ABDELLAH EL AOUAD

CUSTOMER ENGAGEMENT ANALYSIS IN EXCEL PROJECT

DATA ANALYTICS PROJECT

CASE DESCRIPTION

In 2022, there were high expectations for the growth of the 365 company and increased student engagement based on the introduction of new website platform features. Some of these features included an XP system that enabled students to track their progress, level up, and earn rewards by completing various learning objectives.

Additionally, the company expanded its course library, covering a broader range of topics to provide its students with a richer set of skills and attract a larger audience. These enhancements were anticipated to positively impact the student experience, create an effective strategy for customer engagement, and contribute to the company's success in the coming year.

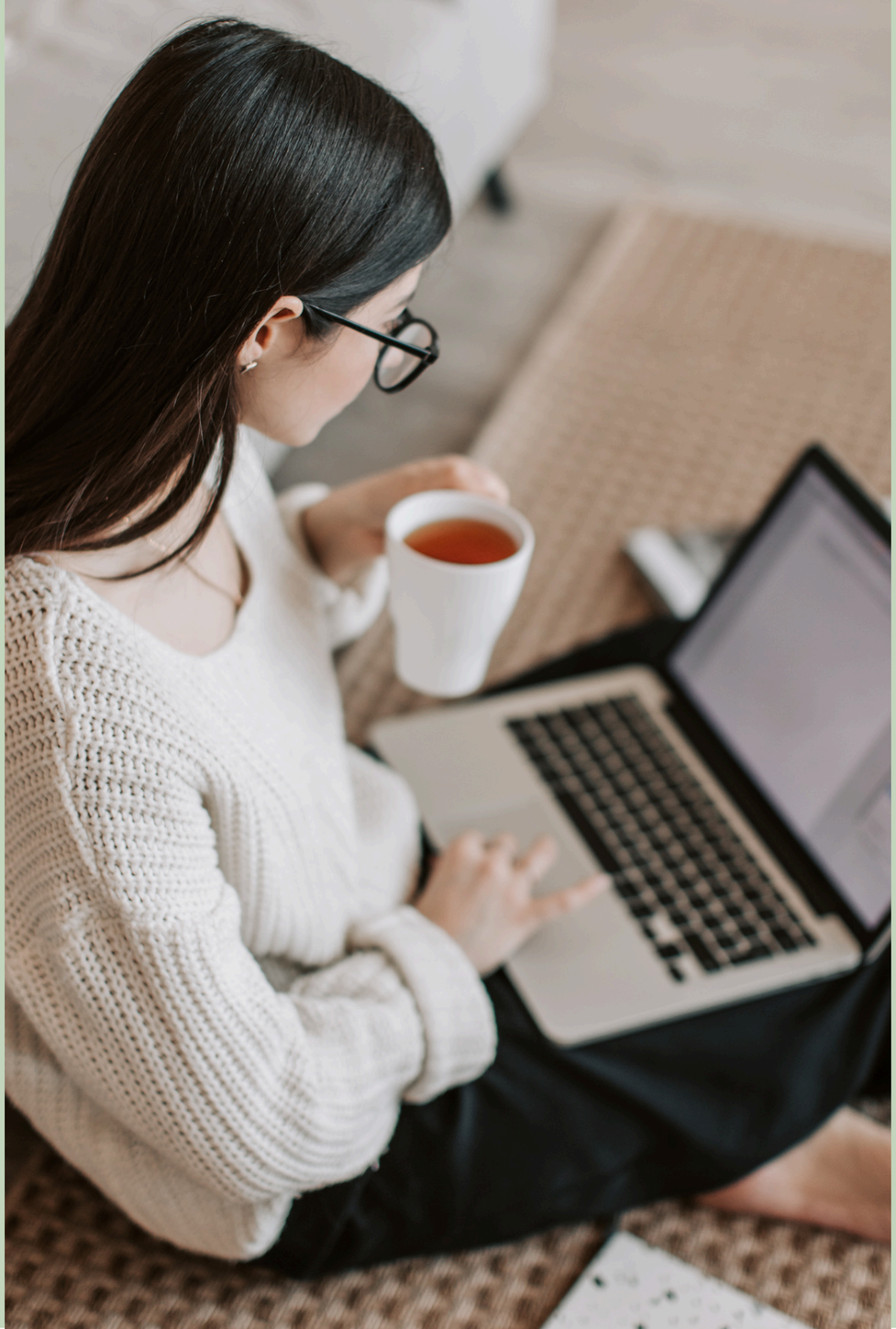
PROJECT'S GOAL

With this Customer Engagement Analysis in Excel, I must analyze whether the new additions to the platform have increased student engagement.

PROJECT FILES

The following information about the column values :

1. **student_id –** the unique identifier for each student in the dataset.
2. **student_country –** identifies the country of each student.
3. **Paid –** indicates whether a student had a paid account during the specified period. It is a binary variable.
4. **minutes_watched_21 –** represents the student's engagement level in Q4 2021.
5. **minutes_watched_22 –** denotes the student's engagement level in Q4 2022.



PROCESS OVERVIEW

To successfully complete this project, I followed a structured approach, which included the following key steps:

Descriptive Statistics

Confidence Intervals

Hypothesis Testing

Conclusion

DESCRIPTIVE STATISTICS

Task 1

My first task was to provide insights into the relative engagement levels in **Q4 2021** and **Q4 2022**. I did focus on low-engagement users (those who watched between **1** and **100** minutes in **2021**).

