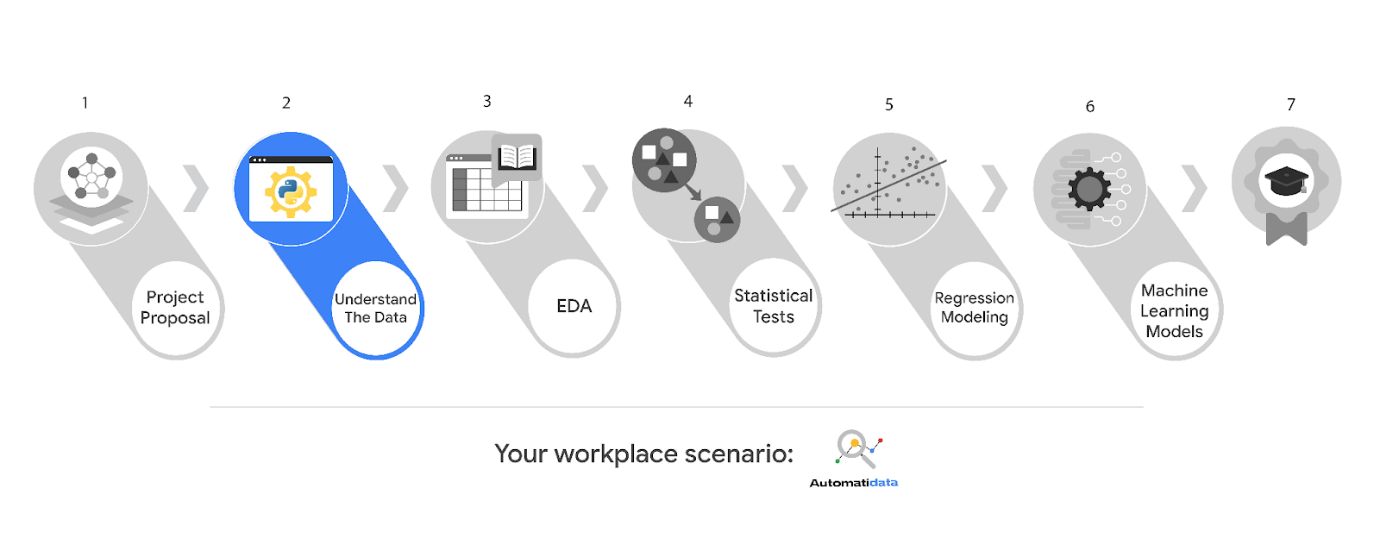
Portfolio Project Overview: Automatidata

**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 predicts taxi and limousine ride durations based on location and time of day 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 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

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 claims 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! Automatidata looks forward to seeing how you communicate your creative work and approach problem-solving!

**Automatidata Project Scenario**



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 Rodriquez” 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 specifically in the Google certificate program you took. 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.com

**Cc:** “Deshawn Washington,” Deshawn@automatidata.com

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 importing the data (attached) and reviewing it 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 showing 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

**Step-By-Step Instructions**



Follow the instructions to complete the activity. Then, go to the next course item to compare your work to a completed exemplar.

**Step 1: Access the templates**



To use the templates for this course item, click each link of the following links and select *Use Template*. Links to templates:

* [Course 2 PACE strategy document](https://docs.google.com/document/d/1JM7h5MAQkD9uxoUhgpVBeT5R15y85LRaMo30I7bIFdk/template/preview)
* [Executive summary templates](https://docs.google.com/presentation/d/1Pps5GKxi1V31y2oRHRzU-xhJubkEYzCgEIfNjlEY3Og/template/preview#slide=id.g1390e7c1ade_0_62)

OR

If you don’t have a Google account, you can download the template directly from the following attachments.

**[Activity Template\_ Course 2 PACE strategy document](https://d3c33hcgiwev3.cloudfront.net/_yBpMx1dSB2uGY2vDfgE7w_03099932865341069e58d99bccdfbaf1_Activity-Template_-Course-2-PACE-strategy-document.docx?Expires=1685577600&Signature=RemphTLk37eFHux4OMb3tAOqvtd~8xwIpTykHOc7SbyChVO1d1psVqCGc6vGiszNDc~I9Riix0jZaiFpbByU0aB-OOFEKbKGYwE-uD87f5Mnla-jbuXFasWHgEoJMQ1DzXo5bg5qZ8ghMYiRE8wrjIHp7kAccbQm5wEs9MEt6O4_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)**

