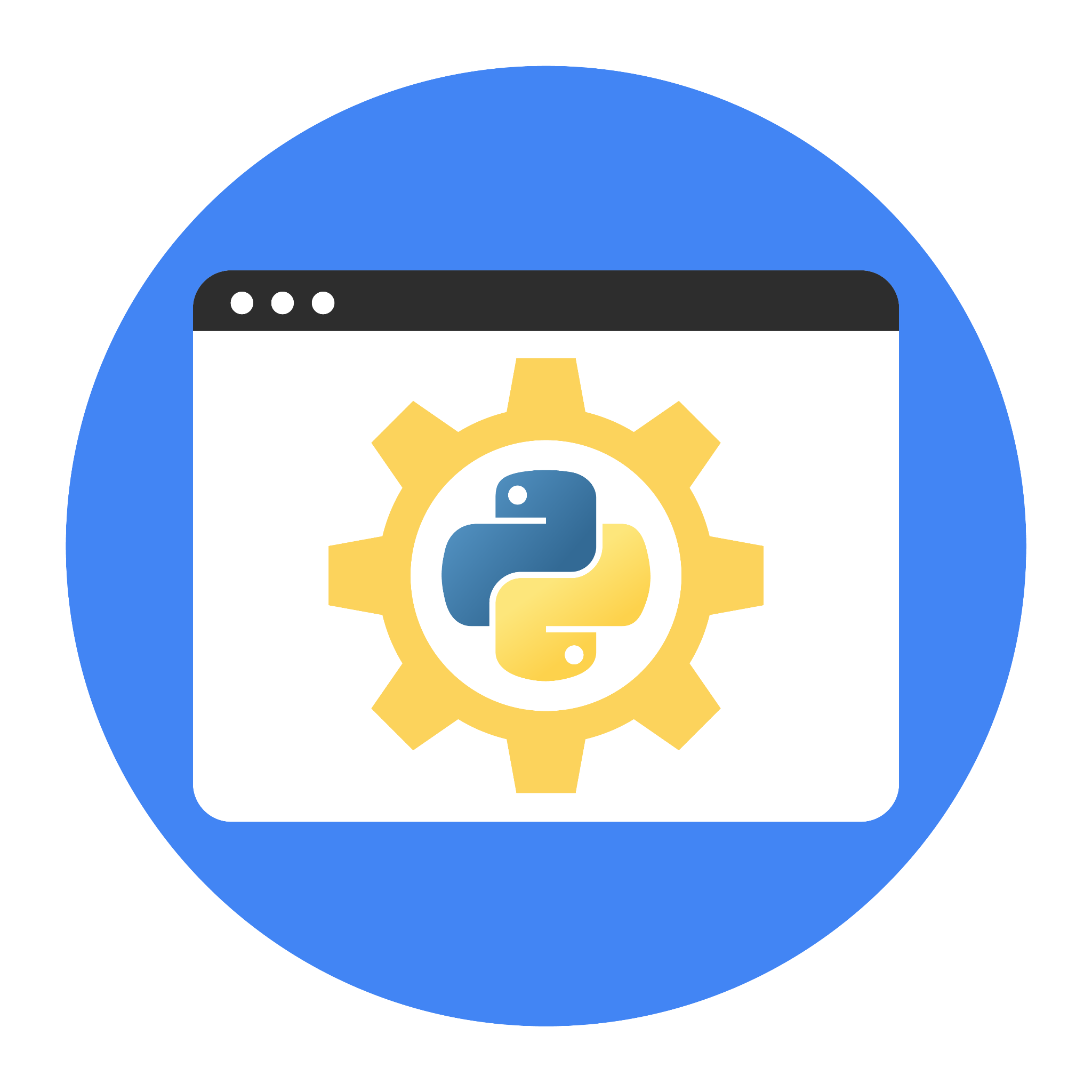
**Course Two**

# 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

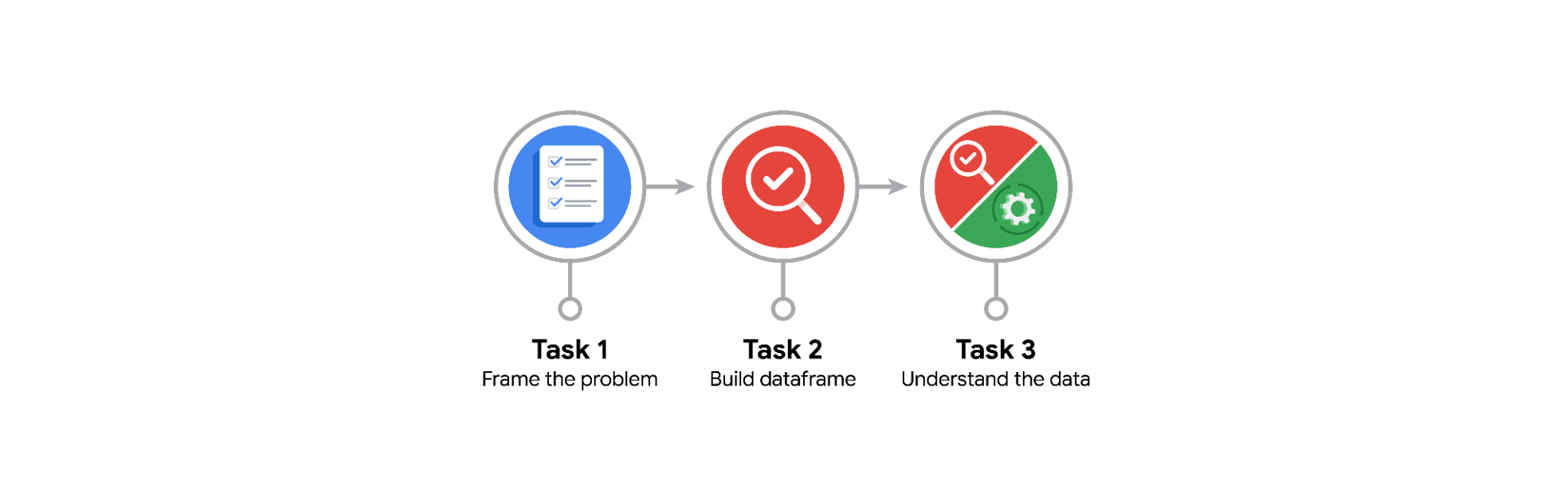
# Relevant Interview Questions

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?

Understand what is the goal of the project and gather information about the data that is provided.

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

The follow-along notebooks that explain the most common data structures in python, as well the notebooks that explain what functions, conditional statements and loops are, will likely be useful for this work.

* What are some additional activities a resourceful learner would perform before starting to code?

Read all the information provided for the project at hand before starting to write code.

**PACE: Analyze Stage**

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

It would be useful to understand how negative value can arise for some of the amounts. But if that information cannot be obtained from TLC, the records with negative values can be filtered out.

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

Use the describe method.

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

Apart from the passenger count, the average of the other variables of type int are meaningless.

Some of the variables that contain monetary values have negative values. These are abnormal and might have to be excluded from the dataset to be used for the next step, unless they are actually possible in certain circumstances.

The maximum total\_amount has a a fare\_amount of 999.99, which seems like it is a dummy value. That record should be excluded from the dataframe before further exploration.

**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?

Investigate whether the negative values are justified or if they should have their sign switched or be excluded.

* What data initially presents as containing anomalies?

The maximum fare\_amount value of 999.99.

The negative values for some of the fare-related variables.

The trip distances that equal 0.

* What additional types of data could strengthen this dataset?

Some data about the drivers and passengers could provide additional insights. For example, number of years of experience of the drivers, country origin of the main passenger, purpose of the trip (work or leisure).