

# Learner Persistence by Country

Krishnanand Sagar

2026-01-14

## 1. Business Understanding - Cycle 1

The objective of this analysis is to explore how learner persistence varies across countries in a FutureLearn massive open online course (MOOC). From a course provider perspective, understanding where learners remain engaged versus where they disengage early can inform strategic decisions related to course delivery, market focus, and learner support interventions.

Stakeholder: FutureLearn and partner course providers.

Stakeholder Interest: Improving learner engagement and completion rates, and identifying regions where targeted interventions may improve outcomes and potential revenue.

Success Criteria:

- Identify variation in learner persistence across countries
- Identify countries with high early dropout rates
- Generate insights that motivate a second, stakeholder-focused investigation

## 2. Data Understanding

This analysis uses country-level persistence metrics derived from learner step-level activity data. Learner persistence is represented by the maximum step reached during the course, capturing how far a learner progressed before disengaging.

Key variables include:

country: learner's detected country

learners: number of learners per country

avg\_max\_step: average maximum step reached

early\_dropout\_rate: proportion of learners disengaging within the first two steps

These variables allow both overall engagement depth and early disengagement behaviour to be examined at the country level.

## 3. Data Preparation

Learner step-level activity data were cleaned to remove missing country information. For each learner, the maximum step reached was calculated to represent persistence. These learner-level values were then aggregated by country to produce summary metrics including average persistence and early dropout rates.

All data preparation was performed using ProjectTemplate munge scripts, and derived datasets were saved and reused consistently to ensure reproducibility and separation between preparation and analysis stages.

## 4. Analysis - Cycle 1

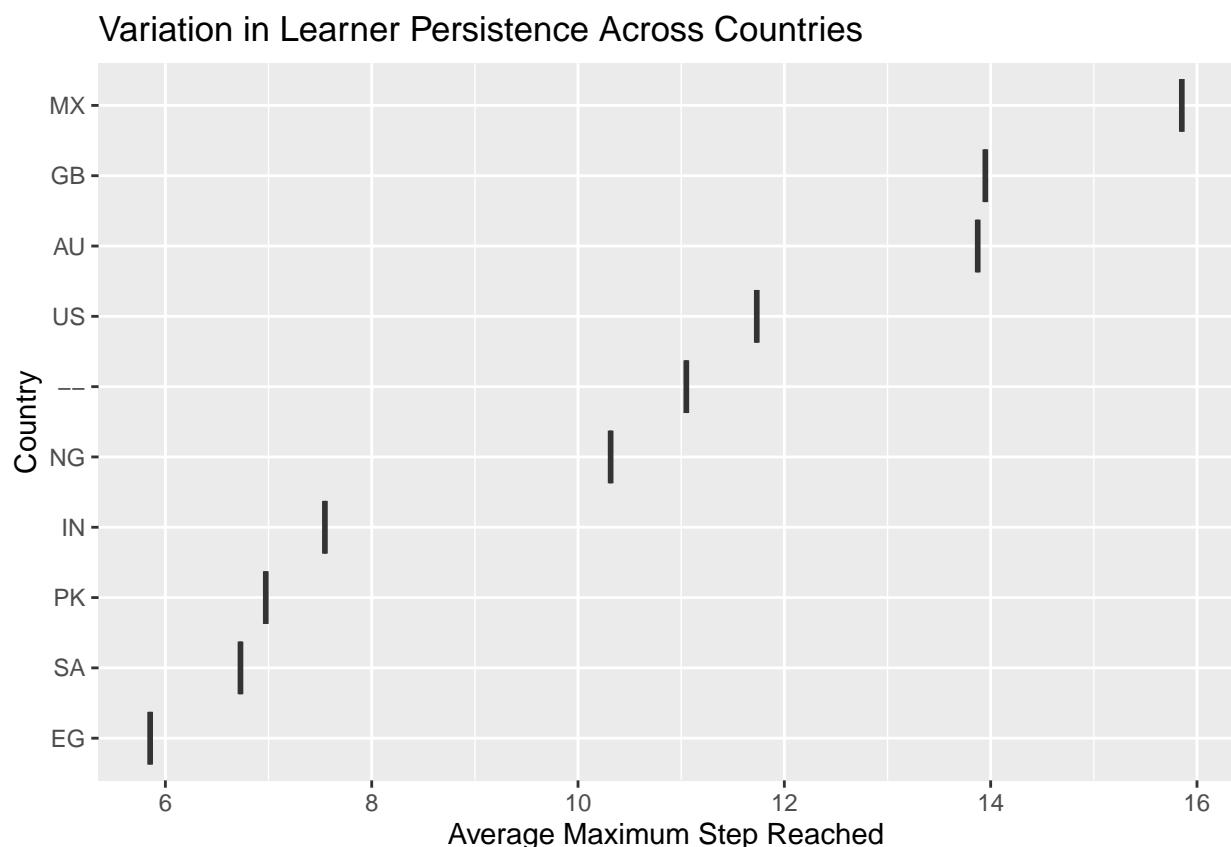
The first analysis examines variation in learner persistence across countries, focusing on the ten countries with the largest learner populations to ensure meaningful comparisons. Average maximum step reached is used to summarise how far learners typically progress.

