**KT 568 -Assignment 2**

**Names: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Due: Oct 2, 11:59 PM**

**Scoring: 30 points total (15% of your grade)**

**This assignment aims to help you solidify your understanding of A/B test and hypothesis testing. With your team, discuss the following questions and answer them. In your submission, delete the question texts and *only* include your answers. Your submission must be written in an M.S. Word file with a Times New Roman, 12 Font. Any sources used in the assignment must be properly cited in the APA format (7th edition). Failing to cite your sources is considered plagiarism.**

**What you submit on the deadline includes (1) a Word document where you have responded to the questions listed in this document and (2) a Python code with the complete code that helps you respond to question B. The Python code must be in the format of a Jupyter Notebook (.ipynb) and without errors. You will lose at least 2 points if your Python files throw errors.**

**Question A. Designing A/B Tests for Launching “VibeFest” – A New Music and Arts Festival in New York City (10 pts)**

VibeFest is an exciting new music and arts festival that will debut next June on Randall’s Island, New York City. The festival aims to bring together emerging musicians, interactive art installations, food trucks, and immersive experiences. Since this will be its first year, VibeFest’s primary goal is to build brand awareness and sell as many tickets as possible to establish itself as a premier cultural event in NYC’s highly competitive festival scene.

To do so, in January, the festival organizers plan to launch an extensive promotional campaign across multiple social media platforms, including Instagram, Facebook, YouTube, and TikTok. The marketing team has developed various advertisements to target the festival’s audience on these platforms. As part of this broader campaign, the marketers are focused on designing a specific advertisement for Instagram. The planned advertisement will be a static image shown to Instagram users in New York State and neighboring areas, similar to an Instagram post with a description below the post that provides information about the event, a link to the event’s website, and ticket information.

The marketing team has created two distinct images for the Instagram advertisement and wants to run an A/B test with a small sample of Instagram users to determine which ad is more effective before rolling out the full campaign.

* Ad A: Emphasizes the unique aspects of the festival—exclusive music performances, interactive art installations, and gourmet food offerings.



* Ad B: Focuses on the overall experience of attending VibeFest, using images of people enjoying themselves and dancing to live music on Randall’s Island.



The goal of the A/B test is to pretest these two advertisements with a small set of Instagram users (1000 users) in November and measure which ad is more effective in driving clicks (whether the customer clicks on the link included in the post description below the image) and ticket purchases (whether the customer buys a ticket to the event after seeing this ad). The results from this small-scale test on Instagram will determine which ad is more successful, and that advertisement will be used in the larger social media campaign in January when many more users are being targeted by the ad. You are responsible for designing this A/B test. Respond to the following questions regarding designing this test:

1. What are the hypotheses in this study? Visually present the hypotheses in boxes and arrows (1 pt) AND write them out in words similar to what we did in class. (1 pt)
2. For each hypothesis you wrote in QA-1, also write the null hypothesis required for the A/B testing. (1 pt)
3. For each hypothesis you wrote in QA-1, determine the independent and the dependent variable in the hypothesis QA-1. (1 pts)
4. Explain how you would run this A/B test on Instagram. For example, how would you run these ads on Instagram from a technical standpoint, who will see the ads, in what timeframe would you run these ads, how will you measure the dependent variables, and how will you randomly assign participants to the two conditions? (3 pts)
5. Consider your answer to QA-4, what challenges may arise during the A/B test? Note that no matter how you design an A/B test, challenges will still exist. (1 pt)
6. For each hypothesis you wrote in QA-1, name thestatistical analysis you need to run to test the hypothesis and explain why. (2 pt)

**Question B. Price Testing for “StreamSmart” Subscription Service (20 points)**

StreamSmart is a popular subscription-based video streaming service, similar to Netflix, that offers a wide variety of TV shows, movies, and original content. The company operates on a monthly subscription model, where users pay a fixed fee to access all content available on the platform. Currently, the primary subscription plan costs $9.99 per month, providing full access to the entire content library.

StreamSmart recently decided to raise its subscription price due to increasing operational costs. This price increase is essential for the company to remain profitable. However, the marketing manager is uncertain whether the price should be raised by $1 (to $10.99) or by $2 (to $11.99) per month. While the company aims to maximize revenue, there are concerns about customer retention, as a significant price increase might lead to subscription cancellations.

