Scenario

The TikTok data analytics team has completed the first three milestones of the claims classification project and is nearing the halfway point. So far, the team has completed a project proposal, and used Python to perform exploratory data analysis on the dataset for the claims classification project. The team also produced data visualizations in both Python and Tableau to share with stakeholders. The next step is to use statistical methods to analyze and interpret the claims classification data.

You receive a new email from Mary Joanna Rodgers, one of TikTok's project management officers. Mary Joanna informs the data team about a new request: to determine whether there is a statistically significant difference in the number of views for TikTok videos posted by verified accounts versus unverified accounts. You also receive follow-up emails from Data Science Manager, Rosie Mae Bradshaw and Data Science Lead, Willow Jaffey. These emails share the details of the analysis. A final email from Data Scientist, Orion Rainier, details your next assignment: to conduct a hypothesis test on verified versus unverified accounts in terms of video view count.

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

Email from Mary Joanna Rodgers, Project Management Officer

Subject: New Request - Hypothesis Test: Verified/Unverified Accounts

From: "Rodgers, Mary Joanna" —maryjoannarodgers@tiktok

Cc: "Rainier, Orion"—orionrainier@tiktok; "Jaffey, Willow" —willowjaffey@tiktok

; "Bradshaw, Rosie Mae" —rosiemaebradshaw@tiktok

Hello Data Team!

Really excellent work so far. The leadership team is impressed with the results–especially the progress and insights shared on the last executive summary report! Thanks so much for the hard work.

On that note, they have requested an additional item to be added to the initial project scope. We are interested in whether there is a statistical difference in the data between verified and unverified accounts. Do you have any indication which variable would be most insightful to test in terms of verified and unverified accounts?

Many thanks!

Mary Joanna Rodgers

Project Management Officer

TikTok

Network with TikTok employees from a variety of teams and locations. Participate in TikTok Tuesdays, every Tuesday @2pm EST.

Email from Rosie Mae Bradshaw TikTok's Data Science Manager

Subject: RE: New Request - Hypothesis Test: Verified/Unverified Accounts

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

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

Thanks for the update, Mary Joanna.

It's great to hear that the leadership team is pleased with the data team's progress and the early insights we have been able to deliver. I never grow tired of being reminded of what a great data team we have assembled here at TikTok!

If you would, please tell them we will be providing this analysis in two weeks time.

@Orion, my initial thought is for us to conduct a hypothesis test to analyze whether there is a significant difference in video views for verified versus unverified accounts. What do you think?

In summary, I think we should do the following:

Compute descriptive statistics on the claims classification data

Conduct a two-sample hypothesis test of verified versus unverified accounts in terms of video view counts

Thanks,

Rosie Mae Bradshaw

Data Analysis Manager

TikTok

Learn about TikTok's Trust & Safety team

Email from Orion Rainier, Data Scientist

Subject: RE: New Request - Hypothesis Test: Verified/Unverified Accounts

From: "Rainier, Orion"—orionrainier@tiktok

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

Hi all,

@Rosie Mae, I agree with you on statistical testing. We'll share a summary of the results before we present it to the client.

We'll get started right away.

Thank you,

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: New Request - Hypothesis Test: Verified/Unverified Accounts

From: "Jaffey, Willow" —willowjaffey@tiktok

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

I agree with everyone’s assessments on this project so far. I look forward to the team’s progress in this milestone. Thank you all.

Willow Jaffey

Data Science Lead

TikTok

Email from Orion Rainier, Data Scientist

Subject: RE: New Request - Hypothesis Test: Verified/Unverified Accounts

From: “Rainier, Orion”—orionrainier@tiktok

Cc:

Hi there, fellow data professional!

You’ve been handling all of this work really well, by the way. Excellent job.

I was wondering if you’d like to try the statistical testing yourself? Based on what you’ve shared with me, I have every confidence you already have all the skills and experience needed for this task.

What do you think? Would you like to try?

Also, like I said in my email to Rosie Mae, you’ll need to write an executive summary of the results so we can present it to Willow before sharing it with the client.

Thanks so much!

Orion Rainier

Data Scientist

TikTok

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 at this stage of the project:

- Explore the project data
- Implement a hypothesis test
- Communicate insights with stakeholders within TikTok

Your assignment

You will conduct hypothesis testing on the data for the claims classification data. You've been asked to investigate TikTok's user claim dataset to determine which hypothesis testing method best serves the data and the claims classification project.

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 4 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
- Consider the different groups of data represented in the dataset
- Implement a hypothesis test
- Create an executive summary to share your results