THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Exploring the Potential of VR in Education

Clustering Analysis of VR Learning Patterns and Student Profiles

Team members: Shabnam Rafat Ummey Sawda, Mohammed Ismail Sarfaraz Shaik, Jahnavi Vallepalli, Uyemaa Gantulga

Introduction

Purpose of the Project: *Understand the impact of VR in education and explore key relationships in the dataset.*

Data

- Source: Kaggle
- Data pre-processing

EDA

- Underlying Distribution
- Relationship

Modeling

- Cluster Creation
- Evaluation

SMART Qs

- Key features
- Cluster characteristics
- VR Usage Impact Analysis

Total Number of Observations:

5000

Missing Values: 00

Total Variables: 20

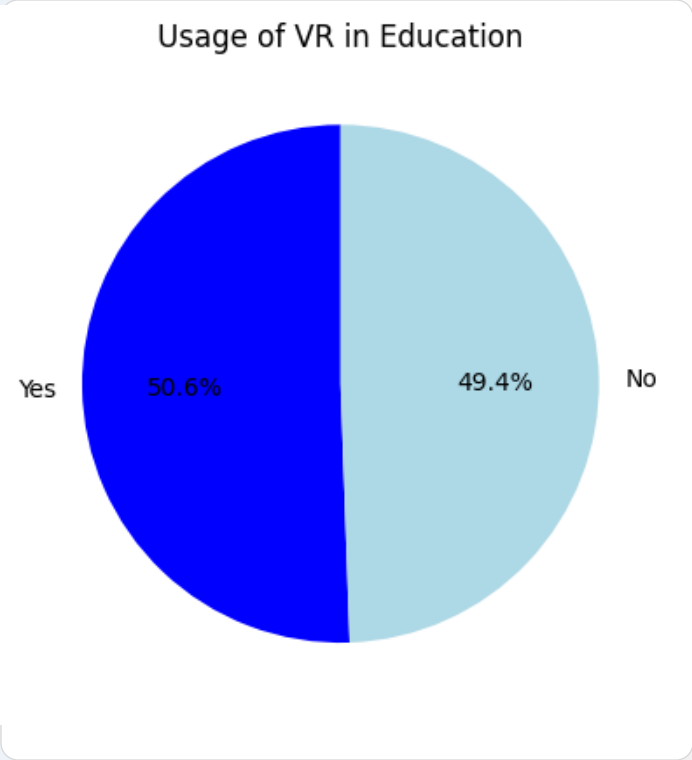
Some Key variables:

- Academic Outcome
- Engagement level
- Creativity Impact
- Stress level
- Duration of use
- Region
- School Support System
- Instructor Proficiency

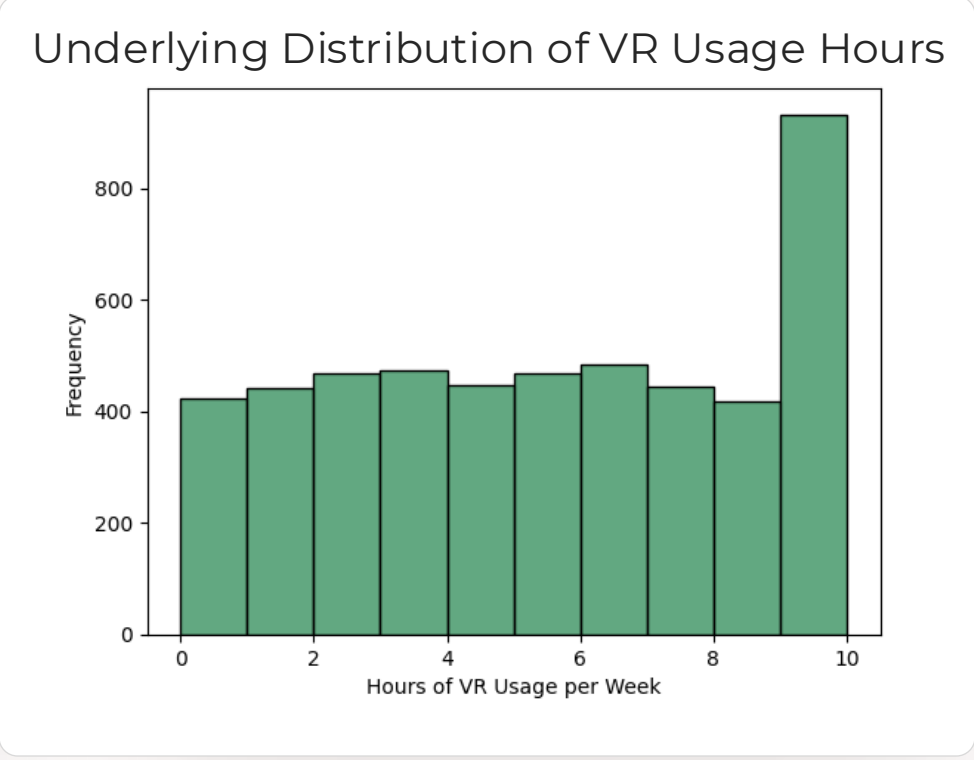
Variable types:

