

# Data Science

## Course Introduction

Prof. Gheith Abandah

أ.د. غيث علي عبندة

# Outline

- Course Information
- Video: What REALLY is Data Science?
- Data Science Skills
- Textbooks
- References
- Course Outline
- Policies

# Course Information

- **Instructor:** Prof. Gheith Abandah
- **Email:** abandah@ju.edu.jo
- **Home page:** <http://www.abandah.com/gheith>
- **Duration:** 140 hours
- **Weekly Meetings (7 hours)**
  - Mondays 6:00 – 9:00 pm
  - Wednesdays 6:00 – 9:00 pm
  - Thursdays 5:30 – 6:30 pm

# What is Data Science?

- YouTube Video from **Joma Tech**

*What REALLY is Data Science? Told by a Data Scientist*

<https://youtu.be/xC-c7E5PK0Y>

# Data Science Hierarchy of Needs

## THE DATA SCIENCE **HIERARCHY OF NEEDS**

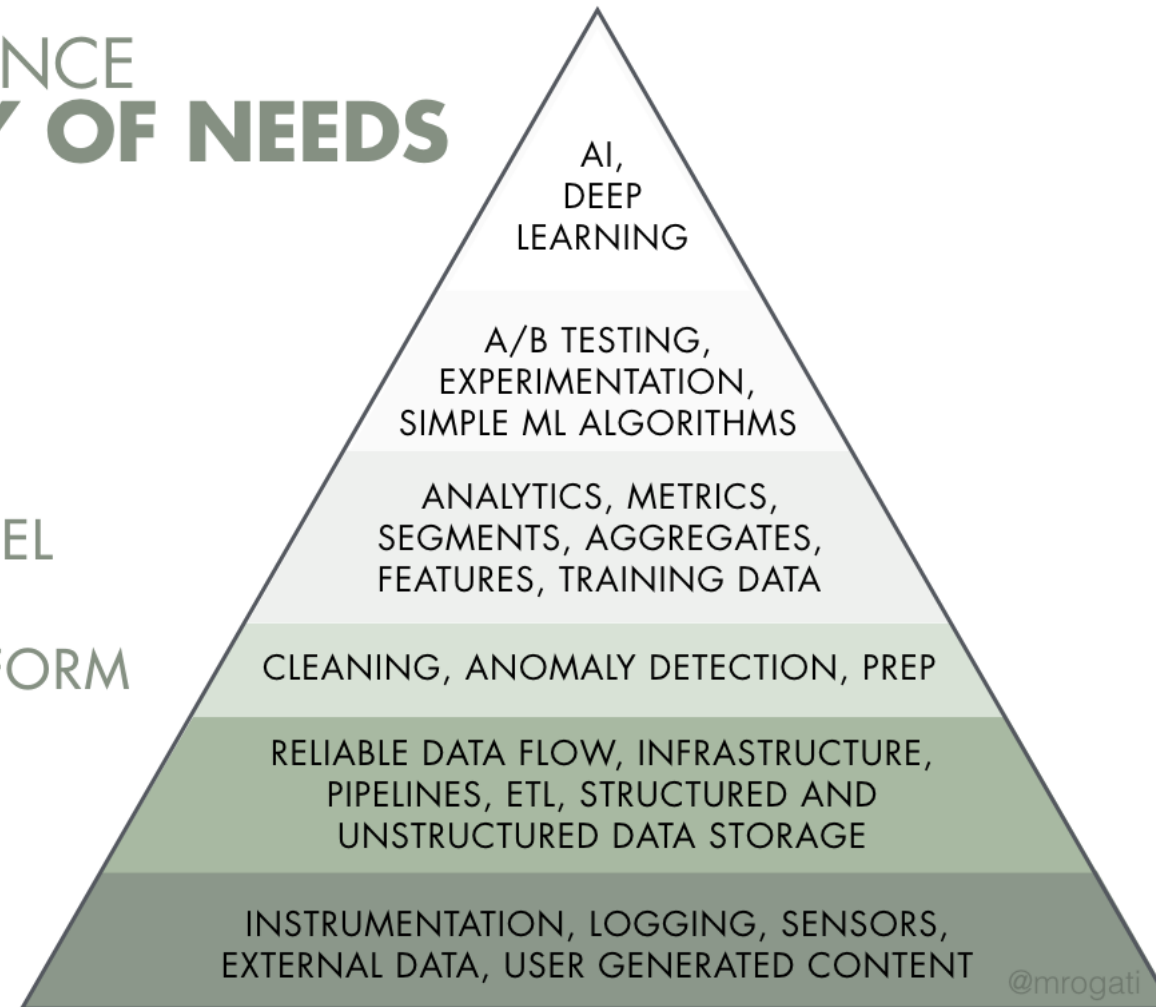
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

