TikTok Data Analysis:

# TRAVEL & VISA NICHE

Ayodele Moses Omolanke

Table of content

# CHAPTER 1

## INTROCUTION

In today's digital world, short video platforms like TikTok have become powerful tools for sharing information, building communities, and influencing decision-making. Many young people depend on TikTok to learn about travel opportunities, visa processes, scholarships, and career pathways abroad. Because of this, creators in the travel and visa niche play an important role in providing quick and relevant information to viewers.

As a content creator in this niche, I produce videos about travel tips, visa sponsorship opportunities, scholarships, and study or work abroad guidance. My videos reach a wide audience across different countries, and each video performs differently depending on several factors such as the topic, hook, posting time, video duration, and content style. Understanding why some videos perform better than others is important not only for improving my content, but also for learning how data analytics can help in real-life digital communication.

The goal of this project is to analyze my TikTok travel/visa videos using the data analysis techniques taught in this course. By collecting and examining the performance of my past videos, I want to identify the key factors that influence engagement and reach. This project follows the structure of a Data Analytics/Explainer Project, as defined in the course guidelines, and applies the basic steps of dataset exploration, cleaning, visualization, and simple predictive modeling.

# CHAPTER 2

## DATASET AND ENTITY DESCRIPTION

The entity I selected for this project is my own TikTok account, where I create travel and visa-related content. This includes videos about visa-free travel, study abroad tips, work opportunities, visa sponsorship, scholarships, and immigration updates. I chose this entity because I already have access to real engagement data, and it allows me to apply data analytics to a real-world situation that I am personally involved in.

For the dataset, I will manually collect information from the TikTok Creator Analytics dashboard. Each row in the dataset represents one video I posted. The following variables will be included:

- Video ID
- Short description or title
- Date posted
- Time posted
- Day of the week
- Niche category (e.g., visa, travel, scholarship, work abroad)
- Video format (talking head, slideshow, green screen, screen recording, etc.)
- Hook style (question, FOMO, story, etc.)
- Video duration (seconds)
- Views
- Likes
- Comments
- Shares
- Saves/Favorites
- Profile visits
- Followers gained
- Watch time percentage (if available)
- Average watch time
- Top three hashtags
- Total hashtag count

The dataset will contain at least 20–30 videos, which is enough to conduct basic exploratory analysis and train a simple prediction model. Once the data is collected, it will be saved in two formats: **Excel (.xlsx)** and **CSV (.csv)**, as recommended for data analysis pipelines.

This dataset is unique because it is not downloaded from an existing platform like Kaggle; instead, it is based on original content and real audience interaction. This makes the project more practical and closely aligned with the “real-world dataset” requirement described in the lecture.

# CHAPTER 3

## DATA CLEANING AND PREPROCESSING PLAN

Before starting the analysis, the dataset needs to be cleaned to make sure all values are consistent and usable. The cleaning steps will include:

1. **Handling missing values**
   Some metrics, like watch time percentage or followers gained, may not appear for every video. Missing values will either be left blank, filled with zero, or marked as "not available," depending on what makes sense for each field.
2. **Formatting dates and times**
   The posting date will be converted into a proper `YYYY-MM-DD` format, and the posting time will be converted into 24-hour format.
   The day of the week will be extracted automatically (e.g., Monday, Tuesday).
3. **Converting numeric fields**
   Columns such as views, likes, comments, shares, and saves will be converted to integer values.
   Any symbols (like "K" for thousands) will be removed.
4. **Categorizing video formats and hook styles**
   These will be manually tagged into a small number of categories.
   For example:
   - Talking head
   - Slideshow
   - Green screen
   - List-style
   - Story-style
5. **Calculating new features**
   - **Engagement rate** = (likes + comments + shares) / views
   - **Hashtag count** will be counted from the caption
   - **Duration group** (short, medium, long) may be created for easier comparison
6. **Ensuring consistency**
   All categorical variables (e.g., "visa", "Visa", "VISA") will be made consistent so that the analysis is accurate.

After cleaning, the dataset will be ready for exploratory data analysis (EDA) and model development as required in the project guidelines.