Hypotheses testing: For Free and Paid Plan

Null Hypothesis (H₀)
Q4 2021 engagement is greater than or equal to Q4 2022 engagement
$H_0: \mu_1 \geq \mu_2$
Alternative Hypothesis (H₁): Q4 2022 engagement is less than Q4 2021 engagement
$H_1: \mu_1 < \mu_2$

F-Test Two-Sample for Variances For Paid Plans between 2021 and 2022

	minutes_watched_21	minutes_watched_22
Mean	332.502508	368.3547139
Variance	236063.3116	355699.1148
Observations	3433	5104
df	3432	5103
F	0.663660104	
P(F<=f) one-tail	0	
F Critical one-tail	0.949796198	

Interpretation:
The F-value of 0.6636 means that the variance in the first group (paid students for 2021) is smaller than the variance in the second group (paid students for 2022), because the F-value is less than 1.

The p-value of 0 indicates that there is a very significant difference in the variances between 2021 and 2022, strongly rejecting the null hypothesis of equal variances.

The F-critical value of 0.947 is the threshold for significance. Since 0.6636 (F-value) is

Conclusion:
1. For both paid and unpaid students, the p-values are very low (0), indicating that there is a significant difference in the variances between 2021 and 2022.
2. Paid students: The variance in 2022 is smaller than in 2021 (since F-value < 1).
3. Unpaid students: The variance in 2022 is larger than in 2021 (since F-value > 1)
Because both p-values are less than 0.05, you reject the null hypothesis of equal variances in both cases. The variances are significantly different for both groups.

F-Test Two-Sample for Variances For Free or UnPaid Plans between 2021 and 2022

	minutes_watched_21	minutes_watched_22
Mean	133.9333129	69.14765544
Variance	134881.7038	65343.34428
Observations	32171	120658
df	32170	120657
F	2.06419958	
P(F<=f) one-tail	0	
F Critical one-tail	1.014667161	

Interpretation:
The F-value of 2.064 means that the variance in the first group (unpaid students for 2021) is greater than the variance in the second group (unpaid students for 2022), because the F-value is greater than 1.

The p-value of 0 again indicates a very significant difference in the variances, strongly rejecting the null hypothesis of equal variances.

The F-critical value of 1.01 is the threshold for significance. Since 2.064 (F-value) is greater than 1.01 (F-critical), this confirms that the variances of unpaid students in 2021 and 2022 are significantly different.

t-Test: Two-Sample Assuming Unequal Variances For Paid Plans

	minutes_watched_21	minutes_watched_22
Mean	332.502508	368.3547139
Variance	236063.3116	355699.1148
Observations	3433	5104
Hypothesized Mean Difference	0	
df	8229	
t Stat	-3.046942872	
P(T<=t) one-tail	0.001159572	
t Critical one-tail	1.645038819	
P(T<=t) two-tail	0.002319144	
t Critical two-tail	1.960252308	

✅ Interpretation:
1. Using p-value method:
p = 0.0012 < 0.05
☞ Reject H₀
✅ This means there's strong evidence that engagement increased in Q4 2022 for paid users.

2. Using t-stat method:
t Stat = -3.047

Compare with -t Critical = -1.645

Since -3.047 ≤ -1.645,
☞ Again, reject H₀

🏆 Conclusion:
✅ Engagement among paid users significantly increased in Q4 2022 compared to Q4 2021.
💡 This supports the idea that the new features or changes had a positive impact on user engagement.

t-Test: Two-Sample Assuming Unequal Variances For free Plans

	minutes_watched_21	minutes_watched_22
Mean	133.9333129	69.14765544
Variance	134881.7038	65343.34428
Observations	32171	120658
Hypothesized Mean Difference	0	
df	40836	
t Stat	29.77523819	
P(T<=t) one-tail	4.7441E-193	
t Critical one-tail	1.644890942	
P(T<=t) two-tail	9.4881E-193	
t Critical two-tail	1.960022079	

✅ Interpretation:
1. Using p-value method: p-value ≈ 0 < 0.05
☞ Reject H₀
✅ There is very strong statistical evidence that engagement in Q4 2022 is higher than in Q4 2021 for free users.

2. Using t-stat method: t Stat = 29.775 > -1.645
Since this is a left-tailed test, this result is actually in the fail to reject H₀ zone.

But here's the important thing: Your t-statistic is extremely positive, which means the mean in 2021 is much greater than in 2022, not the other way around.

😕 Correction: There's a sign conflict
Since: t Stat = positive   And we are testing whether 2022 > 2021 (which would result in a negative t-stat if true),
But we got a large positive t-stat, that means 2022 < 2021