



---

# A Data-Driven Gamification Approach to Monitor and Predict the Students' Academic Performance

Sherif Abdelhamid, Kolby Quigg, Mona Aly

Virginia Military Institute

[abdelhamidse@vmi.edu](mailto:abdelhamidse@vmi.edu)

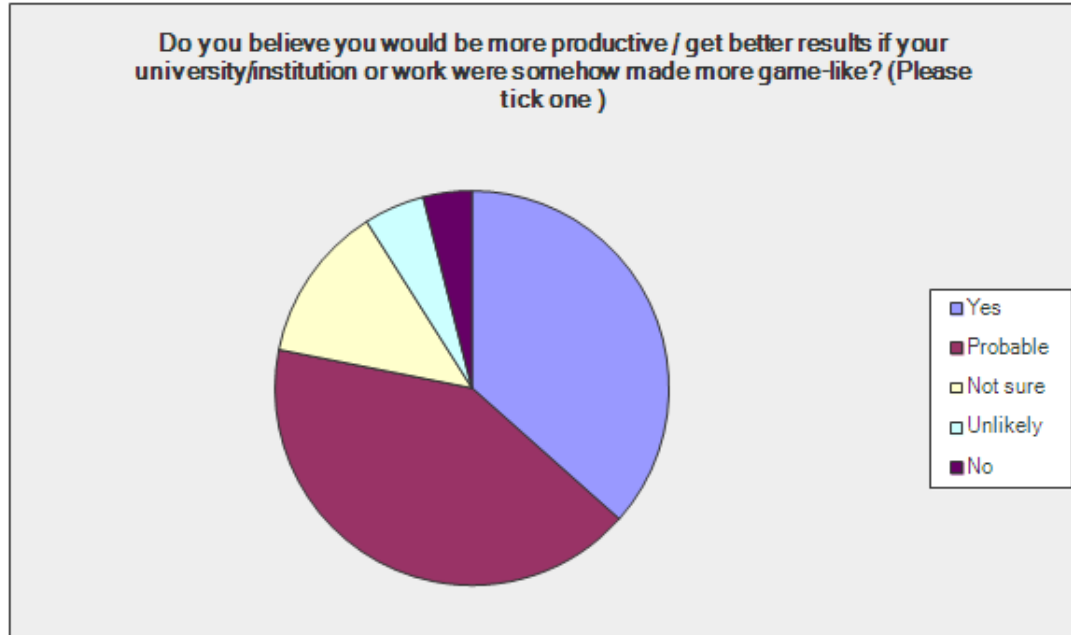


# Overview

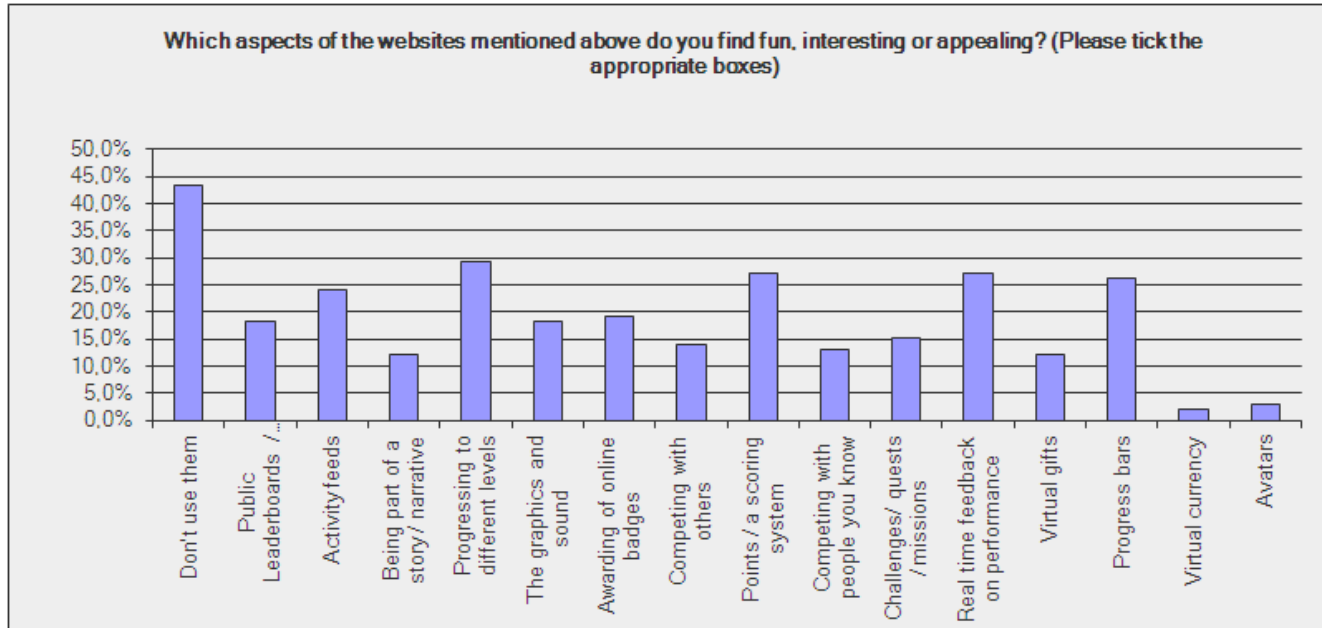
- Learning environments can be more stimulating by incorporating game design elements into the curriculum.
- Gamification improves student motivation, learning, and academic performance.
- Gamification can be implemented in schools at different grade levels, from kindergarten to 12 years of basic education.