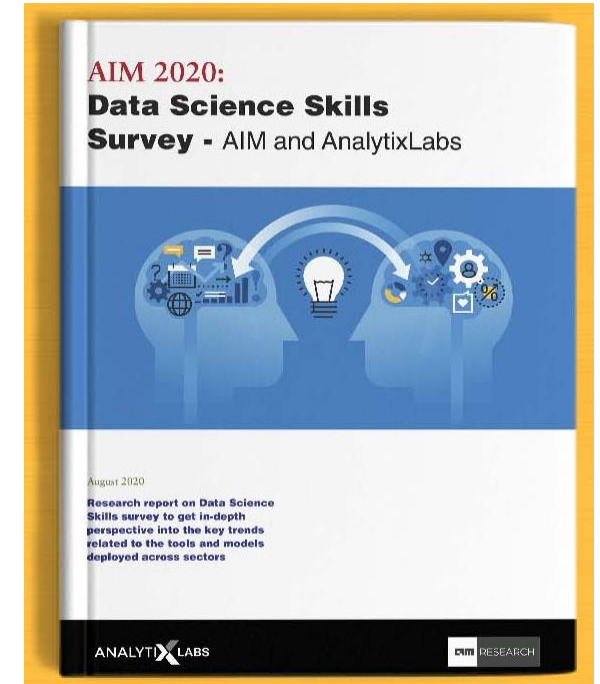
MOVE/STORE

COLLECT

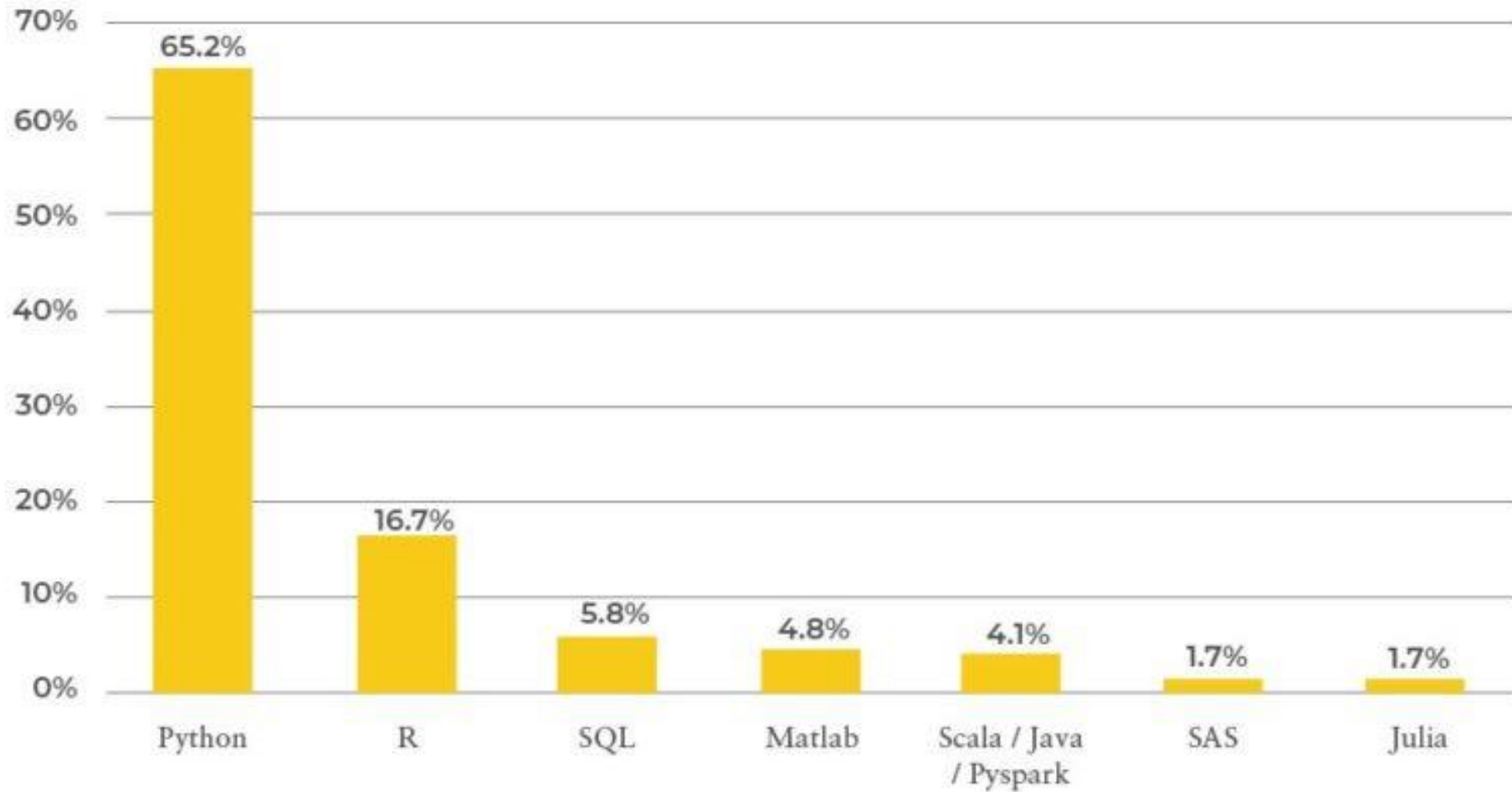


# Data Science Skills

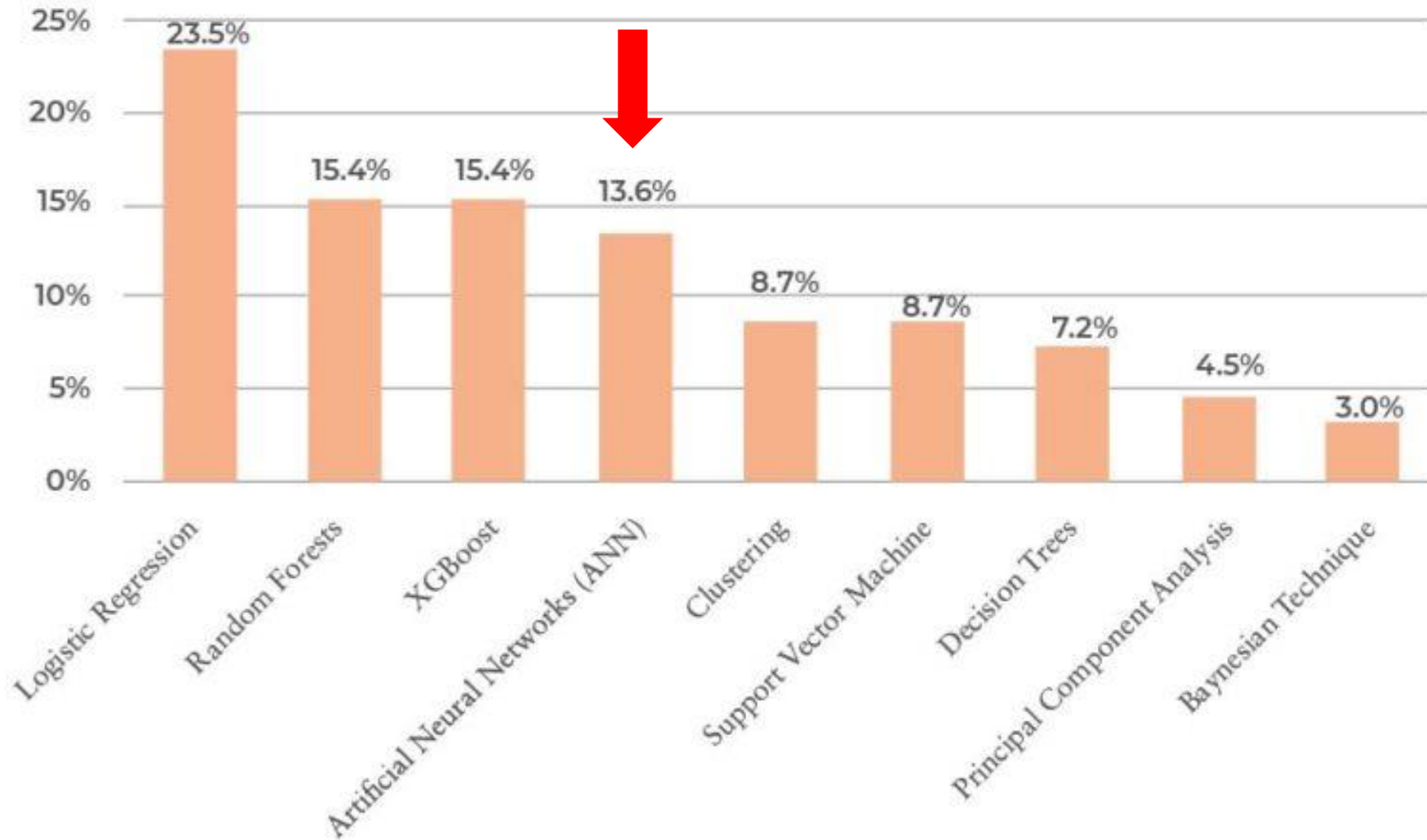
- **Data Science Skills Study 2020**
- <https://analyticsindiamag.com/aim-2020-data-science-skills-survey-aim-and-analytix-labs/>
- By AIM and AnalytixLabs
- Released on 17/8/2020



