**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: Planning Stage**

* How can you best prepare to understand and organize the provided information?

Begin by exploring your dataset and consider reviewing the Data Dictionary. I can prepare to understand the information by reading the taxi cab data fields and ensuring I understand how each one impacts the dataset. Reviewing the fact sheet could also provide helpful background information. However, my primary goal is to get the data into Python, inspect it, and provide DeShawn with my initial observations. Afterwards, I can learn more deeply about the data and check for any anomalies.

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

By exploring your dataset and consider reviewing the Data Dictionary. I can prepare to understand the information by reading the taxi cab data fields and ensuring I understand how each one impacts the dataset. Reviewing the fact sheet could also provide helpful background information.

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

Go through all the documents provided. Try to frame the problem.

******PACE: Analyzing Stage**

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

Yes.

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

For that, I use exploratory data analysis technique to get a summarize their main characteristics of the dataset.

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

No averages look fine but the maximum fare amount is a much larger value ($1000) than the 25-75 percent range of values.

******PACE: Constructing 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. Through EDA we get the summary of the characteristics about data. But for a detailed analysis we might need to use another data analysis techniques.

* What data initially presents as containing anomalies?

There are anomalies in total amount as it contains very large values for a small trip distance.

Negative values for fare amount.

* What additional types of data could strengthen this dataset?

Time difference column between pickup and dropoff.