

# Introduction to the Machine Learning Lab

---

**Week #1 - 4 Sept '24**  
**ML Lab**

The image features a solid black background. In the top right and bottom left corners, there are abstract, organic shapes in shades of pink, magenta, and orange, resembling soft, glowing light or stylized flames. Centered on the black background is the text "What is AI Society??" in a bold, yellow, sans-serif font.

**What is AI Society??**

# Club Mission

AI Society at Arizona State University is committed to demystifying artificial intelligence and machine learning through hands-on, project-based learning. We aim to make these advanced technologies accessible to all, regardless of prior programming experience, fostering a community of innovative thinkers ready to tackle real-world challenges.



The image features a solid black background. In the top right and bottom left corners, there are abstract, organic shapes in shades of pink, magenta, and orange, resembling soft, glowing light or liquid. Centered on the black background is the text "What to expect from this series??" in a bold, yellow, sans-serif font.

**What to expect from this series??**

Learn to manage and prepare data, covering everything from loading data to normalization and handling missing values.

### Data Cleaning and Processing

Explore clustering algorithms and dimensionality reduction to discover patterns and groupings in data.

### Supervised and Unsupervised Learning Techniques

Dive into creating and selecting meaningful features from data, using techniques like tokenization and PCA. Implement and evaluate classification models such as Logistic Regression and Random Forest.

### Feature Engineering and Category Prediction

Combine all learned techniques to create a sophisticated recommender system in our final project.

### Building a Recommender System

# Google Colab Setup

# What is Colab ..... and why are we using it

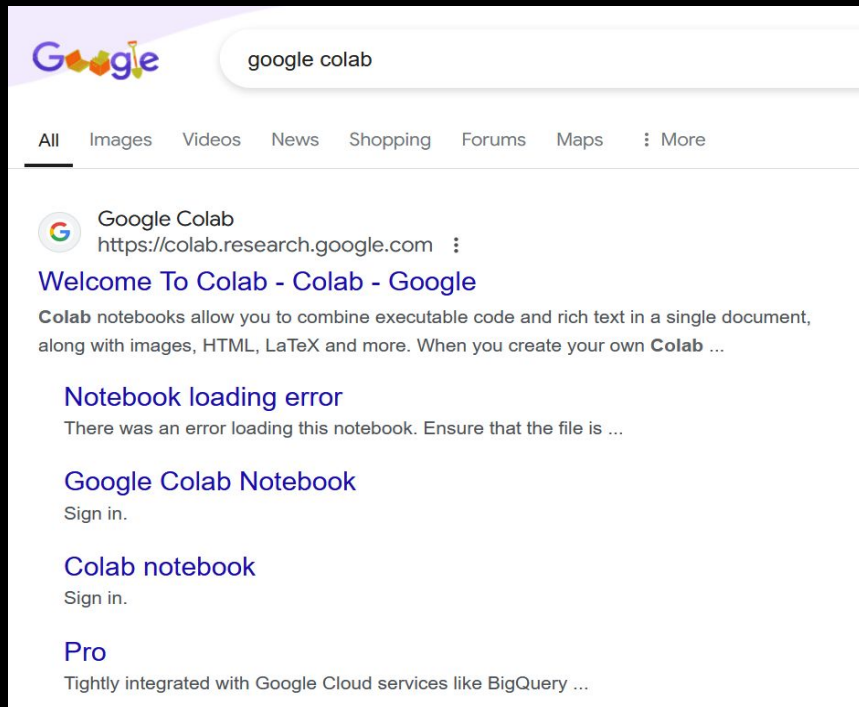
---

Google Colab, or Colaboratory, is a free, browser-based tool that allows users to write and run Python code. It's built around Project Jupyter code and allows users to run Jupyter notebooks without the need for local software installation.

The logo for Google Colab, featuring the word "colab" in a bold, lowercase, orange-yellow sans-serif font. The letters are slightly shadowed, giving them a 3D appearance as if they are floating above the dark background.

colab

# Step 1: Accessing Google Collab



The screenshot shows a Google search interface. At the top left is the Google logo. To its right is a search bar containing the text "google colab". Below the search bar is a horizontal menu with links: "All", "Images", "Videos", "News", "Shopping", "Forums", "Maps", and "More". The "All" link is underlined. Below the menu, the search results for "Google Colab" are displayed. The first result is from "https://colab.research.google.com" and includes a "Welcome To Colab - Colab - Google" heading. The text below the heading states: "Colab notebooks allow you to combine executable code and rich text in a single document, along with images, HTML, LaTeX and more. When you create your own Colab ...". Below this text is a section titled "Notebook loading error" with the message: "There was an error loading this notebook. Ensure that the file is ...". Further down are two more sections: "Google Colab Notebook" and "Colab notebook", both with a "Sign in." link. At the bottom is a section titled "Pro" with the text: "Tightly integrated with Google Cloud services like BigQuery ...".

