

# **Welcome to MSIS 502**

## **Business Data Analysis**

Introduction



# Landscape of Business Data Analysis?

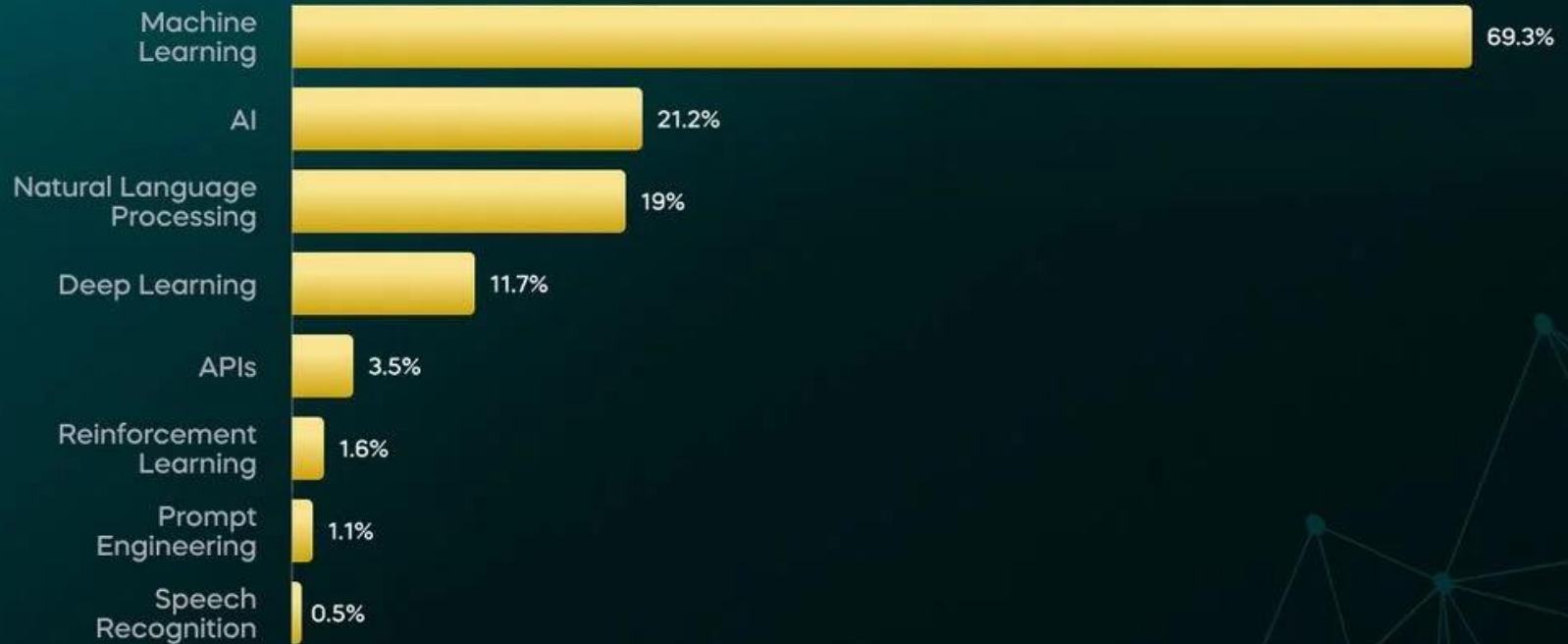


# Age of **Big Data** and **GenAI**



## AI Skills

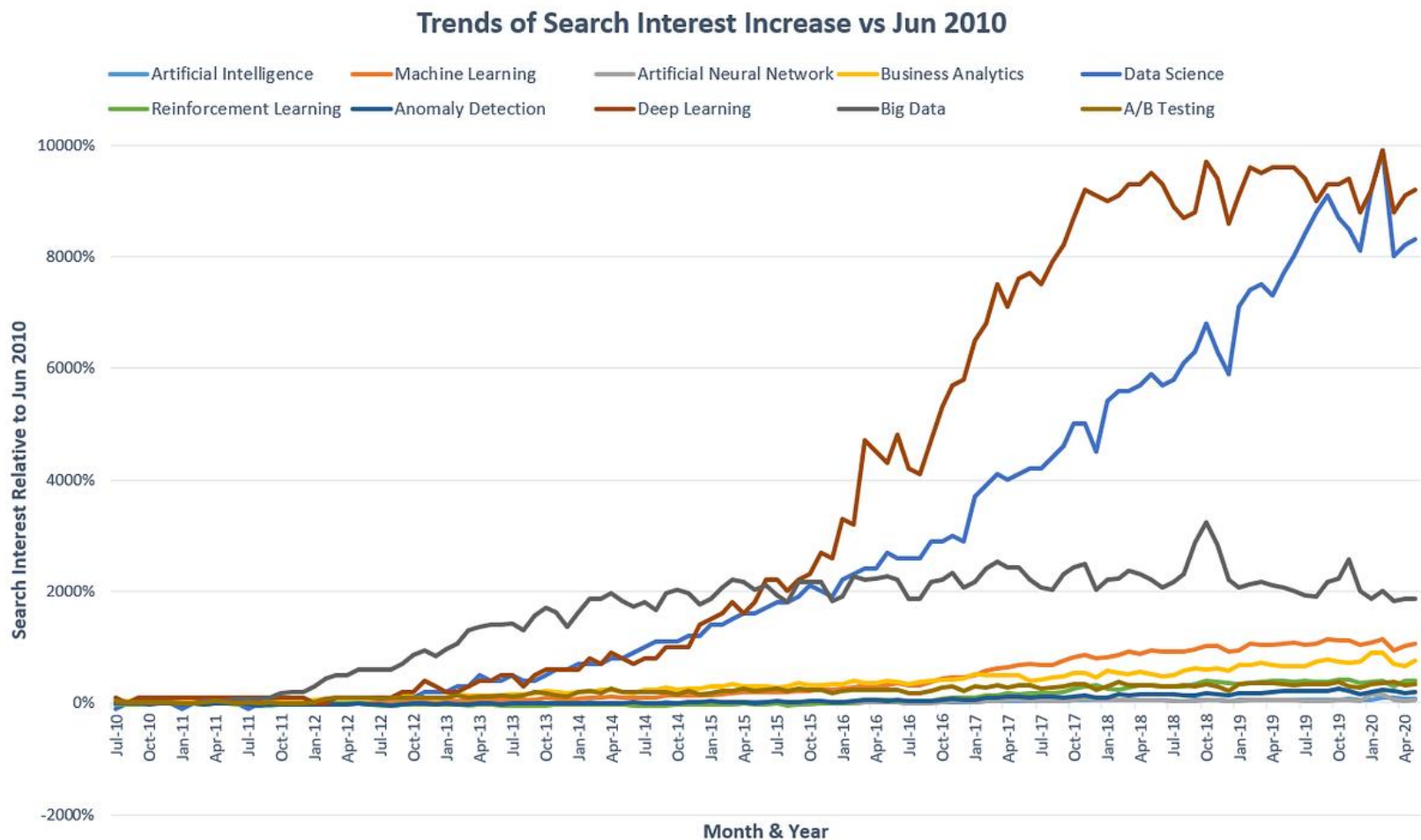
Mentioned in Data Scientist Job Postings