During this task, I used the following functions to provide the desired insights:

=AVERAGE() : Mean

=MEDIAN()

=STDEV.S() : Standard deviation

=SKEW() : Skewness

=KURT() : Kurtosis

	minutes_watched_21	minutes_watched_22
Mean		
Median		
Standard deviation		
Skewness		
Kurtosis		

RESULTS 1

Paid-Plan Students

	minutes_watched_21	minutes_watched_22
Mean	33.79695842	273.0205033
Median	26.33	40.28
Standard deviation	28.213543	854.5752369
skewness	0.625157717	7.065595612
kurtosis	-0.849282209	58.4849581

Free-plan Students

	minutes_watched_21	minutes_watched_22
Mean	25.39167554	117.6374576
Median	14.17	11.83
Standard deviation	26.2315369	468.9346649
skewness	1.16773938	15.06167594
kurtosis	0.360033015	315.7647882

Paid-Plan Students (Low Engagement Group)

- *Mean: Increased from 33.80 to 273.02 minutes*
- *Median: Increased from 26.33 to 40.28 minutes*
- *Standard Deviation: Increased from 28.21 to 854.58 minutes*

Free-Plan Students (Low Engagement Group)

- *Mean: Increased from 25.39 to 117.64 minutes*
- *Median: Decreased from 14.17 to 11.83 minutes*
- *Standard Deviation: Increased from 26.23 to 468.93 minutes*

Distribution Shape

Skewness (Paid): Increased from 0.63 to 7.07 (right-skewed).
Skewness (Free): Increased from 1.17 to 15.06 (right-skewed).
Kurtosis (Paid): Increased from -0.85 to 58.48 (more outliers).
Kurtosis (Free): Increased from 0.36 to 315.76 (more extreme values).

INSIGHTS

- **Paid-Plan Students (Low Engagement Group)**

  Significant engagement increase, but high variability in 2022 suggests diverse engagement levels.

- **Free-Plan Students (Low Engagement Group)**

  Moderate engagement increase, driven by a few students; typical student engagement declined.

- **Paid vs. Free-Plan Comparison**

  Engagement Growth: Higher for paid-plan students.
  Variability: Increased more for free-plan students, suggesting more outliers.

CONFIDENCE INTERVALS

Task 2

The following task was about determining the minute interval for each of the four groups, within which I can be **95%** confident that a randomly selected individual will be situated. During this task, I used the following functions and equations:

=AVERAGE() : Mean

=STDEV.S() : Standard deviation

- Standard Error = STDEV.S()/SQRT(COUNT())
- 95% CI, z0.025 = 1.96
- CI low = MEAN()-(95% CI, z0.025) / Standard Error
- CI high = MEAN()-(95% CI, z0.025) / Standard Error

	minutes_watched_22
Mean	
Standard Deviation	
Standard Error	
95% CI, z0.025	

Z	CI low	CI high
0.95		

RESULTS 2

Paid-Plan Students

	minutes_watched_21
Mean	332.502508
Standard Deviation	485.8634701
Standard Error	8.292344684
95% CI, z0.025	1.96

Z	CI low	CI high
0.95	316.2495124	348.7555036

	minutes_watched_22
Mean	368.3547139
Standard Deviation	596.4051599
Standard Error	8.348069563
95% CI, z0.025	1.96

Z	CI low	CI high
0.95	351.9924976	384.7169303

Paid-Plan Students:

*There was an increase in engagement for paid-plan students 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** (**351.91** 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**.*

RESULTS 2

Free-plan Students

	minutes_watched_21
Mean	133.9333129
Standard Deviation	367.2624454
Standard Error	2.047595853
95% CI, z0.025	1.96

Z	CI low	CI high
0.95	129.920025	137.9466008

	minutes_watched_22
Mean	69.14765544
Standard Deviation	255.6234424
Standard Error	0.735906464
95% CI, z0.025	1.96

Z	CI low	CI high
0.95	67.70527877	70.59003211

Paid-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**.*

INSIGHTS

- **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 **67.71** 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.*

HYPOTHESIS TESTING

Task 3

The null hypotheses should include the following:

- The engagement (minutes watched) in **Q4 2021** is higher than or equal to the one in **Q4 2022** ($\mu 1 \geq \mu 2$). We test free-plan and paying students separately.

To test this hypothesis I will perform a **"t-Test: Two-Sample Assuming Unequal Variances"**

(The test results are in the analysis sheet)

SUMMARY

- **Paid-Plan Students**

  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, so we reject it. Of course, rejecting the null hypothesis does not confirm the alternative hypothesis; it suggests that the data provide enough evidence against the null hypothesis.

- **Free-Plan Students**

  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 μ1μ1 is smaller than μ2μ2. So, the data supports the null hypothesis that μ1μ1 is larger than or equal to μ2μ2.

CONCLUSION

The cost to the company of each type of error would depend on the implications of incorrectly concluding that engagement has increased — potentially leading to over-investment in certain features or complacency about needing to improve features — versus incorrectly concluding that engagement has not increased — potentially missing out on recognizing successful features or identifying areas that need improvement.