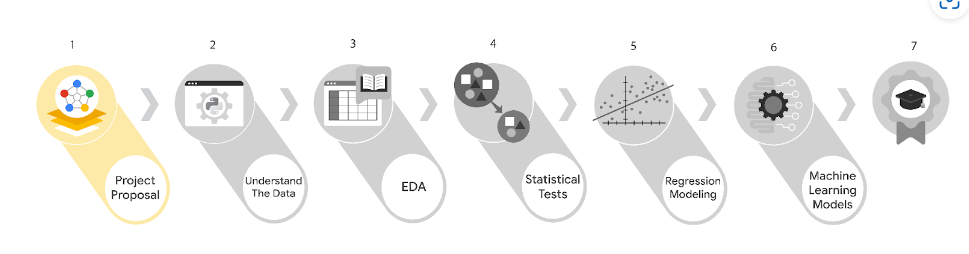
**Automatidata project**

****

[**Project Overview 1**](#_ktz5mlu0b7kz)

[**Project Proposal 1**](#_pyn5blknv22f)

[**PACE strategy document 3**](#_jc0lgu7c89g5)

[**Automatidata workplace scenario 8**](#_3bn06ccecgl5)

## 

# Project Overview

The New York City Taxi and Limousine Commission seeks a way to utilize the data collected from the New York City area to predict the fare amount for taxi cab rides.

# Project Proposal

| **Milestones** | **Tasks** | **Deliverables/Reports** | **Relevant Stakeholder (Optional)** |
| --- | --- | --- | --- |
| **1** | **Establish structure for project workflow (PACE)Plan** | * Global-level project document | Deshawn Washington — Data Analysis Manager |
| **1a** | **Write a project proposal**  **Plan** |  | Uli King — Senior Project Manager |
| **2** | **Compile summary information about the dataAnalyze** | * Data files ready for EDA | Luana Rodriquez — Senior Data Analyst |
| **2a** | **Begin exploring the data**  **Analyze** |  | Deshawn Washington — Data Analysis Manager |
| **3** | **Data exploration and cleaning**  **Plan and Analyze** | * EDA report * Tableau dashboard/visualizations | Luana Rodriquez — Senior Data Analyst |
| **3a** | **Visualization building**  **Construct and Analyze** |  | Uli King — Senior Project Manager |
| **4** | **Compute descriptive statisticsAnalyze** | * Analysis of testing results between two important variables * Share results of testing | Deshawn Washington — Data Analysis Manager |
| **4a** | **Conduct hypothesis testing**  **Analyze and Construct** |  | Udo Bankole — Director of Data Analysis |
| **5** | **Build a regression model**  **Analyze and Construct** | * Review testing results * Determine the success of the model | Luana Rodriquez — Senior Data Analyst |
| **5a** | **Evaluate the model**  **Execute** |  | Udo Bankole — Director of Data Analysis |
| **6** | **Communicate final insights with stakeholders**  **Execute** |  |  |
| **6a** | **Not necessary for this project**  **Select PACE stage** |  |  |

Realistic timelines when working with actual clients and data scientists as a data scientist would most likely have tight deadlines, for example:

Milestone 1: 1-2 days

Milestone 2: 2-3 weeks

Milestone 3: 1 week

Milestone 4: 1 week

Milestone 5: 1-2 weeks

# PACE strategy document

## Instructions

Use this 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?

***If you selected the Automatidata scenario***, your audience will be the New York City 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?

***If you selected the Automatidata scenario***, the problem you are trying to solve is the estimation of taxi fares based on relevant variables that you identify.

* What questions need to be asked or answered?

***For all projects,*** consider the following questions:What is the condition of the provided dataset? 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?

* What resources are required to complete this project?

***For all projects,*** you will needthe project dataset, Python notebook, and input from stakeholders.

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

***For all projects, t***he deliverables include*a*dataset scrubbed for exploratory data analysis, visualizations, statistical model, regression analysis and/or machine learning model.

## 

## **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?

After the model has been constructed, data is run through to evaluate whether it meets the project’s expectations and goals.

1. **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. During the construction phase, the test is carried out.

1. **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 information within it.

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

Why did you select these stages for this task?

Planning takes place when you first make choices about the methods needed. The cleaning process then takes place in the analyzing stage.

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

Why did you select this stage for this task?

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

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

Why did you select this stage for this task?

Communication is necessary at various points throughout a project. Final insights are shared with stakeholders in the execute phase of the data project workflow.

1. **Compute descriptive statistics: Analyze**

Why did you select this stage for this task?

Investigating the statistics within data takes place during analysis.

1. **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.

1. **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.

1. **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.

1. **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.

1. **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.