## Languages Used for Statistical Modelling

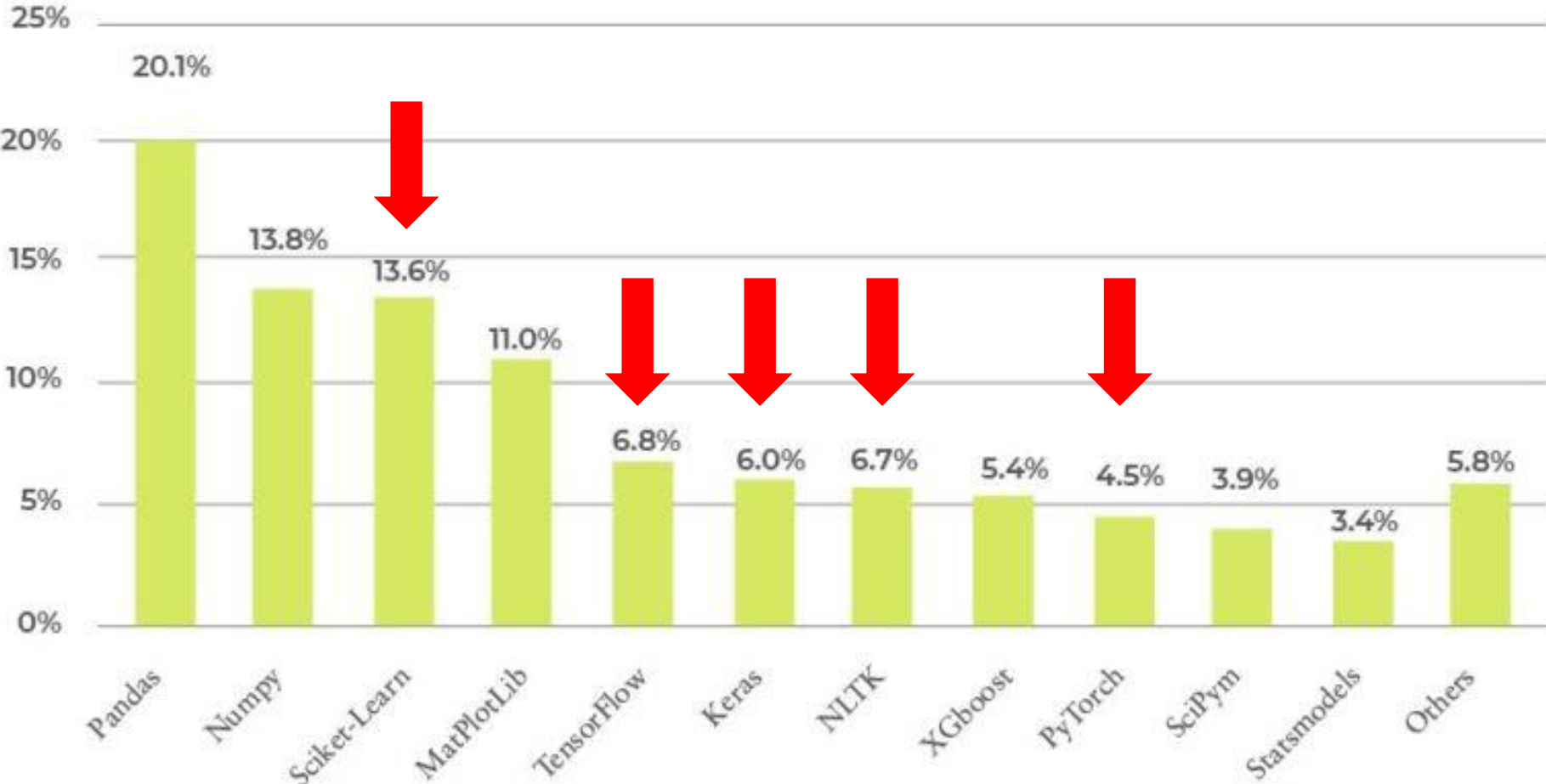


## Data Science Models

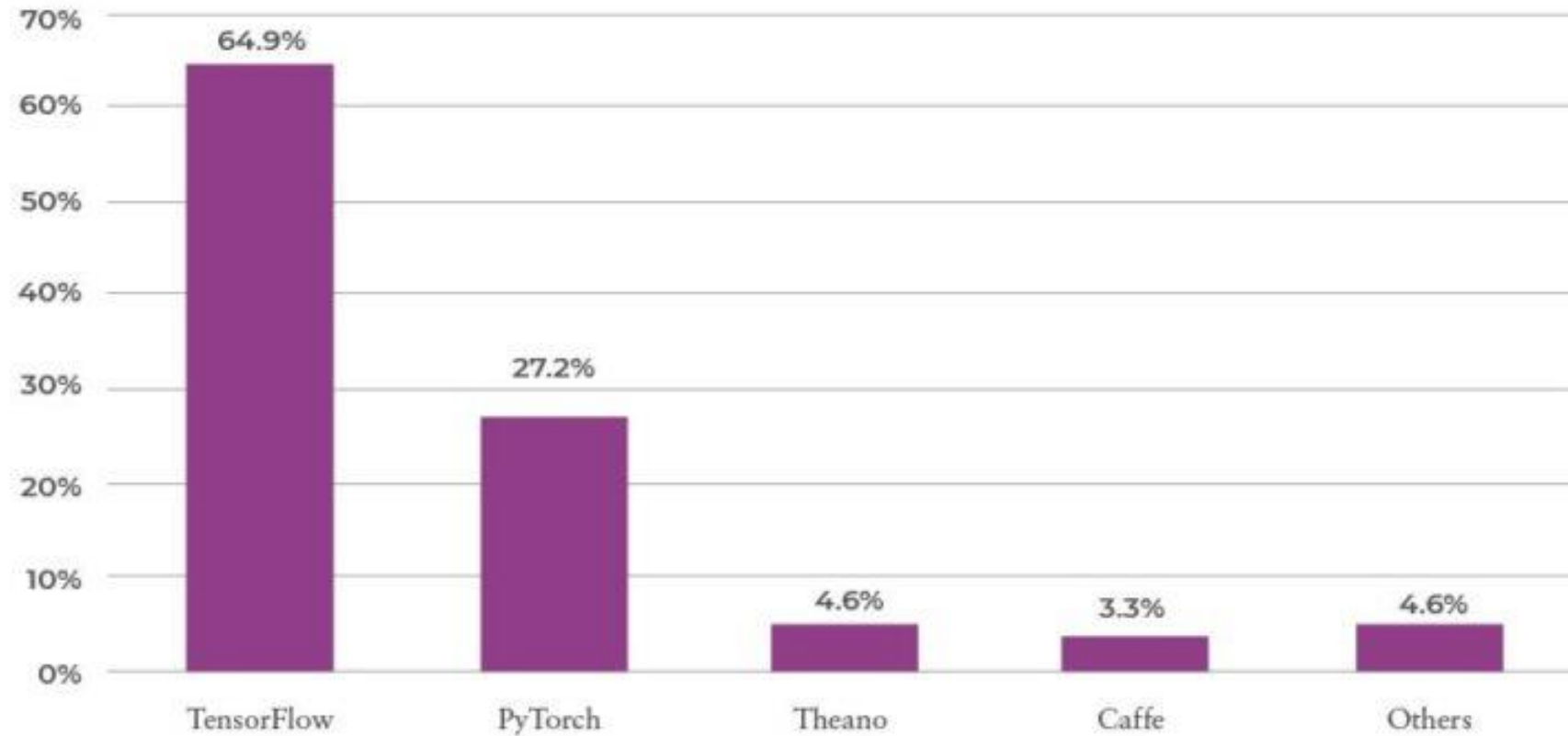




