

# Data Analytics Report

Of FutureLearn  
Cyber-Security Course

Davies Luo

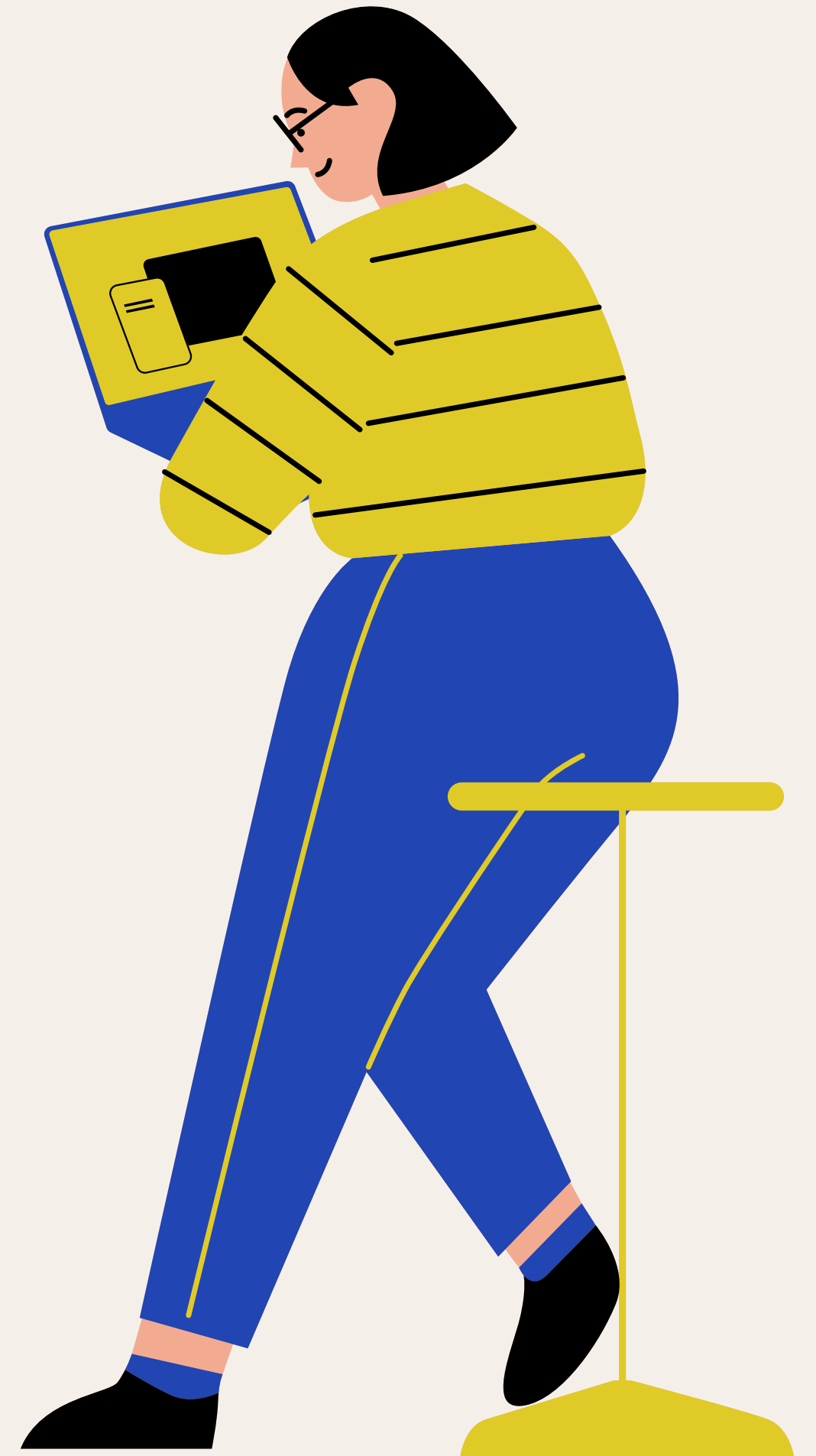


- Introduction
- 1st CRISP-DM Cycle
- 2nd CRISP-DM Cycle
- Conclusions

Data

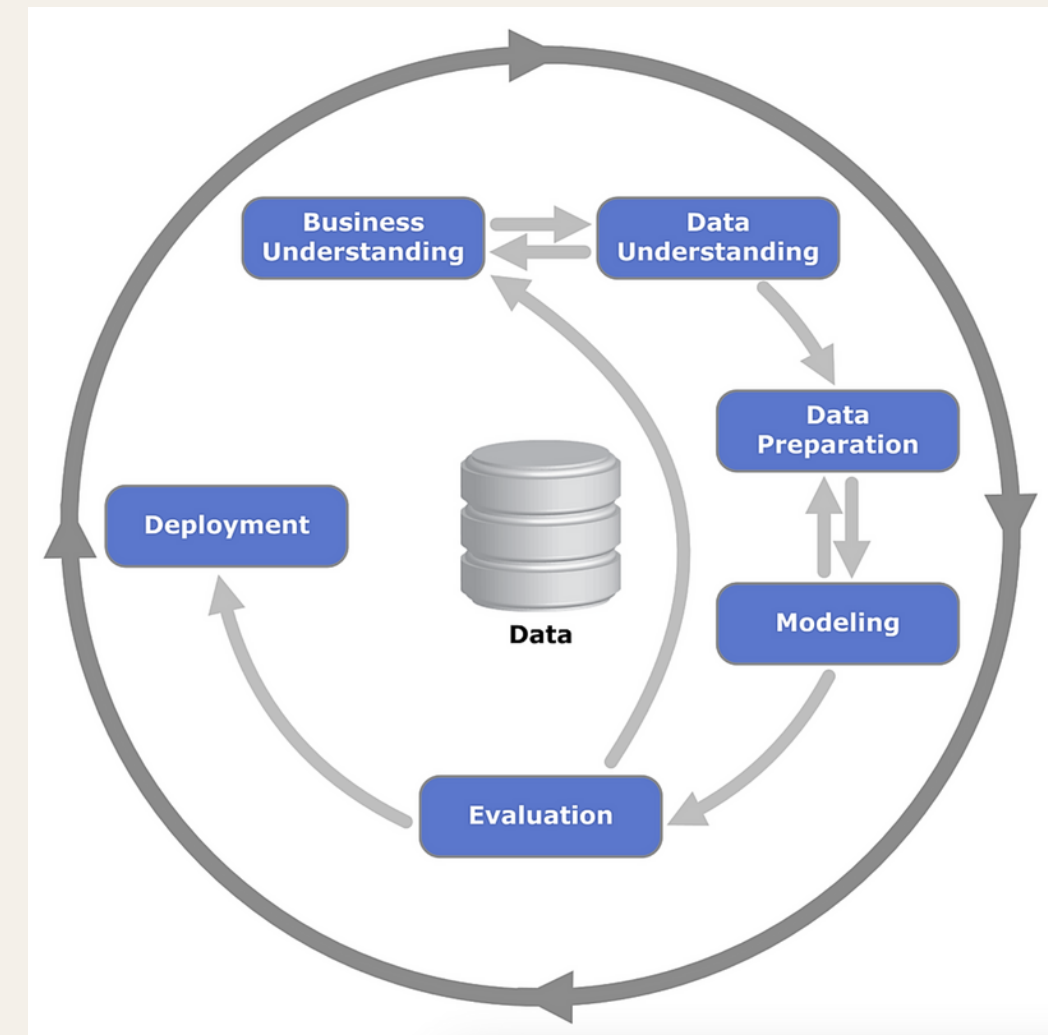
Modelling

Evaluation



# Introduction

- Analysis guided by CRISP-DM methodology
- Stages:
  - **Business Understanding**: Define objectives and project scope.
  - **Data Understanding**: Explore and familiarize with data.
  - **Data Preparation**: Cleanse and refine data.
  - **Modelling**: Apply statistical techniques for insights.
  - **Evaluation**: Assess results against objectives.



# 1st Cycle - Business Understanding

Objective:

Assess course efficacy over time

- Focus on learner engagement and performance.
- Identify patterns and trends indicating areas for improvement in course delivery and content



*Question: Evaluation of the course quality development in 7 years*

# 1st Cycle - Data Understanding

Understand course appeal and effectiveness, with exploring demographic reach.

