Customer Engagement Analysis in Excel Project

A Practical Approach to Analyzing Customer Engagement Metrics

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. The platform also offered in-app coins that could be exchanged for special awards, a leaderboard where students could compete for top positions in different divisions, earning weekly rewards and advancing up the ladder, and streaks to motivate students to maintain consistent learning habits. 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. With this Customer Engagement Analysis in Excel project, you must analyze whether the new additions to the platform have increased student engagement.

Project files

During this Customer Engagement Analysis in Excel project, you'll analyze a dataset from the 365 company. It's important to note that personal user information has been masked to ensure privacy, and the database's volume has been reduced for practical reasons. Nevertheless, you can consider this dataset to accurately represent the company's operations—providing a realistic and relevant context for your analysis.

In addition, consider the following information about the column values while working with the data:

1. **student_id** – the unique identifier for each student in the dataset. The field contains IDs for students who used the 365 Data Science platform with free or paid accounts in **Q4 2021** (October 1, 2021 – December 31, 2021, both included) and **Q4 2022** (October 1, 2022 – December 31, 2022, both included).

1. **student_country** – identifies the country of each student. The field provides information about students' geographic location and can help analyze regional differences or conduct country-specific analyses.

2. **Paid** – indicates whether a student had a paid account during the specified period. It is a binary variable, where '1' represents a paid account and '0' represents a free or unpaid account. It helps differentiate between students who have access to additional features or content through a paid subscription.
3. **minutes_watched_21** – represents the student's engagement level, as expressed by the number of minutes a student has watched in Q4 2021.
4. **minutes_watched_22** – denotes the student's engagement level, as expressed by the number of minutes a student has watched in Q4 2022.

The Excel file 'Engagement Project.xlsx' consists of four sheets: Task 1 and 2, Task 3, Task 4, and Task 5. Each sheet contains specific information regarding the project's tasks and corresponding data. Use the data in these sheets to answer the questions that follow.

Tasks

Descriptive Statistics

Task 1

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. The platform also offered in-app coins that could be exchanged for special awards, a leaderboard where students could compete for top positions in different divisions, earning weekly rewards and advancing up the ladder, and streaks to motivate students to maintain consistent learning habits.

Please open the 'Engagement project.xlsx' file and navigate to the 'Task 1 and 2' sheet using Microsoft Excel. Your first task is to provide insights into the relative engagement levels in Q4 2021 and Q4 2022. You will focus on low-engagement users (those who watched between 1 and 100 minutes in 2021). Low-engagement users often represent the most significant potential for growth.

If 365 can find ways to increase its usage, it could significantly impact the overall use of the platform.

If there are repeated students who watched in Q4 2021 and Q4 2022, how does their engagement compare between the two periods? Compute the mean, median, and standard deviation for these groups. Is there a difference in engagement between paid- and free-plan subscribers?

Task 2

Please open the 'Engagement project.xlsx' file and navigate to the 'Task 1 and 2' sheet using Microsoft Excel. Calculate the skewness and kurtosis of students who watched content in Q4 2021 and Q4 2022. Consider paid- and free-plan students. Does the result contradict the mean and median values previously obtained?

Confidence Intervals

Task 3

Please open the 'Engagement project.xlsx' file and navigate to the 'Task 3 sheet using Microsoft Excel.

You will find information about two pairs of students:

- Students who haven't had a paid-plan subscription
  - Engaged in Q4 2021
  - Engaged in Q4 2022
- Students who have been paid-plan subscribers
  - Engaged in Q4 2021
  - Engaged in Q4 2022

For each of the four groups, determine the minute interval within which you can be 95% confident that a randomly selected individual will be situated.

What conclusions can you draw about the students' engagement in Q4 2021 and Q4 2022?

Hypothesis Testing

Task 4

Use the data in Task 4 of the 'Engagement project.xlsx' to solve the following task.

You want to reach a data-driven customer engagement decision on whether the platform's new features contribute to the increase of minutes watched on the platform for both free-plan and paying students—i.e., the rise in student engagement in their study process. To do that, use hypothesis testing on both groups (free-plan and paying) for 2021 and 2022.

Your 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.

Additionally, make the following assumptions:

- For free-plan students, perform a two-sample t-test assuming unequal variances.
- For paying students, conduct a two-sample t-test assuming unequal variances.

Optional: *Perform a two-sample f-test for variances to support the*

What conclusion can you draw from this test? Comment on the results of committing a Type I or a Type II error in this study. Which one would result in higher costs for the company?

Task 5

Your last task is determining whether the average number of minutes watched in the US is similar to that in India.

Understanding the differences in usage patterns can help in product localization. The platform might need to tailor its content, features, or user interface to better fit the preferences or needs of users in different regions.

You'll focus only on free-plan students in 2022. Use the Excel sheet Task 5 to perform your calculations.

Your null hypotheses should (respectively) include the following:

- The engagement (minutes watched) in the US is higher than or equal to that in India ($\mu1 \geq \mu2$). We test only free-plan students.
- The engagement (minutes watched) in the US is lower than that in India ($\mu1 < \mu2$). We test only free-plan students.