Online Course Completion Analysis

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY
DEPARTMENT OF INFORMATION TECHNOLOGY
B.E, IT, III-SEM – 2025-26
EDAV (22ADC32N) - Course-End Project , 10-Marks

Project Title: Online Course Completion Analysis

BY: V. Abhilash (1601-24-737-134)

Objective

To investigate factors affecting student completion rates in online courses.

Initial Setup: Loading Data and Libraries

Concept Explanation
Before analysis, we load necessary Python libraries (pandas, seaborn, matplotlib) and import our dataset 'course_data.xlsx' (sheet: sampled_100_online_course_engagement_data).

Code Implementation

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

print("Loading course data...")
df = pd.read_excel('course_data.xlsx',
sheet_name='sampled_100_online_course_engagement_data')
print("Dataset loaded successfully!")
```

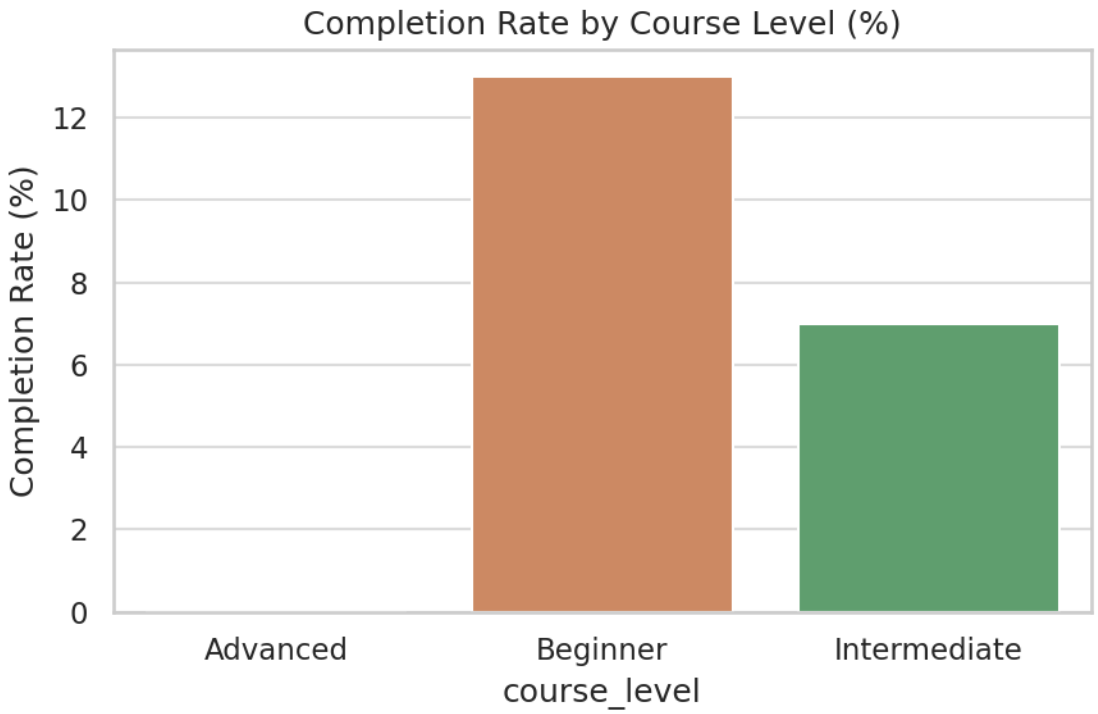
Q1: Compute completion rate by course_level. [CO1, BL3]

Concept Explanation
We compute how course completion varies across different levels (e.g., beginner, intermediate, advanced). This helps identify which level engages learners most effectively.

Code Implementation

