Data Analytics

Welcome to Data Analytics

GENERAL ASSEMBLY

What You'll Learn Today

In this lesson, we'll:

- Outline goals, expectations, and logistics.
- Definition of data analytics vs other data fields
- Identify the skills and mindset of a successful data analyst.
- Discuss the discipline of data analytics, including topics such as data formats and data ethics.
- Identify tools and topics within data analytics

General Assembly is a global community of individuals empowered to pursue work they love.

30+
Global Campuses

70,000+
Course Alumni

500,000+
Workshop Attendees

20,000+
Expert Instructors

Meet Your Instructor

Ng Shu Min

Data Analytics and Data Science Instructor

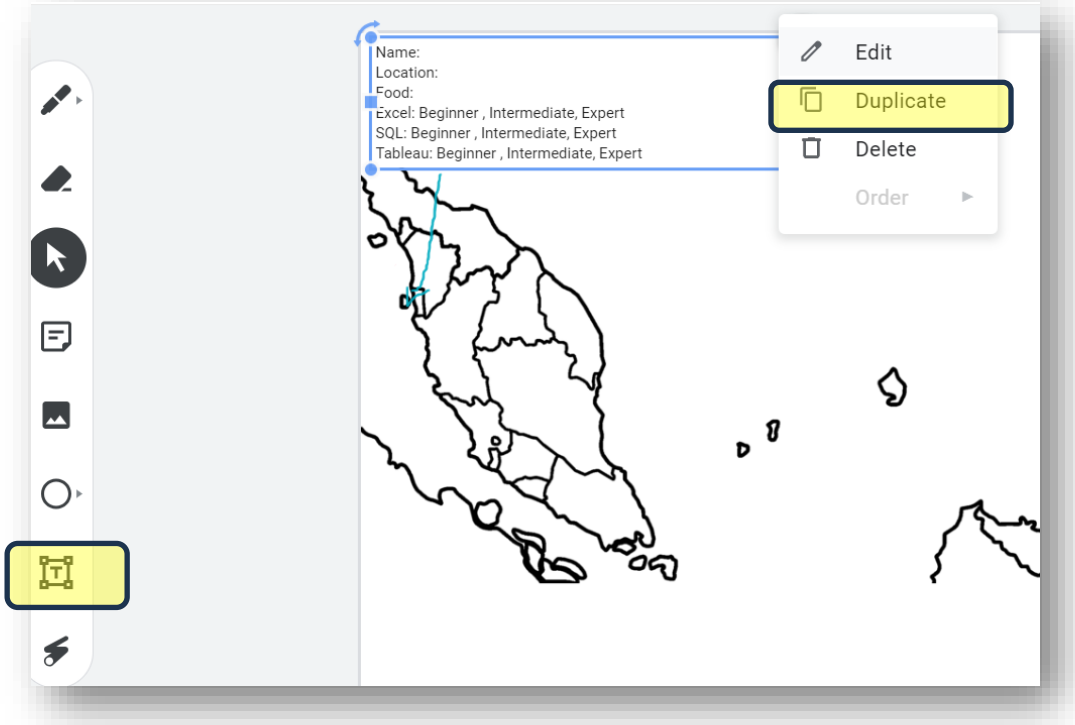
- Passionate educator with over 20 years of experience in teaching and training specialising in **Programming**, **Data Modelling, Data Analytics, Data Science**
- Industry experience in **Manufacturing**, **Media,** and **Education**
- Consulting in **Data Analytics** and **Data Science** for academia and business

Group Exercise:
Getting to Know You! (Ice Breaking!)

15 minutes

Introduce yourself and say hi to your classmates!

- Fill in this Google JamBoard: https://jamboard.google.com/d/1Xi5t__u6MjSN7Tdj2rywVl3lydpT6bn3Nc5f2eoaE0Y/edit?usp=sharing
- You can click on the text box and **Duplicate,** then fill in your info: **name, location and favourite food!**
- Also: your experience with Excel, SQL and Tableau



Orientation

No such thing as a silly question

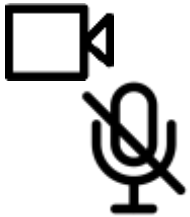
Slides and data files in Google Classroom

Your time - limit distractions

Last class – Fri 6 October

Orientation

Out of class support & working as a team



Camera on, mic off

Pulse Check

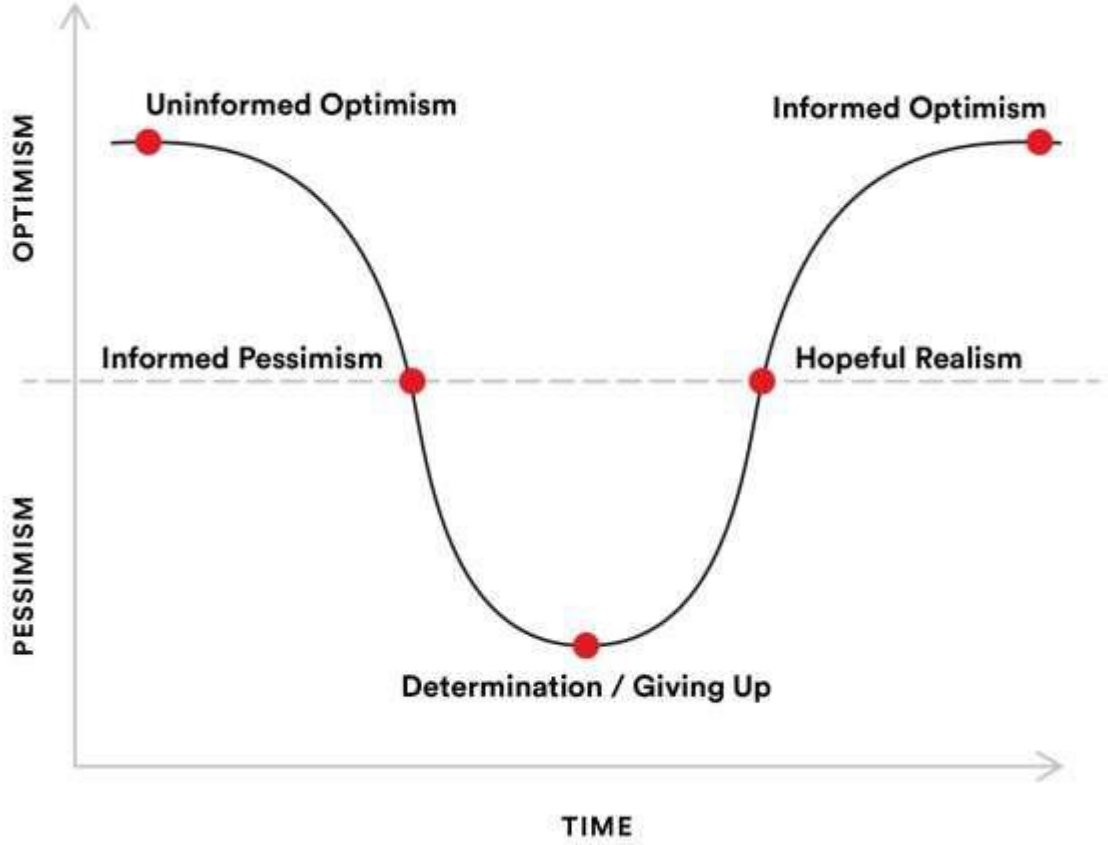
Timing and pace

At GA, we create norms for how we'll work together during the course.

