Activity Overview

In this activity, you will showcase your ability to use Python to build a regression model. You will also update team members and stakeholders 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 5 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

The Automatidata team is more than halfway through their project for the New York City Taxi & Limousine Commission (TLC). Earlier, the team completed a project proposal, used Python to explore the data, create data visualizations and conducted statistical testing. Now, the New York City TLC wants your team to build a regression model for ride fares based on a variety of variables.

In your inbox you discover an email from Titus Nelson, the Operations Manager at the New York City TLC asking for details about regression modeling. You also notice two follow-up emails from your manager, Deshawn Washington. 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 Titus Nelson, Operations Manager (NYC TLC)

Subject: Details on Regression Model

From: "Titus Nelson," Titus.Nelson@tlc.nyc

Cc: "Udo Bankole," Udo@automatidata; "Uli King" Uli@automatidata; "Deshawn Washington," Deshawn@automatidata; "Luana Rodriguez" Luana@automatidata;

Hello Automatidata team,

I really appreciate your work, and thanks for the explanation of the next phase of the algorithm creation.

I was hoping to get a bit more detail on regression. Will you be applying a linear regression or a multiple regression model? It wasn't clear in the meeting, and I wanted to be sure our teams are aligned on expectations.

Thank you,

Titus Nelson

Operations Manager

IT Division

NYC TLC

Learn more about TLC's accessible vehicle initiatives ↗.

Email from Deshawn Washington, Data Analysis Manager (Automatidata)

Subject: RE: Details on Regression Phase

From: "Deshawn Washington," Deshawn@automatidata

Cc: "Udo Bankole," Udo@automatidata; "Uli King" Uli@automatidata; "Luana Rodriguez" Luana@automatidata; "Titus Nelson," Titus.Nelson@tlc.nyc

Thank you for your email, Titus.

To answer your question, we will create and run a multiple linear regression (MLR) model to get the most accurate prediction because we want to predict ride fares based on multiple variables.

Our team will be working on getting you the results of the MLR model this week.

Feel free to reach out with additional questions.

Many thanks,

Deshawn Washington

Data Analysis Manager

Automatidata

Email from Deshawn Washington, Data Analysis Manager (Automatidata)

Subject: RE: Details on Regression Phase

From: "Deshawn Washington," Deshawn@automatidata

Cc: "Luana Rodriguez" Luana@automatidata;

Hello my Data team!

Would you two mind completing the following:

MLR model in a Python notebook

Draft an executive summary of your results

I'd appreciate a chance to review it before you send it over to Uli, but write the summary as if you're addressing the client.

Best regards,

Deshawn Washington

Data Analysis Manager

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 5 PACE strategy document ⤴

Executive summary templates ⤴

OR

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

📎 Activity Templates_ Executive summaries
PPTX File

📎 Activity Template_ Course 5 PACE strategy document
DOCX 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 5 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 build a regression model. You'll find helpful reminders for tasks like:

Model building and evaluation

Checking model assumptions

Interpreting model results

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 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

Column name	Description
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 5 PACE strategy document includes questions that will help guide you through the Course 5 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 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: Prepare an executive summary

Your executive summary will keep your Automatidata teammates and the stakeholders at the New York City TLC 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 variables analyzed in your regression model

The results of your analysis

Recommendations or insights based on your results

Complete your executive summary to effectively communicate your results to external stakeholders.

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 listed. Be sure to address the following elements in your completed activity.

Course 5 PACE strategy document:

Answer the questions in the PACE strategy document

Course 5 Automatidata project lab:

Build a regression model

Course 5 executive summary:

Include regression assumptions

Identify the outcome and impact of your work for this data project