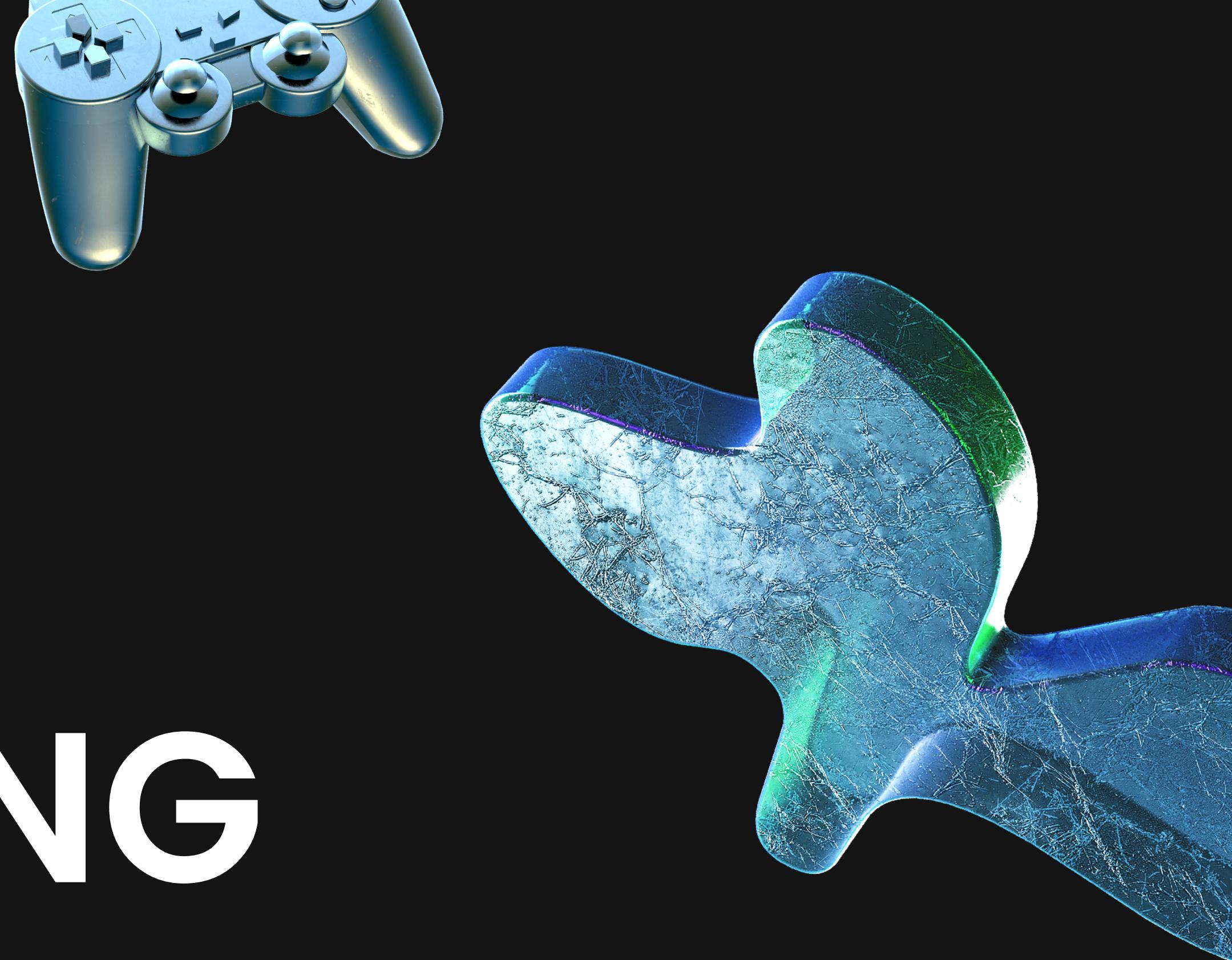


PYEXPO 2022

WEB SCRAPING



#Team No:2

SOWNDARYA S

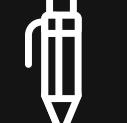
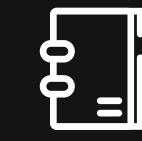
PRATHIBHA P

