

# 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.

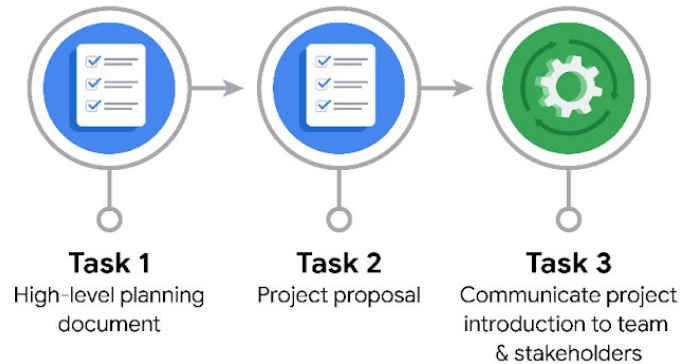
### Relevant Interview Questions

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?

The audience are the users who are going to use the application where they report the videos or comments

- 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 solve the backlog of reports where I want to identify the reports quickly and efficiently to find whether it is a claim or an opinion where I can prioritize the claims

- What questions need to be asked or answered?

What are the characteristics that identifies the claim?  
 What are the characteristics that identifies the opinions?  
 Are the opinion reports going to be reviewed or are they going to be kept in low priority.

- What resources are required to complete this project?

- 1- A data set where they have all the needed data
- 2- the tools needed in python to complete the task

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

- 1- machine learning models that can *identify the claim from reports*
- 2- *a report describing the projects step from the beginning.*
- 3- *visuals to give to the stakeholders along the presentation*

## 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: Construct

Why did you select this stage for this task?

Evaluating the model is a part of the Construct where you build and test the models to find the best suited one for the project

### 2. Conduct hypothesis testing: Plan and Analyze

Why did you select these stages for this task?

The plan is where you give your hypothesis and the analysis is where you test this hypothesis and by that you go back and forth between them in order to find the hypothesis that can be tested true.

### 3. Begin exploring the data: Analyze

Why did you select this stage for this task?

Because in the first stage you only try to gather the data and after that you start exploring it only once you believe you have accumulated the required the data for the project.

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

Why did you select these stages for this task?

It is important to clean the data before testing any hypothesis in the analysis phase and to also reformat the to suit you need in you specific model in the construct phase.



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

Why did you select this stage for this task?

It is the first thing you do in the project where you are

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

Why did you select this stage for this task?

At this point you begin showing you findings and model accuracy to you stakeholders and show them visualization to make it easier for them to understand without knowing the technical stuff.

**7. Compute descriptive statistics: Construct**

Why did you select this stage for this task?

It is a part for building the machine learning model and hence it is part of the construct

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

Why did you select these stages for this task?

Building is for the Construct where you make the machine learning models and the visualization is from the execute where you showcase your results to you stakeholders.

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

Why did you select this stage for this task?

The proposal is part of the planing for the workflow and hence it is from the plan phase



**10. Build a regression model:** Construct and Select PACE stage

Why did you select this stage for this task?

**I selected the Construct because at this stage you begin making you machine learning models.**

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

Why did you select this stage for this task?

**You start to make the summary and the EDA in the analyze stage**

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

Why did you select this stage for this task?

**You begin making your machine learning models in the Construct**