

## Course 4 workplace scenarios

TikTok



### Project goal:

The TikTok data team is developing a machine learning model for classifying claims made in videos submitted to the platform.

### Background:

TikTok is the leading destination for short-form mobile video. The platform is built to help imaginations thrive. TikTok's mission is to create a place for inclusive, joyful, and authentic content—where people can safely discover, create, and connect.

### Scenario:

The TikTok data team has successfully completed exploratory data analysis on the data for the claims classification project. The team is ready to begin the process of hypothesis testing. You've been asked to investigate TikTok's user claim dataset to determine which hypothesis testing method best serves the data and the claims classification project.

### Course 4 tasks:

- Import relevant packages and TikTok data
- Explore the project data
- Implement a hypothesis test
- Communicate insights with stakeholders within TikTok

**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

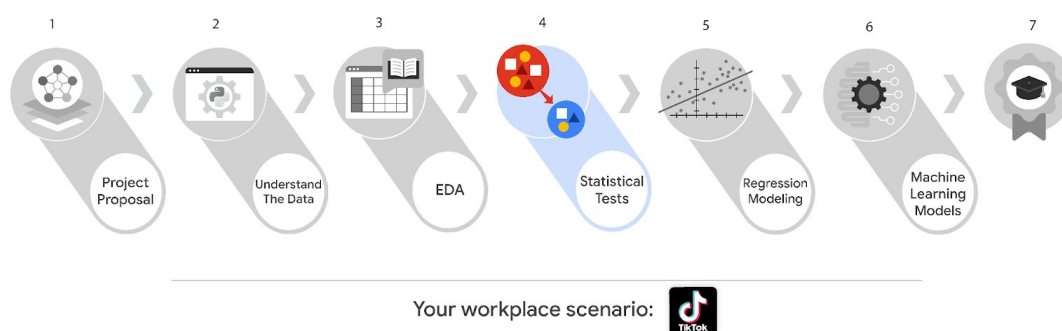
The end-of-course portfolio projects are designed for you to apply your data analytical skills within a workplace scenario. No matter which scenario you work with, you will practice your ability to discuss data analytic topics with coworkers, internal team members, and external clients.

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: TikTok

## Learn about the Course 4 TikTok workplace scenario!

The end-of-course project in Course 3 focuses on your ability to use exploratory data analysis to organize and understand the data within a project. As a reminder, in Course 1 you developed a project proposal that outlined milestones, which progress with each of the end-of-course projects. A visual representation is provided in the graphic shown here:



Learn more about the project, your role, and expectations in this reading.

## Background on the TikTok scenario

At TikTok, our mission is to inspire creativity and bring joy. Our employees lead with curiosity and move at the speed of culture. Combined with our company's flat structure, you'll be given dynamic opportunities to make a real impact on a rapidly expanding company, and grow your career.

TikTok users have the ability to submit reports that identify videos and comments that contain user claims. These reports identify content that needs to be reviewed by moderators. The process generates a large number of user reports that are challenging to consider in a timely manner.

TikTok is working on the development of a predictive model that can determine whether a video contains a claim or offers an opinion. With a successful prediction model, TikTok can reduce the backlog of user reports and prioritize them more efficiently.

## **Project background**

TikTok's data team is working on the claims classification project. The following tasks are needed at this stage of the project:

- Explore the project data
- Implement a hypothesis test
- Communicate insights with stakeholders within TikTok

## **Your assignment**

You will conduct hypothesis testing on the data for the claims classification data. You've been asked to investigate TikTok's user claim dataset to determine which hypothesis testing method best serves the data and the claims classification project.

## **Team members at TikTok**

### **Data team roles**

- Willow Jaffey- Data Science Lead
- Rosie Mae Bradshaw- Data Science Manager
- Orion Rainier- Data Scientist

The members of the data team at TikTok are well versed in data analysis and data science. Messages to these more technical coworkers should be concise and specific.

