

# 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.

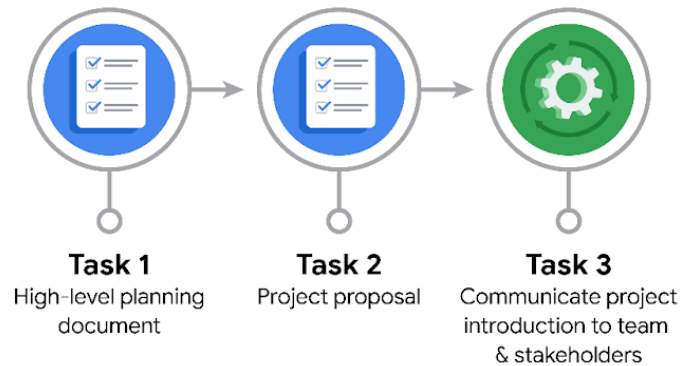
### Relevant Interview Questions

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?

My audience include team members from Data professionals from the Data team and Crossfunctional Team members. The data team includes Willow-Data science lead, Rosie-data science manager, Orion-Data scientist.

The Cross-functional team members inca Mary - Project management officer, Margery- Finance Lead and Maika - Operations Lead.

- 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?

A predictive model that can determine whether a video contains a claim or offers an opinion.

I expect that this work will improve the effectiveness and efficiency of addressing claims thereby reducing the backlog of user reports.



- What questions need to be asked or answered?

I would consider the following questions:

- What variables need to be considered
- What is the condition of the dataset? It is important to learn about when and how the data was collected
- What steps can I take to reduce the impact of bias?
- What trends can be observed in the data.
- What variables are likely to influence or yield better results.
- What is the likely economic impact of this project.

- What resources are required to complete this project?

I would the following resources to complete this project:

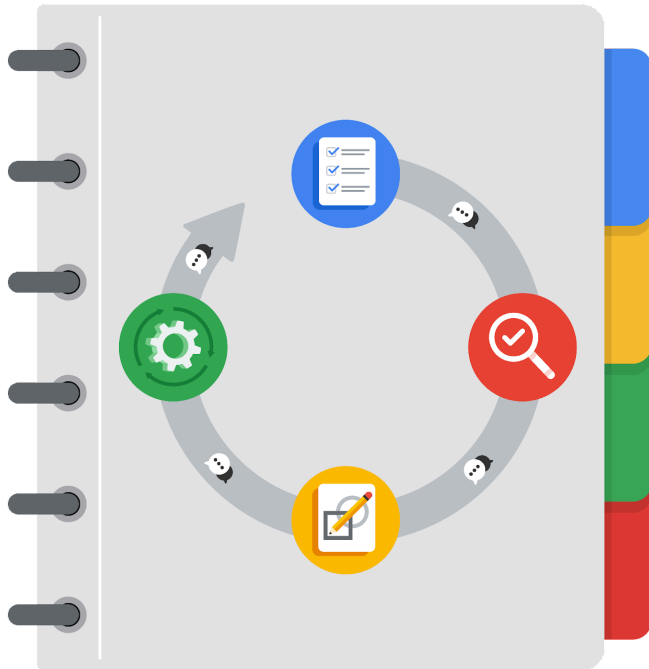
- A python notebook and hardware (computer)
- Dataset
- Input from stakeholders

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

- Project proposal that will include identified deliverables and milestones
- Visualizations
- Machine learning model.
- Statistical model
- Executive summary that will include the key insights and recommendations



## 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, we then put it to action and examine how well it meets project goals..Different evaluation metrics can be employed to check the model performance. For example accuracy, recall, precision to assess how the model performs on test data.



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

Why did you select these stages for this task?

The business needs and machine learning models require some statistical tests to be performed. The tests are performed in the construct stage.

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

Why did you select this stage for this task?

At this stage, I get to perform EDA to understand the data. Learn the structures, uncover patterns and address irregularities. It is at this stage that I get to explore the relationship between the variables, perform feature selection, extraction and transformation.

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

Why did you select these stages for this task?

Because real life data is often messy with lots of irregularities, EDA will include data cleaning. We might find in the construct stage that we need to explore the predictive power of certain features.

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

Why did you select this stage for this task?

It all begins here. Creating a PACE document and outlining a project workflow ensures that my model construction goals align with the actual business needs. This helps inform what decisions should be made.

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

Why did you select this stage for this task?

At the end of the day, we want to communicate the result with stakeholders to get their feedback and see if we address the business needs. Of course communication is necessary at all stages of the project.



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

Why did you select this stage for this task?

Descriptive statistics is performed during the analysis stage.

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

Why did you select these stages for this task?

Gaining insight about the data begins during the analysis stage. Final insight is constructed at the construct stage. The

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

Why did you select this stage for this task?

The project proposal is the first deliverable of the project. Here we highlight the business needs and appropriate method to address the needs

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 the dataset to get a summary overview of the features is part of this stage. I want to ensure I understand the data for accurate predictions.



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

Why did you select this stage for this task?

With a solid plan, we are ready to build a machine learning model. This building of the model happen at this stage.