 2023

Université Sultan Moulay Slimane,

École Nationale des Sciences Appliquées – Beni Mellal

**Filière** : Transformation digitale industrielle

**Module** : Analyse de Données et BIG DATA

**semestre** :4

**Rapport mini-projet**

**Analyse discriminatoire**



Réalisé par :

**LAMHAMDI Anouar**

**OUHAGUA Omar**

Prof

**DR. Mohamed GOUSKIR**

Table des matières

[**I.** **Introduction :** 2](#_Toc134283662)

[**II.** **L'analyse discriminatoire** 2](#_Toc134283663)

[**1.** **Définition Objectifs de l'analyse discriminatoire** 2](#_Toc134283664)

[**2.** **Les différentes techniques d'analyse discriminatoire** 2](#_Toc134283665)

[**3.** **Les hypothèses et les prérequis pour réaliser une analyse discriminatoire** 2](#_Toc134283666)

[**III.** **Etude de cas [Entreprise BETA]** 2](#_Toc134283667)

[**1.** **contexte :** 2](#_Toc134283668)

[**2.** **Problématique à résoudre :** 2](#_Toc134283669)

[**3.** **Méthodologie :** 2](#_Toc134283670)

[**4.** **Description des données utilisées dans le mini-projet** 2](#_Toc134283671)

[**5.** **Préparation des données :** 4](#_Toc134283672)

[a) Importation des données 4](#_Toc134283673)

[b) Transformation des variables qualitatives en variables indicatrices 6](#_Toc134283674)

[c) Transformer la variable « satisfaisant » en facteur 9](#_Toc134283675)

[d) Diviser les données 10](#_Toc134283676)

[**6.** **APPLIQUER L’ANALYSE DISCRIMINANTE** 0](#_Toc134283677)

[1) Choix et justification de la méthode ADL pas à pas : 0](#_Toc134283678)

[2) Les résultats 0](#_Toc134283679)

[**7.** **Interprétation des résultats :** 1](#_Toc134283680)

[**8.** **Conclusion :** 1](#_Toc134283681)

# **Introduction :**

# **L'analyse discriminatoire**

## **Définition Objectifs de l'analyse discriminatoire**

## **Les différentes techniques d'analyse discriminatoire**

## **Les hypothèses et les prérequis pour réaliser une analyse discriminatoire**

.

# **Etude de cas [Entreprise BETA]**

## **contexte :**

BETA est une entreprise de restauration qui fournit des services de restauration dans la cité universitaire de Béni Mellal. Les étudiants ont la possibilité de réserver des repas à partir d'un catalogue fixe pour toute la semaine. Les repas sont ensuite distribués par des agents après vérification de la réservation. L'entreprise rencontre des problèmes de qualité des repas et de service, ce qui conduit à des grèves de la part des étudiants insatisfaits.

## **Problématique à résoudre :**

L'entreprise BETA souhaite améliorer la qualité de son service de restauration pour réduire les grèves. Pour atteindre cet objectif, BETA se demande quels sont les critères qui influencent la satisfaction des étudiants et comment elle peut les améliorer.

## **Méthodologie :**

Pour répondre à cette problématique, une analyse discriminatoire sera utilisée pour identifier les critères qui ont le plus d'impact sur la satisfaction des étudiants et pour déterminer les caractéristiques des étudiants les plus satisfaits.

## **Description des données utilisées dans le mini-projet**

Dans cette étude de cas sur l'entreprise de restauration BETA dans la cité universitaire de Béni Mellal, nous avons utilisé les variables suivantes :

Qualité du service, Qualité des repas : des variables quantitatives continues mesurant respectivement la qualité globale du service, la qualité des repas fournissent par l'entreprise BETA sur une échelle de 1 à 10, où 1 représente une très mauvaise qualité et 10 représente une excellente qualité.

Le temps d'attente (TA): correspond au temps que les étudiants doivent attendre avant de recevoir leur repas sur une échelle de 1 à 10, où 1 représente une grande durée d’attente et 10 représente que l’étudiant reçoit son repas vitement.

Sexe : une variable catégorielle nominale indiquant le sexe des étudiants.

Programme d'étude : une variable catégorielle nominale indiquant le programme d'étude des étudiants.

Age correspond à l'âge des étudiants.

