

Customer Engagement Analysis in Excel Project

Introduction

The 365 Company embarked on an ambitious journey in 2022, introducing numerous new features to its website platform aimed at fostering student engagement and driving company growth. Among these enhancements were an XP system, in-app coins, leaderboards, and streaks, designed to motivate and reward students for consistent learning habits. Additionally, an expanded course library broadened the scope of available topics. The company anticipated that these improvements would significantly enhance the student experience, contribute to effective customer engagement, and propel the company towards success in the coming year.

This portfolio is dedicated to the Customer Engagement Analysis in Excel project, a comprehensive investigation into the impact of the new platform features on student engagement. The dataset provided by the 365 Company is analyzed to draw meaningful insights into user behavior, providing a realistic and relevant context for the analysis.

Problem Statement

The central question addressed in this project is whether the introduced website features succeeded in boosting student engagement in Q4 2022 compared to Q4 2021. To delve into this, various data analysis techniques, statistical tests, and Excel functions will be employed. The specific tasks include:

Task 1: Analyzing the engagement levels of low-engagement users in Q4 2021 and Q4 2022, comparing mean, median, and standard deviation for free- and paid-plan subscribers.

Task 2: Calculating skewness and kurtosis for students who watched content in Q4 2021 and Q4 2022, considering both paid- and free-plan students, and assessing the results in relation to mean and median values.

Task 3: Determining confidence intervals for four groups of students categorized by paid-plan subscription and engagement in Q4 2021 and Q4 2022. Drawing conclusions about changes in student engagement.

Task 4: Conducting hypothesis testing to assess whether the platform's new features contributed to increased engagement in Q4 2021 compared to Q4 2022 for both free-plan and paying students. Assessing the potential costs associated with Type I and Type II errors.

Task 5: Investigating whether the average number of minutes watched in the US is similar to that in India for free-plan students in 2022, utilizing two-sample t-tests and optional f-tests for variances.

Task 1: Low-engagement User Analysis

The 'Task 1 and 2' sheet contains the list of users who watched between 1 and 100 minutes in 2021 for both Paid-plan and Free-plan. We will focus on low-engagement users (those who watched between 1 and 100 minutes in 2021) as low-engagement users often represent the most significant potential for growth. If 365 company can find ways to increase its usage, it could significantly impact the overall use of the platform.

Paid-Plan Students				Free-plan Students			
student_id	paid	minutes_watched_21	minutes_watched_22	student_id	paid	minutes_watched_21	minutes_watched_22
16979	1	13.32	260.72	238865	0	1.43	157.28
207114	1	40.12	387.98	247592	0	3.1	0.1
156680	1	17.57	128.78	195373	0	8.45	12.57
149601	1	42.95	7417.4	229324	0	44.87	1
251499	1	4.92	10.47	198040	0	61.88	0.23
179664	1	45.07	628.05	14672	0	55.05	114.17
145813	1	16.98	949.9	182954	0	3.13	0.07
160274	1	61.97	2480.43	245547	0	63.03	28.9
9305	1	72.33	715.95	37976	0	35.17	30.28
211124	1	1.12	5.7	231774	0	48.85	0.05
172631	1	88.8	217.65	180503	0	16.02	6.3
240248	1	42.17	392.63	117871	0	2.48	0.07
233048	1	12.3	47.92	87495	0	3.9	0.05
1436	1	4.88	1045.55	186757	0	8.02	57.82
150663	1	41.52	253.78	247599	0	20.48	20.42
156542	1	19.33	9214.13	225846	0	66.57	24.17
862	1	57.8	162.23	59019	0	6.25	0.13
240591	1	77.25	293.13	228556	0	3.65	137.2
230669	1	72.52	230.67	172658	0	97.2	13.03
153036	1	92.7	65.33	246855	0	7.12	714.63
226445	1	21.3	17.37	103148	0	7.23	0.68
156941	1	4.15	7.77	212456	0	80.43	2.53

Paid-Plan Students		
	minutes_watched_21	minutes_watched_22
Mean	33.80	273.02
Median	26.33	40.28
Std Dev	28.21	854.58

Free-plan Students		
	minutes_watched_21	minutes_watched_22
Mean	25.39	117.64
Median	14.17	11.83
Std Dev	26.23	468.93

Analysis for Paid Plan Users:

- **Mean:** The average minutes watched by paid-plan users in Q4 2021 is 33.80 minutes. There is a substantial increase in the average engagement to 273.02 minutes for paid-plan users in Q4 2022. This implies a significant rise in involvement within the group of students who initially had low engagement and were enrolled in paid plans.