- 5 numerical
- 15 categorical

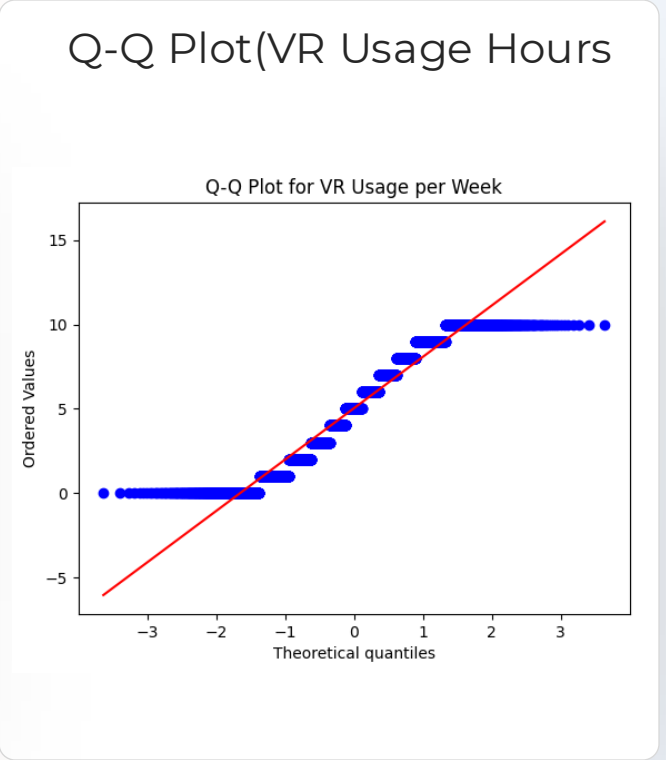
Exploratory Data Analysis



- **50%+** of the students are using VR



- The distribution is **not symmetric**,
- This suggests a potential skewness or that a large number of users have very high VR usage.

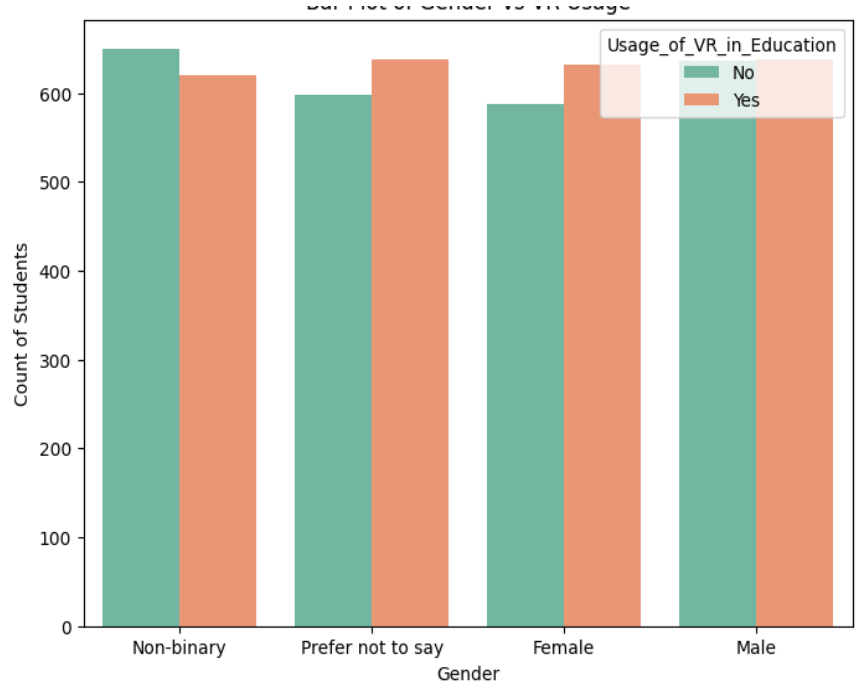


- Points significantly deviate from the red diagonal line, especially in the tails, indicating that the data **does not follow a normal distribution**.

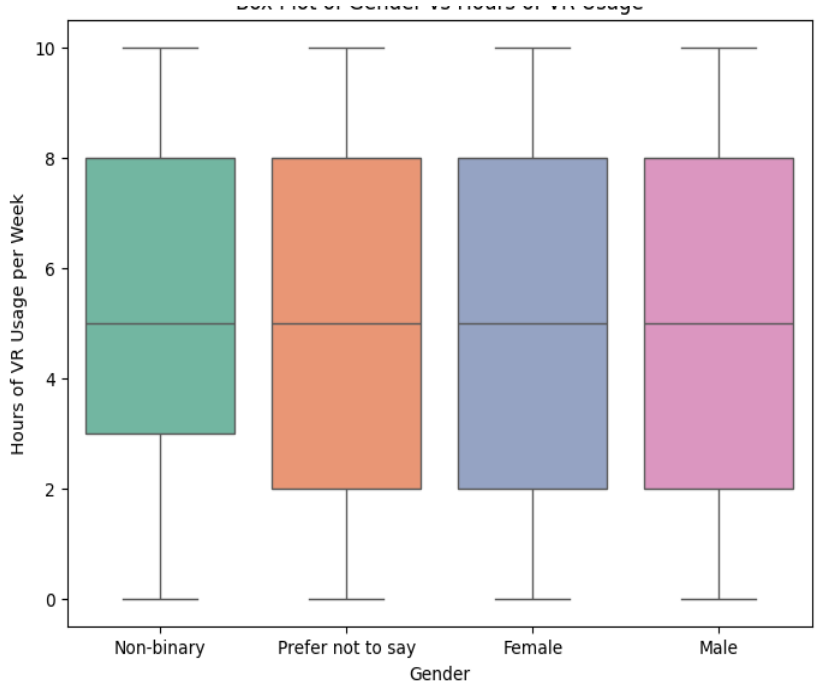
VR Usage Patterns by Gender

- **Engagement Levels**: Comparable for males and females
- **VR Usage Hours**: Slightly higher average for males
- **Access to Equipment**: Males report greater access
- **Stress Levels**: Females report higher stress
- **Learning Outcomes**: Minimal gender differences in perceived improvements.
- **Collaboration**: Higher among females

