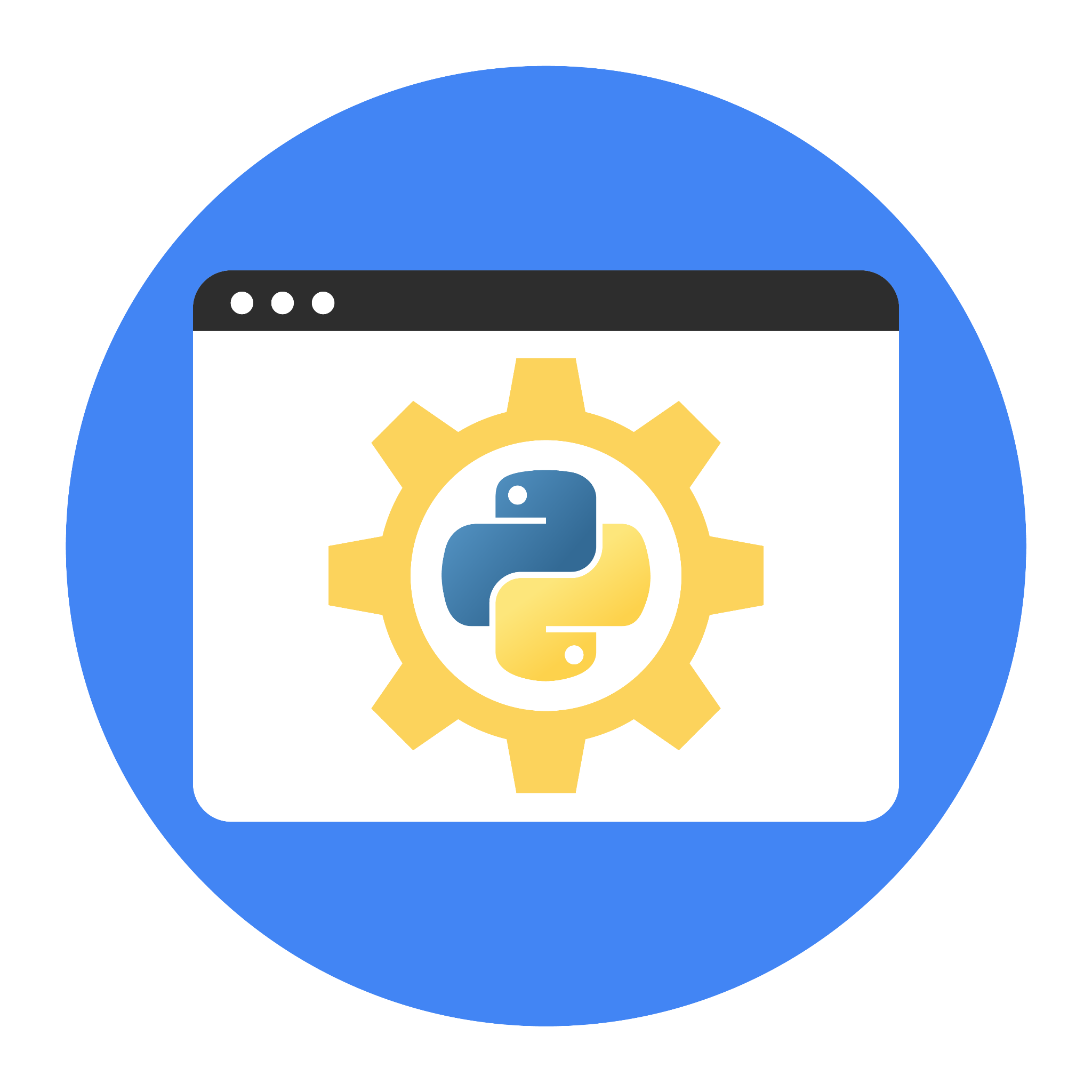
**La Course Two**

# Get Started with Python



# Instructions

Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. You can use this document as a guide to consider your responses and reflections at different stages of the data analytical process. Additionally, the PACE strategy documents can be used as a resource when working on future projects.

# Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

* Complete the questions in the Course 2 PACE strategy document
* Answer the questions in the Jupyter notebook project file
* Complete coding prep work on project’s Jupyter notebook
* Summarize the column Dtypes
* Communicate important findings in the form of an executive summary

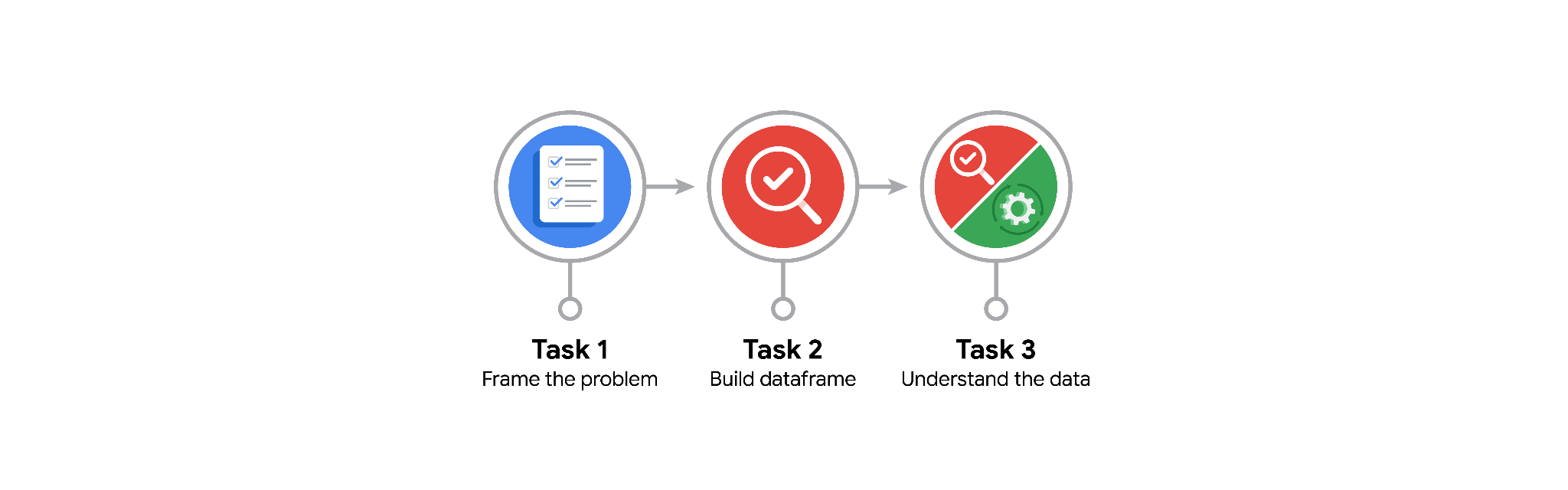
# Relevant Interview Questions

Completing the end-of-course project will help you respond these types of questions that are often asked during the interview process:

* Describe the steps you would take to clean and transform an unstructured data set.
* What specific things might you look for as part of your cleaning process?
* What are some of the outliers, anomalies, or unusual things you might look for in the data cleaning process that might impact analyses or ability to create insights?

**Reference Guide**

This project has three tasks; the visual below identifies how the stages of PACE are incorporated across those tasks.



**Data Project Questions & Considerations**

**PACE: Plan Stage**

* How can you best prepare to understand and organize the provided information?

Identify the overall goal of the project and clarify the specific objectives of my assigned task.

* What follow-along and self-review codebooks will help you perform this work?

Access the dataset, explore the column structure, and review any existing documentation or examples to understand the general scope of the data.

* What are some additional activities a resourceful learner would perform before starting to code?
* Carefully read and understand the project requirements
* Clarify any doubts by reviewing previous readings or asking questions
* Set up the Jupyter Notebook environment
* Load and inspect the CSV file (tiktok\_dataset.csv)
* Review column definitions and identify any potential data issues

**PACE: Analyze Stage**

* Will the available information be sufficient to achieve the goal based on your intuition and the analysis of the variables?

In the existing dataframe there are enough columns to identify work with it and achieve the goal.

* How would you build summary dataframe statistics and assess the min and max range of the data?

Using the describe method.

data.describe()

* Do the averages of any of the data variables look unusual? Can you describe the interval data?

The large gap between the mean and maximum values suggests the presence of outliers, especially in video\_like\_count, video\_share\_count, and video\_comment\_count

**PACE: Construct Stage**

**Note**: The Construct stage does not apply to this workflow. The PACE framework can be adapted to fit the specific requirements of any project.

**PACE: Execute Stage**

* Given your current knowledge of the data, what would you initially recommend to your manager to investigate further prior to performing exploratory data analysis?

I recommend looking deeper into the claim videos from banned and under-review authors. These videos have very high views, likes, and shares, even though there are fewer of them. It would be useful to check if these videos follow the platform’s rules or if they include harmful or false content.

* What data initially presents as containing anomalies?

The view, like, and share counts for banned and under-review users look unusual. Their median values are very high compared to active users. Also, some values in the dataset are missing, like claim status or transcription, which may affect the analysis.

* What additional types of data could strengthen this dataset?

The date the video was posted (to analyze trends over time),

The number of followers of each user (to understand their influence),

A content category or topic label for the video (to know what it’s about)