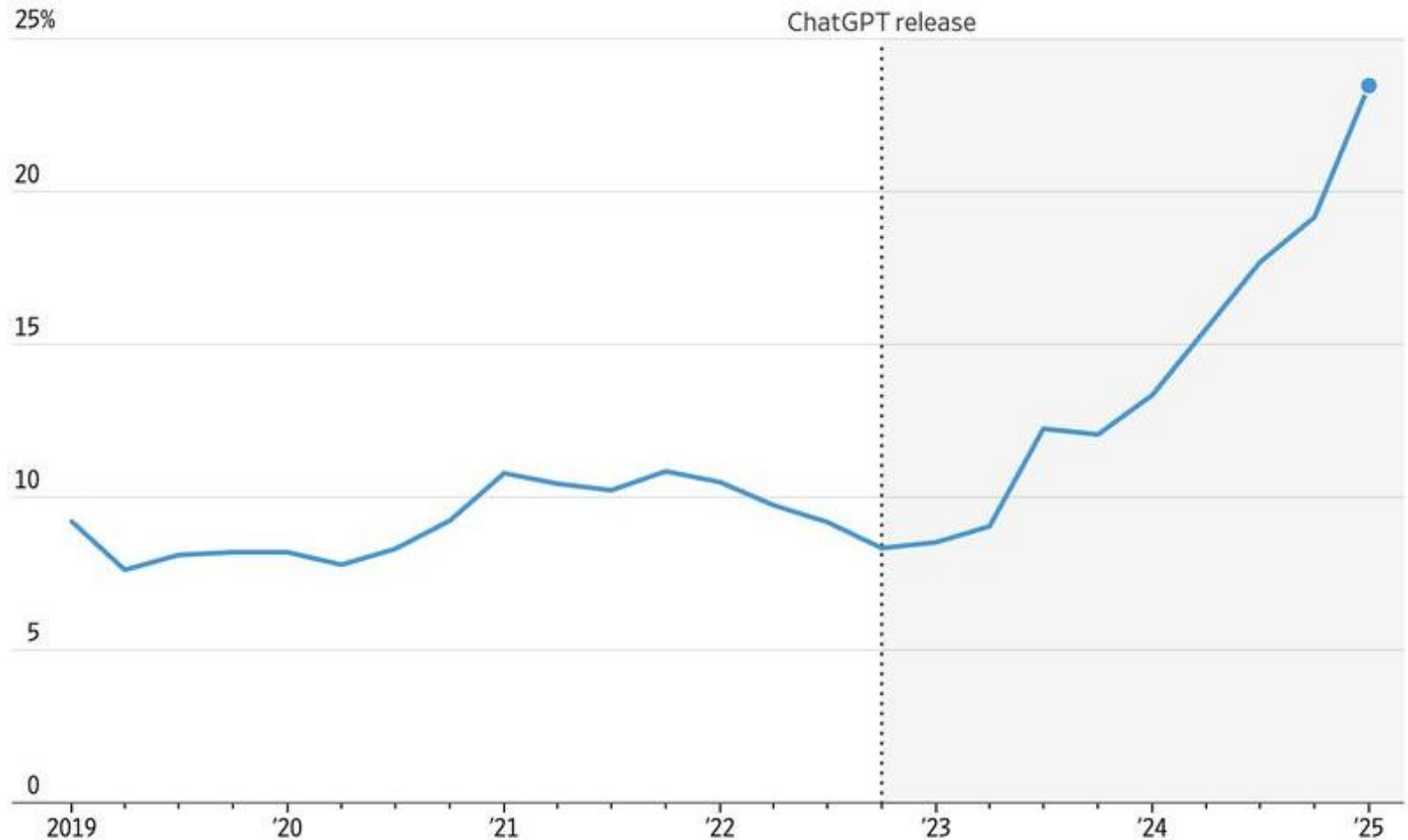
Percentage of Total Job Postings

# Growth in Data Related Searches on Google

> Google Trends Keywords 2010 – 2020 Worldwide



# Newly listed technology jobs that are AI-related

