Course 3 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:

It is now time to begin the process of exploratory data analysis (EDA). As a data analyst on TikTok's data team, you will complete the EDA process for the claims classification project. You'll also use Tableau to create visuals for an executive summary to help non-technical stakeholders engage and interact with the data.

Course 3 tasks:

- Imports of relevant packages and TikTok data into Python
- EDA and cleaning
- Assess Tableau measures and dimensions
- Select and build visualization(s) type
    - Create plots to visualize variables and relationships between variables
- Share your results with the TikTok team

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 3, Go Beyond the Numbers: Translate Data into Insights, you explored the process of exploratory data analysis (EDA). You learned to Identify the core steps, basic methods, and benefits of structuring and cleaning data. Additionally, you investigated raw data using Python, and created data visualizations using Tableau

Course 3 skills:

- Conduct exploratory data analysis
- Create data visualization with Tableau
- Expand knowledge of Python coding
- Share insights and ideas with stakeholders

Course 3 end-of-course project deliverables:

- Complete EDA with workplace scenario dataset using Python
- Executive summary including a Tableau visualization

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.

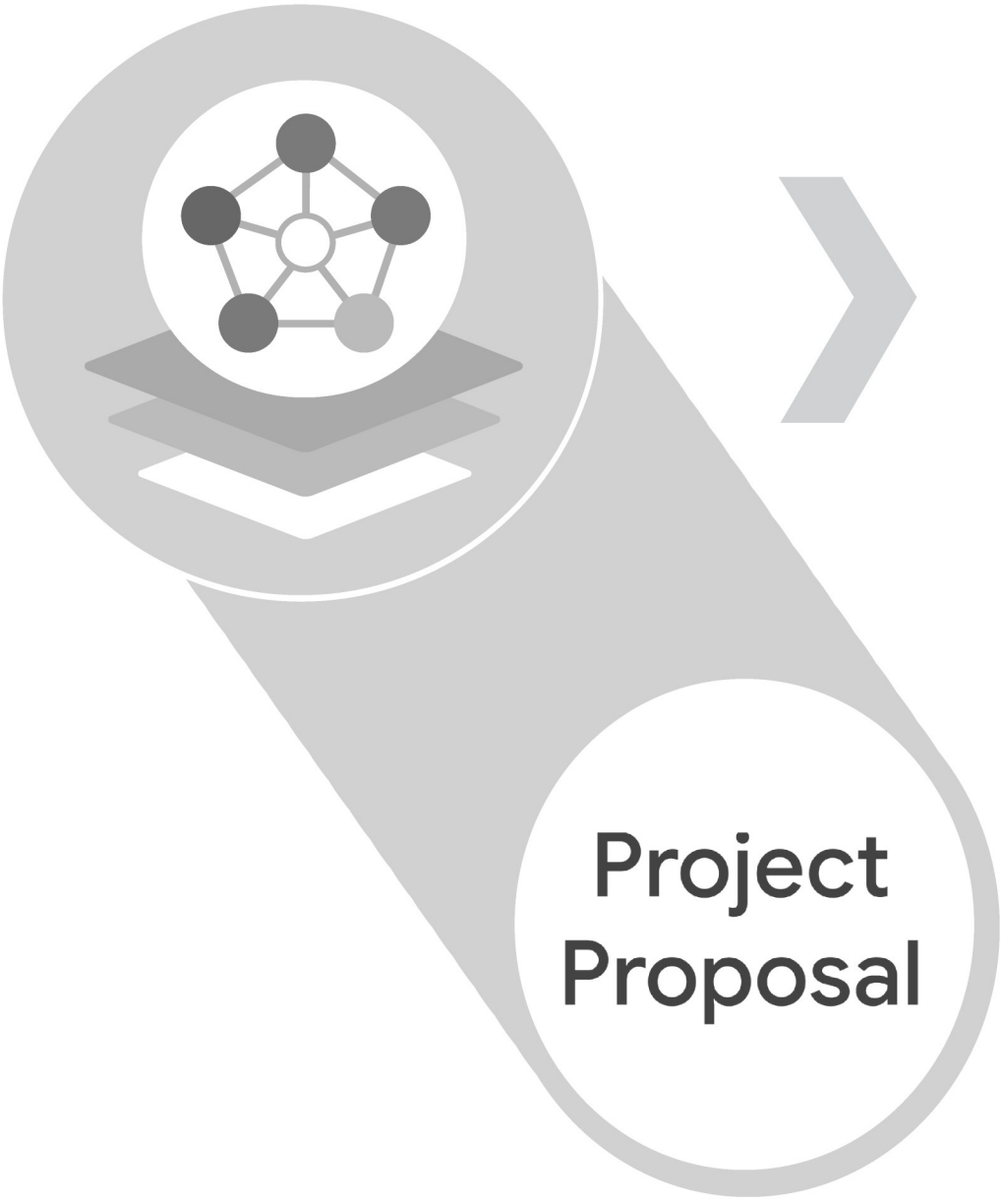
As a reminder, you are required to complete one project for each course. To gain additional practice, or to add more samples to your portfolio, you may complete as many of the scenarios as you wish.