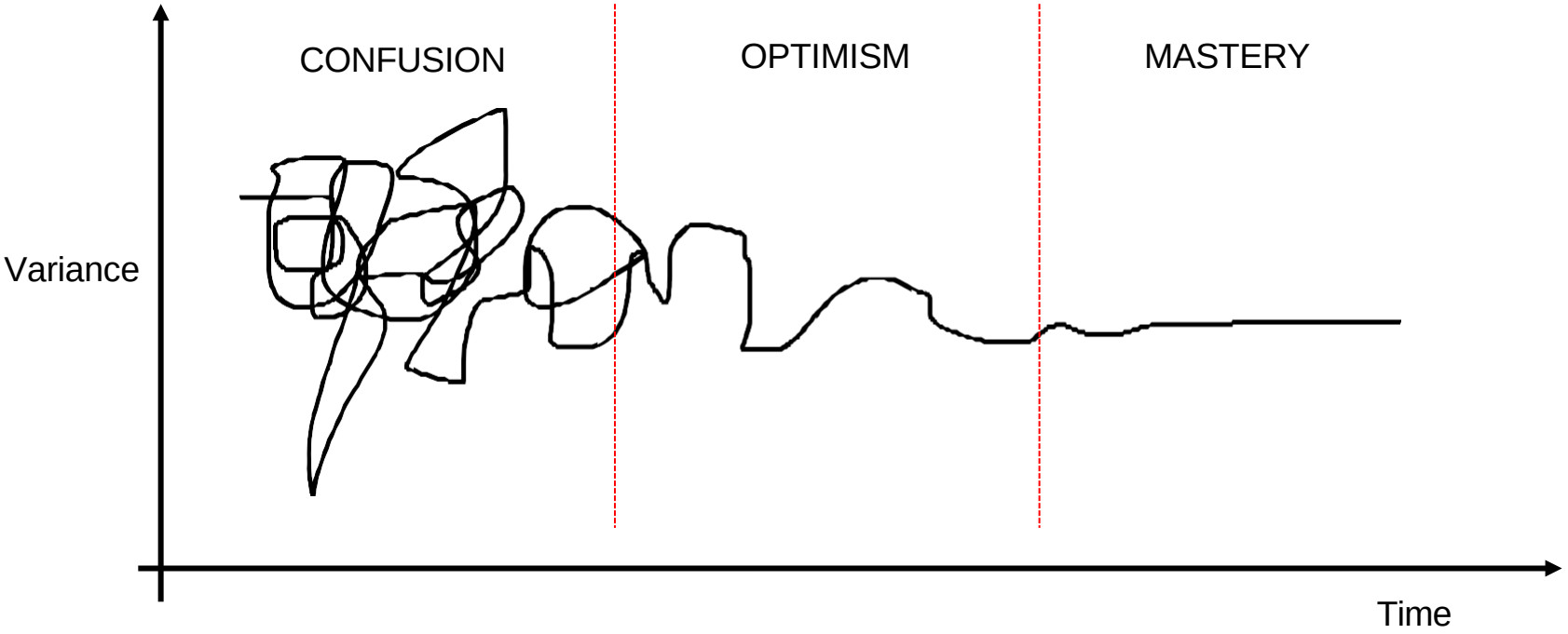
Check out the working norms. Is there anything else we should add to the list?

Be Present	**Contribute Constructively**	**Work Hard**
Ask Questions	**Be Supportive**	**Talk to Us!**

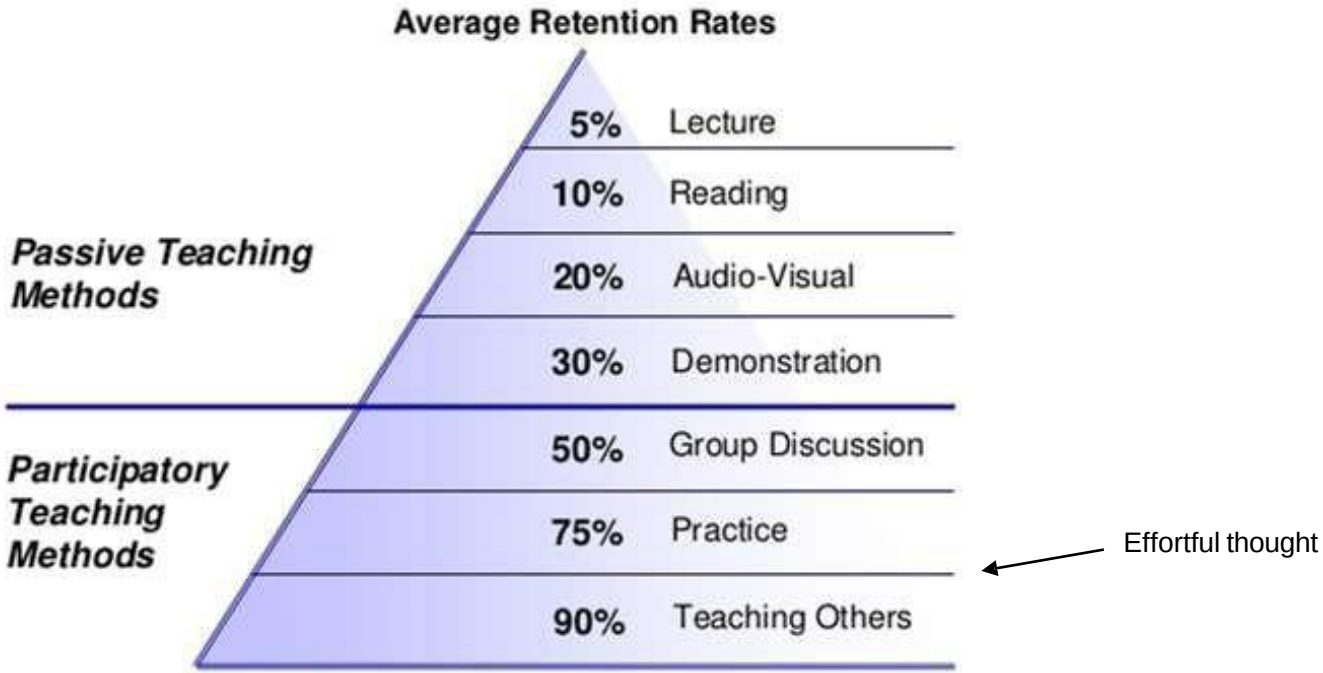
Buckle Up for the Journey Ahead!



Gaining new idea => the path



Our Teaching Method



Where We're Going |Lesson by Lesson

1	Introduction to Data Analytics
2	Introduction to SQL
3	Grouping and CASE WHEN in SQL
4	JOINs and Merges
5	Subqueries in SQL
6	Data Cleaning and Formulas
7	Referencing and Lookups
8	Aggregating Data With PivotTables

9	Communicating With Excel
10	Introduction to Tableau
11	Data Manipulation in Tableau
12	Dashboards in Tableau
13	Data Narratives in Tableau
14	**Final Presentation**

You'll Leave This Course Saying...

"I am no longer intimidated by rows and rows of data! I can combine, clean, and visualize data to gain insights into important business trends."

"I feel empowered to continue learning new techniques and acquiring new ways of working with data."

"Converting numbers into visually appealing, easy-to-understand visualizations is such a creative process. It's a lot of fun!"

Introduction to Data Analytics

Course Logistics

Here for You

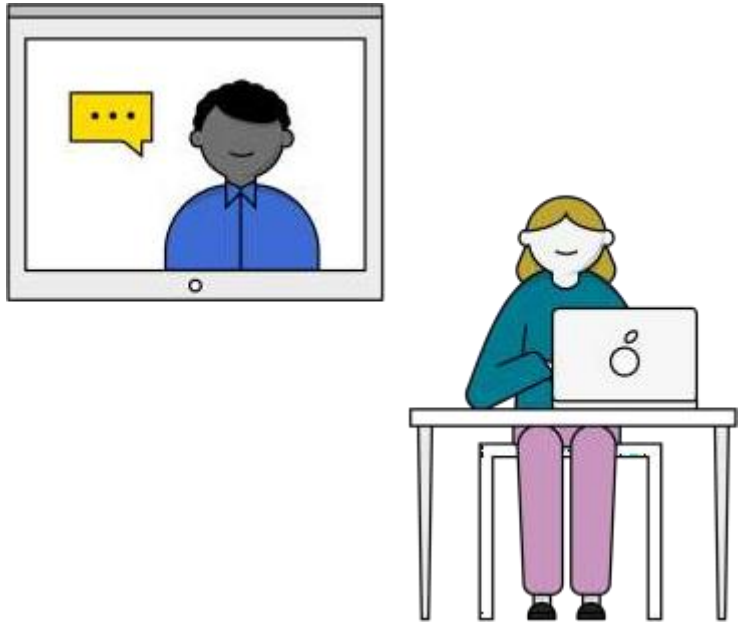
Office Hours

Support and advice

Day and time - refer to:

Google Classroom and Telegram

Used for course content, announcements, and class discussions.



Graduation Requirements

- Complete 80% of homework assignments.
- Maintain consistent attendance.
- Complete and submit the final project.