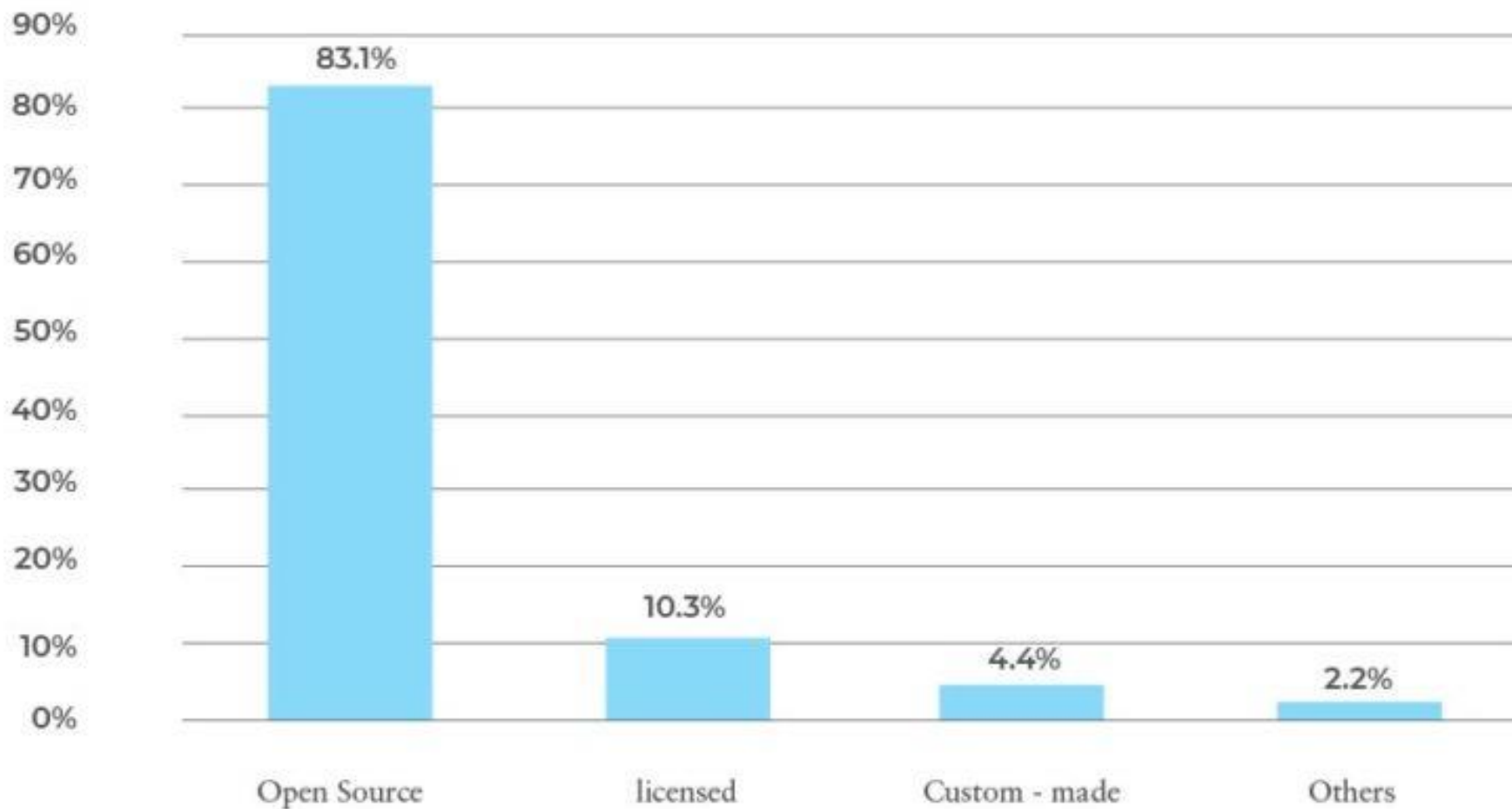
# Python General Purpose Libraries



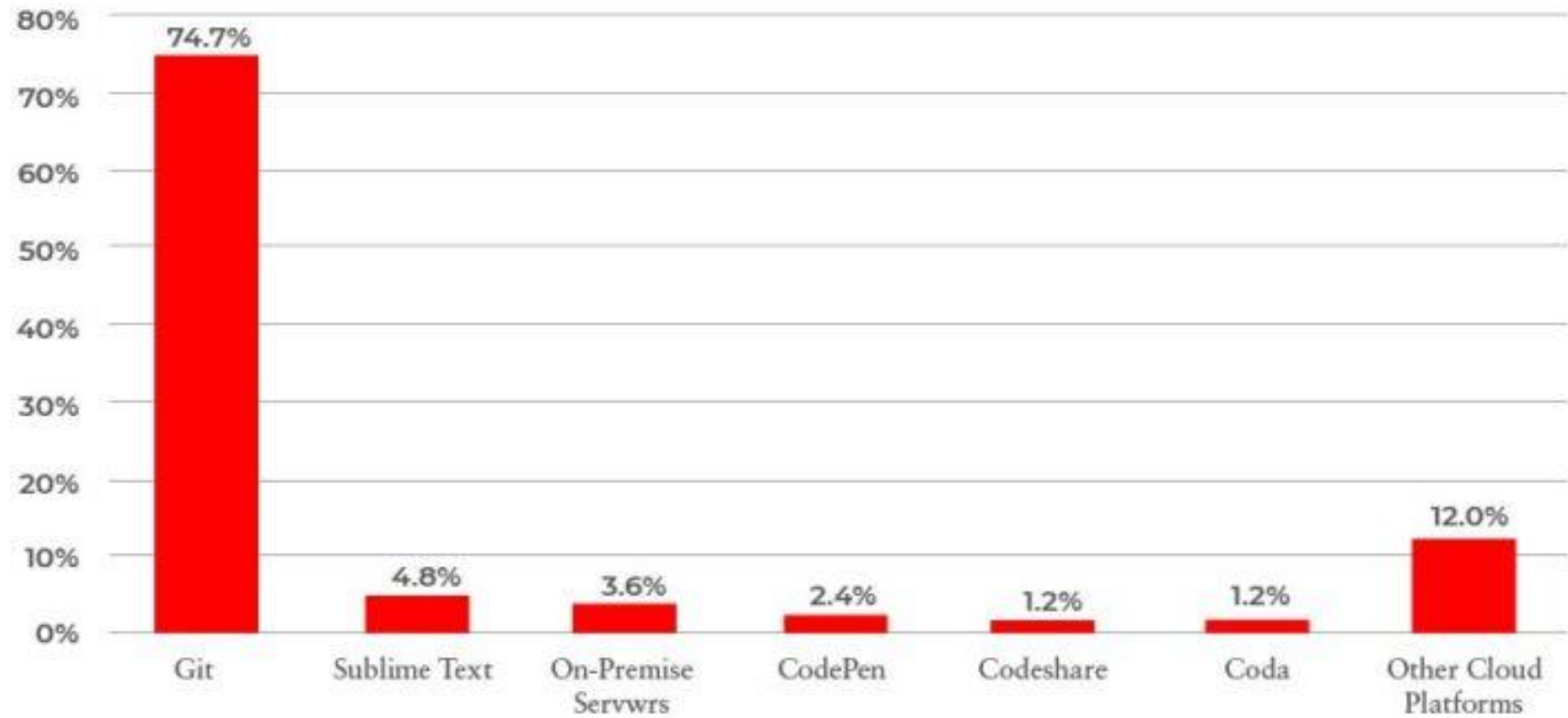
