



An Introduction to Data Science

Women Techsters Bootcamp

What to Expect in this Bootcamp

- An Introduction to Data Science Methodology
- Basics of Excel
- Introduction to Visualizations with Power Bi
- 2 Quizzes
- 1 Group Project

Course Outline

Module 1 :

- Intro to Data Science Methodology

Module 2:

- Introduction to Excel
- Introduction to Power BI

Questions?



What is Data Science?

Data Science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data - Wikipedia

[Source - Wikipedia](#)

What is Data Science?

Data Science has been one of the biggest tech buzz word for the last 6 - 10 years!

Need for Data Science

A DAY IN DATA

The exponential growth of data is undisputed, but the numbers behind this explosion - fuelled by internet of things and the use of connected devices - are hard to comprehend, particularly when looked at in the context of one day

500m

tweets are sent
every day
Twitter



4PB

of data created by
Facebook, including

350m photos
100m hours of video
watch time
Facebook Research

DEMYSIFYING DATA UNITS

From the more familiar 'bit' or 'megabyte', larger units of measurement are more frequently being used to explain the masses of data

Unit	Value	Size
b	0 or 1	1/8 of a byte
B	8 bits	1 byte
KB	1,000 bytes	1,000 bytes
MB	1,000 ² bytes	1,000,000 bytes
GB	1,000 ³ bytes	1,000,000,000 bytes
TB	1,000 ⁴ bytes	1,000,000,000,000 bytes
PB	1,000 ⁵ bytes	1,000,000,000,000,000 bytes
EB	1,000 ⁶ bytes	1,000,000,000,000,000,000 bytes
ZB	1,000 ⁷ bytes	1,000,000,000,000,000,000,000 bytes
YB	1,000 ⁸ bytes	1,000,000,000,000,000,000,000,000 bytes

*A lowercase "b" is used as an abbreviation for bits, while an uppercase "B" represents bytes.

65bn

messages sent over WhatsApp and
two billion minutes of voice and
video calls made
Facebook



463EB

of data will be created every day by 2025
BBC



95m

photos and videos are
shared on Instagram
Instagram Business

294bn

billion emails are sent
Radicati Group

320bn

emails to be sent
each day by 2021

306bn

emails to be sent
each day by 2020

3.9bn

people use emails

4TB

of data produced by a connected car
Intel

ACCUMULATED DIGITAL UNIVERSE OF DATA

4.4ZB

PeW

2013

44ZB

2020

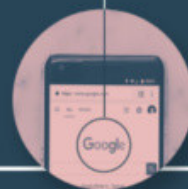
Searches made a day

5bn

Searches made
a day from Google

3.5bn

Smart Insights



28PB

to be generated from wearable
devices by 2020
Statista



Type of Data Science jobs

“Data scientist” is often used as a blanket title to describe jobs that are drastically different.

The Data Engineer



BIG DATA ENGINEER KEY RESPONSIBILITIES, SKILLS, AND TOOLSET

Responsibilities	Skills	Toolset
Performance optimization	Big data-processing architecture	Big data frameworks
Stream processing	Database optimization and data ingestion techniques	NoSQL technologies
Deploying ML models	general purpose and high-level programming languages: Python, R, SQL, and Scala:	ML platforms
Implementing ELT processes and maintaining big data pipeline	Data lake handling	Big Data querying tools



Data Analyst

INTERPRETS data and turns it into information which can offer ways to improve a business, thus affecting business decisions. Data Analyst gathers information from various sources and interpret patterns and trends.

MISSION - import, inspect, clean, transform, validate, collect, process or interpret collections of data to stay connected to customers, drive innovation, and product development.

What do they do?

- Analyse big data to reveal trends
- Apply statistical analysis techniques
- Collect ICT data
- Define data quality criteria
- Establish data processes
- Execute analytical mathematical calculations
- Handle data samples
- Integrate ICT data
- Interpret current data
- Perform data cleansing
- Perform data mining
- Report on research and data analysis

Where can they work?

- Work in an office environment or in the IT department of a company
- Work as service providers for companies

Use of ICT:

- Work on a personal computer
- Various software as demanded by the situation and the current data, from common spreadsheet applications up to business intelligence and specialized data analytical tools (Excel, SPSS, etc.)
- Visualization tools such as graphs, charts, and dashboards

