

Data Science

Introduction

Prof. Gheith Abandah

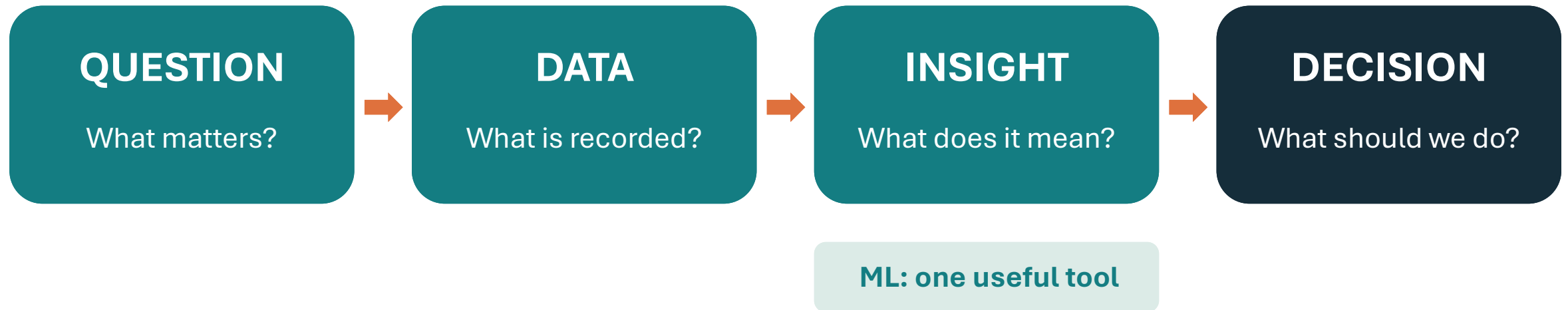
أ.د. غيث علي عبدة

Outline

- Data Science Introduction
- Data Science Skills Survey

Data science goes beyond the model

Data science uses data to answer questions and support decisions.



Describe • Explain • Predict • Support decisions

Start with a decision, then ask a question

Running example: a campus bike-sharing service.

01 Decision

Where should we move bikes before 8 a.m.?



02 Question

Which stations face the highest morning demand?



03 Evidence

Trips + station capacity + time + availability

Define the unit: one trip. Define the window: weekdays, 7–9 a.m.

Messy data can produce confident mistakes

Before analysis: inspect types, missing values, duplicates, and units.

RAW RECORDS

Trip	Station	Duration
101	Library	12 min
101	Library	12 min
102	library	720 s
103	Gym	missing



ANALYSIS-READY RECORDS

Trip	Station	Minutes
101	Library	12
102	Library	12
103	Gym	NA

Validate the result. Keep a record of every cleaning rule.

Turn records into a useful view

Load → clean → join → reshape → group → summarize

TRIPS

Trip	Station
101	Library
102	Library
103	Gym

STATION INFORMATION

Station	Capacity
Library	20
Gym	10



Join on station
Group by station



STATION SUMMARY

Station	Trips	Capacity
Library	2	20
Gym	1	10

Keys and grouping rules
shape the answer.

A chart is a claim: check the comparison

Illustrative example: two stations observed for different numbers of days.

Total trips



Observed for 30 and 10 days

Trips per observed day



$300 \div 30 = 10$; $200 \div 10 = 20$

Same data. Different question. Different conclusion.

Your course: from a file to a data system

Build reliable analysis first; then learn how it scales.

01 Prepare

Pandas + statistics
Files + formats
Cleaning + joins



02 Understand

Visualization
Group operations
Time series



03 Scale

Big data
Architectures + patterns
MapReduce

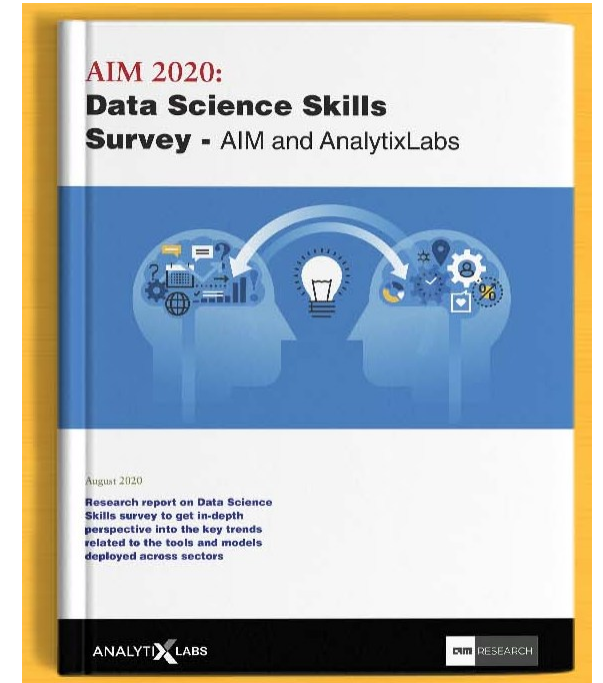
Discuss: what data would justify moving bikes to another station?

