-A KILUVA JAMES REPORT-

Student Satisfaction Survey Analysis Report;

1. General Summary

This report analyzes student satisfaction data from Kaggle datasets, aiming to identify key trends, strengths, and areas for improvement in the teaching-learning process and overall student experience. The analysis reveals generally positive satisfaction levels, with several questions and courses classified as "Excellent" or "Good." Key strengths include teacher communication and preparation, and fairness of internal evaluation. Areas for improvement, though generally in the "Good" category, suggest opportunities to enhance syllabus coverage and engagement in extracurricular activities. Recommendations focus on leveraging strengths, targeted improvements for specific areas, and considering sentiment analysis for future feedback.

2. Introduction

- **Purpose:** The primary objective of this report is to analyze feedback data from a student satisfaction survey to uncover satisfaction trends, assess the effectiveness of various aspects of the educational experience, and propose actionable improvements for future events and ongoing academic enhancements. This analysis aims to provide insights into student perceptions of teaching quality, course content, institutional support, and overall satisfaction.
- **Dataset:** The analysis is based on the Student_Satisfaction_Survey.csv dataset, which was obtained from Kaggle datasets. This dataset contains various metrics related to student feedback, including numerical ratings (Weightage 1-5), total feedback given, total configured questions, specific survey questions, course names, basic course categories, and an "Average/ Percentage" score.
- **Methodology:** The analytical approach involved several key steps: data loading and initial inspection, comprehensive data cleaning (including handling missing values, duplicates, and outliers), descriptive statistical analysis, and performance analysis grouped by course categories. Crucially, a question-level classification was performed to categorize satisfaction levels for individual survey items. The analysis was conducted using Python with libraries such as Pandas for data manipulation, and Matplotlib and Seaborn for data visualization. It is important to note that while the request included "Use NLP tools to score sentiment in comments," the provided script did not include this functionality; therefore, sentiment analysis of qualitative comments is not covered in this report.

3. Data Overview and Preprocessing

- **3.1 Data Loading:** The Student_Satisfaction_Survey.csv dataset was successfully loaded. The script initially attempted a default UTF-8 encoding and then successfully loaded the dataset using the latin1 encoding, due to the character encoding issues in the original file which were resolved during loading.
- **3.2 Initial Inspection:** The DataFrame consists of 580 entries (rows) and 13 columns after initial processing. The columns include:

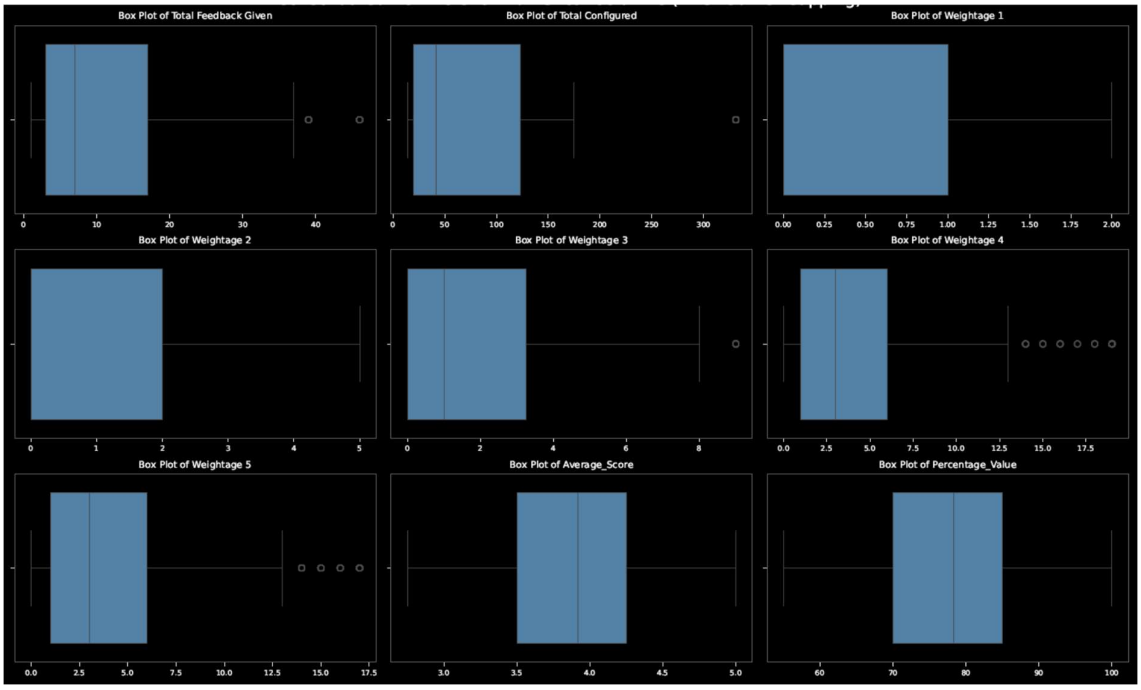
• **Column**	• **Description**
SN	Serial Number of the Survey Question
Total Feedback Given	Selected number of sample students for the feedback per Course
Total Configured	Total Strength of the batch of the Course
Questions	List of 20 Survey Questions in Total
Weightage 1	Lowest grade on scale
Weightage 2	better than lowest but below average rating
Weightage 3	Average