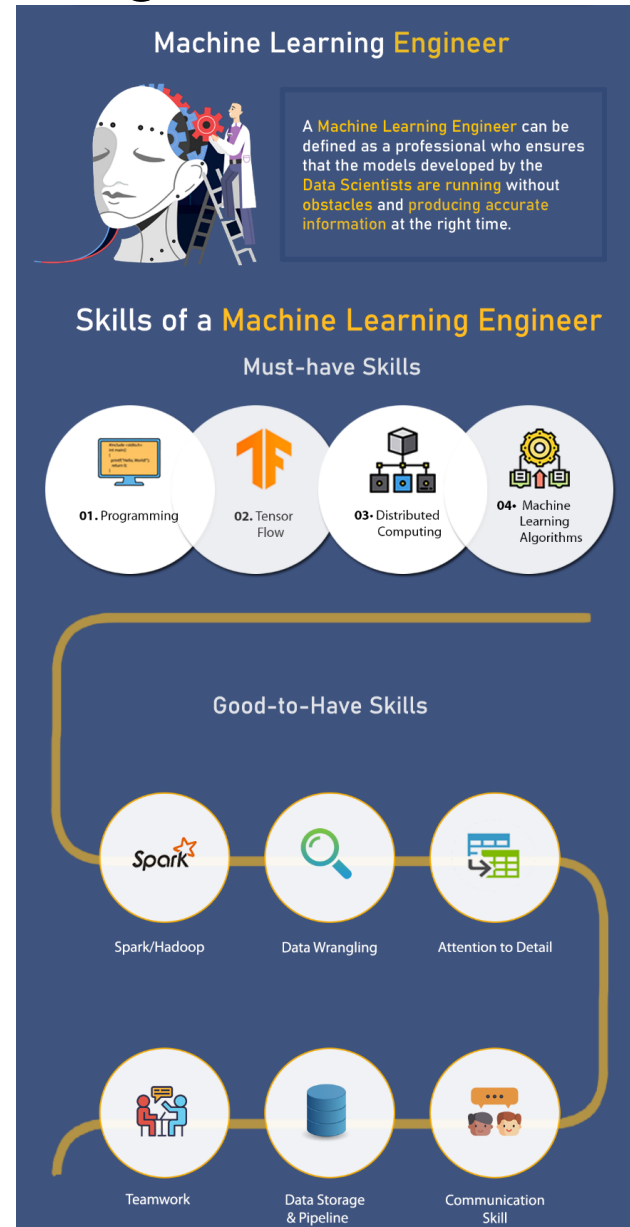


The Data Analyst

The Machine Learning Engineer



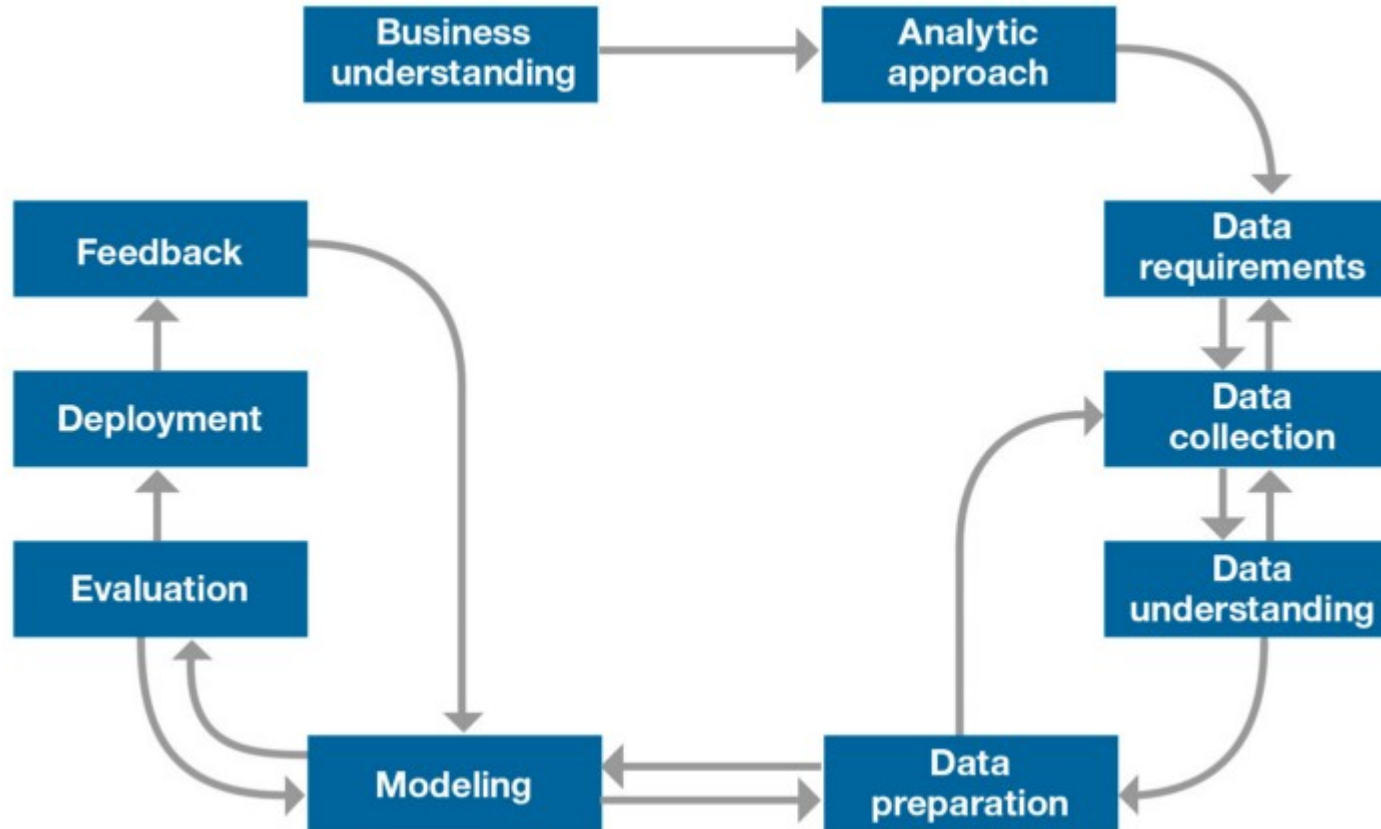
Machine Learning Engineer



Data Science Methodology

“Methodology in Data Science is the best way to organize your work, doing it better, and without losing time.”

Data Science Methodology



Source: [IBM](#)



Thank you for listening

Women Techsters Bootcamp

Questions?