- **Median:** Analyzing the engagement of low-engagement-paid-plan students from Q4 2021 to Q4 2022, we see their typical watch time increased from 26.33 to 40.28 minutes. While this rise in the middle value isn't as big as the overall average increase, it suggests most students in this group upped their engagement. This points to a widespread improvement, not just a few students doing exceptionally well. The starting median of 26.33 minutes indicates a bit of unevenness, likely because a few students had much higher viewing times. Even though the median is lower than the average at 40.28, it tells us outliers still have an impact, making the engagement increase a mix of typical and extreme values in the group.
- **Standard Deviation:** The variation in engagement levels among low-engagement-paid-plan students saw a notable increase, with the standard deviation rising from 28.21 minutes in Q4 2021 to 854.58 minutes in Q4 2022. This substantial leap indicates a much broader range of minutes watched by these students in the Q4 2022 compared to Q4 2021. The initial standard deviation of 28.21 in Q4 2021 pointed to a moderate variability in engagement levels. However, the significant surge to 854.58 in Q4 2022 suggests a more extensive spread, likely influenced by a wider range of engagement levels, including extreme values.

These findings show that paid-plan students who weren't very engaged in 2021 boosted their engagement in 2022. However, the higher standard deviation suggests a wider range of engagement levels among these students in 2022. Further investigation on why some students increased engagement a lot could offer insights to enhance overall engagement. Using the motivations of those who significantly increased engagement might help encourage others to get more involved too.

Analysis for Free Plan Users:

- **Mean:** Free-plan users had an average engagement of 25.39 minutes in Q4 2021. The mean engagement for free-plan users substantially increased to 117.64 minutes in Q4 2022. This shows that low-engagement-free-plan students became more engaged over time. However, their increase isn't as much as low-engagement-paid-plan students, indicating that although free-plan students are watching more, they're still not as engaged as their paid-plan counterparts.
- **Median:** It was noted that the middle point of engagement, represented by the median minutes watched by low-engagement-free-plan students, dropped from 14.17 minutes in Q4 2021 to 11.83 minutes in Q4 2022. This signals a decrease in engagement for the typical student within this group. The significant rise in the mean, however, could be influenced by a small group of free-plan students who significantly increased their engagement in Q4 2022, while the majority either maintained or decreased their engagement levels. The initial lower median of 14.17 in Q4 2021 indicates a distribution that leans towards higher engagement for some users. Similarly, the lower median

compared to the mean at 11.83 in Q4 2022 suggests the presence of potential outliers or a distribution that skews towards higher engagement levels, reinforcing the idea that the mean increase may be driven by a minority of users.

- **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. A moderate standard deviation of 26.23 indicates variability in engagement levels among free-plan users in Q4 2021. The higher standard deviation of 468.93 reflects increased variability in engagement levels among free-plan users in Q4 2022.

Even though the average time spent watching increased, showing an overall engagement boost, the typical student, represented by the median, actually watched less. This difference, along with a higher standard deviation, suggests that a few students increased their engagement a lot, while most didn't. This points to the potential necessity of specific strategies to increase engagement for the larger group of low-engagement-free-plan students.

Overall Implications:

Paid-plan students, on average, significantly increased their watching time more than free-plan students from Q4 2021 to Q4 2022. This suggests that paid-plan students might find more value in the platform, possibly due to premium features or content available to them. In contrast, the median watch time for free-plan students decreased, indicating that the typical student in this group did not increase their engagement. This difference could mean that strategies or features aimed at boosting engagement are more effective for paid-plan students, or it might indicate that the financial investment motivates increased usage to get their money's worth.

The platform appears more successful in increasing engagement among paid-plan students. However, the increased variability, especially among paid-plan students, suggests differences in how individual students respond to the platform. Personalized approaches might be useful for boosting engagement, and further analysis could uncover the factors driving increased engagement among both paid- and free-plan students.

Task 2: Skewness and Kurtosis Analysis

Skewness measures the asymmetry of the watch time distribution. A positive skewness indicates a longer tail on the right side (more outliers with higher watch times), while negative skewness signifies a longer tail on the left (more outliers with lower watch times).

Kurtosis measures represents the "peakedness" of the distribution. High kurtosis suggests a sharp peak with relatively few outliers, while low kurtosis indicates a flatter distribution with more frequent outliers. If a distribution is heavy-tailed—i.e., more data in the tails—it exhibits high

kurtosis. Meanwhile, a low kurtosis occurs when the data is more evenly distributed between the tails and the center or the distribution is light-tailed.