Gender-wise VR Usage

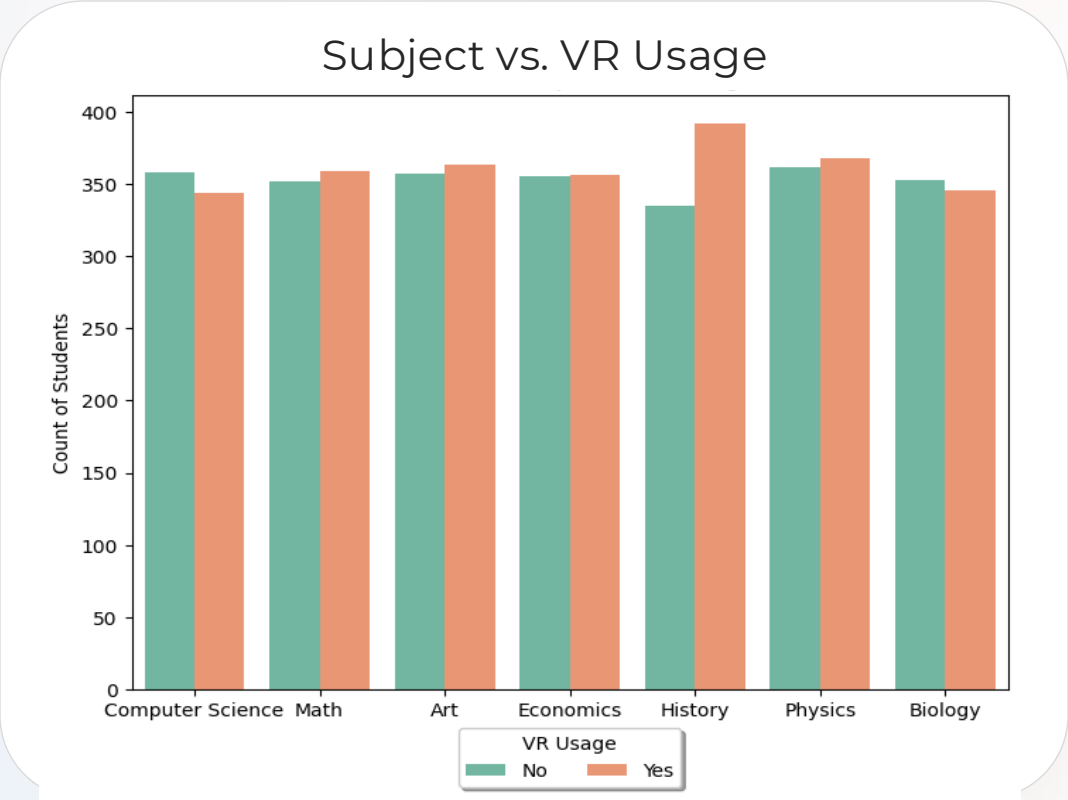


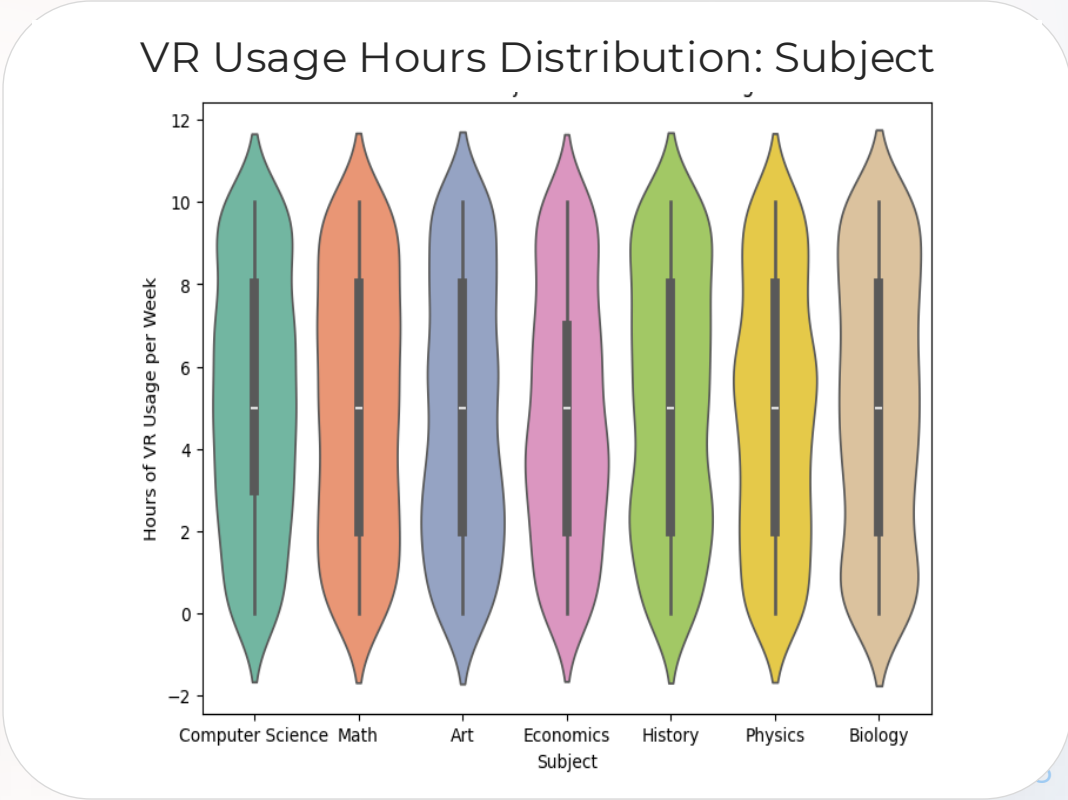
Gender-wise VR Usage Hours Distribution



VR Usage Patterns by Subject

- **Engagement**: High in STEM; lower in Arts and Humanities.
- **VR Usage Hours**: Most hours in Science and Technology; sporadic in Arts.
- **Stress Levels**: Lower in STEM; higher in Arts and Social Sciences.
- **Learning Outcomes**: Strongest improvement in Math and Science.
- **Collaboration**: High in group-oriented subjects like Engineering.

