

MÁSTER EN DATA SCIENCE Y BIG DATA (FINANZAS)

PRÁCTICA DE ESTADÍSTICA

ENTREGABLE

Memoria del ejercicio en formato pdf/Word explicando bien el análisis realizado, gráficos, tablas, resultados, etc. Adicionalmente se deben entregar los códigos utilizados para poder ser ejecutados directamente por el profesor.

FECHA LÍMITE DE ENTREGA: 3 de diciembre de 2017

EJERCICIO

Los datos siguientes contienen los precios y otros atributos de unos 54.000 diamantes y se encuentran en <http://ggplot2.tidyverse.org/reference/diamonds.html>. También allí hay una descripción de lo que significan las variables. El objetivo del ejercicio es hacer un análisis descriptivo, con el fin de entender los datos y las posibles relaciones entre las variables.

Los datos están disponibles en R. Para ello basta con hacer lo siguiente:

```
install.packages("ggplot")  
  
library(ggplot)  
  
data(diamonds)
```

De manera orientativa:

- Estadísticos resumen, gráficos (histogramas, boxplot, pairs,...), ver las posibles relaciones entre las variables (p.e. carat-price).
- Analizar si hay diferencias de precio en función de la variable color (p.e. boxplot) o de otras variables.
- Realizar un contraste de hipótesis para determinar si la media de carat es igual para las categorías cut-Fair y cut-Good).