La propreté du restaurant (PR)**:** correspond à une évaluation de la propreté du restaurant de l'entreprise BETA. Cette évaluation peut inclure des critères tels que la propreté des tables, des chaises, des sols, des toilettes, etc. sur une échelle de 1 à 10, où 1 représente une très mauvaise qualité et 10 représente une excellente qualité.

Nombre d’années de résidence : qui correspond au nombres d’années que l’étudiant a passé dans la cité universitaire qui est un nombre entier.Les données nécessaires pour cette analyse sont collectées à travers des enquêtes auprès des étudiants. Les enquêtes sont diffusées en ligne à travers les groupes Facebook et WhatsApp en demandant aux étudiants de remplir un questionnaire.



Figure Capture de questionnaire diffusé

## **Préparation des données :**

Avant de procéder à l'analyse des données de l'entreprise BETA, nous avons dû préparer les données en effectuant des étapes de nettoyage et de transformation. Les données ont été collectées à l'aide d'un formulaire Google et stockées dans un fichier Excel. Un échantillon des données préparées est présenté ci-dessous.

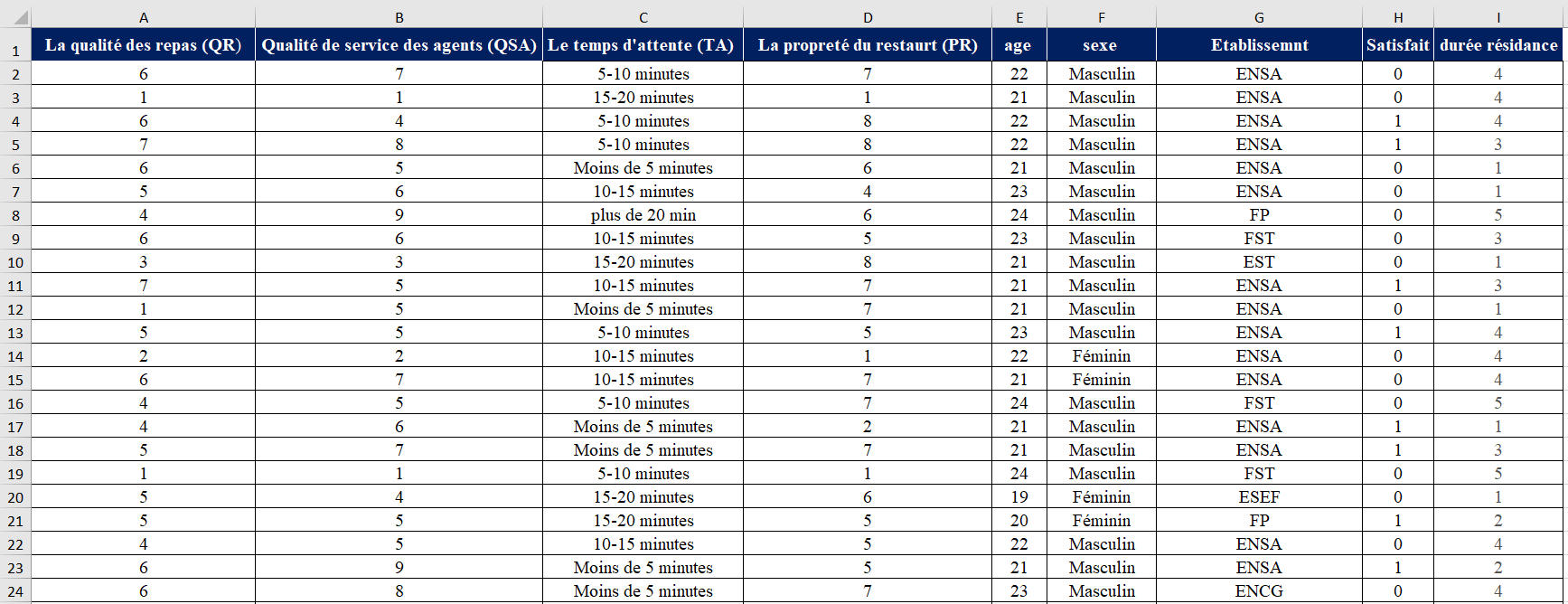
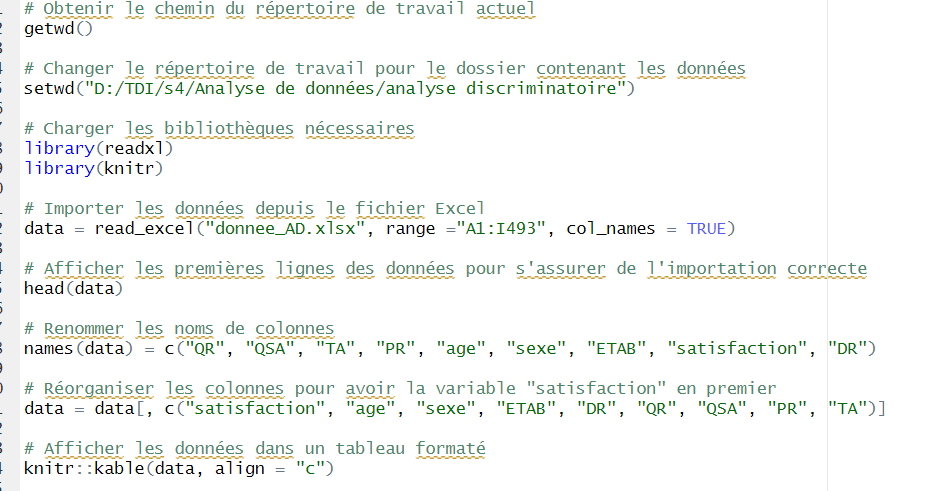
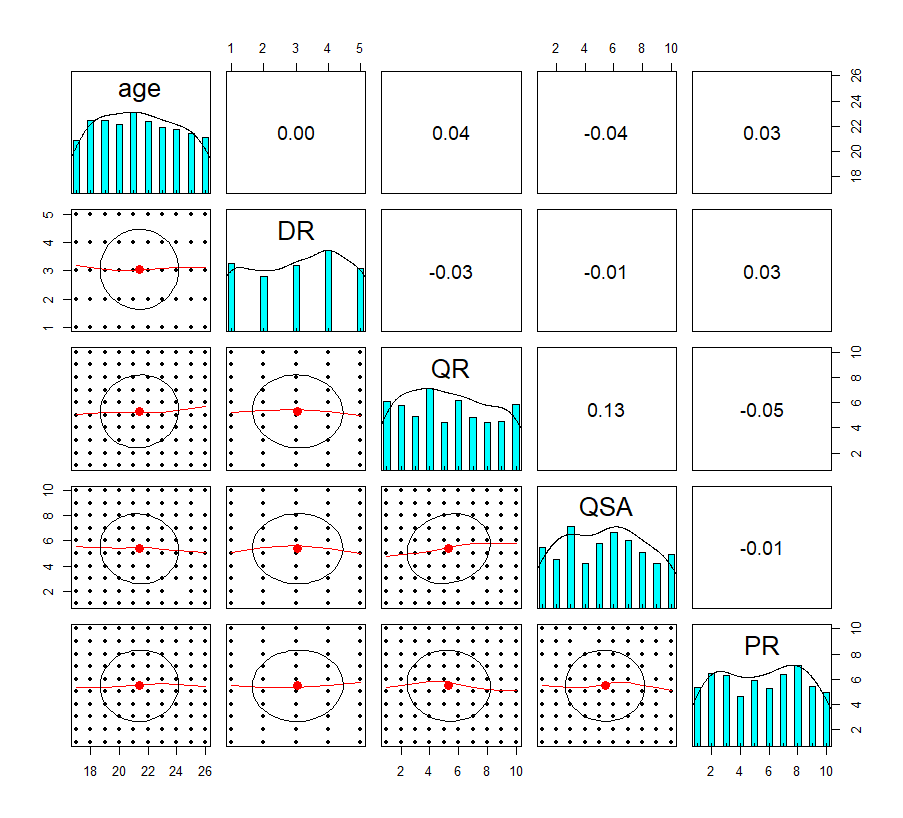
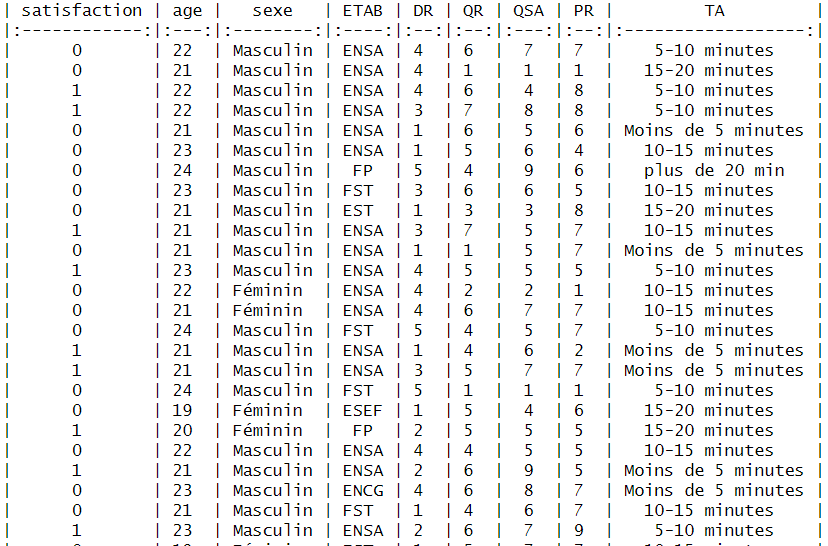


