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

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.

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