Paid-Plan Students		
	minutes_watched_21	minutes_watched_22
Skewness	0.63	7.07
Kurtosis	-0.85	58.48

Free-plan Students		
	minutes_watched_21	minutes_watched_22
Skewness	1.17	15.06
Kurtosis	0.36	315.76

Analysis for Paid Plan Users:

Q4 2021:

- **Skewness:** For paid-plan students, the skewness increased from 0.63 in Q4 2021 to 7.07 in Q4 2022. A positive skewness of 0.63 indicates that the distribution is slightly skewed to the right, with a tail on the positive side. This aligns with the interpretation from Task 1, where the median was lower than the mean, suggesting the presence of higher engagement outliers. Similarly, in Q4 2022, a significantly positive skewness of 7.07 indicates a highly skewed distribution to the right. This supports the findings from Task 1, where there was a substantial increase in the mean, potentially driven by extreme values.
- **Kurtosis:** For paid-plan students, the kurtosis increased from -0.85 in Q4 2021 to 58.48 in Q4 2022. A negative kurtosis of -0.85 suggests a flatter distribution with fewer extreme values than a normal distribution. Similarly, in Q4 2022, a very high positive kurtosis of 58.48 suggests a distribution with heavy tails and more extreme values than a normal distribution. The distribution is highly peaked.

Analysis for Free Plan Users:

- **Skewness:** For free-plan students, the skewness increased from 1.17 in Q4 2021 to 15.06 in Q4 2022. A positive skewness of 1.17 indicates a distribution skewed to the right. This aligns with the interpretation from Task 1, where the median was lower than the mean. An extremely high positive skewness of 15.06 indicates a distribution highly skewed to the right. This aligns with the significant increase in mean and the presence of extreme values observed in Task 1.
- **Kurtosis:** For free-plan students, the kurtosis increased from 0.36 in Q4 2021 to 315.76 in Q4 2022. A positive kurtosis of 0.36 suggests a distribution with a moderate peak and heavier tails than a normal distribution. Similarly, in Q4 2022, the very high kurtosis of 315.76 indicates a distribution with extremely heavy tails and an exceptionally peaked shape.

Overall Implications:

The skewness and kurtosis results support the findings from Task 1, indicating the presence of extreme values in Q4 2022 for both paid- and free-plan users. The high kurtosis values suggest a small group of users contributing to a large portion of the overall watch time, explaining the discrepancies between mean and median, particularly for paid users in Q4 2022.

In both groups, more students started watching a lot more content from Q4 2021 to Q4 2022, as shown by the increasing skewness and kurtosis. This trend is more pronounced among free-plan students in Q4 2022 compared to paid-plan students.

Task 3: Confidence Intervals for Engagement Levels

Confidence intervals provide a range of values where we can reasonably expect the true population mean to lie. Non-overlapping confidence intervals indicate statistically significant differences in mean engagement levels between the specified time periods. Assuming normality due to large sample sizes (>30), the z-score for a two-tailed 95% confidence interval is 1.96. The z-score helps determine the critical value beyond which we can reject the null hypothesis. The margin of error provides a range within which the true population mean is likely to fall.

Paid-Plan Students		
	minutes_watched_21	minutes_watched_22
n	3433	5104
Mean	332.50	368.35
Std Dev	485.86	596.41

Free-plan Students		
	minutes_watched_21	minutes_watched_22
n	32171	120658
Mean	133.93	69.15
Std Dev	367.26	255.62

Paid-Plan Students		
	minutes_watched_21	minutes_watched_22
Significance level	0.05	0.05
z-Score	1.96	1.96
Standard Error	8.29	8.35
Margin of error	16.25	16.36
Lower bound	316.25	351.99
Upper Bound	348.76	384.72

Free-plan Students		
	minutes_watched_21	minutes_watched_22
Significance level	0.05	0.05
z-score	1.96	1.96
Standard Error	2.05	0.74
Margin of error	4.01	1.44
Lower bound	129.92	67.71
Upper Bound	137.95	70.59

Analysis for Paid Plan Users:

Paid-plan students engaged more from Q4 2021 to Q4 2022. The average minutes watched by paid-plan students went up. The 95% confidence interval for the mean engagement level in Q4 2021 for paid-plan users is [316.25, 348.76]. This means we are 95% confident that the true mean engagement level falls within this range. Similarly, the 95% confidence interval for the mean engagement level in Q4 2022 for paid-plan users is [351.99, 384.72]. Similar to Q4 2021, we are 95% confident that the true mean engagement level falls within this range. Moreover, the non-overlapping confidence intervals suggest a significant increase in mean engagement levels for paid-plan users in Q4 2022 compared to Q4 2021.

