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

Project background

Automatidata is near the end of the TLC project. The following tasks are needed at this stage of the project:

- Determine the correct modeling approach
- Build a regression model
- Finish checking model assumptions
- Evaluate the model
- Interpret model results and summarize findings for stakeholders within TLC

Your assignment

You will create a regression model. Determine the type of regression model that is needed and develop one using the TLC data.

Specific project deliverables

In this end-of-course project, you will gain valuable practice of your new skills as you complete the following deliverables:

- Complete a 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
- Build a regression model in Python
- Report the results in an executive summary

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