- **Enrollment Data:** Assess course popularity and growth.
- **Quiz Responses:** Gauge learner performance and engagement.
- **Gender Data:** Understand demographic reach and inclusivity.

## Data Quality Checks

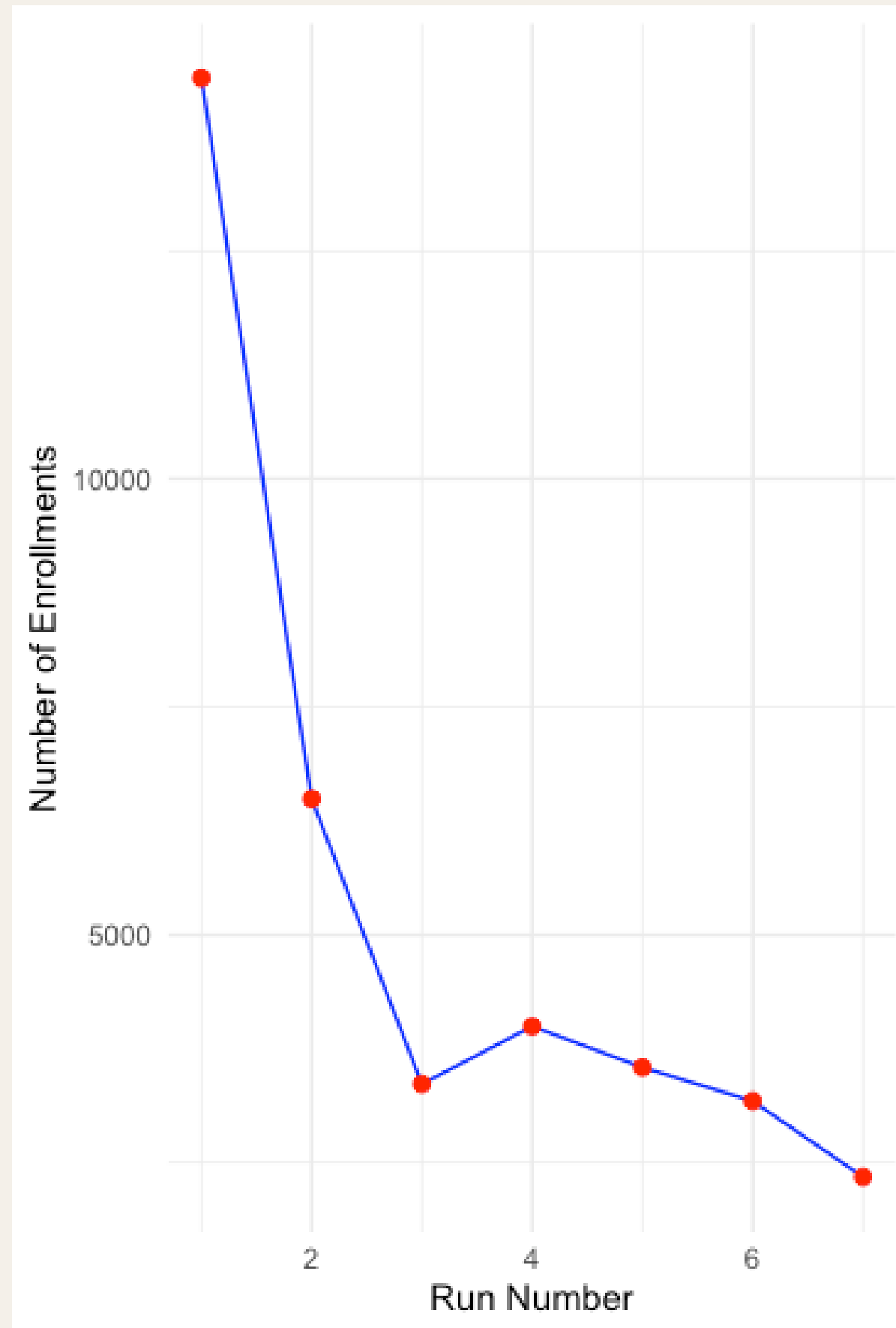
Ensure consistency, completeness, and accuracy.

# 1st Cycle - Data Preparation

- **Data integration:** Merge data using 'learner\_id' across 7 course runs, creating a new column “run\_number”
- **Data cleansing:** Address inconsistencies and missing values
- **Select relevant data points:** Include comprehensive data for detailed analysis
- **Ensure consistency:** Standardize data formats and categories

# 1st Cycle - Data Modelling

Enrollment Trend



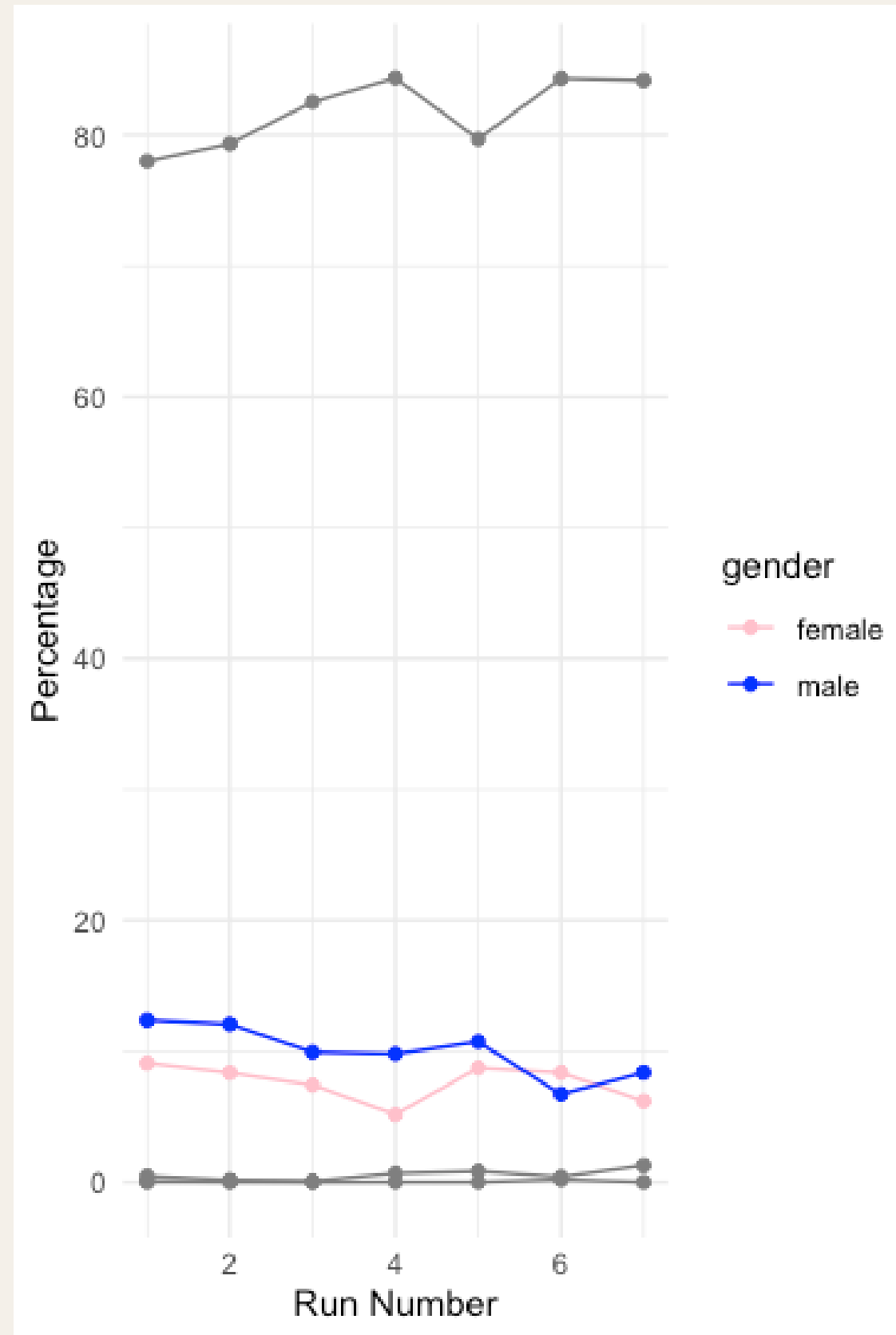
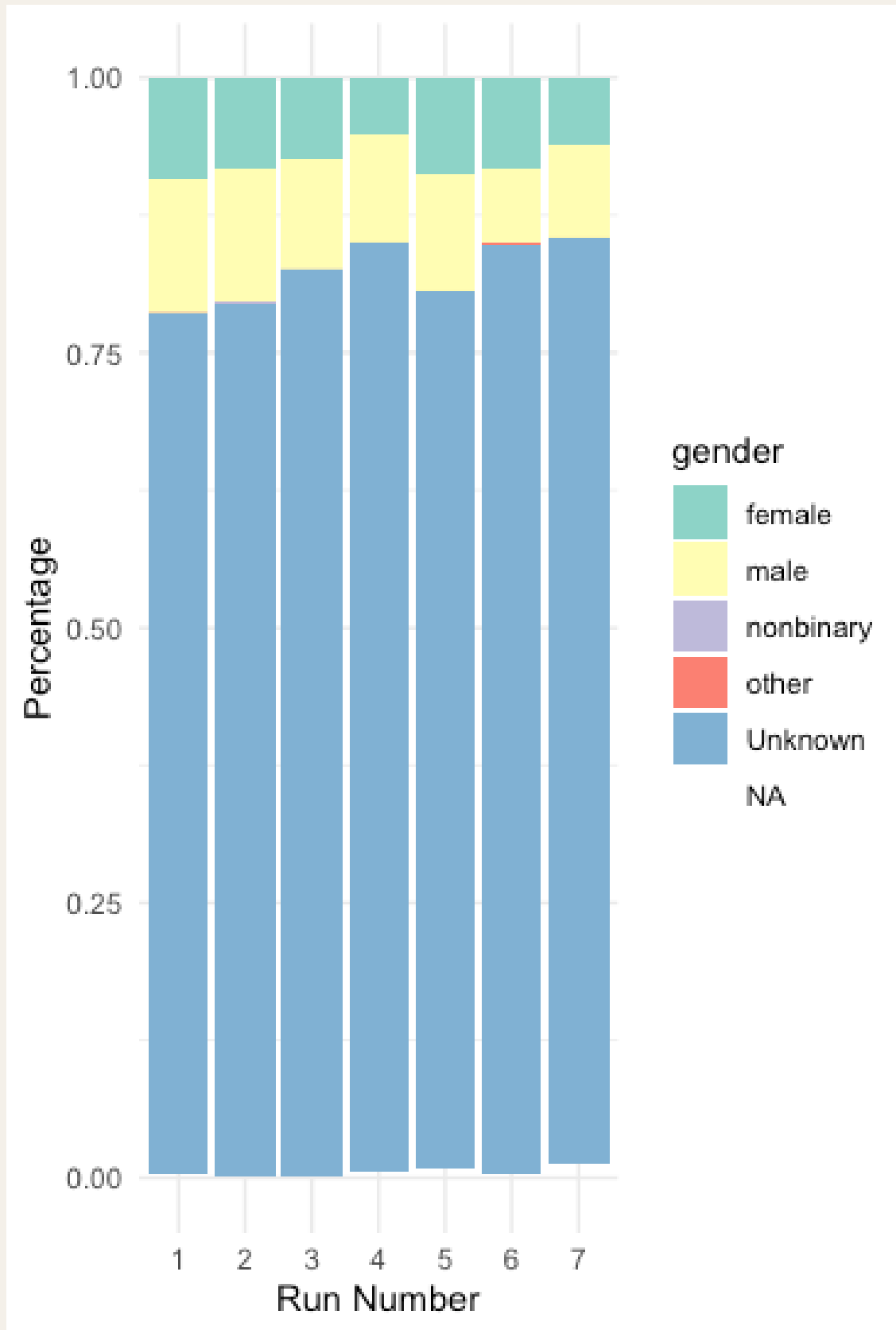
Columns used:

run\_number

enrollment

# 1st Cycle - Data Modelling

Gender Equality



Columns used:

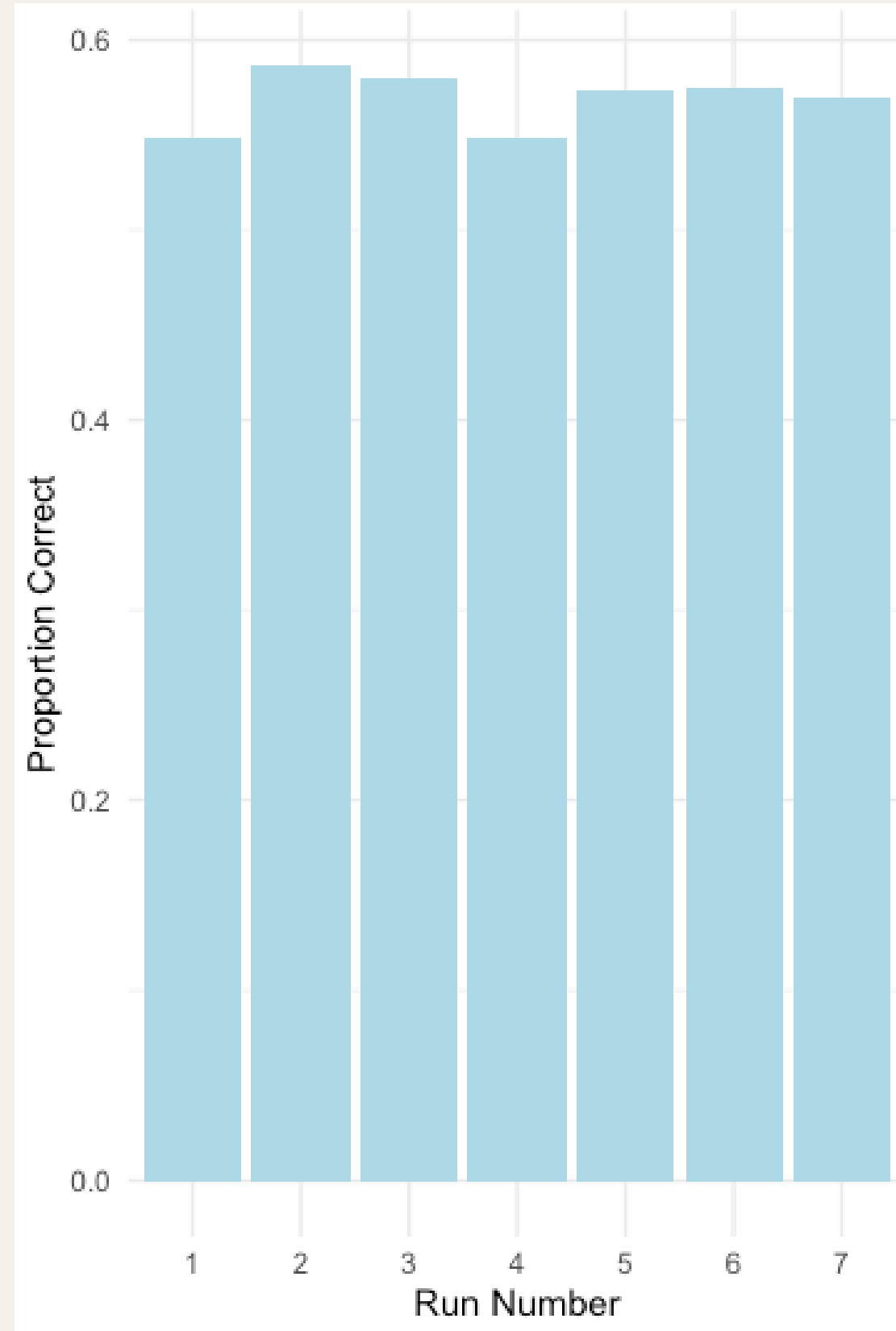
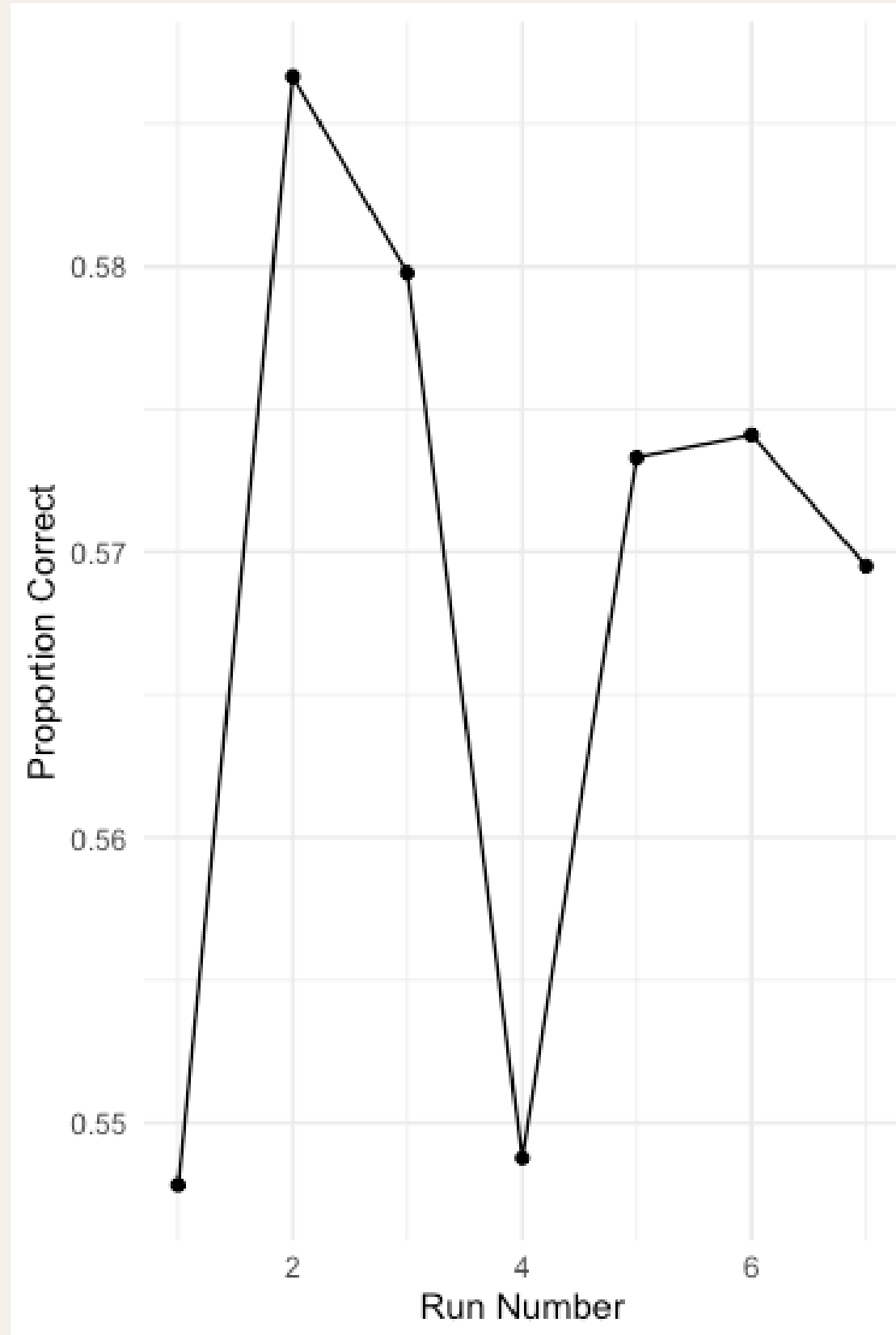
run\_number