Google Colab  
https://colab.research.google.com :  
**Welcome To Colab - Colab - Google**  
Colab notebooks allow you to combine executable code and rich text in a single document, along with images, HTML, LaTeX and more. When you create your own Colab ...

**Notebook loading error**  
There was an error loading this notebook. Ensure that the file is ...

**Google Colab Notebook**  
Sign in.

**Colab notebook**  
Sign in.

**Pro**  
Tightly integrated with Google Cloud services like BigQuery ...



# Step 2 - Creating a notebook

The image shows two screenshots from the Google Colaboratory interface. The left screenshot displays the 'Welcome to Colaboratory' window with the 'File' menu open. The 'New notebook' option is highlighted, and an orange arrow points from it to the 'Recent' section of the 'Open notebook' dialog shown in the right screenshot.

**Left Screenshot: Welcome to Colaboratory**

File Edit View Insert Runtime Tools Help

- New notebook
- Open notebook ⌘/Ctrl+O
- Upload notebook
- Rename
- Save a copy in Drive
- Save a copy as a GitHub Gist
- Save a copy in GitHub
- Save ⌘/Ctrl+S
- Revision history
- Download
- Print ⌘/Ctrl+P

**Right Screenshot: Open notebook**

Examples > Recent >

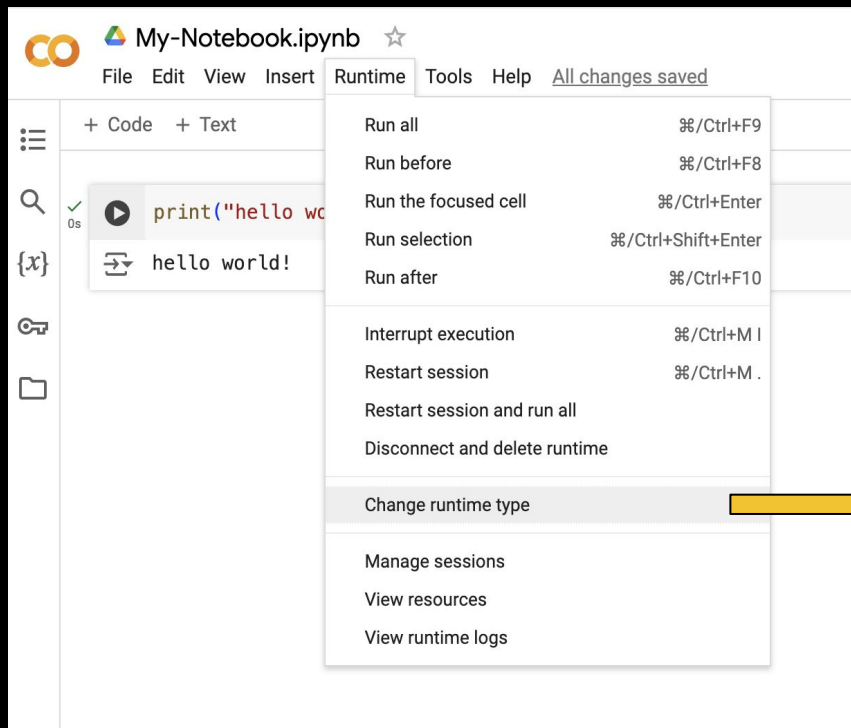
Google Drive > GitHub > Upload >

Search notebooks

Title	Last opened ↓	First opened ↑
Making the most of your colab subscription	08:48	08:48

+ New notebook Cancel

# Step 3 - Changing the Runtime



The screenshot shows the Google Colab interface for a notebook titled "My-Notebook.ipynb". The "Runtime" menu is open, displaying various execution options. The option "Change runtime type" is highlighted, and a yellow arrow points from it to the "Change runtime type" dialog box on the right.