## Python Frameworks Utilized for AI / Deep Learning Projects



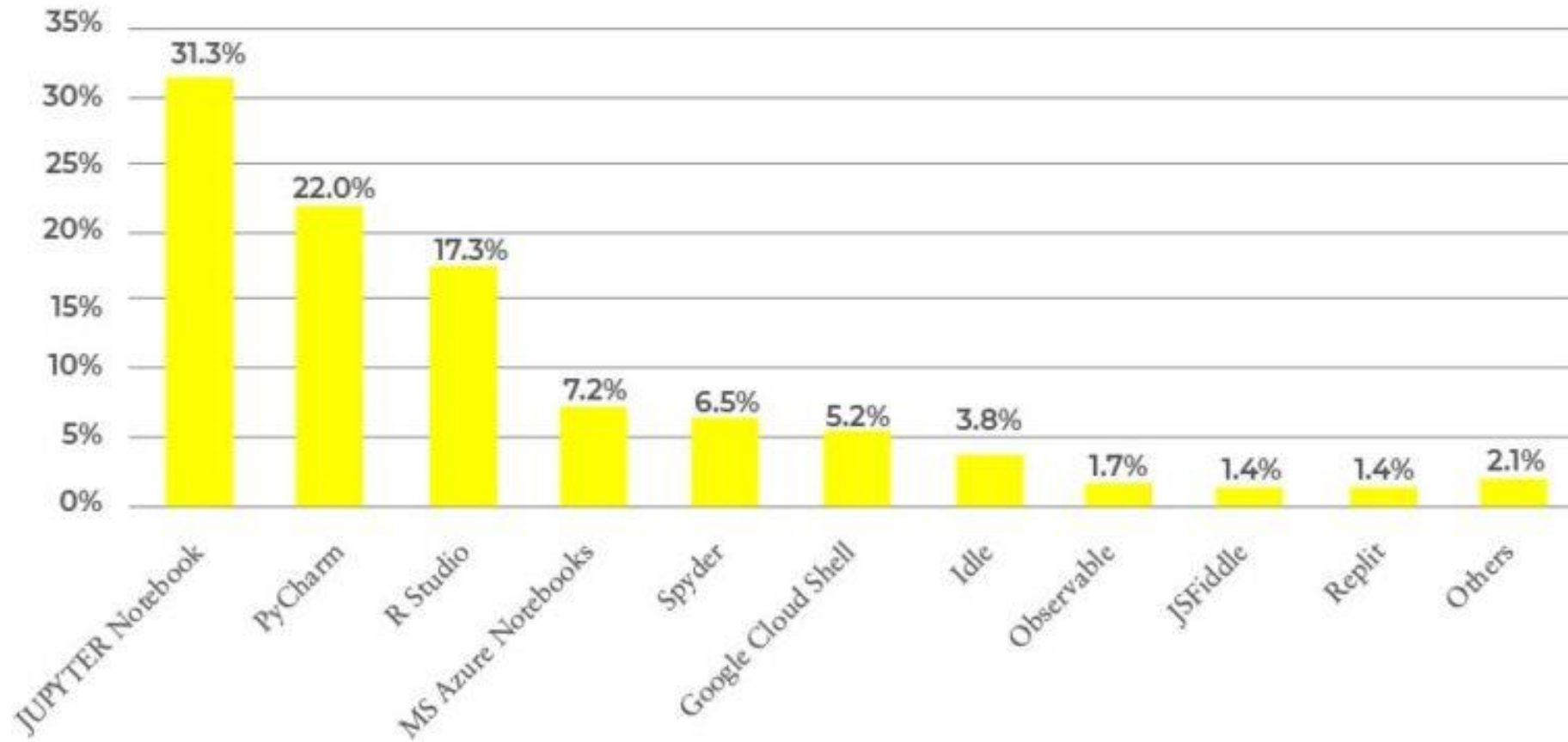
## Preferred Platforms & Tools to Develop AI Models