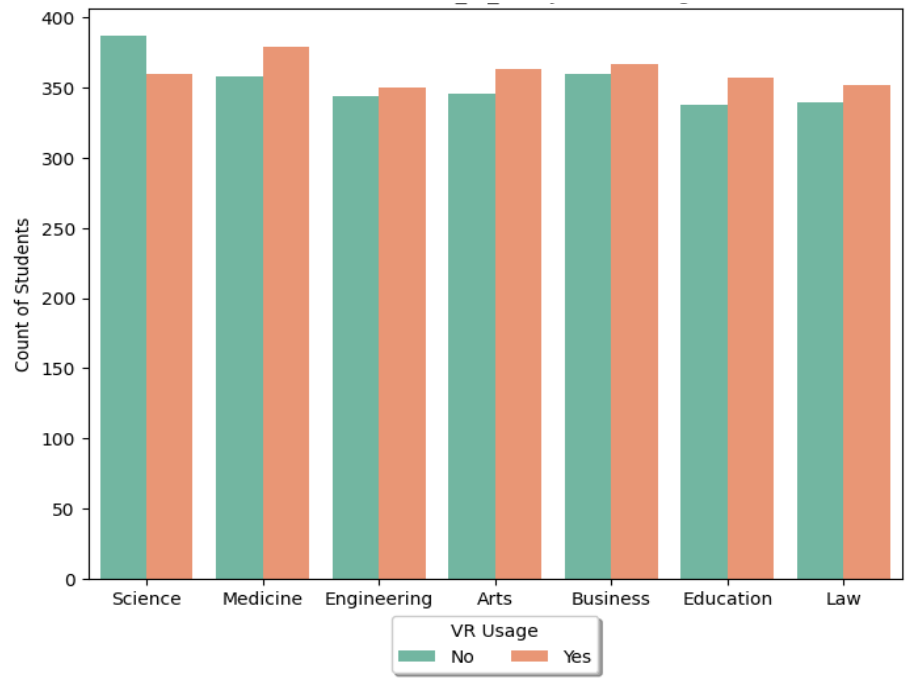




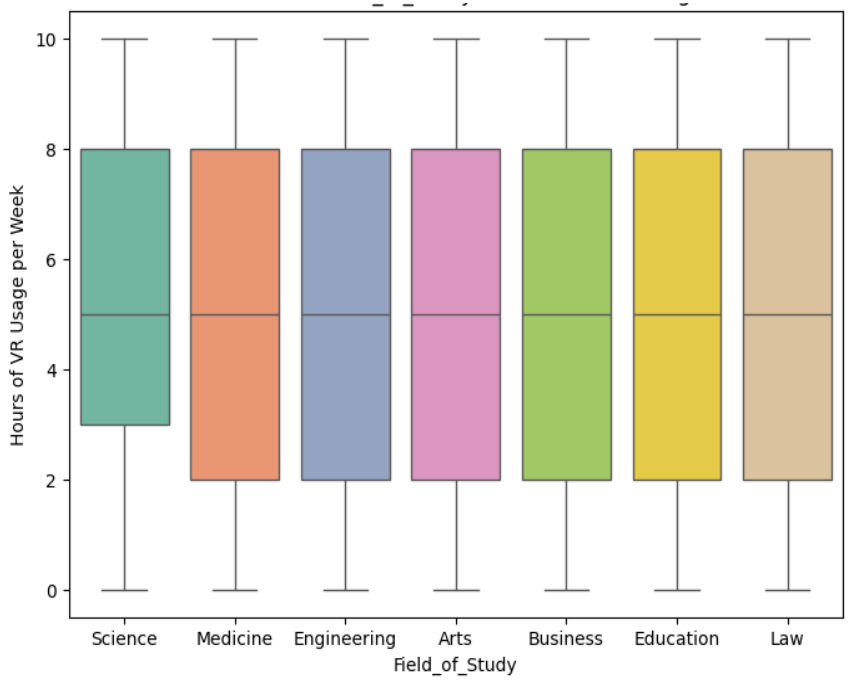
VR Usage Patterns by Field of Study

- **STEM**: High VR engagement and weekly usage; low stress levels.
- **Arts**: Moderate engagement; mixed learning outcomes; higher stress.
- **Business**: High collaboration via VR; steady engagement.
- **Social Sciences**: Lower VR usage; moderate learning improvements.
- **Medicine**: Strong focus on VR for practical learning; high effectiveness reported.