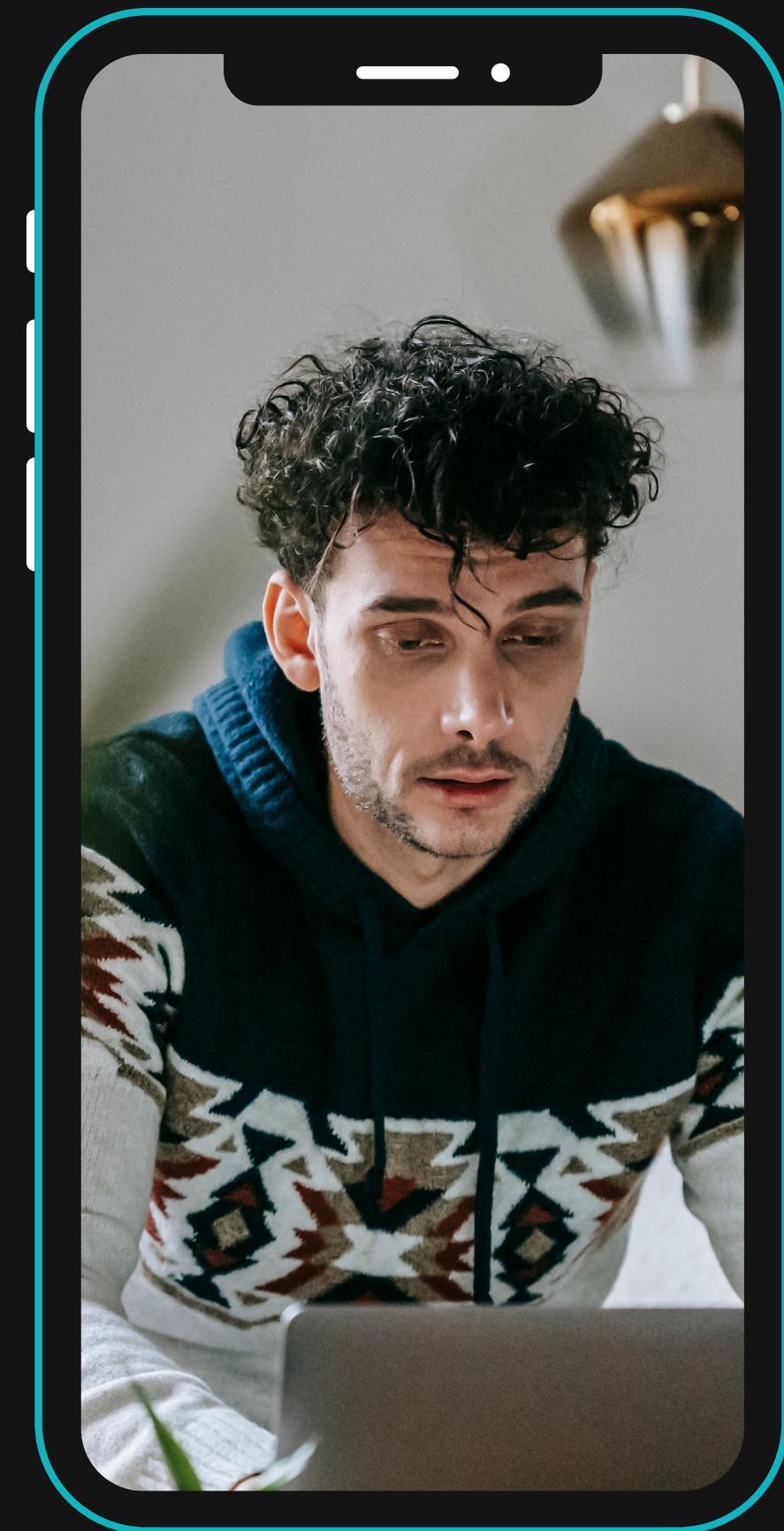
SANTHIYA S

BALAJI P R

SANJAI ROHITH M

SHARMILA S





DATA SCIENCE



Data Science is blended with various tools, algorithms, and machine learning principles. Most simply, it involves obtaining meaningful information or insights from structured or unstructured data through a process of analyzing, programming and business skills.

You need to start from data, its visualization, programming, formulation, development, and deployment of your model.

DATA SCIENCE

HOW DATA SCIENCE WORKS ?



PROBLEM STATEMENT

DATA COLLECTION

DATA CLEANING

DATA ANALYSIS AND
EXPLORATION

DATA MODELLING

OPTIMIZATION AND
DEPLOYMENT

WEB SCRAPING

WHAT IS WEBSRAPING ?

Web scraping is an automatic method to obtain large amounts of data from websites. Most of this data is unstructured data in an HTML format which is then converted into structured data in a spreadsheet or a database so that it can be used in various applications.

HOW IT WORKS ?

When a web scraper needs to scrape a site, first the URLs are provided. Then it loads all the HTML code for those sites. Then the scraper obtains the required data from this HTML code and outputs this data in the format specified by the user.

