Learner Persistence by Country

Krishnanand Sagar

2026-01-13

1. Business Understanding - Cycle 1

The objective of this analysis is to understand how learner persistent varies across countries in an online course from a provider perspective identifying countries where learners remain engaged for longer verses those where learners drop out early can inform strategic decisions across course delivery market focus and engagement interventions

Specifically this cycle aims: Compare learner persistent across Identify countries with high early Provide evidence to support targeted engagement strategies in later analysis cycles

2. Data Understanding

The analysis uses country level persistent metrics derived from learner step activity data each learner's engagement is represented by the maximum step reached during the course capturing how far they progressed before disengaging Key variables used in the Analysis include Country detected learner country Learner's number of learners per country Average max step average maximum step reached by learners In the country Pearly drop out rate proportion of learners disengaging within the first two steps These variables allow both overall persistent and early disengagement behaviour to be examined at the country level

3. Data Preparation

Learner step activity data were cleaned with philtre 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 contrary to produce summary matrix including average persistence and early dropout rate the resulting country level data set was saved to disc and used consistently across all analysis to ensure reproducibility and separation between data preparation and analysis stage

4. Analysis - Cycle 1

The first plot examines variation in learner persistent across countries by focusing on the 10 countries with the largest number of learners using average maximum step reached this visualisation highlights how far learners typically progress in different region while ensuring comparisons are based on sufficiently large learner populations the second plot focuses on early dropout behaviour showing the proportion of learners who disengage early versus those who remain active beyond the initial steps countries are ordered by early dropout rate allowing direct comparison of engagement composition across regionstogether these plots provide complementary perspective one summarising overall persistence depth and other revealing how engagement breaks down at the earliest stage of the course

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.5.2
```

```
## Warning: package 'ggplot2' was built under R version 4.5.2
```

```
## Warning: package 'dplyr' was built under R version 4.5.2
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.1     v stringr   1.5.2
## v ggplot2   4.0.1     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
country_metrics<-read_csv("../cache/country_metrics_cycle1.csv")
```

