

# FAD\_DataAnalysis\_practica

Alvaro Simón Merino, Antonio Fernández Cáceres

13/11/2021

#Instalar Librerias

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
library(stringr)
```

## R Markdown

```
RawData <- read.csv(file = '../00_DatosOriginales/listings.csv')
```

```
#summary(RawData)
```

```
data <- RawData %>%
```

```
  select(id, host_id, host_since, host_is_superhost, neighbourhood_group_cleansed, property_type, room_type)
summary(data)
```

```
##           id           host_id           host_since           host_is_superhost
## Min.      :   6369   Min.      :   7952   Length:18909   Length:18909
## 1st Qu.:19305558   1st Qu.: 27320546   Class :character   Class :character
## Median :33537761   Median : 99268253   Mode  :character   Mode  :character
## Mean    :30558406   Mean    :135905611
## 3rd Qu.:42555334   3rd Qu.:230152535
## Max.    :52182334   Max.    :422253585
##
## neighbourhood_group_cleansed property_type      room_type
## Length:18909                Length:18909      Length:18909
## Class :character            Class :character   Class :character
```

```

## Mode :character          Mode :character  Mode :character
##
##
##
##
##      accommodates      bedrooms      price      minimum_nights
## Min.   : 0.000   Min.   : 1.000   Length:18909   Min.   : 1.000
## 1st Qu.: 2.000   1st Qu.: 1.000   Class :character 1st Qu.: 1.000
## Median : 2.000   Median : 1.000   Mode  :character Median : 2.000
## Mean   : 3.095   Mean   : 1.442                   Mean   : 7.297
## 3rd Qu.: 4.000   3rd Qu.: 2.000                   3rd Qu.: 3.000
## Max.   :16.000   Max.   :18.000                   Max.   :1125.000
##              NA's   :1434
## minimum_nights_avg_ntm maximum_nights_avg_ntm availability_365
## Min.   : 1.000   Min.   : 1   Min.   : 0.0
## 1st Qu.: 1.000   1st Qu.: 365   1st Qu.: 0.0
## Median : 2.000   Median : 1125   Median :114.0
## Mean   : 8.319   Mean   : 1479   Mean   :154.4
## 3rd Qu.: 4.000   3rd Qu.: 1125   3rd Qu.:311.0
## Max.   :1125.000   Max.   :11111111   Max.   :365.0
## NA's   :1       NA's   :1
## review_scores_rating review_scores_accuracy review_scores_cleanliness
## Min.   :0.000   Min.   :1.00   Min.   :1.000
## 1st Qu.:4.500   1st Qu.:4.67   1st Qu.:4.560
## Median :4.750   Median :4.86   Median :4.800
## Mean   :4.549   Mean   :4.72   Mean   :4.658
## 3rd Qu.:4.940   3rd Qu.:5.00   3rd Qu.:4.980
## Max.   :5.000   Max.   :5.00   Max.   :5.000
## NA's   :5032   NA's   :5271   NA's   :5269
## review_scores_checkin review_scores_communication review_scores_location
## Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:4.750   1st Qu.:4.750   1st Qu.:4.730
## Median :4.900   Median :4.920   Median :4.920
## Mean   :4.782   Mean   :4.777   Mean   :4.788
## 3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.000
## Max.   :5.000   Max.   :5.000   Max.   :5.000
## NA's   :5269   NA's   :5269   NA's   :5272
## review_scores_value reviews_per_month
## Min.   :1.000   Min.   : 0.010
## 1st Qu.:4.500   1st Qu.: 0.200
## Median :4.710   Median : 0.690
## Mean   :4.588   Mean   : 1.222
## 3rd Qu.:4.880   3rd Qu.: 1.760
## Max.   :5.000   Max.   :23.290
## NA's   :5273   NA's   :5032

```