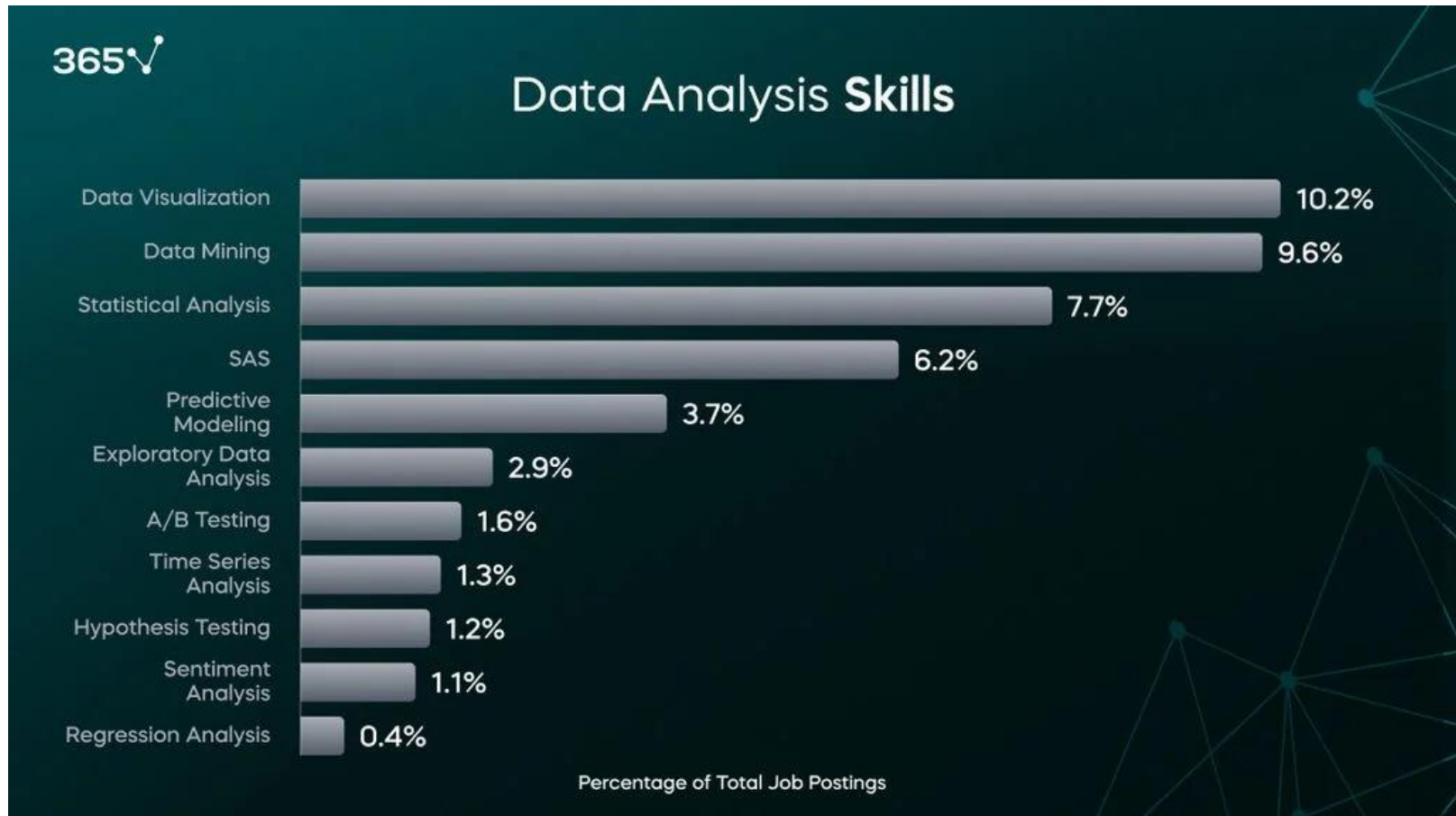


\*Technology jobs are a broad group of computer- and math-related occupations

Note: U.S. only. 1Q 2025 figures are as of January.

Source: UMD-LinkUp AI Maps