Figure échantillon des données collectées

### Importation des données





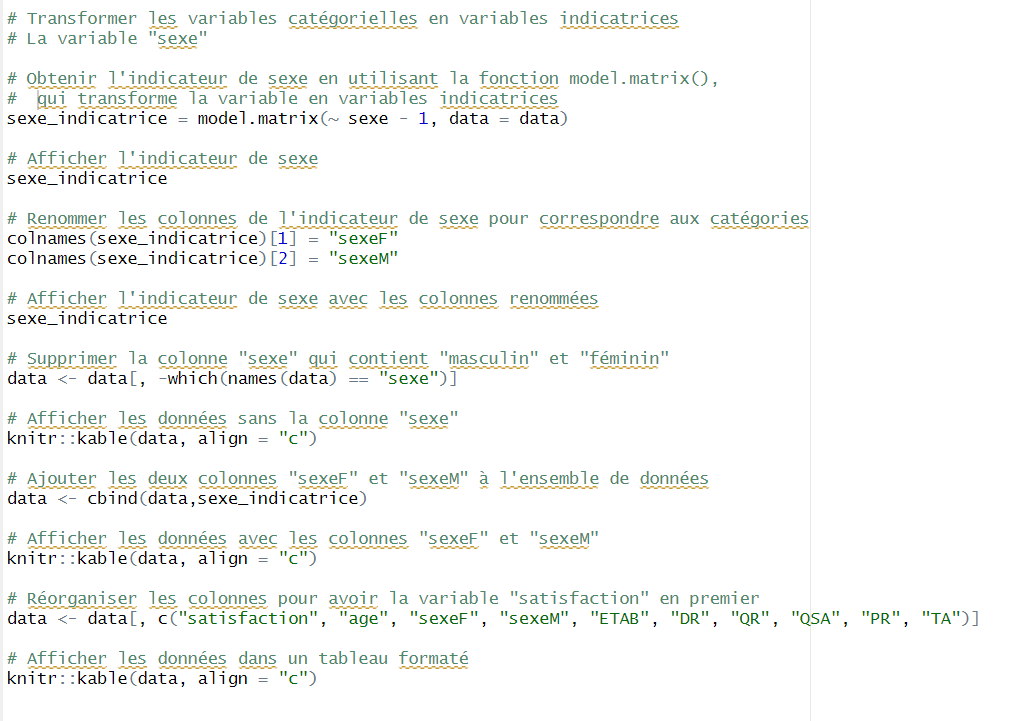


Commentaire

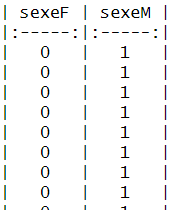
Cette partie de code importe les données à partir d'un fichier Excel, les réorganise et renomme les colonnes. Il affiche ensuite les données dans un tableau formaté en utilisant la fonction **kable**() de la bibliothèque **knitr**

