**Data Project Questions & Considerations**

**PACE: Plan Stage**

* Who is your audience for this project?

### **Technical Team:**

* **Data Science Team** – They need detailed insights on data exploration, statistical testing, and model selection.
* **Data Analysts & Engineers** – They will work on data preprocessing, hypothesis testing, and machine learning model development.

### **Non-Technical Stakeholders:**

* **Project Management Officer** – Needs a high-level overview of project milestones, deliverables, and workflow structure.
* **Operations Lead & Finance Lead** – Require summary reports and visualizations to understand the project’s impact on business processes.
* **TikTok Executives** – Need a final presentation with key insights and the model's business impact to support decision-making.
* What are you trying to solve or accomplish? And, what do you anticipate the impact of this work will be on the larger needs of the client?

We aim to develop a **predictive machine learning model** that can classify TikTok user interactions (videos and comments) as either **claims or opinions**. This will help streamline content moderation by prioritizing reported content for review.

* What questions need to be asked or answered?

### 1. How is the current state of the provided dataset described? 2. Which variables are likely to be considered the most valuable? 3. Are any trends in the data likely to offer useful insights? 4. What actions can be taken to minimize the effects of bias?

* What resources are required to complete this project?

The project dataset, Python notebook, and the information from stakeholders will be needed as well.

* What are the deliverables that will need to be created over the course of this project?

The deliverables for the TikTok claims classification project will include:

1. **Project Proposal:** Outlining the goals, milestones, and resources for the project.
2. **Exploratory Data Analysis (EDA) Report:** Insights on data patterns, distributions, and potential issues.
3. **Data Preprocessing Pipeline:** Steps for cleaning and preparing the data for modeling.
4. **Model Selection & Evaluation:** A report comparing different machine learning models and their performance.
5. **Final Predictive Model:** A trained model capable of classifying claims vs. opinions.
6. **Visualization Dashboards:** Interactive visuals to present key findings and model performance.
7. **Presentation to Stakeholders:** Summarizing results, insights, and recommendations for the leadership team.