Outline

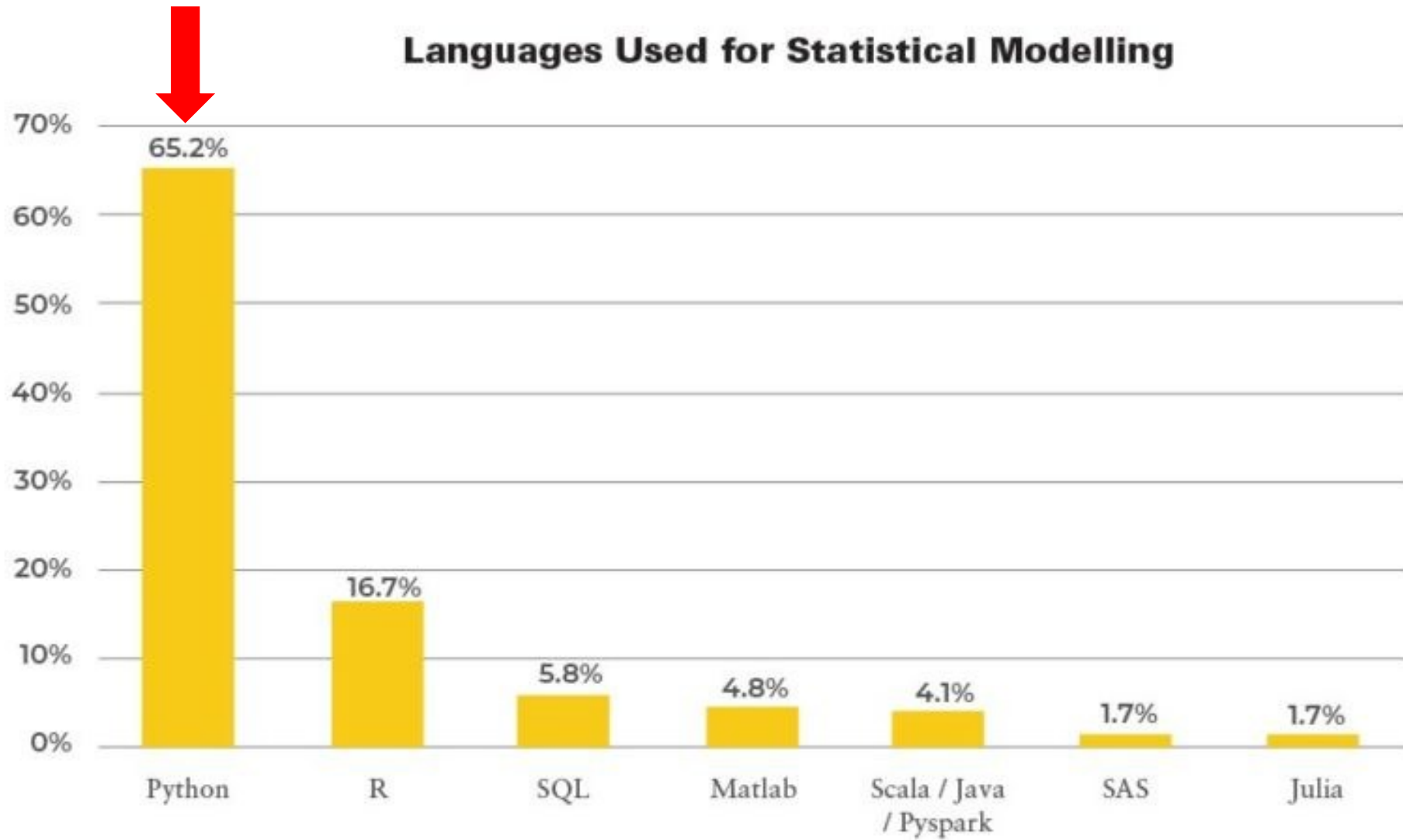
- Data Science Introduction
- Data Science Skills Survey

Data Science Skills

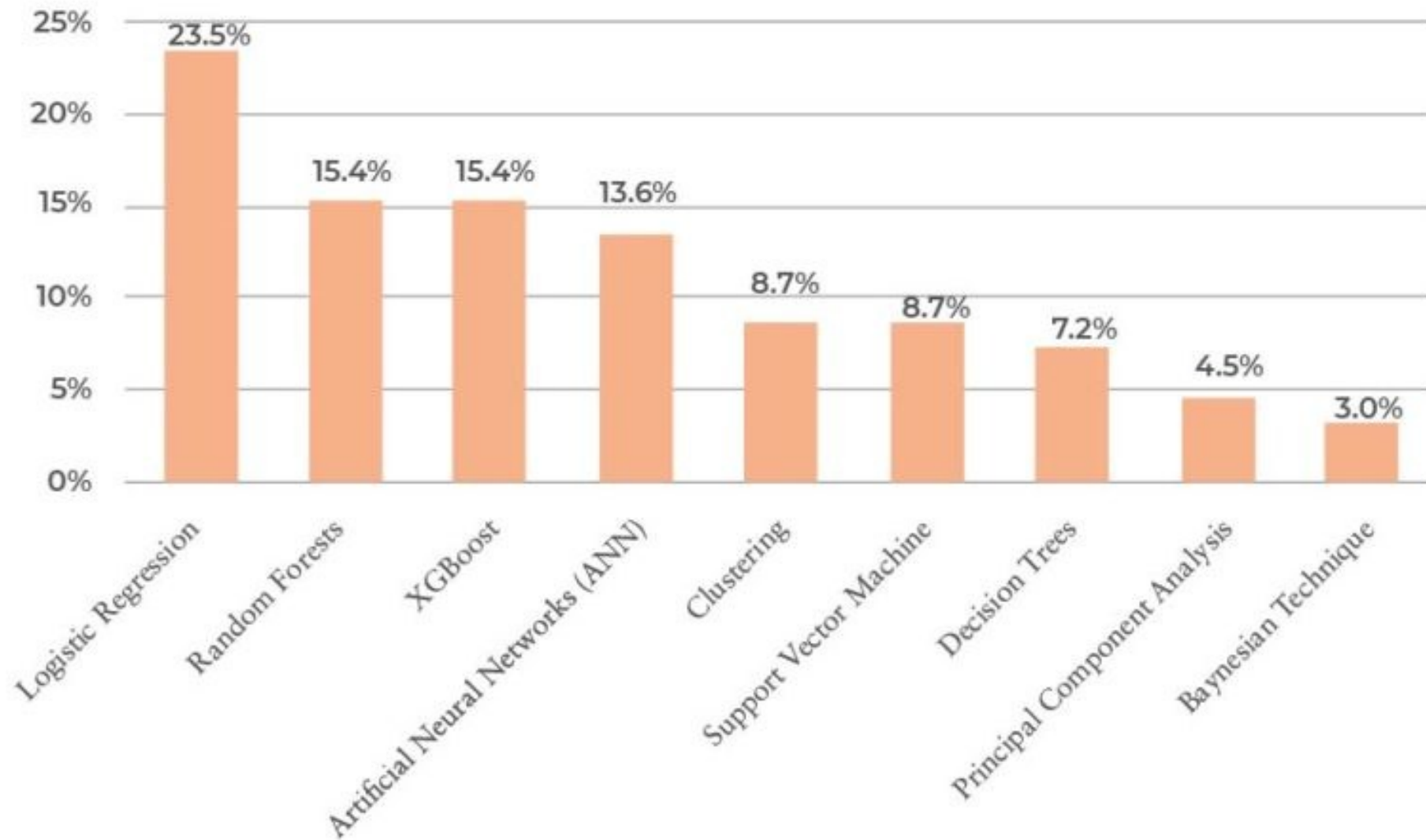
- **Data Science Skills Study 2020**
- <https://analyticsindiamag.com/aim-2020-data-science-skills-survey-aim-and-analytix-labs/>
- By AIM and AnalytixLabs
- Released on 17/8/2020



