Scenario

You are a member of the TikTok data analytics team. The team is currently more than halfway through the claims classification project. Earlier, you completed a project proposal, used Python to explore and analyze the claims classification dataset, created data visualizations, and conducted a hypothesis test.

The TikTok team has reviewed the results of the hypothesis testing. TikTok's Operations Lead, Maika Abadi, is interested in how different variables are associated with whether a user is verified. Earlier, the data team observed that if a user is verified, they are much more likely to post opinions. Now, the data team has decided to explore how to predict verified status to help them understand how video characteristics relate to verified users. Therefore, you have been asked to conduct a logistic regression using verified status as the outcome variable. The results may be used to inform the final model related to predicting whether a video is a claim vs an opinion.

You check your inbox and discover a new email from Maika Abadi asking the data team about the details of a regression model. You also notice two follow-up emails from Rosie Mae. The first lists the specific variables that should be analyzed in the logistic regression model. The second email asks you to help build the model and prepare an executive summary to share your results.

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

_____

Email from Maika Abadi, Operations Lead

Subject: Details on Regression Model

From: "Abadi, Maika,"— maikaabadi@tiktok

Cc: ""Jaffey, Willow" —willowjaffey@tiktok; "Rodgers, Mary Joanna" —maryjoannarodgers@tiktok; "Bradshaw, Rosie Mae" rosiemaebradshaw@TikTok; "Rainier, Orion"—orionrainier@tiktok

Hello Data Team,

I really appreciate your work, and thanks for the explanation of the next phase of the data analysis.

I'm curious to know more about how different variables are associated with whether a user is verified. I was hoping to get a bit more detail on the regression your team is planning to conduct.

Will you be applying a linear regression model or a logistic regression model? It wasn't clear in the meeting, and I wanted to be sure our teams are aligned on expectations.

Thank you,

Maika Abadi

Operations Lead

TikTok

Learn about using
TikTok to do good

Email from Rosie Mae Bradshaw TikTok's Data Science Manager

Subject: RE: Details on Regression Phase

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

Cc: "Jaffey, Willow" —willowjaffey@tiktok; "Rodgers, Mary Joanna" —maryjoannarodgers@tiktok; "Rainier, Orion"—orionrainier@tiktok; "Abadi, Maika,"—maikaabadi@tiktok

Thank you for your email.

Apologies that it was not clear in the meeting.

To answer your question, we've decided to look into how to predict 'verified_status', which we believe will help us understand how video characteristics relate to verified users. To achieve this, the data team will build a logistic regression model using 'verified_status' as the outcome variable. The results of this milestone will inform us as we approach constructing the final claims prediction model.

Feel free to reach out with additional questions.

Many thanks,

Rosie Mae Bradshaw

Data Analysis Manager

TikTok

Learn about TikTok's Trust & Safety team

Email from Rosie Mae Bradshaw TikTok's Data Science Manager

Subject: RE: Details on Regression Phase

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

Cc: "Rainier, Orion"—orionrainier@tiktok

Hello my Data team!

Would you two mind completing the following using the fictional test data set::

Logistic regression model in a Python notebook based 'verified_status' variable in the claims classification dataset

Be sure to include a confusion matrix of the results and the accuracy score of the model

 Draft an executive summary of your results

I'd appreciate a chance to look it over before you send it over to Mary Joanna. Please write the summary as if you're addressing the leadership team.

Best regards,

Rosie Mae Bradshaw

Data Analysis Manager

TikTok

Project background

TikTok's data team is working on the claims classification project. The following tasks are needed at this stage of the project:
- Determine the correct modeling approach
- Build a regression model
- Finish checking model assumptions
- Evaluate the model
- Interpret model results and summarize findings for cross-departmental stakeholders within TikTok

Your assignment

You will create a regression model for the claims classification data. You'll determine the type of regression model that is needed and develop one using TikTok's claim classification 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 5 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
- Create a regression model
- Evaluate the model
- Create an executive summary to share your results