Spotify Personalized Playlist Algorithm

**Problem :**  
The product team believes the new algorithm will increase user engagement by making people listen to more music.

**Null and Alternative Hypotheses:**

* **H₀ (Null Hypothesis):** The new algorithm does **not** increase weekly listening time.
* **H₁ (Alternative Hypothesis):** The new algorithm **increases** weekly listening time.

**Reflection Question:**

1. What makes this initial request problematic for hypothesis testing?

The request is unclear because it doesn’t specify exactly how much of an increase matters as meaningful, like what does meaningful increase in listening looks like in mathematically form 5% or 20% increase

1. How would you refine this into a specific, measurable research question?

Does the new playlist algorithm increase average weekly listening time per user by at least X minutes compared to the current system

1. What additional information would you need from the product team?

Target user segments, and the timeframe for measurement

1. Would a one-sided test or two-sided test be best in this scenario? Why?

Appropriate because we only care if listening time increases.

1. What are the potential risks of choosing a one-sided test?

If a **one-sided test** only looks for an **increase** in listening time, but the algorithm reduces listening time, then the test **fails to detect a real effect in the opposite direction.** This is called a **Type II error**, also known as a **false negative.**

**Key Consideration:**

The platform has 100 million active users worldwide across different subscription types (free and premium). Given here is Spotify's user segmentation:

* **Premium subscribers:** 40% of user base
* **Free tier users:** 60% of user base
* **Geographic distribution:** 30% US, 40% Europe, 30% Rest of World
* **Device usage:**70% mobile, 20% desktop, 10% other

**Reflection Question:**

1. What factors might affect listening time that need to be controlled for?

* User subscription types (free vs premium).
* Geographic location (US, Europe, Rest of World).
* Device usage (mobile, desktop, other).

1. How would you ensure representative sampling across user segments?

Randomly assign users to control (current algorithm)

1. What potential biases could emerge from improper sampling?

Non-random assignment and Over-representing a region could skew results.