# Data Analysis Skills

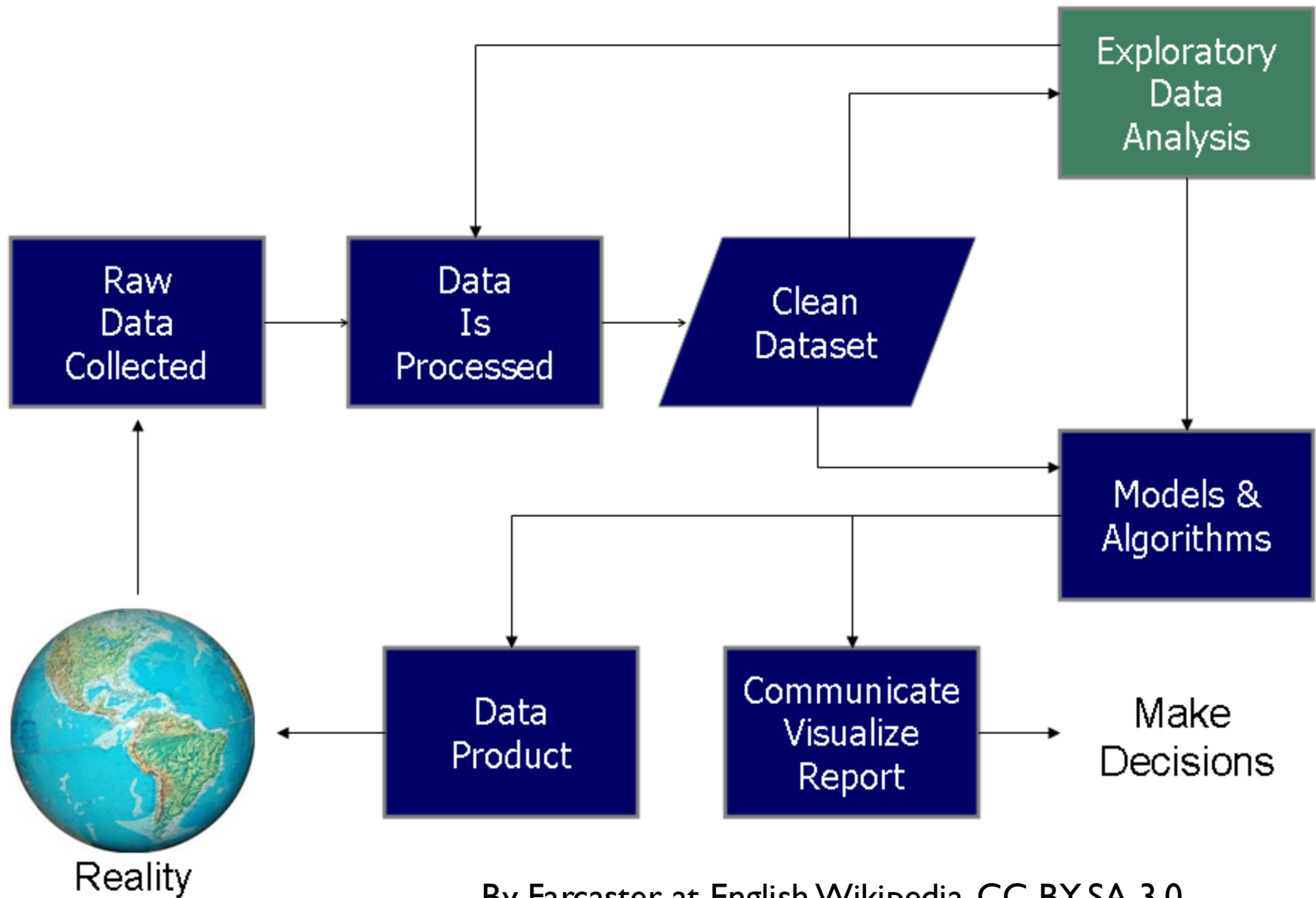


# What is Business Data Analysis?





# Data Science Process



By Farcaster at English Wikipedia, CC BY-SA 3.0,  
<https://commons.wikimedia.org/w/index.php?curid=40129394>