- The use of gamification in education has been demonstrated in several fields, including computer science, mathematics, astronomy, physics, medicine, and law.



# Overview



# Overview



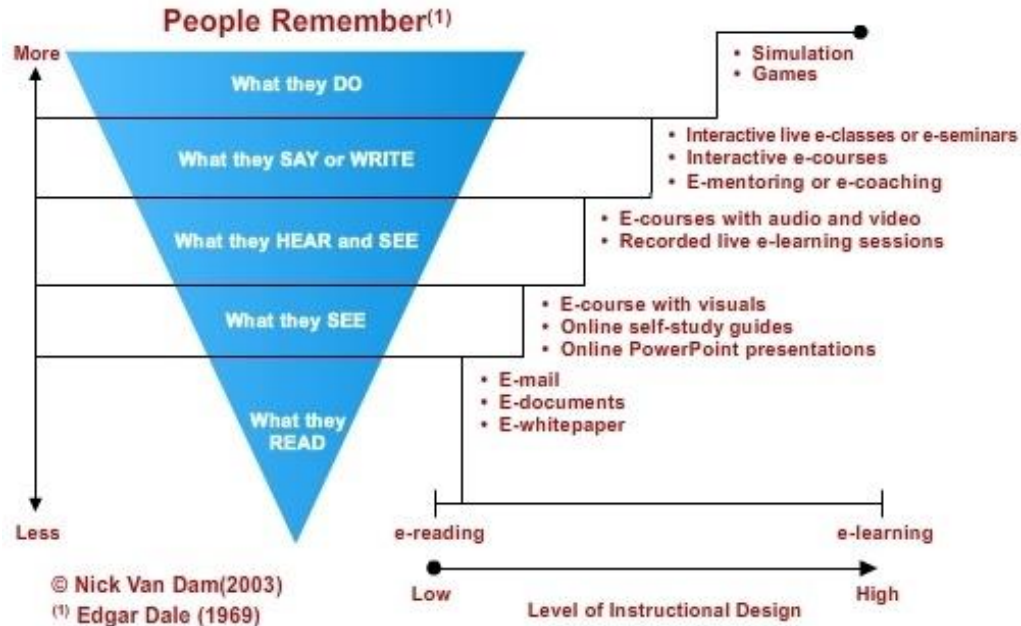


# Overview

## Gamification in workplace training/education

- 33% would like more game-like features in their employee training software.
- 61% of the respondents receive training with gamification.
- 83% of those who receive gamified training feel motivated, while 61% of those who receive non-gamified training feel bored and unproductive.

