**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?

I have import the dataset from the csv file using as a programming language in the notebook workspace, then I can do analysis using python for search anomalies and insights for analysis.

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

The codebooks to perform the analysis are the importance or dataframe and any libraries needed, and use some function such as head(), scribe(), info(). Then the use of Numpy and Pandas will help me very well to analyze data and determine anomalies.

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

Practice the syntax of any function and how’s working with dataframe, That will help me very much in organizing and practicing the code before application.

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

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

The available information provided by the summary of the dataframe is very helpful for understanding the data in general, but that’s not all story, because the relationship of variables is not clear yet, then the data need more cleaning and organizing.

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

For building the summary of dataframe, I used the describe() function from pandas library, that shows many information about the data such as mean, max, min ranges …

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

In General , it looks like all values of variables are correct, however, there’s some little cases for exemple, the trip distance is less than 0 in some cases, maybe the problem comes from the collection of data phase.

******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?

I initially recommend to my manager to investigate the missing values and incorrect values in the data and determine the causes and the how to solve this problem before starting the analyze phase.

* What data initially presents as containing anomalies?

First anomalies exist in the data is the incorrect values in the trip distance variable, and some variable took zero as value.

* What additional types of data could strengthen this dataset?

Margin error, mean, place(specific in the city).