Assignments 101/2

Marcello Haddeman 2152991

# Assignment 101

Data Science is de bestudering van data om hieruit conclusies te kunnen trekken. Door de assignments hebben we op verschillende manieren data verwerkt zodat we hierop analyse kunnen doen. In de eerste aantal assignments hebben we eerst gebruik gemaakt van verschillende grafieken om verschillende soorten data duidelijk te weergeven. Hiermee was het al mogelijk door een paar trekken. In de database van Video Games die ik heb gekozen kon ik bijv. in assignment 6 de conclusie maken dat de critic scores van games gemiddeld rond de 7 liggen, en dat cijfers onder 3.5 uitliggers zijn. Door gewoon een simpel grafiekje te gebruiken heb ik dus duidelijk kunnen zien dat cijfers rond de 5 al heel slecht zijn, terwijl je zou verwachten dat 5 najuist het middelpunt zou zijn.

In de latere hoofdstukken hebben we met decision trees en algoritmes gewerkt om meer geadvanceerde data op te stellen. Dit kan later heel handig zijn voor de herkenning van het soort klanten wat een bedrijf misschien zou hebben, we kunnen hiermee ook bijvoorbeeld churn proberen te voorkomen.

# Assignment 102

Een data scientist heeft de taak om voor een bedrijf conclusies en beslissingen op te stellen voor een bedrijf op basis van de ruwe data die het bedrijf heet. Dit begint door de ruwe data om te zetten naar duidelijk leesbare grafieken en tabellen. Iets wat we in de eerste aantal assignments hebben gedaan, hierna kan er begonnen worden aan het vergelijken van de data met elkaar om te kijken of er verbanden liggen, dit kan doormiddel door bijvoorbeeld een correlatie tabel of door scatter plots. Als laatst kan de data scientist decision trees en clustering op te zetten. Wanneer de data scientist klaar is met zijn analyse kan hij een voorstel opzetten bij andere teams van het bedrijf. Als de data scientist bijv. een klant kan categorizeren op basis van zijn gedrag dan kan hij dit geven aan het marketing team zodat ze specifieke marketing kunnen opzetten voor specifieke klanten.

Om even een voorbeeld te geven: Een nieuws website houdt bij wat voor soort artikelen een gebruiker bezoekt. Op basis van deze nummers kan worden gezien wanneer een klant een sports fan is, een aandeel hobbyist of een typische gebruiker. Met het programmeer team en marketingteam kunnen ze iets opzetten waar er advertenties worden laten zien op basis van de smaak van de gebruiker. Dit is dus een voorstel wat wordt opgezet doormiddel door de Data Scientist en dit wordt voorgelegd aan deze partijen. Deze partijen kunnen dan besluiten wat ze ermee doen.