My-Notebook.ipynb ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

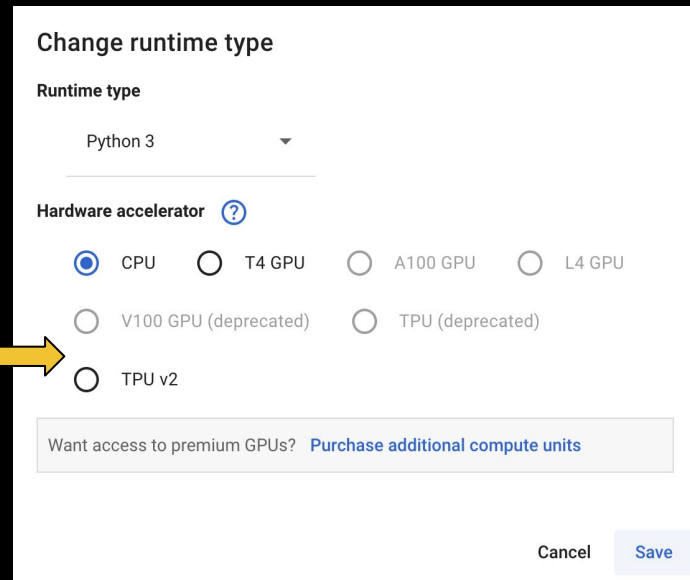
+ Code + Text

0s `print("hello world")`

`hello world!`

Runtime menu options:

- Run all ⌘/Ctrl+F9
- Run before ⌘/Ctrl+F8
- Run the focused cell ⌘/Ctrl+Enter
- Run selection ⌘/Ctrl+Shift+Enter
- Run after ⌘/Ctrl+F10
- Interrupt execution ⌘/Ctrl+M I
- Restart session ⌘/Ctrl+M .
- Restart session and run all
- Disconnect and delete runtime
- Change runtime type**
- Manage sessions
- View resources
- View runtime logs



The "Change runtime type" dialog box is shown. It allows users to select a runtime type and a hardware accelerator. The current runtime type is "Python 3". The hardware accelerator options include CPU, T4 GPU, A100 GPU, L4 GPU, V100 GPU (deprecated), TPU (deprecated), and TPU v2. A link to "Purchase additional compute units" is provided for users wanting access to premium GPUs. The "Save" button is highlighted.

Change runtime type

Runtime type

Python 3

Hardware accelerator ?

☒ CPU ☐ T4 GPU ☐ A100 GPU ☐ L4 GPU

☐ V100 GPU (deprecated) ☐ TPU (deprecated)

☐ TPU v2

Want access to premium GPUs? [Purchase additional compute units](#)

