

## 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 thoughts at different stages of the data analytical process. The PACE strategy documents can also 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.

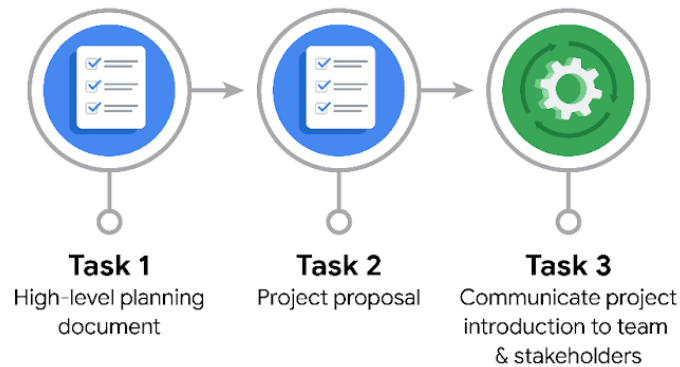
#### Relevant Interview Questions

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 into an analytical solution?
- Why is actively managing data an essential part of a data analytics team's responsibilities?
- What considerations might you 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?

The primary audience for this project proposal 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 more significant needs of the client?

The issue that's being addressed is estimating taxi fares prior to a trip by using relevant predictor variables identified from the data.

- What questions need to be asked or answered?

Timeline:

- What is the overall timeline for the project?
- What are the key milestones and deliverables for each project phase?
- What are the potential risks and challenges to the project timeline?

Data sourcing:

- What are the data sources that will be used for the project?
- How will the data be collected and processed?
- What are the potential limitations of the data sources?

**Data sourcing:**

- What are the data sources that will be used for the project?
- How will the data be collected and processed?
- What are the potential limitations of the data sources?
- How can bias be mitigated?

**Key questions:**

- What are the critical questions that the analysis will seek to answer?
- Why are these questions important?
- How will the answers to these questions help the business?

**Analytical methods:**

- What analytical methods will be used to answer the key questions?
- Why are these methods the best fit for the data?
- How will the results of the analysis be validated?
- What trends exist within the data?

**Expected results:**

- What are the expected results of the analysis?
- How will the results be measured?
- How will the results be used to improve the business?

**Leverage of results:**

- How will the analysis results be communicated to the business stakeholders?
- How will the results be used to make informed decisions?
- How will the results be used to improve the business?

➤ What resources are required to complete this project?

**Data:** The project will need access to data that is relevant to the goals of the project.

- TLS will provide several years' worth of third-party taxi and rideshare trip data.

**People:** The project will need a team with the skills and experience to collect, analyze, and interpret the data.

**Tools:** The project will need access to software tools that can be used to collect, analyze, and visualize the data.

- Python, Tableau, Excel, SQL, and potentially R.

**Time:** The project will need a certain amount of time to complete, which can be broken down by task for more streamlined and efficient execution.

**Budget:** The project will have a budget covering the costs of data collection, personnel, tools, and other expenses.



- What deliverables will need to be created throughout this project?

**Data dictionary:** A data dictionary is a valuable document describing the data being utilized in a project. It includes details such as the data field's type, source, and significance. In a project, a data dictionary is a valuable document that describes the data being utilized.

**Data analysis plan:** A data analysis plan is a document that describes how the data will be analyzed in the project. It includes information such as the analytical methods that will be used, the data that will be analyzed, and the expected results of the analysis.

**Data analysis report:** A data analysis report is a document that presents the results of the data analysis. It includes information such as the findings of the analysis, the conclusions drawn, and the recommendations made.

**Data visualizations:** Data visualization refers to creating graphical representations of data, which can significantly enhance the communication of data analysis results.

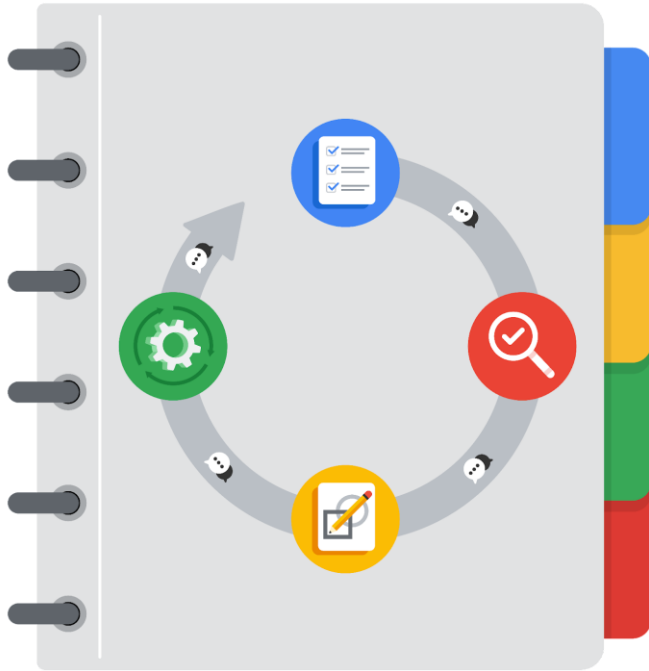
**Models:** Statistical/regression analysis and models.

**Presentations:** Presentations may be given to stakeholders to communicate the data analysis results and get feedback.

**White papers:** White papers are documents that provide an in-depth analysis of a particular topic. They can be used to educate stakeholders about the data analysis results and help them make informed decisions.



## 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 to 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, only some workplace scenarios will require every job. Refer 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 is 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, explain why you selected the location 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 model construction, the data is evaluated to ensure it meets the expectations and goals of the project.



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

Why did you select these stages for this task?

In the construct phase, the machine-learning and regression models are created. In the analysis phase, these models are used, and hypothesis testing occurs.

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

Why did you select this stage for this task?

Exploratory data analysis (EDA) is a critical step in data analysis. It occurs during the analysis phase when the data analyst begins to investigate the data in detail. EDA includes a variety of tasks, such as handling missing values, checking for biases and outliers, reformatting data, and imputation.

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

Why did you select these stages for this task?

Initially, the planning stage involves determining the necessary methods and strategies. Subsequently, during the analysis phase, the cleaning process is initiated.

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

Why did you select this stage for this task?

The project structure is created during the planning phase, when the project workflow is outlined in a PACE document, including all the necessary stages, resources, and dependencies.

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

Why did you select this stage for this task?

In the Execute phase, stakeholders are informed of the project results and any newfound insights.

**7. Compute descriptive statistics: Analyze**

Why did you select this stage for this task?

During data analysis, the examination of statistics is conducted.



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

Why did you select these stages for this task?

The first steps to visualization occur during the analysis phase and are subsequently created during the construct stage.

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

Why did you select this stage for this task?

A project proposal is drafted during the Plan phase, and all significant steps and tasks are outlined in detail.

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

Why did you select this stage for this task?

In the Analyze phase, the data is thoroughly examined, cleaned, and formatted to lay the groundwork for developing a regression model. The complete models are constructed and put into action in the Construct phase.

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

Why did you select this stage for this task?

After cleaning, reformatting and inspecting the data, data compilation occurs during the analysis phase.

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

Why did you select this stage for this task?

The development of a machine learning model takes place in the Construct phase, where it is subjected to multiple tests to ensure that it meets the project's requirements and expectations.