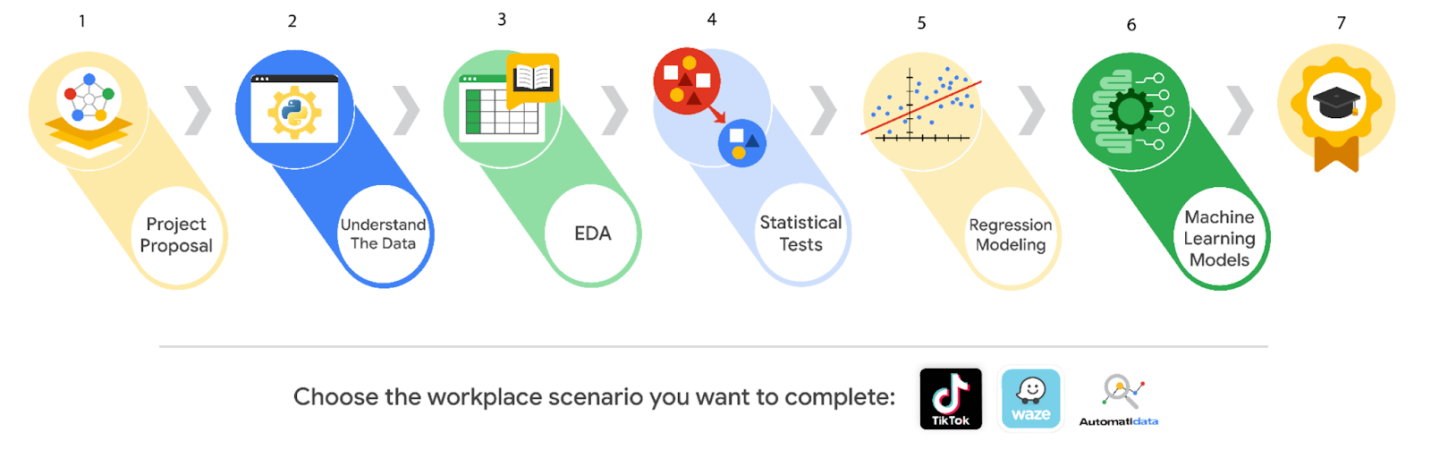
# Portfolio Project Introduction

This project was completed as part of the Google Data Analytics Portfolio project.

Each course in the Google Advanced Data Analytics Certificate concludes with a project that provides hands-on opportunities to practice your knowledge. These end-of-course portfolio projects build across courses to simulate the full lifecycle of a data project, just like job tasks that you will encounter as a data professional.



## ****Importance of communication in the data career space****

End-of-course projects emphasize technical and professional workplace skills, as well as the importance of communication in data analytics. The success of your workflow management, data analysis, data visualizations, statistics, regression analysis, and building machine learning models relies on your ability to communicate with cross-functional team members. So, each project is designed to help you develop critical communication skills that you’ll need to use on the job in order to effectively:

* Ask questions
* Share project needs
* Communicate with stakeholders
* Give and receive feedback
* Stay in contact with team members on the project

To help you complete each end-of-course project, you will receive a PACE (Plan, Analyze, Construct, and Execute) strategy document to guide your workflow. As you answer the questions, you will gain a deeper understanding of the data analysis process, document your growth as a data professional, and prepare the observations and reflections needed to complete each project.

# ****Start your project****

To complete the end-of-course portfolio project, you will first need to:

* Gather information about the business problem or question to be answered.
* Respond to key questions posed in the PACE strategy document.
* Create a project proposal for cross-functional team members.

**AUTOMATIDATA**

**Course 1 end-of-course portfolio project**



***A fictional data consulting firm***

**Project goal:**

In this fictional scenario, the New York City Taxi and Limousine Commission (TLC) has approached the data consulting firm Automatidata to develop an app that enables TLC riders to estimate the taxi fares in advance of their ride.

**Background:**

Since 1971, TLC has been regulating and overseeing the licensing of New York City's taxi cabs, for-hire vehicles, commuter vans, and paratransit vehicles.

**Scenario:**

You are a newly-hired data professional at Automatidata, a fictional data consulting firm. Automatidata’s focus is to help clients transform their unused and stored data into useful solutions. In this scenario, you will consult with The New York City Taxi & Limousine Commission to develop an app that will help users (TLC riders) estimate their taxi fares before their ride. Your first responsibility as a data analytics consultant will be to structure the necessary tasks into a project proposal that establishes milestones for the ride fare data project.

