



Activity Overview

In this activity, you will complete a project that showcases your ability to use Python to import, inspect, and organize data. 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 2 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 team's latest project is in its early stages of developing a machine learning model to classify claims in videos.

Previously, you were asked to complete a project proposal by your supervisor, Rosie Mae Bradshaw. You have received notice that the project proposal submitted by the team has been approved and your team has been given access to TikTok's user data. To get clear insights, the data must be inspected, organized, and prepared for analysis.

You discover two new emails in your inbox: one from your supervisor, Rosie Mae Bradshaw, and one from Willow Jaffey, the data team's Data Science Lead. 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 TikTok.

Email from Rosie Mae Bradshaw, Data Science Manager

Subject: Help with coding notebook?

From: "Bradshaw, Rosie Mae" —rosiemaebradshaw@tiktok

Cc: "Rainier, Orion"—orionrainier@tiktok

Good morning,

I have a couple of updates on our latest project. The leadership team has approved the project proposal that we completed previously. Thanks for all of your great work so far. Additionally, I just received an email from our Project Management Officer, Mary Joanna Rodgers that the data team is clear to proceed.

Before we begin the process of Exploratory Data Analysis (EDA), we could really use your help with coding and prepping the data. During your interview you mentioned that you worked with Python specifically in the Google certificate program you completed. That experience sounds applicable here.

Orion Rainier (Cc'd above) started a Jupyter notebook with the relevant dataset (attached). Orion is very involved in the final stages of another project. I'm sure your assistance in completing the coding and setting up the notebook for the project would be greatly appreciated.

Orion, do you mind sharing the details?

Humblest regards,

Rosie Mae Bradshaw

Data Science Manager

TikTok

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

Email from Orion Rainier, Data Scientist

Subject: RE: Help with coding notebook?

From: "Rainier, Orion"—orionrainier@tiktok

Cc: "Bradshaw, Rosie Mae"—rosiemaebradshaw@tiktok

Nice to meet you (virtually)!

Hope you have enjoyed your first few weeks!

With the project proposal approved, we are ready to begin the process of preparing the claim classification data. The goal of this project is to ultimately build a machine learning model that can streamline the claims process by identifying whether statements made in videos are claims or opinions.

A claim refers to information that is either unsourced or from an unverified source. For example, "The news reported that someone revealed that around 50% of the mined gold on Earth comes from one source."

Opinions refer to the personal beliefs or thoughts of a group or an individual. Here's an example, "In my opinion the most productive work day of the week is Tuesday."

There are a number of data team members committed to adjusting the machine learning developed for the last project, so your help is greatly appreciated!

Until we finish the prior project, there is no need to do a full EDA on this data. We will get to that soon. Do you mind importing the data (attached) and reviewing it for the team? It would be fantastic if you could include a summary of the column Data types, data value nonnull counts, relevant and irrelevant columns, along with anything else code related you think is worth sharing/showing in the notebook? You'll need to select a couple of variables to focus on. Include their minimum and maximum values. I haven't looked closely at the data yet, but it would be really helpful if you can create meaningful variables by combining or modifying the structures given.

Thanks,

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 2 PACE strategy document](#)
- [Executive summary templates](#)

OR

If you do not have a Google account, you can download the templates directly from the following attachments:

[Activity Template](#) [Course 2 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 2 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 inspect and organize your data. You'll find helpful reminders for tasks like:

- Importing data
- Loading necessary packages
- Identifying relevant data structures and summarizing data
- Extracting information from columns
- Combining or modifying data structures to create meaningful variables

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.




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 belief or thought. A “claim” refers to information that is either unsourced or from an unverified source.
video_id	int	Random identifying number assigned to 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 2 PACE strategy document includes questions that will help guide you through the Course 2 TikTok workplace scenario project. Answer the questions in your PACE strategy document to prepare to use 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 your tasks
- Information regarding the results of your data variable assessment
- Identify recommended next steps in order to build a predictive model

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 2 PACE strategy document:

- Answer the questions in the PACE strategy document

Course 2 TikTok project lab:

- Import, inspect, and organize data

Course 2 executive summary:

- A summary of your tasks
- Information regarding the results of your data variable assessment
- Identify recommended next steps in order to build a predictive model