### Transformation des variables qualitatives en variables indicatrices

#### La variable « sexe »



Commentaire

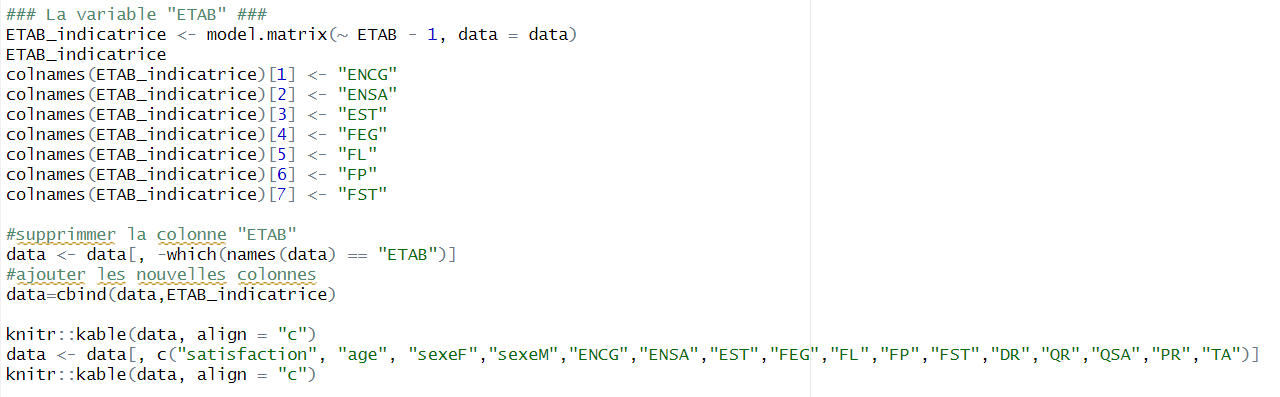


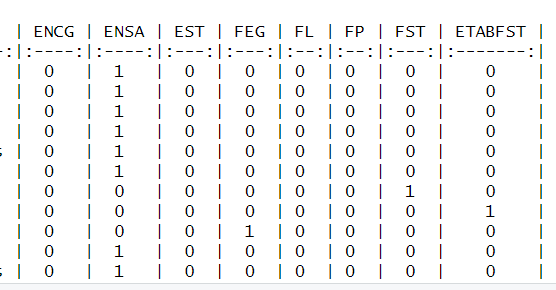
Le code utilise la fonction model.matrix pour transformer la variable catégorielle "sexe" en variables indicatrices et les ajouter aux données. Les colonnes de l'indicateur de sexe sont renommées pour correspondre aux catégories et la colonne "sexe" est supprimée des données. Les nouvelles colonnes indicatrices "sexeF" et "sexeM" sont ajoutées aux données et les colonnes sont réorganisées pour avoir la variable "satisfaction" en premier. Le tableau final est affiché avec la fonction kable.

**Pourquoi transformer les données**

La transformation des variables qualitatives en variables indicatrices est importante pour l'analyse discriminante car elle permet de traiter les variables qualitatives comme des variables numériques, ce qui est nécessaire pour effectuer des calculs mathématiques sur ces variables. Dans l'analyse discriminante, les variables indicatrices sont utilisées pour construire des modèles de discrimination qui permettent de distinguer des groupes en fonction de ces variables. Les variables indicatrices sont également utiles pour effectuer des tests statistiques et des analyses de régression sur des données qualitatives.