TYPES OF WEB SCRAPERS

e. Web Scrapers can be divided on the basis of many different criteria, including Self-built or Pre-built Web Scrapers, Browser extension or Software Web Scrapers, and Cloud or Local Web Scrapers.

TYPES OF WEB SCRAPERS

SELF-BUILT WEB SCRAPERS

Requires advanced knowledge of programming. Pre-built Web Scrapers are previously created scrapers that you can download and run easily. These also have more advanced options that you can customize.

BROWSER EXTENSIONS WEB SCRAPERS

Can be added to your browser. These are easy to run as they are integrated with your browser. But Software Web Scrapers don't have these limitations as they can be downloaded and installed on your computer. These are more complex than Browser web scrapers

CLOUD WEB SCRAPERS

Cloud Web Scrapers run on the cloud, which is an off-site server mostly provided by the company that you buy the scraper from. Local Web Scrapers, on the other hand, run on your computer using local resources.

WHAT ARE THE USES OF WEB SCRAPING ?

#1

Market Research

High-quality web scraped data obtained in large volumes can be very helpful for companies in analyzing consumer trends

#2

News Monitoring

Scraping news sites can provide detailed reports on the current news to a company. This is essential for companies that depend on daily news for their day-to-day functioning.

#3

Sentiment Analysis

To understand the general sentiment for their products among their consumers. Companies can use web scraping to collect data from social media websites.

#4

Price Monitoring

Scrap the product data for their products and competing products. Data to fix the optimal pricing for their products.

#5

Email Marketing

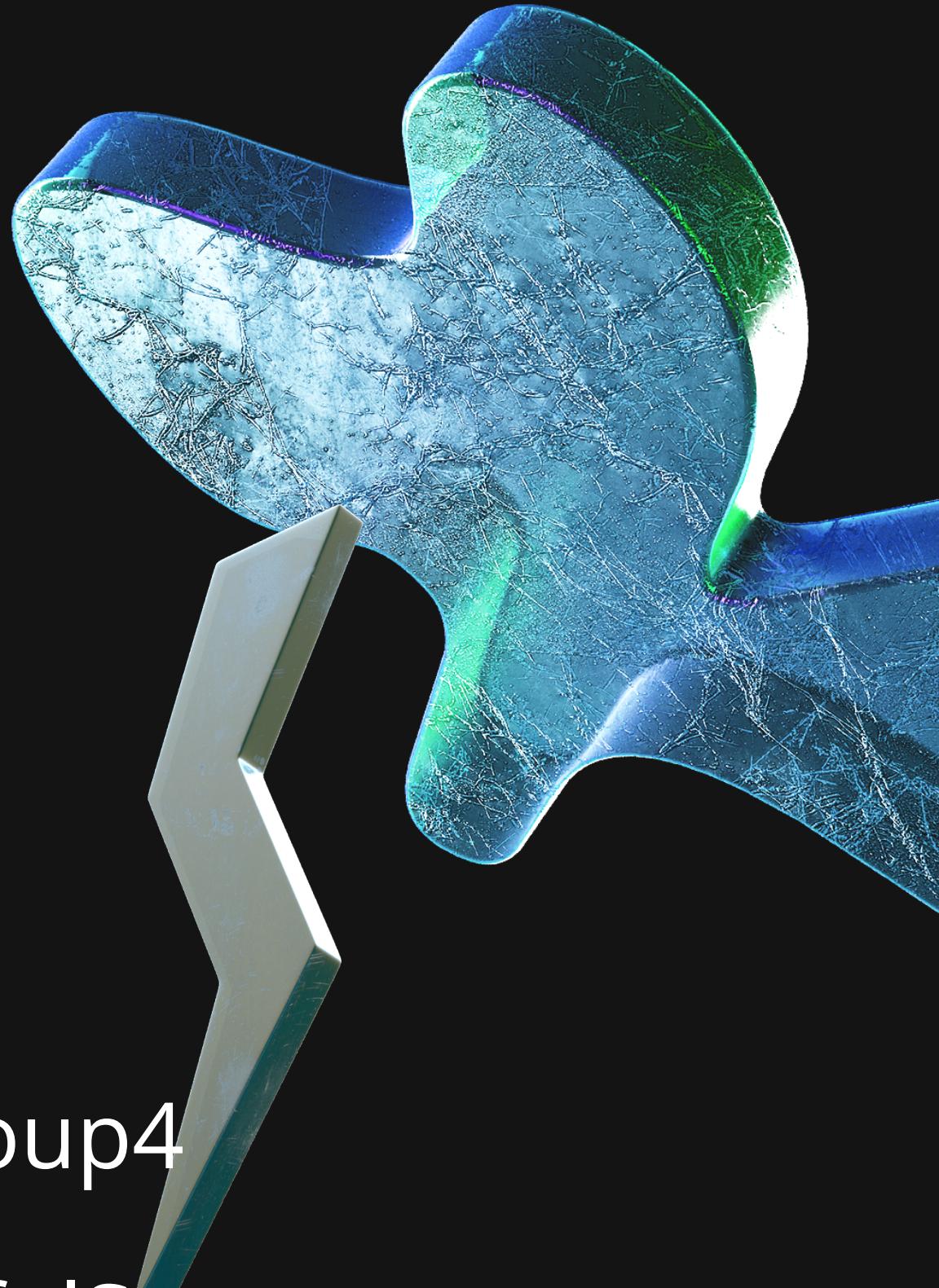
They can collect Email ID's from various sites using web scraping and then send bulk promotional and marketing Emails to all the people owning these Email ID's.