Field of Study vs. VR Usage

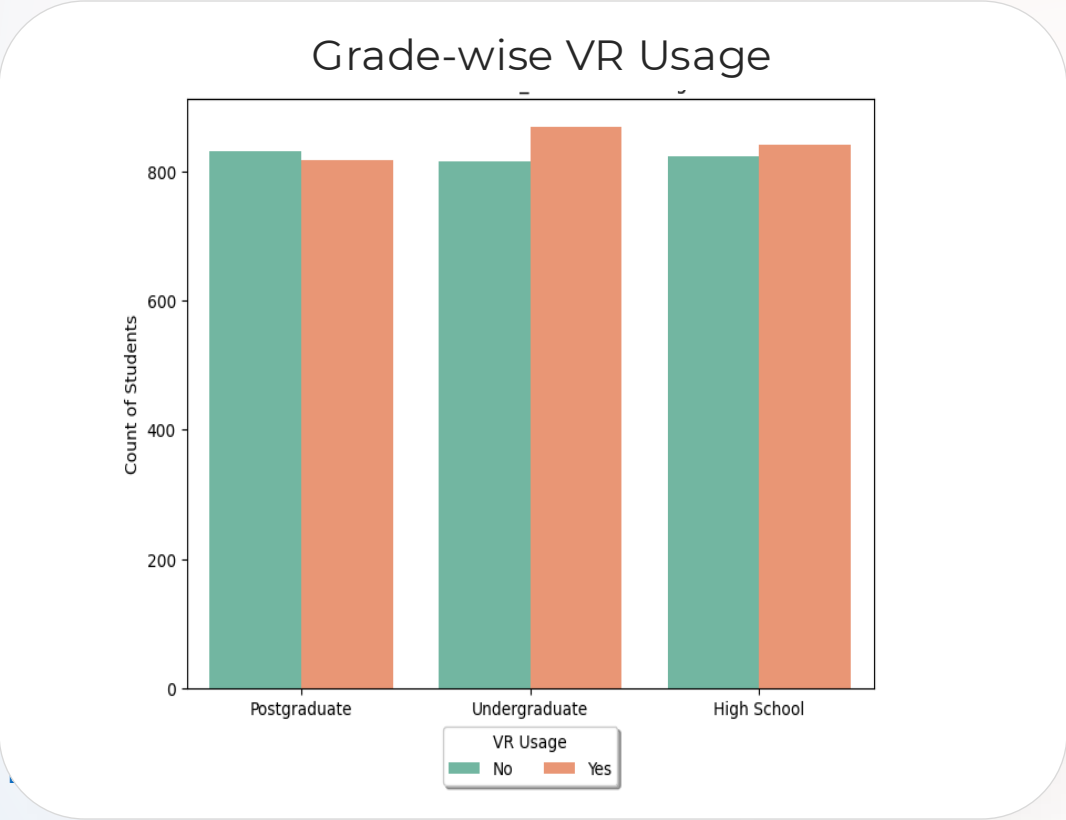


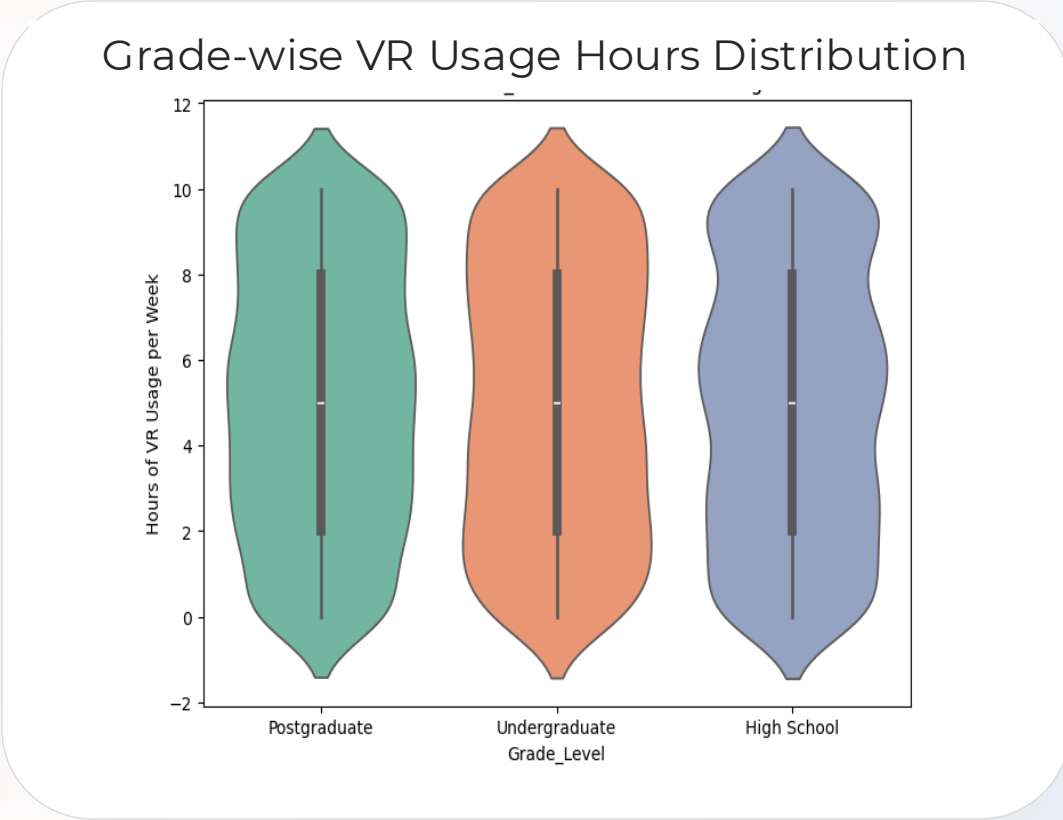
VR Usage Hours Distribution: Field of Study



VR Usage Patterns by Grade level

- **High School**: Moderate VR usage; higher stress levels.
- **Undergraduate**: High engagement and weekly usage; strong collaboration.
- **Postgraduate**: Focused VR usage for advanced learning; highest effectiveness reported.





Relationship between VR Usage and Gender

Is VR usage associated with gender?

Chi-Square Test Result

p-value: 0.39

There is no significant association between VR usage and gender

Is there a significant difference in means in VR usage hours across genders?

ANOVA Test Result

p-value: 0.23

There is no significant difference in VR usage hours across genders.

Relationship between VR Usage and Subject

Is VR usage associated with subject?

Chi-Square Test Result

p-value: 0.63

There is no significant association between VR usage and subject.

Is there a significant difference in means in VR usage hours across subjects?

ANOVA Test Result

p-value: 0.395

There is no significant difference in VR usage hours across subjects.

Relationship between VR Usage and Academic Outcome

Is academic outcome associated with VR Usage?

