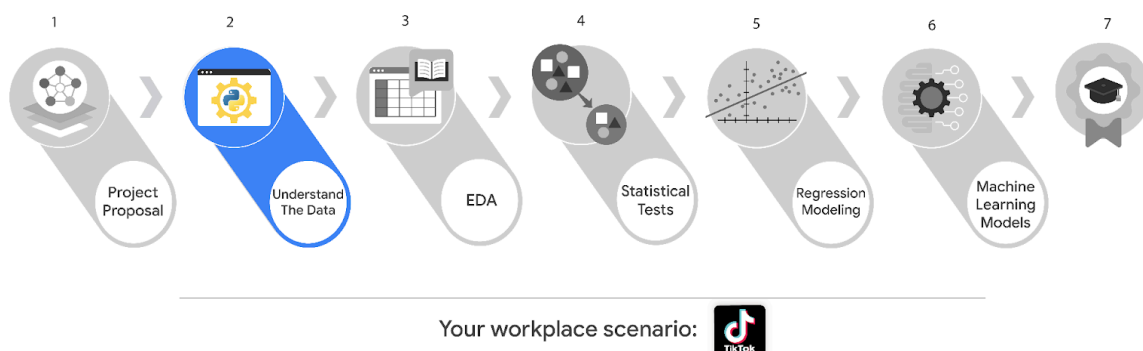


Learn about the Course 2 TikTok workplace scenario!

The end-of-course project in Course 2 focuses on your ability to understand the data needed for 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 in the earliest stages of the claims classification project. The following tasks are needed before the team can begin the data analysis process:

- Build a dataframe for the TikTok dataset
- Examine data type of each column
- Gather descriptive statistics

Your assignment

You will build a dataframe for the claims classification data. After the dataframe is complete, you will organize the claims data for the process of exploratory data analysis, and update the team on your progress and insights.

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 2 PACE Strategy Document to plan your project while considering your audience members, teammates, key milestones, and overall project goal.

- Answer the questions in the Jupyter notebook project file
- Complete coding prep work on project's Jupyter notebook
- Summarize the column Dtypes
- Communicate important findings in the form of an executive summary

TikTok's data team needs you to problem-solve and communicate your findings.
Good luck on your tasks!

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.

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