#### La variable « ETAB »

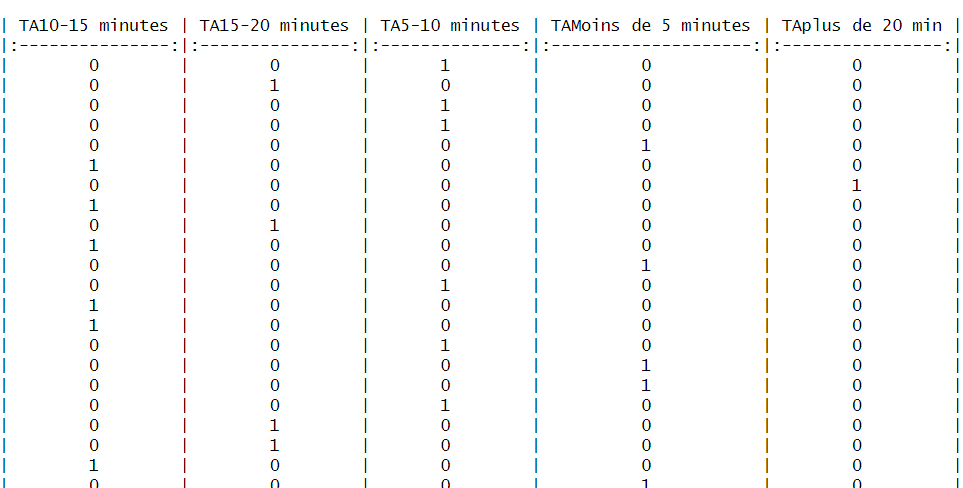


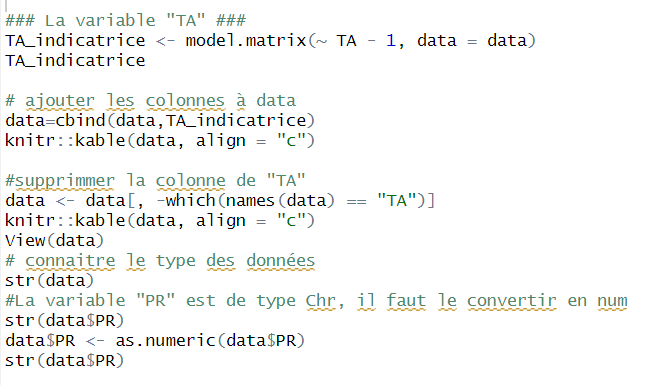


**Commentaire**

Cette partie du code sert à transformer la variable qualitative "ETAB" en une série de variables indicatrices. Elle utilise la fonction "model.matrix" pour créer une matrice d'indicateurs qui représentent chaque niveau unique de la variable "ETAB". Les nouvelles variables indicatrices sont ensuite ajoutées au jeu de données d'origine, puis la variable "ETAB" est supprimée car elle n'est plus nécessaire. Les noms de colonnes sont également renommés pour correspondre aux différentes catégories de "ETAB". Enfin, les colonnes sont réorganisées pour avoir la variable "satisfaction" en premier.

#### La variable « TA »



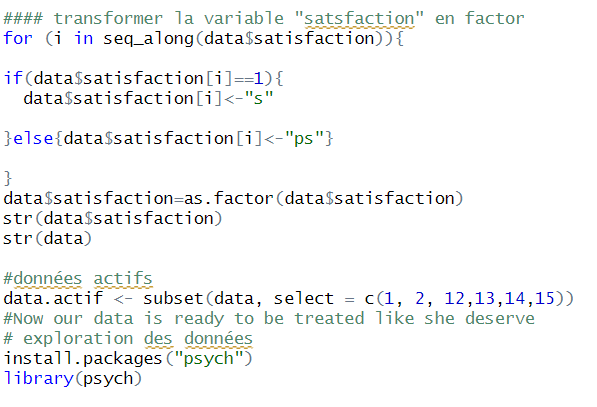
. 

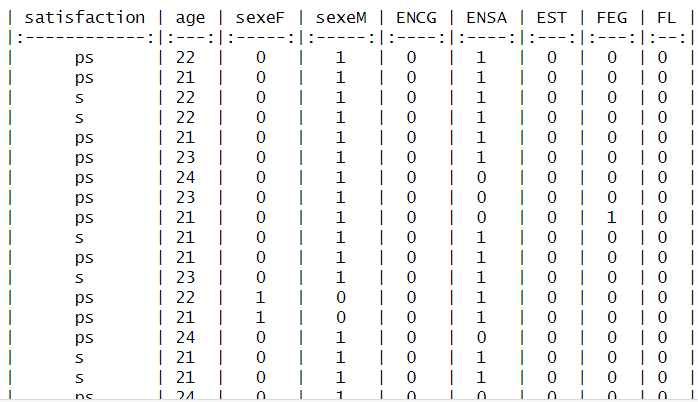
**Commentaire**

La première partie du code utilise la fonction model.matrix pour créer les nouvelles variables indicatrices, renommer les colonnes, supprimer la colonne originale et ajouter les nouvelles colonnes à la dataframe "data".