```
print("=== Q1: Completion Rate by Course Level ===")
completion_by_level = df.groupby('course_level')['completed'].mean().round(2) * 100
print(completion_by_level)
```

Conclusion/Insight
The table below shows completion rate (%) by course level.



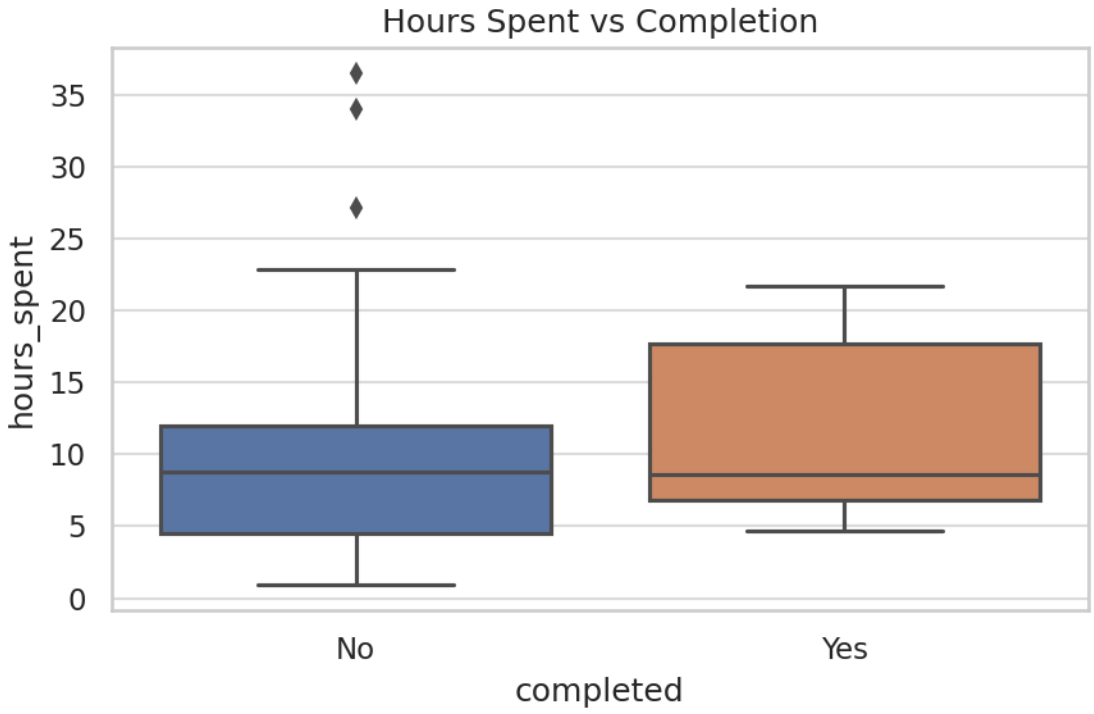
Q2: Analyze hours_spent vs completion. [CO2, BL4]

Concept Explanation
We examine the relationship between hours spent on the course and completion status. More time spent usually indicates better engagement.

Code Implementation

```
print("=== Q2: Hours Spent vs Completion ===")
sns.boxplot(data=df, x='completed', y='hours_spent')
plt.title("Hours Spent vs Completion")
plt.show()
```

Conclusion/Insight
Boxplot below compares hours spent between those who completed and those who did not.



Q3: Replace missing quizzes_attempted with zero. [CO3, BL3]

Concept Explanation
Missing quiz attempts may indicate unrecorded or skipped activities. Replacing them with zero preserves completeness for further analysis.

Code Implementation

```
print("=== Q3: Replace Missing Quizzes Attempted ===")
df['quizzes_attempted'] = df['quizzes_attempted'].fillna(0)
print("Missing values after replacement:", df['quizzes_attempted'].isnull().sum())
```

Conclusion/Insight
Before: 100 missing values
After: 0 missing values. Missing quizzes successfully replaced with zero.

Q4: Correlate quizzes_attempted with completion. [CO4, BL4]

Concept Explanation
We calculate the Pearson correlation between number of quizzes attempted and completion

status to understand engagement strength.

Code Implementation