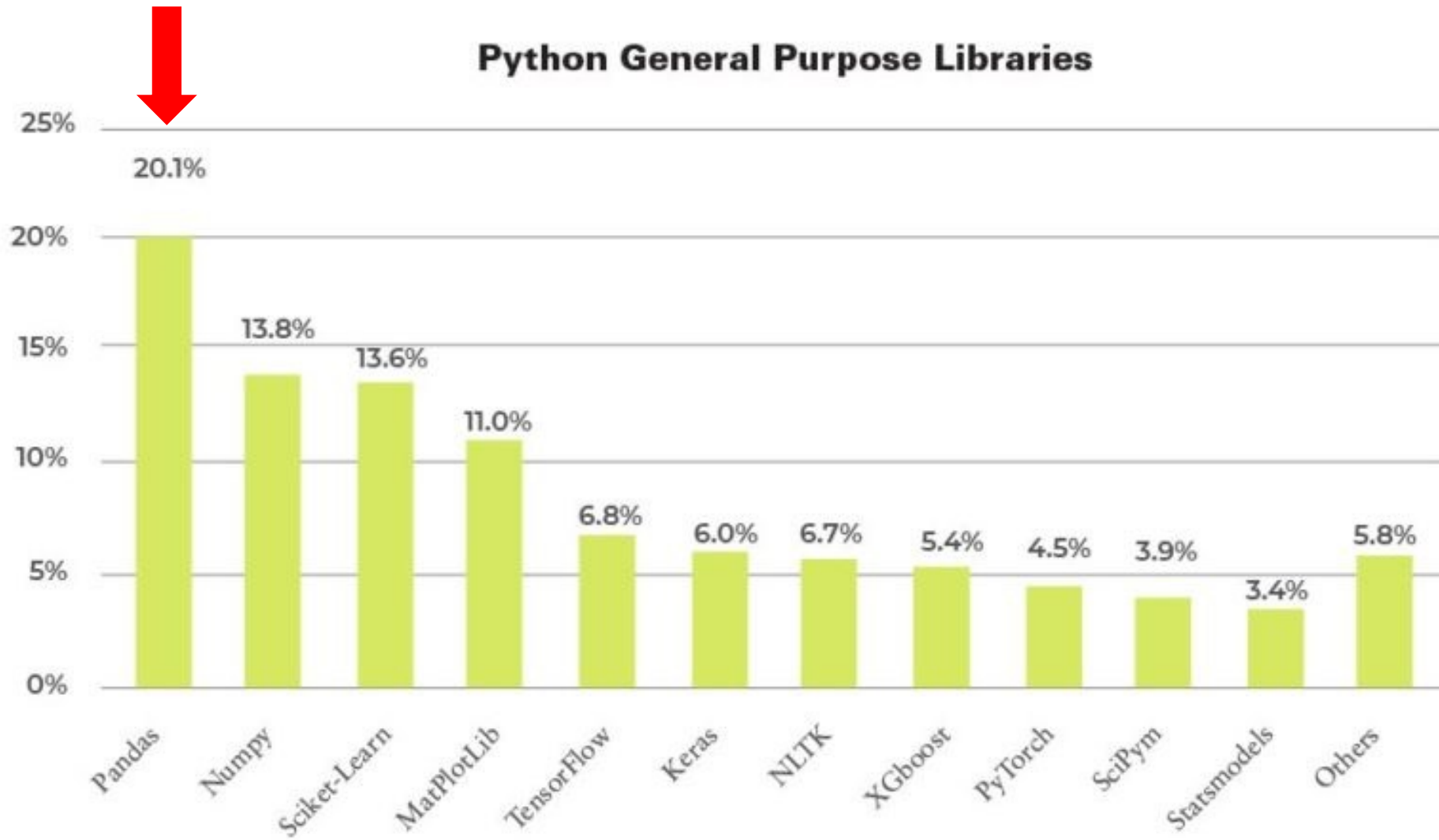
Languages Used for Statistical Modelling



