Activity_Course 4 TikTok project lab

June 10, 2024

1 TikTok Project

Course 4 - The Power of Statistics

You are a data professional at TikTok. The current project is reaching its midpoint; a project proposal, Python coding work, and exploratory data analysis have all been completed.

The team has reviewed the results of the exploratory data analysis and the previous executive summary the team prepared. You received an email from Orion Rainier, Data Scientist at TikTok, with your next assignment: determine and conduct the necessary hypothesis tests and statistical analysis for the TikTok classification project.

A notebook was structured and prepared to help you in this project. Please complete the following questions.

2 Course 4 End-of-course project: Data exploration and hypothesis testing

In this activity, you will explore the data provided and conduct hypothesis testing.

The purpose of this project is to demostrate knowledge of how to prepare, create, and analyze hypothesis tests.

The goal is to apply descriptive and inferential statistics, probability distributions, and hypothesis testing in Python.

This activity has three parts:

Part 1: Imports and data loading * What data packages will be necessary for hypothesis testing?

Part 2: Conduct hypothesis testing * How will descriptive statistics help you analyze your data?

- How will you formulate your null hypothesis and alternative hypothesis?

Part 3: Communicate insights with stakeholders

- What key business insight(s) emerge from your hypothesis test?

- What business recommendations do you propose based on your results?

Follow the instructions and answer the questions below to complete the activity. Then, complete an executive summary using the questions listed on the PACE Strategy Document.

Be sure to complete this activity before moving on. The next course item will provide you with a completed exemplar to compare to your own work.

3 Data exploration and hypothesis testing

4 PACE stages

Throughout these project notebooks, you'll see references to the problem-solving framework PACE. The following notebook components are labeled with the respective PACE stage: Plan, Analyze, Construct, and Execute.

4.1 PACE: Plan

Consider the questions in your PACE Strategy Document and those below to craft your response.

1. What is your research question for this data project? Later on, you will need to formulate the null and alternative hypotheses as the first step of your hypothesis test. Consider your research question now, at the start of this task.

==> ENTER YOUR RESPONSE HERE

Complete the following steps to perform statistical analysis of your data:

4.1.1 Task 1. Imports and Data Loading

Import packages and libraries needed to compute descriptive statistics and conduct a hypothesis test.

Hint:

Be sure to import `pandas`, `numpy`, `matplotlib.pyplot`, `seaborn`, and `scipy`.

```
[ ]:  # Import packages for data manipulation
      ### YOUR CODE HERE ###


      # Import packages for data visualization
      ### YOUR CODE HERE ###



      # Import packages for statistical analysis/hypothesis testing
      ### YOUR CODE HERE ###
```

Load the dataset.

Note: As shown in this cell, the dataset has been automatically loaded in for you. You do not need to download the .csv file, or provide more code, in order to access the dataset and proceed with this lab. Please continue with this activity by completing the following instructions.

2

```
[ ]: # Load dataset into dataframe
     data = pd.read_csv("tiktok_dataset.csv")
```

4.2 PACE: Analyze and Construct

Consider the questions in your PACE Strategy Document and those below to craft your response:
1. Data professionals use descriptive statistics for Exploratory Data Analysis. How can computing descriptive statistics help you learn more about your data in this stage of your analysis?

==> ENTER YOUR RESPONSE HERE

4.2.1 Task 2. Data exploration

Use descriptive statistics to conduct Exploratory Data Analysis (EDA).

Hint:

Refer back to *Self Review Descriptive Statistics* for this step-by-step proccess.

Inspect the first five rows of the dataframe.

```
[ ]: # Display first few rows
     ### YOUR CODE HERE ###
```

```
[ ]: # Generate a table of descriptive statistics about the data
     ### YOUR CODE HERE ###
```

Check for and handle missing values.

```
[ ]: # Check for missing values
     ### YOUR CODE HERE ###
```

```
[ ]: # Drop rows with missing values

     ### YOUR CODE HERE ###
```

```
[ ]: # Display first few rows after handling missing values

     ### YOUR CODE HERE ###
```

You are interested in the relationship between `verified_status` and `video_view_count`. One approach is to examine the mean value of `video_view_count` for each group of `verified_status` in the sample data.

```
[ ]: # Compute the mean `video_view_count` for each group in `verified_status`
     ### YOUR CODE HERE ###
```

4.2.2 Task 3. Hypothesis testing

Before you conduct your hypothesis test, consider the following questions where applicable to complete your code response:

1. Recall the difference between the null hypothesis and the alternative hypotheses. What are your hypotheses for this data project?

==> ENTER YOUR RESPONSE HERE (Double Click)

Your goal in this step is to conduct a two-sample t-test. Recall the steps for conducting a hypothesis test:

1. State the null hypothesis and the alternative hypothesis
2. Choose a signficance level
3. Find the p-value
4. Reject or fail to reject the null hypothesis

==> ENTER YOUR NULL AND ALTERNATIVE HYPOTHESES HERE (Double Click)

You choose 5% as the significance level and proceed with a two-sample t-test.

```
[ ]:  # Conduct a two-sample t-test to compare means
      ### YOUR CODE HERE ###
```

Question: Based on the p-value you got above, do you reject or fail to reject the null hypothesis?

==> ENTER YOUR RESPONSE HERE (Double Click)

4.3 PACE: Execute

Consider the questions in your PACE Strategy Documentto reflect on the Execute stage.

4.4 Step 4: Communicate insights with stakeholders

Ask yourself the following questions:

1. What business insight(s) can you draw from the result of your hypothesis test?

==> ENTER YOUR RESPONSE HERE

Congratulations! You've completed this lab. However, you may not notice a green check mark next to this item on Coursera's platform. Please continue your progress regardless of the check mark. Just click on the "save" icon at the top of this notebook to ensure your work has been logged.