• Column	• Description
Weightage 4	Above Average
Weightage 5	Best rating
Average/ Percentage	Weighted Average shown in absolute as well as in percentage terms
Course Name	Current Year of the Graduation/Post Graduation
Basic Course	Graduation/Post Graduation Stream

- **3.3 Data Cleaning:**
    - **Handling 'Average/ Percentage':** The original 'Average/ Percentage' column was successfully split into two new numerical columns: Average_Score and Percentage_Value. Any missing values introduced during this conversion were imputed using the median of their respective columns.
    - **Missing Values:** No missing values were found across any of the columns after the initial data loading and processing steps.
    - **Duplicate Values:** No duplicate rows were identified in the dataset.
- **3.4 Outlier Management:** Outliers in numerical columns were identified using the Interquartile Range (IQR) method and subsequently handled by capping values at the 5th and 95th percentiles. This process ensured that extreme values, which might

skew the analysis, were adjusted while retaining the overall distribution of the data. Columns where outliers were detected and capped include:

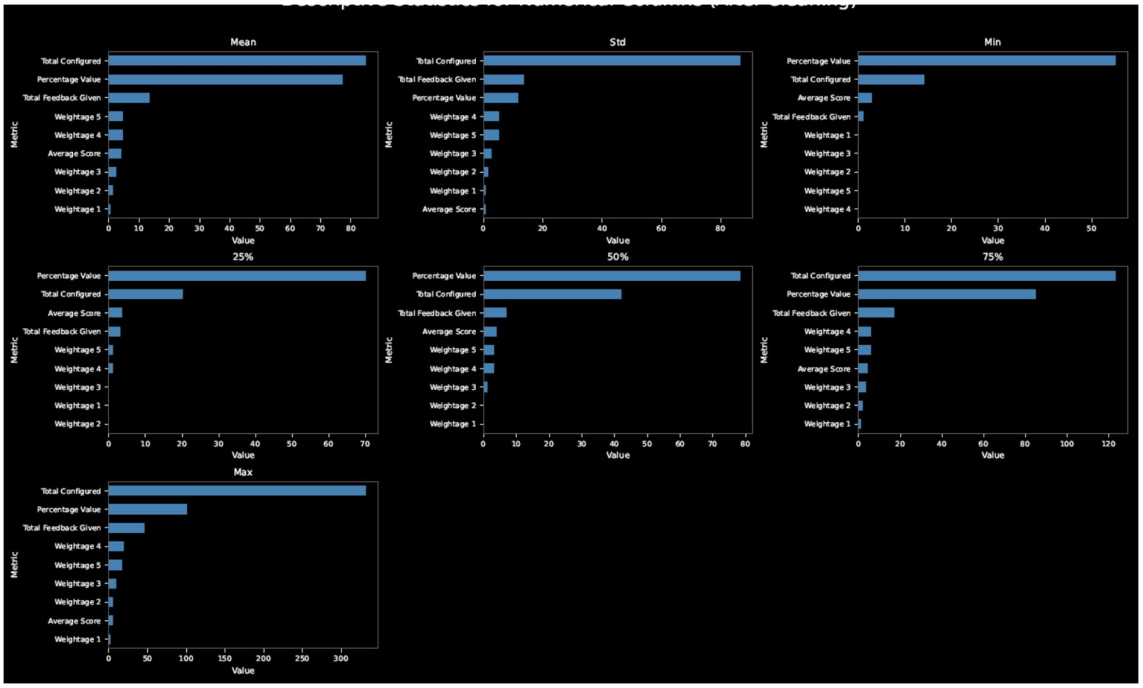
- 'Total Feedback Given' (capped between 1.00 and 46.00)
- 'Total Configured' (capped between 14.00 and 331.00)
- 'Weightage 1' (capped between 0.00 and 2.00)
- 'Weightage 2' (capped between 0.00 and 5.00)
- 'Weightage 3' (capped between 0.00 and 9.00)
- 'Weightage 4' (capped between 0.00 and 19.00)
- 'Weightage 5' (capped between 0.00 and 17.00)
- 'Average_Score' (capped between 2.75 and 5.00)
- 'Percentage_Value' (capped between 55.00 and 100.00)
- Box plots generated after this capping demonstrate a more constrained and less dispersed distribution, as seen in Student_Satisfaction_Analysis_Plots.pdf (Page 1, "Consolidated Box Plots for Numerical Columns (After Outlier Capping)").