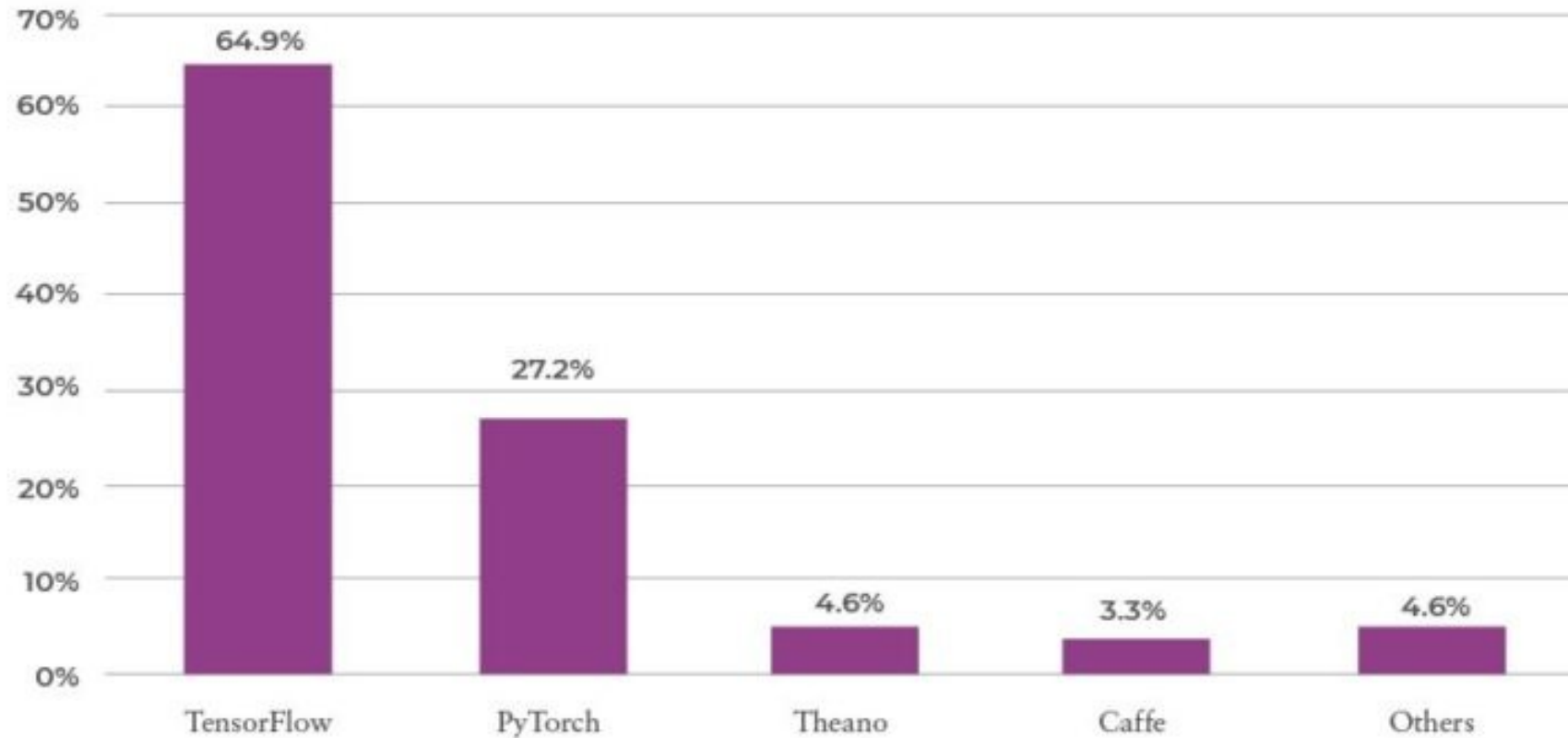
Data Science Models



Python General Purpose Libraries

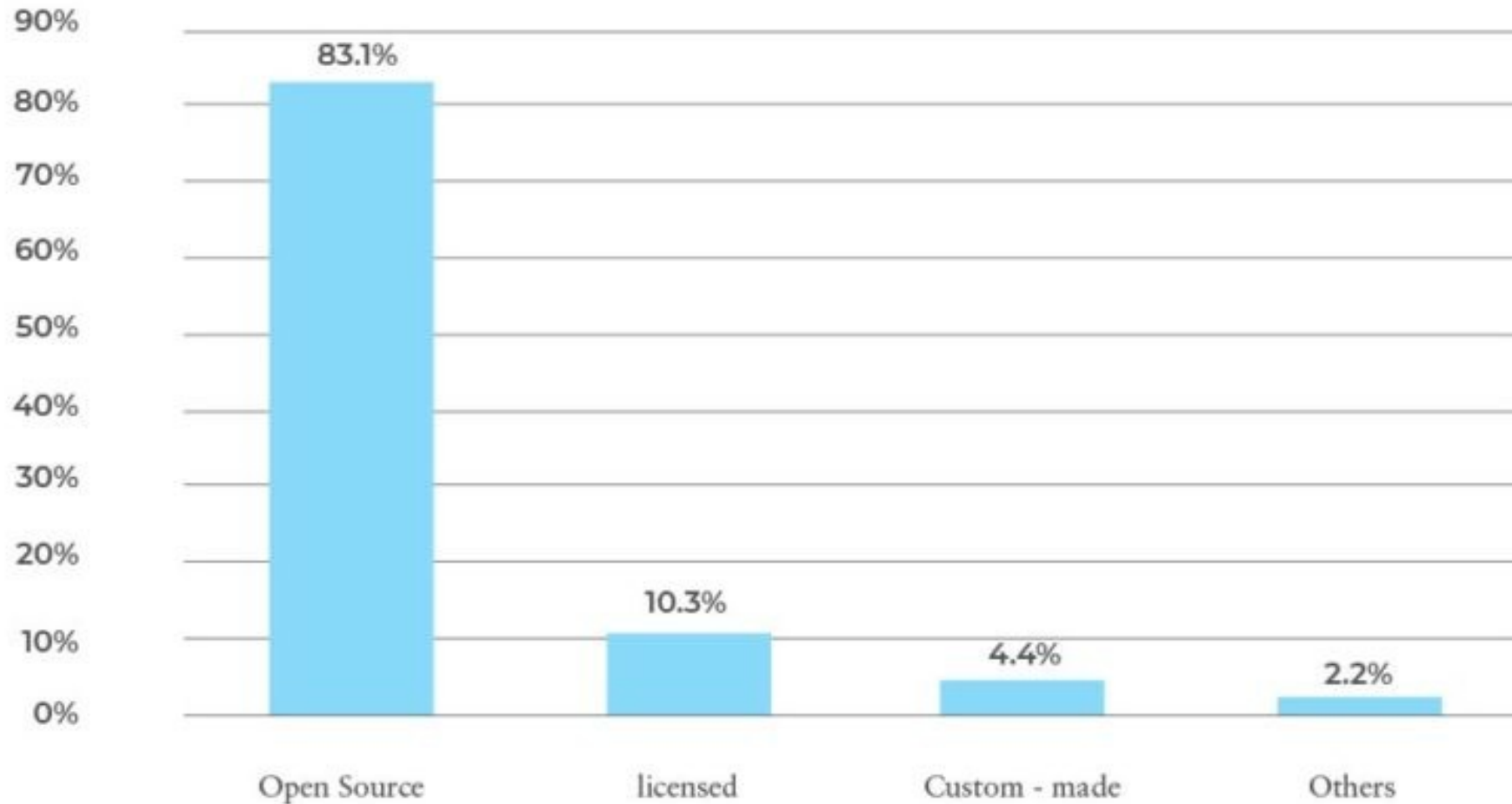


