**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?

Best way to do so is to first read any column names and check data formats. If it is still unknown what the data is it is recommended to contact a person or an organization that provided such data. After that, next step should be to filter out any unnecessary information and start checking through relevant columns one by one.

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

Jupyter notebook is a main tool used for this project.

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

Find out if the data source is reliable and provided data is uncorrupted.

**PACE: Analyze Stage**

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

The amount of provided data should be enough to draw some conclusions about it.

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

I would check if outliers have values that make sense. If some outlier does not make much sense it should be investigated and corrected or removed from the research

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

There is nothing out of the ordinary when looking at the data averages, however some data values look suspicious, like some negative trip distances or very large trip distances.

**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 recommend investigating possible trends, like analyzing connections between revenue and dates (day of week, months).

* What data initially presents as containing anomalies?

Outlier values, especially the very large ones. Hopefully they are just some data corruption errors, not the results of some malicious intent or fraud.