

**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:**

Exploratory data analysis is complete for the project. The New York City TLC would like the data team at Automatidata to analyze the relationship between fare amounts and payment type. The team agrees that the next step is to perform a hypothesis test using the data.

**Course 4 tasks:**

* Compute descriptive statistics
* Conduct a hypothesis test using the New York City TLC dataset
* Create an executive summary for the Automatidata data team before sharing the results with the client

***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.*

**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 at this stage of the project:

* Explore the project data
* Implement a hypothesis test
* Communicate insights with team members and TLC stakeholders

**Your assignment**

You will conduct hypothesis testing on the data for the TLC data. You’ve been asked to investigate TLC’s dataset to determine which hypothesis testing method best serves the data and the TLC project.

**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. The data shared in this project has been altered 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:

* 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
* Statistical testing
* Report results in executive summary

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

**Scenario**



Your team at Automatidata is nearing the midpoint of their project for the New York City Taxi & Limousine Commission (TLC). So far, you’ve completed a project proposal and used Python to explore and analyze the TLC dataset. You’ve also used both Python and Tableau to create data visualizations. The next step is to use statistical methods to analyze and interpret your data.

You receive a new email from Uli King, Automatidata’s project manager. Uli tells your team about a new request from the New York City TLC: to analyze the relationship between fare amount and payment type. You also discover follow-up emails from three other team members: Deshawn Washington, Luana Rodriguez, and Udo Bankole. These emails discuss the details of the analysis. A final email from Luana includes your specific assignment: to conduct an A/B test.

***Notes on the fictional nature of this project and data assumptions:***

*Please note the following considerations when preparing your project. When making data-driven inferences in your professional lives, you will need to perform comprehensive Exploratory Data Analysis and cross-check your own data sources and self-made assumptions. As outlined in the following notes, there is often a gap between theory and practice.*

* The team member names used in this workplace scenario are fictional and are not representative of the New York City TLC.
* The following scenario asks you to conduct an A/B test. An A/B test can only be performed in an experiment with randomly selected groups. In this scenario, this project makes the claim that (fictitiously) randomly grouped riders were asked to pay with a certain payment type in order to make data-driven inferences.
* All riders are assumed to be able to pay with cash or card (in practice, riders might not carry cash and have to pay with card, or vice versa).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Email from Uli King, Senior Project Manager**

**Subject:** New TLC Request -  Taxi Tips Data

**From:** “Uli King” Uli@automatidata

**Cc:** “Deshawn Washington,” Deshawn@automatidata; “Udo Bankole,” Udo@automatidata; “Luana Rodriguez” Luana@automatidata

Hello Data Team!

Really excellent work so far. Everyone over at New York City TLC is impressed with the results–especially the analysis on the last report! Thanks so much for the hard work.

On that note, they have requested an additional item to be added to the initial project scope. They would like a detailed statistical analysis of payment type. That is, do the customers who use a credit card pay higher fare amounts than those who use cash?

That said, the New York City TLC team is asking us to consider the following:

* The relationship between fare amount and payment type.
* Test the hypothesis that customers who use a credit card pay higher fare amounts.
* Should you conclude that there is a statistically significant relationship between credit card payment and fare amount, discuss what the next steps should be: what are your thoughts on strategies our team could implement to encourage customers to pay with credit card?

Many thanks!

Uli King

Senior Project Manager

Automatidata

**Email from Deshawn Washington, Data Analysis Manager**

**Subject:** RE: New TLC Request -  Taxi Tips Data

**From:** “Deshawn Washington,” Deshawn@automatidata

**Cc:**; “Udo Bankole,” Udo@automatidata; “Luana Rodriguez” Luana@automatidata; “Uli King” Uli@automatidata

Thanks, Uli.

It’s great to hear the client is happy. I’m reminded again what a great data team we have!

If you would, please tell the client we will be providing them with this analysis in two weeks’ time.

@Luana, my initial thought is for us to conduct an A/B test to analyze the relationship between fare amount and payment type. What do you think?

Thanks,

Deshawn Washington

Data Analysis Manager

Automatidata

**Email from Luana Rodriguez, Senior Data Analyst**

**Subject:** RE: New TLC Request -  Taxi Tips Data

**From:** “Luana Rodriguez” Luana@automatidata;

**Cc:** “Udo Bankole,” Udo@automatidata; “Uli King” Uli@automatidata; “Deshawn Washington,” Deshawn@automatidata

Hi all,

@Deshawn, I agree with you on the A/B testing. We’ll share a summary of the results with Uli before he presents it to the client.

We’ll get started right away.

Thank you,

Luana Rodriguez

Senior Data Analyst

Automatidata

**Email from Udo Bankole, Senior Data Analyst**

**Subject:** RE: New TLC Request -  Taxi Tips Data

**From:** “Udo Bankole,” Udo@automatidata;

**Cc**: “Uli King” Uli@automatidata; “Deshawn Washington,” Deshawn@automatidata; “Luana Rodriguez” Luana@automatidata;

I support the path forward. Thank you all.

Udo Bankole

Senior Data Analyst

Automatidata

**Email from Luana Rodriguez, Senior Data Analyst**

**Subject:** RE: New TLC Request -  Taxi Tips Data

**From:** “Luana Rodriguez” Luana@automatidata;

**Cc:**

Hi there, fellow data virtuoso!

You’ve been handling all of this work really well, by the way. Excellent job.

I was wondering if you’d like to try the A/B test on the TLC data yourself? Based on what you’ve shared with me, I have every confidence you already have all the skills and experience needed for this task.

What do you think? Would you like to give it a go?

Also, like I said in my email to Deshawn, you’ll need to write an executive summary of the results so we can present it to Udo before he shares it with the client.

Thanks so much!

Luana Rodriguez

Senior Data Analyst

Automatidata

–

*“You can have data without information, but you cannot have information without data.”*

*—-Daniel Keys Moran*