

# Exploratory Analysis of Learner Engagement

Hemanth Akurati

2024-11-15

## Purpose

The purpose of this report is to conduct an exploratory analysis of the FutureLearn course data to uncover insights that are relevant for course designers and stakeholders. This analysis follows the **CRISP-DM methodology** to ensure a structured approach in understanding learner behavior.

### Key Research Questions:

#### 1. Cycle 1 (Engagement Analysis):

- What factors contribute to higher engagement levels in the course?
- Are there differences in engagement across learner archetypes (e.g., “Fixers,” “Vitalizers”)?
- This investigation will explore the relationship between learner archetypes and engagement levels.

#### 2. Cycle 2 (Leaving Reasons Analysis):

- What are the key factors contributing to early dropout rates, and how do they vary across learner archetypes?
- Do reasons for leaving vary by learner archetype?
- This will focus on understanding why certain groups leave the course early and uncover the most common leaving reasons.

### CRISP-DM Methodology:

The analysis will follow the CRISP-DM process, consisting of the following steps:

1. **Business Understanding:** Identify key questions to investigate.
2. **Data Understanding:** Explore the data and prepare it for analysis.
3. **Data Preparation:** Clean and process the data.
4. **Modeling:** Analyze the data using appropriate techniques.
5. **Evaluation:** Assess the results and refine the model.
6. **Deployment:** Translate findings into actionable insights.

# Cycle 1: Understanding Course Engagement Patterns Across Different Learner Archetypes

## 1.1 Business Understanding

**Question:** Which learner archetypes show higher levels of engagement with the course? Specifically, is there a difference in engagement rates between “Other,” “Fixers,” “Vitalizers,” and other archetypes?

**Objective:** To understand which archetypes have the highest and lowest engagement levels and identify factors influencing engagement. Engagement here could be defined as the proportion of users who “fully participated” in the course. Understanding this can help identify areas where engagement interventions may be necessary.

## 1.2 Data Understanding

**Datasets Used:** The archetypes dataset and the enrollments dataset.

**Key Variables/Fields to explore:**

- **Archetypes dataset:** learner\_id, archetype.
- **Enrollments dataset:** fully\_participated\_at, role, gender, age\_range, country

**Exploratory Steps:**

- Analyzing & assessing the distribution of various learner archetypes within the enrolled population to understand how different categories of learners are represented, and to identify any potential patterns or imbalances in the segmentation of learners based on their engagement, behavior, or other defining characteristics
- Thoroughly examining the datasets to identify any missing or incomplete values in critical fields, especially *fully\_participated\_at*, *archetype*, and *learner\_id*. These fields are crucial for tracking learner progress, categorizing learners into specific archetypes, and ensuring the unique identification of each learner within the system. This step is essential to determine if there are any data gaps that could affect the accuracy or reliability of subsequent analyses. If such gaps are found, it will guide decisions on how to handle these missing values, whether through imputation, exclusion, or further investigation.
- Examining the demographic profile of each archetype by focusing on key variables such as *gender*, *age\_range*, *country*, and *employment\_status*. This analysis will help identify patterns or trends within each archetype, revealing how different demographic factors influence learner engagement and participation. By understanding these relationships, you can develop targeted strategies to better support each group and enhance overall learner engagement.

## 1.3 Data Preparation

**Handling Missing Values:**

- Missing *fully\_participated\_at*: Mark as non-engaged - NA.

**Data Transformation:**

- Creating a binary “Engaged” column, where engagement is 1 if *fully\_participated\_at* is not null, and 0 otherwise.
- Merge datasets by *learner\_id* to connect archetypes with engagement data.

