Portfol

afrash ramjit

2151957

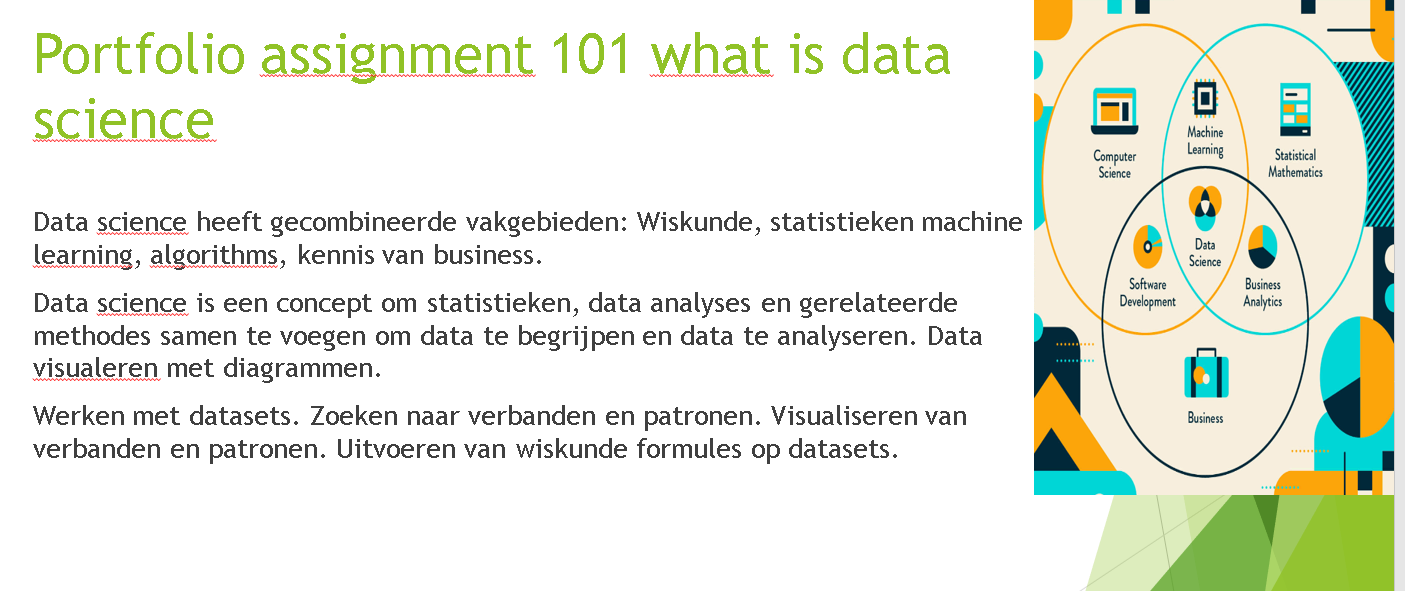
portfolio assignment 101

Afrash Ramjit: 2151957

# Inleiding

Ik zal een zin van de powerpoint pakken. Aan de hand van een zin van de powerpoint zal ik dan uitleggen wat data science is en welke assignment hierbij hoort of waarin dit aspect is uitgevoerd.

Hieronder staat de powerpoint naar waar verwezen wordt.



Inhoud

[Data science heeft gecombineerde vakgebieden: Wiskunde, statistieken machine learning, algorithms, kennis van business. 4](#_Toc66025351)

[Data science is een concept om statistieken, data analyses en gerelateerde methodes samen te voegen om data te begrijpen en data te analyseren. Data visualeren met diagrammen. 4](#_Toc66025352)

[Data science is Werken met datasets. Zoeken naar verbanden en patronen. Visualiseren van verbanden en patronen. Uitvoeren van wiskunde formules op datasets. 4](#_Toc66025353)

# Data science heeft gecombineerde vakgebieden: Wiskunde, statistieken machine learning, algorithms, kennis van business.