La deuxième partie permet de convertir la variable "PR" qui est de type caractère en une variable numérique.

### Transformer la variable « satisfaisant » en facteur



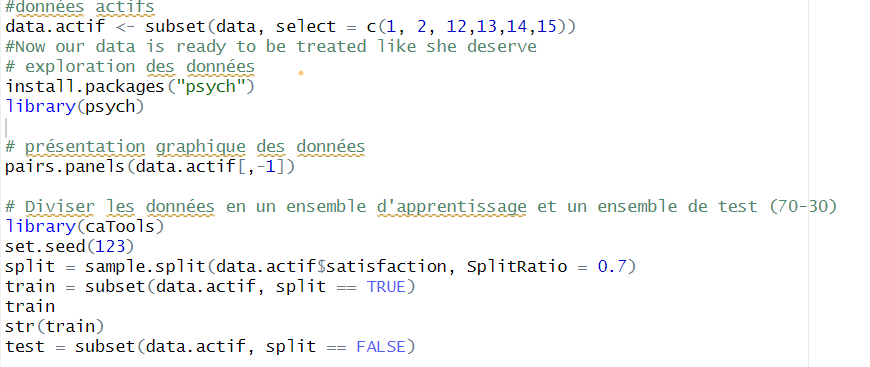
****

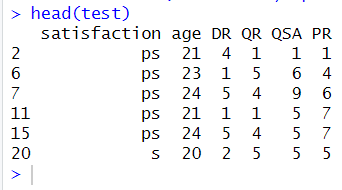
**Commentaire**

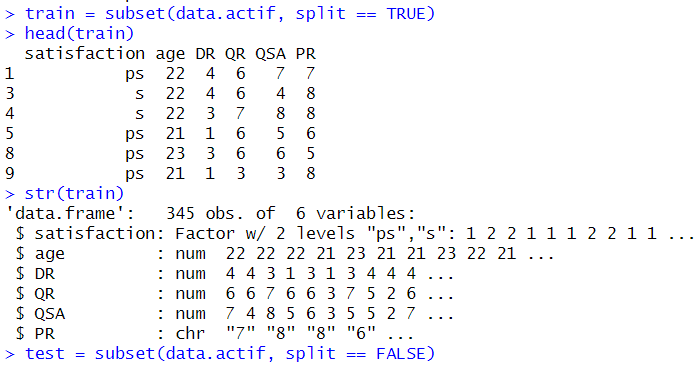
Ce code transforme la variable "satisfaction" en une variable factorielle avec deux niveaux ("s" pour satisfait et "ps" pour pas satisfait). Il utilise une boucle pour parcourir chaque élément de la colonne "satisfaction" et applique la transformation en fonction de la valeur de la satisfaction. La fonction as.factor() est ensuite utilisée pour convertir la variable en un facteur.

Ensuite, la fonction subset() est utilisée pour sélectionner certaines colonnes spécifiques du dataframe "data" et les stocker dans un nouveau dataframe "data.actif" pour une analyse discriminante ultérieure.

### Diviser les données

.





**Commentaire**

Ce code utilise la fonction sample.split du package caTools pour diviser le jeu de données data.actif en deux sous-ensembles : un ensemble d'apprentissage train et un ensemble de test test. La proportion des données dans l'ensemble de formation est fixée à 70% à l'aide du paramètre SplitRatio = 0.7. La fonction set.seed(123) est utilisée pour reproduire les mêmes résultats à chaque exécution du code. La fonction subset() est utilisée pour extraire les données correspondant à l'ensemble d'apprentissage et à l'ensemble de test. La fonction head() est utilisée pour afficher les 6 premières lignes de chaque ensemble de données, et la fonction str() est utilisée pour afficher la structure de l'ensemble d'apprentissage train.