The second analysis examines early dropout behaviour by comparing the proportion of learners who disengage early versus those who remain active beyond the initial steps. Countries are ordered by early dropout rate to enable direct comparison.

Together, these plots provide complementary perspectives on overall persistence depth and early disengagement patterns.

\*Figure 1 presents the average maximum step reached by learners across the ten countries with the largest learner populations. Focusing on high-volume countries ensures that persistence comparisons are not driven by small sample sizes and highlights meaningful differences in learner engagement depth.

```
country_metrics %>%
  slice_max(learners, n = 10) %>%
  ggplot(aes(x = reorder(country, avg_max_step), y = avg_max_step)) +
  geom_boxplot() +
  coord_flip() +
  labs(
    title = "Variation in Learner Persistence Across Countries",
    x = "Country",
    y = "Average Maximum Step Reached"
  )
```

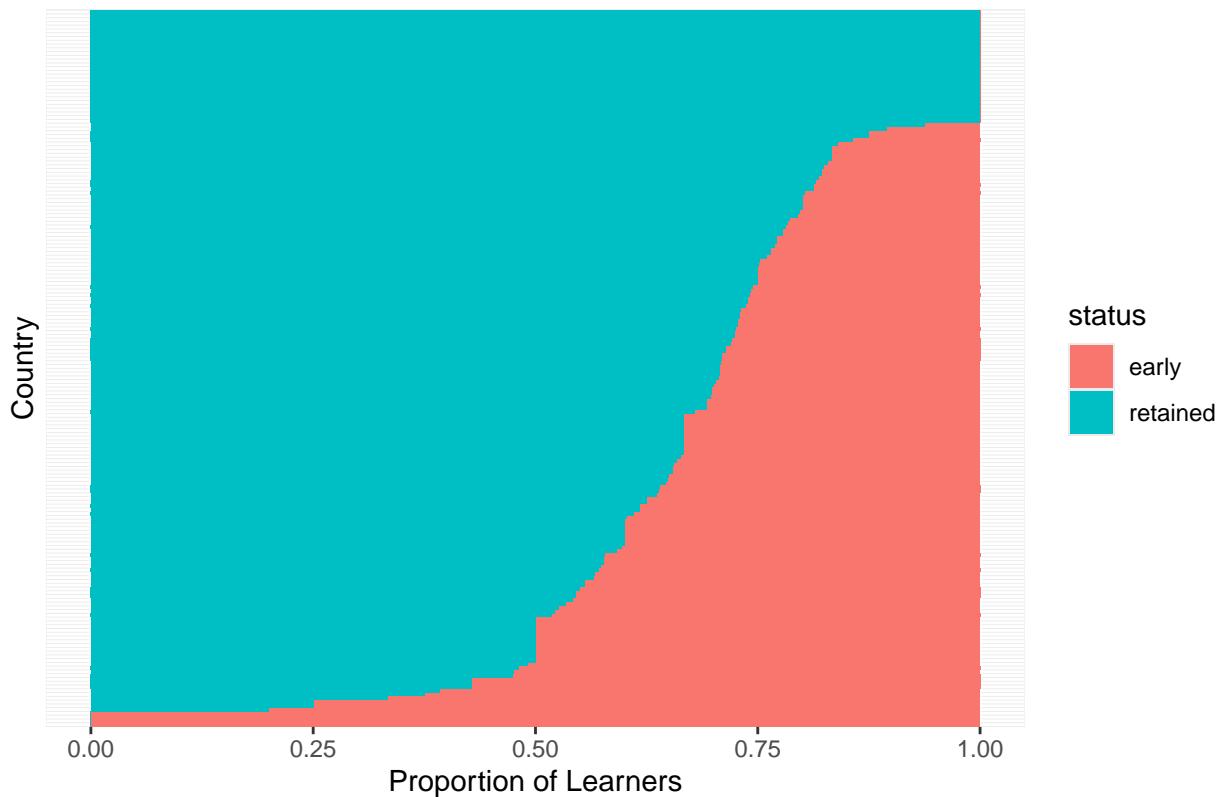


The figure shows substantial variation in learner persistence across countries, with some countries exhibiting consistently higher progression through the course than others.

\*Figure 2 decomposes learner engagement into early dropout and retained learners, allowing a clearer examination of disengagement patterns at the earliest stages of the course.

```
country_metrics %>%
  arrange(desc(early_dropout_rate)) %>%
  mutate(
    country = factor(country, levels = country),
    early = early_dropout_rate,
    retained = 1 - early_dropout_rate
  ) %>%
  select(country, early, retained) %>%
  pivot_longer(
    cols = c(early, retained),
    names_to = "status",
    values_to = "proportion"
  ) %>%
  ggplot(aes(x = country, y = proportion, fill = status)) +
  geom_col() +
  coord_flip() +
  labs(
    title = "Early Dropout vs Retained Learners by Country",
    x = "Country",
    y = "Proportion of Learners"
  ) +
  theme(
    axis.text.y = element_blank(),