```
## Rows: 190 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): country
## dbl (4): learners, avg_max_step, median_max_step, early_dropout_rate
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
country_metrics %>%
  slice_max(learners, n = 10) %>%   # top 10 by volume, not performance
  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"
  )
```
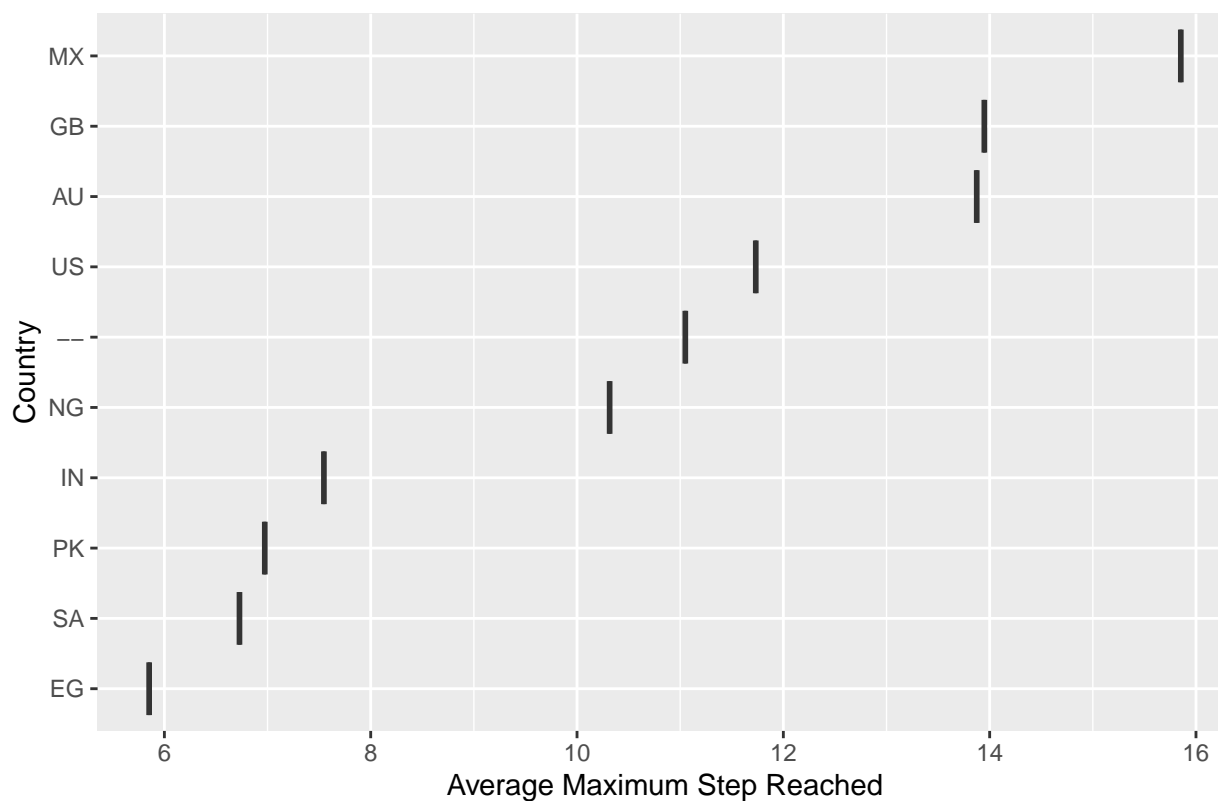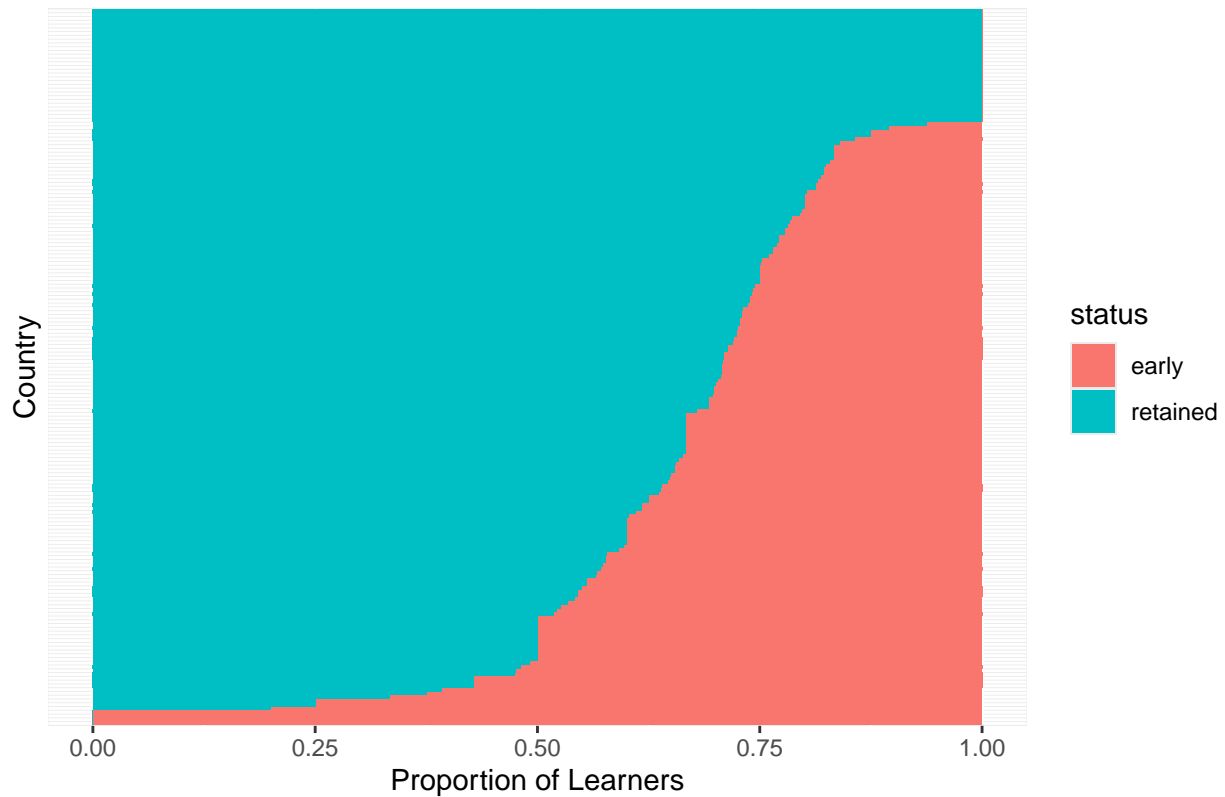
Variation in Learner Persistence Across Countries


```
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

5. Evaluation - Cycle 1

The analysis demonstrates clear difference in learner persistent patterns across country some countries show relatively high average progression through the course indicating sustained learner engagement while other exhibit high early dropout rates suggesting disengagement occurs very quickly From a business perspective Countries with stronger persistent represent stable markets for fully online course delivery Country with high early dropouts may face barriers such as accessibility expectations mismatch or insufficient early support

These findings justify further investigation in cycle two were targeted strategies can be evaluated to improve engagement and reduce early dropout in low persistent regions