BeautifulSoup4

BeautifulSoup is a Python library for pulling data out of HTML and XML files. It works to provide idiomatic ways of navigating, searching, and modifying the parse tree. It commonly saves programmers hours or days of work.

Installing the packages : \$ pip install beautifulsoup4

Accessing the module : from bs4 import BeautifulSoup



TKINTER



Python offers multiple options for developing GUI (Graphical User Interface). Out of all the GUI methods, tkinter is the most commonly used method. It is a standard Python interface to the Tk GUI toolkit shipped with Python. Python with tkinter is the fastest and easiest way to create the GUI applications. Creating a GUI using tkinter is an easy task.



- To create a tkinter app:
 1. Importing the module – tkinter
 2. Create the main window (container)
 3. Add any number of widgets to the main window
 4. Apply the event Trigger on the widgets.

PROJECT INTERFACE



Enter a website:

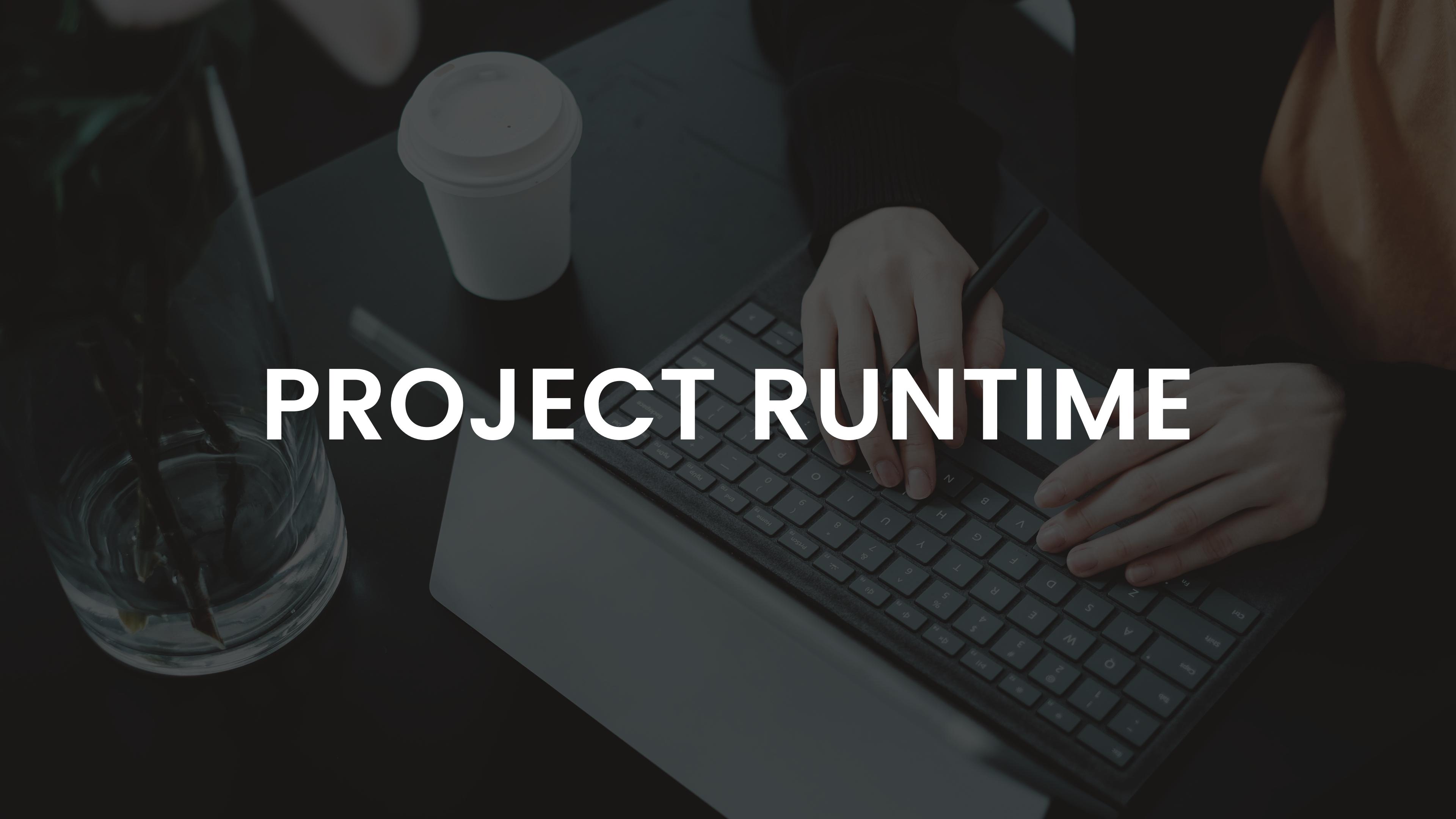
Submit

>

Message icon



PROJECT RUNTIME



Output Reference

```
tkinder
Enter a website: https://en.wikipedia.org/wiki/Data
Data - Wikipedia-->None
Jump to navigation-->#mw-head
Jump to search-->#searchInput
Data (computing)-->/wiki/Data_(computing)
Data (disambiguation)-->/wiki/Data_(disambiguation)
Datum (disambiguation)-->/wiki/Datum_(disambiguation)
Scientific Data (journal)-->/wiki/Scientific_Data_(journal)
-->/wiki/File:Data_types__en.svg
-->/wiki/File:Data_types__en.svg
US-->/wiki/American_English
/'dætə/-->/wiki/Help:IPA/English
UK-->/wiki/British_English
/'dɜːtə/-->/wiki/Help:IPA/English
facts-->/wiki/Facts
statistics-->/wiki/Statistics
information-->/wiki/Information
```

```
collected, reported-->/wiki/Data_reporting
analyzed-->/wiki/Data_analysis
visualizations-->/wiki/Data_visualization
concept-->/wiki/Concept
