In this activity, you will complete a project that showcases your ability to use Python to import, inspect, and organize data. You will also update team members through an executive summary, demonstrating your ability to organize and communicate key information.

For additional information on how to complete this activity, review the previous readings:

End-of-course portfolio project introduction↗ and

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

Be sure to complete this activity before moving on. The next course item will provide you with completed exemplars to compare to your own work. You will not be able to access the exemplars until you have completed this activity.

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 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 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

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↗

Executive summary templates ↗

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
DOCX File

📎 Activity Templates_ Executive summaries
PPTX File

⟩ 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 link and select *Open Lab*.

Course 2 Automatidata project 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.

⟩ 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.

Column name	Description
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 3: Complete your PACE strategy document

The Course 2 PACE strategy document includes questions that will help guide you through the Course 2 Automatidata workplace scenario 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 4: 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 criteria listed below. Be sure to address the following elements in your completed activity.

Course 2 PACE strategy document:

Answer the questions in the PACE strategy document

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