Cancel Save

# Google Collab Link

## Link to Code File

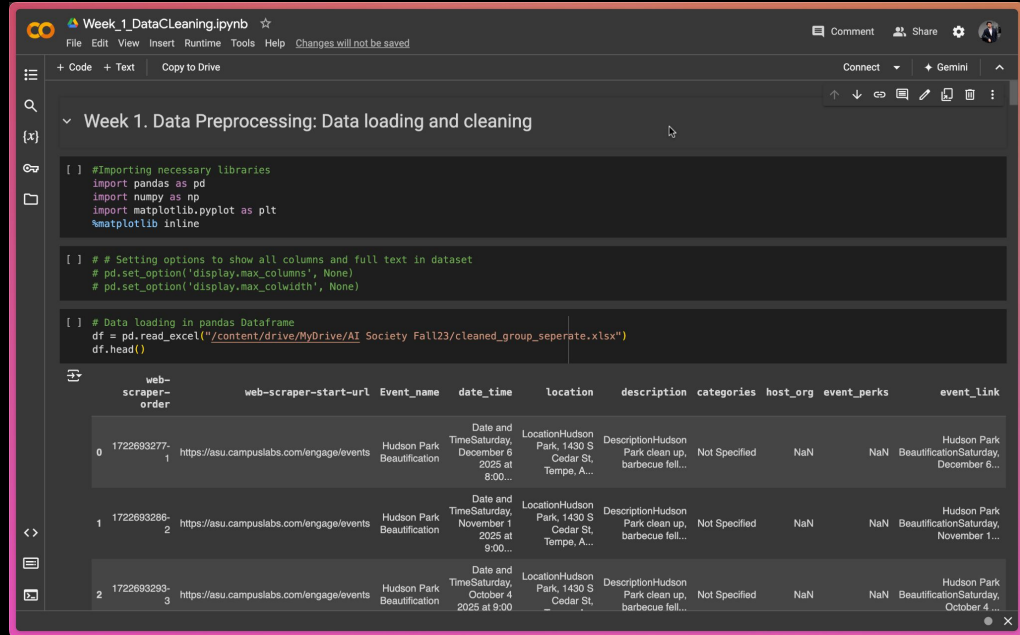
-> <https://bit.ly/4ebcjD3>

## Link to Discord

-> <https://bit.ly/AIS-Links>

## Link to Dataset

-> <http://bit.ly/3MAVZjn>



```
[ ] #Importing necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline

[ ] ## Setting options to show all columns and full text in dataset
# pd.set_option('display.max_columns', None)
# pd.set_option('display.max_colwidth', None)

[ ] # Data loading in pandas Dataframe
df = pd.read_excel("/content/drive/MyDrive/AI Society Fall23/cleaned_group_seperate.xlsx")
df.head()
```

	web-scraper-order	web-scraper-start-url	Event_name	date_time	location	description	categories	host_org	event_perks	event_link
0	1722693277-1	<a href="https://asu.campuslabs.com/engage/events">https://asu.campuslabs.com/engage/events</a>	Hudson Park Beautification	Date and TimeSaturday, December 6 2025 at 8:00...	LocationHudson Park, 1430 S Cedar St, Tempe, A...	DescriptionHudson Park clean up, barbecue fell...	Not Specified	NaN	NaN	Hudson Park BeautificationSaturday, December 6...
1	1722693286-2	<a href="https://asu.campuslabs.com/engage/events">https://asu.campuslabs.com/engage/events</a>	Hudson Park Beautification	Date and TimeSaturday, November 1 2025 at 9:00...	LocationHudson Park, 1430 S Cedar St, Tempe, A...	DescriptionHudson Park clean up, barbecue fell...	Not Specified	NaN	NaN	Hudson Park BeautificationSaturday, November 1...
2	1722693293-3	<a href="https://asu.campuslabs.com/engage/events">https://asu.campuslabs.com/engage/events</a>	Hudson Park Beautification	Date and TimeSaturday, October 4 2025 at 9:00...	LocationHudson Park, 1430 S Cedar St, Tempe, A...	DescriptionHudson Park clean up, barbecue fell...	Not Specified	NaN	NaN	Hudson Park BeautificationSaturday, October 4...

# Python and Its Libraries

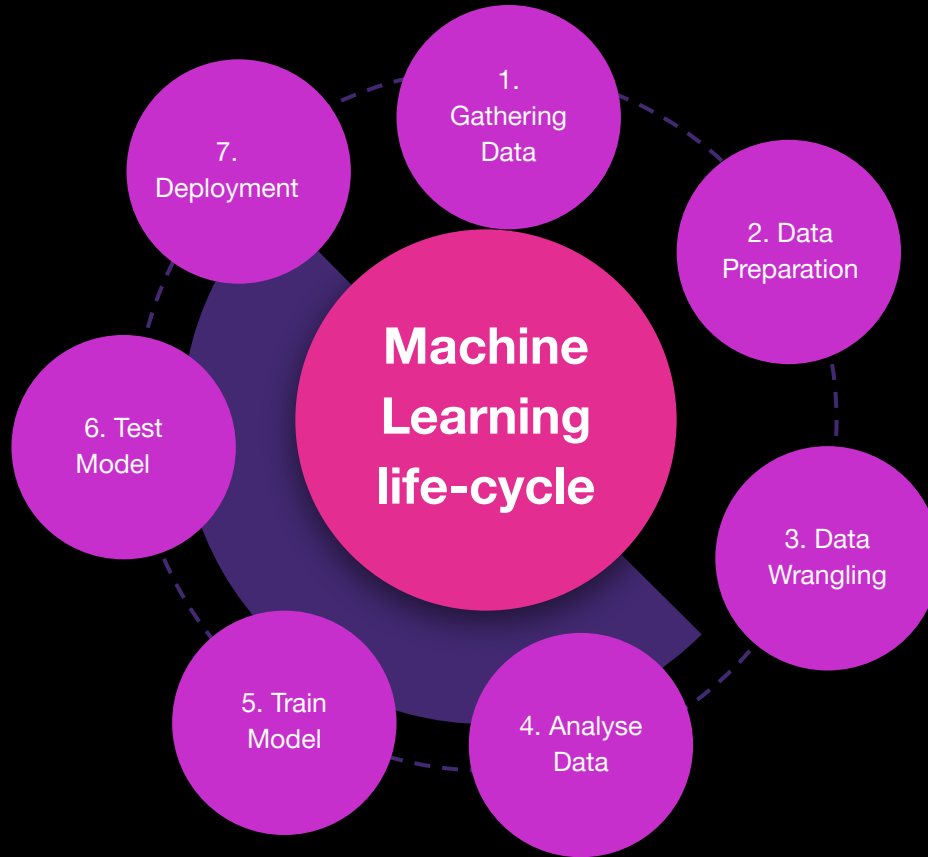
# Why Python??