### **Cross-functional team members**

- Mary Joanna Rodgers- Project Management Officer
- Margery Adebawale- Finance Lead, Americas
- Maika Abadi- Operations Lead

Your TikTok team includes several managers, who oversee operations. It is important to adjust your general correspondence appropriately to their roles, given that their responsibilities are less technical in nature.

**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.

## Specific project deliverables

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

- Course 4 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
- Consider the different groups of data represented in the dataset
- Implement a hypothesis test
- Create an executive summary to share your results

## Key takeaways

The Google Advanced Data Analytics Certificate 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 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. 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 project introduction](#) and [Course 4 end-of-course portfolio project overview: TikTok](#).

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 TikTok data analytics team has completed the first three milestones of the claims classification project and is nearing the halfway point. So far, the team has completed a project proposal, and used Python to perform exploratory data analysis on the dataset for the claims classification project. The team also produced data visualizations in both Python and Tableau to share with stakeholders. The next step is to use statistical methods to analyze and interpret the claims classification data.

You receive a new email from Mary Joanna Rodgers, one of TikTok's project management officers. Mary Joanna informs the data team about a new request: to determine whether there is a statistically significant difference in the number of views for TikTok videos posted by verified accounts versus unverified accounts. You also receive follow-up emails from Data Science Manager, Rosie Mae

Bradshaw and Data Science Lead, Willow Jaffey. These emails share the details of the analysis. A final email from Data Scientist, Orion Rainier, details your next assignment: to conduct a hypothesis test on verified versus unverified accounts in terms of video view count.

**Note:** Team member names used in this workplace scenario are fictional and are not representative of TikTok.

---

**Email from Mary Joanna Rodgers, Project Management Officer**

**Subject:** New Request - Hypothesis Test: Verified/Unverified Accounts

**From:** “Rodgers, Mary Joanna” —maryjoannarodgers@tiktok

**Cc:** “Rainier, Orion”—orionrainier@tiktok; “Jaffey, Willow”—willowjaffey@tiktok

; “Bradshaw, Rosie Mae”—rosiemaebradshaw@tiktok

Hello Data Team!

Really excellent work so far. The leadership team is impressed with the results—especially the progress and insights shared on the last executive summary 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. We are interested in whether there is a statistical difference in the data between verified and unverified accounts. Do you have any indication which variable would be most insightful to test in terms of verified and unverified accounts?

Many thanks!

Mary Joanna Rodgers

Project Management Officer

TikTok

*Network with TikTok employees from a variety of teams and locations. Participate in **TikTok Tuesdays**, every Tuesday @2pm EST.*

**Email from Rosie Mae Bradshaw TikTok’s Data Science Manager**

**Subject:** RE: New Request - Hypothesis Test: Verified/Unverified Accounts

**From:** “Bradshaw, Rosie Mae” —rosiemaebradshaw@TikTok

**Cc:**; “Jaffey, Willow”—willowjaffey@tiktok; “Rainier, Orion”—orionrainier@tiktok; “Rodgers, Mary Joanna”—maryjoannarodgers@tiktok

Thanks for the update, Mary Joanna.

It’s great to hear that the leadership team is pleased with the data team’s progress and the early insights we have been able to deliver. I never grow tired of being reminded of what a great data team we have assembled here at TikTok!

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

@Orion, my initial thought is for us to conduct a hypothesis test to analyze whether there is a significant difference in video views for verified versus unverified accounts. What do you think?

In summary, I think we should do the following:

- Compute descriptive statistics on the claims classification data
- Conduct a two-sample hypothesis test of verified versus unverified accounts in terms of video view counts

Thanks,

Rosie Mae Bradshaw

Data Analysis Manager

TikTok

[Learn about TikTok's Trust & Safety team](#)

**Email from Orion Rainier, Data Scientist**

**Subject:** RE: New Request - Hypothesis Test: Verified/Unverified Accounts

**From:** "Rainier, Orion"—orionrainier@tiktok

**Cc:** "Jaffey, Willow"—willowjaffey@tiktok; "Rodgers, Mary Joanna"—maryjoannarodgers@tiktok ; "Bradshaw, Rosie Mae"—rosiemaebadshaw@tiktok

Hi all,

@Rosie Mae, I agree with you on statistical testing. We'll share a summary of the results before we present it to the client.