Analysis for Free Plan Users:

Among free-plan students, there's a decrease in engagement from Q4 2021 to Q4 2022. The 95% confidence interval for the mean engagement level in Q4 2021 for free-plan users is [129.92, 137.95]. This indicates that we are 95% confident that the true mean engagement level falls within this range. The 95% confidence interval for the mean engagement level in Q4 2022 for free-plan users is [67.71, 70.59]. Once again, we are 95% confident that the true mean engagement level falls within this range. The non-overlapping confidence intervals suggest a significant decrease in mean engagement levels for free-plan users in Q4 2022 compared to Q4 2021.

Overall Implications:

Paid-plan students watch much more than those without a subscription. In Q4 2022, free-plan students watched an average of 67.71 to 70.59 minutes, while paid-plan students watched 351.99 to 384.72 minutes, according to the confidence interval. With 95% confidence, we can say that paid-plan students watched significantly more than free-plan students in Q4 2022. This matches the expectation that those who invest in the platform through a paid plan tend to be more engaged than free-plan users.

A higher standard deviation (596.41 for paid plan users in Q4 2022 compared to Q4 2021) indicates a wider spread of watch times (engagement levels), with some students significantly exceeding or falling below the average. Introduction of new platform features in Q4 2022 might have attracted and motivated paid subscribers to engage more. Free users might not have found the new features as appealing or could be facing limitations due to their subscription type, leading to decreased engagement or due to competition from other platforms, or changes in the user base.

Task 4: Hypothesis Testing for Customer Engagement

A two-sample f-test for variances was performed to prove the assumption of unequal variances between the samples for free- and paid-plan subscribers:

F-Test Two-Sample for Variances

Paid-Plan Students		
	<i>minutes_watched_21</i>	<i>minutes_watched_22</i>
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	

F-Test Two-Sample for Variances

Free-Plan Students		
	<i>minutes_watched_21</i>	<i>minutes_watched_22</i>
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	

The p-value tells us how likely it is to get the observed f-value if the null hypothesis (equal variances) were correct. In both cases, the p-value is 0, showing that the sample variances are not the same.

Analysis for Paid Plan Users:

The mean, standard deviation, and sample size for paid-plan students were calculated.

Paid-Plan Students		
	<i>minutes_watched_21</i>	<i>minutes_watched_22</i>
n	3433	5104
Mean	332.50	368.35
Std Dev	485.86	596.41

Assuming a significance level (α) of 0.05, the critical t-value is approximately 1.65 (for a one-tailed test at a 5% significance level). The Two-Sample t-Test Assuming Unequal Variances with the Data Analysis ToolPak was performed to calculate the T-Statistic and degrees of freedom.

t-Test: Two-Sample Assuming Unequal Variances		
Paid-Plan Students		
	<i>minutes_watched_21</i>	<i>minutes_watched_22</i>
Mean	332.502508	368.3547139
Variance	236063.3116	355699.1148
Observations	3433	5104
Hypothesized Mean	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	

Null Hypothesis: The engagement (minutes watched) in Q4 2021 is higher than or equal to Q4 2022 ($\mu_1 \geq \mu_2$) for paid-plan users.

The t-statistic (-3.05) falls below the critical value (-1.645) and the p-value is less than the significance level (0.05), indicating statistically significant evidence to **reject** the null hypothesis. This suggests that engagement in Q4 2022 was significantly higher than in Q4 2021 for paid-plan users, but it doesn't confirm the alternative hypothesis entirely.

Analysis for Free Plan Users:

The mean, standard deviation, and sample size for free-plan students were calculated.

Free-plan Students		
	<i>minutes_watched_21</i>	<i>minutes_watched_22</i>
n	32171	120658
Mean	133.93	69.15
Std Dev	367.26	255.62

Assuming a significance level (α) of 0.05, the critical t-value is approximately 1.65 (for a one-tailed test at a 5% significance level). The Two-Sample t-Test Assuming Unequal Variances with the Data Analysis ToolPak was performed to calculate the T-Statistic and degrees of freedom.

t-Test: Two-Sample Assuming Unequal Variances

Free-Plan Students		
	<i>minutes_watched_21</i>	<i>minutes_watched_22</i>
Mean	133.9333129	69.14765544
Variance	134881.7038	65343.34428
Observations	32171	120658
Hypothesized Mean	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	

Null Hypothesis: The engagement (minutes watched) in Q4 2021 is higher than or equal to Q4 2022 ($\mu_1 \geq \mu_2$) for free-plan users.

