

# Course One

## Foundations of Data Science



### 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 PACE Strategy Document to plan your project while considering your audience members, teammates, key milestones, and overall project goal.
- ☐ Create a project proposal for the data team.

### Relevant Interview Questions

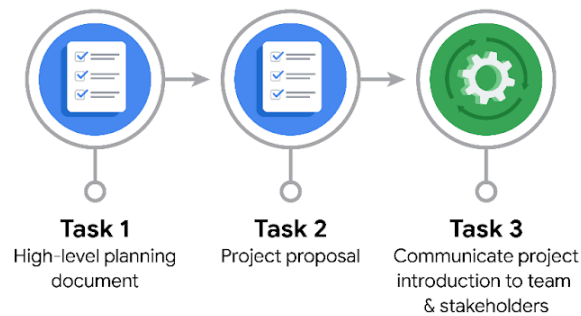
Completing this end-of-course project will empower you to respond to the following interview topics:

- 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: Plan Stage

- Who is your audience for this project?

**If you selected the Automatidata scenario,** your audience will be the New York City Taxi and Limousine Commission.

**If you selected the TikTok scenario,** your audience will be the TikTok data team and cross-functional team members.

**If you selected the Waze scenario,** your audience will be the Waze data team and cross-functional team members.

- 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?

**If you selected the Automatidata scenario,** the problem you are trying to solve is the estimation of taxi fares based on relevant variables that you identify.

**If you selected the TikTok scenario,** you are working on the development of a predictive model that can be used to determine whether a video contains a claim or whether it offers an opinion.

**If you selected the Waze scenario**, working to develop a machine learning model to predict user churn.

- What questions need to be asked or answered?

**For all projects**, consider the following questions: What is the condition of the provided dataset? What variables will be the most useful? Are there trends within the data that can provide insight? What steps can I take to reduce the impact of bias?

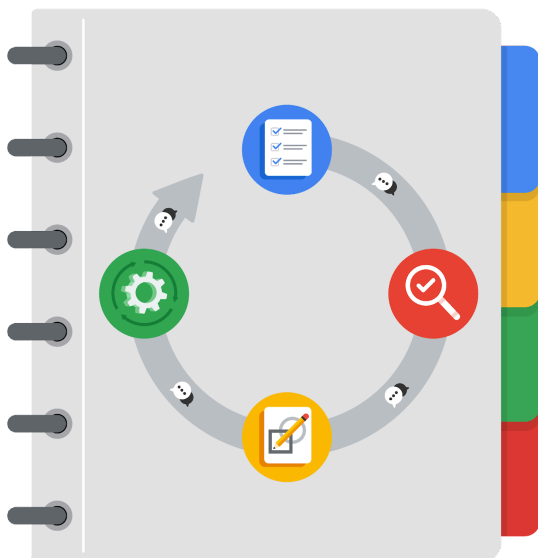
- What resources are required to complete this project?

**For all projects**, you will need the project dataset, Python notebook, and input from stakeholders.

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

**For all projects**, the deliverables include a dataset scrubbed for exploratory data analysis, visualizations, statistical model, regression analysis and/or machine learning model.

## 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 reading 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 and 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?

After the model has been constructed, data is run through to evaluate whether it meets the project's expectations and goals.

**2. Conduct hypothesis testing:** **Analyze** ▾ and **Construct** ▾

Why did you select these stages for this task?

During the analyzing stage, it is determined that a statistical test will be used. During the construction phase, the test is carried out.

**3. Begin exploring the data:** **Analyze** ▾

Why did you select this stage for this task?

During the analysis phase, you will gain a deeper understanding of the dataset and the information within it.

**4. Data exploration and cleaning:** **Plan** ▾ and **Analyze** ▾

Why did you select these stages for this task?

Planning takes place when you first make choices about the methods needed. The cleaning process then takes place in the analyzing stage.



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

Why did you select this stage for this task?

Planning stage. Creating an initial project PACE document outlines the workflow and helps to plan how to best approach a project.

6. Communicate final insights with stakeholders: **Execute** ▾

Why did you select this stage for this task?

Communication is necessary at various points throughout a project. Final insights are shared with stakeholders in the execute phase of the data project workflow.

7. Compute descriptive statistics: **Analyze** ▾

Why did you select this stage for this task?

Investigating the statistics within data takes place during analysis.

8. Visualization building: **Analyze** ▾ and **Construct** ▾

Why did you select these stages for this task?

Visualization begins with data assessment and is created during the construction stage.

9. Write a project proposal: **Plan** ▾

Why did you select this stage for this task?

Planning stage. A project proposal is the initial document used to define a project.

10. Build a regression model: **Analyze** ▾ and **Construct** ▾

Why did you select this stage for this task?

During the analyzing stage, the model is examined in detail to be sure it will meet the needs of the task. The building of the regression model will take place in the construction phase.



**11. Compile summary information about the data:** Analyze ▾

Why did you select this stage for this task?

Inspecting a dataset to compile information would take place in the analysis phase.

**12. Build machine learning model:** Construct ▾

Why did you select this stage for this task?

The building of a data model would take place in the construct stage.