Course 3 workplace scenarios

Automatidata

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:

The New York City TLC data is ready for exploratory data analysis (EDA) in Python. You will need to clean, join, validate, and create a visualization for the taxi commission data. The findings will be shared with internal stakeholders from different departments within Automatidata.

Course 3 tasks:

- Load data, explore, and extract the New York City TLC data with Python
- Use custom functions to organize the information within the New York City TLC dataset
- Build a dataframe for the New York City TLC project
- Create an executive summary for Automatidata for a general audience of internal professionals

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.*

Key Takeaways

In Course 3, Go Beyond the Numbers: Translate Data into Insights, you explored the process of exploratory data analysis (EDA). You learned to Identify the core steps, basic methods, and benefits of structuring and cleaning data. Additionally, you investigated raw data using Python, and created data visualizations using Tableau

Course 3 skills:

- Conduct exploratory data analysis
- Create data visualization with Tableau

- Expand knowledge of Python coding
- Share insights and ideas with stakeholders

Course 3 end-of-course project deliverables:

- Complete EDA with workplace scenario dataset using Python
- Executive summary including a Tableau visualization

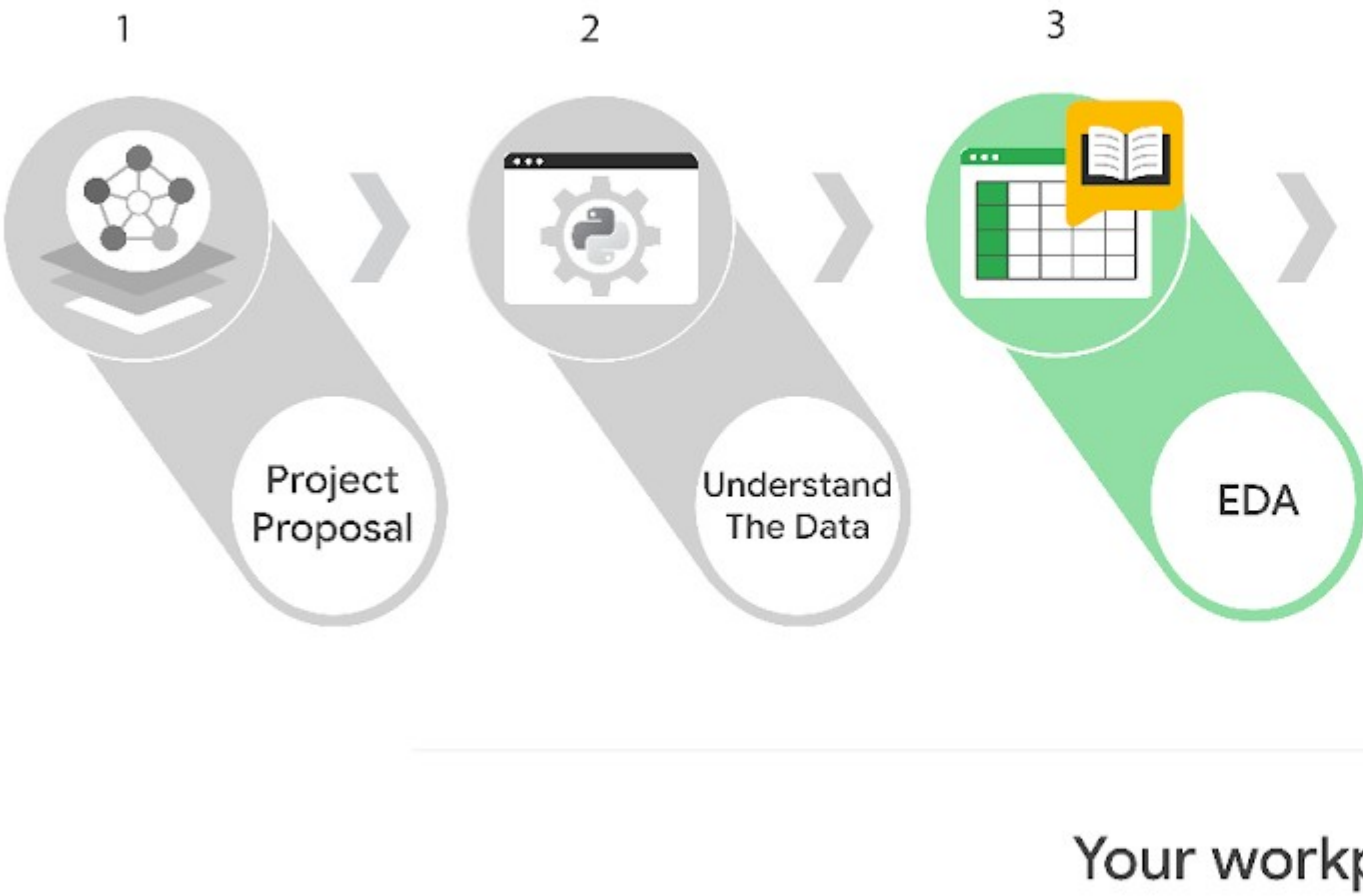
The end-of-course portfolio projects are designed for you to apply your data analytical skills within a workplace scenario. No matter which scenario you work with, you will practice your ability to discuss data analytic topics with coworkers, internal team members, and external clients.

As a reminder, you are required to complete one project for each course. To gain additional practice, or to add more samples to your portfolio, you may complete as many of the scenarios as you wish.

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

Learn about the Course 3 Automatidata workplace scenario!

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.

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.*

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.

The members of 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

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.*

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.

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

Good luck in your role! Automatidata looks forward to seeing how you communicate your creative work and approach problem-solving!

Key takeaways

The 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.

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:

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**

Column name	Description
	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

-

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