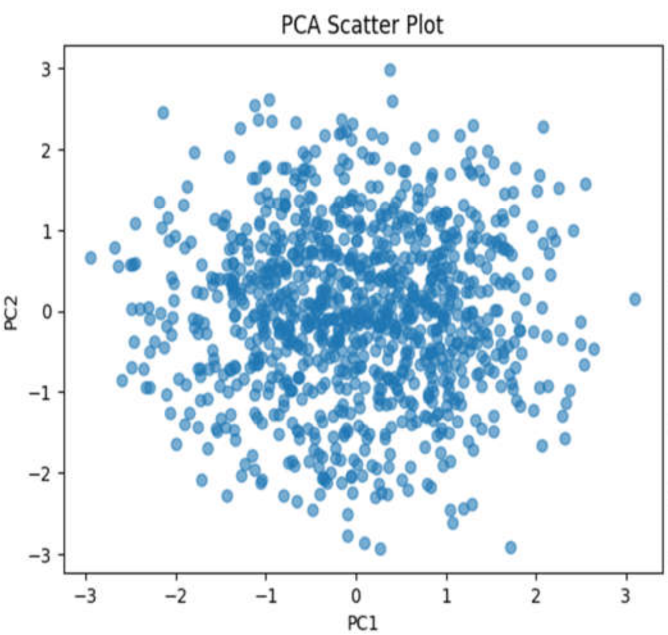
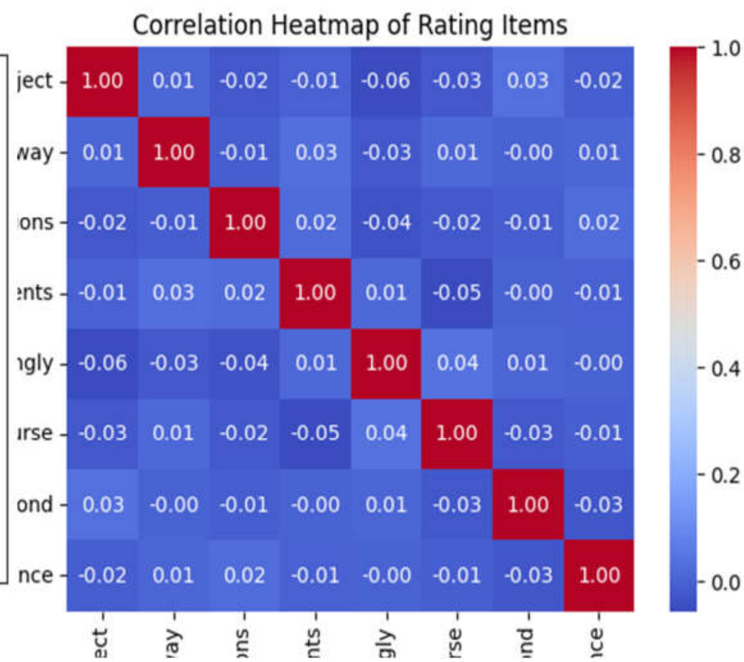
TASK 03 – STUDENT FEEDBACK ANALYSIS REPORT

COLLEGE EVENT & ACADEMIC FEEDBACK

Project Title: College Event Feedback Analysis Using Python, NLP & Data Visualization





Prepared By: Mitiku Dinsamo

Data Science & Analytics Intern

Submitted To: Future Interns – Data Science & Analytics Program

November, 2025

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

Understanding student perceptions is essential for improving the design, delivery, and overall quality of campus activities and academic programs. This project uses real survey data, processed using Python in Google Colab, to analyze how students evaluate different instructional aspects of an academic activity.

Through structured data cleaning, exploration, visualization, and modeling, this study converts raw feedback into meaningful insights. The goal is to support decision-makers in enhancing student satisfaction, refining event formats, and addressing areas that require improvement.