Homework

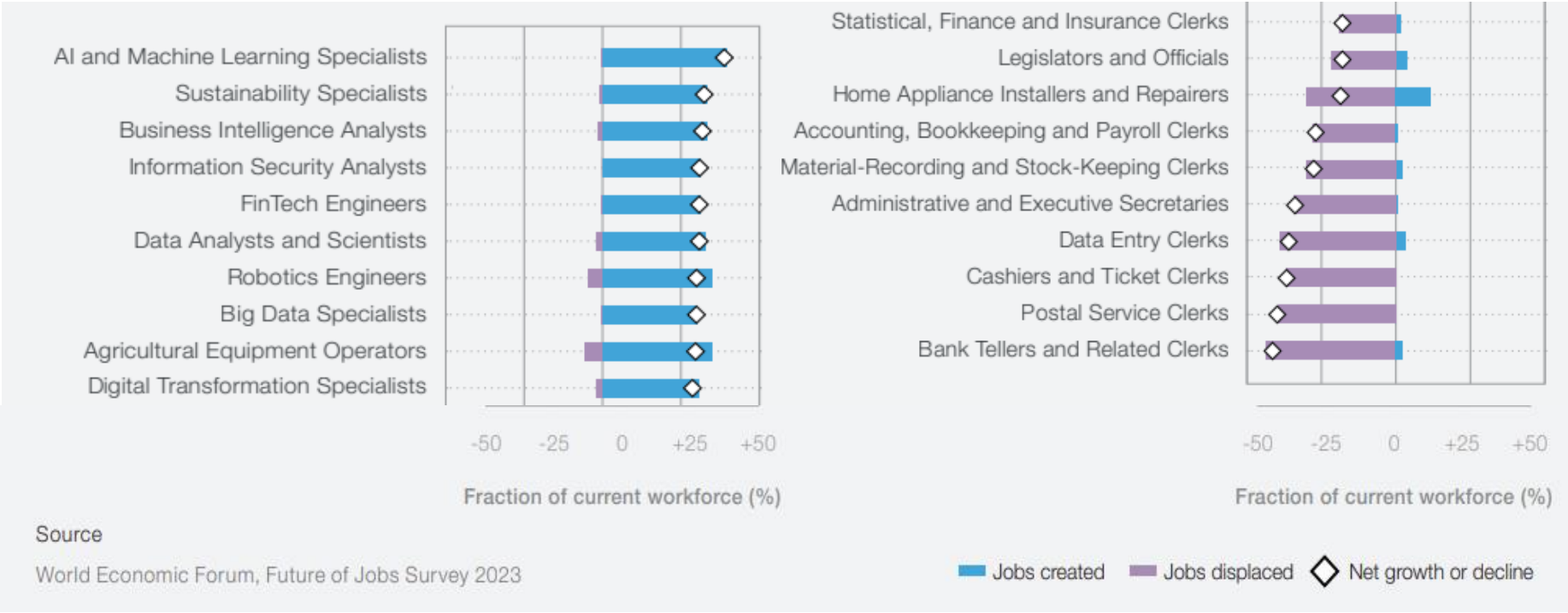
- Homework may be assigned at the end of each lesson. It generally consists of **pre-lesson work** in myGA (optional), additional practice as assigned by your instructor, and **project work** (required).
    - To submit your project work, you'll be provided with a link to your own Google drive and share it with your instructor.
    - This will be your work repository where you upload your project work and any assignments as required by your instructor.
- **Grading**: The pre-lesson work is **not graded**. Project work will be graded after your project presentation.

Final Project Presentation

At the end of the course, you'll give a **5–7 minute presentation** that should address the following:

- Who is your target audience?
- What problem are you solving?
- What is your solution and how does it solve the problem?
- What are some insights and trends you want to share, and why are they important?
- How do you tell the story of your data through data narratives and visualization?

New jobs and lost jobs, 2023 - 2027 (World Economic Forum Future of Jobs Report, 2023)



Source

World Economic Forum, Future of Jobs Survey 2023

Introduction to Data Analytics

The Role of a Data Analyst

Discussion:

What Does a Data Analyst Do?

15 minutes

Maybe you already know what a data analyst does on a daily basis, or maybe you're totally new to the field. Focus on 2–3 of the points below and share your experience:

- What does the typical work day of a DA look like?
- How does a DA collaborate with others?
- What is it like to work for a small startup versus a mid-sized or large company?
- What are some fun things about the job?
- What are some challenges?
- What's the most important skill to have as a DA? (This can also include soft skills.)

Data Analyst

The Storyteller

- Uses Excel, SQL (or NoSQL), Python, R, and visualization software like Tableau, Qliksense, and Power BI.
- Explores data and presents trends and insights.

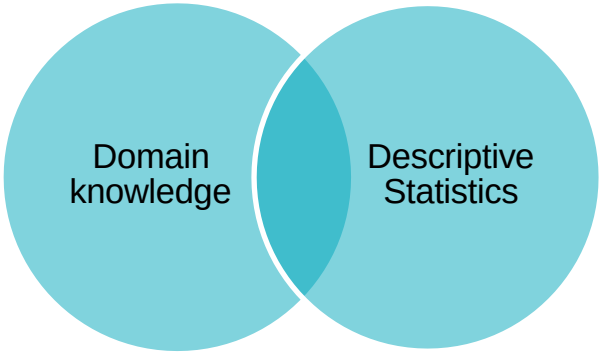
Data Scientist

The Wizard

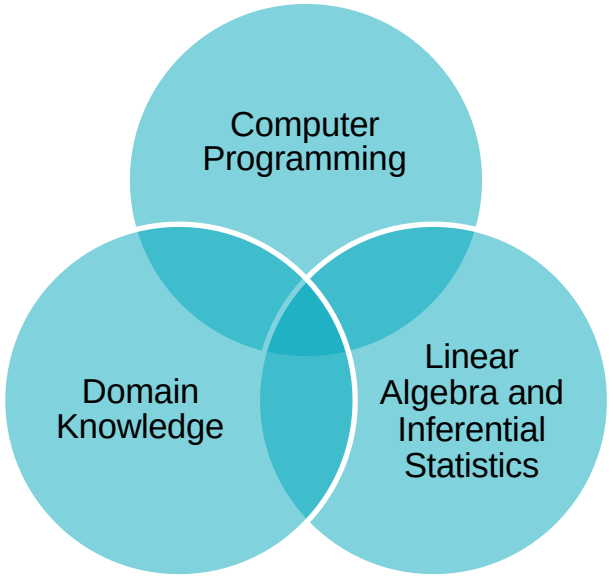
- Uses Python, R, SAS, SQL, Matlab, Hive, Pig.
- Works with data using algorithms, machine learning, and AI.
- Also tells data stories but with more of a focus math and coding.

Data Analytics vs Data Science

Data Analytics

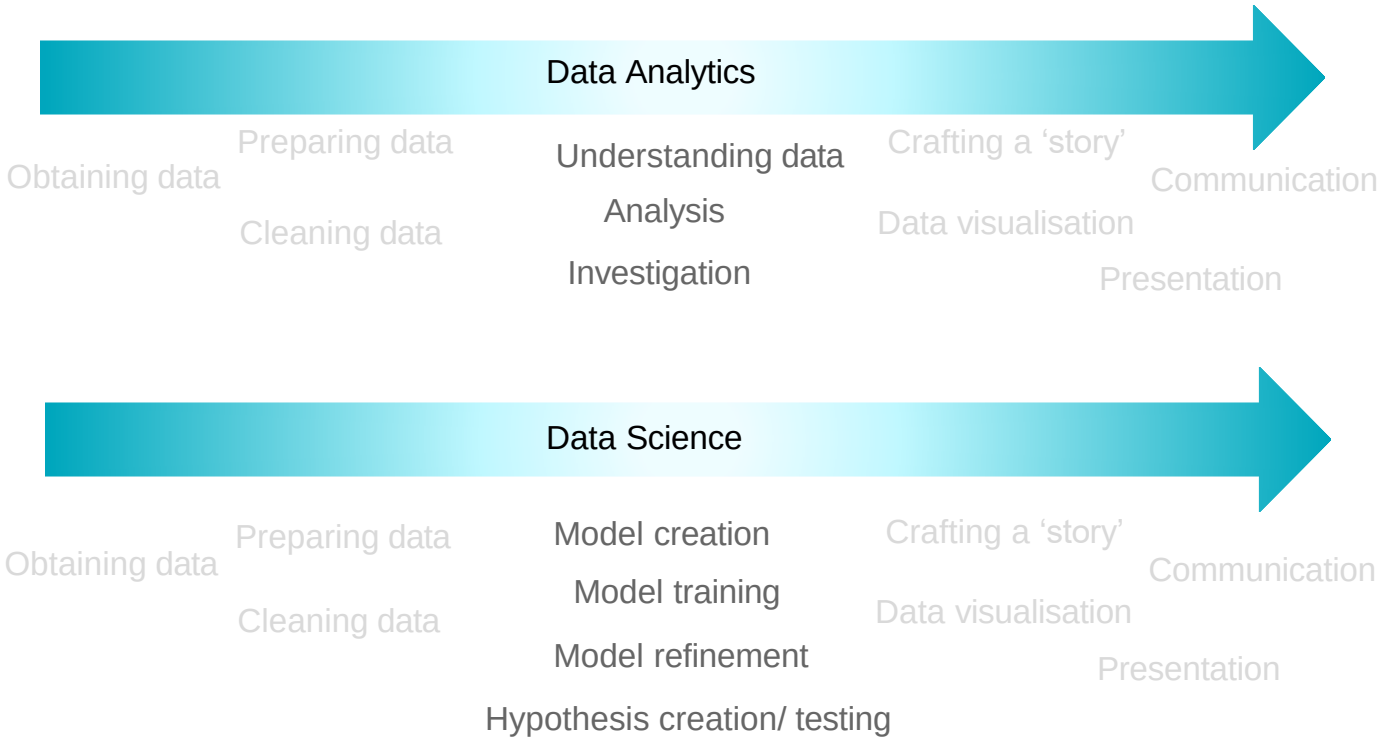


Data Science



Warning: A gross over-simplification!!