#### To make an informed decision, StreamSmart plans to conduct an A/B test to determine which price point—either $10.99 or $11.99—leads to fewer plan cancellations. They will select a random sample of their customers for the experiment and divide them randomly into two groups, each receiving one of the new prices. Throughout the experiment, StreamSmart will track the number of cancellations in each group. After analyzing the results, they will implement the price increase that results in the lowest cancellation rate.

#### The marketing team has designed an A/B test where they selected 1000 users and randomly assigned them to one of the following experimental conditions:

#### Group A: Price increase by $1

#### Group B: Price increase by $2

#### All customers participating in the experiment will receive an email from the company on April 1st. Participants in the “Price Increase by $1” group will receive a message thanking them for their loyalty, explaining that, due to rising costs, the subscription price will increase by $1 starting next month (May 1st) to ensure the company can continue operating in the market. Similarly, those in the “Price Increase by $2” group will receive an email thanking them for their loyalty and informing them that, due to cost increases, the subscription price will rise by $2 beginning next month (May 1st) to maintain the company’s operations.

#### The company ran the experiment in April and May, with all 1,000 customers receiving the email on April 1st. On May 1st—one month after the emails were sent—the company implemented the price increases for these customers: a $1 increase for Group A and a $2 increase for Group B. The marketing team has been monitoring cancellation behavior among these two groups of customers over the past months and has recorded cancellations at two specific points in time:

#### April: Customer cancellations were recorded for one month, starting from the day the email was sent and continuing until the price increase was implemented (covering the entire month of April).

#### May: Customer cancellations were tracked for one month, beginning on the day the price increase took effect and continuing through the end of the month (covering the entire month of May).

#### The data for this analysis can be downloaded from Canvas (StreamSmart.xlsx). Here is the data dictionary:

* Condition: Indicates whether the customer experienced Price Increase A or Price Increase B.
* April\_Cancellation: Indicates whether the customer canceled their plan in April. 1 = The customer canceled their plan in April, 0 = The customer did not cancel their plan in April.
* May\_Cancellation: Indicates whether the customer’s plan status was marked as “canceled” in May. If the customer canceled their plan in either April or May, the value is set to 1. 1 = The customer canceled their plan in April or May; 0 = The customer did not cancel their plan in May.
  + Note: For May\_Cancellation, if a customer canceled their plan in either April or May, the value is set to 1. Therefore, to analyze the effect of the experimental conditions on May cancellations only, you must first filter out the data from customers who canceled their plans in April. This will allow you to focus exclusively on those who canceled their plans in May.

#### Please download the data and respond to the following questions:

#### What are the hypotheses in this A/B test? Visually present all hypotheses in boxes and arrows (1 pts) AND write them out in words similar to what we did in class (1 pts).

#### For each hypothesis you wrote in QB-1, also write the null hypothesis required for the A/B testing (1 pt)

#### For each hypothesis you wrote in QB-1, determine the independent and the dependent variables. (1 pt)

#### Before running any analysis, clean the dataset. In your Word submission, briefly explain the steps you took to clean the data. In your Python file submission, include the code you used to clean the dataset. (2 pts)

#### How many users in each condition canceled their plans in the month of receiving the email (April)? (1 pt)

#### How many users in each condition canceled their plans in the month of the price increase (May)? (1 pt)

#### Run the proper analyses to test whether the two experimental conditions (A vs. B variants or conditions) led to higher cancellations during the month of receiving the email (April). Did Plan B lead to more cancellations during April compared to Plan A? In your response, include the appropriate statistical terms that help you support your answer (p-values and counts or percentages). (5 pts)

#### Run the proper analyses to test whether the two experimental conditions (A vs. B variants or conditions) led to higher cancellations during the month of price increase (May). Did Plan B lead to more cancellations during May compared to Plan A? In your response, include the appropriate statistical terms that help you support your answer (p-values and counts or percentages). (5 pts)

#### Based on the analysis reported in QB-7 and QB-8, which of the two price increases do you recommend to the company? (1 pts)

#### The company currently has 2 million users. The main subscription plan’s price was fixed at $9.99 throughout the past year. Based on the data from the previous year, the average monthly cancellation rate was .00002 per month (40 customers per month). Based on this information, would you recommend that the company increase the plan’s price to plan A or B, or do you recommend that they avoid making a price change altogether? (1 pts)