

## Avocado seizoensprijzen analyse met Rhadoop



Door: Stephan, Sander, Michelle & Joshua  
Datum: 20-6-2020

# Inleiding

De prijzen van verschillende fruiten zijn over het algemeen seizoensgebonden i.v.m. de groeiperiodes van deze planten. Om erachter te komen hoe de verdeling van deze prijzen ligt en welke van de seizoenen het beste is om avocado's te kopen is er een R script ontwikkeld dat, in samenwerking met hadoop, de seizoenen achterhaald voor elke dag van het jaar en de gemiddelde prijs per seizoen berekend.

## Materiaal & Methode

De data voor dit onderzoek is op [deze site](#) te vinden, en bevat voor elke dag in het jaar: de datum, de gemiddelde prijs van een avocado, het totale aantal van avocado's dat is verkocht & de soort avocado dat is verkocht (conventioneel of organisch). Deze data zijn als csv bestand gedownload en ingelezen binnen R, waarna de data geannoteerd zijn met seizoensdata m.b.v. een custom functie. Vervolgens is per seizoen de gemiddelde avocado prijs berekend via een mapreduce in het R script.

## Resultaten

Om de avocado prijs te visualiseren in verband met de verschillende seizoenen is de volgende plot gemaakt:

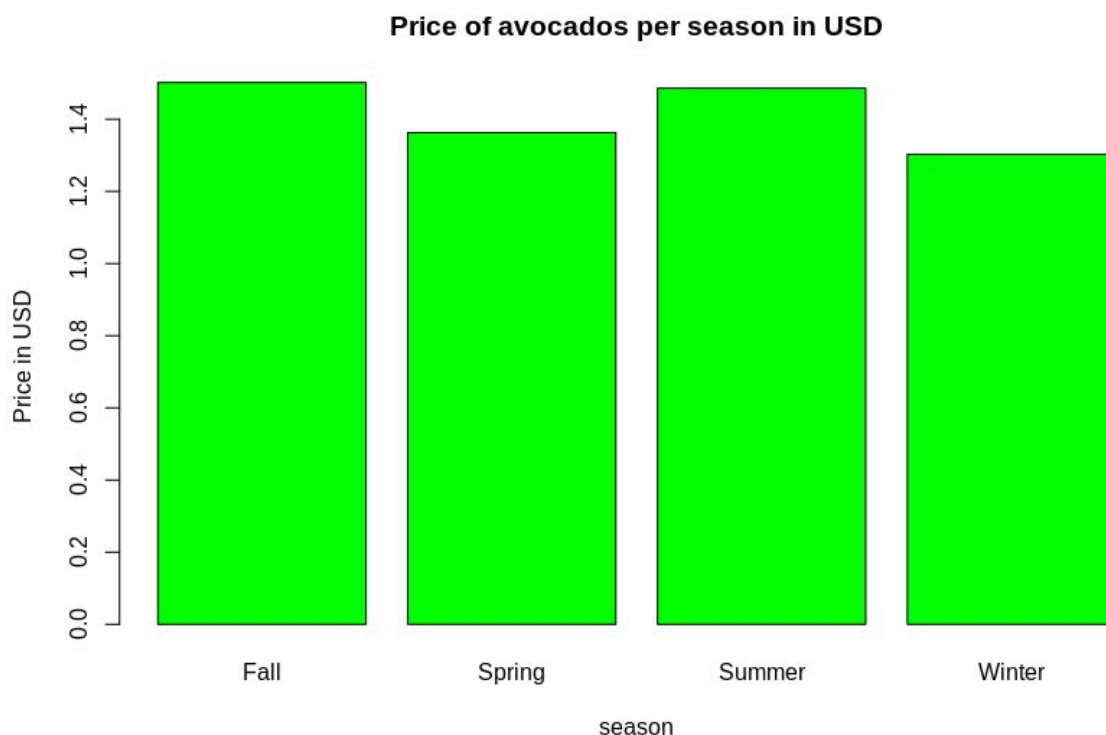


Figure 1: gemiddelde Avocado prijs per seizoen in US dollar.

# Conclusie

Met behulp van R, Rhadoop en hadoop is er succesvol een analyse uitgevoerd om het beste seizoen te vinden om avocado's te kopen. Uit Figuur 1 is af te lezen dat de avocado het goedkoopste is in de winter (1.28 US dollar) en in de lente (1.32 US dollar).