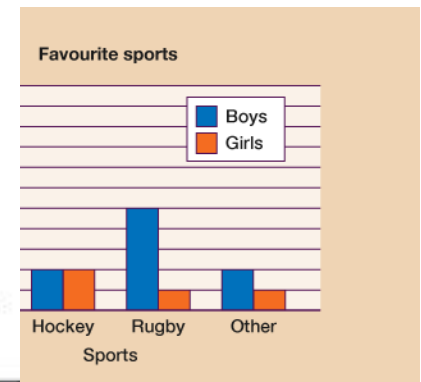
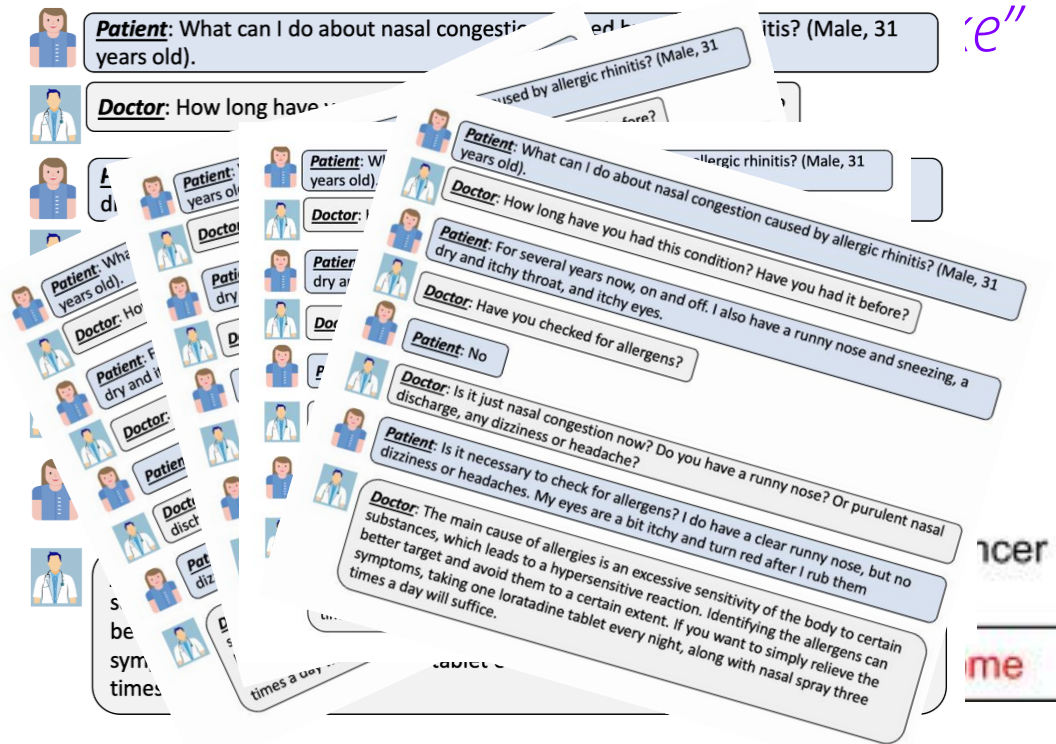
# What Business Data Analysis Can Do?



# Top Data Analysis Techniques

- > Descriptive Analysis, *"What happened"*
- > Diagnostic Analysis, *"Why did this happen?"*
- > Predictive Analysis, *"What is likely to happen in the future"*
- > Prescriptive Analysis, *"What should we do about it?"*

Season	
Summer	
Autumn	
Winter	
Spring	



# Generative AI

- > Generative AI refers to artificial intelligence applications that can generate new content, ranging from text, images, and music to code and synthetic data. This type of AI learns from existing data to produce new items that resemble the training data in structure and features but are original creations.