This report covers:

- Dataset inspection and cleaning
- Summary statistics and rating distribution analysis
- Correlation analysis to understand rating relationships
- Dimensionality reduction using PCA
- Student segmentation using K-Means clustering
- Key insights and targeted recommendations

2. Dataset Overview

The dataset (student_feedback_processed.csv) contains 1,001 student responses across eight rating-based feedback questions. These questions cover key dimensions of academic experience, such as:

- Subject expertise
- Clarity and effectiveness of explanations
- Quality of presentations
- Assignment difficulty
- Supportiveness and doubt resolution
- Course organization
- Relevance of course content

Each variable is rated on a scale of 1 to 10, where higher values reflect greater satisfaction.

Table 2-1. Dataset Structure Summary

Metric	Value
Total Records	1,001
Rating Variables	8
Score Range	1–10
Missing Values	None detected
Outliers	None requiring removal

3. Data Cleaning & Preparation

Data cleaning was conducted using Python (pandas) in Google Colab.

3.1. Preprocessing Steps

1. Removed redundant system-generated columns such as `Unnamed: 0`
2. Ensured all rating variables were properly converted to numeric formats
3. Confirmed that no missing or invalid entries existed
4. Standardized column names for readability
5. Created an **Overall Satisfaction Score** by averaging the eight rating variables

```
df['Overall_Score'] = df[rating_cols].mean(axis=1)
```

3.2. Overall Satisfaction Summary

- **Mean:** 5.92
- **Minimum:** 3.38
- **Maximum:** 8.25

This indicates a moderate satisfaction level, neither exceptionally high nor critically low.

4. Exploratory Data Analysis (EDA)

4.1. Distribution of Ratings

To understand how students rated each item, histograms were generated.

