## Descriptive Statistics

# Mean, Median, & Standard deviation

* **Paid-plan Students**



**Mean**: Among students who watched between 1 and 100 minutes in 2021, the average minutes watched by paid-plan students increased significantly from Q4 2021 to Q4 2022, from approximately 33.80 minutes to about 273.02 minutes. This suggests a substantial increase in engagement among this group of initially low-engagement-paid-plan students.

**Median**: The median minutes these low-engagement-paid-plan students watched increased from Q4 2021 to Q4 2022, from 26.33 minutes to 40.28 minutes. While this increase is not as dramatic as the increase in the mean, it indicates that the typical student in this group (i.e., the student in the middle of the distribution) also increased their engagement. This suggests that the increase in engagement was more widespread among paid-plan students and not solely driven by a few outliers.

**Standard Deviation**: The standard deviation for these low-engagement-paid-plan students increased substantially from 28.21 minutes in Q4 2021 to 854.58 minutes in Q4 2022. This indicates a much larger variability in the minutes watched by these students in Q4 2022 compared to Q4 2021. This could be due to a broader range of engagement levels among the students in Q4 2022, with some students watching very little content and others watching a lot of content.

These results suggest that paid-plan students who were initially low-engagement in 2021 significantly increased their engagement in 2022. But the increased standard deviation indicates a broader range of engagement levels among these students in 2022. Understanding the reasons behind this variability could provide valuable insights for further boosting engagement. For instance, the factors that motivated the students who significantly increased their engagement might be leveraged to encourage increased engagement among other students

* **Free-Plan Students**



**Mean:** Among students who watched between 1 and 100 minutes in 2021, the average minutes watched by free-plan students increased from about 25.39 minutes in Q4 2021 to about 117.64 minutes in Q4 2022. This suggests that overall engagement among these initially low-engagement-free-plan students increased during this period. But the extent of this increase is less than what was observed for similar low-engagement-paid-plan students, suggesting that while these free-plan students are watching more content, they're still not as engaged as the equivalent group of paid-plan students.

**Median:** Interestingly, the median minutes watched by these low-engagement-free-plan students decreased from Q4 2021 to Q4 2022, from 14.17 minutes to 11.83 minutes. This indicates that engagement decreased for the typical student in this group (i.e., the student in the middle of the distribution). The increase in the mean might be driven by a small number of free-plan students who significantly increased their engagement in Q4 2022, while the majority did not increase their engagement or even reduced it.

**Standard Deviation:** The standard deviation for the low-engagement-free-plan students increased from 26.23 minutes in Q4 2021 to 468.93 minutes in Q4 2022. This indicates a more significant variability in the minutes watched by these students in Q4 2022 compared to Q4 2021. The behavior of these students then became more diverse in Q4 2022, with some watching a lot of content and others watching very little.

These results suggest a complex picture for the initially low-engagement-free-plan students. While the mean minutes watched increased—signifying an increase in overall engagement—the median minutes watched decreased, indicating that the typical student in this group did not increase their engagement. This discrepancy and the increased standard deviation suggest that a small number of students within this group might significantly increase their engagement while the majority did not. This might imply the need for targeted strategies to boost engagement among the broader population of initially low-engagement-free-plan students

* **Paid vs Free-Plan Students**

On average, low-engagement-paid students initially increased their watching time more significantly than the free-plan students from Q4 2021 to Q4 2022. This could suggest that paid-plan students find more value in the platform, possibly due to premium features or content that are available to them. In contrast, the median watch time decreased for free-plan students, suggesting that the typical free-plan student in this group did not increase their engagement. This discrepancy might indicate that the strategies or features designed to increase engagement are more effective for paid-plan students. It could also suggest that the monetary investment leads to increased usage due to a desire to get their money's worth.

Based on the findings, the platform is more successful in increasing engagement among students who make a monetary investment (i.e., paid-plan students). But the increased variability, especially among paid-plan students, indicates that there are likely differences in how individual students are responding to the platform's offerings. Therefore, personalized approaches might be beneficial in boosting engagement, and further analysis could help understand the factors that drive increased engagement among paid- and free-plan students.

# Skewness and Kurtosis

For **paid-plan students**, the skewness increased from 0.63 in Q4 2021 to 7.07 in Q4 2022.



The skewness for **free-plan students** increased from 1.17 in Q4 2021 to 15.06 in Q4 2022, indicating positive skewness.



**Positive skew (right-skew)** occurs when the data is not symmetrical around the mean, forming a long tail on its right side. This signifies that most of the distribution's observations are concentrated to the left of the peak. Positive skewness can have several implications.

The mean is larger than the median in a right-skewed distribution because the distribution tail pulls the mean to the right. This observation is confirmed by the mean and median values in the two years. An increasing skewness suggests that more students watch significantly more content than most over time, pulling the mean upwards.

In both cases, the mean is higher than the median (33.80 > 26.33 in 2021 and 273.02 > 40.28 in 2022).

As a result, the mean is no longer a good central tendency indicator, and it cannot accurately reflect the typical value of the dataset. Note that skewness tells us the direction of outliers but doesn’t indicate the number that occurs.

For **paid-plan** students, the kurtosis increased from -0.85 in Q4 2021 to 58.48 in Q4 2022.



The kurtosis increased for **free-plan** students—from 0.36 in Q4 2021 to 315.76 in Q4 2022.



Kurtosis values greater than 0 indicate that the data has heavier tails and a sharper peak than the normal distribution (leptokurtic). A leptokurtic distribution has a high positive kurtosis, suggesting that it’s very peaked and has a relatively large number of outliers. This type has a higher frequency of extreme values or outliers. The increase in kurtosis over time suggests more extreme cases in the data in Q4 2022 than in Q4 2021, particularly for free-plan students.

