

# 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.

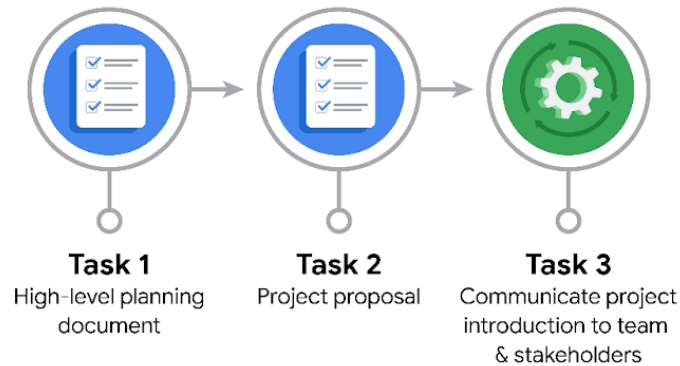
### Relevant Interview Questions

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 for this project are cross-functional team members who are managers that oversee operations whose responsibilities are less technical.

- 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 create a predictive Machine Learning Model that will determine whether a video on TikTok contains a claim or just an opinion, and will be able to differentiate between the two with a high level of accuracy. This is because due to the large amount of users on TikTok there is a massive backlog of claims and so this will help classify them and reduce the backlog.

- What questions need to be asked or answered?

The Questions that need to be asked/answered are:



What is the condition of the provided dataset?  
Is the data reasonably formatted?  
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?  
To what degree of accuracy do you want the ML Model to be?  
What type of regression model will be needed?  
What method of Hypothesis testing will be needed?

- What resources are required to complete this project?

The project dataset, Python Notebook and the input from stakeholders.

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

The deliverables include Data Cleaning, EDA, Visualizations, Statistical Model, Regression Analysis and/or Machine Learning Model, Testing the accuracy of the Model, compiling a presentation for stakeholders.



## 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?



Once the model is built, data is processed through it to assess whether it aligns with the project's objectives and expectations.

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

Why did you select these stages for this task?

Due to the analyze stage is where I would be doing EDA and working with the data to get it ready for the Construct stage which is where the uncovering of relationships in the data happens. And to uncover those relationships between variables in the data, you use hypothesis testing.

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

Why did you select this stage for this task?

During the analysis phase, you will develop a deeper insight into the dataset and the information it contains.

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

Why did you select these stages for this task?

Planning begins when you initially decide on the appropriate methods to use. The data cleaning process follows during the analysis stage.

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

Why did you select this stage for this task?

Planning Stage, Developing an initial project PACE document outlines the workflow and aids in determining the most effective approach to the project.

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

Why did you select this stage for this task?

Communication is essential at multiple stages of a project. In the execute phase of the data project workflow, final insights are presented to stakeholders.



**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 starts with evaluating the data and is developed during the construction stage.

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

Why did you select this stage for this task?

This is where I will define the scope of my project with my team and so therefore where I would create the project proposal.

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

Why did you select this stage for this task?

In the analysis stage, the model is thoroughly reviewed to ensure it meets the requirements of the task. The regression model is then built during the construction phase.

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

Why did you select this stage for this task?

In the Analyze stage is where EDA takes place and so summaries about the data can be formed.

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

Why did you select this stage for this task?

The construction of a data model occurs during the construct stage.