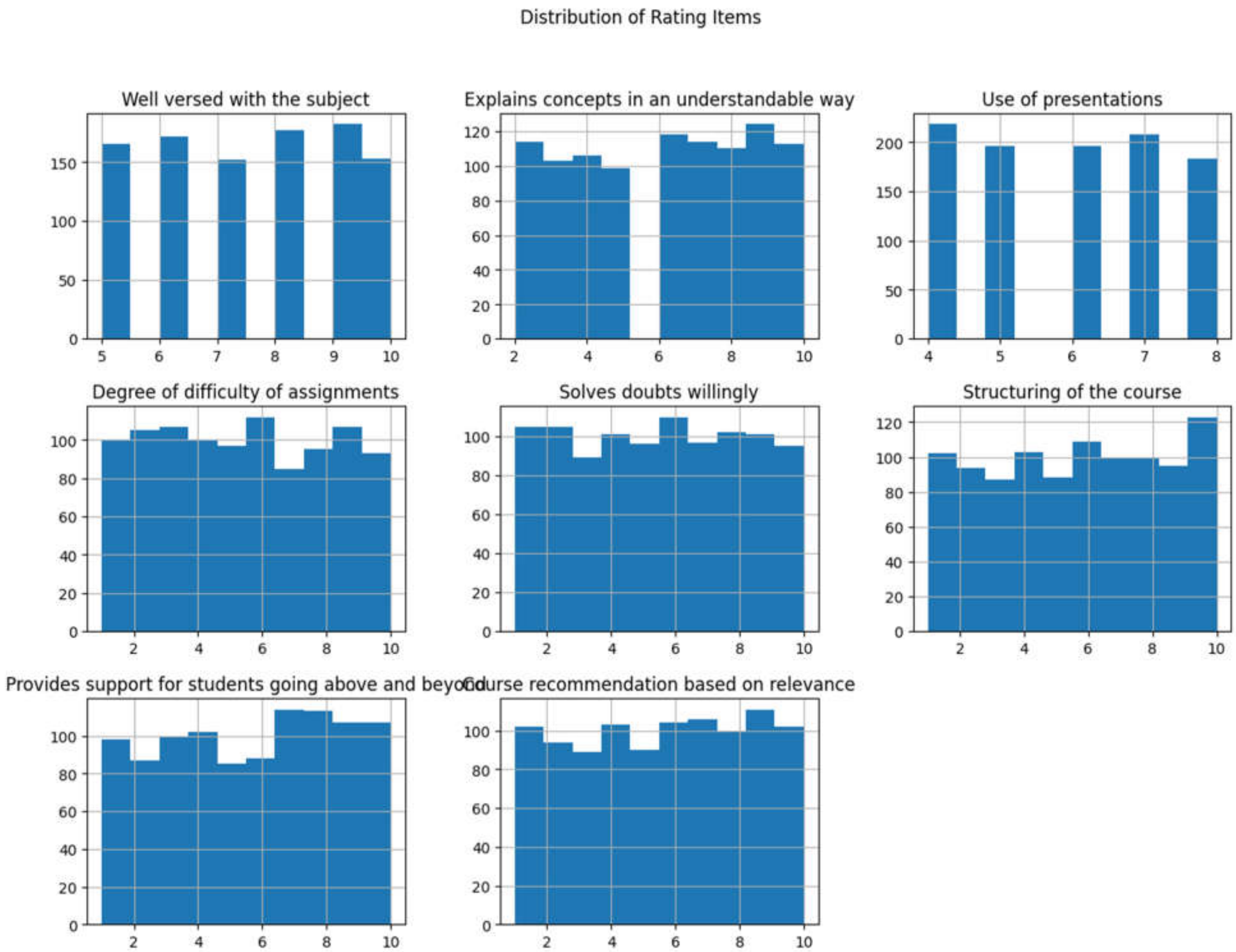


Figure 4-1. Rating Distribution for All Feedback Categories

Interpretation:

- Most rating scores fall in the **4-8 range**, representing moderate satisfaction.
- Extreme highs (9–10) and extreme lows (1–3) are rare, indicating balanced and thoughtful feedback patterns.
- The distribution supports the reliability of feedback, suggesting students provided reasoned evaluations rather than consistently high/low bias.

4.2. Descriptive Statistics

The mean values for each rating criterion provide insight into student perceptions.

Table 4-1. Mean Ratings Across Feedback Categories

Feedback Categories	**Mean Values**
Well versed with the subject	7.49
Explains concepts in an understandable way	6.08
Use of presentations	5.94
Provides support for students going above and beyond	5.66
Structuring of the course	5.64
Course recommendation based on relevance	5.59
Solves doubts willingly	5.47
Degree of difficulty of assignments	5.43

Pictorially, it can be illustrated in the following figure as follows:

