**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?

I need to get the overview of the data, install jupyter notebook, and be ready to explore the data

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

The follow along codebooks will include videos from the data class, online blocks, and some of my previous codes

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

You need to get your data organized, cleaned, and ready for action

**PACE: Analyze Stage**

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

The information given will be sufficient to solve the problem. Information on the number of taxis, how people patronize, price, and distance should be enough to make informed decisions.

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

I will use pandas in python to build a data frame from a dictionary or spreadsheet containing the data. Syntax for min, max, range, etc will be coded on the panda’s data frame.

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

The averages of the data are usual. The trip distance ranges from 0 to 33 miles whilst the time ranges the cost ranges from 111 dollars to 1200 dollars. The amount seems a bit higher for just 33 miles.

**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 recommend my manager to investigate the actual prices of transport since the distance to transport fares showing in the data seems a bit fishy and unrealistic.

* What data initially presents as containing anomalies?

The transport fares seem a bit higher for a maximum of 33 miles distance

* What additional types of data could strengthen this dataset?

Data on car types, number of people on the ride, chances of traffic jam, and other similar apps fares could be useful