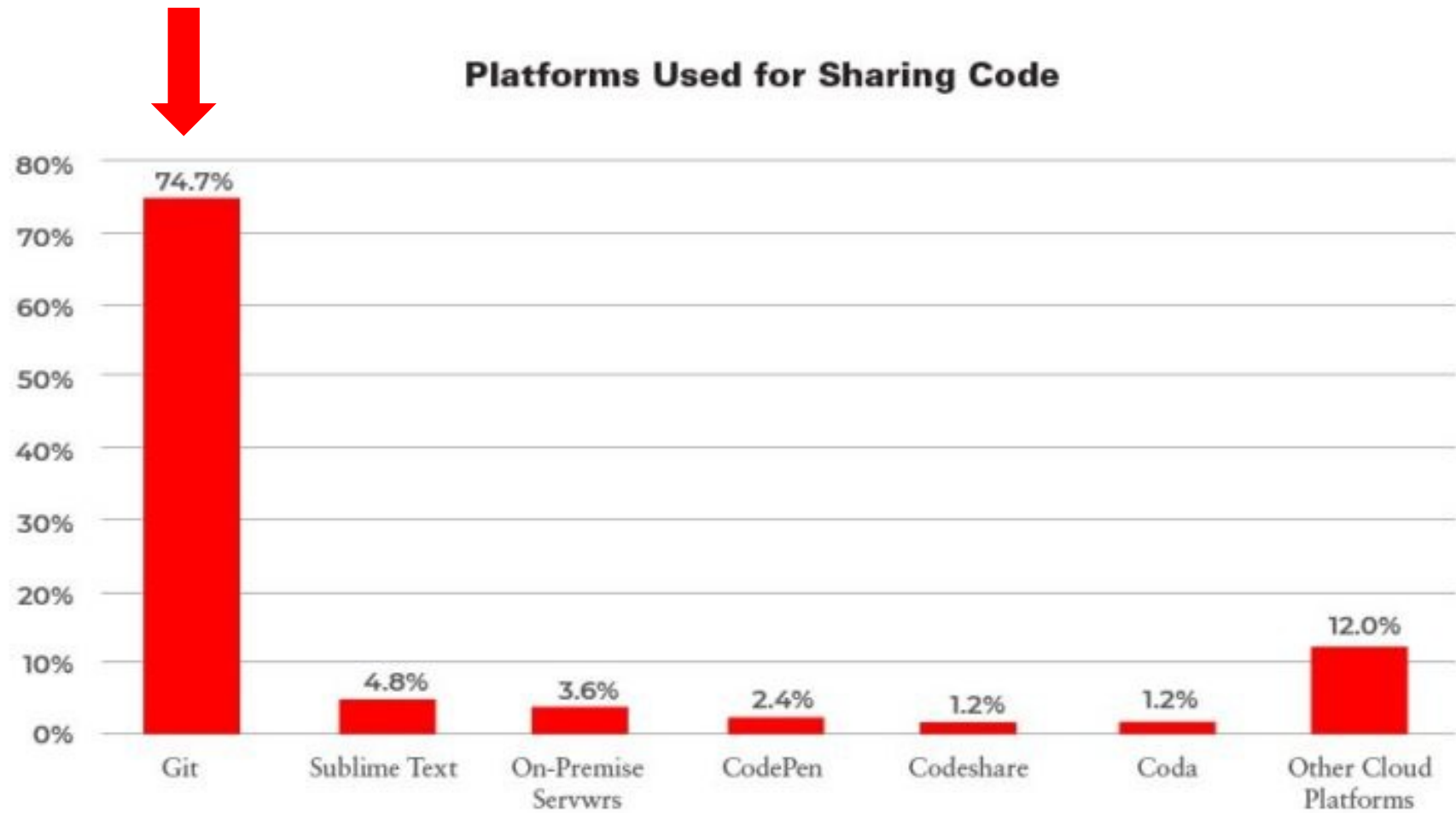
Python Frameworks Utilized for AI / Deep Learning Projects