Course 3 end-of-course portfolio project overview: TikTok

Learn about the Course 3 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:

1



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 before the team can begin the data analysis process:

- EDA and cleaning
- Select and build visualization(s) type
    - Create plots to visualize variables and relationships between variables
- Share your results with the TikTok team

Your assignment

You will conduct exploratory data analysis on data for the claims classification project. You'll also use Tableau to create visuals for an executive summary to help non-technical stakeholders engage and interact with the data.

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 Adebowale- 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 3 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
- Clean your data, perform exploratory data analysis (EDA)
- Create data visualizations
- 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 demonstrate your ability to organize, present, and share the stories within 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 3 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

Your team is still in the early stages of their latest project. So far, you've completed a project proposal and used Python to inspect and organize the TikTok dataset.

You check your inbox and notice a new message from Orion Rainier, a Data Scientist at TikTok. Orion is pleased with the work you have already completed and is requesting your assistance with some Exploratory Data Analysis (EDA) and data visualization. You also notice a follow-up email from the Data Science Lead, Willow Jaffey. Willow suggests including an executive summary of your analysis to share with teammates.

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

Email from Orion Rainier, Data Scientist

Subject: Tik Tok Claims Classification EDA & Vizzes

From: "Rainier, Orion"—orionrainier@tiktok

Cc: "Bradshaw, Rosie Mae" —rosiemaebradshaw@tiktok; "Jaffey, Willow" —willowjaffey@tiktok

Hi there,

Thanks for the amazing work you've done so far.

We're ready to perform EDA on the data. Has Rosie Mae told you what the management team expects when it comes to EDA? If not, think of it as a "show your work" kind of report. They will want to see a Python notebook showing the structuring and cleaning you did, as well as any matplotlib/seaborn visualizations you plotted to help us understand the data. I would suggest at the very least a graph comparing claim counts to opinion counts, as well as boxplots of the most important variables (like "video duration," "video like count," "video comment count," and "video view count") to check for outliers. Also a breakdown of "author ban status" counts. But whatever you think makes most sense works for us.

Additionally, the management team has recently asked all EDA to include Tableau visualizations. We've found these to be particularly helpful in status reports to the client and board members. For this data, I suggest a Tableau dashboard showing a simple claims versus opinions count, as well as stacked bar charts of claims versus opinions for variables like video view counts, video like counts, video share counts, and video download counts. Make sure it is easy to understand to someone who isn't data savvy, and remember that the assistant director is a person with visual impairments. I understand you have some Tableau experience? Let me know if you need help with this.

By the way, I CC'd our Data Science Lead, Willow Jaffey, who is on the senior management team and will be reviewing and approving our analysis before the project manager reports it back to the client. @Willow, I just want to keep you informed on the progress!

Thanks!

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: Tik Tok Claims Classification EDA & Vizzes

From: "Jaffey, Willow" —willowjaffey@tiktok

Cc: "Bradshaw, Rosie Mae" —rosiemaebradshaw@tiktok; "Rainier, Orion"—orionrainier@tiktok

Thank you, Orion!

Welcome to the team, so glad to have you.

Along with the Tableau dashboard and notebook, it would be really helpful if you included an executive summary of your analysis attached via email.

Appreciate your help!

Willow Jaffey

Data Science Lead

TikTok

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 each link below and select *Use Template*.

Link to templates:

- Course 3 PACE strategy document
- Executive summary templates

OR

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

Activity Template_ Course 3 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 3 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 perform EDA and create data visualizations. You'll find helpful reminders for tasks like:

- Importing data
- Reviewing and structuring data as needed for analysis
- Analyzing data for outliers
- Visualizing data

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 upcoming step.

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 belief or thought. A "claim" refers to information that is either unsourced or from a 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

Column name	Type	Description
		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 3 PACE strategy document** includes questions that will help guide you through the Course 3 TikTok workplace scenario project. Answer the questions in your PACE strategy document to prepare for using Python for EDA and both Python and Tableau for data visualization.

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: Open Tableau Public and visualize your data

While using Tableau Public, you will need to create the following deliverable for stakeholders:

- Data visualization in the form of a scatter plot

If you need additional help, review the Course 3 Completed Tableau follow-along guide.

Follow-along guide_ TikTok Project
DOCX File

Step 5: 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 results of your Exploratory Data Analysis (EDA)
- A proposed solution for dealing with outliers in your data

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 3 PACE strategy document:

- Answer the questions in the PACE strategy document

Course 3 TikTok project lab:

- Perform Exploratory Data Analysis (EDA)
- Create data visualizations

Course 3 Tableau visualization:

- Create a scatterplot to enhance the visualization created with Python

Course 3 executive summary:

- Provide a summary of the results of your exploratory data analysis (EDA)
- Propose a solution for dealing with outliers in your data