**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 the data dictionary and look at the data head.

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

The materials in this course, as well as Pandas documentation.

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

Having a window open for consulting documentation, and making sure all of the files are organized.

**PACE: Analyze Stage**

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

Yes, the project could be completed with this data. However, it may represent a subset of drivers who are long haul truckers, in which case we may not get results representing shorter-distance drivers.

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

DataFrame.describe

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

There was nothing unusual in this data except for that 700 data labels were missing.

**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?

Yes. We could benefit from getting data that represents shorter-distance drivers.

* What data initially presents as containing anomalies?

There are 700 missing labels. There was no pattern indicating why these may be missing. Still, these are 700 rows that we can’t use for classification.

* What additional types of data could strengthen this dataset?

Data representative of shorter-distance drivers.