# **Automatidata workplace scenario**

The end-of-course project in Course 1 focuses on your ability to plan for data projects and create a project proposal. The end-of-course projects were designed with you in mind, offering an opportunity for you to practice and apply your data analytic skills. The materials provided here will guide you through discussions with co-workers, internal team members, and external stakeholders.

Learn more about the project, your role, and expectations in this reading.

## **Background on the Automatidata scenario**

Congrats on your new job as a data analyst at a data consulting firm called Automatidata. Automatidata works with its clients to transform their unused and stored data into useful solutions, such as performance dashboards, customer-facing tools, strategic business insights, and more. They specialize in identifying a client’s business needs and utilizing their data to meet those business needs.

Automatidata is consulting for the New York City Taxi and Limousine Commission (TLC). New York City TLC is an agency responsible for licensing and regulating New York City's taxi cabs and for-hire vehicles. The agency has partnered with Automatidata to develop a regression model that helps estimate taxi fares before the ride, based on data that TLC has gathered.

The TLC data comes from over 200,000 taxi and limousine licensees, making approximately one million combined trips per day.

**Note:***This project's dataset was created for pedagogical purposes and may not be indicative of New York City taxi cab riders' behavior.*

### 

### **Project background**

Automatidata is in the earliest stages of the TLC project. The following tasks are needed before the team can begin the data analysis process:

* A project proposal identifying the following:
  + Organize project tasks into milestones
  + Classify tasks using the PACE workflow
  + Identify relevant stakeholders

### **Your assignment**

For your first assignment, Automatidata will need a project proposal that will create milestones for the tasks within the TLC project. Remember to take into account your audience, team, project goal, and PACE stages of each task in planning your project deliverable.

## **Team members at Automatidata and the New York City TLC**

### **Automatidata Team Members**

* Udo Bankole, Director of Data Analysis
* Deshawn Washington, Data Analysis Manager
* Luana Rodriquez, Senior Data Analyst
* Uli King, Senior Project Manager

Your teammates at Automatidata have technical experience with data analysis and data science. However, you should always be sure to keep summaries and messages to these team members concise and to the point.

### **New York City TLC Team Members**

* Juliana Soto, Finance and Administration Department Head
* Titus Nelson, Operations Manager

***Note:*** *The story, all names, characters, and incidents portrayed in this project are fictitious. No identification with actual persons (living or deceased) is intended or should be inferred. The data shared in this project has been altered for pedagogical purposes.*

The TLC team members are program managers who oversee operations at the organization. Their roles are not highly technical, so be sure to adjust your language and explanation accordingly.

## **Meeting notes**

Now that you are working as Automatidata’s latest data analytics professional, you are given access to the company network and set up with a company email account (your first initial and last name, followed by @automatidata.org).

Opening your inbox, you notice an email from your supervisor, Deshawn.

**From:** Deshawn Washington

**Subject:** Review meeting notes

If you are able to read this, then your company accounts have been created! Now is the perfect time to get started. Last week, I attended an internal meeting with our leadership team about a new project we are about to begin. You’ll receive more information in the next few days, but I would like you to be aware of some needs that were identified by our leadership team. Here is an excerpt from the notes I took during the Automatidata leadership team meeting. I’ve organized the points by the person who made them.

Uli King (Senior Project Manager)

* The data team will need a global-level project document to outline the goals and milestones.
* I am working closely with Titus Nelson over at the New York City Taxi and Limo Commission. He has requested some visuals to share with TLC’s executives.

Luana Rodriquez (Senior Data Analyst)

* The dataset from TLC has to be inspected before any analysis can begin.
* Our team needs to determine what information the TLC data provides through exploratory data analysis (EDA).
* Eventually, our team will need to test to find if the model is delivering consistent results.

Udo Bankole (Director of Data Analysis)

* Before we present any insights to TLC, we'll need to determine whether or not the model we produce meets the project requirements.
* Once we have a final model, I'll need to know the main talking points going into our presentation with TLC.

My thoughts and concerns…

* I think it's best to use Python for the TLC project. I'll have someone on my team set that up as soon as we have the plan in place.
* It will be important to establish the relationship between any variables within the TLC data. I'd suggest the data team consider A/B testing, since that will analyze the relationship between the two most useful variables and subsequently provide data-driven support for future business decisions.

Review the meeting notes above to become familiar with the project’s context. I’ll ask you to identify project tasks and come up with a structure to guide the data team through this project. After our discussion about your experience in the certificate program offered by Google, I know that your efficient communication style and problem-solving will enhance the abilities of the data team.

There will be more details sent to you very soon.

Welcome to the team,

Deshawn Washington

Data Analysis Manager

Automatidata

(P.S. There will be muffins in the break room every Tuesday morning. Be early…unless you like bran muffins. LOL)