# Overview





## Problem

- Current research works have extensively explored how gamification can improve students' engagement and motivation.
- Few studies have examined how it can track and predict students' academic performance. This can help educators identify students who may be struggling and provide them with additional support before they fall behind.
- Additionally, tracking and predicting academic performance can help educators tailor their instruction to meet the needs of individual students.



# Literature Review

- Based on the current literature review, we found underrepresented research areas related to the use of gamification in education, including:
  - Performance prediction (the focus of this paper).
  - Long-term effectiveness
  - Individual differences
  - Gamification integration with other teaching methods
  - Ethical considerations and concerns



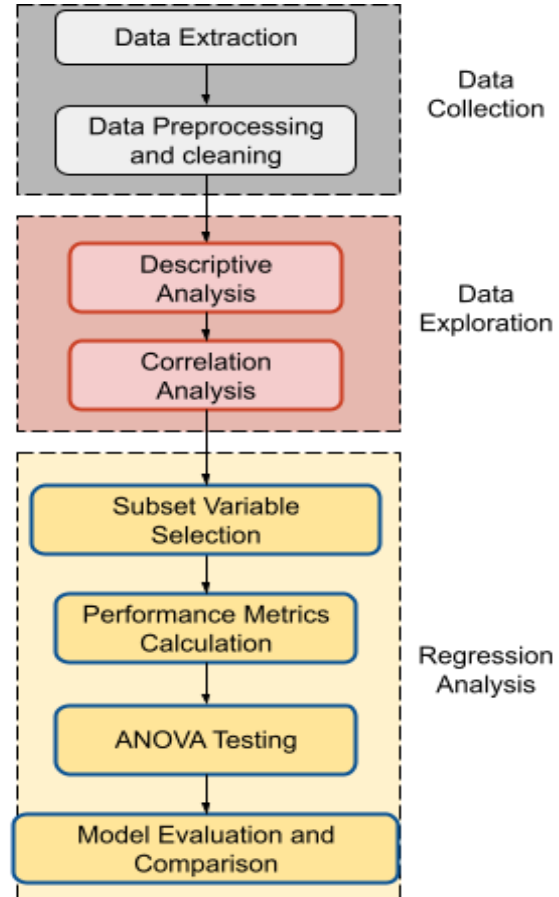
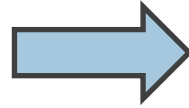


## Solution

- This study investigates how data collected from gamification activities can help instructors monitor and predict students' performance in the classroom.
- We used **Quizizz**, a web-based tool that delivers quiz questions in a game-like manner. By answering questions interactively, students earn points and rewards.



# Methodology





# Data Collection

- We identified two experimental groups representing sophomore students in two computer science courses:
  - Database Management Systems
  - Data Structures & Applications.
- Total number of students was 32.



# Data Collection

Quizizz collects several types of data that can be used to inform educators about their students' performance and engagement with the platform. The data collected by Quizizz includes:

- Performance data
- Engagement data
- Question-level data
- Usage data

# Data Collection

Sample quiz report  
by Quizizz

Instructor-paced quiz

## Database SQL 2 [Edit](#)

February 14th 2022, 11:33 AM (a year ago)

[View quiz](#)

[Flashcards](#)

[Live Dashboard](#)



Accuracy



Questions



Participant Attempts

[Participants](#)

[Questions](#)

[Overview](#)

[Tags](#)

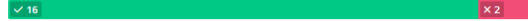


[Share report](#)

Sort By:

Accuracy

[Email all parents](#)



89% Accuracy Score 14780

[Email Parent](#)



94% Accuracy Score 14490

[Email Parent](#)



83% Accuracy Score 13730

[Email Parent](#)



83% Accuracy Score 12740

[Email Parent](#)



78% Accuracy Score 12100

[Email Parent](#)



78% Accuracy Score 11970

[Email Parent](#)



