Background
You are an analyst for an online pharmacy. The company's customers use our online app to order prescription medications, and the company delivers those medications to the customer via courier.

Like most online e-commerce apps, customers interact with the app by adding items (prescription medications, in this case) to a shopping cart. When the customer is ready, they go through a checkout process to place an order for the items in their cart.

During the checkout process, we recently added a feature that allows customers to add additional, non-medication items to their cart (for example, vitamins or Band-Aids). These non-medication items get ordered along with the rest of the cart during checkout, and they are delivered via courier together with the rest of the items in the cart.

Case
The Product Team is considering a change in February 2024 whereby we increase the visibility of the feature that allows customers to add additional, non-medication items to their cart during the checkout process. The Product Team believes the change will result in 20% of customers going through the checkout process to add non-medication items to their cart.

As the analyst supporting this initiative, you are responsible for estimating the expected financial impact of this change and developing an experimentation plan to determine the success of the feature.

Questions
1) How would you evaluate the potential impact of the change being proposed by the Product Team?
2) Suppose the Product Team proceeds with implementing the proposed change. How would you evaluate whether the change is a success?

Assumptions

- **About the Data:**
  - Data range: 11/01/2023 to 01/19/2024.
  - Data quality is ensured (no duplicates, null rows, etc.).
- **About the Business Model:**
  - On February 1, 2024, the Product Team will increase the visibility of the feature that allows customers to add non-medication items to their cart during checkout.
  - Success is defined as 20% (or higher) of customers adding non-medication items to their cart during checkout.
    → **Non-Medication Items Add-to-Cart Rate (Checkout attempts) = %20**

Metrics Analysis

- **Checkout Metrics:**
  - **Non-Medication Items Add-to-Cart Rate (Checkout attempts):**
    - Formula: Number of orders where customer_added_non_med_to_order is True / Total checkout attempts
      - Overall Non-Medication Items Add-to-Cart Rate

```
SELECT
  (COUNT(case when customer_added_non_med_to_order = TRUE THEN 1 else NULL END)+0.0)/COUNT(*)*100 as non_med_added_rate
FROM
  modealto.checkout_attempts
```

        - 3.26%
      - Monthly Breakdown Non-Medication Items Add-to-Cart Rate:

```
SELECT
  EXTRACT(YEAR FROM checkout_started_at) as check_out_year,
  EXTRACT(MONTH FROM checkout_started_at) as check_out_month,
  (COUNT(case when customer_added_non_med_to_order = TRUE THEN 1 else NULL END)+0.0)/COUNT(*)*100 as non_med_added_rate
FROM
  modealto.checkout_attempts
-- WHERE
--   order_id is not NULL
GROUP BY
  EXTRACT(YEAR FROM checkout_started_at),
  EXTRACT(MONTH FROM checkout_started_at)
ORDER BY
  EXTRACT(YEAR FROM checkout_started_at),
  EXTRACT(MONTH FROM checkout_started_at)
```

        - 2023-11: 0%
        - 2023-12: 4.6%
        - 2024-01: 6.43%
  - **Non-Medication Items Add-to-Cart Rate (Completed orders):**
    - Formula: Number of orders where customer_added_non_med_to_order = True / Total number of completed orders

- Overall Non-Medication Items Add-to-Cart Rate

```
SELECT
  (COUNT(case when customer_added_non_med_to_order = TRUE THEN 1 else NULL END)+0.0)/COUNT(*)*100 as non_med_added_rate
FROM
  modealto.checkout_attempts
WHERE
  order_id is not NULL
```

    - 5.46%
- Monthly Breakdown Non-Medication Items Add-to-Cart Rate:

```
SELECT
  EXTRACT(YEAR FROM checkout_started_at) as check_out_year,
  EXTRACT(MONTH FROM checkout_started_at) as check_out_month,
  (COUNT(case when customer_added_non_med_to_order = TRUE THEN 1 else NULL END)+0.0)/COUNT(*)*100 as non_med_added_rate
FROM
  modealto.checkout_attempts
WHERE
  order_id is not NULL
GROUP BY
  EXTRACT(YEAR FROM checkout_started_at),
  EXTRACT(MONTH FROM checkout_started_at)
ORDER BY
  EXTRACT(YEAR FROM checkout_started_at),
  EXTRACT(MONTH FROM checkout_started_at)
```