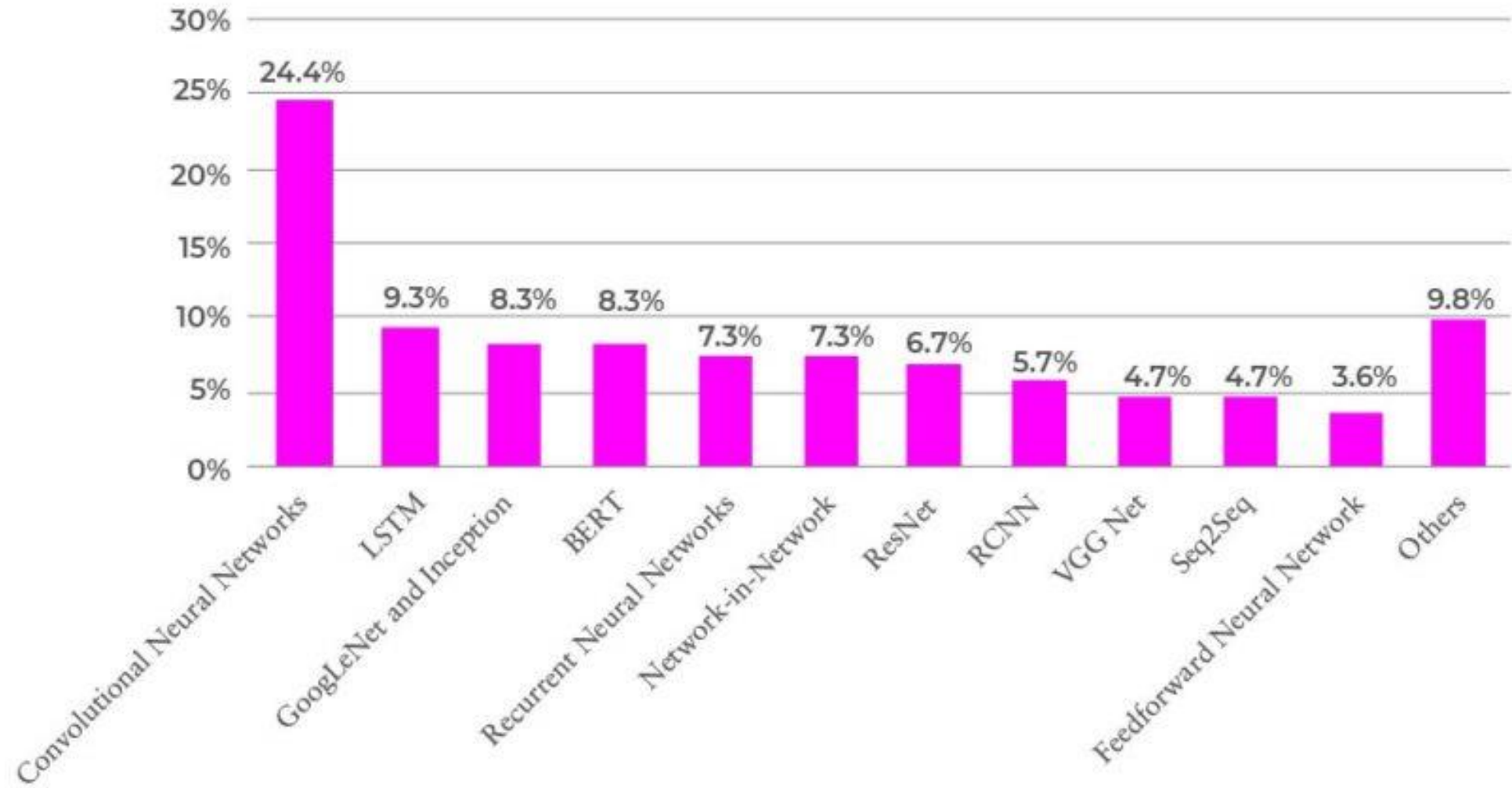
## Platforms Used for Sharing Code



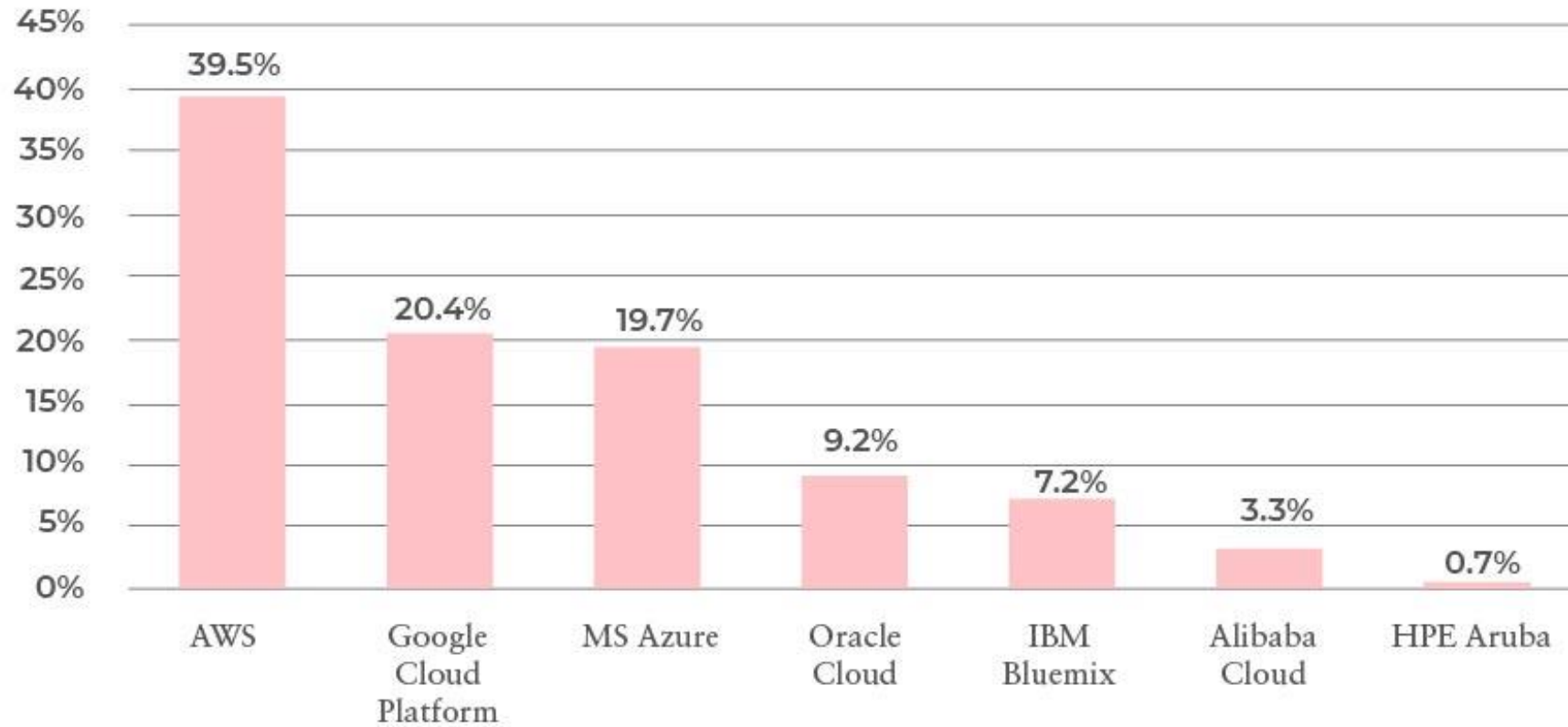
## Integrated Development Environment (IDE) to Streamline Processes