Data Analytics vs Data Science



Data Scientist

- Clean Data
- Massage Data
- Organize Data
- Build Digestible/Predictive /Prescriptive Models
- Build BI models

Data Analyst

- Collect Data
- Process Data
- Analyze Data
- Build Descriptive/Diagnostic Models
- Build BI models

Data Engineer

- Software Engineers
- Develop Database
- Prepare Data for analytics
- Maintain data sources
- Data lake, data warehouse owner

Data Architect

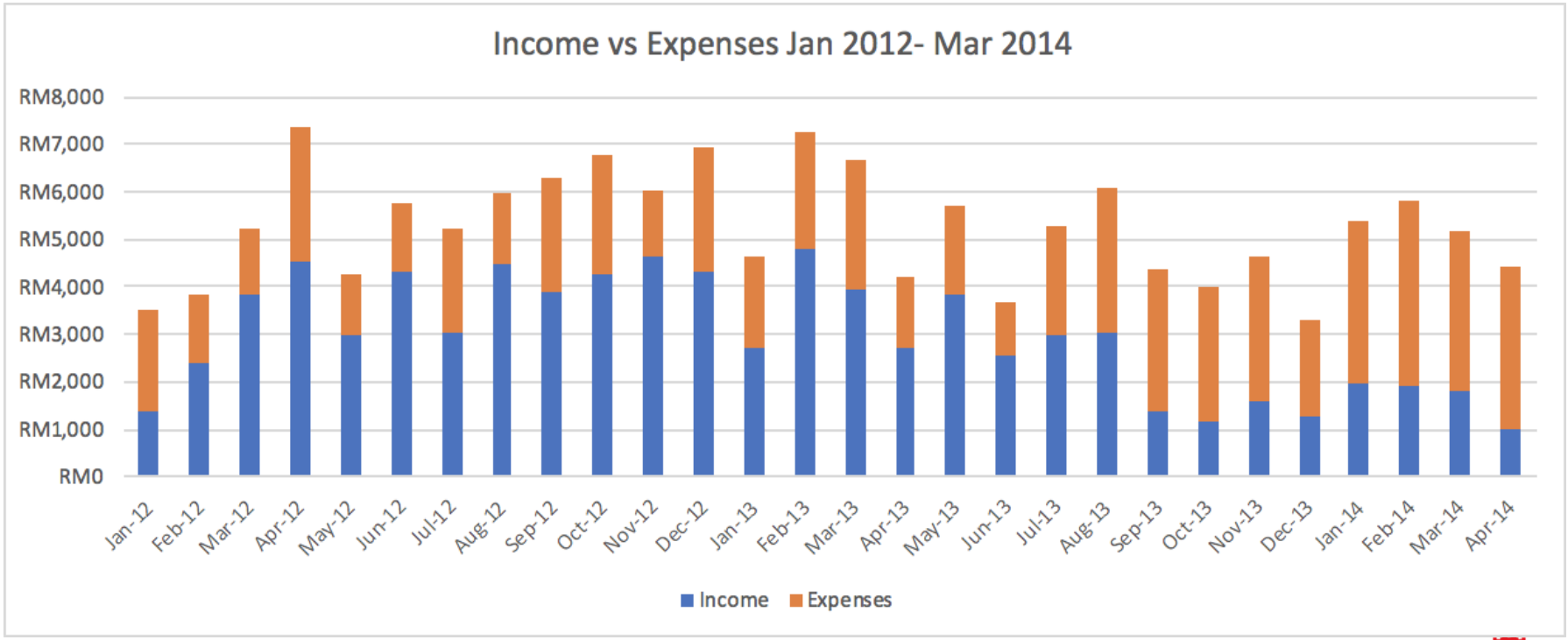
- Design and plan data architecture
- Plan data pipeline
- Plan infrastructure
- Domain expert who understands business problem

Data Admin

- Data Owner
- Ensures pipelines are in place
- Responses to any downtime

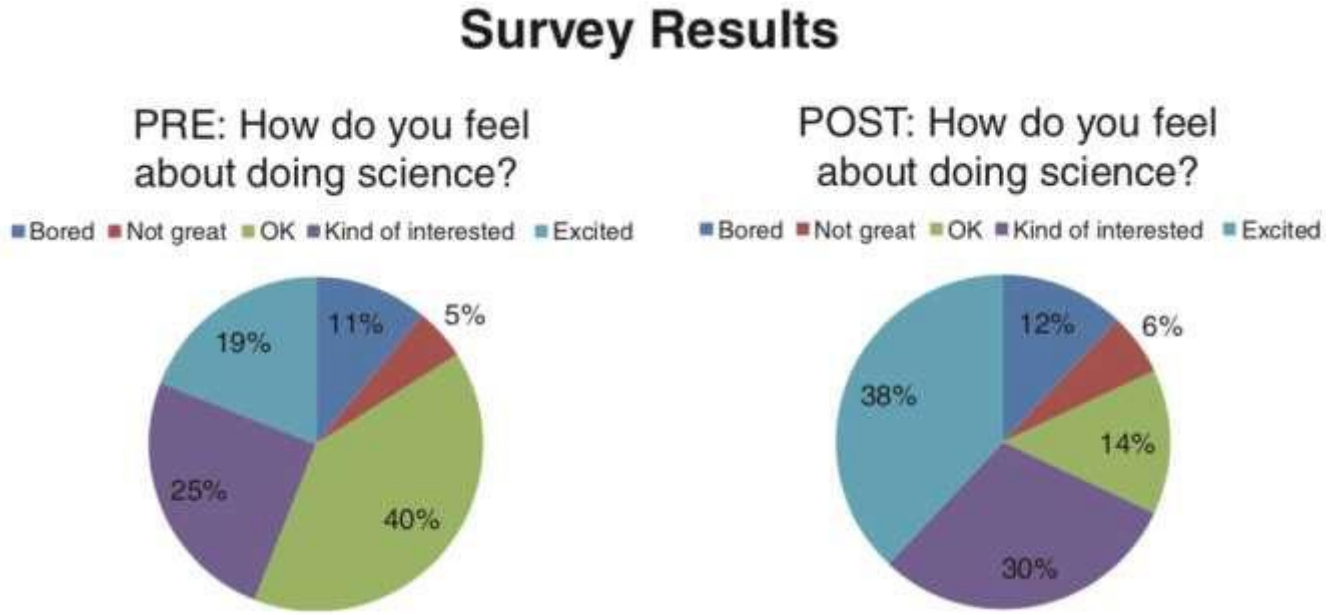
How many of us have encountered this?

Storytelling can make or break your organization



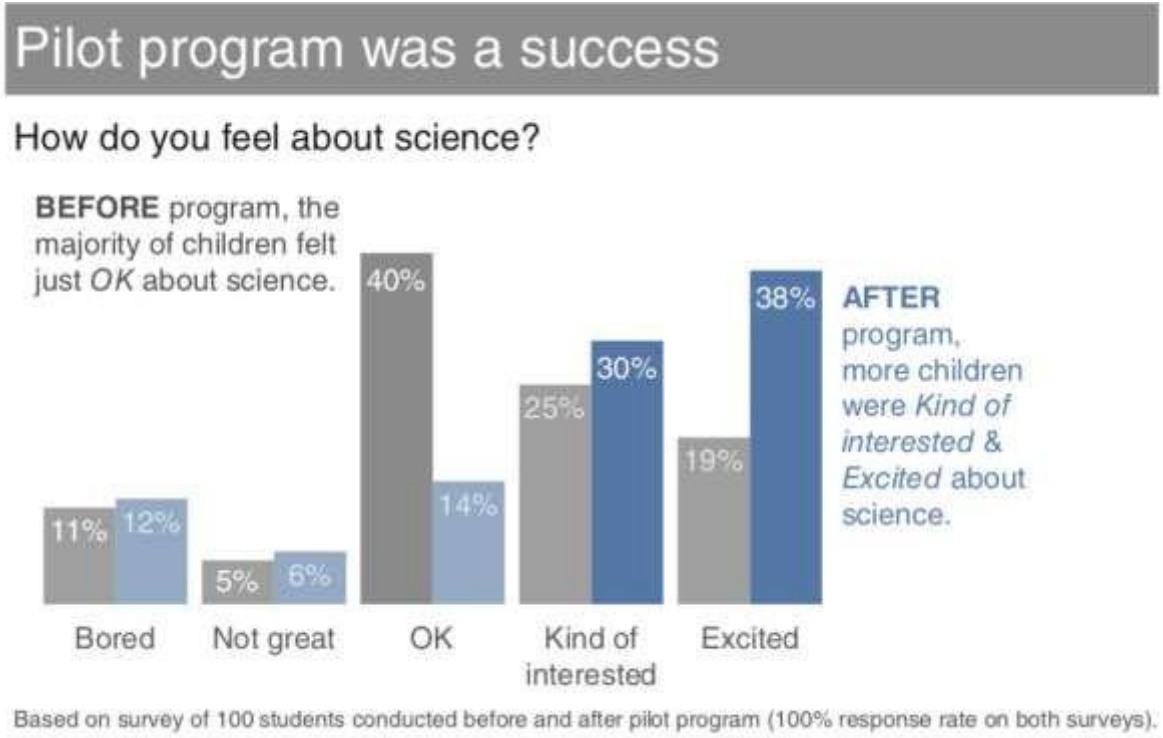
Stop showing the data

This charts shows me the data – but so what?