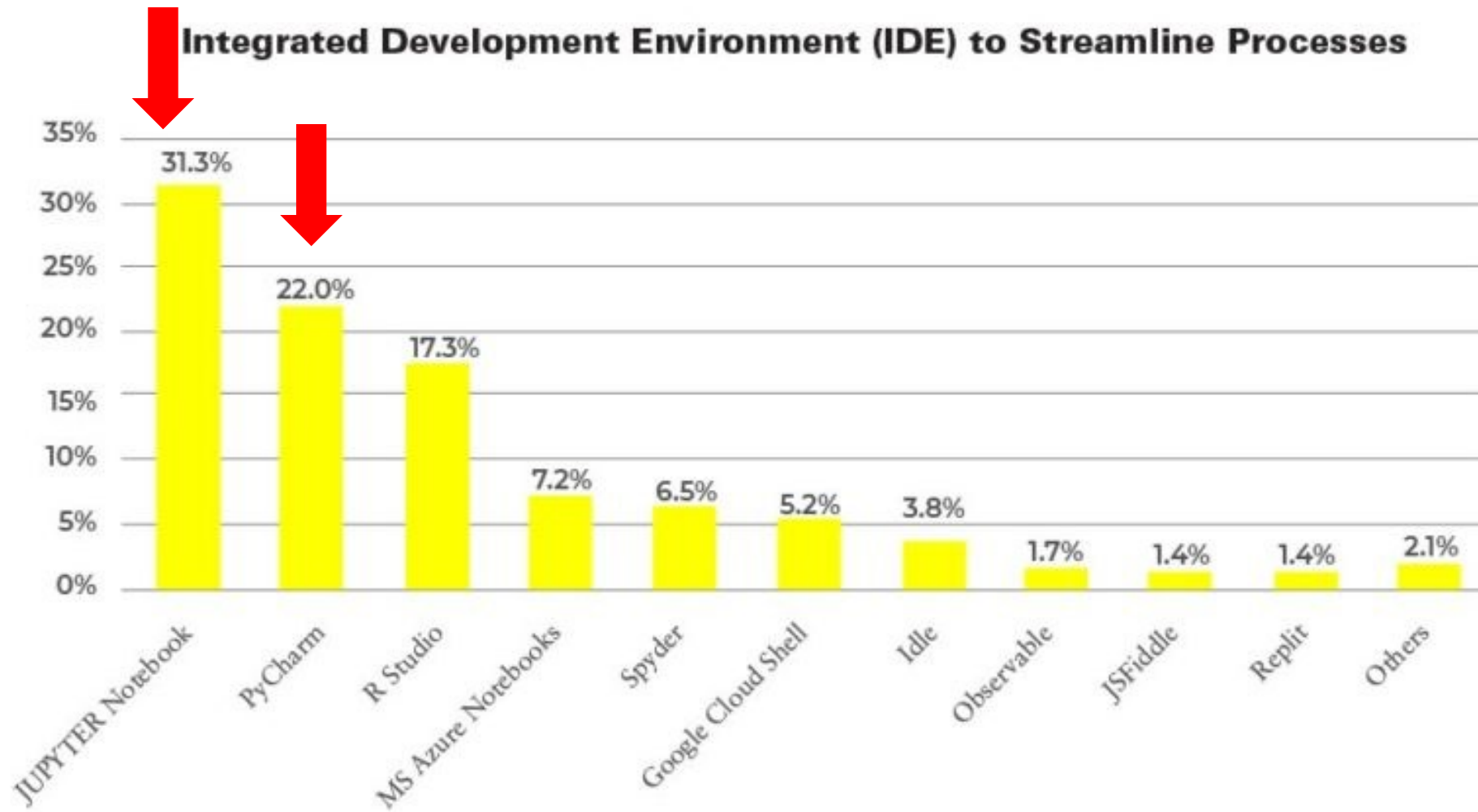




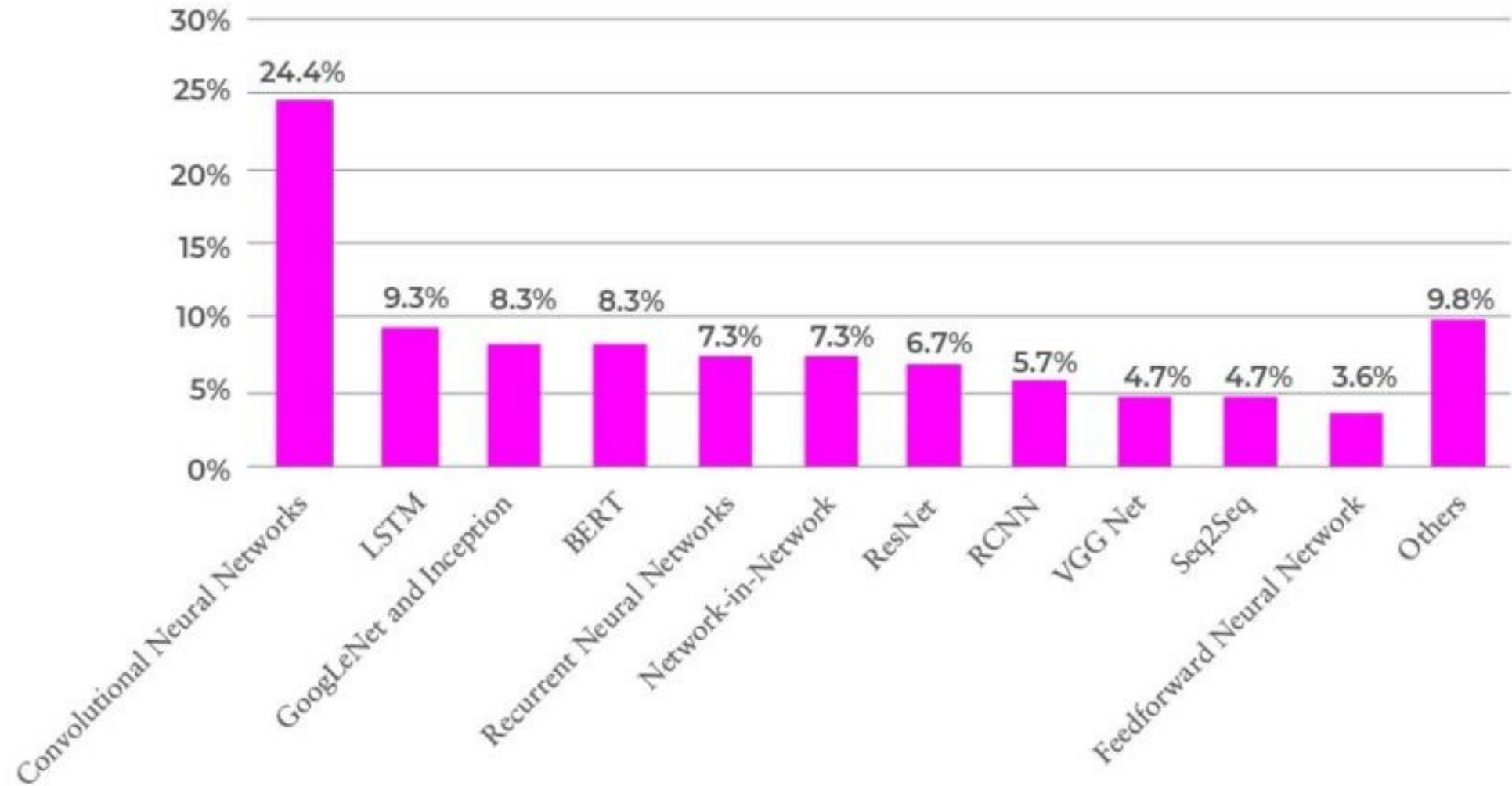
Preferred Platforms & Tools to Develop AI Models



