



Data Analysis Fundamentals

Lecture 1: Career Overview - Data Scientist

Vikas Kumar
Indian Institute of Technology Bombay

04th September 2023

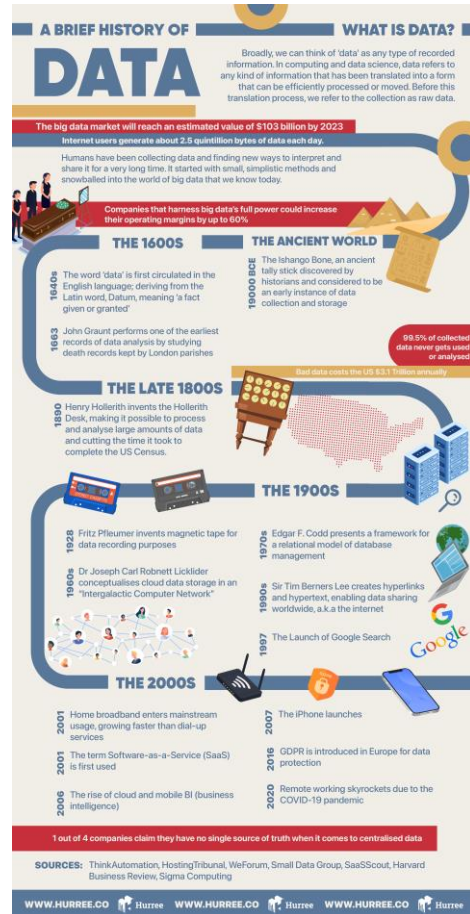
Outline

- A brief history of Evolution of Data
- Data - From a Decision Support System to Strategic Differentiator
- The Data Maturity Hierarchy
- Why choose Data Science as a career option?
- Typical job roles in Data Science
- The Data Science Pipeline
- Toolkit of a Data Analyst
- Your learning journey ahead!
- Excel - Quickest way to start learning Data Analysis

Data

- The Latin word data is the plural of datum meaning “given” or “to give”.
- The first English use of the word "data" is from the 1640s.
- The word "data" was first used to mean "transmissible and storable computer information" in 1946.
- The expression "data processing" was first used in 1954.
- The word data is synonymous with information.
- Data is a collection of discrete or continuous values that convey information, describing the quantity, quality, fact, statistics, other basic units of meaning.

A brief history of Evolution of Data



A Day in Data

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

DEMISTIFYING 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	byte	1 byte
Kb	kilobyte	1,000 bytes
MB	megabyte	1,000 ³ bytes
GB	gigabyte	1,000,000,000 bytes
Tb	terabyte	1,000 ⁴ bytes
Pb	petabyte	1,000,000,000,000 bytes
Eb	exabyte	1,000,000,000,000,000 bytes
Zb	zettabyte	1,000,000,000,000,000,000 bytes
Yb	yottabyte	1,000,000,000,000,000,000,000 bytes

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

463EB

of data will be created every day by 2025
IDC

95m

photos and videos are shared on Instagram
Instagram Business

294bn

billion emails are sent
Radial 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

65bn

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

28PB

to be generated from wearable devices by 2020
Statista

ACCUMULATED DIGITAL UNIVERSE OF DATA

4.4ZB

44ZB

For:

2019

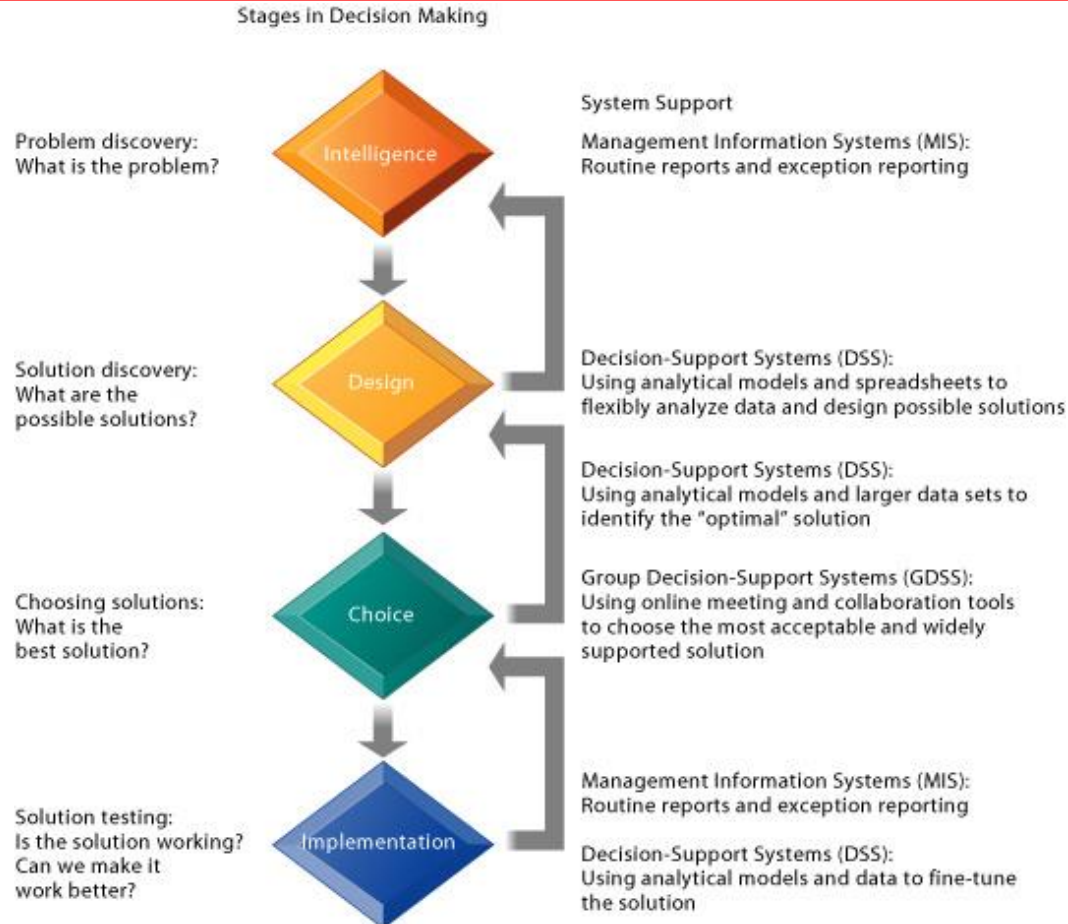
2020

Searches made a day **5bn**

Searches made a day from Google **3.5bn**



Data - From a Decision Support System to Strategic Differentiator



Data - From a Decision Support System to Strategic Differentiator

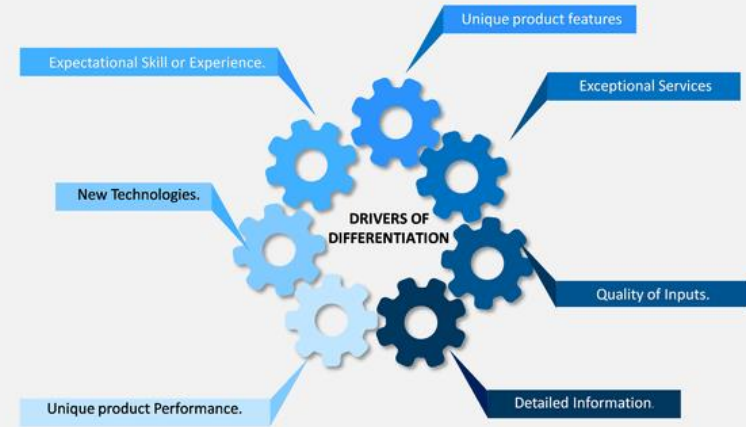
DIFFERENTIATION STRATEGY

Implementing a Differentiation Strategy