Start storytelling with data

Communicate insight through data



Developing an Analytical Mindset

What to Do	How to Do It
Be curious and ready to learn new things.	Ask lots of questions. Explore new techniques and new ways of looking at data.
Practice, practice, practice!	Find data sets online and take them through the DA Workflow, from framing all the way to communicating.
Follow what the data tells you.	Never twist your analysis to support your initial hypothesis.

Developing an Analytical Mindset (Cont.)

What to Do	How to Do It
Stay ORGANIZED!	Keep your spreadsheets tidy and easy to understand.
Document, document, document!	Maintain lists of what you delete or change when cleaning data, and comment your code.

So… What Does Data Actually Look Like?

	Customer City	Customer State		Customer Region		Customer Age Range		Customer Gender	Customer Income Range		Customer Education		Household Count		Housing Type	Customer Marital Status		Lifetime Value	
1	San Francisco	1	CA	1	West	1	below 20	1	Male	1	0 - 20,000	3	Graduate	4	4	2	Owner	2	Married	1	High
4	New York	3	NY	2	East	3	41-60	2	Female	5	80,001 upward	1	High School	3	3	2	Owner	3	Divorced	1	High
4	New York	3	NY	2	East	5	61 and above	2	Female	1	0 - 20,000	4	Other	1	1	2	Owner	3	Divorced	3	Low
5	Washington DC	4	DC	2	East	2	21-40	2	Female	4	60,001 - 80,000	4	Other	3	3	1	Renter	1	Single (Never r	2	Medium
5	Washington DC	4	DC	2	East	4	61-80	2	Female	2	20,001 - 40,000	1	High School	1	1	2	Owner	1	Single (Never r	3	Low
1	San Francisco	1	CA	1	West	5	61 and above	2	Female	5	80,001 upward	4	Other	7	7 or more	2	Owner	2	Married	1	High
4	New York	3	NY	2	East	2	21-40	2	Female	1	0 - 20,000	1	High School	7	7 or more	2	Owner	2	Married	1	High
4	New York	3	NY	2	East	3	41-60	2	Female	3	40,001 - 60,000	3	Graduate	7	7 or more	1	Renter	1	Single (Never r	2	Medium
1	San Francisco	1	CA	1	West	5	61 and above	1	Male	5	80,001 upward	3	Graduate	4	4	1	Renter	2	Married	1	High
2	Los Angeles	1	CA	1	West	1	below 20	2	Female	4	60,001 - 80,000	3	Graduate	3	3	2	Owner	1	Single (Never r	2	Medium
5	Washington DC	4	DC	2	East	2	21-40	2	Female	1	0 - 20,000	1	High School	3	3	2	Owner	1	Single (Never r	3	Low
7	Vienna	5	VA	2	East	1	below 20	1	Male	1	0 - 20,000	4	Other	5	5	1	Renter	3	Divorced	2	Medium
3	Pittsburg	2	PA	2	East	4	61-80	1	Male	2	20,001 - 40,000	3	Graduate	3	3	1	Renter	1	Single (Never r	3	Low
4	New York	3	NY	2	East	1	below 20	1	Male	2	20,001 - 40,000	2	Undergraduate	1	1	1	Renter	2	Married	2	Medium
1	San Francisco	1	CA	1	West	3	41-60	1	Male	3	40,001 - 60,000	4	Other	3	3	1	Renter	1	Single (Never r	2	Medium
1	San Francisco	1	CA	1	West	5	61 and above	2	Female	3	40,001 - 60,000	4	Other	2	2	1	Renter	3	Divorced	2	Medium
2	Los Angeles	1	CA	1	West	2	21-40	2	Female	5	80,001 upward	3	Graduate	2	2	1	Renter	3	Divorced	1	High
4	New York	3	NY	2	East	4	61-80	1	Male	2	20,001 - 40,000	4	Other	1	1	2	Owner	2	Married	2	Medium
5	Washington DC	4	DC	2	East	2	21-40	1	Male	1	0 - 20,000	3	Graduate	7	7 or more	1	Renter	3	Divorced	1	High
4	New York	3	NY	2	East	1	below 20	2	Female	3	40,001 - 60,000	1	High School	4	4	2	Owner	2	Married	1	High
3	Pittsburg	2	PA	2	East	1	below 20	1	Male	5	80,001 upward	3	Graduate	5	5	2	Owner	2	Married	1	High
5	Washington DC	4	DC	2	East	5	61 and above	1	Male	3	40,001 - 60,000	2	Undergraduate	1	1	1	Renter	3	Divorced	2	Medium
4	New York	3	NY	2	East	2	21-40	2	Female	3	40,001 - 60,000	3	Graduate	4	4	2	Owner	2	Married	1	High
1	San Francisco	1	CA	1	West	5	61 and above	1	Male	3	40,001 - 60,000	2	Undergraduate	3	3	2	Owner	2	Married	1	High
1	San Francisco	1	CA	1	West	2	21-40	1	Male	5	80,001 upward	3	Graduate	6	6	1	Renter	2	Married	1	High
1	San Francisco	1	CA	1	West	4	61-80	1	Male	4	60,001 - 80,000	3	Graduate	2	2	1	Renter	1	Single (Never r	2	Medium
4	New York	3	NY	2	East	5	61 and above	1	Male	2	20,001 - 40,000	1	High School	7	7 or more	2	Owner	1	Single (Never r	2	Medium
4	New York	3	NY	2	East	2	21-40	2	Female	5	80,001 upward	2	Undergraduate	2	2	1	Renter	1	Single (Never r	2	Medium
2	Los Angeles	1	CA	1	West	2	21-40	1	Male	5	80,001 upward	1	High School	2	2	2	Owner	1	Single (Never r	2	Medium
4	New York	3	NY	2	East	4	61-80	2	Female	2	20,001 - 40,000	2	Undergraduate	5	5	1	Renter	3	Divorced	2	Medium
1	San Francisco	1	CA	1	West	1	below 20	1	Male	2	20,001 - 40,000	3	Graduate	7	7 or more	1	Renter	1	Single (Never r	2	Medium
4	New York	3	NY	2	East	3	41-60	2	Female	5	80,001 upward	3	Graduate	1	1	1	Renter	3	Divorced	1	High
10	Kansas	8	KA	3	Central	3	41-60	1	Male	3	40,001 - 60,000	3	Graduate	6	6	2	Owner	2	Married	1	High
1	San Francisco	1	CA	1	West	1	below 20	1	Male	1	0 - 20,000	2	Undergraduate	1	1	1	Renter	3	Divorced	3	Low
8	Las Vegas	6	NV	3	Central	5	61 and above	2	Female	3	40,001 - 60,000	2	Undergraduate	6	6	2	Owner	3	Divorced	1	High
7	Vienna	5	VA	2	East	2	21-40	1	Male	2	20,001 - 40,000	2	Undergraduate	6	6	2	Owner	3	Divorced	1	High
4	New York	3	NY	2	East	1	below 20	2	Female	5	80,001 upward	2	Undergraduate	2	2	1	Renter	2	Married	1	High
1	San Francisco	1	CA	1	West	1	below 20	2	Female	5	80,001 upward	2	Undergraduate	1	1	2	Owner	1	Single (Never r	2	Medium
5	Washington DC	4	DC	2	East	2	21-40	2	Female	3	40,001 - 60,000	2	Undergraduate	5	5	1	Renter	3	Divorced	2	Medium
9	Seattle	7	WA	1	West	2	21-40	2	Female	3	20,001 - 40,000	4	Other	2	2	2	Owner	2	Married	2	Medium

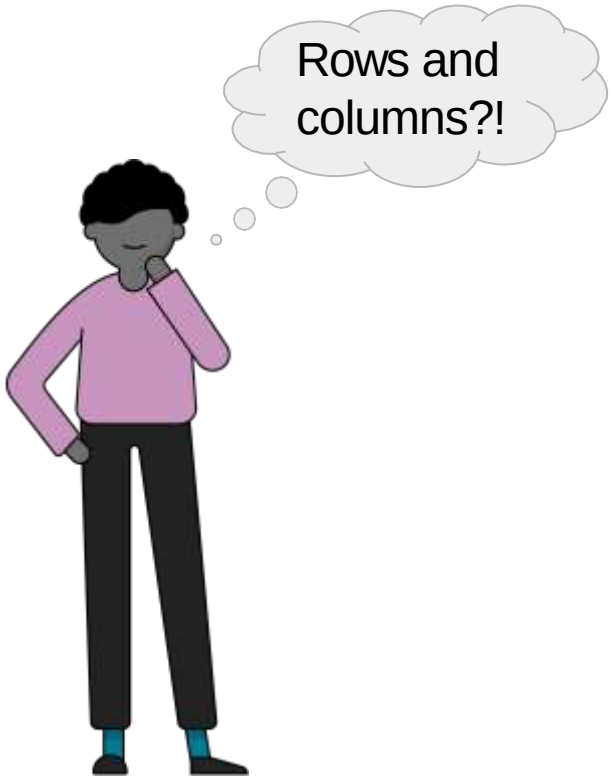


Image source: Data Mining Data Set Reports

The Five Vs (or Characteristics) of Big Data

Volume: Consider the **scale** of the data (big or small, structure).