Overall, the increasing skewness and kurtosis for both groups from Q4 2021 to Q4 2022 suggest a growing number of students watching significantly more content than the majority. This is especially true for free-plan students with a higher skewness and kurtosis in Q4 2022 than paid-plan students.

## Confidence Intervals

* **Paid-Plan Students**:



For paid-plan students, there's an increase in engagement from Q4 2021 to Q4 2022. The confidence interval for the average minutes watched by paid-plan students increased from Q4 2021 (316.25 to 348.76 minutes) to Q4 2022 (351.99 to 384.72 minutes). This suggests that we can be 95% confident that the true average minutes watched by all paid-plan students in the population increased from Q4 2021 to Q4 2022.

* **Free-Plan Students**:



Among free-plan students, there's a decrease in engagement from Q4 2021 to Q4 2022. The confidence interval for the average minutes watched decreased from Q4 2021 (129.92 to 137.95 minutes) to Q4 2022 (67.71 to 70.59 minutes). We then can be 95% confident that the true average minutes watched by all free-plan students in the population decreased from Q4 2021 to Q4 2022.

* **Comparison between Paid- and Free-Plan Students (Q4 2022).**

Students with a paid-plan subscription watch substantially more than those without. The confidence interval for the average minutes watched in Q4 2022 was 61.71 to 70.59 minutes for free-plan students and 351.99 to 384.72 minutes for paid-plan students. We then can be 95% confident that paid-plan students watched significantly more minutes than free-plan students in Q4 2022. This aligns with the expectation that paid-plan students who have invested in the platform tend to be more engaged than free-plan users.

It should be noted that these are just interpretations based on the confidence intervals. Actual cause-effect relationships must be examined further to understand the causes behind these engagement changes.

The fact that paid-plan subscribers watch more doesn't necessarily mean that having a paid-plan subscription encourages them to watch more. For example, the higher engagement among paid-plan students may be due to the additional features or content available or because more engaged students are more likely to choose a paid-plan subscription.

Similarly, the decrease in engagement among free-plan students could be due to various factors, such as changes in the platform, competition from other platforms, or changes in the user base.

## Hypothesis Testing

# Q4/21 VS Q4/22

Given a null hypothesis that states: The engagement (minutes watched) in Q4 2021 is higher than or equal to the one in Q4 2022 i.e. (μ1 ≥ μ2), the hypothesis was tested for free-plan and paying students separately at a confidence level of 95%.

* **Paid-Plan Students**:

Using the Two-Sample f-Test for variances to prove that assumption of unequal variances between the samples for paid-plan subscribers:



The p-value indicates the probability of obtaining the observed f-value if the null hypothesis (equal variances) were true. The sample variances are not identical since the p-value is 0.

Next, I used a left-tailed t-test assuming unequal variances for paying students.



Decision Rule: If p−value ≤ 0.05, Reject H0 (null hypothesis)

**Conclusion:** **Reject**because the p-value is lower than the specified significance level α (0.05).

**Summary:** With a t-statistic of -3.05 (less than the critical value of -1.645), you would reject the null hypothesis because the negative t-statistic indicates that (the mean minutes watched by students in Q4 2021) is significantly smaller than (the mean minutes watched by students in Q4 2022). This is contrary to the null hypothesis, so we reject it. Of course, rejecting the null hypothesis does not confirm the alternative hypothesis; it only suggests that the data provided enough evidence against the null hypothesis.

* **Free-plan students:**

Using the Two-Sample f-Test for variances to prove that assumption of unequal variances between the samples for Free-plan students:



Since the p-value is 0, The sample variances are not identical.

Using a left-tailed t-test assuming unequal variances for free-plan students;



**Conclusion:** **Fail to Reject**because the calculated **t-statistic** is higher than the critical value.

For free-plan students: With a t-statistic of 29.78 (greater than the critical value of -1.645), you would fail to reject the null hypothesis. This means there’s not enough evidence to conclude that (μ1) is smaller than (μ2). So, the data supports the null hypothesis that (μ1) is larger than or equal to (μ2).

These results align with previous findings from the confidence intervals and further underscore the difference in engagement patterns between paid- and free-plan students.

# US VS India (Free-plan students)

Estimating the similarity in the number of minutes watched between the free-plan students in the US and in India, given a null hypothesis that states: The engagement (minutes watched) in the US is higher than or equal to that in India (μUS ≥ μIN) at a confidence level of 95%.

Using the Two-Sample f-Test for variances to prove the assumption of unequal variances between the samples for free-plan students in both countries:

.

The sample variances are not identical since the p-value is 0.

Using a left-tailed t-test assuming unequal variances:



**Conclusion:** **Fail to** **Reject** because the calculated **t-statistic** is higher than the critical value and the p-value is higher than the specified significance level α (0.05).

If the hypothesis that US students watch more or an equal amount of content as Indian students is rejected, this suggests that US students watch less content on average than students in India.

This could have the following implications.

**Market Differences:**These details might indicate that the platform is more engaging or relevant to students in India than e US students. Understanding the reasons behind this could be valuable. Are the platform’s specific features, content, or aspects particularly appealing to Indian students? Such questions need to be addressed further but are beyond the scope of this analysis.

**Growth Opportunities:** If US engagement is lower, this could represent a growth opportunity. The 365 company might seek ways to increase engagement among US students, involving marketing efforts, adding more content relevant to US students, or other strategies.

**Resource Allocation**: This information could be helpful when deciding where to allocate resources. For example, if Indian students are more engaged, investing in more content and features targeted toward this audience might make sense.