Exploratory Data Analysis of Online Learning Activity
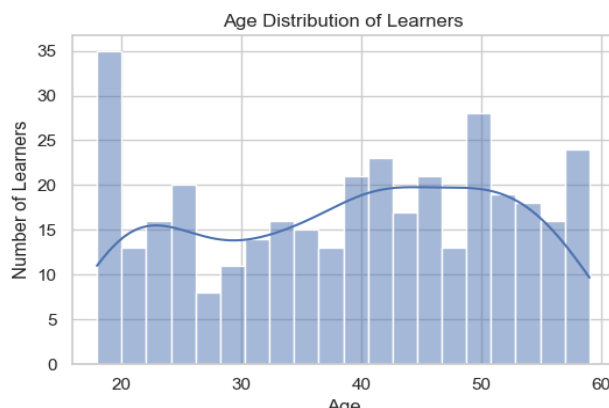
1. Dataset Overview

The dataset represents learner activity on an online learning platform. It includes demographic information, engagement metrics (videos watched, assignments submitted), academic performance (quiz scores), and course completion status. The analysis aims to understand learner behavior and identify patterns related to engagement and drop-off.
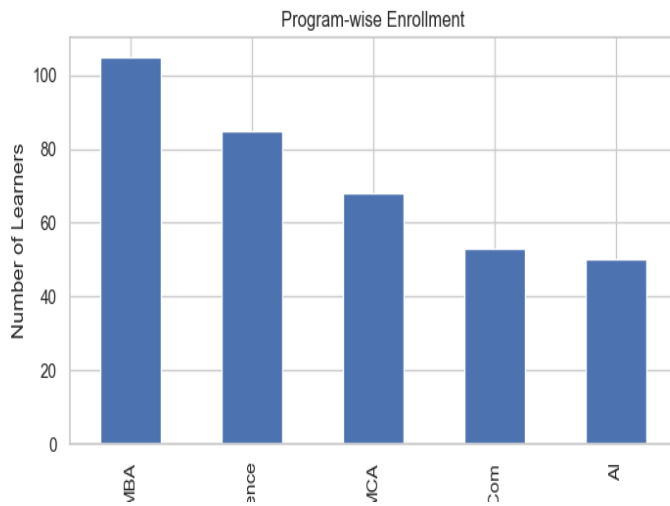
2. Data Cleaning Summary

- Unrealistic age values were removed
- Negative engagement values were filtered out
- Missing numerical values were filled using median values
- Missing support ticket values were assumed to be zero
- Records missing last login dates were removed
- A new feature, days_active, was created to measure learner engagement duration
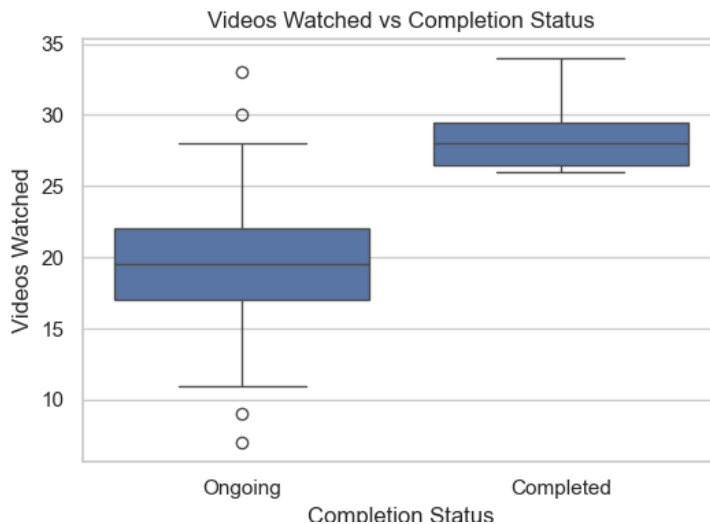
3. Key Findings