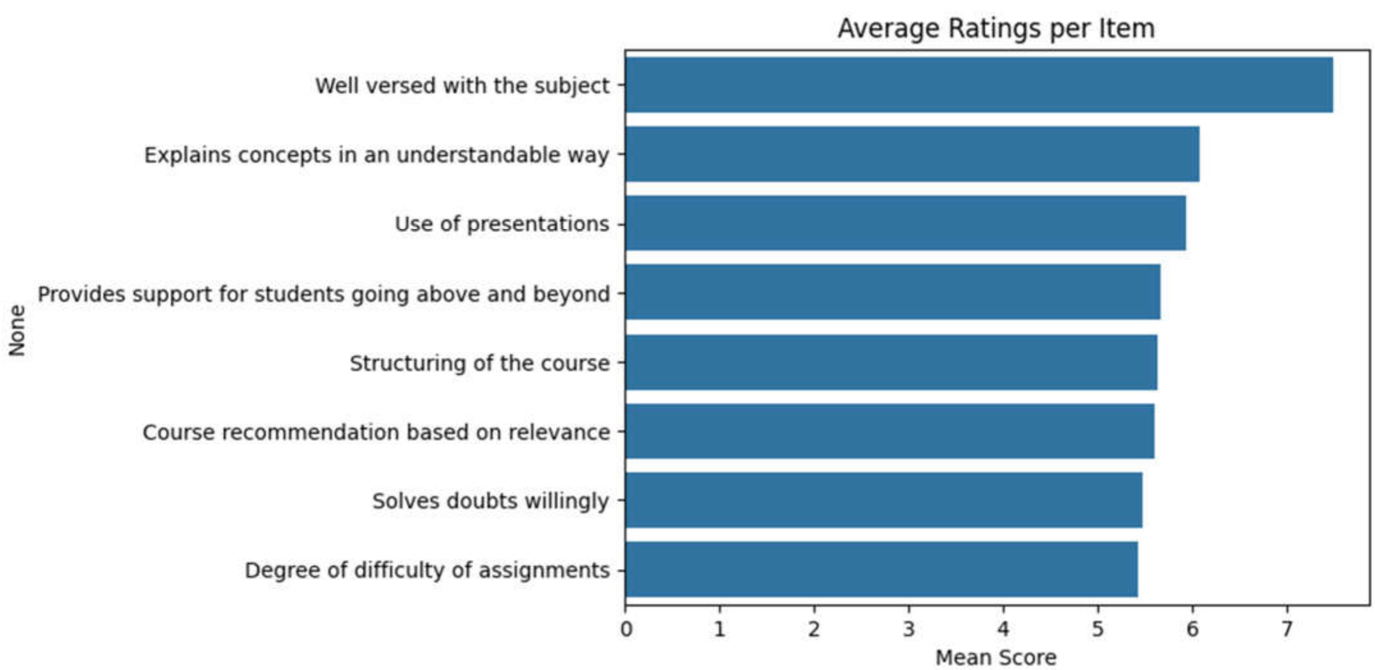


Figure 4-2. Bar Chart Showing Average Scores for Each Rating Topic

Key observations:

- **Highest score:** “Well versed with the subject” (Mean ≈ 7.49)

→ Students value instructors’ knowledge and recognize strong subject mastery.

- **Lowest scores:**
    - “Degree of difficulty of assignments” (Mean ≈ 5.43)
    - “Solves doubts willingly” (Mean ≈ 5.47)

    → These areas require immediate improvement.

4.3. Correlation Analysis

A correlation matrix was generated to measure how feedback dimensions relate to each other.