For free-plan users, the t-statistic (29.78) falls significantly above the critical value (-1.645) and the p-value (4.74) is much greater than the significance level (0.05). This suggests that there is not enough evidence to reject the null hypothesis. Therefore, we fail to reject the null hypothesis, meaning the difference in engagement between Q4 2021 and Q4 2022 for free-plan users is not statistically significant.

Overall Implications:

The results align with our previous findings using confidence intervals, reinforcing the distinct engagement patterns between paid- and free-plan students. This emphasizes the importance of considering subscription plans when studying user engagement dynamics.

In addressing potential errors, a Type I error, or a false positive, occurs if we incorrectly assert that engagement increased in 2022 when it actually didn't. The likelihood of committing this error rests on the significance level (α) chosen for the hypothesis test—a decision solely within the researcher's control.

This significance level is closely linked to the confidence level, representing our confidence in the obtained results. A 5% significance level corresponds to a 95% confidence level, indicating a 5% chance of erroneously claiming increased engagement when it hasn't occurred.

Conversely, a Type II error, or a false negative, transpires if we fail to detect an actual increase in engagement in 2022. The implications of each error type are contingent on the context of the business, involving potential misjudgments in feature investments and the assessment of areas needing improvement.

Determining the cost associated with each error type is crucial, as it hinges on the consequences of either overinvesting based on incorrect conclusions or overlooking positive developments that demand attention. Striking the right balance ensures a nuanced interpretation of engagement trends, minimizing the impact of misjudgments on strategic decision-making.

Task 5: Two-Sample t-Test for Average Minutes Watched in the US vs. India (Free Plan Users - 2022)

A two-sample f-test for variances was performed to prove that assumption of unequal variances between the samples for free-plan subscribers of US and India.

Paid-plan Students		
	<i>minutes_watched_22_US</i>	<i>minutes_watched_22_IN</i>
Mean	73.07053569	78.42208628
Variance	95208.64187	101975.5527
Observations	6459	21210
df	6458	21209
F	0.933641833	
P(F<=f) one-tail	0.000347535	
F Critical one-tail	0.967314359	

The mean, standard deviation, and sample size for free-plan students of both US and India in 2022 were calculated.

Paid-Plan Students		
	<i>minutes_watched_22_US</i>	<i>minutes_watched_22_IN</i>
n	6459	21210
Mean	73.07	78.42
Std Dev	308.56	319.34

t-Test: Two-Sample Assuming Unequal Variances

	<i>minutes_watched_22_US</i>	<i>minutes_watched_22_IN</i>
Mean	73.07053569	78.42208628
Variance	95208.64187	101975.5527
Observations	6459	21210
Hypothesized Mean Difference	0	
df	11001	
t Stat	-1.210387573	
P(T<=t) one-tail	0.113078106	
t Critical one-tail	1.644992151	
P(T<=t) two-tail	0.226156213	
t Critical two-tail	1.960179649	

Null Hypothesis (H0): The average engagement (minutes watched) in the US is higher than or equal to that in India ($\mu_1 \geq \mu_2$).

The t-statistic (-1.21) which is below the critical value (-1.645) falls within the non-rejection region for a one-tailed test. The p-value (0.11) is also higher than the significance level (0.05). Therefore, we fail to reject the null hypothesis, meaning we do not have sufficient evidence to conclude that the average engagement in India is greater than the average engagement in the US for free-plan users.

Overall Implications:

The mean minutes watched in the US (73.07) and India (78.42) are relatively close, suggesting similar average engagement levels for free plan users in 2022. The small difference in means indicates that, on average, users in both countries spend a comparable amount of time on the platform.

While statistical significance is not observed, it's essential to consider practical significance. The small absolute difference in means may not have a substantial impact on the user experience or business outcomes.

Conclusion

The introduction of new platform features in Q4 2022 had a substantial impact on user engagement, with statistically significant changes observed across various user segments. While extreme values and variability were noted, suggesting diverse user behaviors, the overall trend indicated increased engagement. These findings provide valuable insights for the 365 Company to refine its strategies and continue enhancing the user experience based on observed user preferences and behaviors.