Developing data products pitch

Predicting the price of a diamond based on its carat size

Bruno Hoste

2-9-2020

Introduction

This is the pitch for the final peer graded assignment in the Developing Data Products class.

The goal of our little shiny app is to determine the price of a diamond based on its carat size.

The app can be found on following link:
http://brunohoste.shinyapps.io/Assignment/

The code (UI.R and server.R) can be found on github:
http://github.com/BrunoHoste/developing-data-products

Method

We started with the diamond dataset from the UsingR package. It has observations for carat size and price of 48 diamonds.

```
library(UsingR)
data("diamond")
dim(diamond)
```

```
## [1] 48  2
```

```
head(diamond)
```

```
##   carat price
## 1  0.17   355
## 2  0.16   328
## 3  0.17   350
## 4  0.18   325
## 5  0.25   642
## 6  0.16   342
```

Regression

The outcome of the regression model used for our predictions turns out to be quite satisfying: significant coefficients and an R-squared of almost 98%.

```
model <- lm(price~carat,data=diamond)
summary(model)
```

```
##
## Call:
## lm(formula = price ~ carat, data = diamond)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -85.159 -21.448  -0.869  18.972  79.370
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -259.63      17.32  -14.99   <2e-16 ***
## carat        3721.02      81.79   45.50   <2e-16 ***
```

Predictions

We therefore decided to use the model to make the predictions in our little shiny app. In the app, we also added a plot and regression line.

```
plot(diamond$carat,diamond$price,xlab = "Carat size",
                          ylab = "Diamond price (in USD)")
                    abline(model,col='red',lwd=2)
```