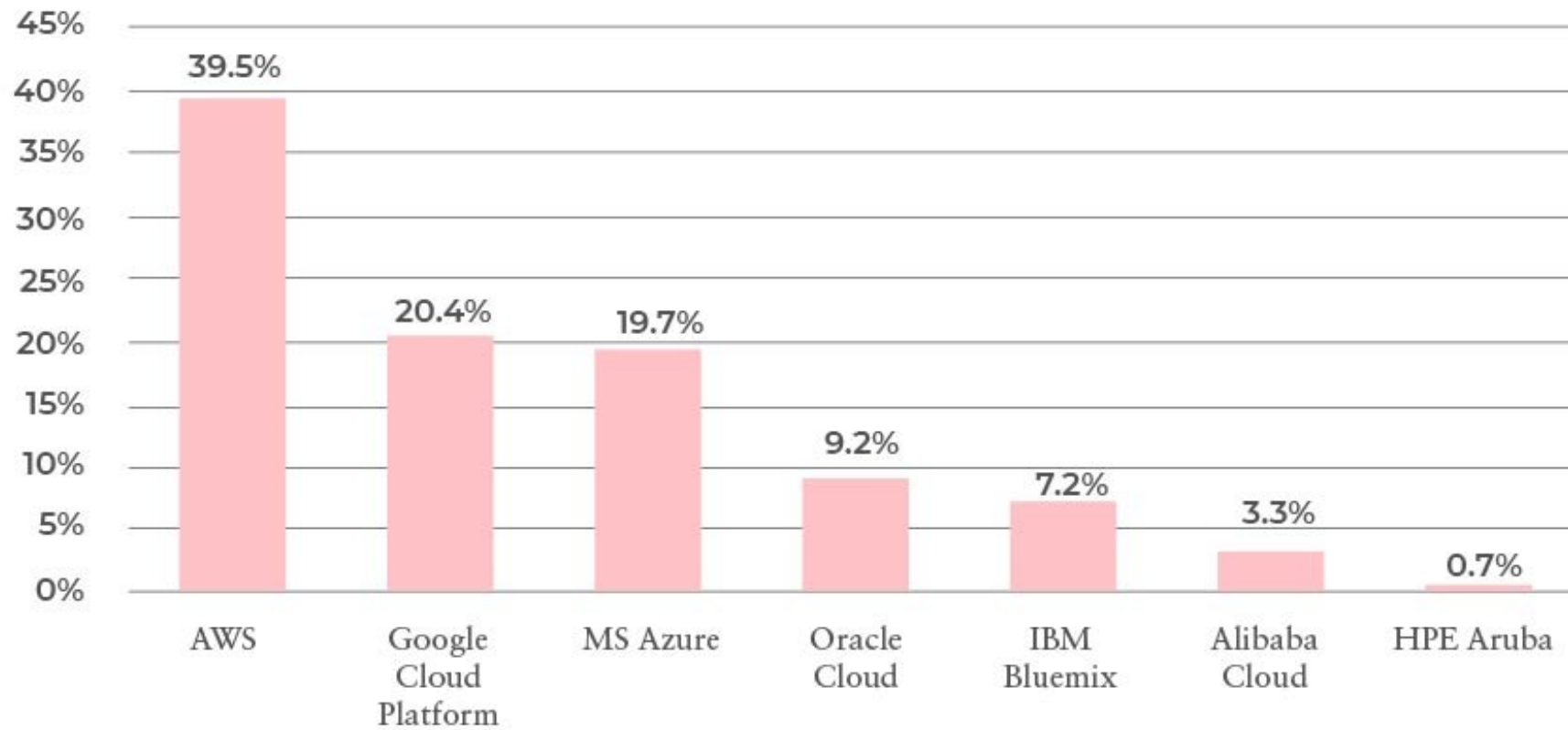




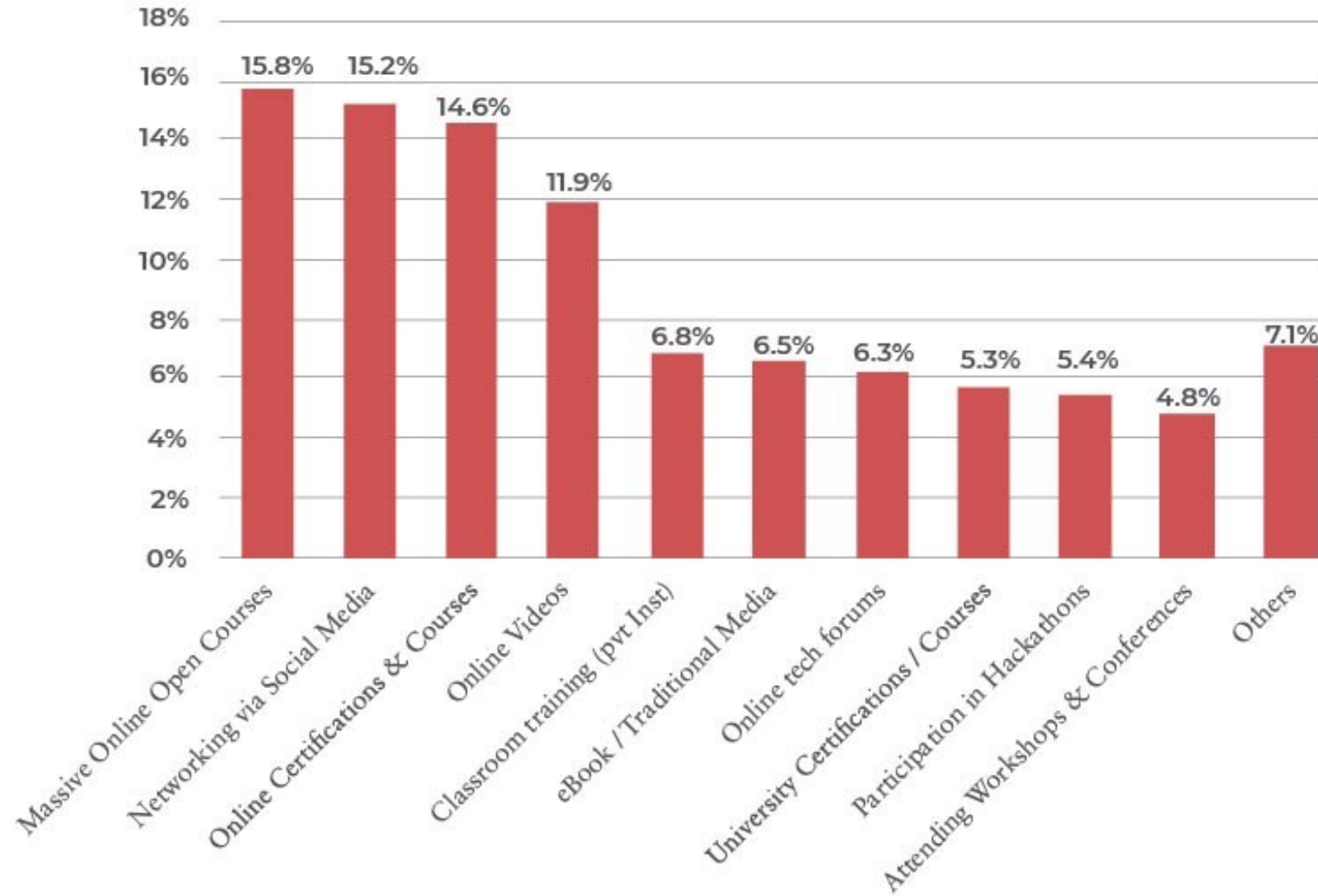
Neural Network Architectures



Cloud Service Platforms to Develop AI / ML Models



Learning Resources Utilized to Upskill



Summary

- Data Science Introduction
- Data Science Skills Survey