Chi-Square Test Result

p-value: 0.84

There is no significant association between academic outcome and VR usage

Do VR usage hours significantly predict academic outcomes?

Logistic regression Result

p-value: 0.856

The number of hours of VR usage per week does not significantly predict academic outcomes in this dataset.

Relationship between VR Usage and Engagement Level

Is engagement level associated with VR Usage?

Chi-Square Test Result

p-value: 0.75

There is no significant association between engagement level and VR usage.

Does VR usage hours significantly influence engagement levels?

Linear Regression Result

p-value: 0.679

VR usage hours do not have a statistically significant influence on engagement levels

Relationship between Instructor VR Efficiency and VR Usage

Is instructor VR efficiency associated with perceived effectiveness of VR?

Chi-Square Test Result

p-value: 0.41

There is no significant association between instructor VR efficiency and perceived effectiveness of VR

Is instructor VR efficiency associated with students' interest in continuing of VR based learning?

Chi-Square Test Result

p-value: 0.54

There is no significant association between instructor VR efficiency and students' interest in continuing VR-based learning.

Research Questions

1. How does the instructor's VR proficiency affect students' improvement in learning outcomes?
2. What are the key distinguishing features between high-performing and low-performing clusters?
3. How do cluster characteristics vary across different regional and support system contexts?
4. How do engagement levels in VR correlate with academic outcomes within each identified cluster?

Q1: How does the instructor's VR proficiency affect students' improvement in learning outcomes?

Objective:

Perform one-way anova test on instructor's VR proficiency and students' improvement in learning outcomes to find the significance between them

1. Label the Improvements in learning outcomes variable
2. Group the data by instructor VR proficiency and extract the improvement in learning outcomes
3. Perform anova test and find out the p-value

ANOVA Test Result :

p-value: 0.972

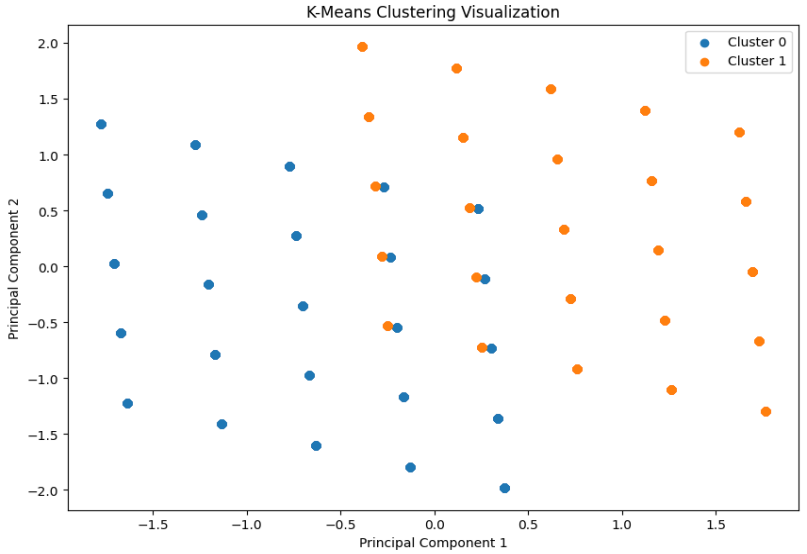
F-statistic: 0.280

There is no statistical significance between instructor's VR proficiency and students' improvement in learning outcomes

Q2: What are the key distinguishing features between high-performing and low-performing clusters?

Objective:

Verify if there are disparities in features between high-performing and low-performing clusters



Silhouette Score: 0.35

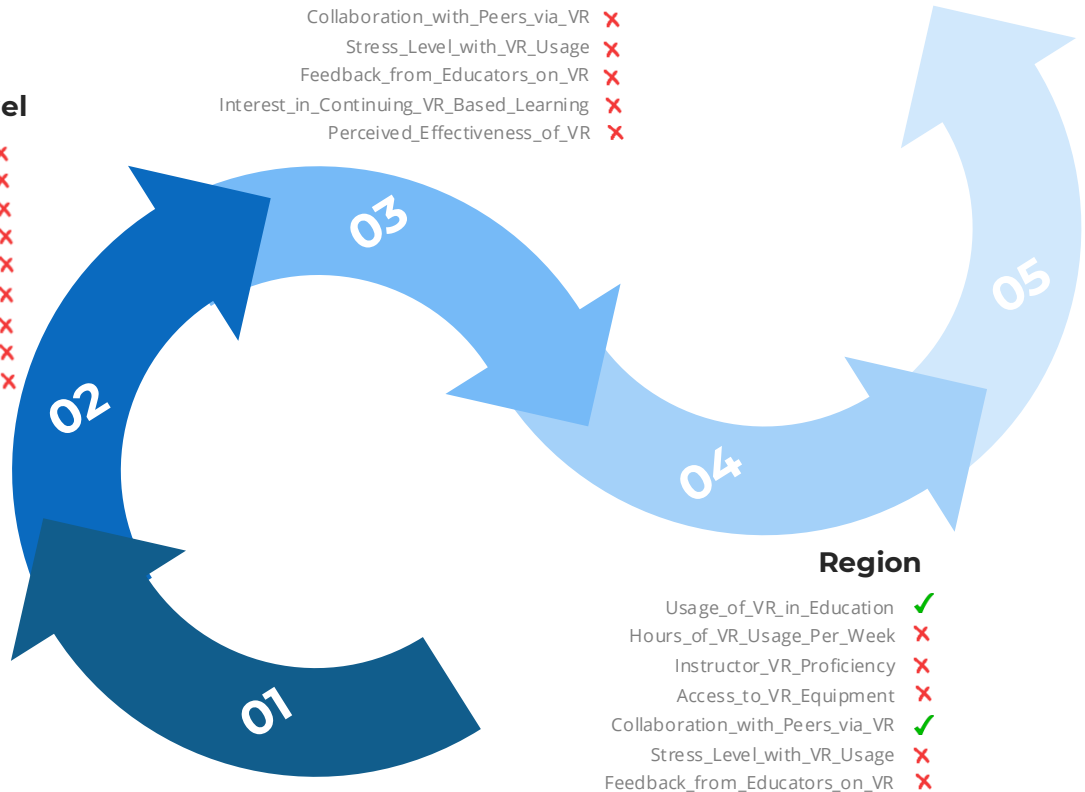
1. Define Performance Indicators:
   - Improvement_in_Learning_Outcomes
   - Impact_on_Creativity
2. Cluster creation using K-means Clustering

