**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?

By understanding the project goals of the project, looking at the data manually like the fields values etc.

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

Fact sheets.

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

Read about python libraries and functions that can be useful for this task.

**PACE: Analyze Stage**

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

Fortunately, the dataset is already automatically loaded for us. So far looks like yes it is sufficient.

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

For the first question I can use df.describe() for the second I can use sort\_values function in descending order.

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

Yes, the unusual values are that the first two are much higher than the rest of the values.

**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 him to perform EDA, data clearning a discovery of anomalies, use descriptive statistics and lastly create and run the model and see how it performs, it does not matter if its not as expected in the first run as the PACE model is an iterative model.

* What data initially presents as containing anomalies?

There are a few short distance trips but with high charges, also there are some with negative values in fair\_amount.

* What additional types of data could strengthen this dataset?