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.

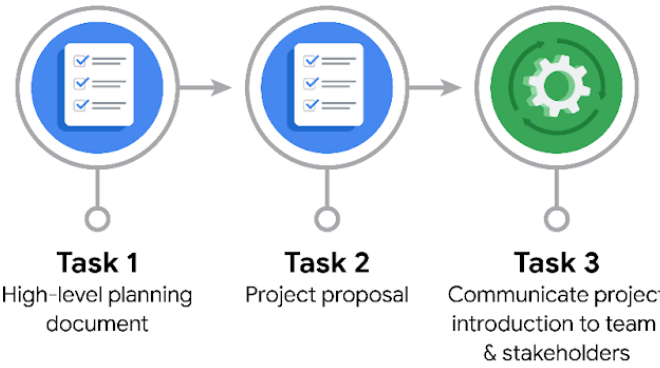
Relevant Interview Questions

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 Automatidata team and New York City Taxi and Limousine Commission (TLC) members.

- 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?

This project is to use the TLC's data to develop the regression model to predict the taxi fares before each ride. By developing the model, the TLC members will be able to forecast each of the fare trips based on the features that might be influenced the most. Then, they might have the appropriate measures to take.

- What questions need to be asked or answered?

- What variables will be the most useful to consider?
- What specific steps do we need to take in the EDA process?

- How do we deal with missing values based on the given data?
- Do we need to consider using Tableau to enrich our visualizations?
- Besides using the A/B test, what other kind of statistical testing will we approach?
- What are the considerations of the machine-learning model(s) we might choose?

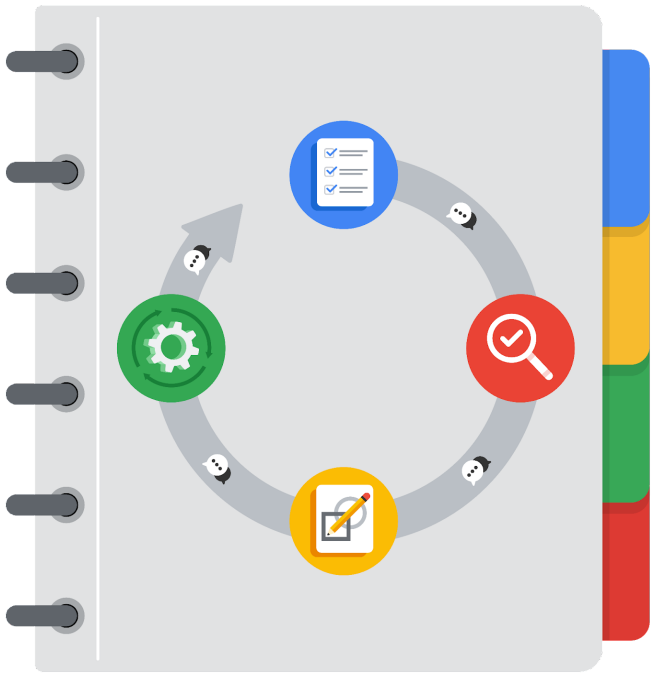
- What resources are required to complete this project?

TLC data was provided, and Jupyter Notebook to complete the workbook.

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

- In the **Plan** stage, the project's scope should be done, and the whole project's proposal should be outlined with the tasks specified in each stage.
- In the **Analyze** stage, the data needs to be well-clean and formatted and be ready for developing model(s).
- In the **Construct** phase, determine the best model to build, and test the accuracy and performance.
- In the **Execute** stage, key insights and results of the model will be reported to the stakeholders. Also, address the feedback 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?

   This stage will be used after the model has been completed. We want to test and evaluate the model's performance, and address the next steps if the model is outperformed or needs to revise.

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

   Why did you select these stages for this task?

   In the **Analyze** stage, we determine the useful variables to be ready for testing (and before that is to perform necessary tasks which include cleaning, EDA, etc). The **Construct** phase is when we're ready to test the hypothesis, which contain t-test, ANNOVA, etc.

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

   Why did you select this stage for this task?

   When we have the dataset in hand, we can jump right into the exploration of data to know and understand about what's given.

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

Why did you select these stages for this task?

From the **Plan** stage, we determine on the methods to perform on the **Analyze** stage (using Python for instance). The majority of exploring and cleaning data is from the **Analyze** stage, which we try to understand every single variables, its meaning, and perform necessary cleaning tasks (missing values, outliers, etc.)

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

Why did you select this stage for this task?

This will align with the project's proposal, which is outlined the whole project's process from the beginning to end, at every single stage.

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

Why did you select this stage for this task?

This stage is when we'll deliver key inisghts/information to stakeholders after going through the analyzing process, EDA, building model, testing the performance, and evaluating. Based on the model's results, we address the specific insights to the stakeholders and recommend the next steps.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

This stage is whe the majority we dig deeply into the dataset, and compute descriptive statistics is to understand what each variable represents, normally with numerical.

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

Why did you select these stages for this task?

In the **Analyze** stage, it's necessary to build visualizations to better understand the relationship between variables and look for the important insights. In the **Construct** phase, for instance of building regression model, we visualize how the relationship between variables, and model's assumptions.

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

Why did you select this stage for this task?

This is the first starting point of the project, which is to outline the milestones and tasks at each stage specifically. This will be tremendously helpful for the teammates to know what their responsibilities are.

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

Why did you select this stage for this task?

In the **Analyze** stage, we choose the best mode's approach based on the dataset is given, and check the assumptions to determine if it is met. Then from the **Construct** phase, we jump right into developing our chosen regression model.

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

Why did you select this stage for this task?

This stage is when after we perform necessary tasks with the dataset (cleaning, EDA, visualizations, etc.), we then compile what we get and learn from the data.

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

Why did you select this stage for this task?

This stage is used when we finally choose the appropriate model(s) approach to develop.