Understanding User Patterns for an EdTech company

The Kaggle dataset provides valuable insights into the customer base of an EdTech company. It includes details about how users interact with the learning platform, the courses in which they are enrolled, the duration of these courses, and various other relevant metrics. This wealth of information enables us to identify patterns in user behavior, allowing us to make data-driven decisions that can enhance the platform's overall quality and expand its users.

Objectives:

After examining the datasets and gaining an in-depth understanding of the provided data, we defined the project's scope by framing a series of inquiries:

- Which courses exhibit the highest levels of engagement and completion, and how do these metrics relate to the duration of the courses?

- Does the structure or nature of the lessons influence user engagement? For instance, is there a noticeable distinction in user engagement between courses categorized as 'practice,' emphasizing question practice, and 'session,' emphasizing content comprehension?

- By analyzing user feedback, are we adeptly addressing users' queries in a meaningful manner?

Data Transformation

We obtained our dataset from the following source -

https://www.kaggle.com/datasets/nwanalytics/customer-activity?select=feedback_details.csv

As mentioned in the project checkpoint we will be considering the following five tables to conduct our analysis-

1. users_basic_details: Contains basic details of the enrolled users.

2. day_wise_user_activity: Contains the details of the day-wise learning activity of the users.
3. learning_resource_details: Contains the details of learning resources offered to the enrolled users
4. discussion_details: Contains the details of the discussions created by the user for a particular lesson.
5. feedback_details: Has ratings for the lessons where users have provided ratings.

The primary table in our dataset is "day_wise_user_activity,". It provides crucial insights into how users interact with the platform and the specific courses they are concentrating on. It has the columns ‘user_id’ and ‘lesson_id’ which act as foreign keys using which we integrated the relevant columns from ‘users_basic_details’ and ‘learning_resource_details’ tables respectively.

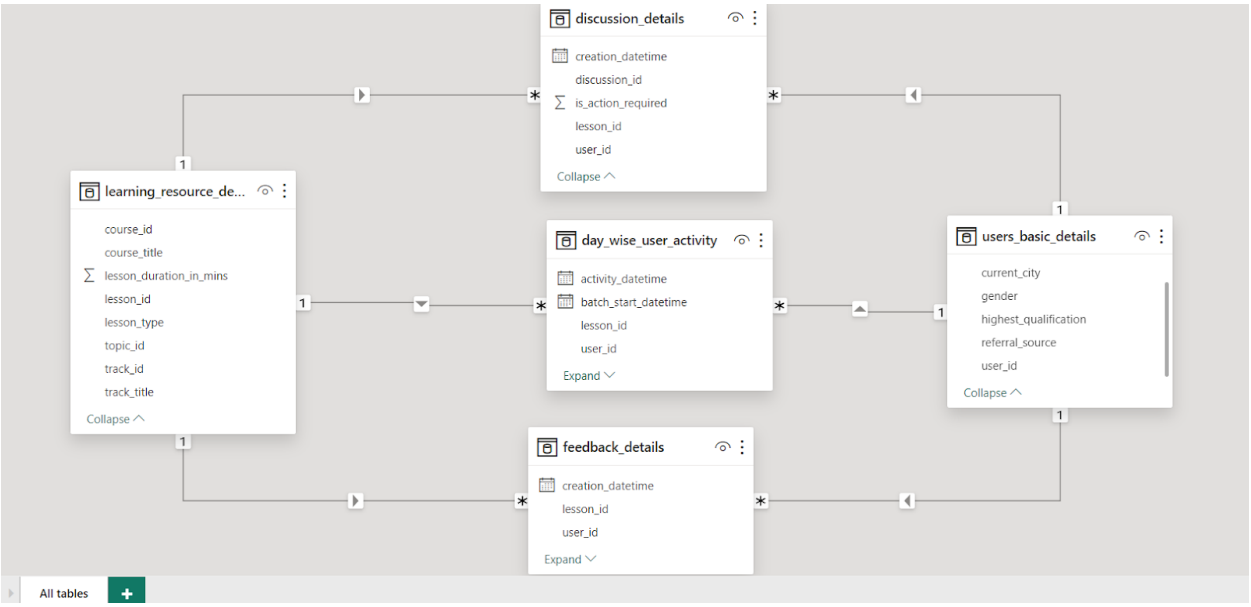

Fig 1: Data model

After creating the model, we followed the certain data transformation steps to clean and modify the data in a way that is suitable to create visualizations and perform analysis. Some of the steps we followed are mentioned below:

- Reformatted certain fields that were not conducive to analysis. For instance, the "user_id" column originally contained values like "user_1" and "user_2," which we modified to represent whole numbers such as 1, 2.