# Data Collection

Individual student data  
collected by Quizizz for sample quiz



Ethan

Jan 23, 2023, 11:43:am

Windows

Chrome

Print

Delete



✓ 8 Correct

✗ 1 Incorrect



89%

Accuracy

8/9

Points

7550

Score

Evaluate responses



✓ Correct

Multiple choice

1 point

3 secs

Question

1. The purpose of a database is to help people keep track of things.

Ethan's response

True

Correct Answer

True

✓ Correct

Multiple choice

1 point

5 secs

Question



# Data Collection

Participant performance data for all students

Attempt #	Accuracy	Score	Correct	Incorrect	Unattempted	Total Time Taken	Started At	Info
18	88 %	14780	16	2	0	01:54	Mon 14 Feb 2022,04:34 PM	hrome on Window
18	94 %	14490	17	1	0	03:25	Mon 14 Feb 2022,04:34 PM	Safari on Mac OS X
18	83 %	13730	15	3	0	02:03	Mon 14 Feb 2022,04:34 PM	hrome on Window
18	83 %	12740	15	3	0	03:41	Mon 14 Feb 2022,04:34 PM	hrome on Window
18	77 %	12100	14	4	0	03:10	Mon 14 Feb 2022,04:34 PM	obile Safari on iO
18	77 %	11970	14	4	0	03:12	Mon 14 Feb 2022,04:34 PM	me Mobile on Anc
18	77 %	11660	14	4	0	04:02	Mon 14 Feb 2022,04:34 PM	hrome on Window
18	77 %	11660	14	4	0	03:25	Mon 14 Feb 2022,04:34 PM	hrome on Window
17	55 %	9400	10	7	1	01:33	Mon 14 Feb 2022,04:34 PM	obile Safari on iO
18	61 %	8940	11	7	0	04:11	Mon 14 Feb 2022,04:34 PM	hrome on Window

# Data Collection

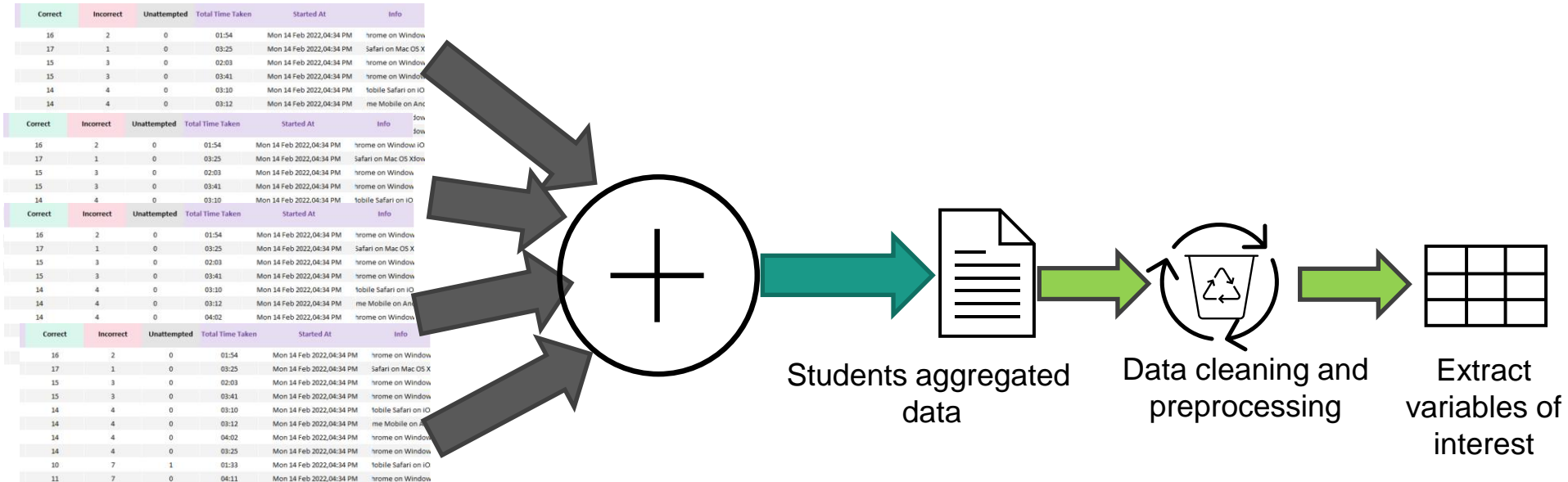
## Participant performance data for all questions

