

# README

## Comparison between haircare products

To create Shiny App I scraped the website LookFantastic to retrieve data about four haircare products from the brands Cantu, Aveda, Shae Moisture and As I Am.

The aim of my project was to show which brand carries the best product to buy. In order to investigate this I analyzed the prices and the customer reviews of the four products.

What emerged from my analysis is that Cantu provides both the cheapest and the best reviewed product, which makes it the best one to choose.

## Data Acquisition

Here is the code I used to scrape the data from the website:

```
library(rvest)

library(stringr)

library(ggplot2)

library(shiny)

# CANTU

# scraping product price

price_ca <- read_html("https://www.lookfantastic.it/cantu-shea-butter-for-natural-hair-moisturizing-curl")
  html_nodes("#mainContent > div.newYorkProductPage_topRow > div.newYorkProductPage_lastColumn > div.newYorkProductPage_lastColumn")
  html_text()

price_ca <-gsub('[\t\n\ ]', '', price_ca)

price_ca <- str_replace_all(price_ca, "€", "")

price_ca <- str_replace_all(price_ca, ",", ".")

# scraping product rating

rating_ca <- read_html("https://www.lookfantastic.it/cantu-shea-butter-for-natural-hair-moisturizing-curl")
  html_nodes("#productReviewsComponent > div > div.productReviews_summary-columns > div.productReviews_summary-columns")
  html_text()

rating_ca <-gsub('[\t\n\ ]', '', rating_ca)

# AVEDA
```

```

# scraping product price

price_av <- read_html("https://www.lookfantastic.it/aveda-be-curly-curl-enhancer-200ml/10304622.html")%>%
  html_nodes("#mainContent > div.newYorkProductPage_topRow > div.newYorkProductPage_lastColumn > div.newYorkProductPage_lastColumn")%>%
  html_text()

price_av <-gsub('[\t\n\ ]', '', price_av)

price_av <- str_replace_all(price_av, "€", "")

price_av <- str_replace_all(price_av, ",", ".")

# scraping product rating

rating_av <- read_html("https://www.lookfantastic.it/aveda-be-curly-curl-enhancer-200ml/10304622.html")%>%
  html_nodes("#productReviewsComponent > div > div.productReviews_summary-columns > div.productReviews_summary-columns")%>%
  html_text()

rating_av <-gsub('[\t\n\ ]', '', rating_av)

# SHAE MOISTURE

# scraping product price

price_sm <- read_html("https://www.lookfantastic.it/shea-moisture-raw-shea-cupuacu-frizz-defense-gel-cream-200ml/10304622.html")%>%
  html_nodes("#mainContent > div.newYorkProductPage_topRow > div.newYorkProductPage_lastColumn > div.newYorkProductPage_lastColumn")%>%
  html_text()

price_sm <-gsub('[\t\n\ ]', '', price_sm)

price_sm <- str_replace_all(price_sm, "€", "")

price_sm <- str_replace_all(price_sm, ",", ".")

# scraping product rating

rating_sm <- read_html("https://www.lookfantastic.it/shea-moisture-raw-shea-cupuacu-frizz-defense-gel-cream-200ml/10304622.html")%>%
  html_nodes("#productReviewsComponent > div > div.productReviews_summary-columns > div.productReviews_summary-columns")%>%
  html_text()

rating_sm <-gsub('[\t\n\ ]', '', rating_sm)

# AS I AM

# scraping product price

price_aia <- read_html("https://www.lookfantastic.it/as-i-am-curling-jelly-coil-and-curl-definer-227g/10304622.html")%>%
  html_nodes("#mainContent > div.newYorkProductPage_topRow > div.newYorkProductPage_lastColumn > div.newYorkProductPage_lastColumn")%>%
  html_text()

price_aia <-gsub('[\t\n\ ]', '', price_aia)

price_aia <- str_replace_all(price_aia, "€", "")

```

```
price_aia <- str_replace_all(price_aia, ",", ".")

# scraping product rating

rating_aia <- read_html("https://www.lookfantastic.it/as-i-am-curling-jelly-coil-and-curl-definer-227g/
  html_nodes("#productReviewsComponent > div > div.productReviews_summary-columns > div.productReviews_
  html_text()

rating_aia <-gsub('[\t\n\ ]', '', rating_aia)
```