gender



# 1st Cycle - Data Modelling

Correct Response



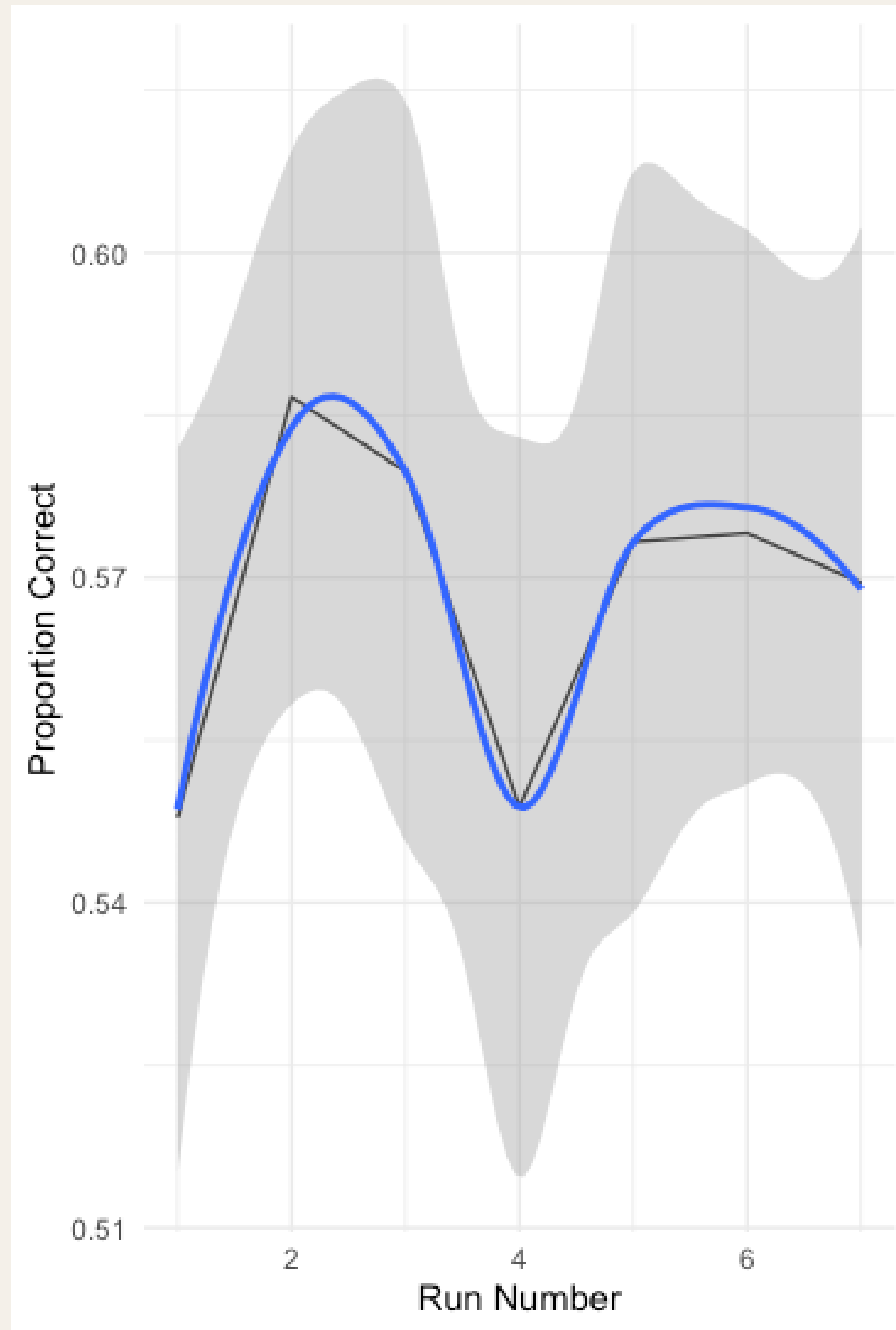
Columns used:

run\_number

correct

# 1st Cycle - Data Modelling

Correct Response



Columns used:

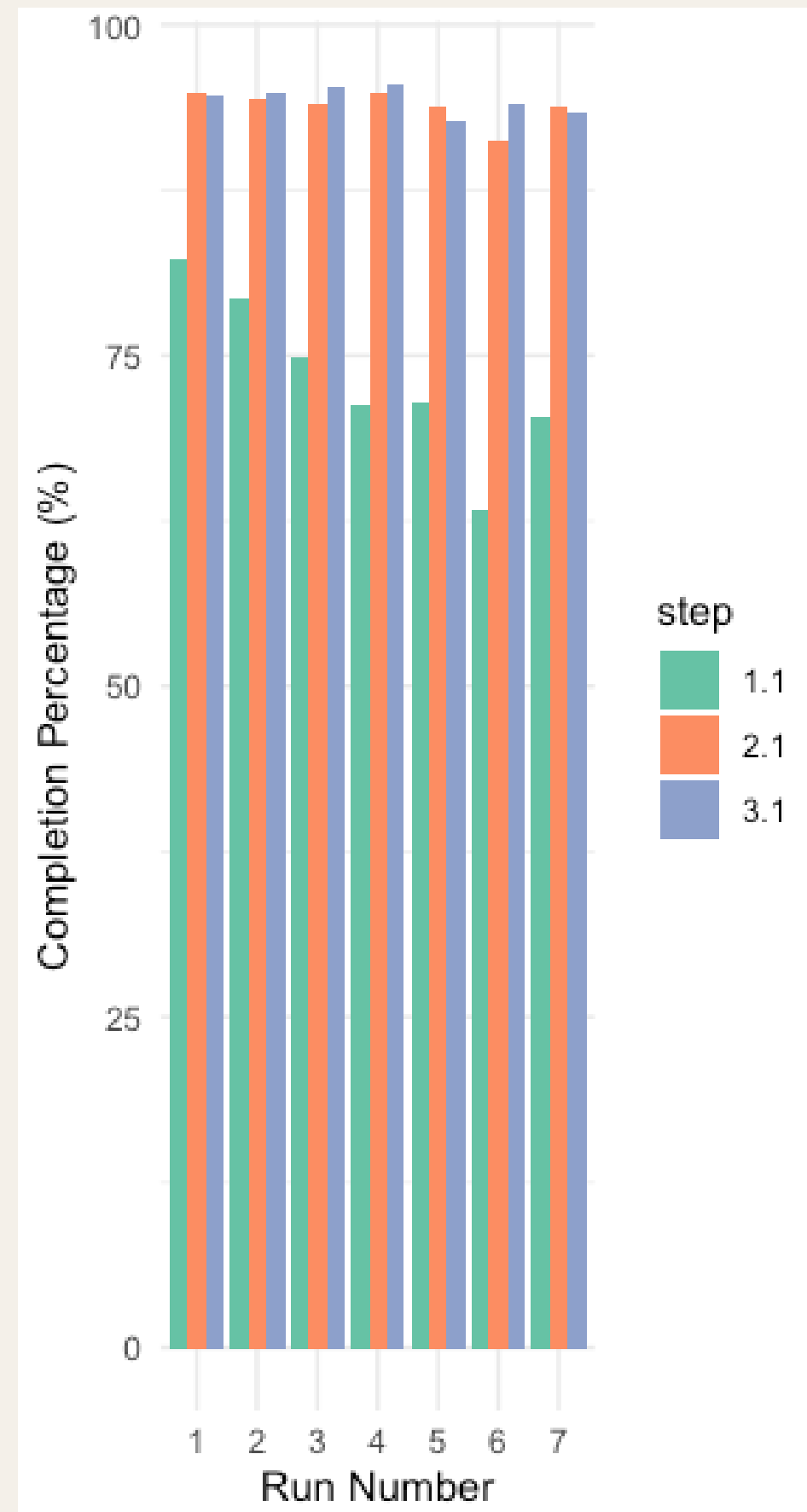
run\_number

first\_visited\_at

last\_completed\_at

# 1st Cycle - Data Modelling

Step Completion Rate



Columns used:

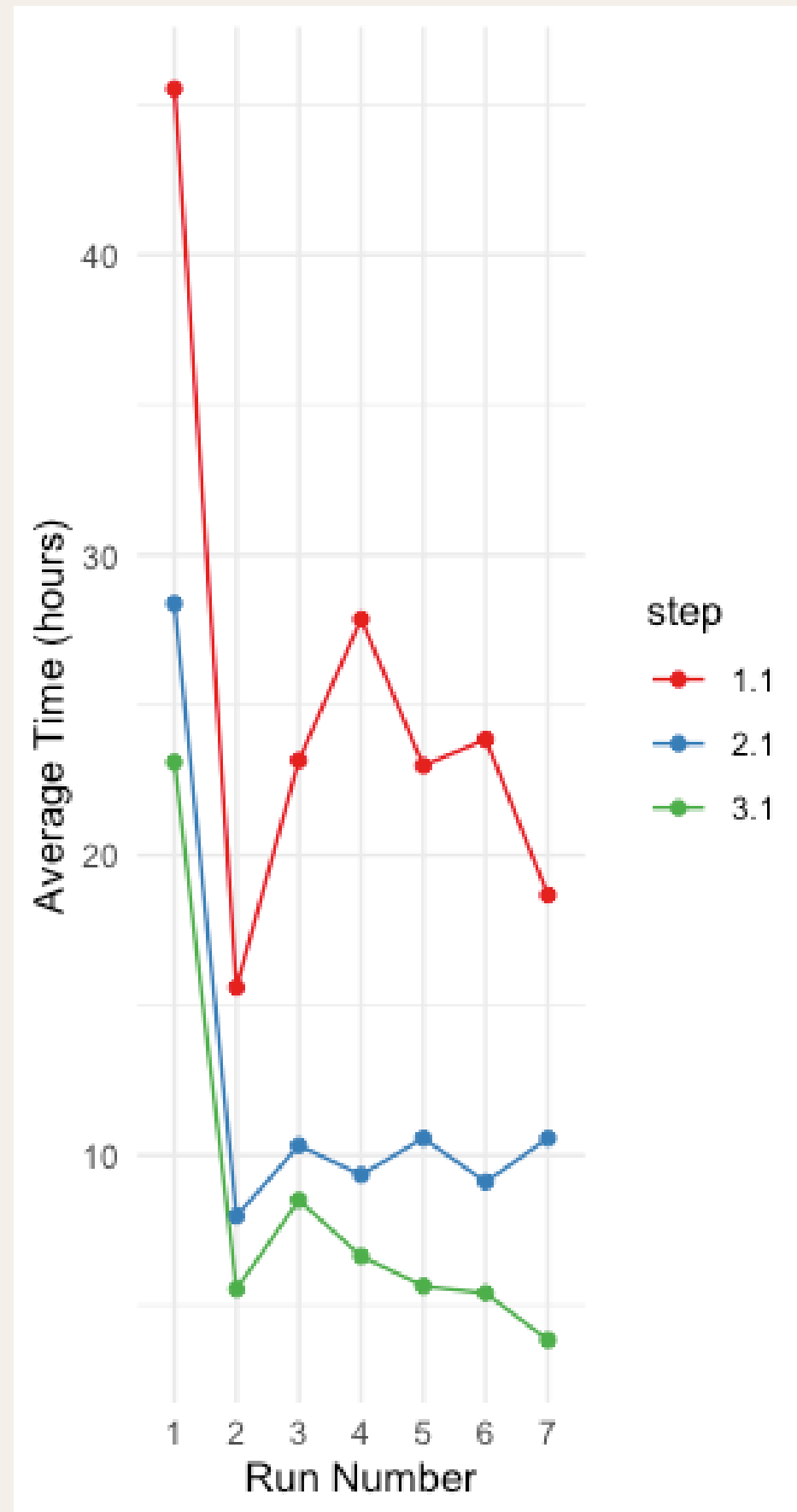
run\_number

first\_visited\_at

last\_completed\_at

# 1st Cycle - Data Modelling

Step Completion Time



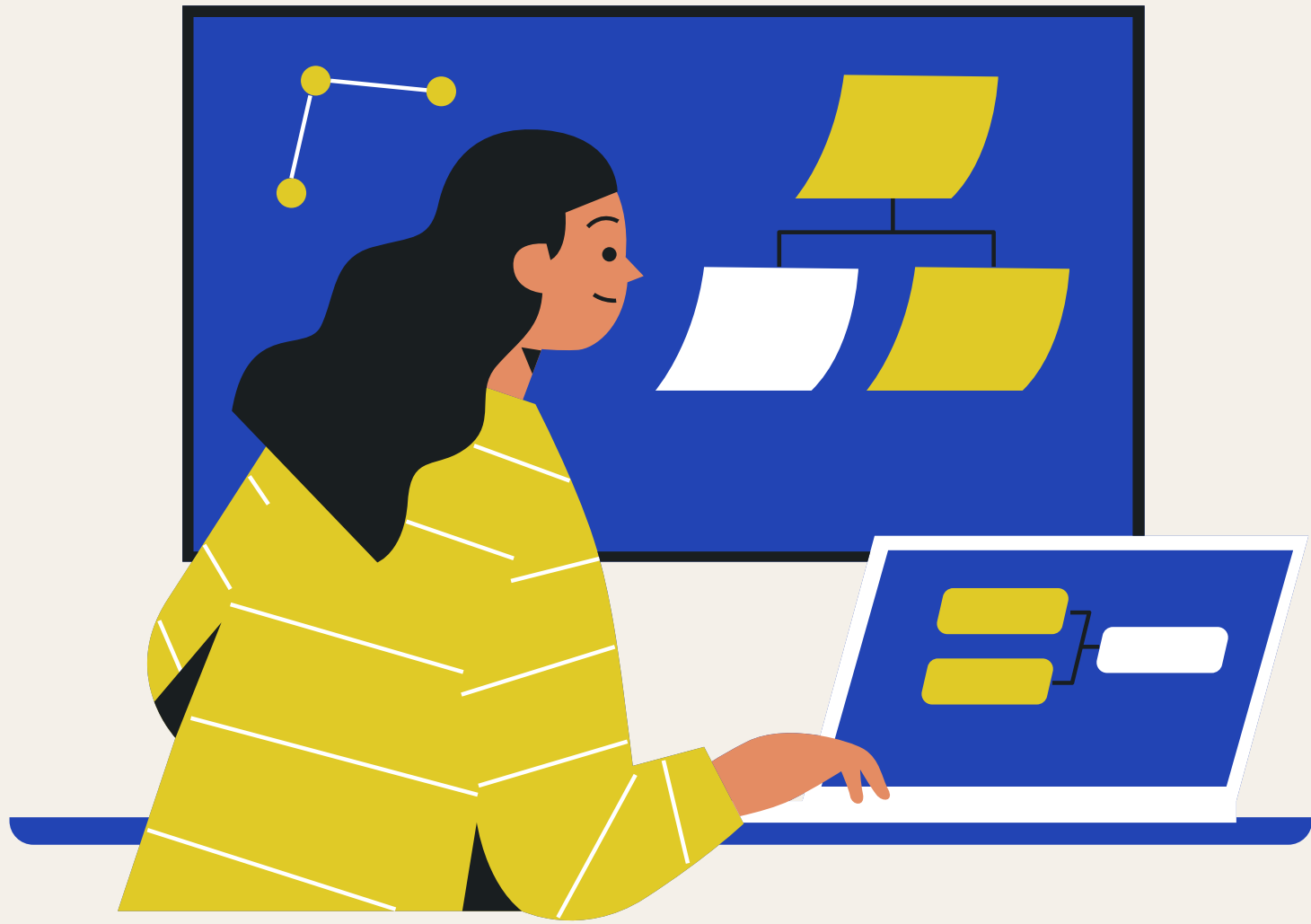
Columns used:

run\_number

first\_visited\_at

last\_completed\_at

# 1st Cycle - Evaluation



- **Enrollment Trend:** Observed decline after initial interest, indicating issues in sustaining engagement.
- **Correct Response Trend:** Stability in questions performance, but with minor fluctuations suggesting changes in learner demographics or course content.
- **Step Completion Rate/Time:** Declining in early course sections, hinting at engagement or content relevance issues.

Decision to focus on qualitative feedback in the second cycle to understand the reasons behind these trends and to identify specific areas for improvement.

*Next Step: Where to improve the course quality based on feedback*

# 2nd Cycle - Business Understanding

## Objective:

Deepen understanding of learner experiences and perceptions

- Focus on qualitative feedback from learners.
- Explore underlying reasons behind engagement trends and dropout rates.



*Question: Where to improve the course quality based on feedback?*

# 1st Cycle - Data Understanding

To get a comprehensive understanding of learner experiences and identify areas for improvement:

- **Archetype Survey:** Understand different learner personalities.
- **Leaving Survey:** Identify reasons behind course dropouts.
- **Sentiment Survey:** Gauge learner experience and reviews.

## Data Quality Checks

Ensure consistency, completeness, and accuracy.

# 2nd Cycle - Data Preparation

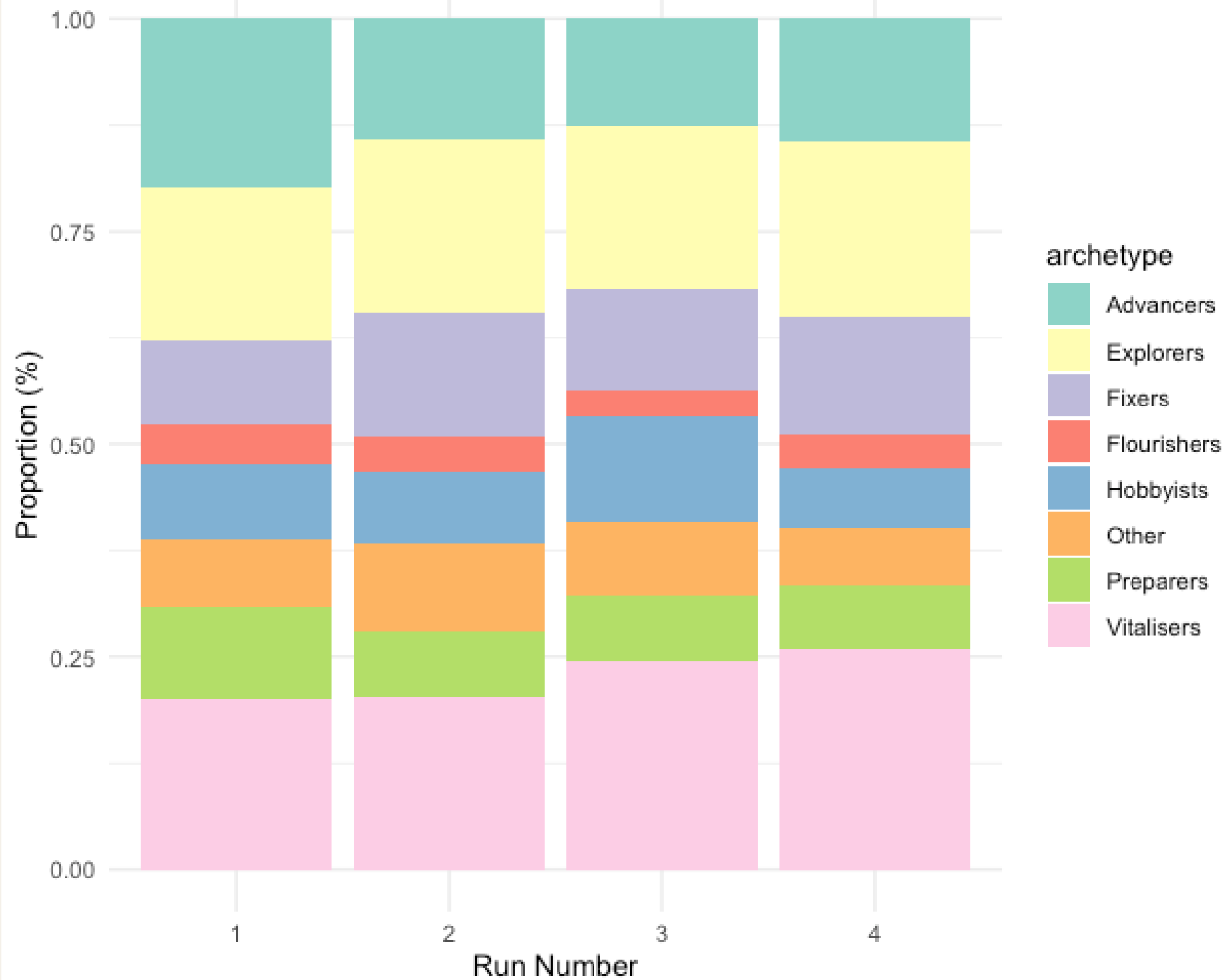
- **Data integration:** Merged archetype surveys, discontinuation reasons, and sentiment surveys using 'id', creating a new column "run\_number"
- **Data Cleaning:** Standardized IDs and text responses, addressed missing or incomplete entries for accuracy.
- **Variable Creation:** Developed new variables to categorize discontinuation reasons and quantify sentiment from surveys.
- **Data Standardization:** Ensured consistency in data formats and categories for reliable analysis.



