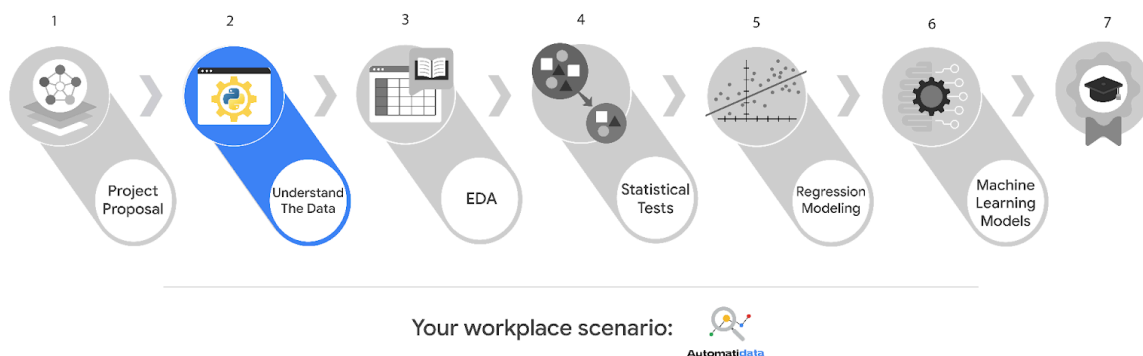


Course 2 end-of-course portfolio project overview: Automatidata

Learn about the Course 2 Automatidata workplace scenario!

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. A visual representation is provided in the graphic shown here:



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

Background on the Automatidata scenario

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.

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 Rodriguez, 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

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.

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.

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 dataframe for the TLC dataset
- Examine data type of each column
- Gather descriptive statistics

Your assignment

You will build a dataframe for the TLC data. After the dataframe 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

Good luck with this project! Automattic looks forward to seeing how you communicate your creative work and approach problem-solving!

Key takeaways

The Google Advanced Data Analytics Certificate end-of-course project is designed for you to practice and apply course skills in a fictional workplace scenario. By completing each course's end-of-course project, you will have work examples that will enhance your portfolio and showcase your skills for future employers.

Scenario

You are the newest member of Automatidata's data analytics team. Your team is still in the early stages of their project for the New York City Taxi & Limousine Commission (TLC).

Previously, you were asked to complete a project proposal by your supervisor, Deshawn Washington. You have received notice that your project proposal has been approved and that TLC has given the Automatidata team access to their data for research purposes. Congratulations! To get clear insights, TLC's data must be analyzed, key variables identified, and the dataset ensured it is ready for analysis.

You discover two new emails in your inbox: one from your supervisor, Deshawn Washington, and one from your teammate, Luana Rodriguez. Review the emails, then follow the provided instructions to complete the PACE strategy document, the code notebook, and the executive summary.

Note: Team member names used in this workplace scenario are fictional and are not representative of the New York City TLC.

Email from Deshawn Washington, Data Analysis Manager

Subject: Help with coding notebook?

From: "Deshawn Washington," Deshawn@automatidata.com

Cc: "Luana Rodriguez" Luana@automatidata.com

Good morning team,

I have a couple of updates on the TLC project. The project proposal that you completed previously has been approved. Thanks for all of your great work so far. Additionally, I just received an email from our Senior Project Manager, Uli King, that TLC has given our team access to their data.

Before we begin the process of Exploratory Data Analysis (EDA), we could really use your help with coding and prepping the data. During your interview you mentioned that you worked with Python during the Google Career Certificate program. That experience sounds applicable here.

Luana (Cc'd) started a Jupyter notebook with the relevant dataset from TLC (attached). She is busy in the final stages of another project currently. I'm sure she

could use your assistance in completing the coding and setting up the notebook for the TLC project.

Luana, do you mind sharing the details?

Humblest regards,

Deshawn Washington

Data Analysis Manager

Automatidata

Email from Luana Rodriquez, Senior Data Analyst

Subject: RE: Help with coding notebook?

From: "Luana Rodriquez" Luana@automatidata

Cc: "Deshawn Washington," Deshawn@automatidata

Nice to meet you (virtually)!

Hope you have enjoyed your first few weeks!

The project proposal you helped prepare covered the major points of this project, so I'll get right to how you can assist the team. There are a number of us making adjustments to the machine learning developed for the last client, so your help is greatly appreciated!

Until we finish the prior project, there is no need to do a full EDA on this data. We will get to that soon. Do you mind reviewing the TLC data we received for the team? It would be fantastic if you could include a summary of the column Dtypes, data value nonnull counts, relevant and irrelevant columns, along with anything else code related you think is worth sharing in the notebook? It would be really helpful if you can create meaningful variables by combining or modifying the structures given.

Thanks,

Luana Rodriquez

Senior Data Analyst

Automatidata