Paige Leeds

UNDERSTANDING TIKTOK CLAIMS

PACE STRATEGY

Google Advanced Data Analytics Professional Certificate

Project Documentation

2026

Table of Contents

00

Overview

Objective

Develop a machine learning model to distinguish factual TikTok claims from opinion-based content.

Business Question

Can engagement, creator, and video characteristics accurately predict whether a video contains a factual claim?

Dataset

- 19,382 TikTok videos
- 12 variables
- Engagement, creator metadata, and claim status

PACE Workflow

Plan → Analyze → Construct → Execute

01

Plan

Approach

Established a structured workflow to inspect the dataset, validate data quality, define key variables, and prepare the data for exploratory analysis and downstream modeling.

Business Context

Understand the dataset, identify potential quality issues, and establish a structured analytical approach before exploratory analysis begins.

Preparation

- Review project objectives and data dictionary
- Import and inspect the dataset
- Validate data types and missing values
- Identify potential outliers
- Confirm the target variable and feature definitions

Resources

Python • pandas • NumPy

02

Analyze

Dataset Assessment

- 19,382 TikTok videos • 12 variables
- Reviewed data types, missing values, and duplicates

Key Observations

- Verified overall data quality
- Explored variable distributions and potential outliers
- Examined relationships between engagement metrics and claim status

Exploratory Data Analysis

- Generated descriptive statistics
- Validated key variables for downstream modeling
- Prepared the dataset for feature exploration

Deliverables

- Cleaned dataset
- Summary statistics
- Exploratory analysis notebook

03

Construct

Model Development

Built and compared supervised machine learning models to classify TikTok claims as factual or opinion-based.

Feature Engineering

- Selected predictor variables
- Prepared training and testing datasets
- Encoded categorical variables
- Standardized model inputs

Model Evaluation

- Compared model performance using accuracy, precision, and confusion matrices
- Examined feature importance

Deliverables

- Trained classification model
- Performance metrics and model evaluation report

04

Execute

Key Findings

- Engagement patterns differed by claim type
- Data quality supported reliable downstream modeling
- Exploratory analysis identified informative predictor variables

Business Recommendations

- Prioritize engagement metrics as model inputs
- Continue monitoring data quality before retraining
- Expand features with creator and content attributes

Project Outcome

Developed a structured analytical workflow to support machine learning model development and business decision-making.

Deliverables

- Business recommendations
- Final project documentation
- Analytical workflow