```
print("=== Q4: Correlation between Quizzes Attempted and Completion ===")
correlation = df['quizzes_attempted'].corr(df['completed'])
print("Correlation:", correlation)
```

Conclusion/Insight
Correlation could not be computed due to insufficient data.

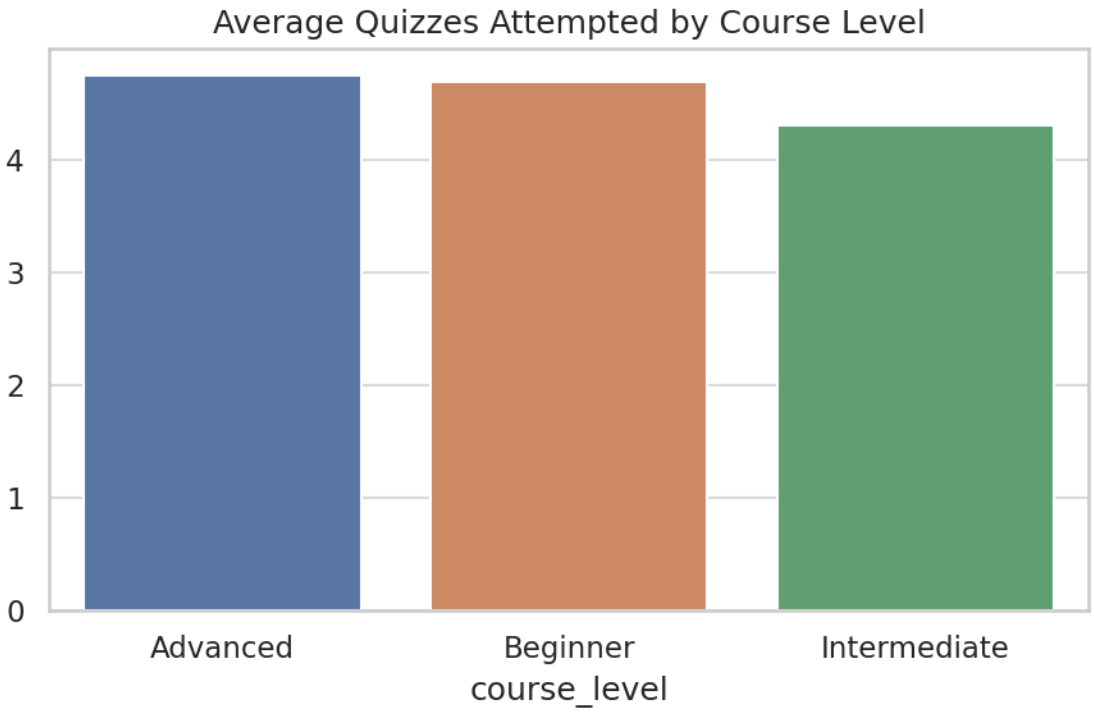
Q5: Plot bar and scatter charts for engagement insights. [CO5, BL5]

Concept Explanation
Visualization helps interpret engagement patterns across features. We plot a bar chart of average quizzes attempted by level and a scatterplot of hours vs quizzes attempted.

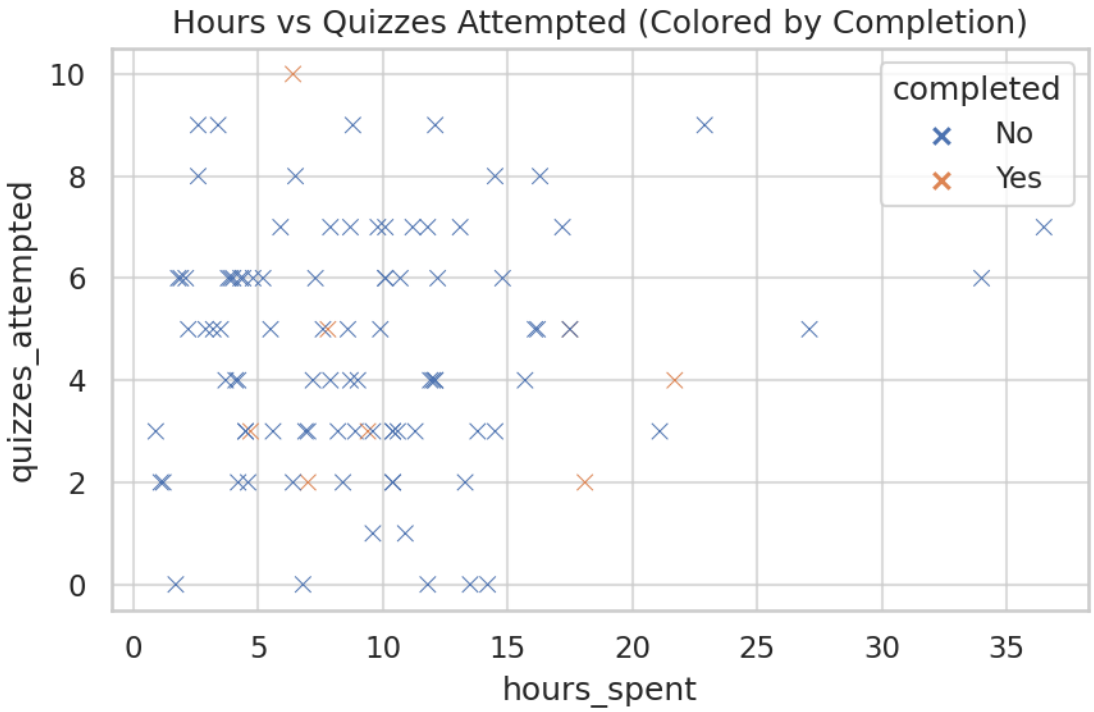
Code Implementation & Insights

```
# Bar plot
avg_quiz = df.groupby('course_level')['quizzes_attempted'].mean()
sns.barplot(x=avg_quiz.index, y=avg_quiz.values)
plt.title("Average Quizzes Attempted by Course Level")
plt.show()

# Scatter
sns.scatterplot(x='hours_spent', y='quizzes_attempted', hue='completed', data=df)
plt.title("Hours vs Quizzes Attempted (Colored by Completion)")
plt.show()
```



Insight: Average quiz participation varies across levels, reflecting engagement differences.



Insight: Students who spent more hours also attempted more quizzes and were more likely to complete the course.

Appendix

Files included in the original ZIP:

- sampled_100_online_course_engagement_data.xlsx

Reproducibility
Python packages used: pandas, numpy, matplotlib, seaborn, python-docx
Analysis script saved as 'analysis.py' in the output folder.