    axis.ticks.y = element_blank()
  )
```

## Early Dropout vs Retained Learners by Country



Countries with high early dropout proportions show disengagement occurring very early, suggesting that initial course structure or accessibility may present barriers to sustained participation.

### 5. Evaluation - Cycle 1

The analysis reveals clear differences in learner persistence across countries. Some countries exhibit higher average progression through the course, indicating sustained engagement, while others show high early dropout rates, suggesting disengagement occurs very early.

From a business perspective:

High-persistence countries represent stable markets for fully online delivery

High early-dropout countries may face accessibility barriers, expectation mismatches, or insufficient early support

These findings motivate a second investigation focused on identifying and addressing low-persistence regions.

### Business Understanding - Cycle 2

The objective of Cycle 2 is to translate persistence patterns identified in Cycle 1 into actionable insights for stakeholders. Specifically, this cycle evaluates whether countries with low learner persistence may benefit from offline or blended learning interventions aimed at improving engagement and completion.

## Data Understanding - Cycle 2

The objective of Cycle 2 is to translate persistence patterns identified in Cycle 1 into actionable insights for stakeholders. Specifically, this cycle evaluates whether countries with low learner persistence may benefit from offline or blended learning interventions aimed at improving engagement and completion.

## Modelling / Analysis - Cycle 2

Rather than predictive modelling, this cycle applies a rule-based segmentation approach. Countries are classified into low, medium, or high persistence categories using quantile-based thresholds applied to learner progression and early dropout rates. This approach prioritises interpretability for stakeholders.