- Changed the data type of certain columns and renamed them to make it more user friendly.

- Wrote the following DAX query that splits the lesson_duration values into categories of 10 minutes

```
Lesson Duration Category =
    if [lesson_duration_in_mins] >= 0 and [lesson_duration_in_mins] <= 10  then "0-10"
    else if [lesson_duration_in_mins] > 10 and [lesson_duration_in_mins] <= 20 then
"10-20"
    else if [lesson_duration_in_mins] > 20 and [lesson_duration_in_mins] <= 30 then
"20-30"
    else if [lesson_duration_in_mins] > 30 and [lesson_duration_in_mins] <= 40 then
"30-40"
    else if [lesson_duration_in_mins] > 40 and [lesson_duration_in_mins] <= 50 then
"40-50"
    else ">50"
```

Key Findings

User Engagement by Course - Dashboard 1

- Highest Engagement: The ‘English for beginners course has the highest user engagement, followed closely by ‘Backend Developer’.
- Significant Interest: The "Programming' courses also show significant user interest.
- Moderate to Low Engagement: Other courses such as ‘Understanding 5G’, 'XPM 4.0 Fundamentals’ , and various specialized tracks show moderate to low engagement levels.

Referral Sources and Batch Start Dates - Dashboard 2

- Primary Referral Source: Instagram serves as the primary referral source, accounting for approximately 45% of the user base. This indicates a strong social media presence and word-of-mouth impact.

- Batch Preferences: The start dates of 2/2/2022 is the most preferred, suggesting seasonal or periodic peaks in enrollment
- However we have to keep in mind that there are a total of 53 users surveyed in this dataset. This might not be a large enough sample to draw conclusions from.

Lesson Engagement by Type and Duration - Dashboard 3

- Session Engagement: 'Session' type lessons dominate user engagement across all duration categories, particularly in the 10-20 minute range.
- Practice and Exam Engagement: 'Practice' and 'Exam' lessons have fewer users, with the highest engagement in the 10-20 minute duration for 'Practice' lessons.
- Completion Rates: Average completion rates are consistently high (>90%) across all lesson types and durations. However, 'Session' lessons exhibit a slight dip in completion rates for durations >50 minutes, while 'Practice' lessons maintain higher completion percentages in Longer duration categories (>50 minutes).
- Exams showed a 100% completion rate indicating that all the users completed the exams once they took it.

Observing discussion resolution

- We decided to examine the 'is_action_required' column in the 'discussion_details' table to assess the resolution status of discussions initiated by users. Upon analysis, we observed that, for the majority of lessons (approximately 95%), this column contains null values. These null values suggest either missing data or the absence of user-initiated discussions for these lessons. This observation could potentially indicate issues such as a lack of engagement or user hesitancy in raising discussions etc.
- For the lessons which had values in the ‘is_action_required’, 90% of them were resolved completely with 10% still requiring further action on the discussion threads.

Visualizations

- The bar graphs, donut charts, and line graphs provided clear visualizations of the aforementioned trends, aiding in the comprehension of user engagement patterns across various dimensions.

Findings

- Retention: Strategies such as personalization and gamification may improve retention, especially for longer-duration sessions.
- Engagement: Interactive and shorter lessons seem to resonate more with users, as evidenced by the completion rates.

Conclusions and Recommendations

- Course Development: There is a clear preference for Beginner level English courses and career-oriented courses such as 'Backend Developer'. Expanding similar course offerings could further increase user engagement.
- Marketing Strategies: Instagram's effectiveness as a tool to attract customers suggests that SEO and online marketing efforts are successful. Investing in these channels could attract more users.
- Outreach Strategy: Leverage the substantial user acquisition through referrals by incentivizing existing users to recommend the platform to others, facilitating organic growth and user-driven advertising.
- User Retention: The dip in completion rates for longer 'Session' lessons suggests a need for engaging content delivery methods to retain user attention for extended periods. Also, users seem to prefer longer Practice sessions (>50 minutes) indicating that its content is engaging and users do not have an issue spending more time in practice sessions.
- Enhance user engagement strategies and provide timely support for discussion posts: Consider incentivizing participation or introducing features that facilitate easier interaction with the platform. Also, recognize that 10% of discussions still require further action

Further Research

To refine these insights, further research could explore the correlation between course completion rates and user success stories, the impact of user demographics on course selection, and a deeper analysis of user engagement trends over a more extended time frame. We can also try to conduct this study over a larger user base and get more data.