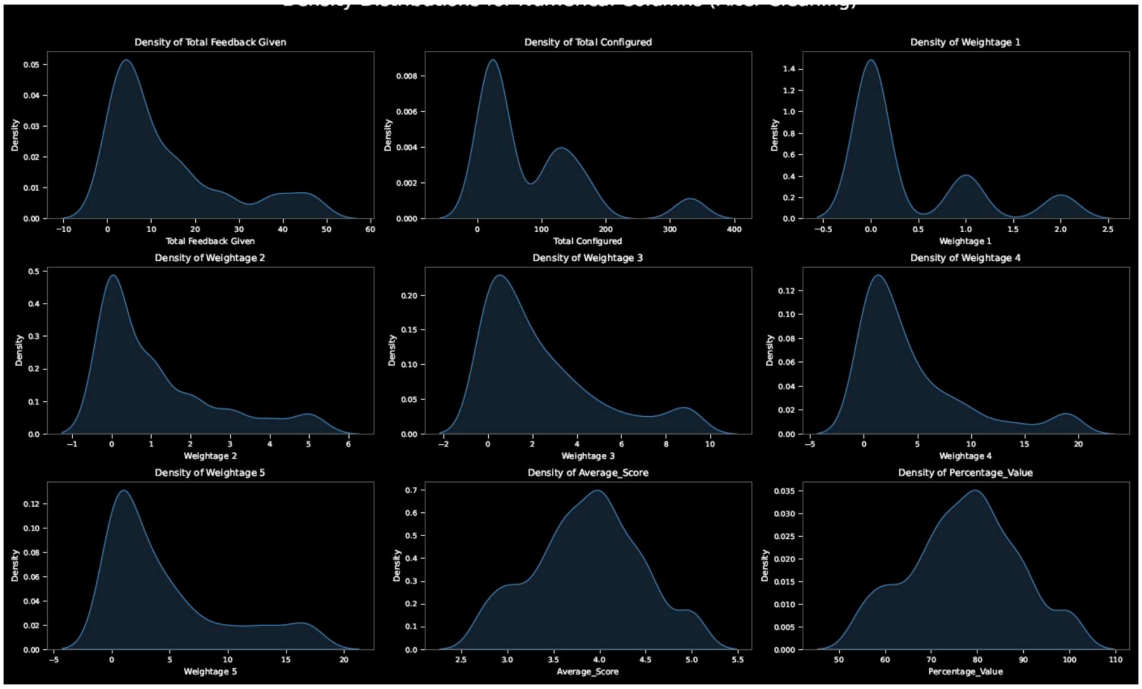
4. Descriptive Statistics and Distributions

- **4.1 Summary Statistics:** The descriptive statistics for numerical columns after outlier capping provide a clear overview of the central tendency and spread of the data. The Average_Score has a mean of approximately 3.86 (out of 5), with a standard

deviation of 0.58, indicating a relatively consistent level of satisfaction across the board. The Percentage_Value shows a mean of 77.27%, with a standard deviation of 11.54%.



