**İş Problemi:**

Bir eğitim platformu, kurslara verilen puanları kullanarak kurs skorları elde ediyor ve bu puanları sıralamada kullanıyor. Ancak bir çalışan, kursları az izleyenlerin daha az puan verdiğini ve sağlıklı bir skorlamayı negatif etkilediğini iddia ediyor. Bu nedenle kursu izleme miktarının puanlamaya gerçekten bir etkisi olup olmadığı anlaşılarak buna göre aksiyon alınmak isteniyor.

Veriseti Örneği:

|  |  |  |  |
| --- | --- | --- | --- |
| Rating | Progress | Question Asked | Question Answered |
| 5 | 5 | 0 | 0 |
| 4.5 | 1 | 0 | 0 |
| 3.5 | 10 | 0 | 0 |
| 5 | 25 | 1 | 1 |
| 4 | 5 | 0 | 0 |

Not! :

Bu kıyaslama basitçe şöyle yapılabilir:

df[(df["Progress"] > 75)]["Rating"].mean()

= 4.86

df[(df["Progress"] < 25)]["Rating"].mean()

=4.72

Görüldüğü gibi kursun 75%’ inden fazlasını izleyenler 25%’inden azını izleyenlerden 0,14 puan daha yüksek oy ortalamasına sahip. Buradan direkt olarak “kursu daha çok izleyenler daha yüksek puan vermiştir, dolayısı ile kursu izleme oranına göre oylamaların katkısı ayarlanarak skor hesabı yapılmalıdır” sonucuna varamayız. Bu durum istatistiksel yöntemlerle şansa yer bırakmayacak şekilde analiz edilerek sonuca varılmalı ve aksiyon alınmalıdır.

**Çözüm :**

**Adım-1:** Puanlama yapanları 25%’ten az ve 75%’ ten fazla olmak üzere iki gruba ayırın.

**Adım-2:** Hipotezi kurun.

H0: M1 = M2 , iki grubun ortalamaları arasında anlamlı bir farklılık yoktur. Yani bir fark görünse bile rastlantısaldır, dikkate değmez.

H1: M1 != M2 , iki grubun ortalamaları arasında anlamlı bir fark vardır. Yani kursu izleme oranı arttıkça verilen puan değiştiğinden dolayı kursu izleme oranı yapılan puanlamaları değerlendirirken bir katsayı ile dikkate alınmalıdır.

**Adım-3:** Hangi istatistiksel testin yapılacağına karar vermek için grupların varyans homojenliği ve normallik kontrollerini yapın.

**Normallik kontrolü**:

İki grup **Shapiro-Wilk Testi** ile test edildi. Her iki grup da normal dağılıma sahip değildir. (p = 0.00 < 0.05)

**Varyans homojenliği kontrolü:**

İki grubun varyansları **Levene Testi** ile test edildi**.** Varyanslar homojen değildir. (p = 0.00 < 0.05)

Parametrik T-Testi'nin uygulanabilmesi için her iki grubun da normal dağılıma ve varyans homojenliğine sahip olması gerekir. Bu nedenle Parametrik Olmayan Test (Mann-Whitney U) uygulanacaktır.

**Adım-4:** Testi hipoteze uygulayın ve sonuçları yorumlayın.

Parametrik Olmayan Test sonucunda H0 reddedilir. Kursun çoğunu izleyenler ile kursun az bir kısmını izleyenlerin puanlamaları arasında anlamlı bir farklılık **vardır.**

**Sonuçların analizi:**

Kursun çoğunu izlemeyenlerin kursun çoğunu izleyenlerden puanlama olarak düşük olması, kursun içeriğinin bilinmeden puan verildiğini ve bu puanın kursu olduğundan kötü görünmesine neden olduğunu gösterir. Bu istatistiksel metotlarla ispat edilmiş sonuç, kursu az izleyenlerin çok izleyenlere kıyasla puanlamalarının daha az dikkate alınması gerektiğini gösterir. Bunun için izleme oranlarına göre katsayılar belirlenerek puanlamalar hesaplanabilir.

Örnek olarak;

* 0-24% için 0.20 ,
* 25-49% için 0.23 ,
* 50-74% için 0.27 ,
* 75-100% için 0.30

Katsayıları kullanılırsa izleme oranını puanlamaya dahil etmiş oluruz. (0.20 + 0.23 + 0.27 + 0.30 = 1)