Limpieza de Datos: Ejercicio en Clase

Introducción

En un proyecto de Ciencia de Datos es muy importante contar con una estructura de datos que facilite su comprensión y manipulación. Es por eso, que una parte muy importante dentro de un proyecto de ciencia de datos es la **limpiza de datos**.

El término de **limpieza de datos** corresponde a un proceso que permite asegurar la calidad de los datos que se van a emplear, para cumplir con el objetivo de un proyecto y poder contestar la pregunta planteada. (En el caso del proyecto final del curso: ¿Cuántas unidades de cada producto se van a vender, en cada punto de venta, el siguiente mes de registro?). En otras palabras, la **limpieza de datos** es el proceso que se encarga de **corregir los problemas de calidad** en los datos que se detectaron en la etapa 2 del CRISP-DM, para, al finalizar, contar con un conjunto de datos listos para su análisis e implementación en modelos como solución a una problemática de Ciencia de Datos.

Retomando, CRISP-DM:

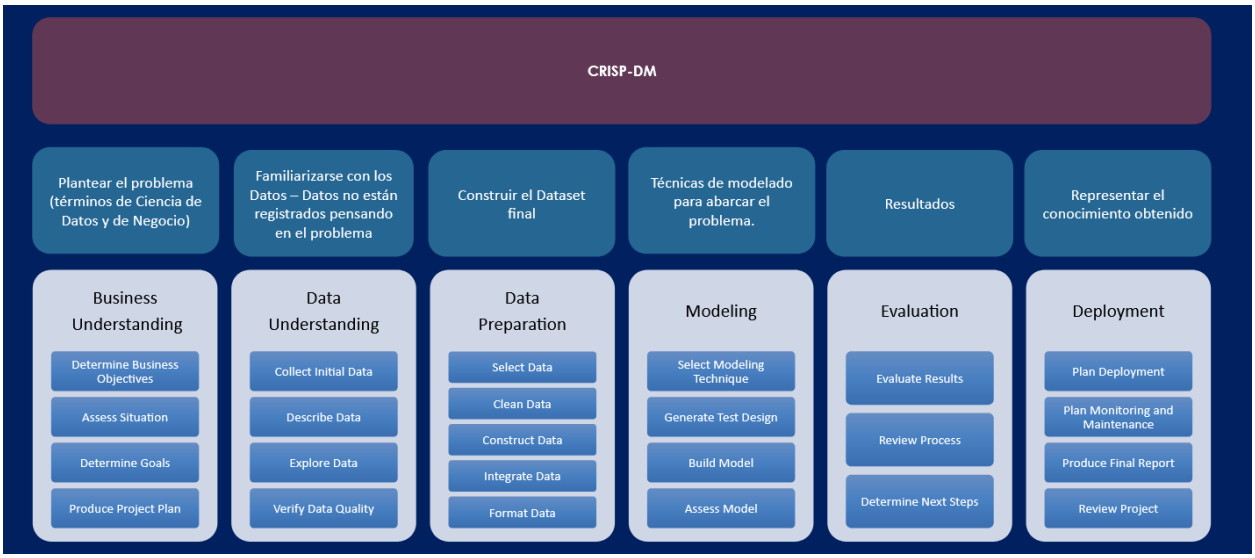

Figure 1: CRISP-DM.

CRISP-DM - Etapa 2: Data Understanding

Algunos de los pasos que se hacen para comprender los datos con más detalle son:

1. Lectura de Datos.
2. Análisis general de los datos. (Formato, dimensión, variables, etc.)
3. Detectar problemas de calidad en todo el conjunto de datos:

- Valores faltantes.
- Datos no homogéneos (Mayúsculas y Minúsculas / Variables Númericas y Categóricas)
- Faltas de ortografía.
- Incongruencias en los datos,
- Valores repetidos.
- Caracteres especiales.

CRISP-DM - Etapa 3: Data Preparatation

1. Lectura de los datos.
2. Detectar problemas de calidad en todo el conjunto de datos.
3. Corregir los problemas de calidad.

- Imputar valores faltantes.
- Homogeneizar registros.
- Corregir registros.
- Eliminar caracteres especiales y valores repetidos.

4. Guardar el nuevo archivo con los datos limpios.

Con relación al CRISP - DM, la **limpieza de datos** entra dentro de la tercera etapa **Data Preparation**. ¡Importante! Los pasos siguientes a la limpieza de datos son: **Análisis Exploratorio de los Datos** e **Ingeniería de Características**. Pasos que aún se consideran dentro de la tercera etapa, **Data Preparation** del CRISP-DM.

