**Course One**

# Foundations of Data Science



# Instructions

Use this PACE strategy document to record your decisions and reflections as you work through this end-of-course project. As a reminder, this document is a resource that you can reference in the future and a guide to help consider responses and reflections posed at various points throughout projects.

# Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

* Understand and assess the proposed scenario
* Demonstrate foundational knowledge of the data science workflow - PACE
* Articulate a data project proposal in the planning stage for cross-functional team members

# Relevant Interview Questions

Completing this end-of-course project will empower you to respond to the following interview topics:

* As a new member of a data analytics team, what steps could you take to get 'up to speed' with a current project? What steps would you take? Who would you like to meet with?
* How would you plan an analytics project?
* What steps would you take to translate a business question to an analytical solution?
* Why is actively managing data an important part of a data analytics team's responsibilities?
* What are some considerations you might need to be mindful of when reporting results?

**Reference Guide**

This project has three tasks; the following visual identifies how the stages of pace are incorporated across those tasks.



**Data Project Questions & Considerations**

**PACE: Planning Stage**

* Who is your audience for this project?

My audience is the New York City Taxi and Limousine Commission.

* What are you trying to solve or accomplish? And, what do you anticipate the impact of this work will be on the larger needs of the client?

I am trying to figure out what is the estimated taxi fares based in relevant potentially correlated variables that are to be identified.

* What questions need to be asked or answered?

What is the state of the dataset provided?

What variables have more influence in the taxi fares?

What trends can I detect from key insights?

Did I do something to account for bias if there is any?

* What resources are required to complete this project?

- The dataset.

- Jupyter notebooks or similar.

- Stakeholders input and feedback.

* What are the deliverables that will need to be created over the course of this project?

## **THE PACE WORKFLOW**



**[Alt-text: The PACE Workflow with the four stages in a circle: plan, analyze, construct, and execute.]**

You have been asked to demonstrate for the company's data team how you would use the PACE workflow to organize and classify tasks for the upcoming project. Select a PACE stage from the dropdown buttons. A few tasks involve more than one stage of the PACE workflow. Additionally, not every workplace scenario will require every task. Refer back to the [Course 1 end-of-course portfolio project overview: Automatidata](https://www.coursera.org/learn/foundations-of-data-science/supplement/XxgHa/course-1-end-of-course-portfolio-project-overview-automatidata) if you need more information about the tasks within the project.

### **Project tasks**

Following are a group of tasks your company’s data team has determined need to be completed within this project. The data analysis manager has asked you to organize these tasks in preparation for the project proposal document. First, identify which stage of the PACE workflow each task would best fit under using the drop down menu. Next, give an explanation of why you selected the stage for each task. Review the following readings to help guide your selections and explanation: [The PACE stages](https://www.coursera.org/learn/foundations-of-data-science/supplement/4OtHr/the-pace-stages) and [Communicate objectives with a project proposal](https://www.coursera.org/learn/foundations-of-data-science/supplement/79Ysh/communicate-objectives-with-a-project-proposal). You will later reorder these tasks within a project proposal.

1. **Evaluating the model: Execute**

Why did you select this stage for this task?

Once the data model is done, the model is put to test with data to see if it meets the project goals.

1. **Conduct a hypothesis test: Analyze** **and** **Construct**

Why did you select these stages for this task?

In the Analyze phase, the statistical test to be used is chosen. However, the test is perform during the construction phase.

1. **Understanding the data: Analyze**

Why did you select this stage for this task?

Following the PACE model the Analysis phase is the next logical stage in which one get to know more in deep the dataset and what information can provide us.

1. **Data exploration and cleaning: Plan** **and Analyze**

Why did you select these stages for this task?

In the planning phase the methods needed are determined. Furthermore, in the analyse stage the cleaning process is performed.

1. **Establish structure for project workflow (PACE): Plan**

Why did you select this stage for this task?

This is the initial phase in which a document is created to show the workflow and from angles aboard the problem.

1. **Communicate final insights with stakeholders: Execute**

Stakeholders always like to be communicated of the latest developments of the project and is recommended. All the key insights from the results are shared with them according to their profile.

1. **Compute descriptive statistics: Analyze**

Why did you select this stage for this task?

Because looking at the statistics in details of the data takes place in this stage,

1. **Visualization building: Analyze and Construct**

Why did you select these stages for this task?

Because visualization starts and is build with the data analysis and construction stage respectively.

1. **Write a project proposal: Plan**

Why did you select this stage for this task?

Because as described by the PACE model is the first stage where the first document to define the project is created.

1. **Build a regression model: Analyze and Construct**

Why did you select this stage for this task?

A carefully examination of various models is one to see which one is more suitable for the task projects goals. Then in the construction stage the regression model is built.

1. **Inspect the data set for missing data: Analyze**

Why did you select this stage for this task?

Because in here is where the dataset is checked in detail to check for any anomalies.

1. **Build machine learning model: Construct**

Because in here is where the construction of the data model begins.