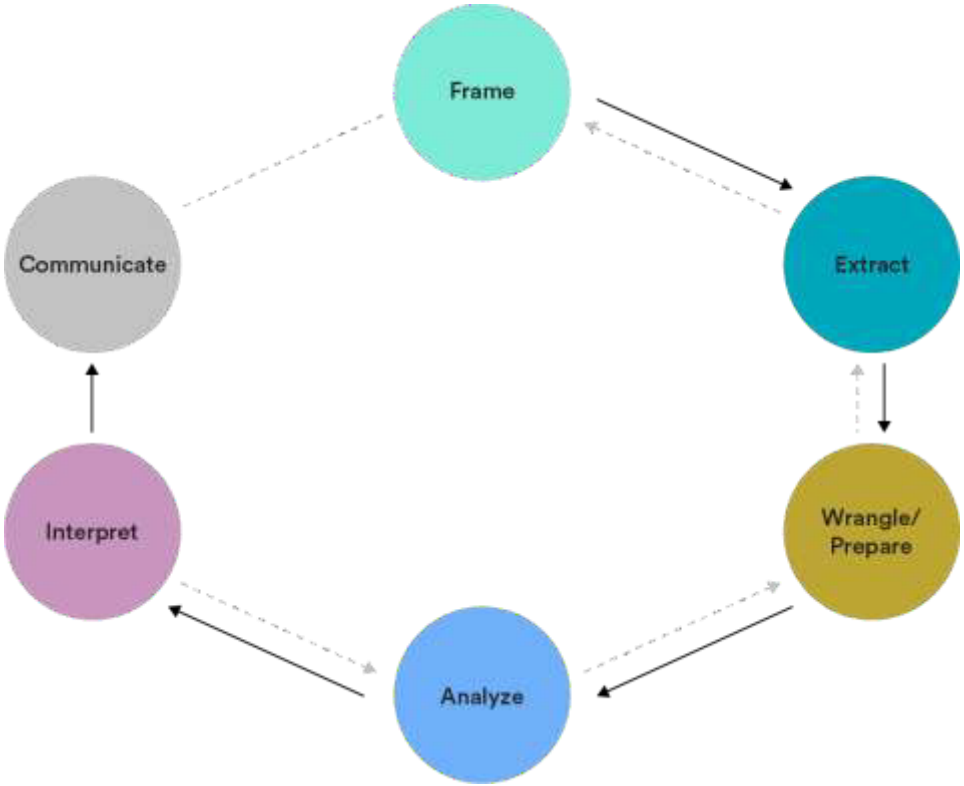
Velocity: Understand data **sources**, **timing**, and **flow**.

Variety: What **forms** and **types** are required to answer questions?

Veracity: Verify the **quality**, **accuracy**, and **reliability** of sources.

Value: What are the **metrics** or **measurements** for desired outcomes?

The DA Workflow



Frame: Develop hypothesis-driven questions for your analysis.

Extract: Select and import relevant data.

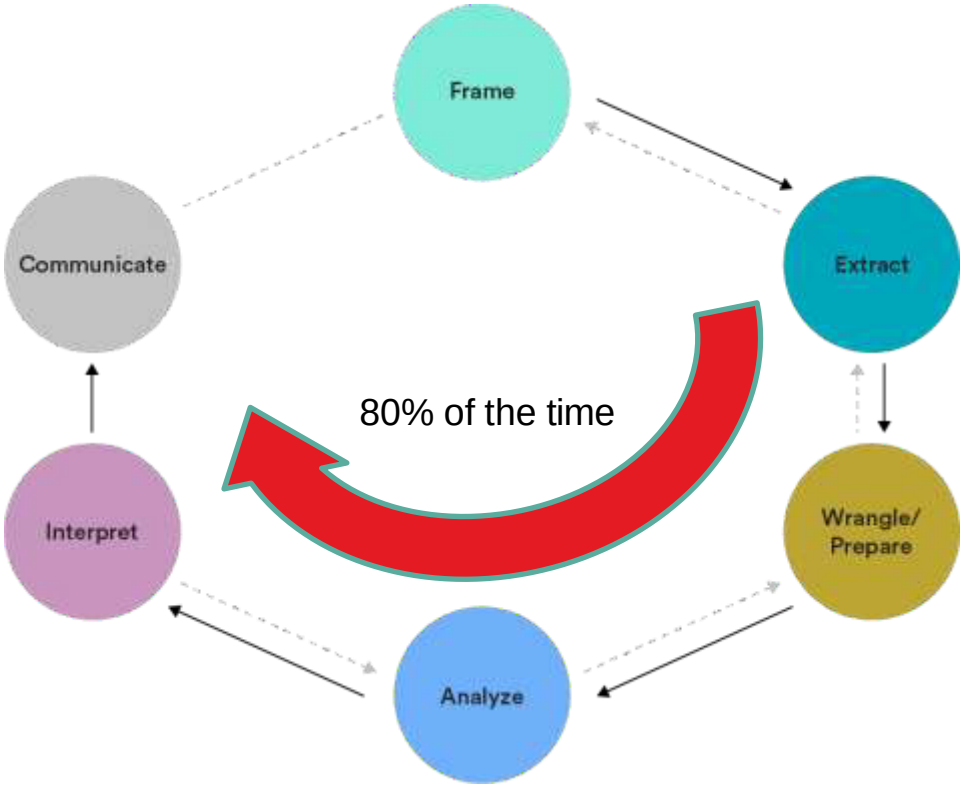
Wrangle/Prepare: Clean and prepare relevant data.

Analyze: Structure, comprehend, and visualize data.

Interpret: Leverage your analysis to make decisions and recommendations.

Communicate: Present data-driven findings and insights in a compelling manner.

The DA Workflow



Frame: Develop hypothesis-driven questions for your analysis.

Extract: Select and import relevant data.

Wrangle/Prepare: Clean and prepare relevant data.

Analyze: Structure, comprehend, and visualize data.

Interpret: Leverage your analysis to make decisions and recommendations.

Communicate: Present data-driven findings and insights in a compelling manner.

Introduction to Data Analytics

Data Formats

Let’s Talk About Data Formats

You’ll be looking at a lot of data throughout this course.

The formatting of data can make a real difference in your work as a DA!

Discussion:

Data Formats | Columns

2 minutes

Take a look at this example:

- What do you notice?
- What can we do to improve it?

Share your answers with the class.

Street Address
1234 Main Street, Sacramento, CA, 95822
5678 Broadway Avenue, Denver, CO, 80122
9810 Poplar Street, Philadelphia, PA 19108

Discussion:

Data Formats | Columns

2 minutes

How about now?

- What's changed?
- What makes this version better?

Street Address	City	State	Zip
1234 Main Street	Sacramento	CA	95822
5678 Broadway Avenue	Denver	CO	80122
9810 Poplar Street	Philadelphia	PA	19108

Partner Exercise:

Data Formats | Columns and Rows

5 minutes

Pair up with a classmate and take a look at the example below.

- What do you notice? What does this data set tell you?
- What can we do to improve it?

Country Name	Country Cod	Indicator Name	1960	1961	1962	1963
Aruba	ABW	Life expectancy at birth, total (years)	65.5693658	65.9880243	66.3655365	66.7139756
Andorra	AND	Life expectancy at birth, total (years)				
Afghanistan	AFG	Life expectancy at birth, total (years)	31.5800487	32.0959756	32.6118780	33.1273170
Angola	AGO	Life expectancy at birth, total (years)	32.9848292	33.3862195	33.7875853	34.1884634
Albania	ALB	Life expectancy at birth, total (years)	62.2543658	63.2734634	64.1628536	64.8870975
Arab World	ARB	Life expectancy at birth, total (years)	46.7626948	47.3886012	48.0024362	48.6075914
United Arab Emirates	ARE	Life expectancy at birth, total (years)	52.2432195	53.2865609	54.327	55.3635122
Argentina	ARG	Life expectancy at birth, total (years)	65.2155365	65.3385122	65.4326097	65.5093902
Armenia	ARM	Life expectancy at birth, total (years)	65.8634634	66.2843902	66.7098536	67.1378536

Data Formats |Columns and Rows

Here's a better way!