Ejercicio en Clase

```
#Librerias
library(tidyverse)
```

```
#data()
```

```
#Conjunto de datos
iris_dataset <- iris
```

```
#Primeros registros
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
#Dimensión del conjunto de datos
dim(iris_dataset)
```

```
## [1] 150   5
```

```
#guardar un archivo
write.csv(iris_dataset, file="iris_dataset.csv", row.names = FALSE)
```

1. Lectura de datos

```r
#C:\Users\anamh\Desktop\TEC\Semestre_AD_2020\Laboratorio_DYO_De_Operaciones\Semana4
iris_ejercicio <- read.csv("iris_dataset.csv")
```

2. Análisis general de los datos

```r
#tipo de variables
str(iris_ejercicio)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 ...
```

```r
#dimesión
dim(iris_ejercicio)
```

```
## [1] 150   5
```

```r
#descripción general de las variables
summary(iris_ejercicio)
```

```
##   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
##  Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
##  1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
##  Median :5.800   Median :3.000   Median :4.350   Median :1.300
##  Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
##  3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
##  Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
##        Species
##  setosa    :50
##  versicolor:50
##  virginica :50
##
##
##
```

3. Detectar (en este caso, generar) problemas de calidad:

```r
head(iris_ejercicio)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

Problema 1: Faltas de ortografía, valores mal escritos. (POR COLUMNA)

```r
#cambiar variable Species - como ya esta "bonito" en factor, hay que cambiarlo para generar problemas d
iris_ejercicio$Species <- as.character(iris_ejercicio$Species)
```

```r
iris_ejercicio[1,5] <- "set"
iris_ejercicio[2,5] <- "SETOSA"
iris_ejercicio[3,5] <- "cetósa"

iris_ejercicio[150,5] <- "Virg"
iris_ejercicio[148,5] <- "VIRGINICA"
```

```r
head(iris_ejercicio)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2     set
## 2          4.9         3.0          1.4         0.2  SETOSA
## 3          4.7         3.2          1.3         0.2  cetósa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```r
diferentes_species <- iris_ejercicio %>% select(Species)%>%unique()
diferentes_species
```

```
##        Species
## 1          set
## 2       SETOSA
## 3       cetósa
## 4       setosa
## 51  versicolor
## 101  virginica
## 148  VIRGINICA
## 150       Virg
```

Solución 1: Homogeneizar registros (POR COLUMNA)

```r
#Pasar a minúsculas POR COLUMNA
iris_ejercicio$Species <- tolower(iris_ejercicio$Species)
```

```r
iris_ejercicio %>% select(Species)%>%unique()
```

```
##        Species
## 1          set
## 2       setosa
## 3       cetósa
## 51  versicolor
## 101  virginica
## 150       virg
```

```r
#Quitar caracteres especiales: acentos, espacios, / , - , _ _, ñ, etc
#_ no es malo, depende de cada quien como maneje sus datos (iris_ejercico -> irisejercicio)

iris_ejercicio$Species <- str_replace(iris_ejercicio$Species, "á", "a") %>%
  str_replace("é","e")%>%
  str_replace("í", "i") %>%
  str_replace("ó","o")%>%
  str_replace("ú", "u")#%>%
  #str_replace("ñ","n")%>%
  #str_replace(" - ", " ")%>%
  #str_replace("-"," ")%>%
  #str_replace("  "," ")
```

```r
iris_ejercicio %>% select(Species)%>%unique()
```

```
##        Species
## 1          set
## 2       setosa
## 3       cetosa
## 51  versicolor
## 101  virginica
## 150       virg
```

```r
#corregir faltas de ortografía
iris_ejercicio$Species <- str_replace(iris_ejercicio$Species, "set", "setosa")%>%
  str_replace("cetosa","setosa") %>%
  str_replace("virg","virginica")
```

IMPORTANTE:

```r
iris_ejercicio %>% select(Species)%>%unique()
```

```
##            Species
## 1           setosa
## 2        setosaosa
## 51      versicolor
## 101 virginicainica
## 150      virginica
```

```r
iris_ejercicio$Species <- str_replace(iris_ejercicio$Species, "virginicainica","virginica") %>%
  str_replace("setosaosa","setosa")
```

```r
iris_ejercicio %>% select(Species)%>%unique()
```

```
##        Species
## 1       setosa
## 51  versicolor
## 101  virginica
```

```r
diferentes_species <- iris_ejercicio %>% select(Species)%>%unique()
diferentes_species
```

```
##        Species
## 1       setosa
## 51  versicolor
## 101  virginica
```

```r
iris_ejercicio$Species <- as.factor(iris_ejercicio$Species)
```

```r
class(iris_ejercicio$Species)
```

```
## [1] "factor"
```

```r
levels(iris_ejercicio$Species)
```

```
## [1] "setosa"     "versicolor" "virginica"
```

Problema 2: Valores incongruentes

```r
head(iris_ejercicio)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```r
summary(iris_ejercicio)
```

```
##   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
##  Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
##  1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
##  Median :5.800   Median :3.000   Median :4.350   Median :1.300
##  Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
##  3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
##  Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
##        Species
##  setosa    :50
##  versicolor:50
##  virginica :50
##
##
##
```

?iris - saber más información sobre datos

```r
iris_ejercicio[6,1] <- 54
iris_ejercicio[5,2] <- 360
iris_ejercicio[3,1] <- 54
```