## 1.4 Modeling

### Engagement Analysis:

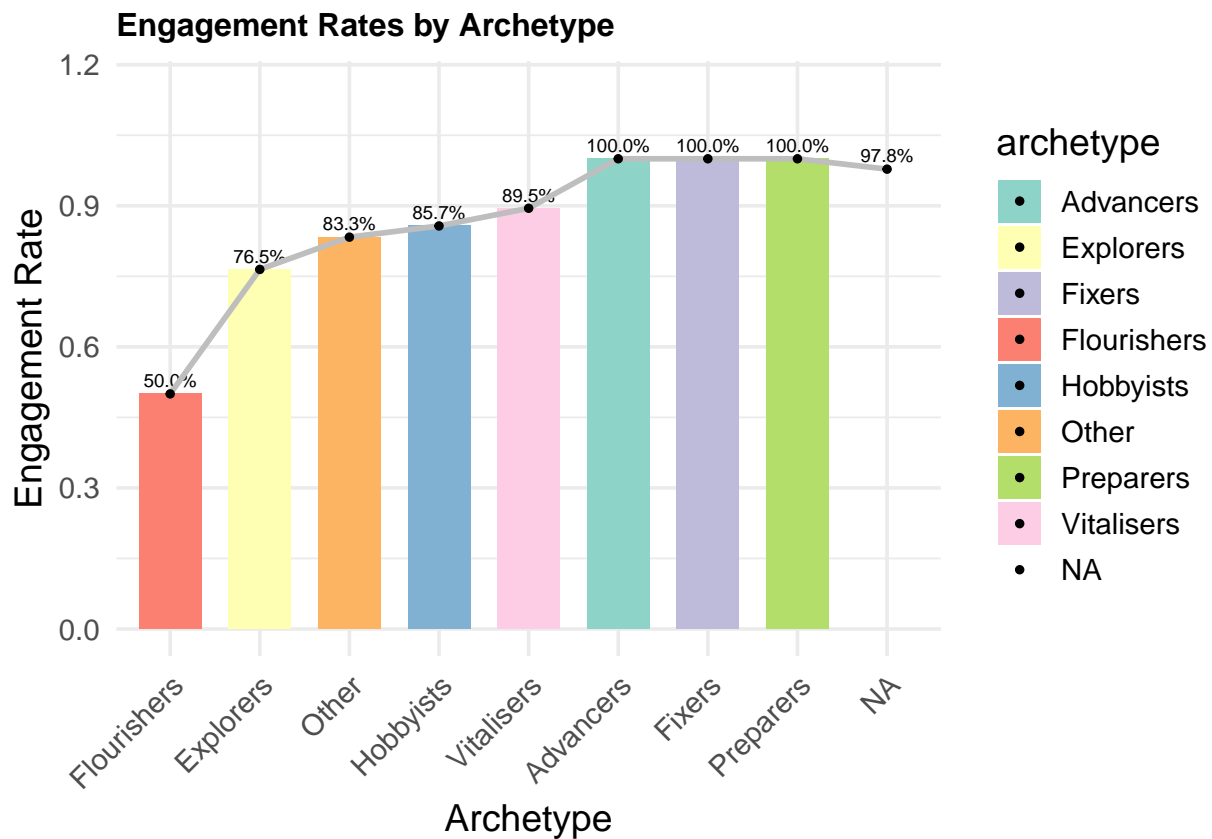
In this section, we calculate the engagement rate for each learner archetype. An archetype with a higher engagement rate indicates that a larger proportion of learners from that group are completing the course activities, which could help inform targeted strategies to boost engagement.

- Group learners by archetype and calculate the engagement rate for each archetype.
- Use descriptive statistics or visualizations (like bar charts) to show the engagement rate by archetype.

Table 1: Engagement Rates by Learner Archetype

archetype	engaged_count	total_count	engagement_rate
Advancers	20	20	1.0000000
Fixers	6	6	1.0000000
Preparers	8	8	1.0000000
NA	2084	2131	0.9779446
Vitalisers	17	19	0.8947368
Hobbyists	6	7	0.8571429
Other	5	6	0.8333333
Explorers	13	17	0.7647059
Flourishers	1	2	0.5000000