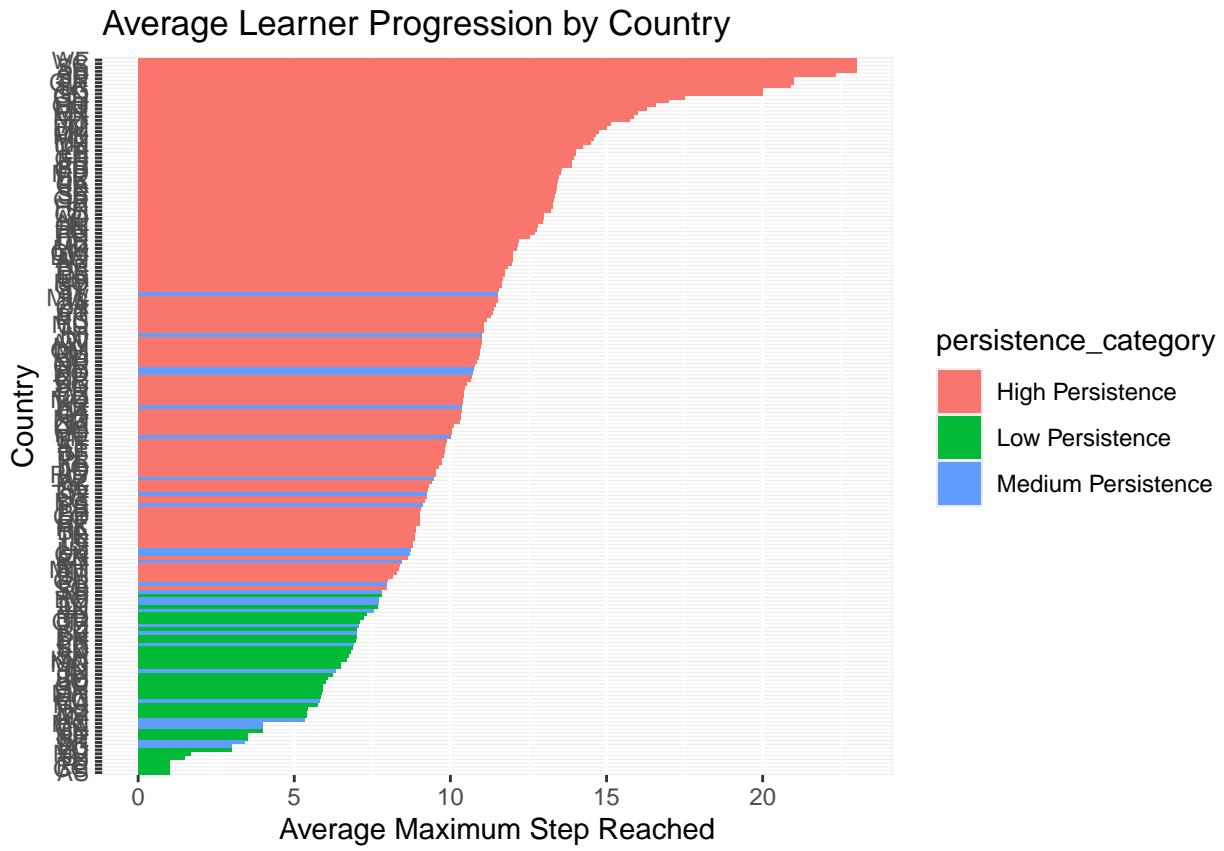
```
country_cycle2
```

```
## # A tibble: 190 x 6
##   country learners avg_max_step median_max_step early_dropout_rate
##   <chr>      <dbl>        <dbl>          <dbl>                <dbl>
## 1 --          382         11.0           8.5                 0.280
## 2 AD           1            23             23                  0
## 3 AE          113         9.86            7                 0.354
## 4 AF           15           7.8             4                 0.467
## 5 AG           2            12             12                  0
## 6 AL           62           9.31            5                 0.290
## 7 AM           18           5.39            3                 0.444
## 8 AO           15           13             11                  0
## 9 AR           55           8.62            5                 0.345
## 10 AS          1            1              1                  1
## # i 180 more rows
## # i 1 more variable: persistence_category <chr>
```

## Results and Visualisation - Cycle 2

\*Figure 3 summarises average learner progression by country, segmented by persistence category. This visualisation highlights countries identified as low persistence and supports stakeholder prioritisation of intervention targets.