I then stored this data in vectors, and I manually created a vector containing the brands' names and a list containing the words "Prices" and "Ratings". I created a data frame containing my data.

```
##   prices ratings
## 1    8.45    4.73
## 2   26.45    4.50
## 3   12.45    3.50
## 4   11.45    4.31

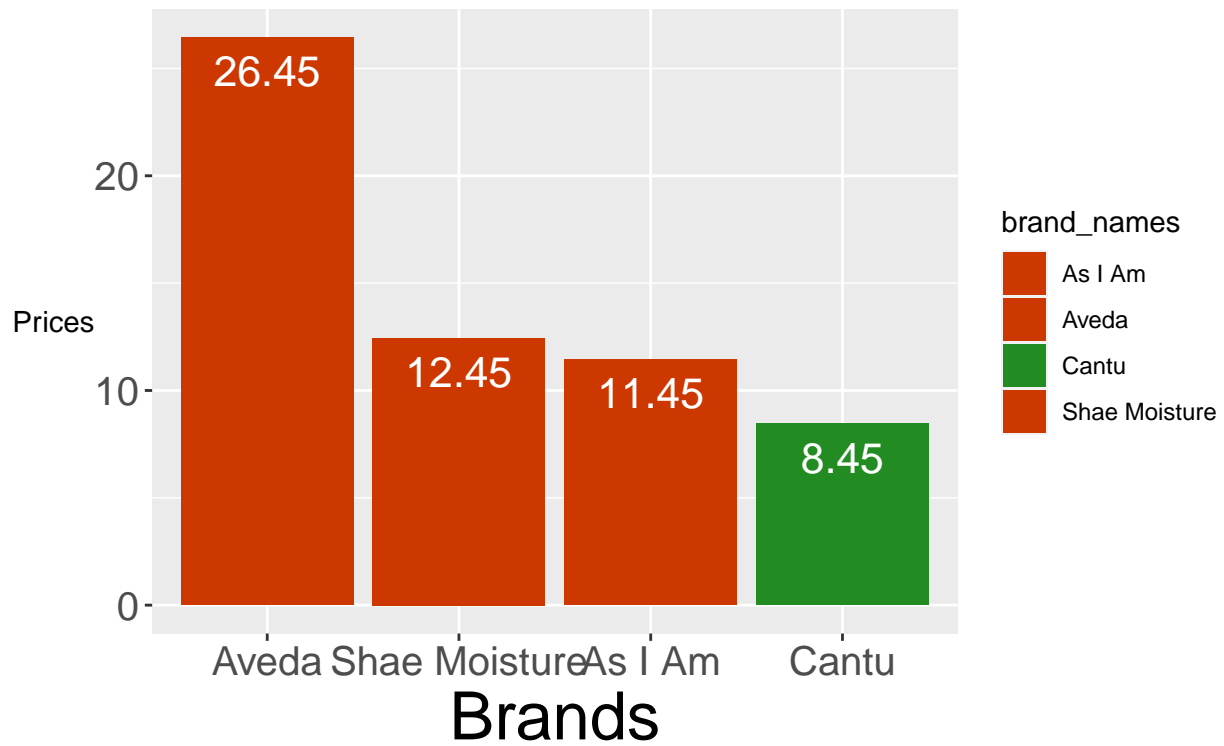
## [[1]]
## [1] "Prices"
##
## [[2]]
## [1] "Ratings"
```

## Data Visualization

To visualize the data I plotted two barplots using the ggplot2 library. One plot compares the prices of the four products and the other compares the ratings of the four products.

