Parte 6.

**6. Ejecutar el experimento**

* Implementar la versión A y B a los usuarios seleccionados.
* Recoger los datos **sin manipularlos ni filtrar durante la ejecución**.
* **Duración**: Debe ser suficiente para alcanzar el tamaño de muestra calculado y cubrir ciclos de comportamiento (día de la semana, hora, estacionalidad).

**7. Monitorear sin hacer peeking**

* No analizar resultados antes de completar el tamaño de muestra, ya que esto **inflaría la tasa de falso positivo**.
* Monitorear solo por problemas técnicos, no por tendencias estadísticas.

**8. Analizar los datos**

Una vez finalizado el experimento:

1. **Calcular estadísticas descriptivas**:
   * Media o proporción de la métrica en cada grupo
   * Desviación estándar o varianza
2. **Elegir test estadístico adecuado**:
   * Para proporciones: **test Z para dos proporciones**
   * Para medias: **t-test para dos muestras independientes**
3. **Calcular estadístico de prueba**:
   * **Z-test para proporciones**:

Z=p^B−p^Ap^(1−p^)(1nA+1nB)Z = \frac{\hat{p}\_B - \hat{p}\_A}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n\_A} + \frac{1}{n\_B}\right)}}Z=p^​(1−p^​)(nA​1​+nB​1​)​p^​B​−p^​A​​

donde p^=xA+xBnA+nB\hat{p} = \frac{x\_A + x\_B}{n\_A + n\_B}p^​=nA​+nB​xA​+xB​​ es la proporción combinada.

1. **Obtener p-value**: Probabilidad de observar un efecto igual o más extremo bajo H₀.
2. **Comparar con α**:
   * Si **p ≤ α → rechazas H₀** (hay evidencia de efecto)
   * Si **p > α → no rechazas H₀** (no hay evidencia suficiente)

**9. Interpretar resultados**

* **Significativo**: La versión B es diferente de A según tu métrica y umbral de α.
* **No significativo**: No hay evidencia suficiente para afirmar que B sea mejor o peor.

Nota: No significa que la diferencia sea cero, solo que no es detectable con tu experimento.

**10. Reportar resultados**

Un buen reporte incluye:

1. Objetivo y KPI principal
2. Hipótesis (H₀ y H₁)
3. Tamaño de muestra y cómo se calculó
4. Detalles de aleatorización y ejecución
5. Resultados estadísticos:
   * Medias/proporciones de cada grupo
   * Diferencia observada
   * Intervalo de confianza (95%)
   * Estadístico de prueba y p-value
6. Conclusión y recomendaciones
7. Posibles limitaciones