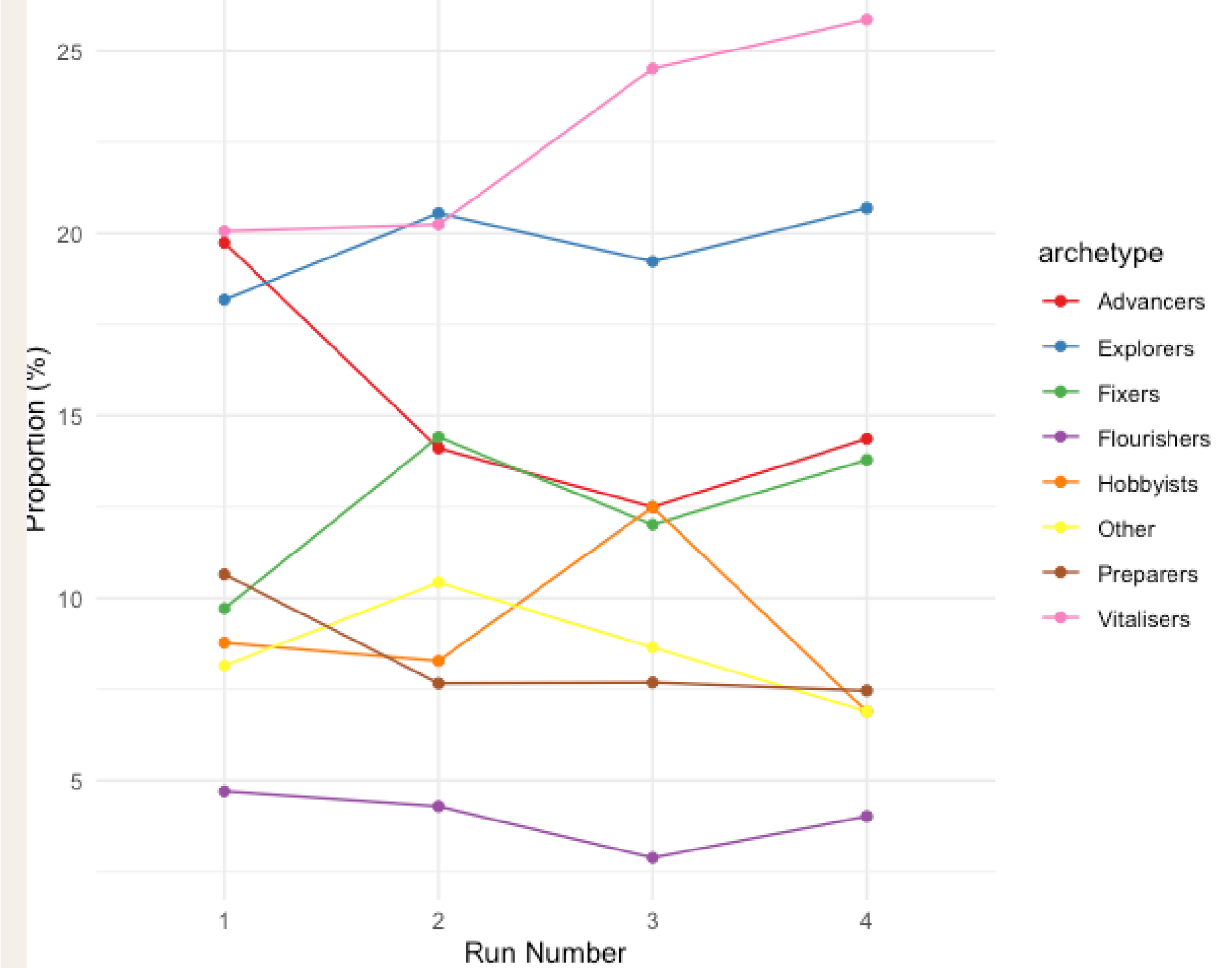
# 2nd Cycle - Data Modelling

## Archetype Proportions

Proportion of Archetypes in Runs 4-7

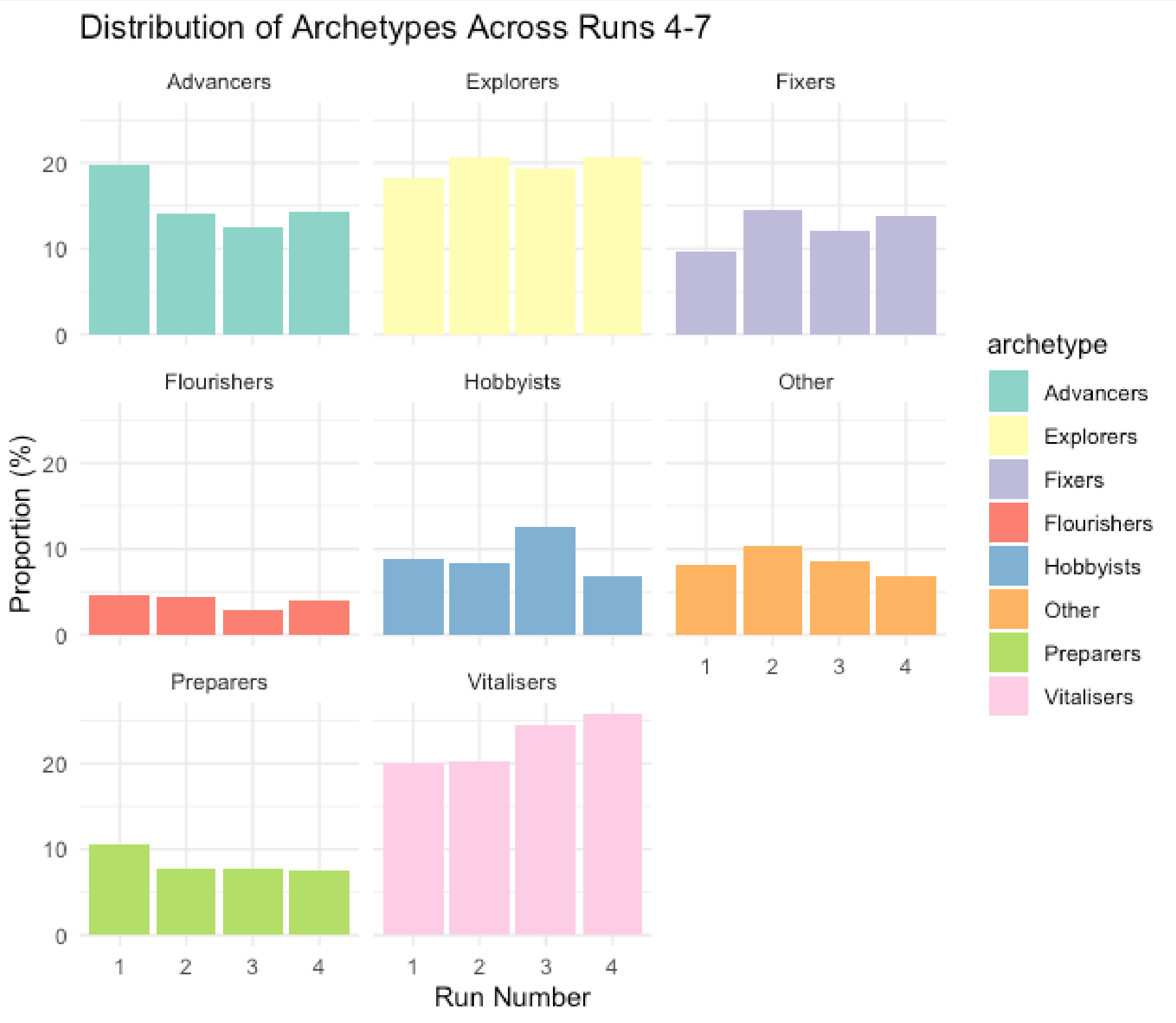


Trend of Archetype Proportions Across Runs 4-7



# 2nd Cycle - Data Modelling

Archetype Proportions



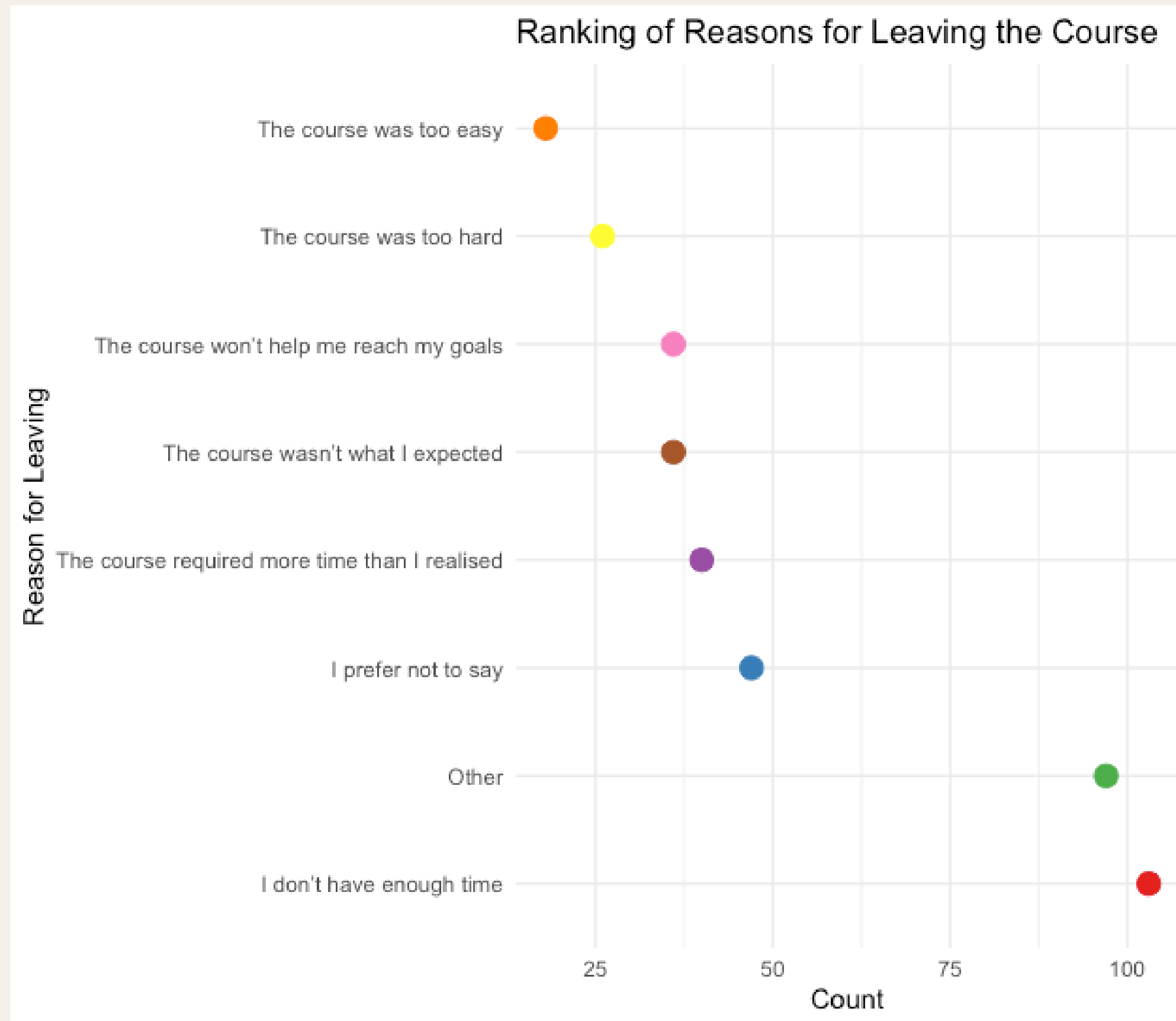
Columns used:

run\_number

archetype

# 2nd Cycle - Data Modelling

Leaving Reasons



Columns used:

run\_number

leaving\_reason

# 2nd Cycle - Data Modelling

Course Reviews



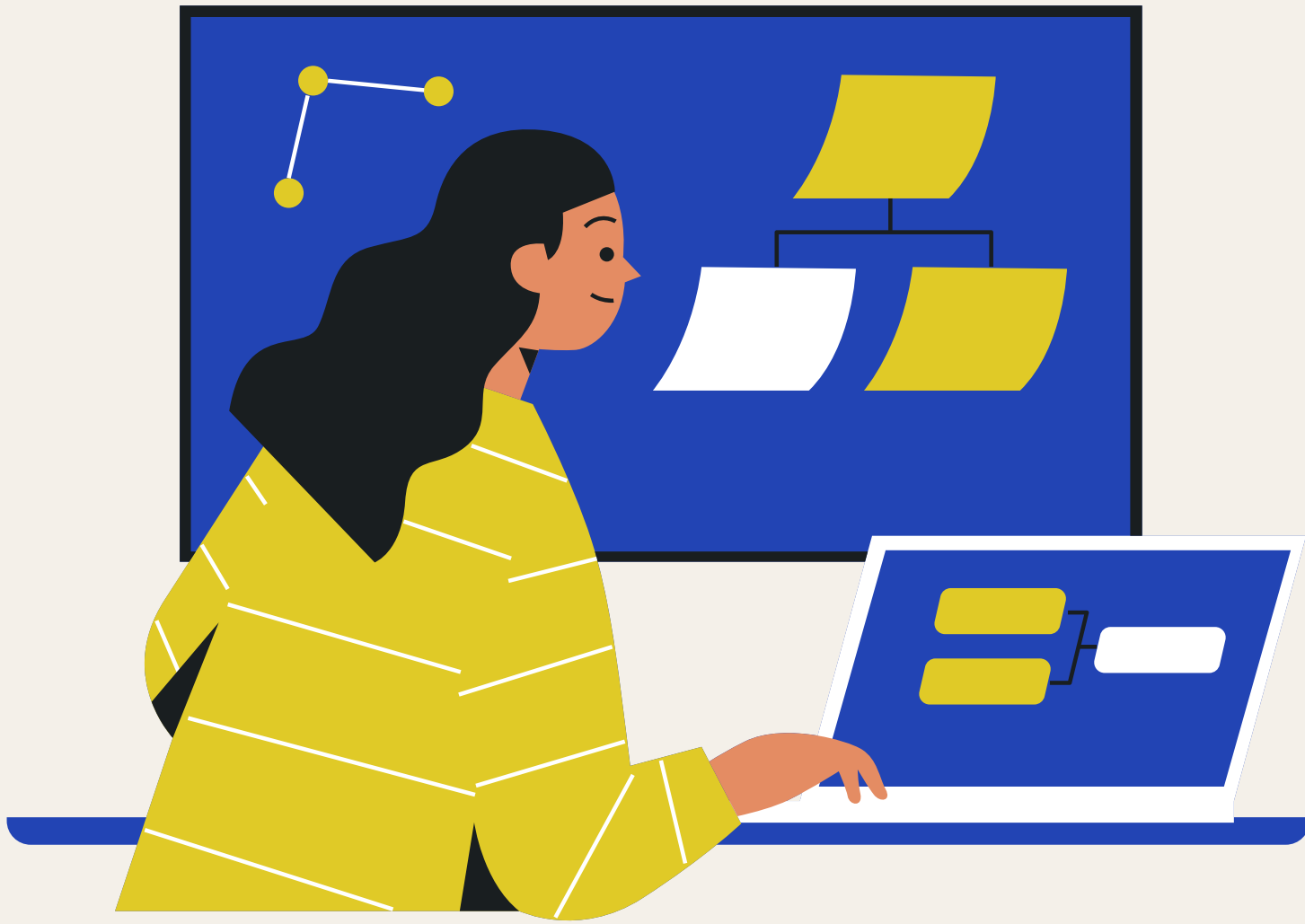
A word cloud of terms from course reviews. The words are arranged in a cluster, with 'information' being the largest and most central. Other prominent words include 'learned', 'content', 'practical', 'easy', 'lot', 'online', 'informative', 'learning', 'privacy', 'protect', and 'answered'. The colors of the words are: 'practical' (pink), 'content' (green), 'learn' (blue), 'easy' (pink), 'lot' (brown), 'online' (pink), 'learned' (grey), 'information' (brown), 'informative' (pink), 'protect' (blue), 'learning' (pink), 'privacy' (blue), and 'answered' (blue).

Columns used:

run\_number

reason

# 2nd Cycle - Evaluation



- **Archetype Proportions:** Shifts in learner archetypes like 'Vitalisers' increasing and 'Preparers' decreasing, reflecting evolving learner demographics and preferences.
- **Leaving Reasons Analysis:** Key reasons for course dropout included time constraints and mismatched expectations, highlighting barriers to learner retention.
- **Course Reviews:** Positive reception of course content, yet indicated areas needing clarity and more engagement.

In-depth analysis of learner feedback revealed a need for course adaptation to meet changing learner needs and market dynamics.

*The course's success relies on adapting to our learners' evolving needs.*

# Thanks