# Examples

- > **GPT** (Generative Pre-trained Transformer): A text-based model that generates human-like text based on the input it receives. It can write essays, create poetry, generate code, or even simulate dialogue based on the patterns it has learned from the vast amounts of text it was trained on.
  - Chatbot, Content Creation
  - Software Development, Product Design
  
- > **DALL-E**: An AI program created by OpenAI that generates images from textual descriptions. It can create new visual concepts by combining attributes and styles in novel ways.
  - Marketing, Arts, Fashion

# DALL-E

Can you generate an image from the following twitter message?  
Autumn's fan favorite is officially on our fall menu! Introducing the  
Iced Pumpkin Cream Chai Tea Latte. 🍏🍂 (US & Canada)





# DALL-E

Can you blend in a cat in this image?



# DALL-E

I like this style! Can you change the background to a sea?






# AI assisted programming

ChatGPT, Google Colab (Gemini), Github Copilot...

Can you show me how to generate a line chart in python?

python

 Copy code

```
import matplotlib.pyplot as plt

# Sample data
x = [1, 2, 3, 4, 5]
y = [2, 3, 5, 7, 11]

# Create the plot
plt.figure(figsize=(10, 5))
plt.plot(x, y, marker='o')

# Add title and labels
plt.title('Sample Line Chart')
plt.xlabel('X Axis')
plt.ylabel('Y Axis')

# Show the plot
plt.grid(True)
plt.show()
```

This code will create a simple line chart with `x` values on the X-axis and `y` values on the Y-axis.

The `marker='o'` parameter adds markers at each data point.

# Is GPTs smart? OR Smarter than Human?

- > Is GPTs smart?
- > Smarter than Human?

how many letter r's in strawberry



The word "**strawberry**" contains **two** letter "r"s.

are you sure



Let's double-check! The word "**strawberry**" indeed contains **two** letter "r"s. Here's the breakdown:

- s
- t
- r
- a
- w
- b
- e
- r
- r
- y

✓ Two "r"s! 😊

# Prompts

- > Help Me Write Code
  - Write a Python function that returns the average of a list of numbers.
- > Help Me Understand
  - What does the `zip()` function do in Python? Give an example.
- > Help Me Debug
  - This code throws a `TypeError: 'int' object is not iterable`. Can you help me fix it?
- > Help Me Clean Data
  - How do I fill missing values in a pandas DataFrame with the median of each column?
- > Help Me Visualize
  - Create a boxplot comparing total bill amounts by day of the week.

# Prompt Engineering

- > Please first take a deep breath...
- > Image you are an expert in machine learning.
- > Explain xxx to me like I'm 6 years old.

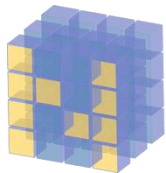
# What You Will Learn?



# Syllabus



Session	Date/Time	Topic	Due
1&2	07/12/2025 Saturday 9:00a ~ 12:15p	Course Introduction Python Basic	
3&4	07/23/2025 Wednesday 6:00p ~ 9:15p	Python for Data Science Data Representation Data Visualization	Founders 170
5&6	07/28/2025 Monday 6:00p ~ 9:15p	Exploratory Data Analysis	Individual Assignment 1 In-Class Assignment
7&8	08/04/2025 Monday 6:00p ~ 9:15p	Principles of Regression Analysis	Individual Assignment 2
9&10	08/11/2025 Monday 6:00p ~ 9:15p	Final Project Presentation	Final Project



NumPy



matplotlib

pandas  
 $y_{it} = \beta x_{it} + \mu_i + \epsilon_{it}$



# Individual Assignments

- > Two assignments 10 points each
- > Deduction will be applied to incorrect answers, missing or incomplete requirements
- > Due at the beginning of the class on due dates

## In-Class Assignment

- > 20 points each
- > Deduction will be applied to incorrect answers, missing or incomplete requirements

# Team Project

- > 50 points
- > Done by a team of 4-5 students
- > Individual score adjusted based on peer-evaluation
- > Due at the beginning of last class
- > Present in the last class

**Appeals:** If you believe that an error was made in grading, you should first contact the TA to make an argument and request for regrade. If the disagree persists, let the professor know. You have **one week** after each deliverable is graded to submit the request.

**The (required) target median GPA in this class is 3.4 – 3.6.** Your GPA (out of 4.0) will be determined by the above weighted total grade and your rank in the class.



# Late Policy

- > Each student will be allotted three free days which can be used to turn in individual assignments late without penalty.
- > Once your free days are used up, late homework will be penalized 15% of the overall mark for each 24-hour period. (For instance, a homework turned in two days late will receive only 70% credit.)
- > If you decide to use your free days on an assignment, remember to indicate the number of free days you're using on that assignment at the time of submission.

# Late Policy

- > Assignments turned in after 3 days (i.e. 72 hours after the deadline) will receive 0 credit.
- > Free late days **CANNOT** be used for in-class assignment and team project.
- > All the deliverables need to be turned in (even with 0 credit) to earn a passing grade in this course.