Which statement is used to extract data from a database?	0	Select	Select	Select	Select	Get	Select	Select	Get	Select	Get
If you were wanting to filter data, which clause would you use?	0	Where	Order by	Where	Where	Where	Where	Order by	Where	From	Order by
Which word is missing from the following SQL query: SELECT * FROM Designer ORDER BY Designer	0	From	From	From	From	From	From	From	From	From	From
How do we select all rows for the "Designer" table?	0	SELECT * FROM Designer	SELECT * FROM Designer	SELECT * FROM Designer	SELECT * FROM Designer	SELECT * FROM Designer	SELECT * FROM Designer	SELECT * FROM Designer	SELECT * FROM Designer	SELECT * FROM Designer	SELECT * FROM Designer
How would we script a SQL query to select "Designer" items?	0	SELECT Designer	SELECT Designer	SELECT Designer	SELECT Item From Designer	SELECT Designer	SELECT Designer	SELECT Item From Designer	SELECT Designer	SELECT Item From Designer	SELECT Designer
Which query would return only the fields Designer and Item?	0	SELECT Designer	SELECT Designer	SELECT Designer	SELECT Designer	SELECT Designer	SELECT Designer	SELECT Designer	SELECT Designer	SELECT Designer	SELECT Designer
How would you display Chairs in the Items table?	0	SELECT * FROM Chairs	SELECT * FROM Chairs	SELECT * FROM Chairs	SELECT * FROM Chairs	SELECT * FROM Chairs	SELECT * FROM Chairs	SELECT * FROM Chairs	SELECT * FROM Chairs	SELECT * FROM Chairs	SELECT * FROM Chairs
Which symbol would you use to represent more than 50?	0	> 50	> 50	> 50	> 50	> 50	> 50	> 50	> 50	> 50	< 50
Which of the queries is most likely to have produced the following results?	0	ANSWER: SELECT * FROM Designer	ANSWER: SELECT * FROM Designer	ANSWER: SELECT * FROM Designer	ANSWER: SELECT * FROM Designer	ANSWER: SELECT * FROM Designer	ANSWER: SELECT * FROM Designer	ANSWER: SELECT * FROM Designer	ANSWER: SELECT * FROM Designer	ANSWER: SELECT * FROM Designer	ANSWER: SELECT * FROM Designer
What is missing in this code? SELECT * FROM Designer	0	Table name	Table name	Table name	Table name	Table name	Table name	Table name	Table name	Table name	Table name
What is the correct order of clauses in a SQL statement?	0	SELECT, FROM	SELECT, FROM	SELECT, FROM	SELECT, FROM	SELECT, FROM	SELECT, FROM	SELECT, FROM	SELECT, FROM	SELECT, FROM	SELECT, FROM
The HAVING clause does which of the following?	0	Acts like a WHERE clause	Acts like a WHERE clause	Acts like a WHERE clause	Acts EXACTLY like a WHERE clause	Acts like a WHERE clause	Acts like a WHERE clause	Acts like a WHERE clause	Acts like a WHERE clause	Acts like a WHERE clause	Acts like a WHERE clause
AVG, COUNT, MAX, MIN and SUM are known as what?	0	Aggregate functions	Aggregate functions	Aggregate functions	Aggregate functions	Aggregate functions	Aggregate functions	Aggregate functions	Aggregate functions	Aggregate functions	Aggregate functions
Which SQL function is used to find the average of a set of numbers?	0	Avg()	Avg()	Avg()	Avg()	Average()	Average()	Avg()	Avg()	Average()	Average()