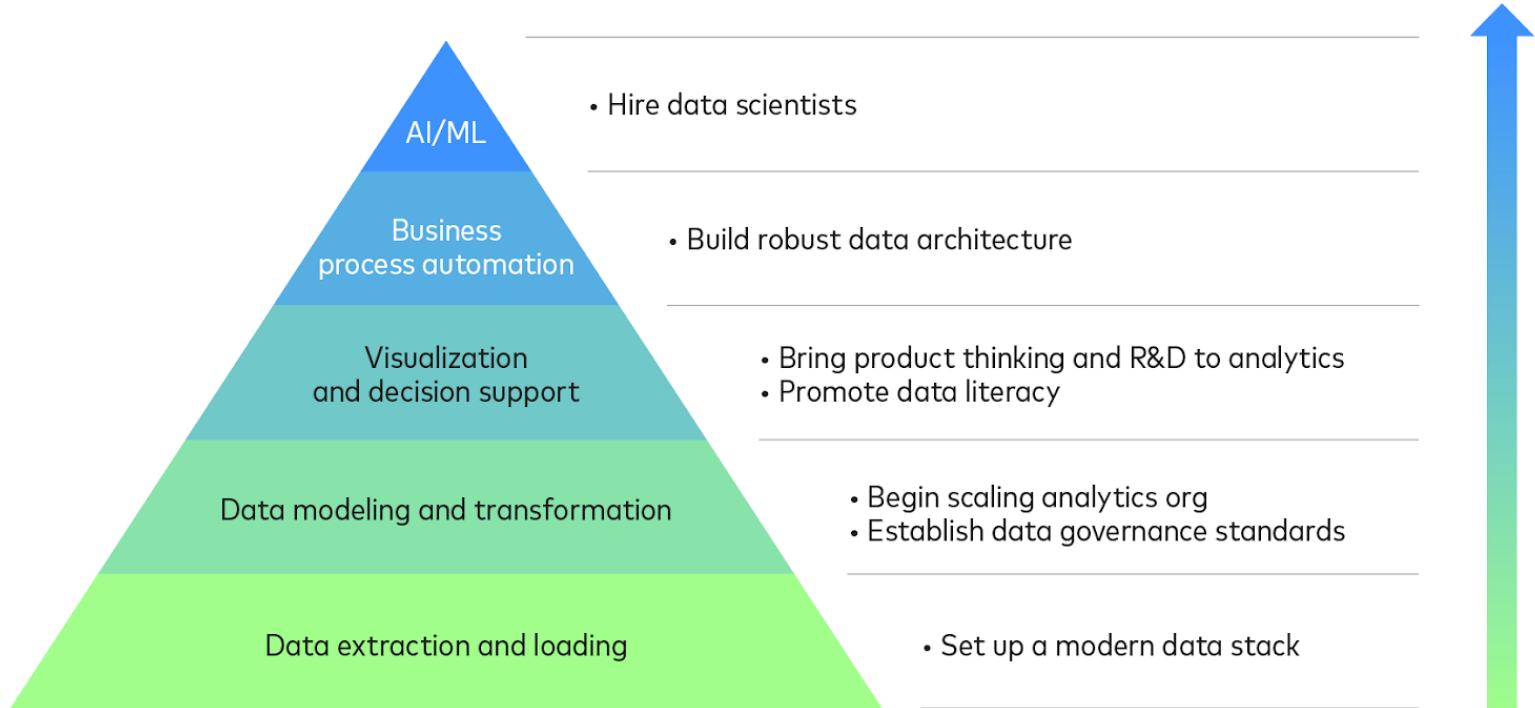


DIFFERENTIATION STRATEGY

Drivers of Differentiation



The Data Maturity Hierarchy



The Data Maturity Hierarchy

4 STAGES OF DATA MATURITY MODEL



**DATA
COLLECTION**



**DATA
PROFICIENCY**

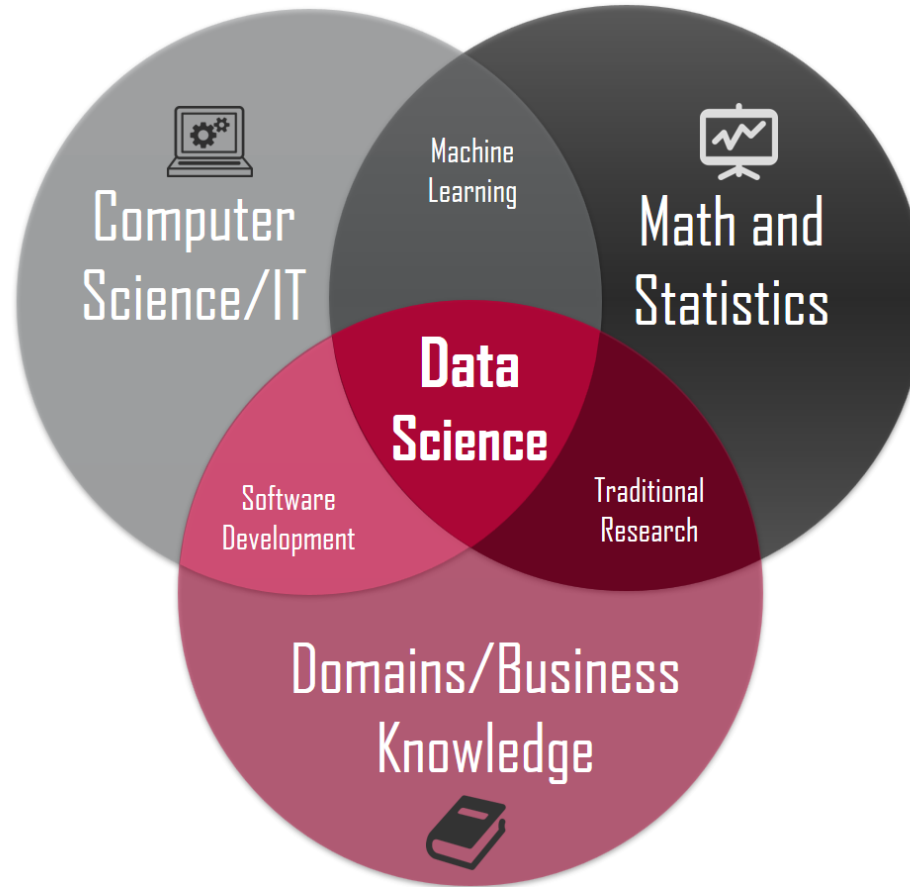


**DATA
CONSUMPTION**



**DATA
AS A CULTURE**

Data Science



Why choose Data Science as a career option?