**Course 1 tasks:**

* Gather information from the notes from the last executive meeting of Automatidata
* Assign PACE stages to the requested tasks
* Organize tasks into milestones
* Create a project proposal for the executive team’s approval

***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. And, the data shared in this project has been created for pedagogical purposes.*

# ****Learn about the Course 1 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****

* 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)

## ****Specific project deliverables****

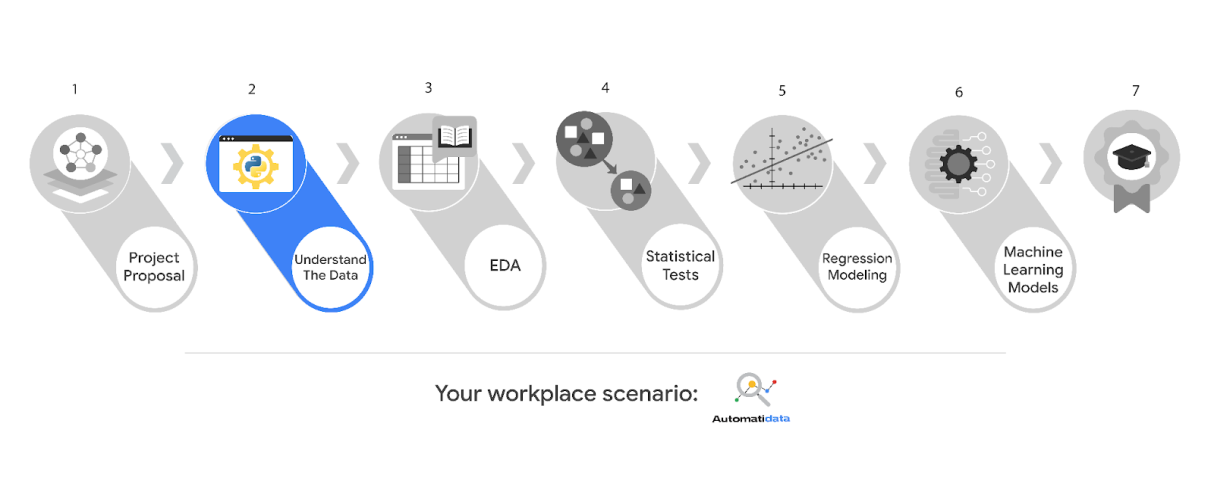
With this end-of-course project, you will gain valuable practice and apply your new skills as you complete the following:

* Course 1 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.

**Course 2 workplace scenarios**

**Course 2 end-of-course portfolio project overview: Automatidata**

The end-of-course project in Course 2 focuses on your ability to understand the data needed for a project. As a reminder, in Course 1 you developed a project proposal that outlined milestones, which progress with each of the end-of-course projects.



**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:

* Build a data frame for the TLC dataset
* Examine data type of each column
* Gather descriptive statistics

**Your assignment**

You will build a data frame for the TLC data. After the data frame is complete, you will organize the data for the process of exploratory data analysis, and update the team on your progress and insights.

**Specific project deliverables**

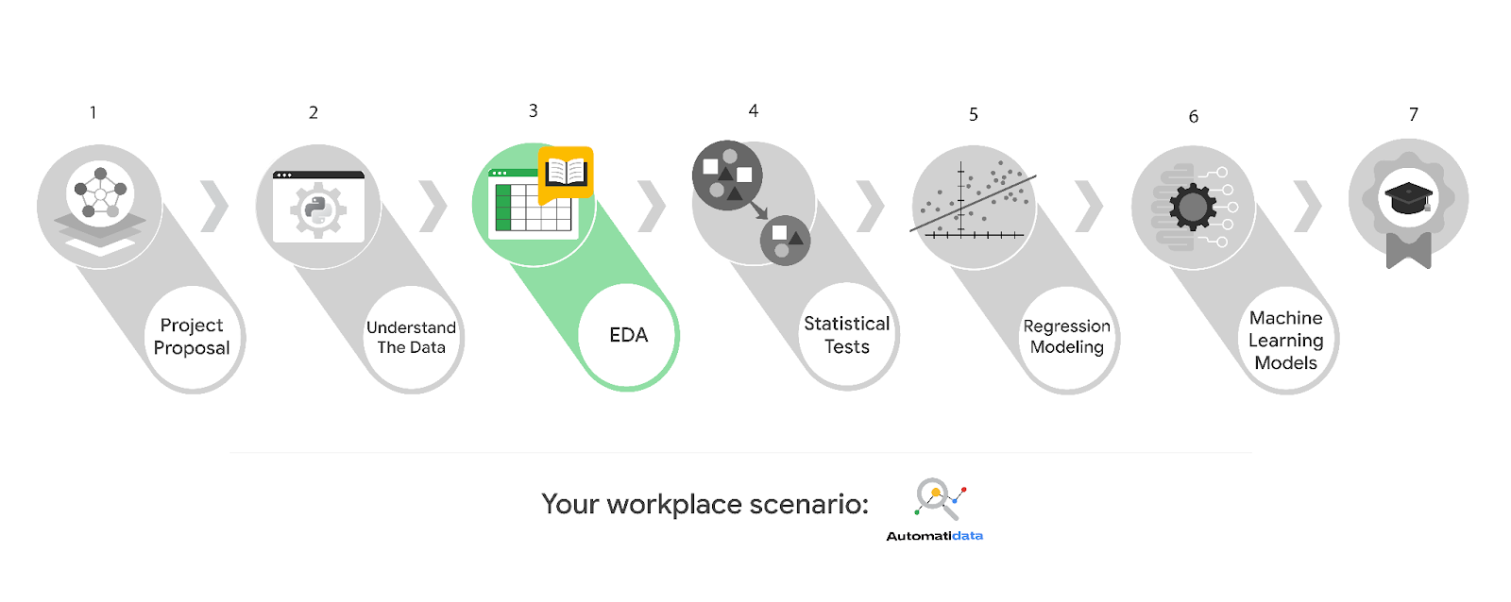
With this end-of-course project, you will gain valuable practice and apply your new skills as you complete the following:

* Complete the questions in the Course 2 PACE strategy document
* Answer the questions in the Jupyter notebook project file
* Complete coding prep work on project’s Jupyter notebook
* Summarize the column Dtypes
* Communicate important findings to DeShawn and Luana in the form of an executive summary

**Course 3 workplace scenarios**

**Course 3 end-of-course portfolio project overview: Automatidata**

The end-of-course project in Course 3 focuses on your ability to use exploratory data analysis to organize and understand the data within a project. 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.

**Project background**

Automatidata is working on the TLC project. The following tasks are needed before the team can begin the data analysis process:

* EDA and cleaning
* Select and build visualization(s) type
  + Create plots to visualize relationships between relevant variables
* Share your results with the Automatidata team

**Your assignment**

You will conduct exploratory data analysis on data for the TLC project. You’ll also use Tableau to create visuals for an executive summary to help non-technical stakeholders engage and interact with the data.

**Specific project deliverables**

With this end-of-course project, you will gain valuable practice of your new skills as you complete the following deliverables:

* Course 3 PACE Strategy Document to consider questions, details, and action items for each stage of the project scenario
* Answer the questions in the Jupyter notebook project file
* Create a Jupyter notebook of full EDA
* Create a Tableau visualization showing two important variables
* Write an executive summary of results and include a visualization

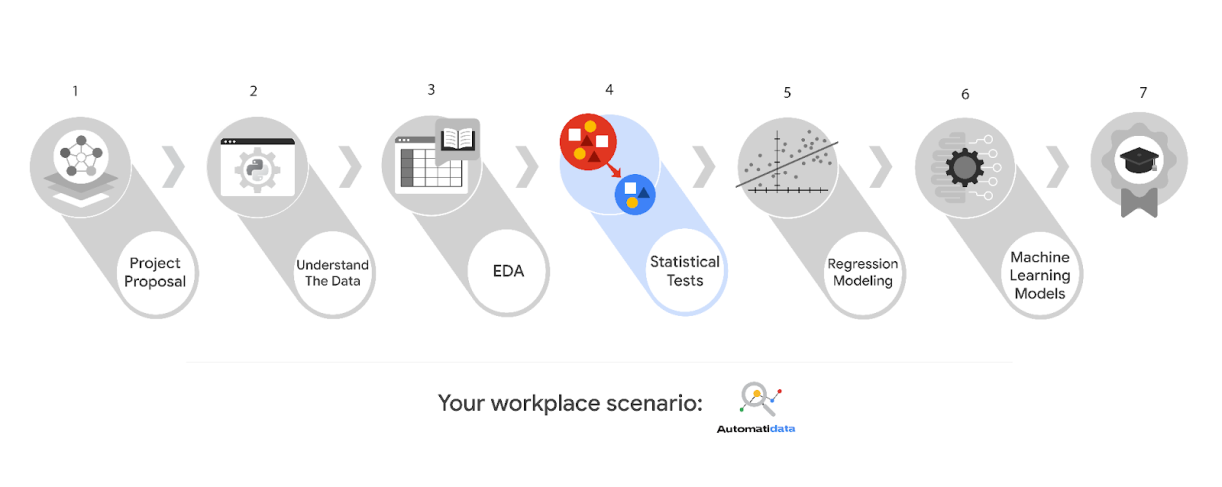
**Course 4 workplace scenarios**

**Scenario:**

Exploratory data analysis is complete for the project. The New York City TLC would like the data team at Automatidata to analyze the relationship between fare amounts and payment type. The team agrees that the next step is to perform a hypothesis test using the data.