	p-value
Usage_of_VR_in_Education	0.82
Hours_of_VR_Usage_Per_Week	0.86
Instructor_VR_Proficiency	0.97
Access_to_VR_Equipment	0.57
Collaboration_with_Peers_via_VR	**0.01**
Stress_Level_with_VR_Usage	0.62
Feedback_from_Educators_on_VR	0.51
Interest_in_Continuing_VR_Based_Learning	0.99
Perceived_Effectiveness_of_VR	0.26

- **"Collaboration with Peers via VR" is the only feature significantly distinguishing high-performing and low-performing clusters,**
- **Other features show no significant differences.**

VR Usage and Academic Performance

Does it vary across Genders, Grade Levels, Field of Study, Regions, School Support System?



Gender

- Usage_of_VR_in_Education ✓
- Hours_of_VR_Usage_Per_Week ✗
- Instructor_VR_Proficiency ✗
- Access_to_VR_Equipment ✗
- Collaboration_with_Peers_via_VR ✗
- Stress_Level_with_VR_Usage ✗
- Feedback_from_Educators_on_VR ✗
- Interest_in_Continuing_VR_Based_Learning ✗
- Perceived_Effectiveness_of_VR ✗

Grade Level

- Usage_of_VR_in_Education ✗
- Hours_of_VR_Usage_Per_Week ✗
- Instructor_VR_Proficiency ✗
- Access_to_VR_Equipment ✗
- Collaboration_with_Peers_via_VR ✗
- Stress_Level_with_VR_Usage ✗
- Feedback_from_Educators_on_VR ✗
- Interest_in_Continuing_VR_Based_Learning ✗
- Perceived_Effectiveness_of_VR ✗

Field of Study

- Usage_of_VR_in_Education ✓
- Hours_of_VR_Usage_Per_Week ✗
- Instructor_VR_Proficiency ✗
- Access_to_VR_Equipment ✗
- Collaboration_with_Peers_via_VR ✗
- Stress_Level_with_VR_Usage ✗
- Feedback_from_Educators_on_VR ✗
- Interest_in_Continuing_VR_Based_Learning ✗
- Perceived_Effectiveness_of_VR ✗

Region

- Usage_of_VR_in_Education ✓
- Hours_of_VR_Usage_Per_Week ✗
- Instructor_VR_Proficiency ✗
- Access_to_VR_Equipment ✗
- Collaboration_with_Peers_via_VR ✓
- Stress_Level_with_VR_Usage ✗
- Feedback_from_Educators_on_VR ✗
- Interest_in_Continuing_VR_Based_Learning ✗
- Perceived_Effectiveness_of_VR ✗

School Support for VR Education

- Usage_of_VR_in_Education ✓
- Hours_of_VR_Usage_Per_Week ✗
- Instructor_VR_Proficiency ✗
- Access_to_VR_Equipment ✗
- Collaboration_with_Peers_via_VR ✗
- Stress_Level_with_VR_Usage ✗
- Feedback_from_Educators_on_VR ✗
- Interest_in_Continuing_VR_Based_Learning ✗
- Perceived_Effectiveness_of_VR ✗

Q3: How do cluster characteristics vary across different regional and support system contexts?

Objective:

- *Analyze the impact of regional contexts on cluster features.*
- *Assess the influence of support system contexts on cluster characteristics.*

6 Regional Clusters

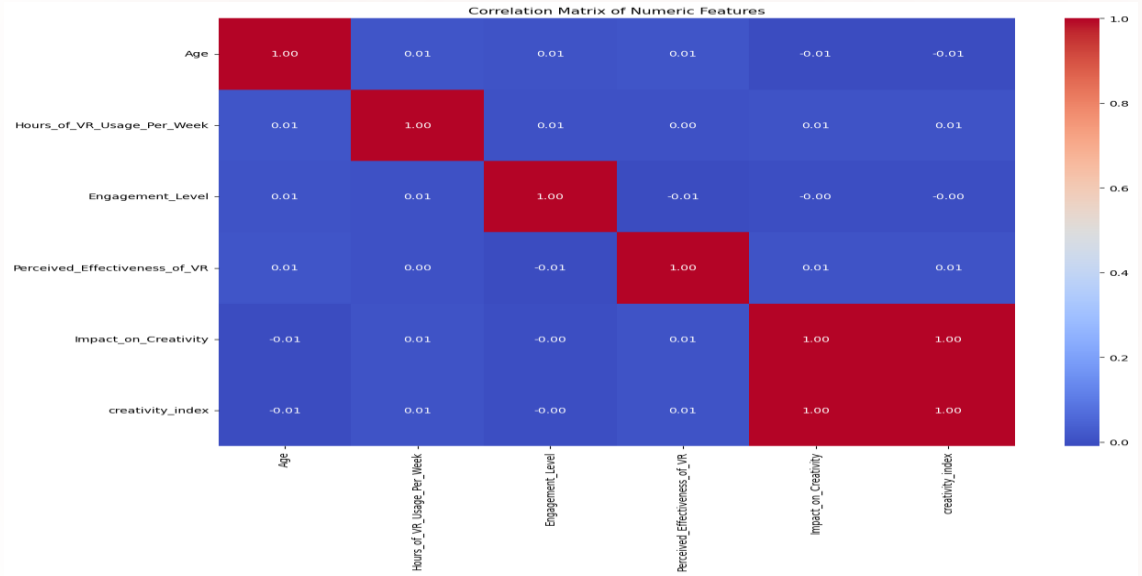
	p-value
Usage_of_VR_in_Education	0.81
Hours_of_VR_Usage_Per_Week	0.31
Instructor_VR_Proficiency	0.61
Access_to_VR_Equipment	0.77
Collaboration_with_Peers_via_VR	**0.01**
Stress_Level_with_VR_Usage	0.26
Feedback_from_Educators_on_VR	0.78
Interest_in_Continuing_VR_Based_Learning	0.72
Perceived_Effectiveness_of_VR	0.08