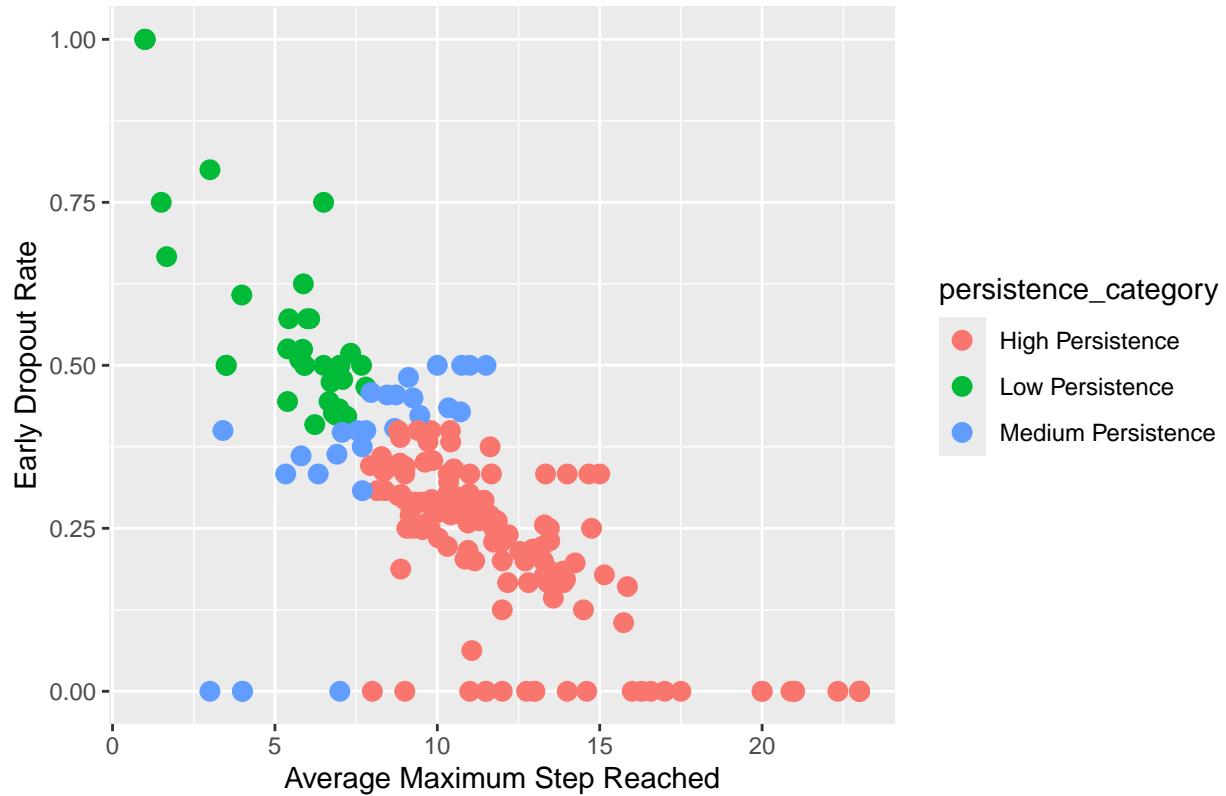
```
ggplot(
  country_cycle2,
  aes(x = reorder(country, avg_max_step),
  y = avg_max_step,
  fill = persistence_category)
) +
  geom_col() +
  coord_flip() +
  labs(
    title = "Average Learner Progression by Country",
    x = "Country",
    y = "Average Maximum Step Reached"
)
```



\*Figure 4 illustrates the relationship between average learner progression and early dropout rates, providing a complementary view of engagement dynamics across persistence categories.

```
ggplot(
  country_cycle2,
  aes(x = avg_max_step,
      y = early_dropout_rate,
      color = persistence_category)
) +
  geom_point(size = 3) +
  labs(
    title = "Early Dropout Rate vs Learner Engagement",
    x = "Average Maximum Step Reached",
    y = "Early Dropout Rate"
)
```

## Early Dropout Rate vs Learner Engagement



\*Table 1 lists countries classified as low persistence, representing priority candidates for targeted offline or blended learning interventions.

```
low_persistence_countries
```

```
## # A tibble: 35 x 6
##   country learners avg_max_step median_max_step early_dropout_rate
##   <chr>     <dbl>        <dbl>           <dbl>                <dbl>
## 1 AS          1            1              1                 1
## 2 CG          1            1              1                 1
## 3 RE          1            1              1                 1
## 4 TD          1            1              1                 1
## 5 LC          5            3              2                 0.8
## 6 BI          4            6.5             1                 0.75
## 7 BJ          4            1.5             1                 0.75
## 8 MV          3            1.67            1                 0.667
## 9 MN          8            5.88            2                 0.625
## 10 TN         51            3.98            1                 0.608
## # i 25 more rows
## # i 1 more variable: persistence_category <chr>
```

## Evaluation - Cycle 2

Countries classified as low persistence exhibit both low average learner progression and high early dropout rates. These patterns suggest the presence of structural or contextual barriers rather than a lack of learner interest.

## **Deployment and Recommendations**

It is recommended that course providers pilot offline or blended learning initiatives in selected low-persistence countries identified in this analysis. Cycle 1 persistence metrics should be retained as a baseline, and the same analytical framework reapplied post-intervention to evaluate changes in learner engagement, dropout rates, and potential revenue impact.