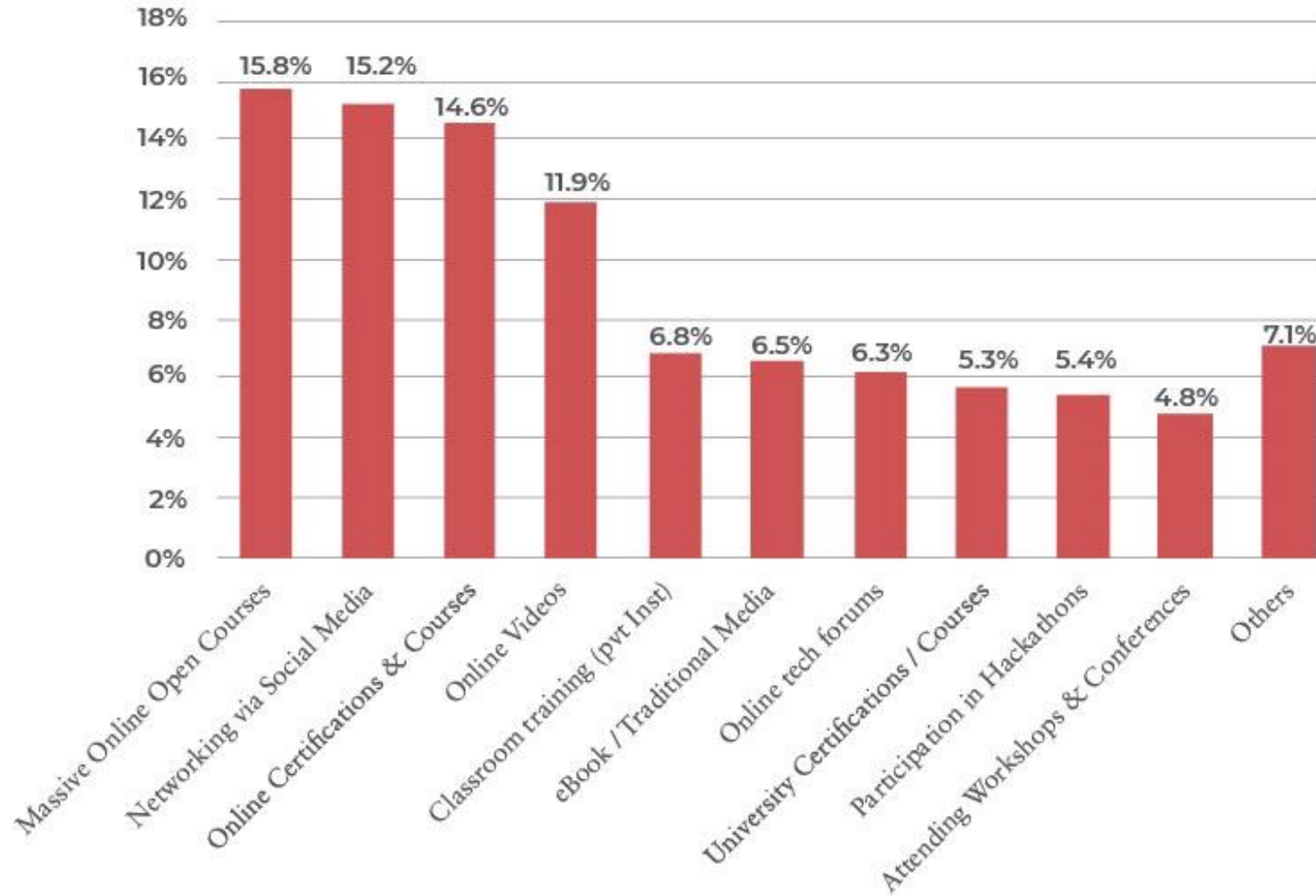
## Neural Network Architectures



## Cloud Service Platforms to Develop AI / ML Models



## Learning Resources Utilized to Upskill





# Textbooks

1. Wes McKinney, **Python for Data Analysis**: Data Wrangling with Pandas, NumPy, and Ipython, O'Reilly Media, 2nd Edition, 2018.
2. Aurélien Géron, **Hands-On Machine Learning** with Scikit-Learn, Keras and TensorFlow: Concepts: Tools, and Techniques to Build Intelligent Systems, 2nd Edition, O'Reilly Media, Oct 2019.
3. Arshdeep Bahga and Vijay Madisetti, **Big Data Analytics**: A Hands-On Approach, 2019.

