TEXAS ADVANCED COMPUTING CENTER

WWW.TACC.UTEXAS.EDU

TEXAS
The University of Texas at Austin

What Is Data Science, and Why Does It Matter?

"From Caveman Art to Spotify Streams"

Goal:

Introduce students to the concepts, history, tools, and relevance of Data Science, preparing them to explore it hands-on.

PRESENTED BY:

The Oldest Profession You've Never Heard Of

Data Science Isn't New—We've Been Doing It for Thousands of Years

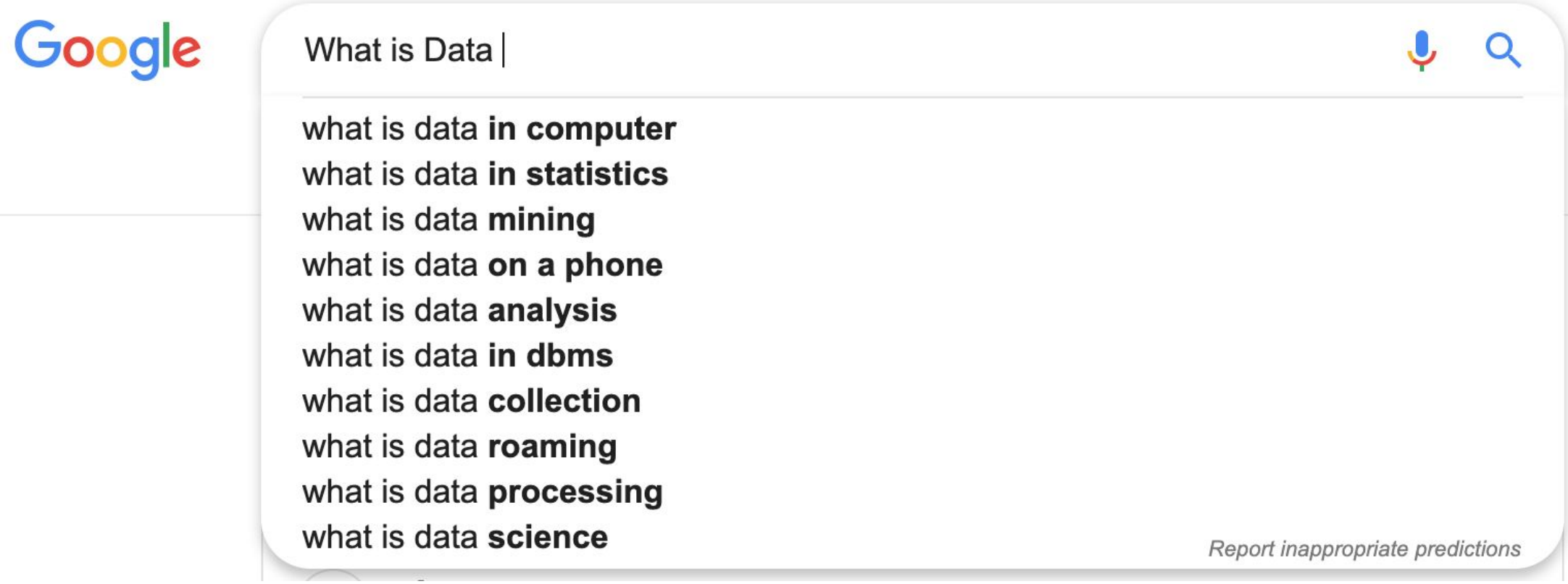
- **Cave paintings** were early data logs—recording what was hunted or important
- **Ancient civilizations** used census data to allocate resources
- **Humans** are natural data collectors, always looking for patterns

Like tracking scores in a game or keeping a diary, we've always collected data to understand the world better.

Data Science 101

What is Data?

