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.

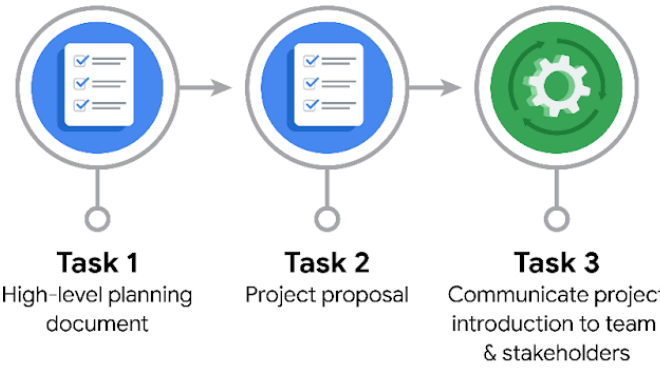
Relevant Interview Questions

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 for this project is one, the stakeholders: the New York City Taxi and Limousine Commission (TLC) members who are 1. Juliana Soto, Finance and Administration Department Head

2. Titus Nelson, Operations Manager

and

Two, the Automatidata Team Members

namely:

1. Udo Bankole, Director of Data Analysis
2. Deshawn Washington, Data Analysis Manager
3. Luana Rodriquez, Senior Data Analyst
4. Uli King, Senior Project Manager
5.
6.

- 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 main aim of the project is to develop a regression model that will accurately predict taxi fare prices before each ride in New york city. It will use historical data provided by the new york city taxi including information like distance and time of day.

By providing a more predictable and transparent fare system, the TLC can improve customer satisfaction and potentially attract new riders. The insights gained from the project can help the TLC make informed decisions about fare regulations, taxi fleet management, and overall service improvement. The success of this project could pave the way for further technological advancements in the taxi industry, such as dynamic pricing models or real-time fare estimates.

With a better understanding of fare pricing, drivers can optimize their routes and earnings potential.

- What questions need to be asked or answered?

Questions like what are the specific objectives for the development of this regression model and against what metrics will the success be measured.

What are the limitations of the project, is there too little data available or are there issues with the model being too complex?

What are the ethical considerations involved in using this data and model?Like is there too much bias in the data or are there any legal issues associated with the data that need to be checked on.

Data understanding and exploration:

What variables or features are present in the data and which ones will be relevant when predicting taxi fares?

Are there missing values, outliers or inconsistencies that need to be taken into account before the actual construction phase reaches?

Model development and evaluation:

Based on the fact that this is a predictive project challenge, regression analysis and therefore regression models are going to be used. So we can ask questions like, what regression models are going to be used. Is it linear or random forests?

What evaluation metrics will be used to evaluate the model performance?

Communication and deliverables:

There are key deliverables that are obviously going to be happening, like project proposal, interim reports, final model and final presentation. So what are the key deliverables of the project and when are they due?

What are the communication needs of the different stakeholders?

What is the estimated timeline for completing the project? What resources are needed to complete the project?

What are the potential risks and mitigation strategies?

What training and support will the TLC team need to understand the project or rather models to be used.

How can we ensure long-term maintenance of the model?

- What resources are required to complete this project?

The specific resources needed will vary depending on the complexity of the project, the funds available and the size of the team.

Human resources;

Data analyst- the one who's responsible for data cleaning, exploration , model development, evaluation and reporting

Data scientist- who may be needed for more complex modeling tasks.

Project manager - Ensures projects stay on track, manages resources and communicates with stakeholders.

Domain expert - someone who is familiar with the taxi industry and its nuances to provide insights and feedback.

Software tools

Data resources like the dataset.

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

Several deliverables will be created to document progress, communicate insight and ensure a successful outcome.

A project strategy document PACE that outline the project tasks, timelines considering team needs, will be created.

A project proposal that summarizes the project goals, the methodology, budget and the timeline, seeking approval from the stakeholders.

A data acquisition plan that outlines how data will be obtained, cleaned and accessed securely.

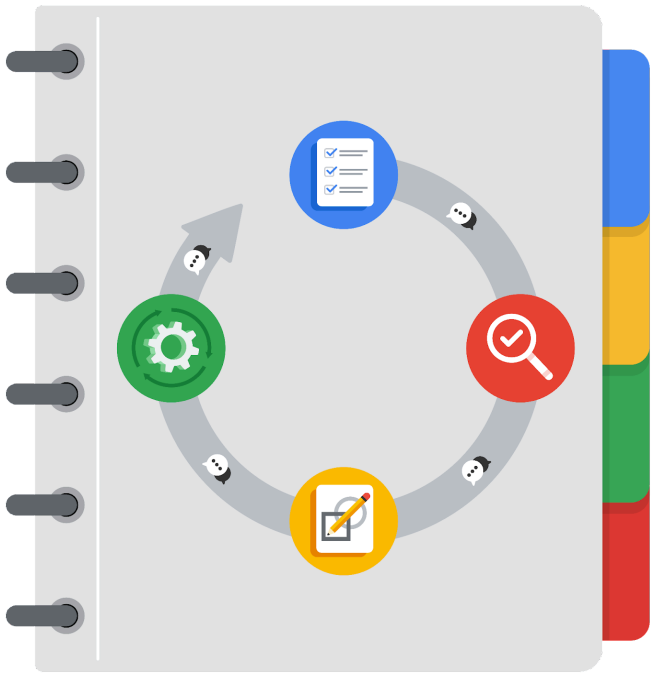
An EDA report that summarizes the key qualities of data like data quality and feature distributions.

Model development report that explains the modelling approach chosen, the hyperparameter tuning process and the chosen evaluation approach.

Model evaluation report that presents the model's performance on various metrics including accuracy, precision and recall.

A clear and concise presentation of key findings and recommendations for the 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?

Only happens after the model has been constructed . A run through of the data is done to determine if it meets the projects' expectations and goals.

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

   Why did you select these stages for this task?

During the analyzing stage, it is determined that a statistical test will be used and during construction stage, the test is carried out.

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

   Why did you select this stage for this task?

It is at the analysis stage that you gain deeper understanding of the data and discover patterns or insights that might be helpful during the model building process.

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

   Why did you select these stages for this task?

Planning takes place when you determine when you make choices of the methods needed then the cleaning process takes place when in the process of analyzing.

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

Why did you select this stage for this task?

It is at the planning stage that you come up with a document that outlines a workflow that will help approach a project.

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

Why did you select this stage for this task?

Inorder for findings to be available first, we'll have to do the construction of the model and evaluate it to come up with findings . All this is part of execution.

7. **Compute descriptive statistics:** Construct

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 begins with data assessment and is created during the construction stage.

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

Why did you select this stage for this task?

Planning stage. A project proposal is the initial document used to define a project.

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

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 a dataset to compile information would take place in the analysis phase.

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

Why did you select this stage for this task?

The building of a data model would take place in the construct stage.

This whole process is iterative.