**Course Two**

Get Started with Python



Instructions

Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. You can use this document as a guide to consider your responses and reflections at different stages of the data analytical process. Additionally, the PACE strategy documents can be used as a resource when working on future projects.

Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

* Complete the questions in the Course 2 PACE strategy document
* Answer the questions in the Jupyter notebook project file
* Complete coding prep work on project’s Jupyter notebook
* Summarize the column Dtypes
* Communicate important findings in the form of an executive summary

Relevant Interview Questions

Completing the end-of-course project will help you respond these types of questions that are often asked during the interview process:

* Describe the steps you would take to clean and transform an unstructured data set.
* What specific things might you look for as part of your cleaning process?
* What are some of the outliers, anomalies, or unusual things you might look for in the data cleaning process that might impact analyses or ability to create insights?

**Reference Guide**

This project has three tasks; the visual below identifies how the stages of PACE are incorporated across those tasks.



**Data Project Questions & Considerations**

**PACE: Plan Stage**

* How can you best prepare to understand and organize the provided information?

First, I would review the goal of the project and the project proposal. Then, I would reach the Data Science Manager to identify which are the details expected for the analysis conducted in this first stage of the project.

* What follow-along and self-review codebooks will help you perform this work?
* What are some additional activities a resourceful learner would perform before starting to code?

Before coding, I would create a simple document outlining the different steps I want to perform (ideally on top of the script or Jupyter notebook to make the code self-explanatory and clear).

**PACE: Analyze Stage**

* Will the available information be sufficient to achieve the goal based on your intuition and the analysis of the variables?

Yes, the preliminary analysis of the TikTok dataset reveals evident trends in the data which could be useful to classify the claims and opinions in the videos.

* How would you build summary dataframe statistics and assess the min and max range of the data?

Using the .describe method.

* Do the averages of any of the data variables look unusual? Can you describe the interval data?

The averages do not look unusual. However, some of the columns may contain outliers as per huge max values compared to the mean of their distribution. The intervals for the engagement metrics range from hundreds to thousands.

**PACE: Construct Stage**

**Note**: The Construct stage does not apply to this workflow. The PACE framework can be adapted to fit the specific requirements of any project.

**PACE: Execute Stage**

* Given your current knowledge of the data, what would you initially recommend to your manager to investigate further prior to performing exploratory data analysis?

First, I would tackle the null values present in some of the columns in the dataset. Then, I would focus on the engagement measurements which were derived as a rate per view.

* What data initially presents as containing anomalies?

Some of the columns include null values. Others seem to contain outliers as revealed by the huge max values in comparison to the mean of their distribution.

* What additional types of data could strengthen this dataset?

Information regarding the count of times a video was reported could be useful for the objective established in this project.