Scenario

You are the newest member of Automatidata's data analytics team. Your team is close to completing their project for the New York City Taxi & Limousine Commission (TLC). Previously, you completed a project proposal and used Python to explore and analyze the TLC dataset, create data visualizations, and conduct an A/B test. Most recently, you built an MLR model for fare amounts based on a variety of variables.

The New York City TLC is impressed with your work so far. Now, they want your team to identify which variables or factors influence the amount of gratuity a rider gives a driver. Your work will help TLC stakeholders make informed business decisions that will increase gratuities and subsequently improve driver satisfaction.

At a meeting with New York City TLC stakeholders, your team suggests building a random forest model to predict whether or not a rider will be a generous tipper (>= 20%). At the end of the meeting, Titus Nelson, the Operations Manager at the New York City TLC, says that he will share the suggestion with his organization's leadership team.

A few days after the meeting, you receive an email from Juliana Soto, a Department Head at the New York City TLC. Juliana says that TLC leadership likes the idea of using a random forest model to predict gratuity and asks the team to share more details about the model. You also receive a follow-up email from Udo Bankole, the Director of Data Analysis at Automatidata. Udo asks you to build the random forest model and to prepare an executive summary to share your results.

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

Email from Juliana Soto, Finance and Administration Department Head (NYC TLC)

Subject: NYC TLC Approval of Algorithm

From: "Juliana Soto," Juliana.Soto@tlc.nyc

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

Hello Automatidata Team,

Thank you for providing the details for the final phase of the prediction algorithm we have requested. I apologize for missing many of the weekly project meetings, but Titus has kept me

informed of your progress. We discussed in detail your proposal for using a random forest model for prediction, and we are in agreement with you.

If you would, please commence with the creation of the algorithm. It would be very helpful to provide a summary of what data indicators the algorithm is basing its results on and an idea of the confidence your team has in the accuracy of the result.

Thank you for your great work,

–

Juliana Soto

Finance and Administration Department Head

New York City Taxi & Limousine Commission

Learn more about
TLC's accessible vehicle initiatives
.

Email from Udo Bankole, Director of Data Analysis (Automatidata)

Subject: RE: NYC TLC Approval of Algorithm

From: "Udo Bankole," Udo@automatidata

Cc: "Luana Rodriguez" Luana@automatidata;

Hello data pros!

You have done great work so far. We are excited to find out what else you can discover in the data and for you to help us make data-driven business decisions.

If you would please build the random forest model we discussed using the data New York City TLC has provided. As you're aware, you have already cleaned and run this data through a MLR model, but you always need to validate your variables and data. So please revisit the dataset.

Once complete, please send an executive summary to Deshawn and myself of what wording you plan to send to the client. Be sure to include what Juliana requested, a summary of the variables used, and an indication of how we can test the accuracy of the model.

I look forward to exploring your build!

Udo Bankole

Director of Data Analysis

Automatidata

Project background

Automatidata is ready to create a machine learning model for TLC. The following tasks are needed to complete the project:

- Model building
- Model evaluation
- Summarize findings for Automatidata and the stakeholders at TLC

Your assignment

You will create a machine learning model for the TLC data. You will be responsible for leading these tasks, which include feature engineering, model development, and evaluation.

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 planner mapping questions, details, and action items for each stage of the project scenario
- Answer the questions in the Jupyter notebook project file
- Design and implement a machine learning model
- Draft an executive summary of your results

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