

titanic_dataset_project

Manuel Herrera Lara y Anahí Berumen Murillo

15/9/2020

The data domain

Usaremos Machine Learning para crear un modelo que predice cuáles pasajeros sobrevivieron al naufragio del titanic y/o qué tipo de personas tenían mas probabilidades de sobrevivir, usando información de los pasajeros que viajaban en el titanic; como su nombre, edad, sexo, clase socioeconómica, etc. Como breve descripción podemos decir que el hundimiento del titanic fue uno de los naufragios mas infames y recordados de la historia. El RMS Titanic fue un crucero de pasajeros británico que se hundió en el Océano Atlántico Norte y esto sucedio el 15 de abril de 1912, durante su viaje inaugural; y el RMS Titanic, considerado “insumergible”, se hundió tras chocar con un iceberg. Desafortunadamente, no había suficientes botes salvavidas para todos a bordo, lo que resultó en la muerte de 1502 de los 2224 pasajeros y la tripulación. Al parecer algunos grupos de personas tenían más probabilidades de sobrevivir que otros. Y por último destacamos que el Titanic era el barco más grande a flote en el momento y fue construido por el astillero Harland and Wolff en Belfast.

Describe each of the variables

» Variable, definición y/o posibles valores

PassengerId Número de identificación al pasajero.

Survived Indica si el pasajero sobrevivió. 0 = No, 1 = Yes.

Pclass Define la clase socioeconómica del pasajero. 1 = Baja, 2 = Media y 3 = Alta.

Name Nombre del pasajero.

Sex Género del pasajero. Masculino y/o Femenino.

Age Edad del pasajero.

SibSp Número de hermanos y/o cónyuges a bordo.

Parch Número de padres y/o niños a bordo.

Ticket Número de boleto.

Fare Tarifa de pasajero.

Cabin Número de cabina.

Embarked Porción de embarcación. (C = Cherbourg; Q = Queenstown; S = Southampton).

Notas adicionales para algunas variables

pclass Indica el status o clase socioeconómica del pasajero.

1 = Baja

2 = Media

3 = Alta

sibsp El dataset define las relaciones familiares de esta forma:

sibling = hermano, hermana, hermanastro, hermanastra.

spouse = esposo y/o esposa.

parch El dataset define las relaciones familiares de esta forma:

parent: mamá o papá.

child: hijo, hija, hermanastro y/o hermanastra.

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

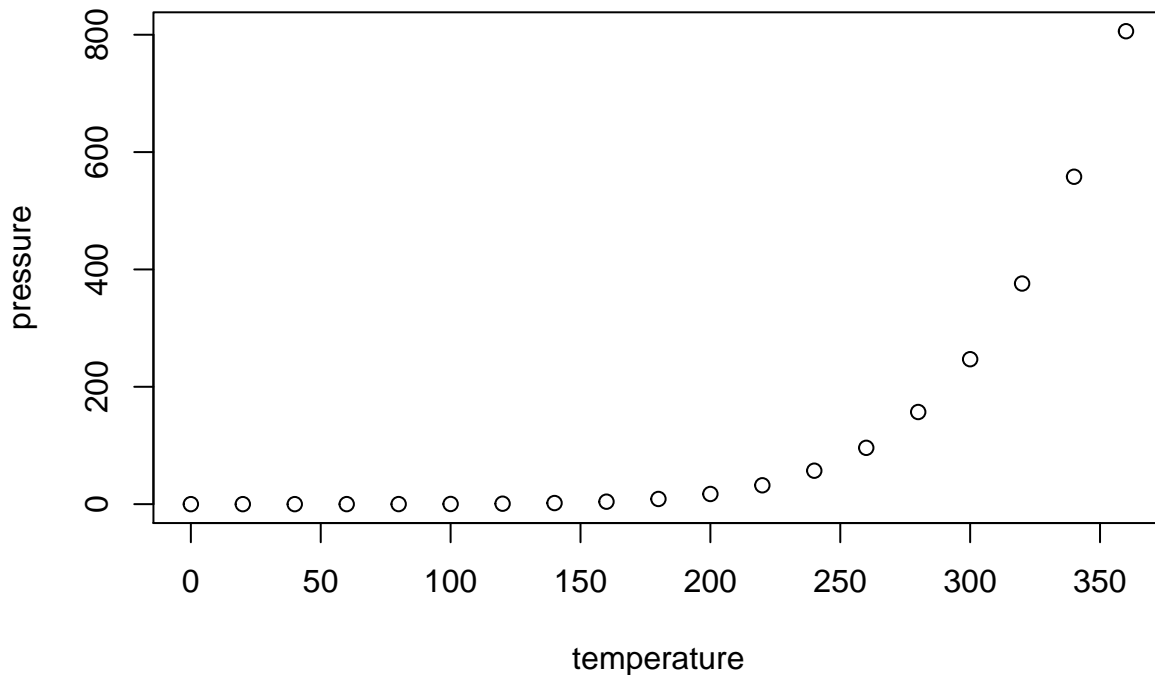
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.    :  2.00
## 1st Qu.:12.0    1st Qu.: 26.00
##  Median:15.0    Median : 36.00
##   Mean  :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
##   Max.  :25.0    Max.    :120.00
```

Including Plots

You can also embed plots, for example:



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