

# Web Scraping - DataSet

*Byron Vinicio Lima Rojas*

*22 de octubre, 2017*

## Contents

Título del DataSet	1
Subtítulo del DataSet	1
Imagen del Dataset	1
¿Cual es la materia del conjunto de datos?	2
¿Qué campos incluye? ¿Cuál es el período de tiempo de los datos y cómo se ha recogido?	3
Agradecimientos. ¿Quién es el propietario del conjunto de datos? (Incluir citas de investigación o análisis anteriores)	3
Inspiración. ¿Por qué es interesante este conjunto de datos? ¿Qué preguntas le gustaria responder a la comunidad?	3
Licencia.	3
Código.	3
Documento CSV.	4
References	4

## Título del DataSet

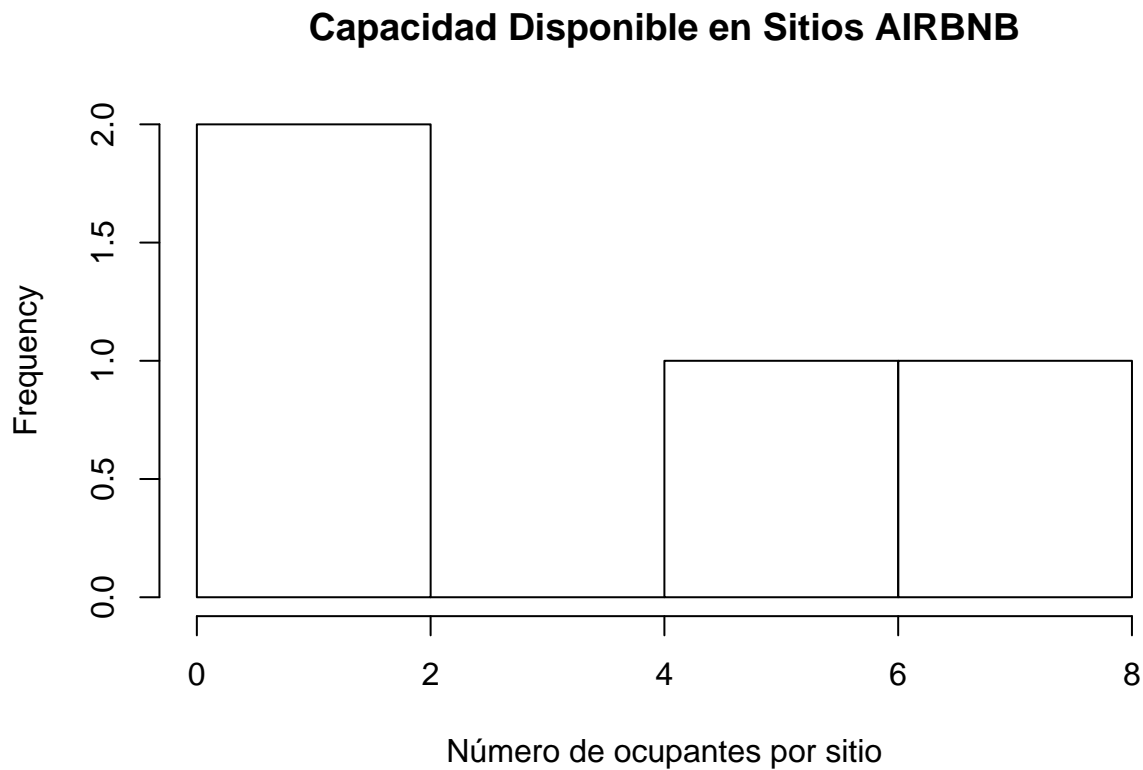
Comparativa de habitaciones ofrecidas mediante AirBNB

## Subtítulo del DataSet

Obtención de habitaciones disponibles en la Ciudad de Loja - Ecuador, para el Festival de Artes Vivas 2017

## Imagen del Dataset

```
# data exploration example using used car data
bnb <- read.csv("bnb.csv", stringsAsFactors = FALSE)
# the min/max of used car prices
hist(bnb$Capacidad, main = "Capacidad Disponible en Sitios AIRBNB",
      xlab = "Número de ocupantes por sitio")
```



## ¿Cual es la materia del conjunto de datos?

El conjunto de datos es analizado en su mayor parte por la estadística, en donde cada columna representa el valor de una variable y cada fila es considerada un registro o información de un elemento del conjunto, también conocido como dato. En la actualidad esta información puede ser extraída desde una base de datos o mediante técnicas de extracción de información, la misma tendrá que pasar un ciclo de vida hasta llegar a la publicación de datos y conclusiones de la misma (Wiki 2017).

En la extracción de datos de la página **AirBNB** es importante señalar que, en este proceso se ejecutaran las fases de captura y almacenamiento de datos.

¿Qué campos incluye? ¿Cuál es el período de tiempo de los datos y cómo se ha recogido?

Agradecimientos. ¿Quién es el propietario del conjunto de datos? (Incluir citas de investigación o análisis anteriores)

Inspiración. ¿Por qué es interesante este conjunto de datos? ¿Qué preguntas le gustaria responder a la comunidad?

Licencia.

Código.

```
from scrapy.item import Item
from scrapy.item import Field
from scrapy.spiders import CrawlSpider, Rule
from scrapy.loader import ItemLoader
from scrapy.linkextractors import LinkExtractor
from scrapy.loader.processors import MapCompose

autor = 'ByronLima'

class Habitacion(Item):
    tipo = Field()

    capacidad = Field()

class Airbnb(CrawlSpider):
    name = 'Habitaciones'

    start_urls = 'https://es.airbnb.com/s/loja'

    allowed_domains = ['airbnb.com']

    rules = (

        Rule(LinkExtractor(allow= r'page-')),

        Rule(LinkExtractor(allow=r'/rooms'), callback='parse_items'),
    )

    def parse_items(self, response):

        item = ItemLoader(Airbnb(), response)

        item.add_xpath('tipo', '//*[@id="summary"]/div/div/div[1]/div/div/div/div[2]/div[2]/div/div[1]/text
```

```
item.add_xpath('capacidad', '//*[@id="summary"]/div/div/div[1]/div/div/div/div[2]/div[2]/div/div[2]  
yield item.load_item()
```

## Documento CSV.

## References

Wiki, UMAIC. 2017. “Conjunto de Datos.” Accessed October 22. [https://wiki.umaic.org/wiki/Conjunto\\_de\\_datos](https://wiki.umaic.org/wiki/Conjunto_de_datos).