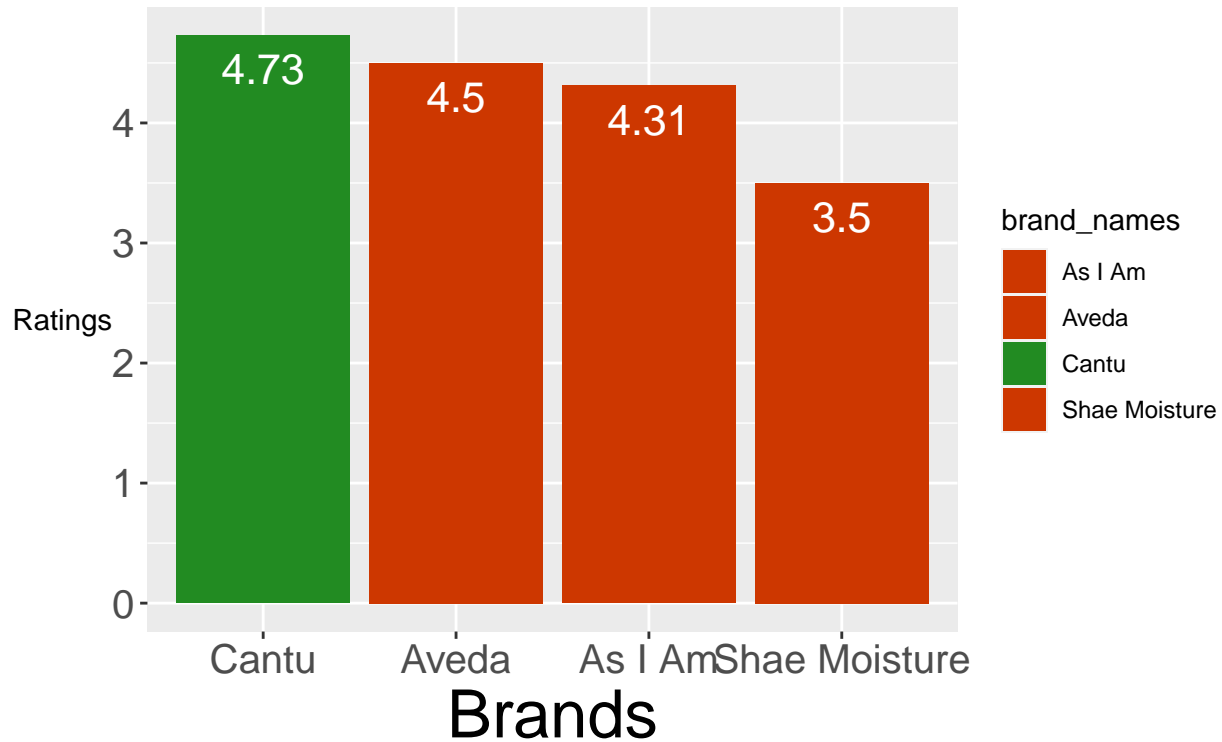
```
gg <- ggplot(mydata, aes(x = reorder(brand_names, -prices), y = prices, fill = brand_names)) + geom_bar
print(gg)
```

# Price comparison across bra



```
gg1 <- ggplot(mydata, aes(x = reorder(brand_names, -ratings), y = ratings, fill = brand_names)) + geom_bar()
print(gg1)
```

# Rating comparison across br



## Data Analysis

From the graphs it seems clear that Cantu carries the product with the lowest price and the highest reviews.

I decided to investigate this aspect even further and wrote two functions that return which is the brand that has the product at the lowest price and with the highest reviews.

```
best_price <- function(x){  
  minimum <- min(x)  
  return(minimum)  
}
```

```
lowest_price <- best_price(prices)
```

```
print(lowest_price)
```

```
## [1] 8.45
```

```
lowest_priced_brand <- function(x){  
  for (i in x){  
    if(i==8.45){print("The best priced brand is Cantu")}  
    else{break}  
  }  
}
```

```
lowest_priced_brand(prices)
```

```
## [1] "The best priced brand is Cantu"
```

```

best_rating <- function(x){
  maximum <- max(x)
  return(maximum)
}

highest_rating <- best_rating(ratings)

print(highest_rating)

## [1] 4.73

best_rated_brand <- function(x){
  for (i in x){
    if(i==4.7){print("The best reviewed brand is Cantu")}
    else{break}
  }
}

best_rated_brand(ratings)

```

## The Shiny App

In the app we find a box that we can use to select which aspect of the product we want to investigate between the price and the rating.

If we select Price the barplot that compares prices will appear, and underneath the result of the `lowest_priced_brand` function.

If we select Ratings the barplot that compares ratings will appear, and underneath the result of the `best_rated_brand` function.