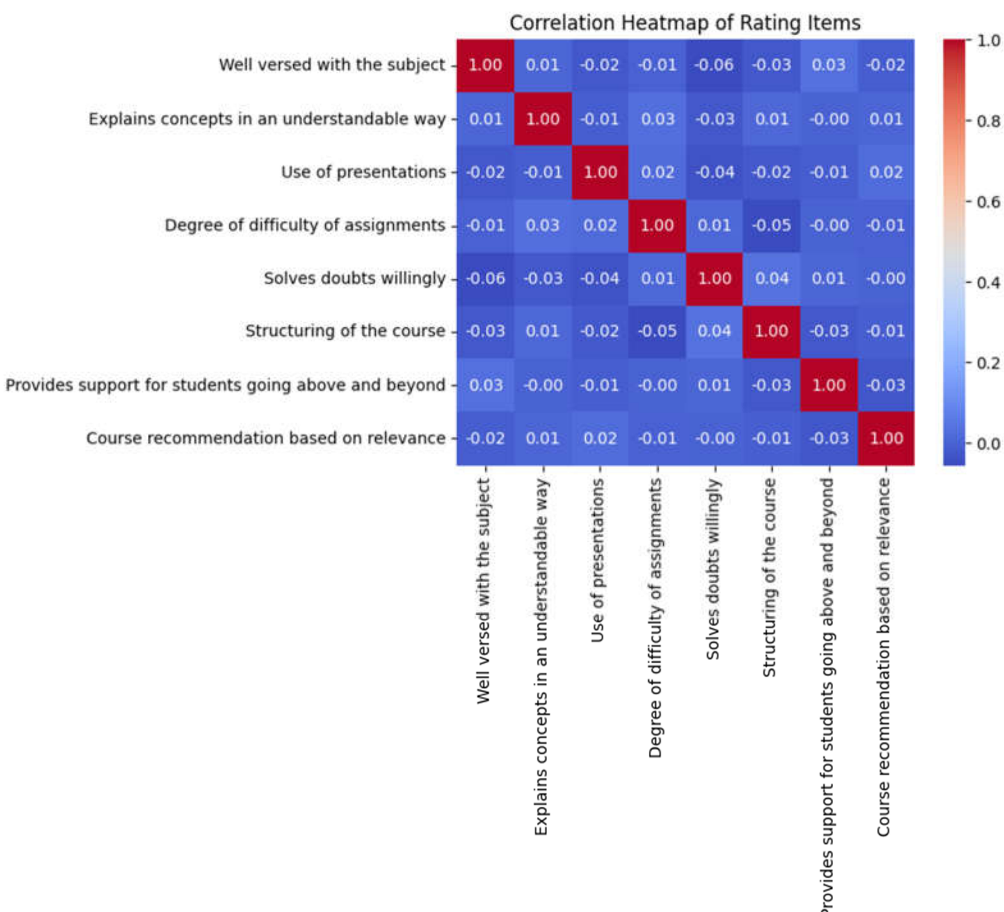


Figure 4-3. Rating Distribution for All Feedback Categories

(*Heatmap Visualization Depicting Correlation Strengths Between Variables*)

Key findings:

- Variables related to teaching clarity, presentation usage, and **course structuring** exhibit strong positive correlations, indicating they collectively influence perceived instructional quality.
- **Assignment difficulty** shows minimal correlation with other scores, meaning perceptions of difficulty are independent of students' views on teaching quality.
- Support-related attributes correlate moderately with clarity, suggesting students expect clearer explanations to be coupled with better doubt resolution.

5. Dimensionality Reduction Using PCA

Principal Component Analysis (PCA) was conducted to uncover underlying patterns and simplify visualization.