# Participation

- > Active participation to class discussions will constitute the participation portion of your grade
- > In-class practice and discussion, sharing your thoughts and the solution (FCFS)
- > Random call if no volunteer
- > 10 points

# AI/ChatGPT

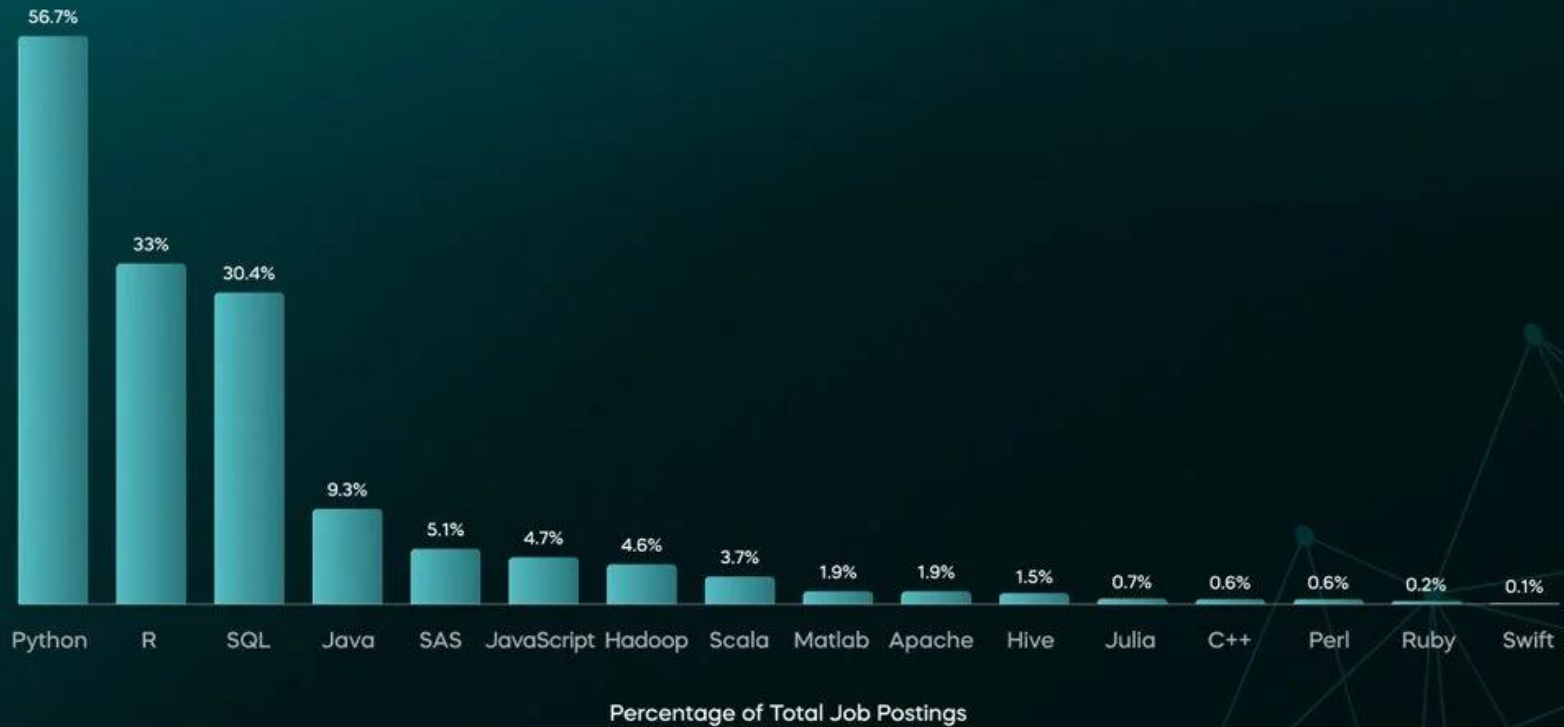
- > Artificial intelligence (AI) language models, such as ChatGPT, may be used for any assignment with **appropriate citation**. The general practice of citation is that you cite anything that comes from somewhere else; anything that isn't your original thought, isn't common knowledge, and/or is a place where you pulled information from. Hence you **must reference all the content from tool that you include** in your assignment. Failure to reference externally sourced, non-original work can result in scholastic dishonesty. References should provide clear and accurate information for each source and should identify where they have been used in your work.
- > A.I. Archives  
The website extension saves your ChatGPT conversations and creates a URL, which allows readers to reference the original conversation used by the author. This tool is particularly useful for creating citations.
- > [MLA created a resource](#) on how to cite ChatGPT and other AI sources.