# Data Aggregation

We collected the data for all students from the two courses for all the quizzes



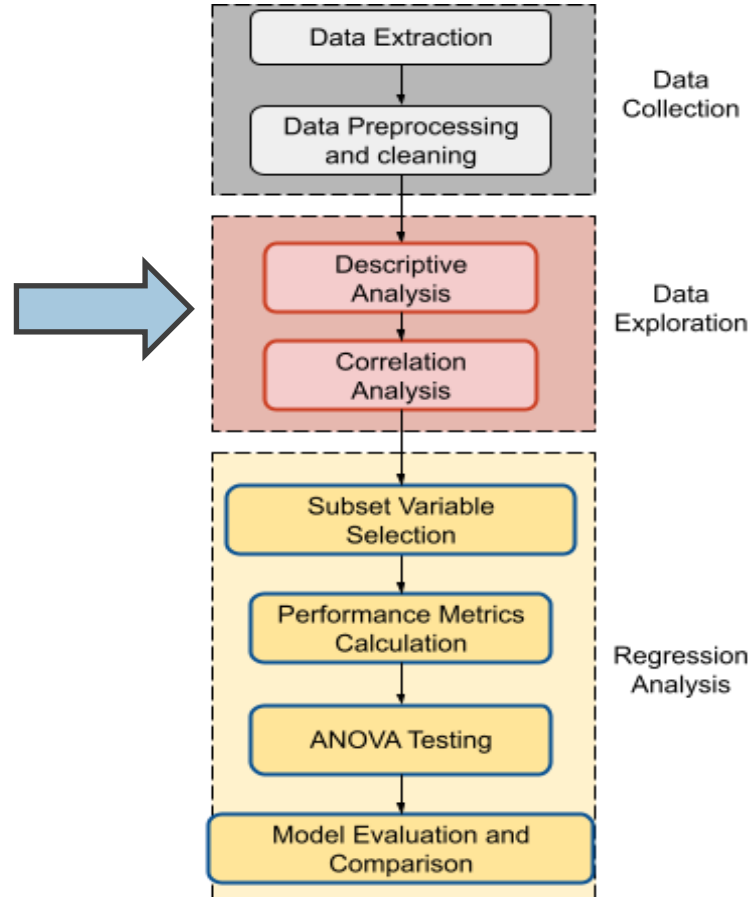


## Variables of interest

Variable	Description	Extracted or Calculated
Final.Grade	Student final test grade in the course	Extracted from Canvas LMS
Accuracy	Ratio of correct answers to the total number of questions	Extracted from Quizizz data
Total.Time.Taken	Average time taken by student to complete a quiz	Extracted from Quizizz data
Attempted	Average number of attempted questions in all quizzes	Extracted from Quizizz data
Unattempted	Average number of unattempted questions in all quizzes	Extracted from Quizizz data
Unattempted.Ratio	Ratio of answered to unanswered questions	Calculated based on Quizizz data
Score	Average points collected by a student based on speed and correctness	Extracted from Quizizz data
Timed.Accuracy	Ratio of accuracy to average time taken to complete a quiz	Calculated based on Quizizz data
Incorrect.Ratio	Ratio of incorrect answers to correct ones	Calculated based on Quizizz data

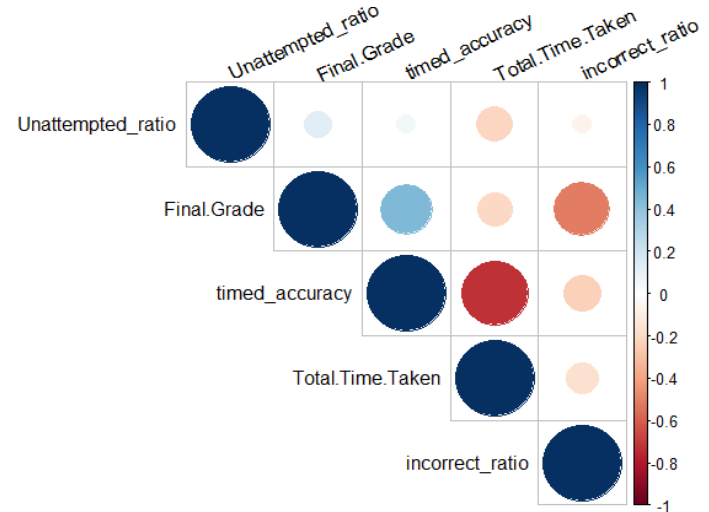
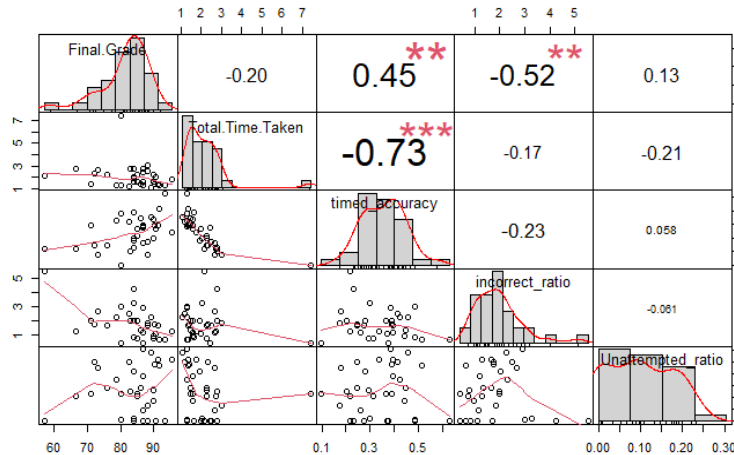