One row for each *variable*: country name, country code, year, and life expectancy.

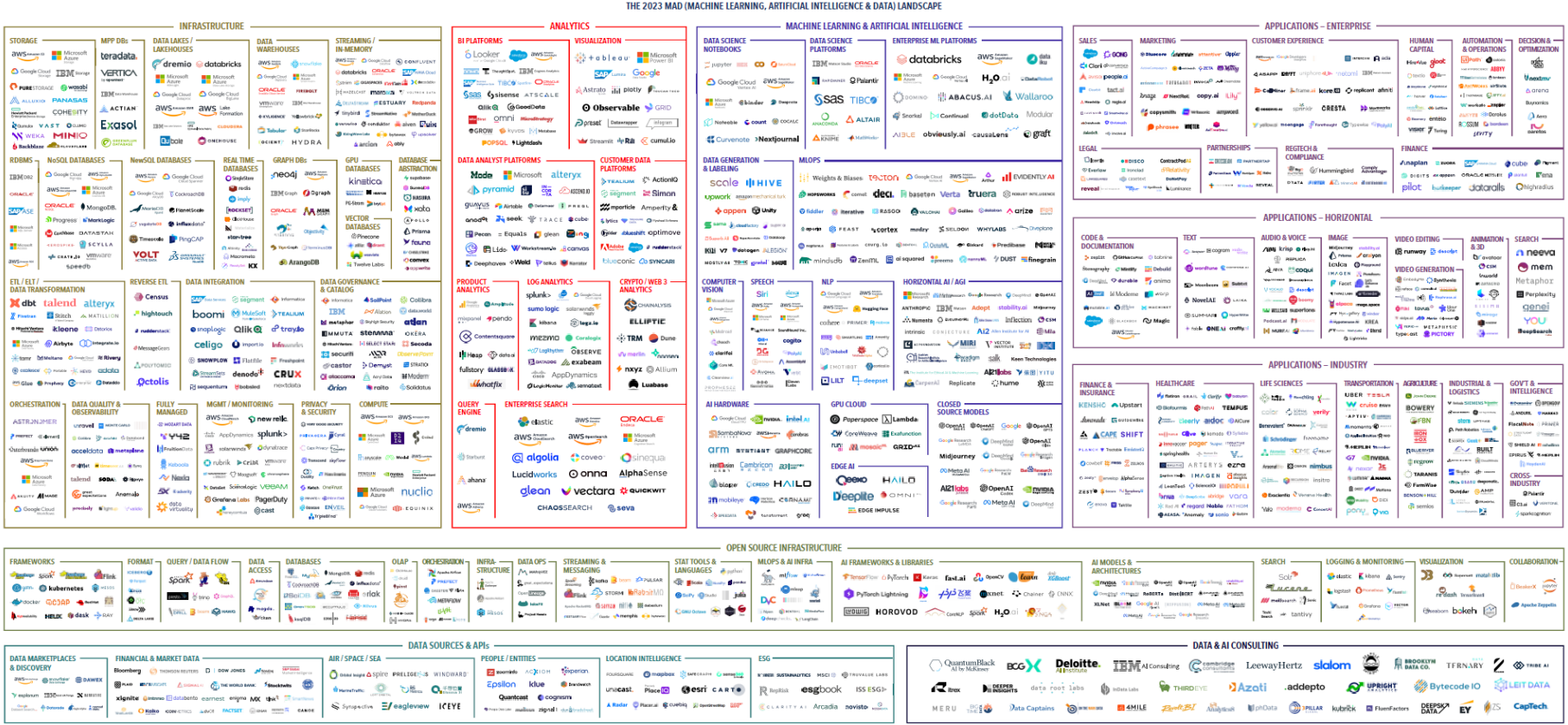
It's OK if some data are repeated!

Country Name	Country Code	Year	Life Expectancy
Aruba	ABW	1960	65.56936585
Aruba	ABW	1961	65.98802439
Aruba	ABW	1962	66.36553659
Aruba	ABW	1963	66.71397561
Afghanistan	AFG	1960	31.58004878
Afghanistan	AFG	1961	32.09597561
Afghanistan	AFG	1962	32.61187805
Afghanistan	AFG	1963	33.12731707
Angola	AGO	1960	32.98482927
Angola	AGO	1961	33.38621951
Angola	AGO	1962	33.78758537
Angola	AGO	1963	34.18846341

Data Tools..

How many you think there is??

Data Tools..



Version 1.0 - Feb 2023

Blog post: mattturck.com/MAD2023

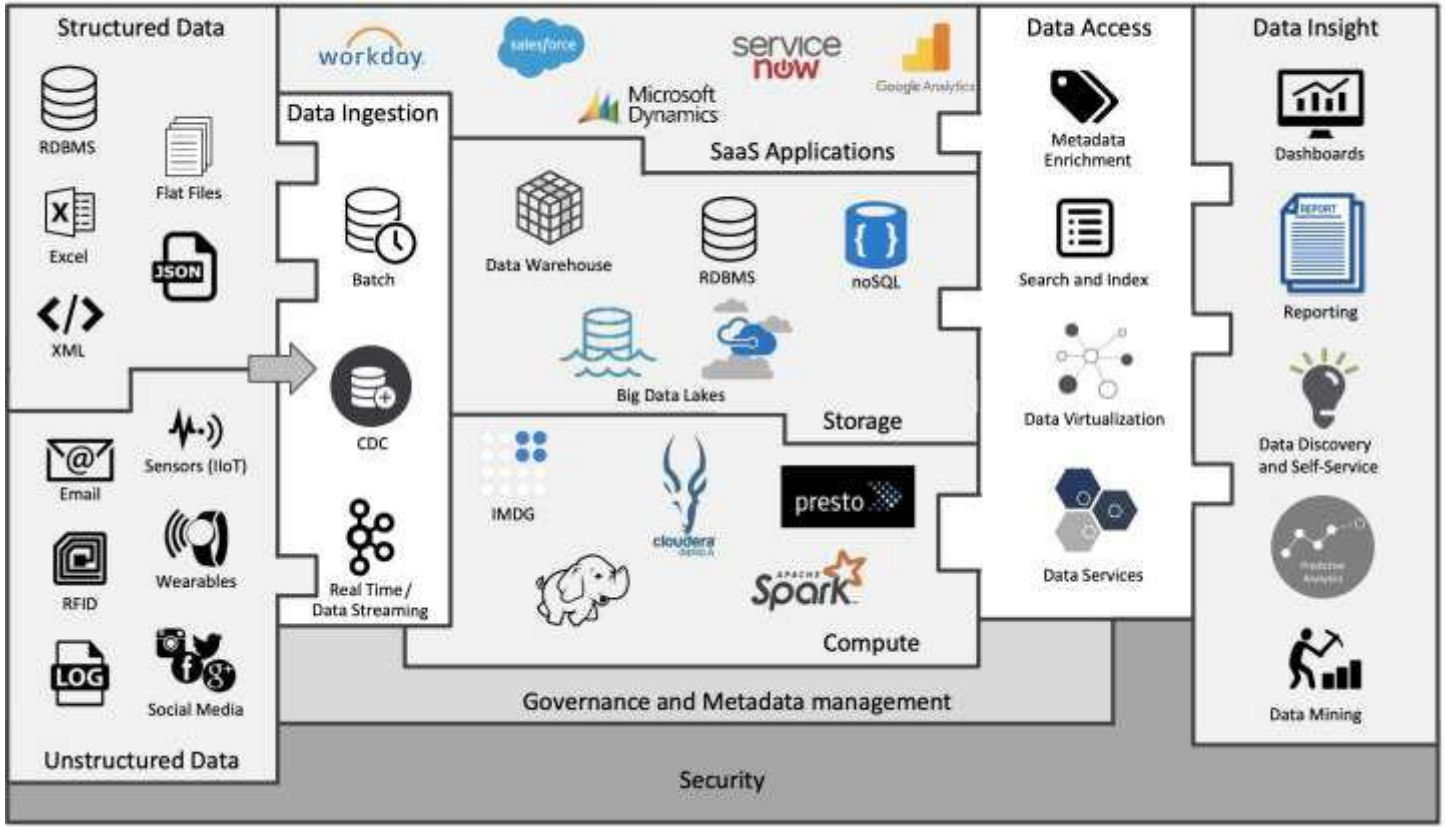
Interactive version: MAD.firstmarkcap.com

Comments? Email MAD2023@firstmarkcap.com

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

Link to pdf

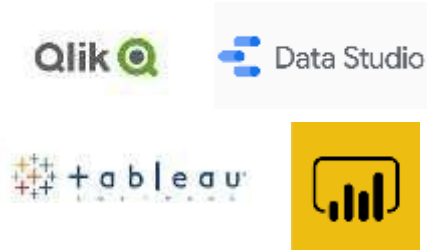
Agile Information Architecture



Example of Tools for Analytics

Workbook/Sheets

Analytic Databases



Visualization Tools

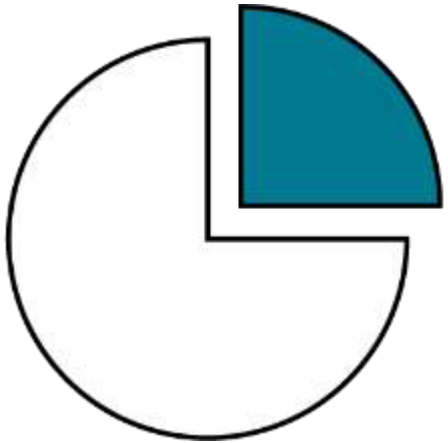
Introduction to Data Analytics

Data Ethics

Data Ethics

Data ethics is about the **responsible and sustainable use of data** —doing the right thing for people and society.

