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.
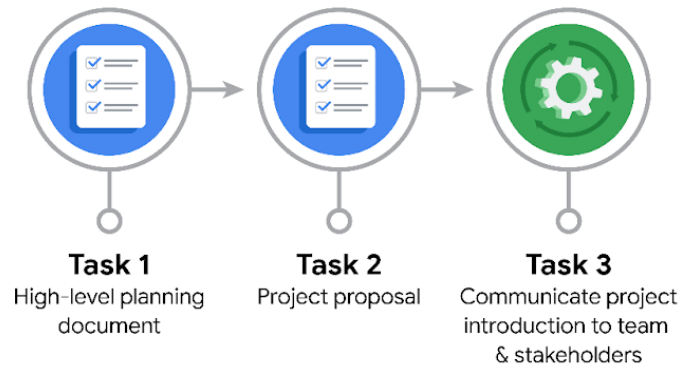
Relevant Interview Questions

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.

Task 1
High-level planning document

Task 2
Project proposal

Task 3
Communicate project introduction to team & stakeholders

Data Project Questions & Considerations

PACE: **Plan Stage**

- Who is your audience for this project?

Moderation team at TikTok

- 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?

The reduction in backlog of user reports and to prioritize them efficiently. As a result of this, the TikTok userbase would be satisfied with the efficiency of the moderation team in reviewing the reports which can have a positive impact on retaining the users.

- What questions need to be asked or answered?

1. What is the condition of the provided datasets?
2. What variables will be the most useful?
3. Are there trends within the data that can provide insight?

4. What steps can I take to reduce the impact of Bias?

- What resources are required to complete this project?

Multiple datasets, tools such as Python and Tableau, input and feedback from stakeholders.

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

1. Project document
2. Data prepared for modeling
3. Machine learning model
4. Update to Stakeholders
5. Visualizations
6. Executive Summary
7. Results

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?

   > The reason is because in the construct stage we build and revise the Machine Learning models.

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

   Why did you select these stages for this task?

   > The reason being that in the Analyze stage, we indulge in hypothesis testing to draw out inferences from the data. In the Construct stage, we carry out hypothesis testing during the development of our Machine Learning models.

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

   Why did you select this stage for this task?

   > The reason is that in the Analyze stage, we process the data, conduct methodological testing and EDA on the data.

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

Why did you select these stages for this task?

The reason is that during the Planning phase, do we draw out a project proposal, get a brief of the data, its background, the problem, the business task and we further explore the data in the Analyze stage where we also clean it using various tools and methods to get it ready for analysis.

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

Why did you select this stage for this task?

In the planning stage, we ask questions about our goals, strategies, and we establish a structure which translate to the timeline, the milestones and the tasks within and the deliverables that we expect from our milestones.

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

Why did you select this stage for this task?

In the Execute stage, we share final actionable recommendations, visualizations, and the Machine Learning model itself to get feedback.

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

Why did you select this stage for this task?

We engage in statistical inference which involves the use of statistical methods including descriptive statistics.

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

Why did you select these stages for this task?

The reason is that in the Analyze stage, visualization is particularly helpful in displaying data as well as in identifying trends and patterns. Similarly, in the Construct phase, visualization is using to assess the Machine Learning models.

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

Why did you select this stage for this task?

Project proposal sets up the workflow structure for our project which is the initial document used to define the project.

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

Why did you select this stage for this task?

In the Analyze phase, based on the findings, we decide upon a model to be pursued in the Construct phase and the construct phase itself is where the regression model is built upon.

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

Why did you select this stage for this task?

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

Why did you select this stage for this task?

The reason is because as the name suggest, it's the construction phase where the Machine learning model is developed.