# Python for Data Analysis

DATA WRANGLING WITH PANDAS,  
NUMPY, AND IPYTHON



powered by



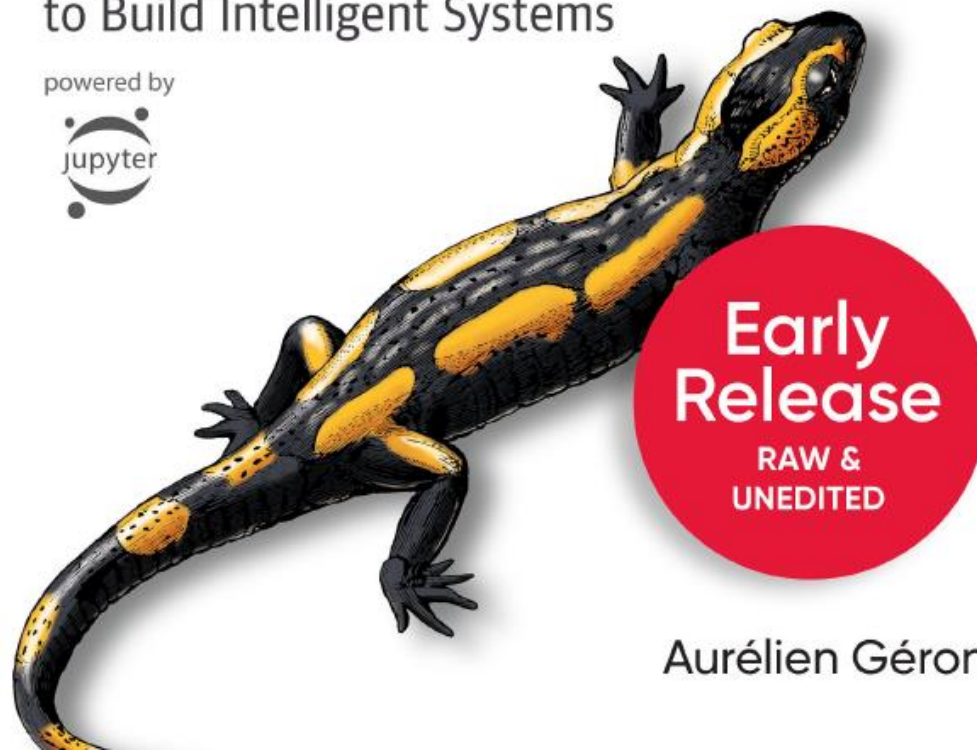
Wes McKinney

O'REILLY®

# Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow

Concepts, Tools, and Techniques  
to Build Intelligent Systems

powered by



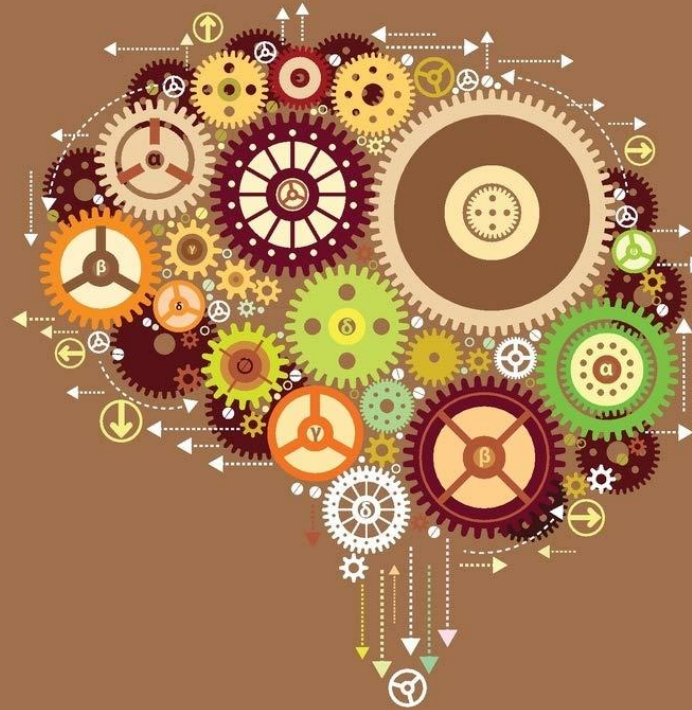
**Early  
Release**  
RAW &  
UNEDITED

Aurélien Géron

2nd Edition  
Updated for  
TensorFlow 2

# Big Data Science & Analytics

A Hands-On Approach



Arshdeep Bahga • Vijay Madisetti

# References

1. Jake VanderPlas, **A Whirlwind Tour of Python**, O'Reilly Media, 2016.
2. Joel Gurs, **Data Science from Scratch**, O'Reilly Media, 2015.
3. Prateek Joshi, **Artificial Intelligence with Python**, Packt Publishing, 2017.
4. François Chollet, **Deep Learning with Python**, Manning Pub. 2018.

# Course Outline

1. Course Introduction	1
2. Python Programming Language	11
3. Data Analysis	38
4. Machine Learning and Deep Learning	43
5. Big Data Analytics	24
6. Final Project Presentations and Evaluation	3
7. Discussion and Applications (1 hr per week)	20

# Course Outline

No	Topic	Hours	Total
1	Course Introduction		1
2	Python Programming Language		11
2.1	Introduction to Python, Anaconda and PyCharm	2	
2.2	Python Basics, IPython and Jupyter Notebooks	3	
2.3	Built-in Data Structures, Functions and Files	3	
2.4	Object Oriented Programming in Python	2	
2.5	Exercises	1	

# Course Outline

No	Topic	Hours	Total
<b>3</b>	<b>Data Analysis</b>		<b>38</b>
<b>3.1</b>	NumPy Basics: Arrays and Vectorized Computation	6	
<b>3.2</b>	Pandas Data Structures, Essential Functionality & Descriptive Statistics	6	
<b>3.3</b>	Data Loading, Storage and File Formats	5	
<b>3.4</b>	Data Cleaning and Preparation	6	
<b>3.5</b>	Data Wrangling: Join, Combine and Reshape	3	
<b>3.6</b>	Plotting and Visualization With Matplotlib and Seaborn	6	
<b>3.7</b>	Data Aggregation and Group Operations	3	
<b>3.8</b>	Time Series	3	



# Course Outline

No	Topic	Hours	Total
<b>4</b>	<b>Machine Learning and Deep Learning</b>		<b>43</b>
<b>4.1</b>	Introduction Machine Learning and Deep Learning	2	
<b>4.2</b>	End-to-End Machine Learning Project	6	
<b>4.3</b>	Classification: Basics, Evaluation Metrics and Advanced Topics	3	
<b>4.4</b>	Course Project Guidelines	1	
<b>4.5</b>	Training Models and Regression	3	
<b>4.6</b>	Classical Techniques	3	
<b>4.7</b>	Unsupervised Learning and Clustering	3	

# Course Outline

No	Topic	Hours	Total
4.8	Neural Networks	2	
4.9	Artificial Neural Networks With Keras	5	
4.10	Deep Neural Networks	3	
4.11	Deep Computer Vision Using Convolutional Neural Networks	4	
4.12	Recurrent Neural Networks	2	
4.13	LSTM Sequence to Sequence Translation	2	
4.14	Reinforcement Learning	3	
4.15	Recommender Systems	1	

# Course Outline

No	Topic	Hours	Total
<b>5</b>	<b>Big Data Analytics</b>		<b>24</b>
<b>5.1</b>	Introduction to Big Data	3	
<b>5.2</b>	Big Data Stack Setup and Examples	3	
<b>5.3</b>	Big Data Architectures and Patterns	3	
<b>5.4</b>	MapReduce Patterns	3	
<b>5.5</b>	NoSQL Databases	3	
<b>5.6</b>	Data Acquisition	3	
<b>5.7</b>	Big Data Storage	1	
<b>5.8</b>	Batch Data Analysis	2	
<b>5.9</b>	Real-Time Analysis	2	
<b>5.10</b>	Interactive Querying	1	

# Course Outline

No	Topic	Hours	Total
6	Final Project Presentations and Evaluation		3
7	Discussion and Applications (1 hour per week)		20
	<b>Total</b>		<b>140</b>

# Policies

- Pass-fail program
- Attendance is required
- Prepare before class
- Solve assignments
- Experiment and try things
- Open-book exams