Data Science 101



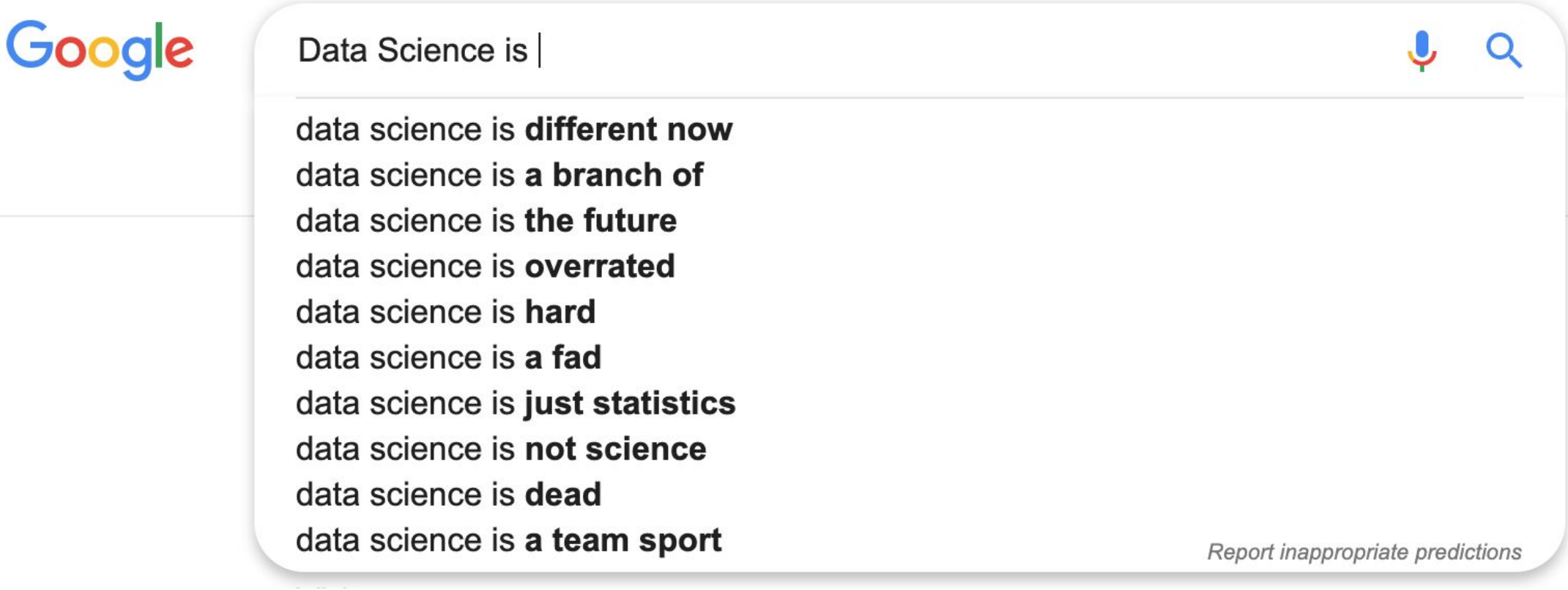
Data Science 101

Data is a set of values of subjects with respect to qualitative or quantitative variables. Data and information or knowledge are often used interchangeably; however data becomes information when it is viewed in context or in post-analysis. Wikipedia

Data Science 101

What makes data important?

Data Science 101



or….


OH, YOU'RE A DATA SCIENTIST?
TELL ME ALL ABOUT HOW YOU DO YOUR ANALYTICS WITH EXCEL
memegenerator.net

Data Science 101

Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data. Wikipedia

Data Science 101

Data Science is the ability to understand that there is a story hidden in the data.

What Is Data Science? - A Deeper Dive

Turning Raw Data into Insight and Action

Data Science is the process of using math, coding, and knowledge to extract meaning from data - finding the story!

It blends:

- Math & Stats to **analyze**
- Programming to **process**
- Domain knowledge to **understand context**

data alone is just numbers

The Three Cornerstones of Modern Data Science

Scale, Speed, Ubiquity

- **Scale**: We create 2.5 quintillion bytes of data daily (that's 2.5 million terabytes!)
- **Speed**: Real-time fraud detection, instant recommendations, live traffic updates
- **Ubiquity**: Data is everywhere—from your phone to city cameras

Data is like oxygen—unseen but everywhere and essential.

Three Cornerstones: SCALE

Massive amounts of data being generated

- **Netflix** processes hundreds of millions of viewing sessions per day to recommend shows.

  Every second of your binge-watch is analyzed to find patterns.

- **Walmart** collects over 2.5 petabytes of customer data per hour—from scanners, inventory, and purchases.
- **NASA** satellites generate terabytes of climate data daily, monitoring everything from hurricanes to ice caps.