---

Python is preferred in data science due to its simplicity and readability, vast library ecosystem, and strong community support.

## Core Libraries:





# Understanding Data

# What is DATA ?

---

**Data** refers to any information that can be used to train models or make predictions. This includes numbers, words, measurements or even images and sounds.

## Types of data

- Structured
- Unstructured
- Semi-Structured





# How important is Data??

## Data: The Lifeblood Of AI

At the core of every AI system lies a fundamental truth: The quality and quantity of data it ingests are paramount to its effectiveness. In essence, data is the lifeblood that fuels AI algorithms, allowing them to learn, adapt and make decisions.

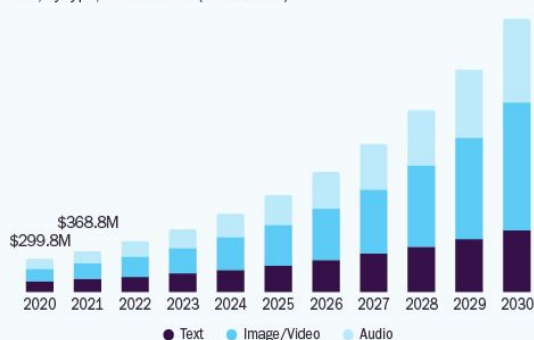
NOVEMBER 1, 2023

## AI is only as good as the data: Q&A with Satish Jayanthi of Coalesce

AI systems obey the golden rule: garbage in, garbage out, Want good results, feed it good data.

### Europe AI Training Dataset Market

size, by type, 2020 - 2030 (USD Million)



GRAND VIEW RESEARCH

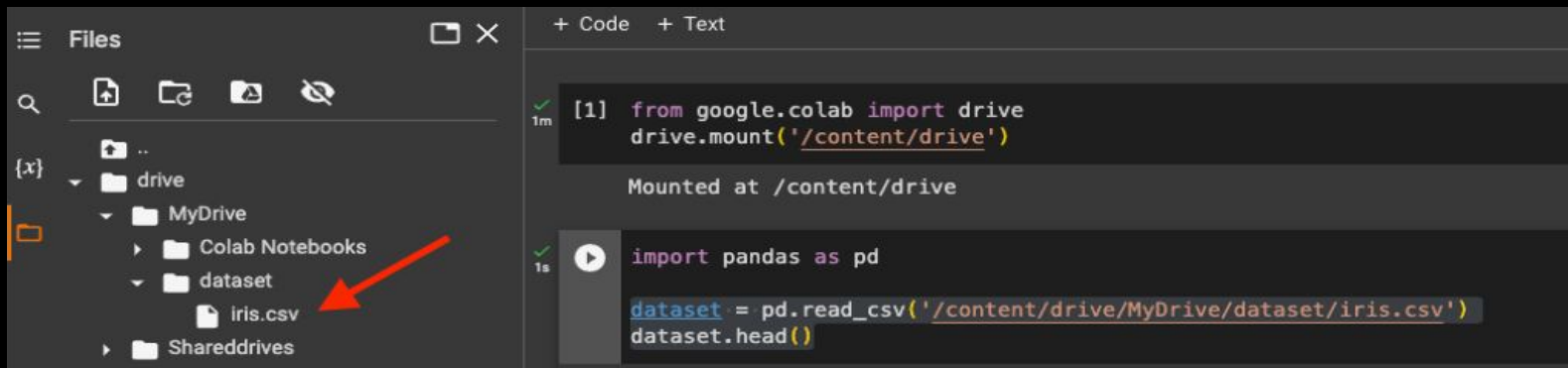
23.2%

Europe Market CAGR,  
2023 - 2030

Source:  
[www.grandviewresearch.com](http://www.grandviewresearch.com)