```
ncol(data)
```

```
## [1] 22
```

```
str(data)
```

```
## 'data.frame': 18909 obs. of 22 variables:
```

```
## $ id : int 6369 21853 23001 24805 26825 28200 30320 30478 30959 37859 ...
## $ host_id : int 13660 83531 82175 346366726 114340 82175 130907 130502 132883 ...
## $ host_since : chr "2009-04-16" "2010-02-21" "2010-02-17" "2020-05-15" ...
## $ host_is_superhost : chr "t" "f" "f" "f" ...
## $ neighbourhood_group_cleansed: chr "Chamartín" "Latina" "Arganzuela" "Centro" ...
## $ property_type : chr "Private room in rental unit" "Private room in rental unit" "Entire home/apt" ...
## $ room_type : chr "Private room" "Private room" "Entire home/apt" "Entire home/apt" ...
## $ accommodates : int 2 1 6 3 1 4 2 1 2 5 ...
## $ bedrooms : int 1 1 3 NA 1 3 1 1 1 2 ...
## $ price : chr "$60.00" "$31.00" "$50.00" "$92.00" ...
## $ minimum_nights : int 1 4 15 5 2 1 5 1 3 3 ...
## $ minimum_nights_avg_ntm : num 1 4 15 5 2 1 5 1 3 3 ...
## $ maximum_nights_avg_ntm : num 1125 40 730 1125 1125 ...
## $ availability_365 : int 180 364 222 115 349 223 314 315 0 365 ...
## $ review_scores_rating : num 4.87 4.58 NA 4.92 4.68 4 4.64 5 4.38 NA ...
## $ review_scores_accuracy : num 4.91 4.72 NA 4.75 4.8 4.2 4.73 4.33 4.14 NA ...
## $ review_scores_cleanliness : num 4.81 4.56 NA 4.83 4.92 4.2 4.89 4.33 4.38 NA ...
## $ review_scores_checkin : num 4.8 4.75 NA 5 4.78 4.8 4.84 4.67 4.63 NA ...
## $ review_scores_communication : num 4.89 4.82 NA 5 4.71 4.4 4.8 4.5 4.63 NA ...
## $ review_scores_location : num 4.77 4.21 NA 5 4.7 4.2 4.9 5 4.88 NA ...
## $ review_scores_value : num 4.85 4.67 NA 4.83 4.72 4.4 4.71 4.33 4.25 NA ...
## $ reviews_per_month : num 1.21 0.39 NA 0.51 1.65 0.05 2.2 4.19 0.12 NA ...
```

A continuación convertiremos los tipos que no sean correctos:

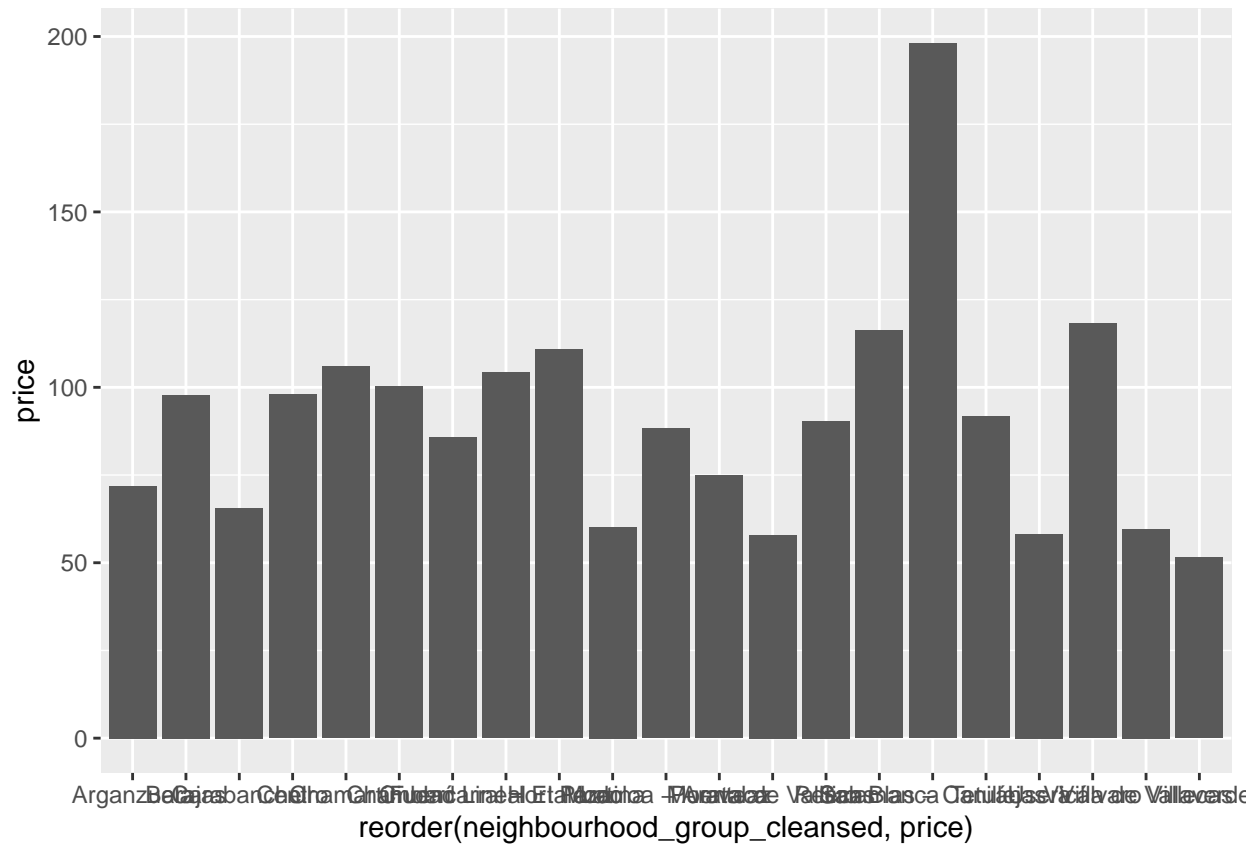
```
data$price<-as.numeric(substring(data$price, 2)) #Convertimos los precios en números
```