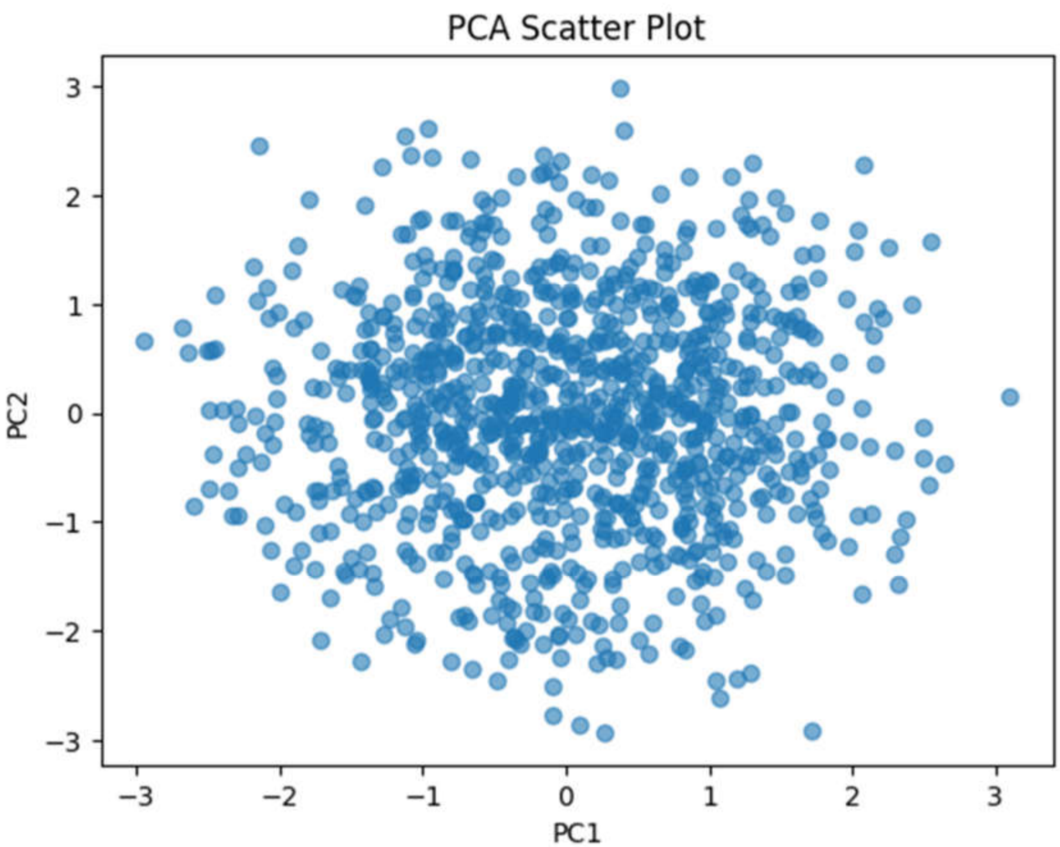


Figure 5-1. PCA Scatter Plot of Students Based on Feedback Patterns

(*2D scatter plot showing PCA Component 1 vs Component 2*)

Interpretation:

- Students naturally group into distinguishable segments based on their overall rating profiles.
- PCA shows that clarity-related features (explanation clarity, presentation use) and support-related attributes (doubt resolution, supportiveness) strongly influence variance.
- This confirms that feedback contains structure, making it suitable for clustering.

6. Student Segmentation Using K-Means Clustering

Clustering helps identify distinct groups of students with similar feedback tendencies. The Elbow Method indicated that k = 3 is a suitable number of clusters.