    - 2023-11: 0%
    - 2023-12: 7.7%
    - 2024-01: 10.57%
- **Checkout Rate (1 - Abandon Rate):**
    - Formula: Total number of completed orders / Total checkout attempts
        - Overall Checkout Rate:

```
SELECT
    ROUND((COUNT(checkout_completed_at) + 0.0) / COUNT(*) * 100, 2) AS check_out_rate
FROM
    modealto.checkout_attempts;
```

        - 59.67%
        - Monthly Breakdown Checkout Rate

```
SELECT
  EXTRACT(YEAR FROM checkout_started_at) as check_out_year,
  EXTRACT(MONTH FROM checkout_started_at) as check_out_month,
  ROUND((COUNT(checkout_completed_at)+0.0)/COUNT(*)*100,2) as check_out_rate
FROM
  modealto.checkout_attempts
GROUP BY
  EXTRACT(YEAR FROM checkout_started_at),
  EXTRACT(MONTH FROM checkout_started_at)
ORDER BY
  EXTRACT(YEAR FROM checkout_started_at),
  EXTRACT(MONTH FROM checkout_started_at)
```

        - 2023-11: 58.96%
        - 2023-12: 59.68%
        - 2024-01: 60.85%

→ Let's consider Non-Medication Items Add-to-Cart Rate (Completed orders)
Non-Medication Items Add-to-Cart Rate (Completed orders)
= Number of orders where customer_added_non_med_to_order is True / Total number of completed orders
= Number of orders where customer_added_non_med_to_order is True / Total checkout attempts * Checkout Rate
= Non-Medication Items Add-to-Cart Rate (Checkout attempts) / Checkout Rate

- ## Delivery Metrics:
  - **Successful Delivery Rate:**
    - Formula: Number of completed deliveries / Total number of completed orders
      - Overall Successful Delivery Rate:

```sql
-- Delivery Rate
SELECT
    ROUND((COUNT(delivered_date)+0.0)/COUNT(*)*100,2) AS delivery_rate
FROM
  modealto.checkout_attempts ca
JOIN
  modealto.order_status os
USING (order_id)
```

      - 93.14%
      - Monthly Successful Delivery Rate:

```sql
-- Delivery Rate Break Down By Month
SELECT
  EXTRACT(YEAR FROM checkout_completed_at) as check_out_year,
  EXTRACT(MONTH FROM checkout_completed_at) as check_out_month,
  ROUND((COUNT(CASE WHEN delivery_status = 'delivered' THEN 1 ELSE NULL END)+0.0)/COUNT(*)*100,2) AS delivery_rate
FROM
  modealto.checkout_attempts ca
JOIN
  modealto.order_status os
USING (order_id)
GROUP BY
  EXTRACT(YEAR FROM checkout_completed_at),
  EXTRACT(MONTH FROM checkout_completed_at)
ORDER BY
  EXTRACT(YEAR FROM checkout_completed_at),
  EXTRACT(MONTH FROM checkout_completed_at)
```

      - 2023-11: 93.36%
      - 2023-12: 94.12%
      - 2024-01: 91.12%

- **Other General Metrics:**
  - **Average Order Value:**
    - Formula: Revenue / Number of delivered orders
      - Breakdown by weather non-med was added

```sql
SELECT
  AVG(CASE WHEN customer_added_non_med_to_order = TRUE THEN order_total_dollars ELSE NULL END) as non_med_added_avg_order_value,
  AVG(CASE WHEN customer_added_non_med_to_order = FALSE THEN order_total_dollars ELSE NULL END) as non_med_not_added_avg_order_value
FROM
  modealto.checkout_attempts ca
JOIN
  modealto.order_status os
USING(order_id)
WHERE
  delivery_status = 'delivered';
```

