



Activity Overview

In this activity, you will demonstrate your ability to organize, present, and share data stories on Tableau Public. 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 3 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

Your team is still in the early stages of their project for the New York City Taxi and Limousine Commission (TLC). So far, you've completed a project proposal and used Python to inspect and organize the TLC dataset.

You check your inbox and notice a new message from Luana Rodriguez, the Senior Data Analyst at Automatidata. Luana is pleased with the work you have already completed and requests your assistance with some exploratory data analysis (EDA) and data visualization. You also notice a follow-up email from Udo Bankole, the Director of Data Analysis. Udo suggests including an executive summary of your analysis to share with teammates.

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

Email from Luana Rodriguez, Senior Data Analyst

Subject: New York City TLC EDA & Vizzes

From: "Luana Rodriguez" Luana@automatidata

Cc: "Deshawn Washington," Deshawn@automatidata; "Udo Bankole," Udo@automatidata

Hi there,

Thanks for the amazing work you've done so far.

We're ready to perform EDA on the taxi data from the New York City TLC to get a general understanding of what taxi ridership looks like. Has Deshawn told you what the management team expects when it comes to EDA? If not, think of it as a "show your work" kind of report. They will want to see a Python notebook showing the structuring and cleaning you did, as well as any matplotlib/seaborn visualizations you plotted to help us understand the data. I would suggest at the very least a box plot of the ride durations and some time series plots, like a breakdown by quarter or month? Whatever you think makes most sense.

Additionally, the management team has recently asked all EDA to include Tableau visualizations. We've found these to be particularly helpful in status reports to the client and board members. Make sure it is easy to understand to someone who isn't data savvy, and remember that the assistant director at the New York City TLC is a person with visual impairments. I understand you have some Tableau experience? Let me know if you need help with this.

By the way, I Cc'd our director, Udo, who is on the management team and will be reviewing/approving our analysis before the project manager reports it back to the client. @Udo, I just want to keep you informed on the progress!

Thanks!

Luana Rodriguez

Senior Data Analyst

Automatidata

Email from Udo Bankole, Director of Data Analysis

Subject: RE: New York City TLC EDA & Vizzes

From: "Udo Bankole," Udo@automatidata;

Cc: "Deshawn Washington," Deshawn@automatidata; "Luana Rodriguez" Luana@automatidata

Thank you, Luana!

Welcome to the team, so glad to have you.

Along with the Tableau visualization and notebook, it would be really helpful if you included an executive summary of your analysis attached via email.

I appreciate your help!

Udo Bankole

Director of Data Analysis

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 the following links and select *Use Template*.


Link to templates: [↗](#)


[Course 3 PACE strategy document](#) [↗](#)

[Executive summary templates](#) [↗](#)

OR

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

 [Activity Template Course 3 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 3 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 perform EDA and create data visualizations. Here are some helpful reminders for tasks:

Import unstructured data

Review and structure data

Filter out unneeded data

Clean the data

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.

> 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:

108,294 rows – each row represents a different trip

18 columns

Column name	Description
trip_id	Trip identification number
vendorID	A code indicating the TPEP provider that provided the record. 1= Creative Mobile Technologies, LLC; 2= VeriFone Inc.
pickup_datetime	The date and time when the meter was engaged.
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 3: Complete your PACE strategy document

The Course 3 PACE strategy document includes questions that will help guide you through the Course 3 Automatidata project. Answer the questions in your PACE strategy document to prepare for using Python for EDA and both Python and Tableau for data visualization.

As a reminder, the PACE strategy document is designed to help you complete the contents for each of the templates provided. You may 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: Open Tableau Public and visualize your data

While using [Tableau Public](#)[↗], you will need to create the following deliverable for stakeholders:
Data visualization in the form of a scatter plot

If you need additional help, review the [Course 3 Tableau follow-along guide: Automatidata project](#)[↗]

OR

If you don't have a Google account, you can download the guide directly from the following attachment:

 [Course 3 Tableau follow-along guide: Automatidata project](#)
[DOCX File](#)

Use Tableau Public to create a scatterplot for your executive summary.

> Step 5: 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 might 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 the results of your exploratory data analysis (EDA)

- A proposed solution for dealing with outliers in your data

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 3 PACE strategy document:

- Answer the questions in the PACE strategy document

Course 3 Automatidata project lab:

- Perform exploratory data analysis (EDA)

- Create data visualizations

Course 3 Tableau visualization:

- Create a scatterplot to enhance the visualization created with Python

Course 3 executive summary:

Provide a summary of the results of your exploratory data analysis (EDA)

Propose a solution for dealing with outliers in your data