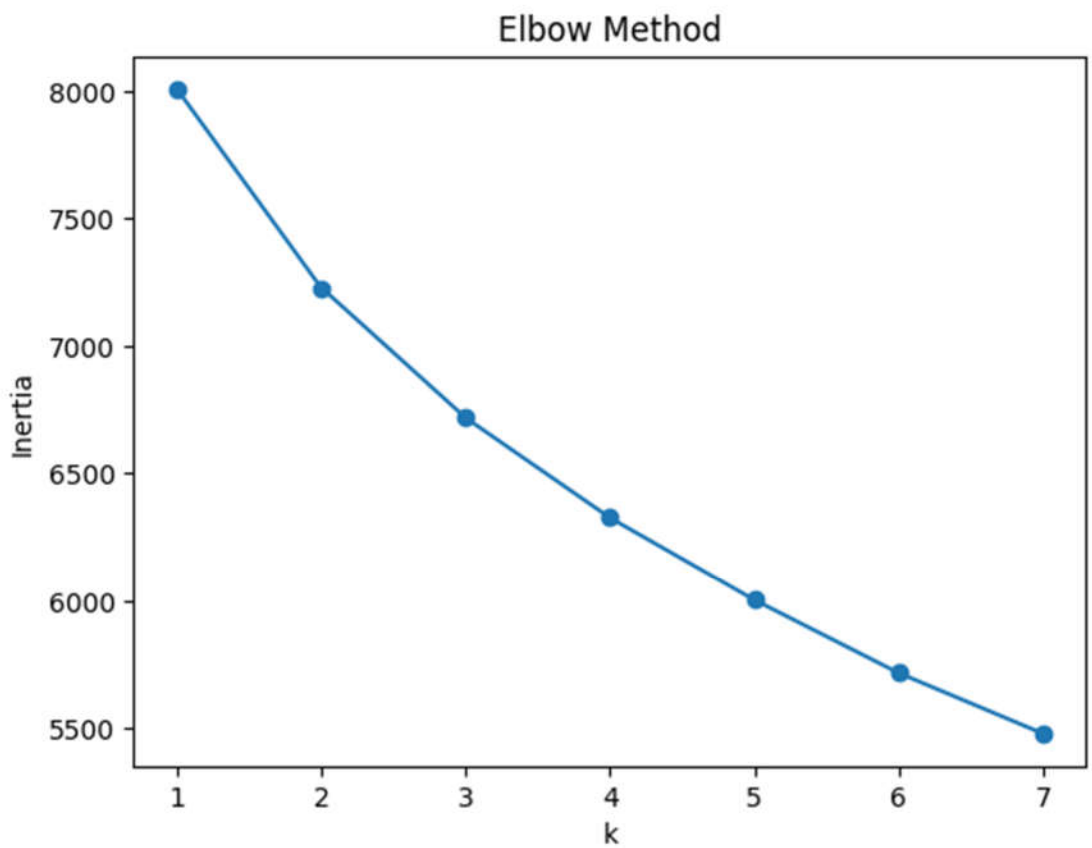


Figure 6-1. Student Clusters Identified through K-Means (k=3)

(*Scatter plot with clusters color-coded on PCA dimensions*)

Table 6-1. Cluster Profiles

Cluster	Description	Interpretation
Cluster 0	Appreciates clarity and presentation quality	Students satisfied with teaching approach but find assignments challenging
Cluster 1	High support rating but inconsistent overall scores	Students value doubt-solving but want more structured instruction
Cluster 2	Moderate scores but low support feedback	Students feel instructors are knowledgeable but not sufficiently supportive

Average Overall Scores

- **Cluster 0:** 6.38
- **Cluster 1:** 5.82
- **Cluster 2:** 5.54

Cluster 0 reflects the most positively engaged student group, while Cluster 2 displays the most concern, particularly about supportiveness.

7. Key Insights

⭐ Strengths Identified

1. Strong Instructor Knowledge – highest-rated dimension.
2. Clear and structured teaching methodology.
3. Effective presentation delivery, improving learning engagement.

⚠️ Weaknesses Identified

1. High assignment difficulty reduces satisfaction.
2. Inconsistent doubt resolution impacts student comfort and learning confidence.
3. Lower supportiveness scores for a significant student group (Cluster 2).

🔍 Patterns & Observations

- Satisfaction is driven primarily by clarity, structure, and subject expertise.
- Students who struggle with assignment difficulty tend to rate supportiveness poorly.
- Instructor interaction quality significantly shapes the student experience.

8. Recommendations for Improvement

A. Simplify and Structure Assignments

- Provide tiered difficulty options
- Share reference solutions
- Include pre-assignment walkthroughs

B. Strengthen Doubt Resolution Support

- Introduce dedicated doubt-clearing hours
- Encourage proactive follow-up
- Create online discussion forums or chat groups

C. Improve Student Support Experience

- Regular feedback loops
- Mid-course check-ins
- Peer support initiatives

D. Maintain and Enhance Teaching Strengths

- Continue using presentations effectively
- Preserve clear, structured explanations
- Encourage instructors to enhance subject enthusiasm

E. Provide Targeted Support for Cluster 2

- Offer personalized learning assistance
- Increase instructor-student engagement
- Mentor-based follow-up systems

9. Conclusion

This comprehensive analysis reveals valuable insights into students' academic experiences. While instructors excel in subject expertise and presentation quality, there are clear opportunities to improve assignment manageability and student support.

Using a combination of descriptive analysis, correlation assessment, PCA visualization, and cluster modeling, the feedback data uncovers actionable themes. These insights empower faculty coordinators and event organizers to refine course delivery and elevate student satisfaction in future academic activities.

The project demonstrates proficiency in:

- Data preparation & cleaning
- Statistical analysis
- Visualization techniques
- Machine learning (PCA & Clustering)
- Insight-driven reporting