Three Cornerstones: SPEED

Real-time or near real-time processing

- **Spotify** adapts your “Daily Mix” while you're still listening to music.
  - *Skip a song, and it starts reshaping your playlist immediately.*
- **Delta Airlines** uses real-time data to reroute flights and reschedule crews to minimize delays.
- **Uber Eats** tracks both driver location and restaurant prep time to adjust delivery estimates on the fly.
- **Amazon**’s dynamic pricing engine changes product prices millions of times a day based on demand, stock, and competition.

Three Cornerstones: UBIQUITY

Data is everywhere

- Your **Smart Watch** tracks your steps, sleep, and heart rate—generating health data 24/7.
- **Smart thermostats** like Nest learn your habits and adjust temperature automatically.
- **Traffic lights** in smart cities adapt timing based on real-time congestion data.
- **Grocery stores** use loyalty card data to understand customer behavior and send you personalized coupons.

Why Does Data Science Matter?

It Helps Us Understand, Predict, and Decide

- **Understand behavior**: What songs do you like?
- **Predict outcomes**: When will traffic be heaviest?
- **Make smarter choices**: When to restock a store or launch a new product

TikTok uses your viewing time to customize your “For You” page, tailoring content just for you.

Data Science in Action

Real World Applications

Company	Use Case
Amazon	Product recommendations based on your browsing and purchase history
Spotify	Creates “Discover Weekly” playlists just for you
Uber	Predicts demand to match drivers and riders
CDC	Tracks and predicts disease outbreaks using data
City of Austin	Monitors traffic incidents to improve safety and flow***

The Data Science Pipeline

How Do We Work with Data?

Steps:

1. **Ask a question (e.g., When is traffic worst in Austin?)**
2. **Collect data (files, sensors, APIs)**
3. **Clean data (fix errors, remove duplicates)**
4. **Explore data (summarize, visualize)**
5. **Model data (make predictions)**
6. **Share insights (dashboards, reports)**

The Tools of the Trade

Meet Pandas and Matplotlib

Pandas: Like Excel but much more powerful for data wrangling

Matplotlib: A way to create charts and graphs from data

Pandas = your data chef's knife

Matplotlib = your data artist's paintbrush

What Can You Do With Data Science?

The Possibilities Are Endless

- Analyze cafeteria usage to reduce waste
- Build an app to find best driving times
- Study climate impacts on neighborhoods
- Discover trends in music or social media

Data science lets you tell stories with data and influence change

Jobs You Could Have

There’s a wide range of careers—no one path

Role	What You Do
Data Analyst	Look for trends and summaries
Data Scientist	Build models and automate decisions
Machine Learning Engineer	Deploy AI-powered systems
Public Policy Analyst	Use data to influence fair laws
Entrepreneur	Create data-driven apps and products

Data Justice: Why Ethics Matter

Data Isn’t Neutral

- Algorithms can reinforce bias if not designed carefully
- Ethical data science is about fairness, transparency, and inclusion

The quiet part out loud:

“How can data be biased?”

How to Get Started

Let's do this thing

- Start with curiosity and ask questions
- Learn Python and explore data projects
- Use free resources like DataCamp, Kaggle, HPC-ED, or ScienceGateways.org

Final Thoughts

The Future Belongs to Data-Literate People

“Data Science is a superpower. With it, you can discover patterns, solve problems, and shape the world… But there's a catch.”

- ***- Anonymous TACCster***

The Catch

The Future Belongs to Data-Literate People

"Data without context is just noise."

Data is powerful, but ***only*** if we understand:

- How it was collected
- Why it was collected
- Who collected it
- What was left out

The Catch

The Future Belongs to Data-Literate People

Real-World Examples:

1. Salary Data Misinterpretation

"The average salary is $100K"
But: Does that include the CEO? Interns? Is it median or mean?

2. Facial Recognition Bias

AI performed poorly on darker-skinned individuals
Why? The training dataset had mostly lighter-skinned faces.

3. Austin Traffic Spike at 5 PM

Are drivers more reckless then?
Or is it rush hour + bad weather + school zones?

The Catch

The Future Belongs to Data-Literate People

Data Science isn’t just about crunching numbers. It’s ***about asking thoughtful questions and understanding the story behind the data.***

What's Next?

The Future Belongs to Data-Literate People

Your Questions & Ideas

- What data projects interest you?
- Anything surprise you today?