We'll get started right away.

Thank you,

Orion Rainier

Data Scientist

TikTok

—

*"Big data isn't about bits, it's about talent." — Douglas Merrill*

**Email from Willow Jaffey, Data Science Lead**

**Subject:** RE: New Request - Hypothesis Test: Verified/Unverified Accounts

**From:** "Jaffey, Willow"—willowjaffey@tiktok

**Cc:** "Rodgers, Mary Joanna"—maryjoannarodgers@tiktok; "Bradshaw, Rosie Mae"—rosiemaebadshaw@tiktok; "Rainier, Orion"—orionrainier@tiktok

I agree with everyone's assessments on this project so far. I look forward to the team's progress in this milestone. Thank you all.

Willow Jaffey

Data Science Lead

TikTok

**Email from Orion Rainier, Data Scientist**

**Subject:** RE: New Request - Hypothesis Test: Verified/Unverified Accounts

**From:** “Rainier, Orion”—orionrainier@tiktok

**Cc:**

Hi there, fellow data professional!

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 statistical testing 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 try?

Also, like I said in my email to Rosie Mae, you’ll need to write an executive summary of the results so we can present it to Willow before sharing it with the client.

Thanks so much!

Orion Rainier

Data Scientist

TikTok

—

*“Big data isn’t about bits, it’s about talent.” — Douglas Merrill*

## 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](#)
- [Course 4 Executive summary](#)

OR

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

[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 TikTok 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 a statistical test. You'll find helpful reminders for tasks like:

- Computing descriptive statistics
- Conducting a hypothesis 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.

## Data Dictionary



This project uses a dataset called `tiktok_dataset.csv`. It contains synthetic data created for this project in partnership with TikTok. Examine each data variable gathered.

**19,383 rows** – Each row represents a different published TikTok video in which a claim/opinion has been made.

### 12 columns

Column name	Type	Description
#	int	TikTok assigned number for video with claim/opinion.
claim_status	obj	Whether the published video has been identified as an “opinion” or a “claim.” In this dataset, an “opinion” refers to an individual’s or group’s personal beliefs or thoughts. A “claim” refers to information that is either unsourced or from an unverified source.



Column name	Type	Description
video_id	int	Random identifying number assigned to a video upon publication on TikTok.
video_duration_sec	int	How long the published video is measured in seconds.
video_transcription_text	obj	Transcribed text of the words spoken in the published video.
verified_status	obj	Indicates the status of the TikTok user who published the video in terms of their verification, either “verified” or “not verified.”
author_ban_status	obj	Indicates the status of the TikTok user who published the video in terms of their permissions: “active,” “under scrutiny,” or “banned.”
video_view_count	float	The total number of times the published video has been viewed.
video_like_count	float	The total number of times the published video has been liked by other users.
video_share_count	float	The total number of times the published video has been shared by other users.
video_download_count	float	The total number of times the published video has been downloaded by other users.
video_comment_count	float	The total number of comments on the published video.

### Step 3: Complete your PACE strategy document



The **Course 4 PACE strategy document** includes questions that will help guide you through the Course 4 TikTok workplace scenario 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 TikTok informed of your progress. The one-page format is designed to respect teammates and stakeholders who may 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 testing
- The results of your statistical testing
- 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 TikTok project lab:**

- Compute descriptive statistics
- Conduct a hypothesis test

### **Course 4 executive summary:**

- State the statistical test results clearly
- Identify recommended next steps in order to build a predictive model