# Methodology



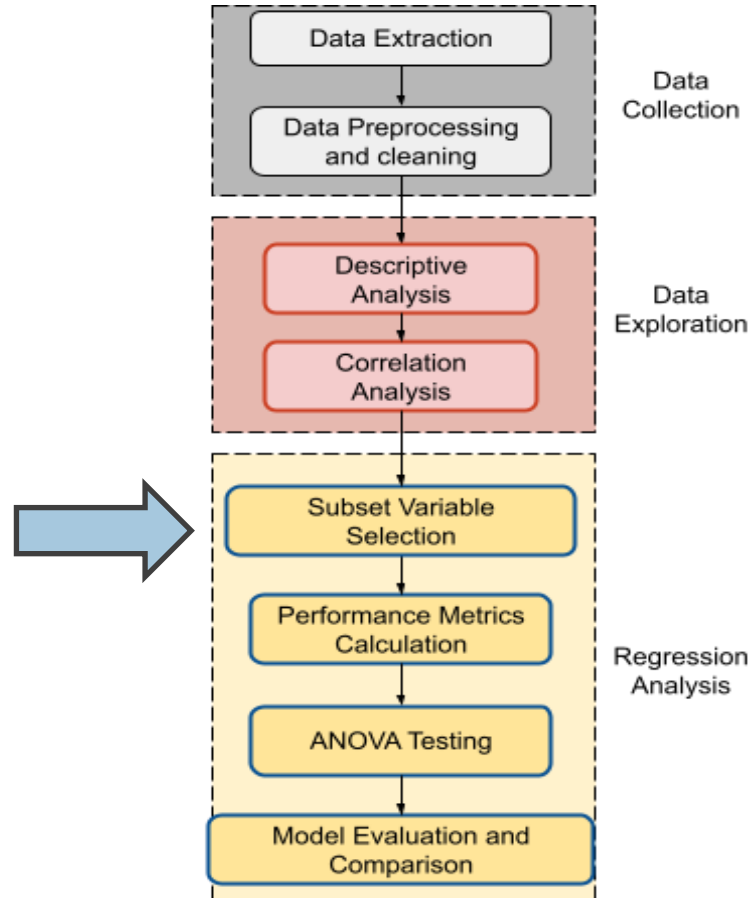
# Results and Findings

The “**Course Final.Grade**” and “**Timed.Accuracy**” variables are **positively correlated**, while the “**Course Final.Grade**” and “**Incorrect.Ratio**” are **negatively correlated**. A better view of variable correlation is achieved using the correlogram below.





# Methodology





## Results and Findings

R<sup>2</sup>: measures how well the data fit the regression line (higher values is better).

