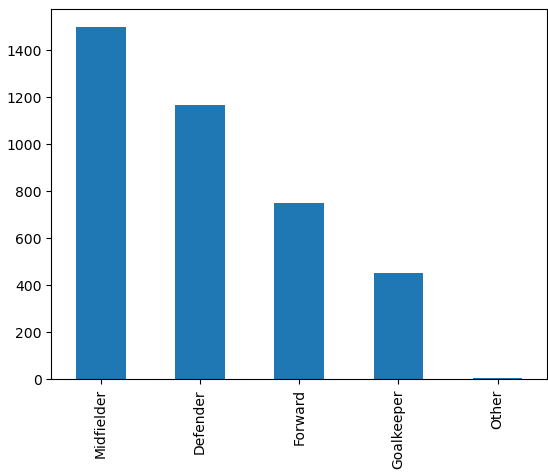
portfolio assignment 101

# Wat is datascience?

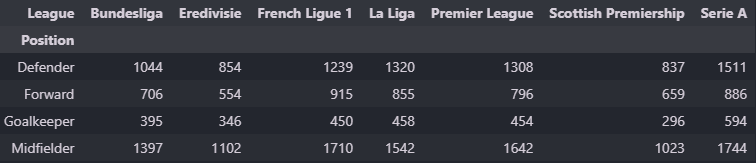
Data science is het vakgebied dat, door gebruik te maken van statistische en wiskundige methodes, kennis probeert te halen uit data. De data die hierbij gebruikt word, komt uit een dataset. De in de assignments gebruikte dataset is een excel bestand, met daarin de (voornamelijk aanvallende) statistieken van alle spelers in de top 5 competities in de jaren 2014 tot 2020.  
 Een van de eerder beschreven statische methodes is om de data om te zetten naar diagrammen. Een simpel voorbeeld van zo’n diagram is te vinden in Figuur 1. In deze simpele ‘bar plot’ is een overzicht gemaakt van het aantal spelers op elke positie. Naast statistieken kunnen er ook berekeningen worden gemaakt met de data. In Figuur 2 is zo’n berekening gemaakt. Deze berekening groepeert de rijen van de dataset op de waardes van 2 gegeven kolommen en telt vervolgens de groepen op. Het uitvoeren van deze functie resulteert in de tabel uit Figuur 3. Vervolgens kunnen deze gegevens gebruikt worden om in een diagram weer te geven. De combinatie tussen wiskundige en statistische methodes is dus erg belangrijk om de data inzichtelijk te kunnen maken.  
 Het doel van data science is het opdoen van nuttige, bruikbare kennis en inzichten uit de gegeven data. Deze kennis en inzichten kunnen worden gebruikt om bijvoorbeeld beslissingen te nemen, problemen op te lossen of prestaties te verbeteren van bedrijven.



Figuur 1: Bar plot van het aantal spelers per positie uit de gekozen dataset van assignment 4



Figuur 2: Wiskundige berekening die gecreëerd en gebruikt wordt in assignment 14



Figuur 3: Tabel die gevormd wordt na het uitvoeren van de functie gegeven in Figuur 2