Course 4 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:

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.*

Key Takeaways

In Course 4, The Power of Statistics, you explored fundamental concepts such as descriptive and inferential statistics, probability, sampling, confidence intervals, and hypothesis testing. Additionally, you learned the fundamentals, methods, and benefits of structuring and cleaning data and how to apply statistical methods using Python.

Course 4 skills:

- Conduct statistical analysis
- Use probability distributions
- Compute descriptive statistics
- Conduct and interpret statistical analyses using Python
- Perform a hypothesis test to identify insights about data
- Share insights and ideas with stakeholders

Course 4 end-of-course project deliverables:

- Hypothesis test prepared with Python
- Executive summary

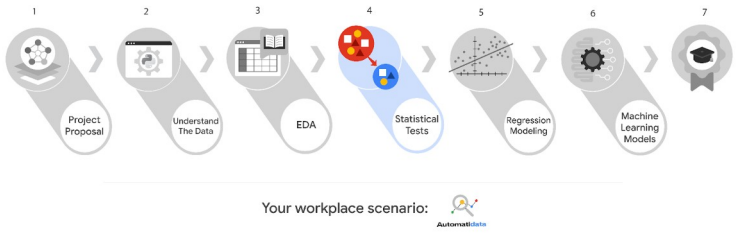
The end-of-course portfolio projects are designed for you to apply your data analytical skills within a workplace scenario. No matter which scenario you work with, you will practice your ability to discuss data analytic topics with coworkers, internal team members, and external clients.

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 4 end-of-course portfolio project overview: Automatidata

Learn about the Course 4 Automatidata workplace scenario!

The end-of-course project in Course 4 focuses on your ability to conduct statistical testing 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 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!

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: Create your Course 4 Automatidata project

Review Learning Objectives

Ready to review what you've learned before taking the quiz? I'm here to help.

Help me practiceLet's chat

Submit your assignment

Resume assignment

Receive grade

To Pass100% or higher

Your grade

-Not available

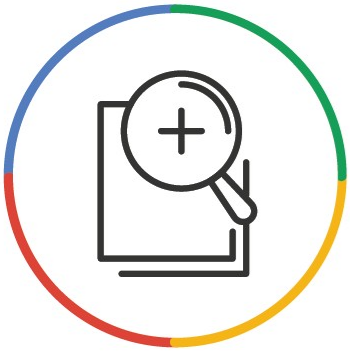
Like
Dislike
Report an issue

Back

Activity: Create your Course 4 Automatidata project

Practice Quiz. • 30 min. • 1 total point available.1 total point

English
DueApr 15, 11:59 PM JST
To pass this practice quiz, you must receive 100%, or 1 out of 1 point, by completing the following activity. You can learn more about the graded and practice items in the course overview.



Activity Overview

In this activity, you will showcase your ability to use statistical methods to analyze and interpret data. In particular, you will use descriptive statistics and hypothesis testing to conduct an A/B test. 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 4 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 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

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*.

Links to templates:_

- Course 4 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 4 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 4 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 run an A/B test using statistical methods. You'll find helpful reminders for tasks like:

- Computing descriptive statistics
- Conducting a hypothesis test (t-test)
    - **Reminder:** An A/B test is a two-sample t-test. You have all the information you need to complete an A/B test by following the same protocols in a two-sample t-test.

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**

Column name	Description
	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 4 PACE strategy document** includes questions that will help guide you through the Course 4 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 teammates at Automatidata 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 statistical methods involved in your A/B test
- The results of your A/B test
- Recommendations or insights based on your results

Complete your executive summary to effectively communicate your results to your teammates.

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

Course 4 PACE strategy document:

- Answer the questions in the PACE strategy document

Course 4 Automatidata project lab:

- Compute descriptive statistics

- Conduct a hypothesis test

Course 4 executive summary:

- State the A/B test results clearly
- List recommended next steps for the data project