# 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

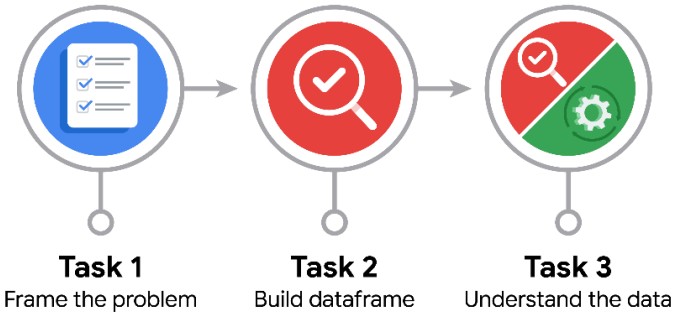
## Relevant Interview Questions

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?

I would reread the previous documents done in Course 1 to refresh the goal of the project. Then, I would review the data provided and see if any cleaning step is required.

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

I would use the last part, where we use Panda, NumPy, boolean mask and groupby to perform the tasks for this project.

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

It's better to review what you learn in Python, take notes and look at any doubt you have on the internet, because there are forums where people have the same doubt or it helps you to have an idea.

## PACE: Analyze Stage

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

| |
|---|
| Yes, because I ran the head, describe and info codes to have an overview of the data provided. Then, I checked what data type the file has and if there was any non-null value or inconsistency may have in the format. So we can |

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

| |
|---|
| I would use the following codes: head, describe and info.<br><br>For the min and max, I would use the 'group by' code. |

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

| |
|---|
| There are negative values in the fare amount column, unusual since it's money. |

## 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?

| |
|---|
| The process on how they obtained the data, so when you observe the anomalies, you can have an idea why some data doesn't fit or is unusual. |

- What data initially presents as containing anomalies?

| |
|---|
| The fare amount presents negative values and the maximum value is unusual because it's above $1000. Also the longest trip doesn't correspond to the highest total amount. |

- What additional types of data could strengthen this dataset?

| |
|---|
| |