        - Added: ≈ $30
        - Not Added: ≈ $20
  - **Monthly Average Revenue**
    - Formula: Sum of order_total_dollars (successfully delivered order)
  - **Number of successful Order**
    - Formula: Total Number of successfully delivered order

```sql
SELECT
  EXTRACT(YEAR FROM checkout_completed_at) as check_out_year,
  EXTRACT(MONTH FROM checkout_completed_at) as check_out_month,
  SUM(order_total_dollars) as Revenue,
  COUNT(*) as Number_of_Order
FROM
  modealto.checkout_attempts ca
JOIN
  modealto.order_status os
USING(order_id)
WHERE
  delivery_status = 'delivered'
GROUP BY
  EXTRACT(YEAR FROM checkout_completed_at),
  EXTRACT(MONTH FROM checkout_completed_at)
ORDER BY
  EXTRACT(YEAR FROM checkout_completed_at),
  EXTRACT(MONTH FROM checkout_completed_at)
```

      - Breakdown by month
        - 2023-11: Revenue: $50145, Number of Order: 2518
        - 2023-12: Revenue: $55172, Number of Order 2639
        - 2024-01: Revenue: $31662, Number of Order: 1508

Questions

1) How would you evaluate the potential impact of the change being proposed by the Product Team?

- We will estimate the Revenue increase by the change. Let consider the Month of December 2024 since we don't have the full month data for January 2025. In December 2023, the metrics is as follow:
    - **Total Number of Order:** 2639
    - **Non-Medication Items Add-to-Cart Rate (After Checkout):** %7.7
    - **Revenue:** $55172
    - **Avg order value:**
        - Non-med Added: ≈ $30
        - Non-med Not Added: ≈ $20

- Let's assume there's change in any other metrics from December 2023 to February 2024 to isolate the impact from the change
    - **Total Number of Order** = 2639
    - **Non-Medication Items Add-to-Cart Rate (After Checkout):**
      = Target Non-Medication Items Add-to-Cart Rate (Before Checkout) / Avg Checkout rate (Calculated above)
      = 20% / 60% = 30%

  → Therefore, we can calculate number of or with and without Non-med
        - Number of order with Non-med: 2639*30% ≈ 791 (orders)
        - Number of order without Non-med: 2639 - 791 = 1848 (orders)

  → Therefore, the estimate revenue will be
        - 791*30+1848*20 ≈ $60690

  → % Increase = ($60690- $55172)/$55172*100 ≈ %10

  → Therefore, the monthly revenue will increase by ≈ %10 solely by the feature change

2) Suppose the Product Team proceeds with implementing the proposed change. How would you evaluate whether the change is a success?

In short, we will create a simple hypothesis test, not necessary an A/B test since we have already identify our base line metrics. Let's consider success as **20% or higher in the add-to-cart rate for non-medication items (Checkout attempts)**

1. **Step 1: Formulate Hypothesis**
    - **Null Hypothesis (H_0):** add-to-cart rate for non-medication items (Checkout attempts) <= 20%
    - **Alternative Hypothesis (H_1):** add-to-cart rate for non-medication items (Checkout attempts) > 20%
2. **Step 2: Random Sampling**

- Select a random sample of users' add-to-cart rates for non-medication items to ensure unbiased data.

3. **Step 3: Choose Significance Level**
    - Set a significance level (e.g., $\alpha = 0.05$) to define the threshold for rejecting the null hypothesis.
4. **Step 4: Perform Hypothesis Test**
    - Conduct a one-tailed z-test or t-test to compare the sample's add-to-cart rate with the 20% baseline.
5. **Step 5: Interpret Results**
    - If p-value < α, reject H_0; otherwise, fail to reject H_0. Determine if the rate significantly exceeds 20%.