# Introduction to Python



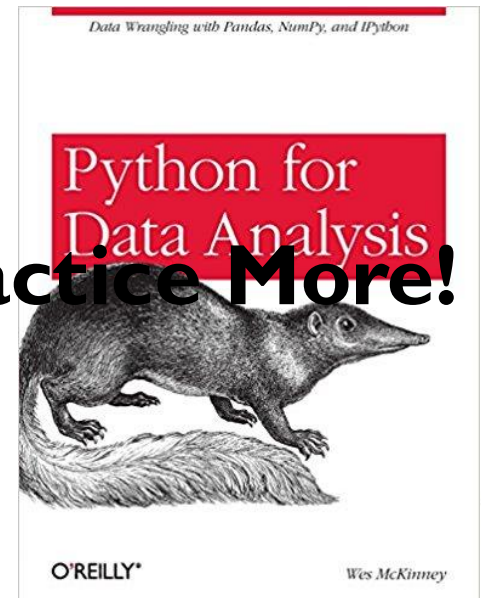
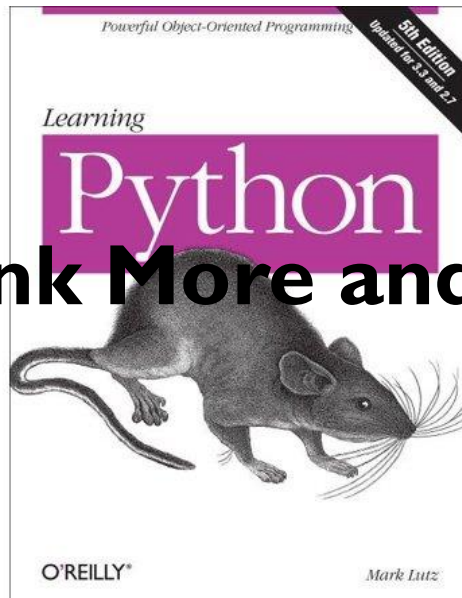
365

## Programming Languages Mentioned in Data Scientist Job Posting



# Learning Python

- ❑ All good programmers are learning from **reading others' codes and reading documentations!**
- ❑ The best way to learn is through **imitation and practice**
- ❑ There are many tutorials on the website
  - <https://www.datacamp.com/courses/intro-to-python-for-data-science>



**Read More, Think More and Practice More!**

# Learning Python

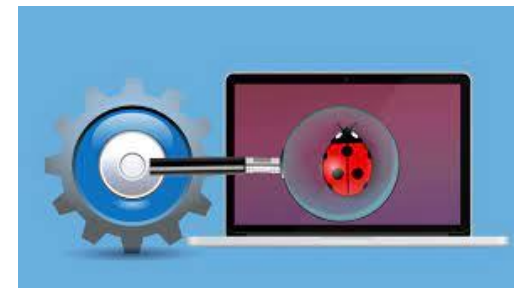
❑ You **won't** need to memorize all the syntax.

- Google
- AI assisted programming (ChatGPT, GitHub Copilot, google colab)
- However, more you “memorize”, more proficient

❑ Debugging is a tedious, yet **crucial part** of any programming

Debugging is the process of finding and resolving defects or problems within a computer program that prevent correct operation of computer software or a system.

- **READ YOUR ERROR MESSAGE!**
- **BE PATIENT!**
- Google it
- AI assisted debugging
- Debugging is your responsibility





# THINGS TO PLACE ON YOUR CV

- > Business Analytics Project Experience
- > Skills: Data Analytics, Quantitative Methods, Regression Analysis, Business Intelligence (BI), Data Visualization, Experimental Design & Analysis, Statistical Modeling, data-driven decision making
- > Coding: Python, Numpy, Pandas, Matplotlib, sklearn

# More Career Choices

- ✓ Data Scientist
- ✓ Software Development Engineer
- ✓ Applied Scientist
- ✓ Data Architect
- ✓ Business Analyst
- ✓ Marketing Analysts
- ✓ Customer Insight
- ✓ Business Intelligence
- ✓ Data Engineer
- ✓ Strategist
- ✓ Big Data Developer
- ✓ Algorithmic Trader
- ✓ Big Data Consultant



**Manager**



Propel your career now and for the long-haul

W

**Foster**  
School of Business  
UNIVERSITY of WASHINGTON

# Who Needs Data Analytics

- > Examining data to answer questions, identify trends, and extract insights.
- > **Marketers** - utilize customer data, industry trends, and performance data from past campaigns to plan marketing strategies
- > **Product managers** - analyze market, industry, and user data to improve their companies' products
- > **Finance professionals** - use historical performance data and industry trends to forecast their companies' financial trajectories
- > **Human resources and diversity, equity, and inclusion professionals** - gain insights into employees' opinions, motivations, and behaviors to make meaningful changes within their organizations
- > Artists, musician, journalists, software engineers, product manager ...
- > everyone !