

Foundations of Data Science



Project Proposal and PAYS Strategy Document

Foundation to Study Data Science - Google x Coursera

This document presents a comprehensive project proposal and PACE (Plan, Analyze, Construct, and Execute) strategy as part of the "Foundation of Data Science" course by Google on Coursera. Through this project, I demonstrate my ability to apply structured data science workflows, implement PACE strategy, and propose actionable recommendations tailored to specific business challenges, and a deep understanding of data workflows.

Scenario Context: Automotive Industry – Automate data and New York City Taxi and Limousine Commission (TLC)

The project is about a collaboration between Automatedata, a data consulting firm, and the New York City Taxi and Limousine Commission (TLC). Automatedata specializes in transforming clients' unused and stored data into actionable insights, such as performance dashboards, customer-facing tools, and strategic business solutions.

In this project, TLC is collaborating with Automatedata to build this project, the project involves developing a regression model using TLC's data to predict taxi fares before a ride begins. This project aims to improve the user experience by providing transparent and reliable fare estimates, fostering passenger confidence, and enhancing the overall trust and convenience of New York City's taxi services.

This document displays my structured approach to the business demands of this project. It shows my understanding of data preparation, exploratory analysis and predictive modeling while showcasing my ability to align technical solutions with real-world business needs and objectives.



Data Project Questions & Considerations



PACE: Plan Stage

- Who is your audience for this project?

NYC Taxi and Limousine commission

- 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?

estimate cab fares before each ride based on distance, time of day, and any additional variables we find necessary.

- What questions need to be asked or answered?

- 1- To understand the data set there will be many questions, such as: what does this column mean? Why do these numbers present? Is conversion rate being considered? Etc....
- 2- What are the variables that will be most useful ?
- 3- Are there any trends within the data that can provide insight?
- 4- What steps can I take to reduce the bias?

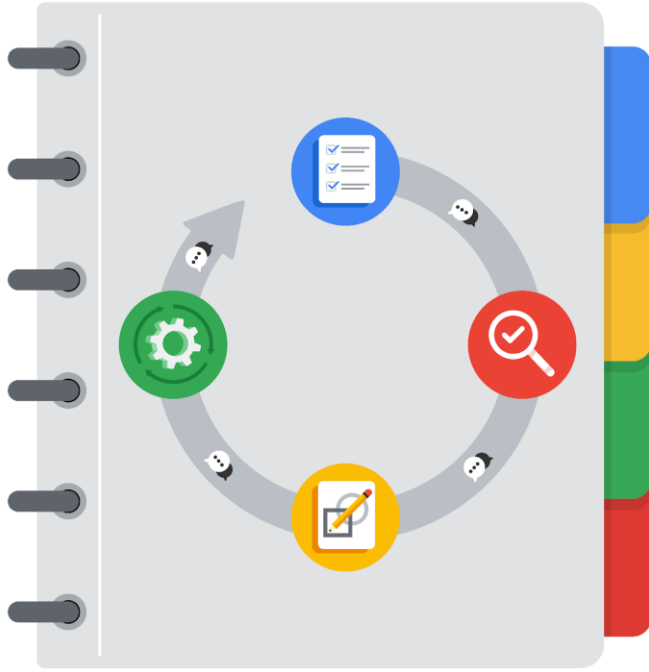
- What resources are required to complete this project?

A data set, python, input from stakeholders

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

A regression model that predicts taxi fares, visualizations, a data set after being scrubbed.

THE PACE WORKFLOW



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.

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, explain why you selected the stage for each task.

1. Evaluating the model: **Execute**

Why did you select this stage for this task?

After we build the model, we evaluate it to see if it meets the project's expectations and goals.

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

Why did you select these stages for this task?

Analyze -> in this phase, it is determined that a statistical test will be used.



Construction -> the testing will be carried out in this phase.

3. Begin exploring the data: Analyze

Why did you select this stage for this task?

During the analysis phase, you will gain a deeper understanding of the dataset and the content inside of it.

4. Data exploration and cleaning: Plan and Analyze

Why did you select these stages for this task?

There are methods that you will use to clean the data, these will be decided in the plan phase, and they will take place in the analyze phase.

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

Why did you select this stage for this task?

Creating an initial project PACE document outlines the workflow and helps to plan how to best approach a project.

6. Compute descriptive statistics: Analyze

Why did you select this stage for this task?

Investigating the statistics within the data takes place during the analysis.

7. Visualization building: Analyze and Execute

Why did you select these stages for this task?

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

8. Build a regression model: Analyze and Construct

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

9. 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.