

Parcial 3

Moreno - Ochoa - Casarubio

30/11/2021

3 Examen Parcial Ciencias de Datos

NOMBRES:

- Miguel Moreno.
- Esteban Ochoa.
- David Casarrubia.

1- Elija un conjunto de datos relacionados con el sector salud. Idealmente, el conjunto de datos debería tener alrededor de ~ 200 filas (Observaciones) y debería tener variables (predictores) tanto categóricas como numéricas.

Fuentes potenciales de datos: TidyTuesday: <https://github.com/rfordatascience/tidytuesday> · KAGGLE. <https://www.kaggle.com/general/168211> · <https://medium.com/@ODSC/15-open-datasets-for-healthcare-830b19980d9>

2- Debe referenciar la fuente donde extrajo los datos:

Estoy usando el conjunto de datos de 'Students Performance in Exams' de los datos disponibles en el sitio web de kaggle 2018. <https://www.kaggle.com/spscientist/students-performance-in-exams>

3- Tenga en cuenta que la mayoría de estos son archivos .csv. Escriba el código para cargar los archivos desde csv para cada uno de los conjuntos de datos y una breve descripción de las variables, o puede cargar el .csv archivo en su carpeta de datos.

R// no se pudieron extraer los datos de la web entonces subire el data set a github

```
library(MASS); library(dplyr); library(caret)
library(mlbench); library(tidyverse)
library(modelr); library(broom); library(pROC)
library(ggplot2); library(mice); library("gridExtra")
library(PerformanceAnalytics); library(readr)
```

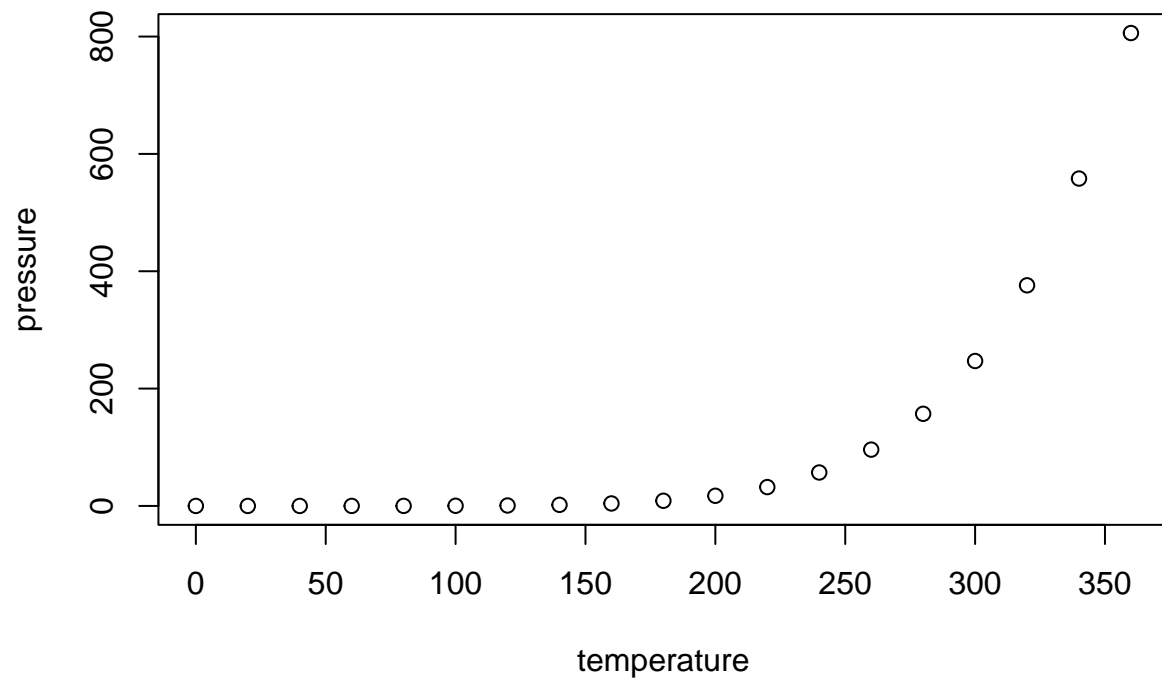
R// cargar la base de datos

```
getwd()
```

```
## [1] "C:/Users/Miguel Angel/Documents/materias/Ciencias de datos/ML-Class"
```

```
setwd("C:/Users/Miguel Angel/Documents/materias/Ciencias de datos/DS - 2021/tareas en clase/Parcial 3")
Estudiantes <- read_csv("StudentsPerformance.csv")
Estudiantes <- data.frame(Estudiantes)
```

4 - Defina su pregunta de investigación a continuación. ¿Qué te interesa de los datos? ¿Cuál es una pregunta específica que desea averiguar sobre los datos? (10 Puntos).



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.