**Project Porfolio #1 Pace Framework Questions.**

# Go Beyond the Numbers: Translate Data into Insights



**Data Project Questions & Considerations**

**PACE: Plan Stage**

* What are the data columns and variables and which ones are most relevant to your deliverable?

Data columns:

Tiktok Number

Claim status

video \_ID

Video\_Duration\_Sec

Video\_transcription\_Text

Verified\_status

Author Ban Status

VIdeo\_View\_Count

Video\_like\_count

Video\_comment\_count

The most relevant columns are claim counts and opinion counts; video duration, video like count; video comment count and video view count.

* What units are your variables in?

The units in this project are Integers , objects and floats.

* What are your initial presumptions about the data that can inform your EDA, knowing you will need to confirm or deny with your future findings?

There are a few nulls in a few different categories that I have to clean up before I can properly perform EDA. Most of the data is numerical so I can effectively manipulate it but I might have to convert the object data types to numerical to manipulate it.

* Is there any missing or incomplete data?

Yes, there are a few Nans in a few of the columns.

* Are all pieces of this dataset in the same format?

No, there are different data types; integers, objects and floats.

* Which EDA practices will be required to begin this project?

I will have to remove, replace or find the correct data as the beginning steps of EDA.

**PACE: Analyze Stage**

* What steps need to be taken to perform EDA in the most effective way to achieve the project goal?

I will react to the data for the first time and figure out if I need to join, replace or remove the data as much as needed. Then I will clean, reorganize, and initially analyze the data for the goal in mind.

* Do you need to add more data using the EDA practice of joining? What type of structuring needs to be done to this dataset, such as filtering, sorting, etc.?

I don't need to add more data but I do need to filter out the bad data first.

* What initial assumptions do you have about the types of visualizations that might best be suited for the intended audience?

Given that I have several Tiktok managers that are less technical in nature, I will create visuals that are concise and straight to the point. I will make the boxplots pie charts that my higher-ups are asking for as stated in the Jupyter notebook they provided me.

**PACE: Construct Stage**

* What data visualizations, machine learning algorithms, or other data outputs will need to be built in order to complete the project goals?

As I said in the ‘Analyze’ stage of this project's PACE framework. I will make pie charts, stacked bar charts and box plots in tableau to compare the necessary columns to each other because it was a necessity in this project.

* What processes need to be performed in order to build the necessary data visualizations?

I need to build the visualizations in Python with the visualization packages to manipulate the data then replicate the visualizations in Tableau.

The variables that my Lead Data Scientist told me to act upon were claim counts, opinion counts, video duration, video like count, video comment count, video view count & author ban status

* Which variables are most applicable for the visualizations in this data project?

The variables that my Lead Data Scientist told me to act upon were claim counts, opinion counts, video duration, video like count, video comment count, video view count & author ban status

* Going back to the Plan stage, how do you plan to deal with the missing data (if any)?

I will simply remove the data, given that we have such a large amount of data that it will not affect the results.

******PACE: Execute Stage**

* What key insights emerged from your EDA and visualizations(s)?

After performing EDA it was clear that almost every video has the status of ‘claim’ if the video went viral. Most authors are active in correlation to the video they posted and I can also conclude that it was almost split right down the middle in terms of videos being ‘claims’ and ‘opinions’.

* What business and/or organizational recommendations do you propose based on the visualization(s) built?

I can recommend we look at the type of content that is posted. Why are the opinionated videos so low in like and view count? Why are claimed videos so sporadic in view and like count? I feel like it has to do with the type of content the author created.

* Given what you know about the data and the visualizations you were using, what other questions could you research for the team?

I think we could research the type of content that goes viral and why are the author being striked as claimed.

* How might you share these visualizations with different audiences?

I could share these visuals just how they are given that they are simple to understand by nature. But i can also create a more in-depth version for people that are familiar with statistics and data analysis.