- Most learners belong to the young to mid-career age group
- Certain programs show higher enrollment, indicating greater demand
- Learners who watch more videos are more likely to complete courses
- Assignment submission is positively related to quiz performance
- Engagement levels vary across programs

Age Distribution of Learners
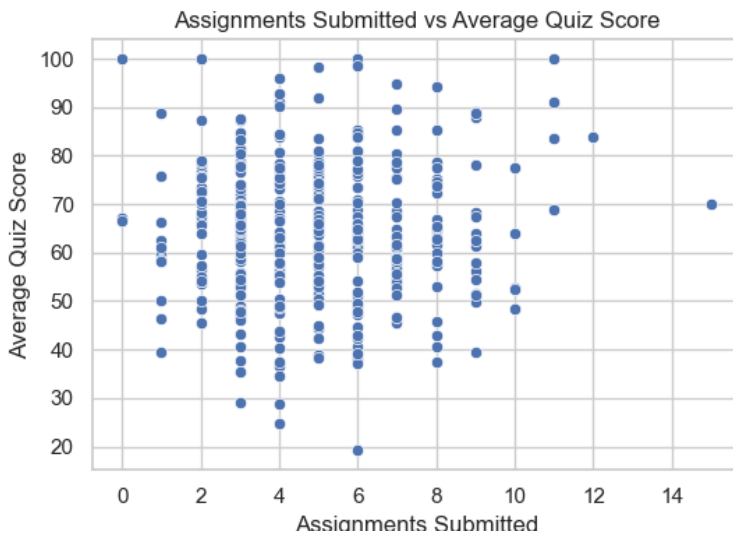
Program-wise Enrollment
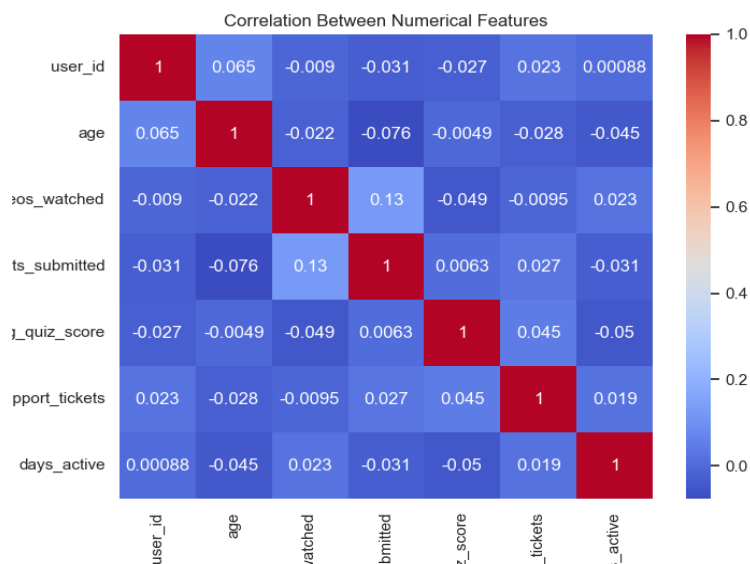
4. Completion vs Drop-off Insights

- Learners who completed courses show higher engagement in terms of videos watched, assignments submitted, quiz scores, and days active

- Dropped learners tend to have lower engagement across all metrics

- Early signs of drop-off include fewer videos watched and fewer assignments submitted

Videos Watched vs Completion Status

Assignments Submitted vs Average Quiz Score

5. Data Quality Observations

- Some unrealistic values (e.g., age and engagement metrics) were present and corrected

- Missing values were handled appropriately based on context

- Engagement duration was derived using login dates

Correlation Between Numerical Features

The heatmap highlights positive relationships between assignments, quiz scores, and videos watched, reinforcing that active participation improves outcomes.

6. Recommendations

- Identify low-engagement learners early and send reminders or motivational nudges

- Encourage assignment completion through progress tracking

- Monitor quiz participation as an indicator of learner commitment

- Provide targeted support for learners who raise multiple support tickets