[DOCX File](https://d3c33hcgiwev3.cloudfront.net/_yBpMx1dSB2uGY2vDfgE7w_03099932865341069e58d99bccdfbaf1_Activity-Template_-Course-2-PACE-strategy-document.docx?Expires=1685577600&Signature=RemphTLk37eFHux4OMb3tAOqvtd~8xwIpTykHOc7SbyChVO1d1psVqCGc6vGiszNDc~I9Riix0jZaiFpbByU0aB-OOFEKbKGYwE-uD87f5Mnla-jbuXFasWHgEoJMQ1DzXo5bg5qZ8ghMYiRE8wrjIHp7kAccbQm5wEs9MEt6O4_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

**[Activity Templates\_ Executive summaries](https://d3c33hcgiwev3.cloudfront.net/UfSUFQO0T1ujTYN1-bg7sg_cc7e225098b64eb28b4a4f19a57ea3f1_Activity-Templates_-Executive-summaries.pptx?Expires=1685577600&Signature=YW15CeNVBoZFQxLHz-PRNalwslK63MjWAIUYESTDq6uRRWUx~ZySvpwUE64LoBISwWm4QqNXSKnbM7q2rhWh~wgZ9syINBUFyVKkCzlLFpA643eGouywuhOnfVjhUZoTvKRWffYi-a1x6PczMKKrDljYASQlD-iP9T4J3W5Uies_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)**

[PPTX File](https://d3c33hcgiwev3.cloudfront.net/UfSUFQO0T1ujTYN1-bg7sg_cc7e225098b64eb28b4a4f19a57ea3f1_Activity-Templates_-Executive-summaries.pptx?Expires=1685577600&Signature=YW15CeNVBoZFQxLHz-PRNalwslK63MjWAIUYESTDq6uRRWUx~ZySvpwUE64LoBISwWm4QqNXSKnbM7q2rhWh~wgZ9syINBUFyVKkCzlLFpA643eGouywuhOnfVjhUZoTvKRWffYi-a1x6PczMKKrDljYASQlD-iP9T4J3W5Uies_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

**Step 2: Access the end-of-course project lab**

***Note****: The following lab is also the next course item. Once you complete and submit your end-of-course project activity, return to the lab instructions’ page and click* ***Next*** *to continue on to the exemplar reading.*

To access the end-of-course project lab, click the following linkand select *Open Lab*.

* [Course 2 Automatidata project lab](https://www.coursera.org/learn/get-started-with-python/ungradedLab/FrE1A/activity-course-2-automatidata-project-lab)

**Step 3: Access the dataset**



To complete the assignment, you’ll need to import the dataset and answer the questions in the provided Python notebook. For this project, you’ll be utilizing data records from yellow taxi trips in New York City during 2017.

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

Link to dataset: [2017\_Yellow\_Taxi\_Trip\_Data](https://docs.google.com/spreadsheets/d/1do1IEEN-J54ug0Gl0d8tMdv9SutdEhQEIAjIQ-6UC20/template/preview?resourcekey=0-1UPqaU_qmf8Pl-0Twd_b7Q#gid=0)

OR

If you don’t have a Google account, you can download the dataset directly from the following attachment

**[2017\_Yellow\_Taxi\_Trip\_Data](https://d3c33hcgiwev3.cloudfront.net/gvJoe37aS_6HZqnLUmiajw_592fe280dd804403b6e33fd9d9ffa9f1_2017_Yellow_Taxi_Trip_Data.csv?Expires=1685577600&Signature=RWC0mrNrM-qkhkkSkHmlBbXDP~Bd6cSRj6luYRzyErZWi5-cu33Vae5HS5k1VHFD2TCTn-nONo-y36QcMwTdLpzDETLmayFNYNZ-SG3~D3M3LmbrGJhU~jW-5ETuB4tKz1o2rIgDvk8moHa2iY7zuttqK36GWcU2~a2o07lssTk_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)**

[CSV File](https://d3c33hcgiwev3.cloudfront.net/gvJoe37aS_6HZqnLUmiajw_592fe280dd804403b6e33fd9d9ffa9f1_2017_Yellow_Taxi_Trip_Data.csv?Expires=1685577600&Signature=RWC0mrNrM-qkhkkSkHmlBbXDP~Bd6cSRj6luYRzyErZWi5-cu33Vae5HS5k1VHFD2TCTn-nONo-y36QcMwTdLpzDETLmayFNYNZ-SG3~D3M3LmbrGJhU~jW-5ETuB4tKz1o2rIgDvk8moHa2iY7zuttqK36GWcU2~a2o07lssTk_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

**Data Dictionary**



This project uses a dataset called 2017\_Yellow\_Taxi\_Trip\_Data.csv. It contains data gathered by the New York City Taxi & Limousine Commission. For each trip, there are many different data variables gathered.