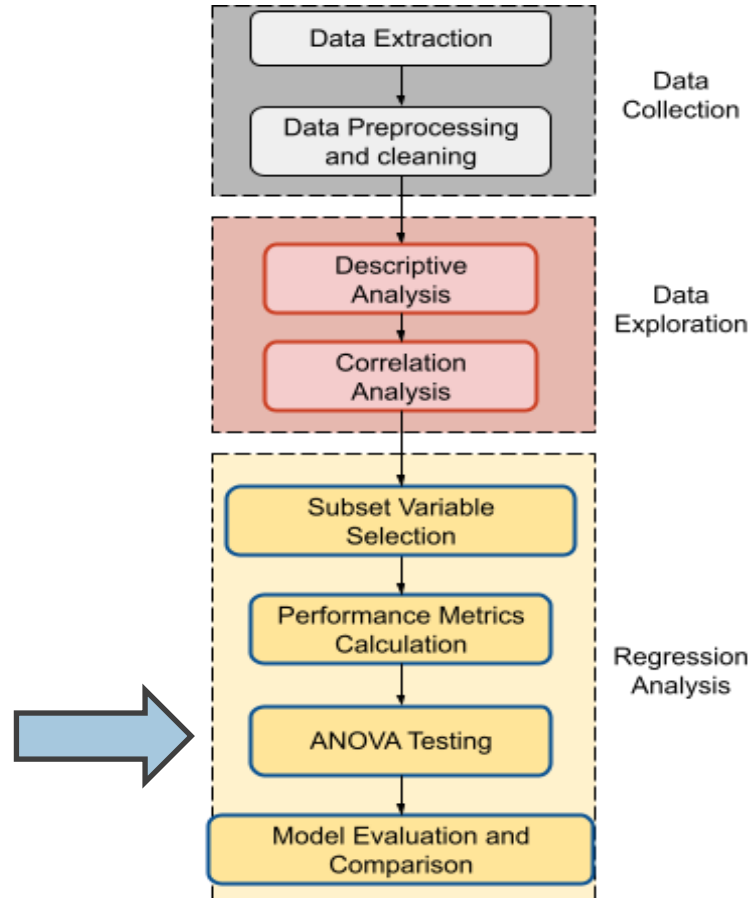
BIC: is the Bayesian Information criterion for model selection that penalizes models with more parameters or complexity (lower BIC value indicates a better model)

Cp: Mallows' Cp measures unexplained error (smaller Cp values are better).

Model	Variables	R <sup>2</sup>	BIC	Cp
1-variable	Incorrect.Ratio	0.24	-3.31	4.43
<b>2-variable</b>	<b>Incorrect.Ratio+ Timed.Accuracy</b>	<b>0.33</b>	<b>-5.32</b>	<b>1.30</b>
3-variable	Incorrect.Ratio+ Timed.Accuracy+ Unattempted.Ratio	0.32	-2.16	3.01
4-variable	Incorrect.Ratio + Timed.Accuracy + Unattempted.Ratio + Total.Time.Taken	0.29	1.31	5.0



# Methodology





## Results and Findings

ANOVA Test	P-value	Interpretation	Decision
1-variable vs. 2-variable models	0.0265 ( < 0.05).	Adding the “Timed.Accuracy” led to a significantly improved fit and better model.	(reject null hypothesis)
2-variable vs. 3-variable models	0.5878 ( > 0.05).	Adding the “Unattempted.Ratio” variable to the model led to no significant improvement fit	(accept null hypothesis)
2-variable vs. 4-variable models	0.8606 ( > 0.05)	Adding the “Unattempted.Ratio” and “Total.Time.Taken” variables to the 2-variable model led to no significant	(accept null hypothesis).





## Conclusion

**Limited research on how gamification can be used to track and predict academic performance.**

This study examined how **gamified classroom activities can be used to track and predict students' academic performance** represented by their final course grades.

We used Quizizz and focused on **two groups of sophomore computer science students.**

The study **explored the relationship between students' final course grades and their scores, interactions, and timings** during the weekly gamified activities through data-driven correlational analysis.

Results indicated that the **2-variable (“Incorrect.Ratio” and “Timed.Accuracy”) regression** model has the best fit for the data, the lowest prediction error, and the lowest complexity compared to the other models.



## Future Work

Increase the dataset sample size to achieve more robust results.

Share the data online.

Include other academic performance metrics and examine other gamification tools and platforms.

Integrate as software service within learning management systems.



Questions Please...