```r
summary(iris_ejercicio)
```

```
##   Sepal.Length     Sepal.Width       Petal.Length    Petal.Width
##  Min.   : 4.300   Min.   :  2.000   Min.   :1.000   Min.   :0.100
##  1st Qu.: 5.100   1st Qu.:  2.800   1st Qu.:1.600   1st Qu.:0.300
##  Median : 5.800   Median :  3.000   Median :4.350   Median :1.300
##  Mean   : 6.496   Mean   :  5.433   Mean   :3.758   Mean   :1.199
##  3rd Qu.: 6.400   3rd Qu.:  3.300   3rd Qu.:5.100   3rd Qu.:1.800
##  Max.   :54.000   Max.   :360.000   Max.   :6.900   Max.   :2.500
##        Species
##  setosa    :50
##  versicolor:50
##  virginica :50
##
##
##
```

Solución 2:

```r
valores_anormales_1 <- iris_ejercicio %>% filter(Sepal.Length >= 50)
valores_anormales_1
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1           54         3.2          1.3         0.2  setosa
## 2           54         3.9          1.7         0.4  setosa
```

```r
which(iris_ejercicio$Sepal.Length > 50)
```

```
## [1] 3 6
```

```r
iris_ejercicio[c(3,6),]
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 3           54         3.2          1.3         0.2  setosa
## 6           54         3.9          1.7         0.4  setosa
```

```r
iris_ejercicio[c(3,6),1] <- 5.4
head(iris_ejercicio)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          5.4         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0       360.0          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

Hacer esto para cada uno de los valores fuera de rango.

Problema 3: NA's

```
iris_ejercicio[1,5]<-NA
```

Solución 3: Detectar NA'S imputar valores

```
na_dataframe <- is.na(iris_ejercicio)
na_dataframe
```

```
##        Sepal.Length Sepal.Width Petal.Length Petal.Width Species
##   [1,]        FALSE       FALSE        FALSE       FALSE    TRUE
##   [2,]        FALSE       FALSE        FALSE       FALSE   FALSE
##   [3,]        FALSE       FALSE        FALSE       FALSE   FALSE
##   [4,]        FALSE       FALSE        FALSE       FALSE   FALSE
##   [5,]        FALSE       FALSE        FALSE       FALSE   FALSE
##   [6,]        FALSE       FALSE        FALSE       FALSE   FALSE
##   [7,]        FALSE       FALSE        FALSE       FALSE   FALSE
##   [8,]        FALSE       FALSE        FALSE       FALSE   FALSE
##   [9,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [10,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [11,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [12,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [13,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [14,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [15,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [16,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [17,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [18,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [19,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [20,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [21,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [22,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [23,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [24,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [25,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [26,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [27,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [28,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [29,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [30,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [31,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [32,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [33,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [34,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [35,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [36,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [37,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [38,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [39,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [40,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [41,]        FALSE       FALSE        FALSE       FALSE   FALSE
##  [42,]        FALSE       FALSE        FALSE       FALSE   FALSE
```

```
##  [43,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [44,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [45,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [46,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [47,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [48,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [49,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [50,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [51,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [52,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [53,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [54,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [55,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [56,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [57,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [58,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [59,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [60,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [61,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [62,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [63,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [64,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [65,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [66,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [67,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [68,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [69,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [70,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [71,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [72,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [73,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [74,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [75,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [76,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [77,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [78,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [79,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [80,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [81,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [82,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [83,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [84,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [85,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [86,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [87,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [88,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [89,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [90,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [91,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [92,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [93,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [94,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [95,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [96,]       FALSE       FALSE       FALSE       FALSE   FALSE
```

```
##  [97,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [98,]       FALSE       FALSE       FALSE       FALSE   FALSE
##  [99,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [100,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [101,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [102,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [103,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [104,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [105,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [106,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [107,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [108,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [109,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [110,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [111,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [112,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [113,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [114,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [115,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [116,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [117,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [118,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [119,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [120,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [121,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [122,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [123,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [124,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [125,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [126,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [127,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [128,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [129,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [130,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [131,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [132,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [133,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [134,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [135,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [136,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [137,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [138,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [139,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [140,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [141,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [142,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [143,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [144,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [145,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [146,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [147,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [148,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [149,]       FALSE       FALSE       FALSE       FALSE   FALSE
## [150,]       FALSE       FALSE       FALSE       FALSE   FALSE
```

```r
class(na_dataframe)
```

```
## [1] "matrix"
```

```r
which(is.na(iris_ejercicio$Species))
```

```
## [1] 1
```

```r
iris_ejercicio[1,]
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2    <NA>
```

```r
iris_ejercicio %>% filter(Sepal.Length==5.1)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 1          5.1         3.5          1.4         0.2       <NA>
## 2          5.1         3.5          1.4         0.3     setosa
## 3          5.1         3.8          1.5         0.3     setosa
## 4          5.1         3.7          1.5         0.4     setosa
## 5          5.1         3.3          1.7         0.5     setosa
## 6          5.1         3.4          1.5         0.2     setosa
## 7          5.1         3.8          1.9         0.4     setosa
## 8          5.1         3.8          1.6         0.2     setosa
## 9          5.1         2.5          3.0         1.1 versicolor
```

```r
iris_ejercicio[1,5] <- "setosa"
```

```r
head(iris_ejercicio)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          5.4         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0       360.0          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```