The dataset contains: **408,294 rows** – each row represents a different trip **18 columns**

| **Column name** | **Description** |
| --- | --- |
| ID | Trip identification number |
| VendorID | A code indicating the TPEP provider that provided the record.  **1= Creative Mobile Technologies, LLC; 2= VeriFone Inc.** |
| tpep\_pickup\_datetime | The date and time when the meter was engaged. |
| tpep\_dropoff\_datetime | The date and time when the meter was disengaged. |
| Passenger\_count | The number of passengers in the vehicle.  This is a driver-entered value. |
| Trip\_distance | The elapsed trip distance in miles reported by the taximeter. |
| PULocationID | TLC Taxi Zone in which the taximeter was engaged |
| DOLocationID | TLC Taxi Zone in which the taximeter was disengaged |
| RateCodeID | The final rate code in effect at the end of the trip.  **1= Standard rate**  **2=JFK**  **3=Newark**  **4=Nassau or Westchester**  **5=Negotiated fare**  **6=Group ride** |
| Store\_and\_fwd\_flag | This flag indicates whether the trip record was held in vehicle memory before being sent to the vendor, aka “store and forward,”  because the vehicle did not have a connection to the server.  **Y= store and forward trip**  **N= not a store and forward trip** |
| Payment\_type | A numeric code signifying how the passenger paid for the trip.  **1= Credit card**  **2= Cash**  **3= No charge**  **4= Dispute**  **5= Unknown**  **6= Voided trip** |
| Fare\_amount | The time-and-distance fare calculated by the meter. |
| Extra | Miscellaneous extras and surcharges. Currently, this only includes the $0.50 and $1 rush hour and overnight charges. |
| MTA\_tax | $0.50 MTA tax that is automatically triggered based on the metered rate in use. |
| Improvement\_surcharge | $0.30 improvement surcharge assessed trips at the flag drop. The  improvement surcharge began being levied in 2015. |
| Tip\_amount | Tip amount – This field is automatically populated for credit card tips. Cash tips are not included. |
| Tolls\_amount | Total amount of all tolls paid in trip. |
| Total\_amount | The total amount charged to passengers. Does not include cash tips. |

**Step 4: Complete your PACE strategy document**



The **Course 2 PACE strategy document** includes questions that will help guide you through the Course 2 Automatidata project. Answer the questions in your PACE strategy document to prepare for using Python to inspect and organize your data.

As a reminder, the PACE strategy document is designed to help you complete the contents for each of the templates provided. You might navigate back and forth between the PACE strategy document and the Python notebook. Make sure your PACE strategy document is complete before preparing your executive summary.

**Step 5: Complete the guided lab**



Your Python notebook for this project includes a guided framework that will assist you with the required coding. Input the code and answer the questions in your Python notebook to inspect and organize your data. You’ll find helpful reminders for tasks like:

* Importing data
* Loading necessary packages
* Identifying relevant data structures and summarizing data
* Extracting information from columns
* Combining or modifying data structures to create meaningful variables

You will also discover questions in this Python notebook designed to help you gather the relevant information you’ll need to write an executive summary for your team.

Use your completed PACE strategy document and Python notebook to help you prepare your executive summary in the next step.

**Step 6: Prepare an executive summary**



Your executive summary will keep your teammates at Automatidata informed of your progress. The one-page format is designed to respect teammates and stakeholders who may not have time to read and understand an entire report.

First, select one of the executive summary design layouts from the provided template. Then, add the relevant information. Your executive summary should include the following:

* A summary of your tasks
* Information regarding the results of your data variable assessment
* Identify recommended next steps in order to build a predictive model

Complete your executive summary to effectively communicate your results to your teammates.

**Pro Tip: Save the templates**

Finally, be sure to save a blank copy of the templates you used to complete this activity. You can use them for further practice or in your professional projects. These templates will help you work through your thought processes and demonstrate your experience to potential employers.

**What to Include in Your Response**



Later, you will have the opportunity to self-assess your performance using the following criteria. Be sure to address the following elements in your completed activity.

**Course 2 PACE strategy document**:

* Answer the questions in the PACE strategy document.
* Identify the proper stages of PACE for each task.

**Course 2 Automatidata project lab:**

* Import, inspect, and organize data

**Course 2 executive summary:**

* A summary of your tasks
* Information regarding the results of your data variable assessment
* Identify recommended next steps in order to build a predictive model