# Course 4 end-of-course portfolio project overview: Automatidata

The end-of-course project in Course 4 focuses on your ability to conduct statistical testing within a project. 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.

### ****Your assignment****

You will conduct hypothesis testing on the data for the TLC data. You’ve been asked to investigate TLC’s dataset to determine which hypothesis testing method best serves the data and the TLC project.

## ****Specific project deliverables****

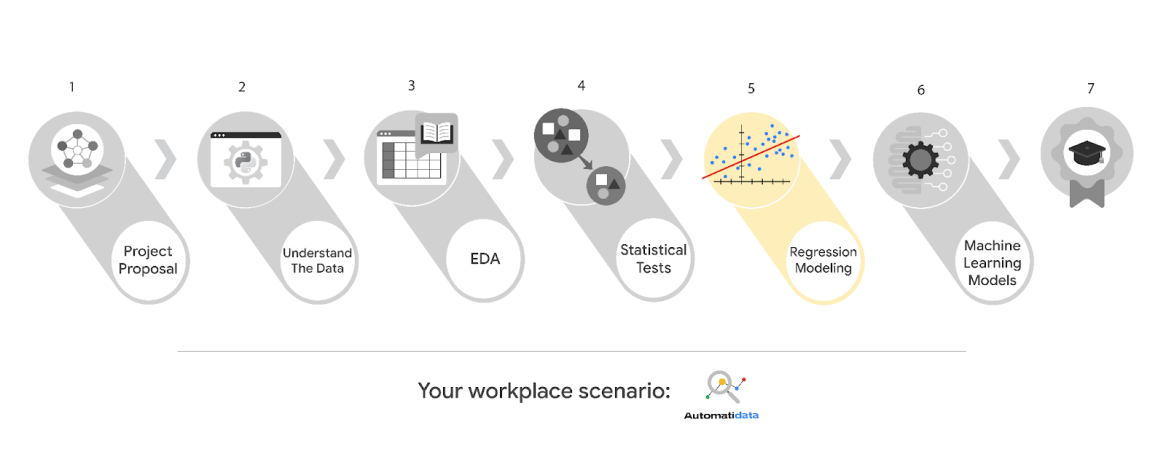
With this end-of-course project, you will gain valuable practice of your new skills as you complete the following deliverables:

* Complete a PACE Strategy Document to consider questions, details, and action items for each stage of the project scenario.
* Answer the questions in the Jupyter notebook project file
* Statistical testing
* Report results in executive summary

**Course 5 workplace scenarios**

**Course 5 end-of-course portfolio project overview: Automatidata**

The end-of-course project in Course 5 focuses on your ability to build regression models using Python. 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.

**Project background**

Automatidata is near the end of the TLC project. The following tasks are needed at this stage of the project:

* Determine the correct modeling approach
* Build a regression model
* Finish checking model assumptions
* Evaluate the model
* Interpret model results and summarize findings for stakeholders within TLC

**Your assignment**

You will create a regression model. Determine the type of regression model that is needed and develop one using the TLC data.

**Specific project deliverables**

In this end-of-course project, you will gain valuable practice of your new skills as you complete the following deliverables:

* Complete a PACE Strategy Document to consider questions, details, and action items for each stage of the project scenario
* Answer the questions in the Jupyter notebook project file
* Build a regression model in Python
* Report the results in an executive summary

**Course 6 workplace scenarios**

**Course 6 end-of-course portfolio project overview: Automatidata**

The end-of-course project in Course 6 focuses on your ability to build a machine learning model for a project. As a reminder, in Course 1 you developed a project proposal that outlined milestones, which progress with each of the end-of-course projects. A visual representation is provided in the graphic shown here:



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

**Project background**

Automatidata is ready to create a machine learning model for TLC. The following tasks are needed to complete the project:

* Model building
* Model evaluation
* Summarize findings for Automatidata and the stakeholders at TLC

**Your assignment**

You will create a machine learning model for the TLC data. You will be responsible for leading these tasks, which include feature engineering, model development, and evaluation.

**Specific project deliverables**

In this end-of-course project, you will gain valuable practice of your new skills as you complete the following deliverables:

* Complete a PACE planner mapping questions, details, and action items for each stage of the project scenario
* Answer the questions in the Jupyter notebook project file
* Design and implement a machine learning model
* Draft an executive summary of your results