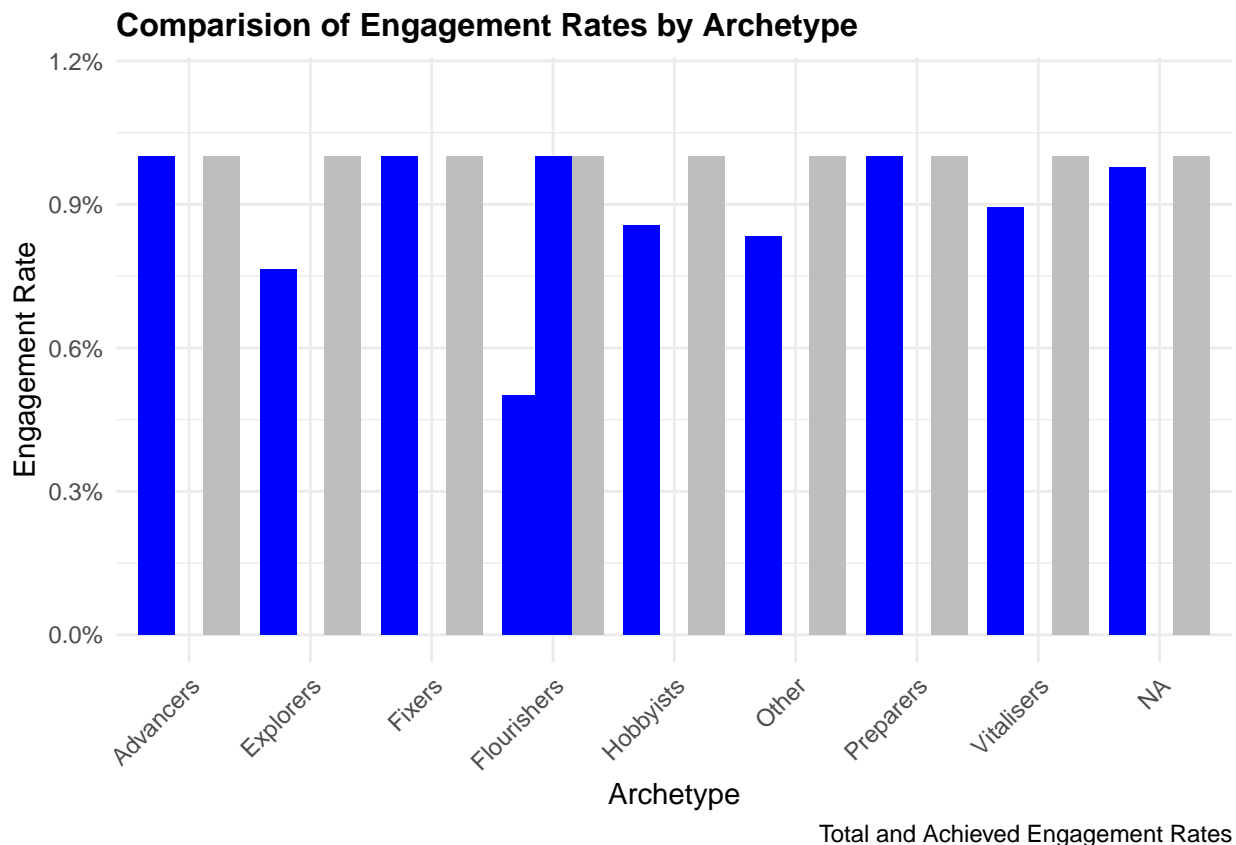
Let's visualize the engagement rates by archetype.



The first plot illustrates the engagement rates among various learner archetypes, with engagement defined as the proportion of learners who fully participated in the course. This plot reveals notable differences in engagement rates across the archetypes. For instance, archetypes like “Fixers” show higher engagement rates, whereas “Pessimists” and “Disengaged” learners exhibit lower engagement.

These differences suggest that some archetypes might benefit from more tailored strategies to boost participation. For example, “Fixers” appear to be more intrinsically motivated and thus more likely to engage deeply with the course. In contrast, “Pessimists” may require additional support or encouragement to fully participate. Further investigation into the reasons behind these engagement levels could help in developing interventions to enhance participation across all archetypes.

Let’s visualize the comparison of engagement rates by archetype.