## **APPLIQUER L’ANALYSE DISCRIMINANTE**

### Choix et justification de la méthode ADL pas à pas :

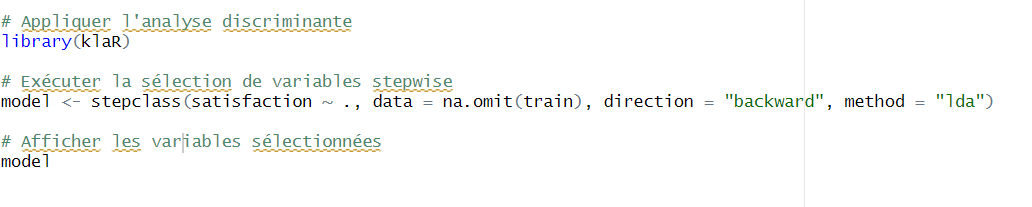
Nous avons choisi la méthode ADL (Analyse discriminante linéaire) pour notre analyse discriminante car elle est l'une des méthodes les plus couramment utilisées. Elle permet de séparer les groupes de données en utilisant des combinaisons linéaires des variables explicatives.

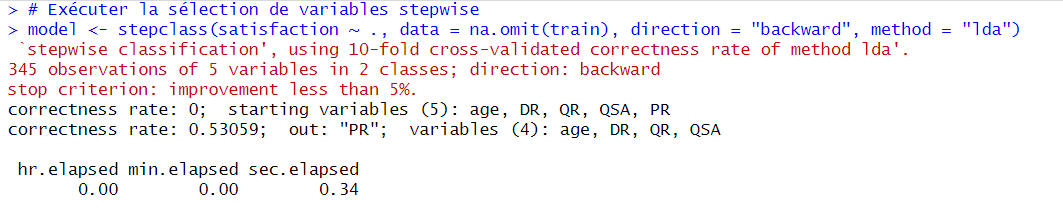
Pour analyser les données, nous avons utilisé la méthode ADL pas à pas, qui est une approche itérative qui sélectionne de manière séquentielle les variables explicatives les plus importantes pour construire le modèle final.

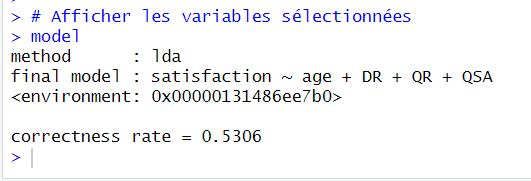
Dans le cadre de notre analyse de données de restauration, nous avons voulu comprendre les facteurs qui impactent plus sur la satisfaction des étudiants.

En résumé, la méthode ADL pas à pas de l'analyse discriminante s'est avérée être une méthode efficace pour identifier les variables les plus importantes pour discriminer entre les groupes et elle a été particulièrement utile pour notre cas.

### Les résultats







**Commentaire**

Ce code utilise la méthode de sélection de variables stepwise backward pour choisir les variables les plus importantes pour prédire la variable de réponse "satisfaction" à partir des autres variables du jeu de données d'entraînement. La fonction stepclass() de la bibliothèque MASS est utilisée pour effectuer cette sélection, en utilisant le modèle d'analyse discriminante linéaire (LDA). La fonction renvoie les variables sélectionnées pour le modèle final.

## **Interprétation des résultats :**

Le résultat de la fonction montre que la méthode a été exécutée avec succès avec une amélioration de performance supérieure à 5% à chaque étape de la sélection. Le modèle final sélectionné contient 4 variables explicative, "age", "DR", "QR" et "QSA". Cela veut dire que l’entreprise BETA doit agir sur La qualité des repas et la qualité de services des agents pour améliorer la satisfaction des étudiants. D’autre part elle doit comprendre que les étudiants anciens sont moins satisfaits que les autre donc elle doit penser à donner des distinctions pour améliorer leurs satisfactions. La performance de ce modèle final est estimée à 53,06% de taux de classification correcte sur les données d'entraînement.

## **Conclusion :**

Cependant, il est important de souligner que notre modèle n'est qu'une approximation de la réalité, et que d'autres facteurs peuvent influencer la satisfaction des étudiants qui n'ont pas été pris en compte dans cette analyse. Par conséquent, nous recommandons une approche plus globale pour comprendre les attentes et les besoins des étudiants, en incluant des enquêtes de satisfaction et d'autres formes de feedback des étudiants.

En fin de compte, cette analyse discriminante peut être un outil utile pour BETA Restauration pour prédire la satisfaction des étudiants et améliorer leur expérience. Cependant, il ne doit pas être utilisé comme la seule méthode pour évaluer la satisfaction des étudiants et des méthodes supplémentaires doivent être utilisées pour une meilleure compréhension des attentes des étudiants