```
## Warning: NAs introducidos por coerción
```

```
# data$host_is_superhost %>%str_replace_all(c('f' = "0", "t"="1"))
class(data$host_is_superhost) = "Logical" #Convertimos esta variable a lógica. TODO Verificar si esta c
```

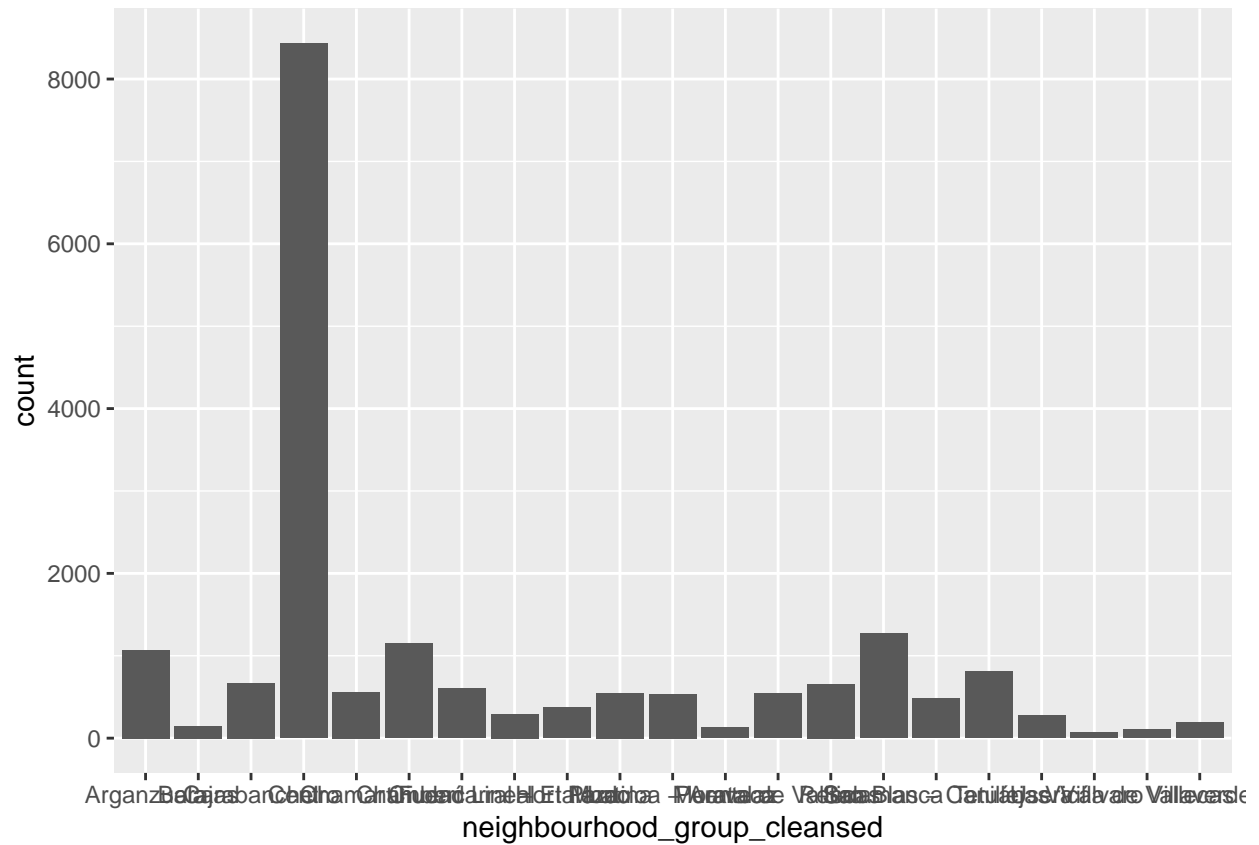
```
ggplot(data, aes(x= reorder(neighbourhood_group_cleansed, price), y= price)) +
  geom_bar(stat = "summary", fun = "mean")
```

```
## Warning: Removed 272 rows containing non-finite values (stat_summary).
```



Esta gráfica no es clara, ya que están mezclados pisos y habitaciones. La intuición dice que en San Blas-Canillejas debe haber todo pisos, puesto que no es un lugar céntrico.

```
ggplot(data, aes(x=neighbourhood_group_cleansed)) + geom_bar()
```



Este gráfico revela la inmensa cantidad de AirBnbs que hay en el centro frente al resto de Madrid.

TODO:

Separar por tipos de Airbnb: piso o habitación