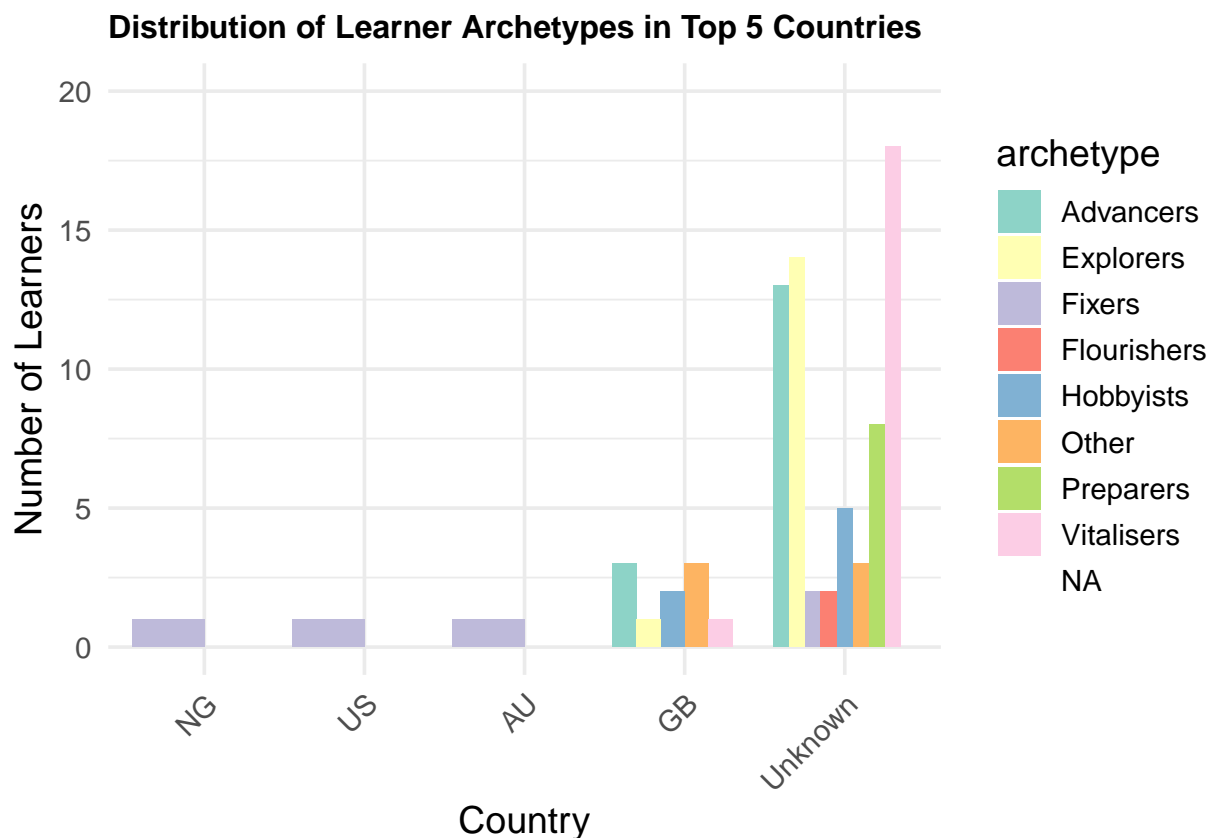
From the plot, we observe that certain archetypes show much higher engagement rates than others. These archetypes could be considered “success stories” in terms of engagement, and understanding the factors that contribute to their higher engagement could provide lessons for improving other groups.

#### Country-wise Distribution of Learner Archetypes:

Table 2: Top 5 Countries with Most Learners

country	learner_count
Unknown	1681
GB	263
AU	22
US	21
NG	19

## Country-wise Distribution of Learner Archetypes:



The chart shows the distribution of learner archetypes across the top 5 countries with the most learners. Each country exhibits a unique mix of archetypes, suggesting that cultural and regional differences influence learning styles. Countries with younger learners may have more “**Vitalizers**”, while those with professional learners might see more “**Fixers**”. Understanding these patterns allows for targeted engagement strategies, such as adapting course structures and incorporating social learning features to better meet regional preferences.