information-->/wiki/Information
knowledge-->/wiki/Knowledge
represented-->/wiki/Knowledge_representation_and_reasoning
coded-->/wiki/Code
processing-->/wiki/Data_processing
Raw data-->/wiki/Raw_data
numbers-->/wiki/Number
characters-->/wiki/Character_(computing)
outliers-->/wiki/Outlier
field data-->/wiki/Field_work
in situ-->/wiki/In_situ
Experimental data-->/wiki/Experimental_data
scientific-->/wiki/Science
```

```
-->/w/index.php?title=File:Philosophical_Transactions_-_Volume_001.djvu&page=60
-->/wiki/File:Philosophical_Transactions_-_Volume_001.djvu?page=60
Adrien Auzout-->/wiki/Adrien_Auzout
a 1665 article-->https://en.wikisource.org/wiki/Philosophical_Transactions/Volume_1/Number_4#56
Philosophical Transactions-->/wiki/Philosophical_Transactions
DIKW pyramid-->/wiki/DIKW_pyramid
information-->/wiki/Information
knowledge-->/wiki/Knowledge
wisdom-->/wiki/Wisdom
[8]-->#cite_note-8
Shannon entropy-->/wiki/Shannon_entropy
Knowledge-->/wiki/Knowledge
Mount Everest-->/wiki/Mount_Everest
altimeter-->/wiki/Altimeter
[9]-->#cite_note-9
[10]-->#cite_note-10
sign-->/wiki/Sign
[11]-->#cite_note-11
[12]-->#cite_note-12
```

```
[5]-->#cite_note-5
1 Etymology and terminology-->#Etymology_and_terminology
2 Meaning-->#Meaning
3 Data documents-->#Data_documents
3.1 Data collection-->#Data_collection
4 In other fields-->#In_other_fields
5 See also-->#See_also
6 References-->#References
7 External links-->#External_links
edit-->/w/index.php?title=Data&action=edit&section=1
Data (word)-->/wiki/Data_(word)
[6]-->#cite_note-EOL-6
[6]-->#cite_note-EOL-6
```

CONCLUSION

ID LIKE TO CONCLUDE THAT WEB SCRAPERS CAN BE A SIMPLE CONCEPT BUT IT CAN BE FURTHER ADVANCED TO USED IN MULTIPLE PROJECTS FOR AN EXAMPLE IN CASE OF ROBOTS IT CAN SCRAP PARTICULAR DATA FOR USE FROM A WEBSITE....

Do you have
any questions?



THANK YOU