# File Handling

A CSV file is a Comma Separated Value file that stores tabular data in plain text. Each row represents a record, and each column represents a field in that record. The values in each row are separated by commas or other delimiters, such as semicolons or tabs

A screenshot of the Google Colab interface. On the left, the 'Files' pane shows a directory structure: 'drive' > 'MyDrive' > 'dataset' > 'iris.csv'. A red arrow points to the 'iris.csv' file. On the right, the code editor shows two code cells. The first cell contains code to mount the drive, and the second cell contains code to import pandas and read the CSV file. The output of the first cell is 'Mounted at /content/drive'.

```
+ Code + Text

[1] from google.colab import drive
    drive.mount('/content/drive')

Mounted at /content/drive

import pandas as pd

dataset = pd.read_csv('/content/drive/MyDrive/dataset/iris.csv')
dataset.head()
```

```
[ ] # Let's export the data for later use.
    result.to_csv('cleaned_data.csv', index=False)
```

# Data Aggregation

---

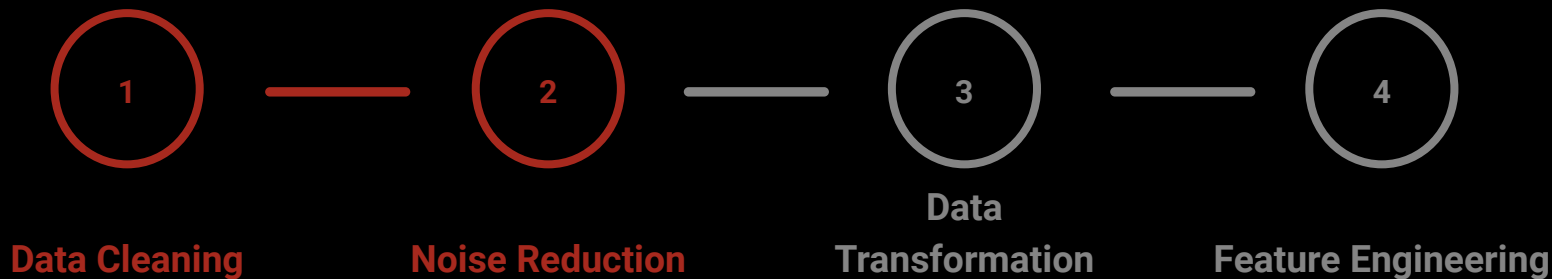
Data aggregation is the process of removing noise from the data by combining and organizing data from multiple sources into a single, unified body.

```
[ ] # Group by 'Event_name' and 'datetime', then aggregate categories and location_extracted
result = df.groupby(['Event_name', 'datetime', 'description']).agg({
    'host_org': lambda x: ', '.join(x[x != 'Unknown'].unique()),
    'event_perks': lambda x: ', '.join(x[x != 'None'].unique()),
    'categories': lambda x: ', '.join(x[x != "Not Specified"].unique()),
    'location_extracted': lambda x: ', '.join(x[x != 'Unknown'].unique()) # Exclude 'Unknown' locations
}).reset_index()
```

# Data Pre-processing

---

It involves transforming raw data into an understandable format. It is a critical step before any machine learning model training, ensuring the data is clean and suitable for analysis.



# Feature Engineering

---

## Definition:

Feature engineering is the process of using domain knowledge to select, modify, or create new features from raw data, enhancing the machine learning model's performance.

## Importance :

- Better Model Performance
- Reduce Overfitting
- Improved accuracy
- Faster Computation

# Additional Resources

---

**Python:** <https://www.w3schools.com/python/>

**Python Video:** [https://www.youtube.com/watch?v=\\_uQrJ0TkZlc](https://www.youtube.com/watch?v=_uQrJ0TkZlc)

**Libraries:** <https://www.geeksforgeeks.org/libraries-in-python/>

**Pandas:** <https://www.w3schools.com/python/pandas/default.asp>

**Join our Discord for more updates and resources!**

<https://www.bit.ly/AIS-Links>

# Kahoot Time!



The image features a solid black background. In the top right and bottom left corners, there are abstract, organic shapes in shades of bright pink and orange, resembling soft, glowing light or perhaps stylized flames. These shapes are partially cut off by the edges of the frame.

# Group Picture!