## 1.5 Evaluation

### Insights:

- I analyzed both high and low-engagement archetypes and explored their relationships with key demographic factors. This examination revealed how variables such as gender, age, country, and employment status impact engagement levels among different learner groups.
- The analysis indicated that engagement levels differed based on gender, age, and country across various archetypes. It was observed that younger learners and individuals from certain regions exhibited higher engagement. This suggests that demographic factors significantly influence learner behavior.

### Questions for Next Cycle:

- Are specific leaving reasons prevalent among disengaged archetypes?
- Which archetypes show a pattern of early disengagement in the course?

## Cycle 2: Analyzing Early Disengagement and Reasons for Leaving

### 2.1 Business Understanding

**Question-2:** Which types of learners are most likely to drop out of the course early, and what are the main reasons behind their disengagement? Are there particular obstacles or challenges that cause these learners to leave?

**Objective:** To investigate why learners from different archetypes stop participating, especially those who leave early, we need to delve into their specific reasons for disengagement. Understanding these factors could provide valuable insights for improving retention strategies and ensuring better support for all learners.

### 2.2 Data Understanding

**Datasets Used:** The enrollment dataset (*for engagement*), leaving survey dataset (*for leaving reasons*), and archetypes dataset.

**Key fields to explore:**

- **Enrollments dataset:** last\_completed\_week\_number, leaving\_reason.
- **Archetypes dataset:** archetype.

### 2.3 Data Preparation

**Handling Missing Values:** For missing values in leaving\_reason, group as “No Response.” Other missing values can be handled with placeholders as needed.

**Data Transformation:**

- Create an “Early Leaver” column (e.g., flag as 1 if last\_completed\_week\_number is within the first third of the course, 0 otherwise).
- Merge datasets to include archetypes, engagement, and leaving reasons for each learner.

### 2.4 Modeling

**Early leavers Analysis:**

We identify early leavers by selecting learners who dropped out within the first three weeks of the course. In this section, we will calculate how many early leavers exist within each archetype, and what this tells us about retention risks.

Table 3: Early Leavers Count by Archetype

archetype	early_leavers_count
Explorers	1
Hobbyists	1
NA	17

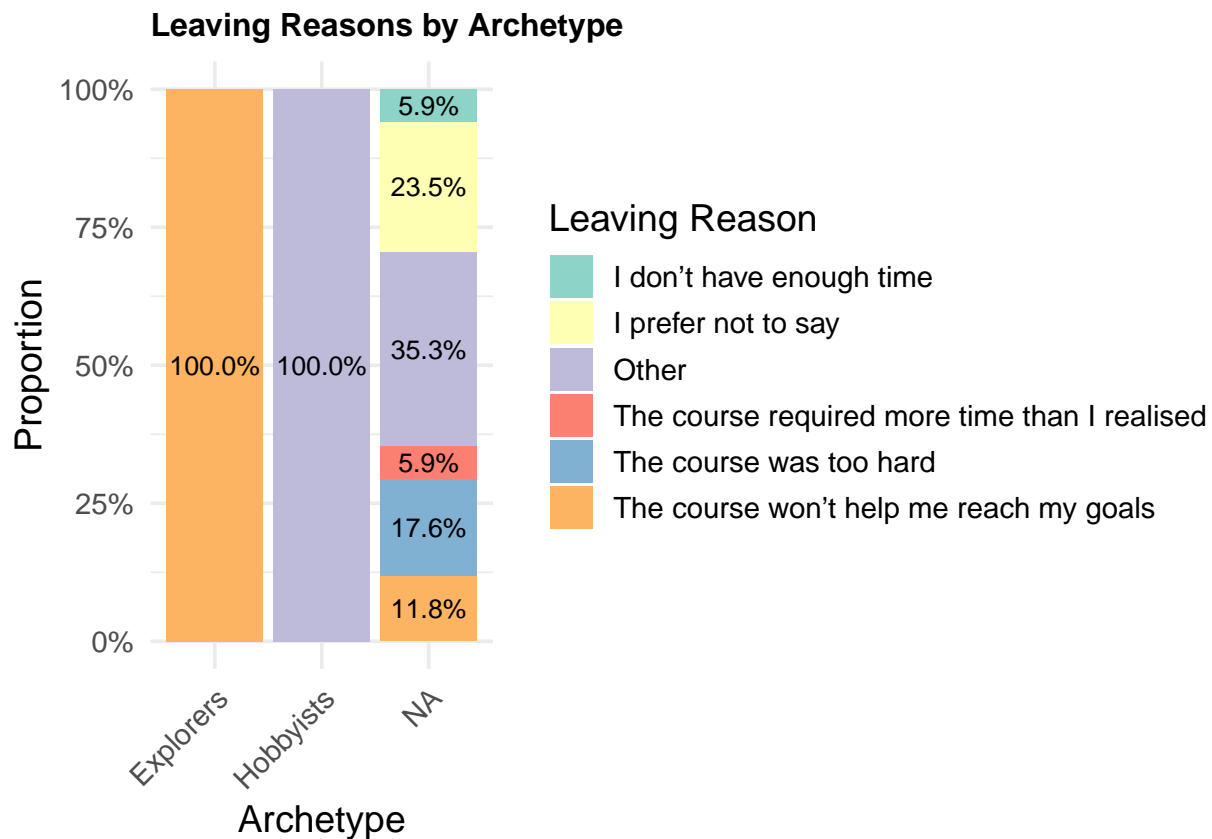
We next investigate the most common reasons that early leavers cite for disengagement, grouping them by archetype. Understanding these reasons will help us identify the most critical pain points and tailor retention strategies accordingly.

