# 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

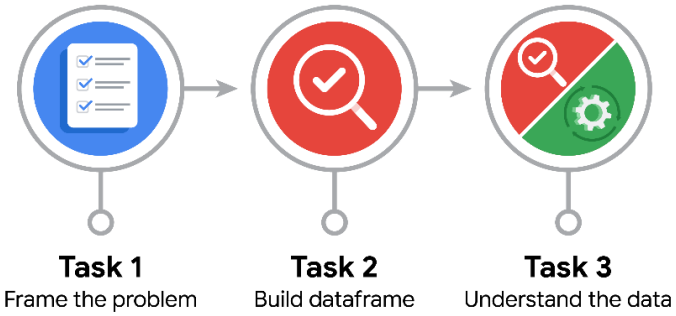
## Relevant Interview Questions

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?

Firstly, I would upload the data and see the data fully without doing anything to it and try to identify the structure of the data and what things can be done to facilitate it. Then I would try to get a piece of summary information like the null values and duplicates and clear them. I can also organize the data in a way to make it easier such as adjusting the formatting of the data and also changing the table names.

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

I will use the Jupyter Notebook where I can make the analysis and brief reports for the steps taken in the exploring

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

I can see the source of the data and how it was gathered to inspect any form of bias.

## 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 it would give us the ability to organize the data depending on the metadata and the source it was taken from. |

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

| |
|---|
| ○ Upload the data to Python<br>○ Import pandas<br>○ Use the summary statistics function on the data |

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

| |
|---|
| The # is a duplicate of the index where there are two indices in the data frame on in the original and the other on is a column called “#”.the data is heavily skewed to the right. |

## 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 would tell them to investigate the reason why the banned videos generally have more engagement.

- What data initially presents as containing anomalies?

The “#” column is redundant.

- What additional types of data could strengthen this dataset?

The reason behind the report would be beneficial.