**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?

Read and understand the data, view the data dictionary, and explore the dataset to identify key variables for the stakeholder.

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

Revise module 2 of Google Advanced Data Analytics course

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

Brush up on key Python commands for aggregating and summarising data. Have a cheat sheet beside.

**PACE: Analyze Stage**

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

Yes, it will be sufficient.

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

Use aggregate function to obtain mean, median, count and groupby() to group two or more columns.

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

It has a very large mean and maximum values compared to the quartiles.

**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?

Engagement level is strongly correlated with claim status. This should be a focus of further inquiry.

* What data initially presents as containing anomalies?

Videos with banned authors have significantly higher engagement than videos with active authors. Videos with authors under review fall between these two categories in terms of engagement levels.

* What additional types of data could strengthen this dataset?