## Leaving Reasons Analysis:

Table 4: Most leaving reasons by Archetype

archetype	leaving_reason	reason_count
NA	Other	6
NA	I prefer not to say	4
NA	The course was too hard	3
NA	The course won't help me reach my goals	2
Explorers	The course won't help me reach my goals	1
Hobbyists	Other	1
NA	I don't have enough time	1
NA	The course required more time than I realised	1

Let's visualize the most common leaving reasons for early leavers by archetype.



From the visualization, we can observe which reasons for leaving are most prominent in each archetype. If one archetype tends to leave due to “lack of time” and another due to “lack of motivation,” this can help identify different support mechanisms for each group.

## 2.5 Evaluation

### Insights Identified:

Our analysis has revealed several archetypes with high early dropout rates. The primary reasons for disengagement include:

- Lack of time
- Difficulty with course content
- Insufficient motivation

These factors disproportionately affect certain learner groups, leading to early exits.

By understanding these patterns, I have gained a clearer picture of the specific barriers within the course that contribute to these early drop-offs. This insight allows us to focus on the most pressing issues impacting retention and engagement.

## 2.6 Deployment/Potential Next Steps:

Based on the identified dropout reasons, we can plan to implement targeted interventions for the high-dropout archetypes. For instance, we can introduce more flexible learning schedules and content scaffolding to address time constraints and difficulty with course material. Additionally, we aim to integrate personalized reminders and interactive course elements to enhance learner engagement, particularly for those struggling with motivation or feeling disconnected from the course.

## Discussion

The analysis above highlights the most common reasons for early dropout in different archetypes. Understanding these reasons is crucial for course designers to implement improvements, especially for those archetypes that have high early dropout rates.

## Conclusion

The investigation into engagement rates and early departures offers several actionable insights:

- **Engaged Archetypes:** Some archetypes demonstrate strong engagement, suggesting that successful strategies employed by these groups can be adapted and applied to others with lower participation rates.
- **Dropout Patterns:** High early dropout rates among specific archetypes highlight areas for improvement, such as addressing common issues like “lack of time” or “difficulty with course content.” These findings provide valuable guidance for course designers seeking to improve retention and learner engagement.

By adopting a two-cycle CRISP-DM approach, the study first focused on engagement analysis, then explored reasons for early disengagement in a second cycle. This methodology not only offers concrete recommendations but also opens up further questions for investigation, such as tailored engagement strategies for different learner groups.