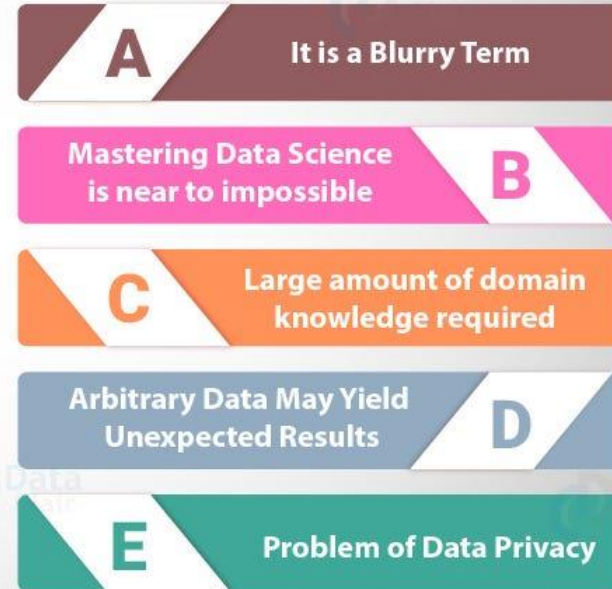


Data Science

Advantages



Disadvantages

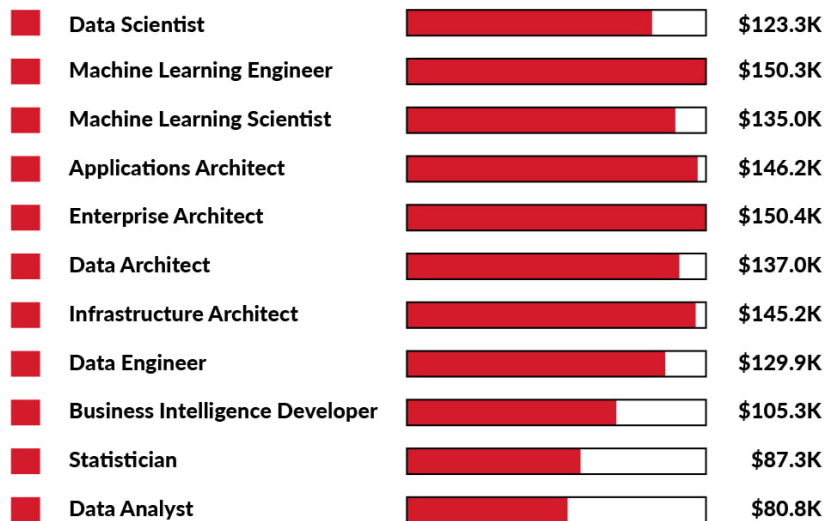


Typical job roles in Data Science



Northeastern University

Salaries of In-Demand Data Science Jobs



Source: Lightcast™ Analyst, 2023

www.northeastern.edu

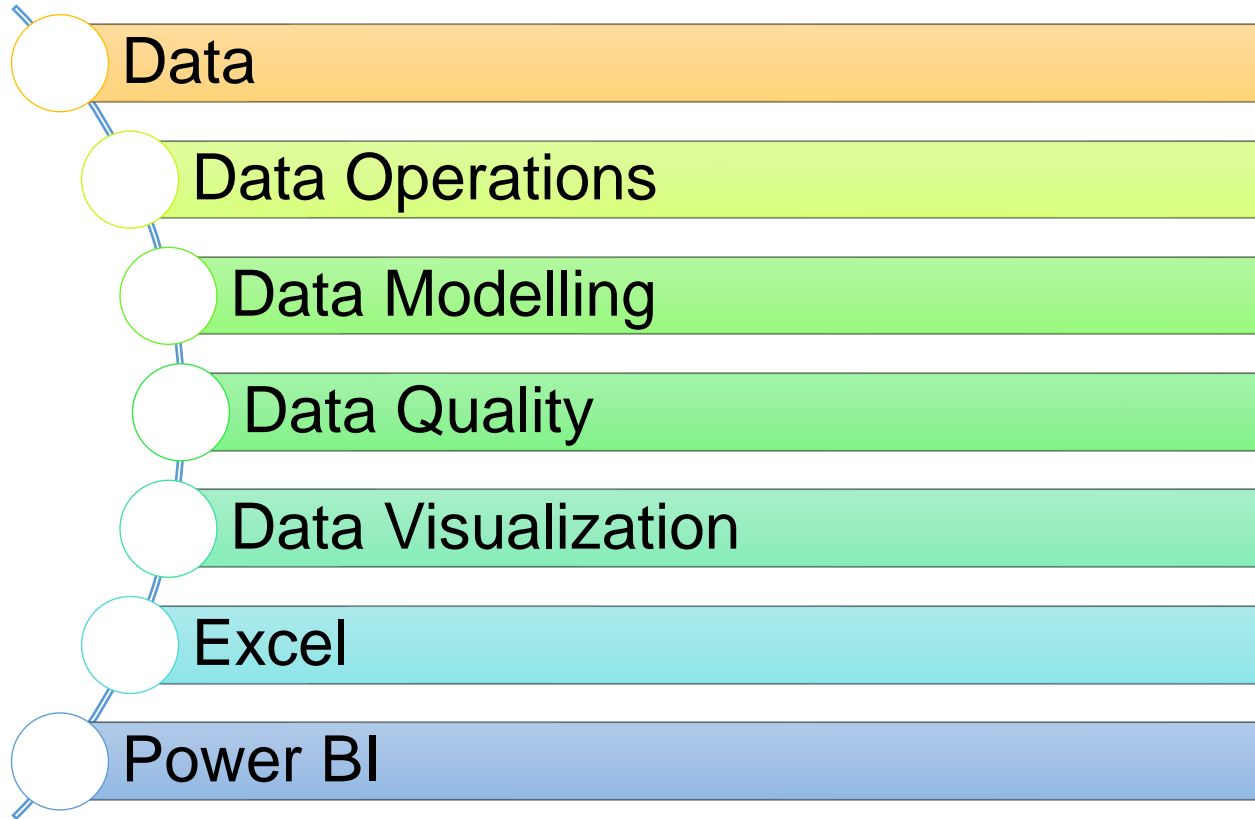
The Data Science Pipeline



Toolkit of a Data Analyst



Your learning journey ahead!



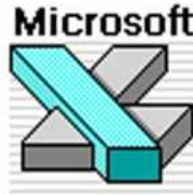
Excel - Quickest way to start learning Data Analysis

- Excel is a spreadsheet editor developed by Microsoft in 1985.
- It features calculation or computation capabilities, graphing tools, pivot tables, and a macro programming language called Visual Basic for Applications (VBA).
- Excel 2016 has 484 functions.
- Microsoft classifies these functions into 14 categories.
- Additional features are available using add-ins such as Analysis ToolPak, Solver Add-In.

Excel - Quickest way to start learning Data Analysis



1985 - 1990



1990 - 1995



1995 - 1999



1999 - 2003



2003 - 2007



2007 - 2010



2010 - 2013



2013 - 2019



2019 - NOW

10