Dit kun je terug zien in eigenlijk alle opdrachten. In alle opdrachten wordt eigenlijk wel geprogrammeerd en worden er algoritmes toegepast. Ook word er bijna bij elke opdracht wiskunde toegepast zo kijken we bij opdracht 6 naar gekeken naar de min & max, de mean & medium, de standaard deviation en de outliers. Bij opgave 6 is dus een analyse uitgevoerd. Deze analyse heet de Univeriate analyse. Hierbij is dan ook gekeken naar de numerical data en de catgorical data. Dit komt van wiskunde af. Bij univeriate analyse kijk je naar de data per kolom. In deze opdracht keek ik dan naar lv 50 Hp kolom, lv50 Atk, Stage en de attribuut kolom. Bij assignment 5 word ook een univeriate analyse toegepast op de life expectancy gefiltert op 2019.

Bij portfolio opdracht 8 word de confidence interval berekend en wordt er gekeken wat de confidence interval is als je dit berekent met minder data. De confidence interval wordt gedaan op de lv 50 Hp kolom en op de lvl 50 atk kolom.

Data science heeft met statistieken te maken omdat je in een aantal opdrachten gebruik maakt van wiskundige formules om naar statistieken te kunnen kijken. Bij opdracht 13 en 14 wordt er bijvoorbeeld een chi squared test uitgevoerd en wordt er dus gekeken wat de kans is dat er een correlatie is tussen 2 kollomen. Dit is dus kansberekening en hoort bij statistieken en wiskunde. Bij opdracht 15, 16, 17 en 18 wordt er een decision tree gemaakt en wordt deze geanalyseerd, ook word er data getrained en getest en dit behoort eigenlijk tot machine learning. Er is een bivariate analyse uitgevoerd bij de opdrachten: 10, 11, 12, 13, 14.

Bij opdracht 19 word machine learning ook toegepast en wordt er een algoritme toegepast om een cluster model te maken. Bij opdracht 19 word het k means algoritme gebruikt om dit te maken. De Silhouette Coefficient word berekend voor dit cluster en daarna wordt deze geplot.

# Data science is een concept om statistieken, data analyses en gerelateerde methodes samen te voegen om data te begrijpen en data te analyseren. Data visualeren met diagrammen.

Dit zie je terug in meerdere opdrachten. Er zijn verschillende analyses gedaan waaronder univariate analyses, bivariate analyses en multivariate analyses. Deze data is dan ook visueel weergeven in de verschillende portfolio opdrachten. In opdracht 3 en 6 zijn univeriate analyses gedaan. In opdracht 3 is dit gedaan op de pinguin dataset en bij opdracht 6 is dit gedaan op de digimon dataset. Dit is gedaan voor categorical data en numerical data. Hiervan is data gevisualiseerd met een histogram en in opdracht 7 wordt gekeken naar de distributie.

# Data science is Werken met datasets. Zoeken naar verbanden en patronen. Visualiseren van verbanden en patronen. Uitvoeren van wiskunde formules op datasets.

Het werken met datasets zie je terug in eigenlijk elke opdracht. In mijn portfolio is gebruik gemaakt van mijn eigen dataset de digimon dataset, de penguin dataset en de life expactancy dataset. Hier zijn dan verschillende analyses op gedaan en zijn er bepaalde verbanden en patronen gevisualiseerd.

Bij opdracht 9 wordt gekeken naar de correlatie, hier kijk je dus naar de verbanden die kollommen met elkaar kunnen hebben. Je kijkt hier dan naar de strongest positive correlation, de strongest negative correlation en de weakest correlation.

Bij opdracht 10 word gekeken naar de correlatie voor de digimon dataset. Correlatie heeft te maken met verbanden in data.

Bij opdracht 13 en 14 wordt op 2 verschillende datasets een bivariate analyse gedaan. Hier wordt dan gekeken naar 2 kollommen, en hier wordt dan gekeken hoe groot de kans is op een correlatie tussen de 2 kollommen. Dus er wordt nogmaals gekeken naar verbanden in de data.