It refers to the principles and values on which **human rights and personal data protection** laws are based.



—DataEthics.eu

Real Cases:

Netflix and… Data Ethics

According to a McKinsey report, **75% of Netflix viewing decisions result from product recommendations**. This raises ethical implications such as:

- Addictiveness
- Radicalized content
- Privacy

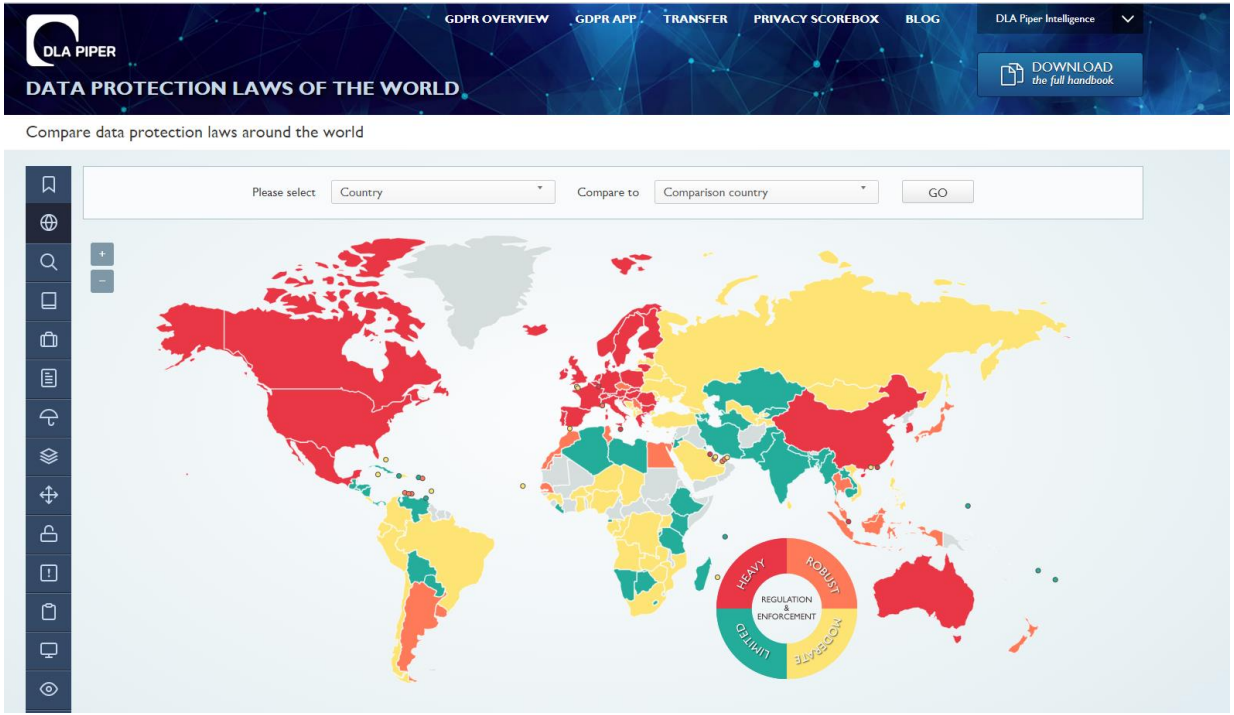
Source: How Retailers Can Keep Up With Consumers

Data Ethics Principles

- Privacy
- Transparency
- Faireness
- Accountability
- Responsiblity

Data Privacy

In order to protect individual data privacy, governments have implemented data protection laws:



Source: DLA Piper

Deep Fakes

Poses ethical problems everywhere

Regulating deep fakes: legal and ethical considerations

Edvinas Meskys, Aidas Liaudanskas, Julija Kalpokiene, Paulius Jurcys

Journal of Intellectual Property Law & Practice, Volume 15, Issue 1, January 2020, Pages 24–31, https://doi.org/10.1093/jiplp/jpz167

Published: 17 January 2020

ETHICS AND TRANSFORMATIVE TECHNOLOGIES

SU Home > Ethics and Transformative Technologies > Viewpoints > Deepfakes and the Value-Neutrality Thesis

ETHICS AND TRANSFORMATIVE TECHNOLOGIES

ABOUT

EDUCATIONAL RESOURCES

EVENTS

Deepfakes and the Value-Neutrality Thesis

Nathan Colaner and Michael J. Quinn

February 10, 2020

Anticipating and Addressing the Ethical Implications of Deepfakes in the Context of Elections

New Media & Society

27 Pages • Posted: 1 Nov 2019 • Last revised: 23 Apr 2020

Nicholas Diakopoulos

Northwestern University - School of Communication

Deborah Johnson

University of Virginia - School of Engineering & Applied Science

Date Written: October 21, 2019

ORF

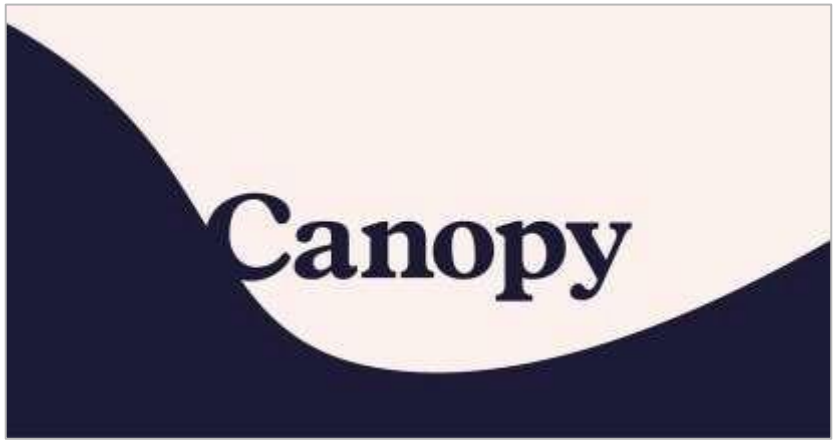
Debating the ethics of deepfakes

Real Cases:

A Healthier Landscape for Product Recommendations

It’s certainly possible! Canopy is working on a recommendation system that:

- Looks for **signs of quality**.
- Makes suggestions **without centralized data collection**.
- Runs the recommendation algorithms on **a person’s device**.
- Shares only **anonymized usage data** with company servers.



Source: The People Trying to Make Internet Recs Less Toxic

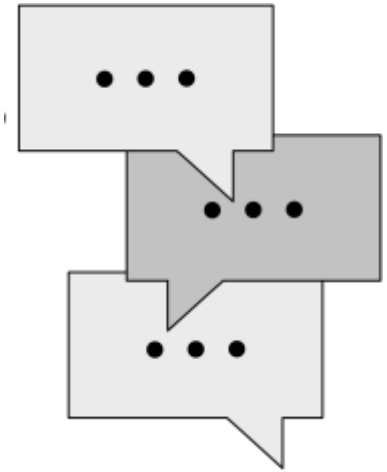
Partner Exercise:

Doing Our Part as Data Analysts

10 minutes

Discuss the following with your partner. We'll regroup after five minutes.

- What does it mean to use data ethically?
- What are some ways you've seen data being used unethically?
- What role do data analysts play in using data ethically?



Group Exercise:

More examples of data

15 minutes

- How Google/Facebook faced problems with Data?
- Biased data reporting
- Prediction models
- The accessibility of datasets and what you do with them?
- Also check out deep fakes, and give an example of good usage of deep fakes

Introduction to Data Analytics

Wrapping Up

Recap

Today, we:

- Outlined goals, expectations, and logistics.
- Identified the skills and mindset of a successful data analyst.
- Discussed the discipline of data analytics, including topics such as data formats and data ethics.

Looking Ahead

Homework

- Optional myGA lessons:
  - **Exploring Data** (unit)
    - Data Profiling
    - Probing Data With Logical Functions
  - **Data Wrangling** (unit)
    - Cleaning Your Data

Up Next: Introduction to SQL

Q&A


GA