

Mid-Level Project Report

TikTok Data Analysis Project – Travel/Visa Niche

Course: Data Analytics for AI Applications
Faculty of Computer Science, TU Chemnitz

Author: Ayodele Moses Omolanke

Date: 17th December, 2025

1. Project Overview

The goal of this project is to analyze TikTok video performance using real data from my own TikTok account in the travel and visa niche. TikTok is a platform where content performance depends on many factors such as hook style, content format, video duration, and audience engagement. However, many creators rely on intuition rather than data-driven decisions.

In this project, I aim to use data analytics techniques to understand which factors contribute to higher engagement on TikTok videos related to travel, visas, scholarships, internships, and international opportunities. The final goal is to build a simple machine learning model that can help predict whether a video is likely to perform well based on its characteristics.

The project follows a full data analytics pipeline, starting from dataset design and data collection, moving through exploratory data analysis, and later applying machine learning techniques.

2. Dataset Description

The dataset used in this project is collected manually from TikTok Analytics. It consists of real performance data from my own TikTok videos. At the current mid-level stage, the dataset contains **10 recent high-engagement videos** published in November 2025.

Each data entry represents one TikTok video and includes both numerical and categorical features. The collected features include:

- Views
- Likes
- Comments
- Shares
- Saves (favorites)

- Profile visits
- Followers gained
- Video duration (seconds)
- Watch time percentage
- Average watch time (seconds)
- Posting date, time, and day of the week
- Hook style (fomo, value, tutorial, question)
- Content format (text overlay, voiceover, screen recording)
- Niche category (visa, scholarship, internship, conference, study)
- Hashtag information

In addition, a new derived feature called **engagement_rate** was created using the formula:

$$\text{engagement_rate} = (\text{likes} + \text{comments} + \text{shares} + \text{saves}) / \text{views}$$

This feature provides a better measure of audience interaction than views alone, since it captures how actively viewers engage with the content.

The dataset is stored in both Excel and CSV formats and is version-controlled using GitHub.

3. Milestones Completed

These milestones were complete ahead of the mid-level deadline:

3.1 Dataset Design and Collection

- Designed a structured dataset suitable for data analytics and machine learning.
- Collected real TikTok analytics data manually for 10 videos.
- Ensured data consistency and correctness across all features.

3.2 Feature Engineering

- Created a new engagement metric (**engagement_rate**) to better measure performance.
- Categorized videos by hook style, content format, and niche topic.

3.3 Exploratory Data Analysis (EDA)

- Performed initial exploratory analysis using Excel.
- Created visualizations to compare:
 - Engagement rate across individual videos.
 - Average engagement rate by hook style.
- Identified early patterns, such as higher engagement for FOMO and value-based hooks.

3.4 Project Organization and Documentation

- Created a well-structured GitHub repository including folders for data, documentation, notebooks, source code, and presentation.
- Uploaded dataset files, report drafts, and presentation slides.
- Prepared and delivered a successful progress presentation, which received positive feedback from the lecturer.

These completed milestones demonstrate that the project is progressing according to plan and that the foundational data analysis work is already in place.

4. Upcoming Milestones and Plan

For the remainder of the project, the following milestones are planned:

4.1 Dataset Expansion

- Increase the dataset size from 10 videos to at least 30–50 videos.
- Include a wider range of performance levels (not only high-engagement videos).

4.2 Extended Exploratory Data Analysis

- Create additional charts, such as:
 - Engagement rate by content format.
 - Video duration versus views.
 - Posting day and time versus engagement.
- Analyze hashtag usage patterns.

4.3 Machine Learning Model

- Define a binary target variable (high-performing vs. normal-performing videos).
- Train a Logistic Regression model to predict video performance.
- Optionally compare results with a Random Forest model.
- Evaluate models using accuracy and basic performance metrics.

4.4 Final Report and Presentation

- Update the report with full results, visualizations, and model evaluation.
- Prepare a final presentation summarizing the entire project.
- Submit the final project by the official deadline.

5. Conclusion

This mid-level report shows that the project has made strong and structured progress. The dataset has been designed and collected, key features have been engineered, and initial exploratory analysis has already produced meaningful insights. The remaining work focuses on expanding the dataset, applying machine learning techniques, and refining the analysis for the final submission.

The project demonstrates how data analytics can be applied to real-world social media content to support better, data-driven decision-making.