Support System-Based Clusters

	p-value
Usage_of_VR_in_Education	0.74
Hours_of_VR_Usage_Per_Week	0.76
Instructor_VR_Proficiency	0.20
Access_to_VR_Equipment	0.50
Collaboration_with_Peers_via_VR	0.29
Stress_Level_with_VR_Usage	0.78
Feedback_from_Educators_on_VR	0.27
Interest_in_Continuing_VR_Based_Learning	0.82
Perceived_Effectiveness_of_VR	0.68

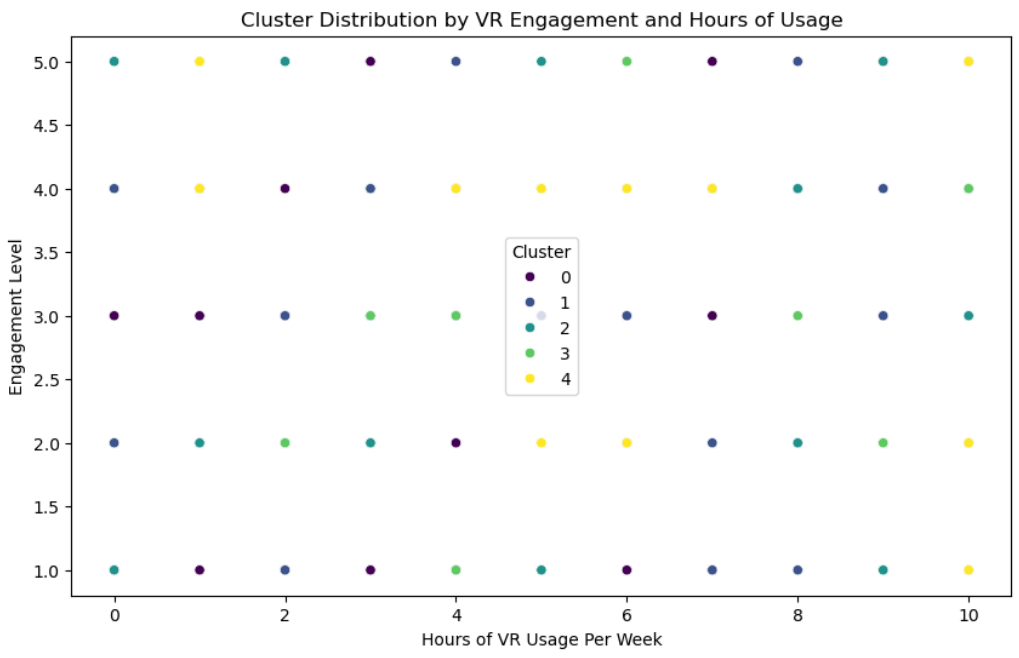
- "Collaboration with Peers via VR" significantly varies across regions. Other variables show no significant regional variation.
- In the Support System-Based Clusters, none of the variables exhibit significant variation, suggesting that support systems may not strongly differentiate the characteristics of VR usage in education.

Q4: How do engagement levels in VR correlate with academic outcomes within each identified cluster?



Correlation Matrix of Numeric Features

Visualizing the clusters based on VR engagement and usage hours



Prediction and evaluation per cluster using Linear Regression

- Cluster 2 - MSE: 0.26, R2: -0.06
- Cluster 3 - MSE: 0.25, R2: -0.01
- Cluster 0 - MSE: 0.25, R2: -0.01
- Cluster 1 - MSE: 0.25, R2: -0.02
- Cluster 4 - MSE: 0.25, R2: 0.01

Cluster 0 - Correlation between Engagement Level and Academic Outcomes:
Engagement_Level = 1.000000
Improvement_in_Learning_Outcomes = -0.037576

Cluster 1 - Correlation between Engagement Level and Academic Outcomes:
Engagement_Level = 1.000000
Improvement_in_Learning_Outcomes = 0.021024

Cluster 2 - Correlation between Engagement Level and Academic Outcomes:
Engagement_Level = 1.000000
Improvement_in_Learning_Outcomes = 0.028056

Cluster 3 - Correlation between Engagement Level and Academic Outcomes:
Engagement_Level = 1.00000
Improvement_in_Learning_Outcomes = -0.01695

Cluster 4 - Correlation between Engagement Level and Academic Outcomes:
Engagement_Level = 1.000000
Improvement_in_Learning_Outcomes = -0.051333

Correlation results per cluster using Random Forest

CONCLUSION

- ✓ More than 50% of students are using VR in the educational system.
- ✓ No significant association between VR usage and gender, subject, field of study, or grade level.
- ✓ No statistical significance between the instructor's VR proficiency and students' improvement in learning outcomes.
- ✓ Most features show no significant differences in distinguishing factors between high and low-performing clusters.
- ✓ The variables mostly did not exhibit significant variation across regions or the school support system.