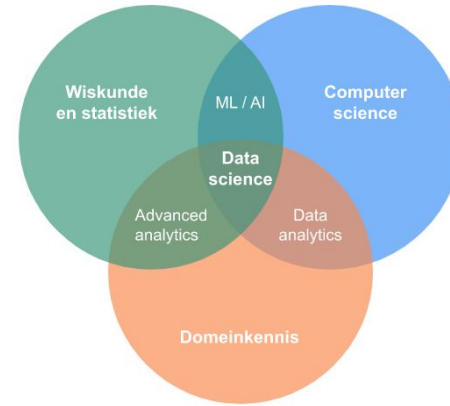


# What is Data Science?

Data science is het raakvlak tussen statistiek, informatica en domeinkennis. Vanuit de informatica kan data verzameld, geprepareerd en verwerkt worden in algoritmen waarmee patronen ontdekt kunnen worden.

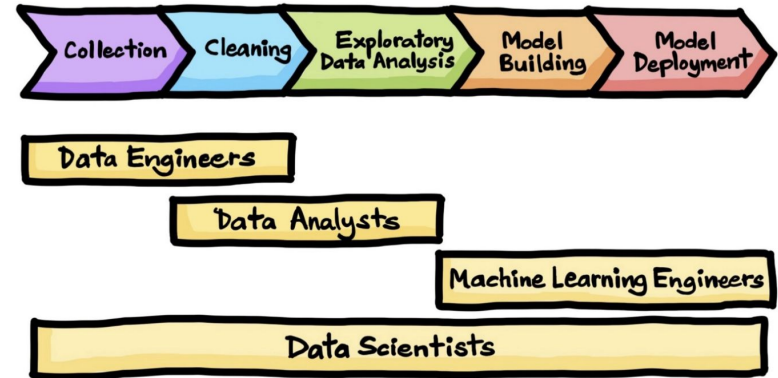
De wiskunde en statistiek zijn benodigd om te weten hoe modellen toegepast kunnen worden om waardevolle inzichten te delven uit de data, en de domeinkennis is cruciaal voor de vertaling van de inzichten naar de organisatie.



# What does a Data Scientist do?

Een Data Scientist is iemand binnen een bedrijf die data verzamelt, opschooft en analyseert en vervolgens rapporten schrijft met daarin de conclusies van het onderzoek.

- Domeinkennis en IT voor verzamelen.
- Informatica voor opschonen en data modeling.
- Wiskunde en statistiek voor data analyse.



# What tools does a Data Scientist use?

1. Matplotlib => Python-pakket voor het lezen, importeren en visualiseren van gegevens.
2. NumPy => Laat programmeurs toe te werken met complexe arrays en wetenschappelijke berekeningen uit te voeren.
3. Apache Hadoop => Open-source framework voor het opslaan en managen van grote datasets.
4. Excel => Excel heeft functionaliteit voor data science en is bij veel mensen al bekend, waardoor het een heel populaire optie is.
5. TensorFlow => Open-source library voor het trainen van machine-learning algoritmes



# What characteristics make these tools popular?

1. Matplotlib:  
Genereren grafiek en visualiseren
2. NumPy:  
Voegt datastructuren toe om code te versnellen.
3. Apache Hadoop:  
Data-verzameling/-management
4. Excel:  
Data opslaan, verschoonen en grafieken maken.  
Toegankelijker dan programmeren.
5. TensorFlow:  
Erg effectief in trainen van Machine/Deep learning