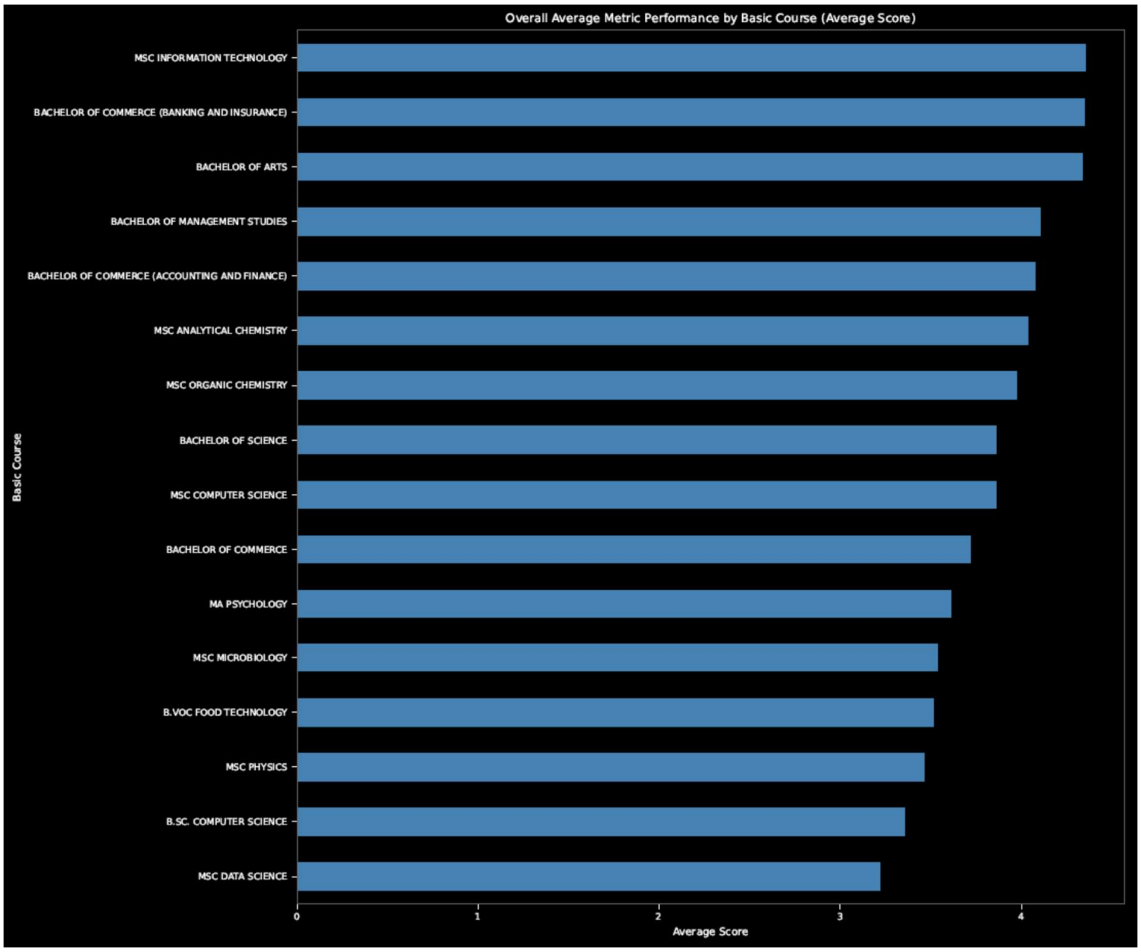
- **4.2 Data Distributions:** The density plots (Page 2, Student_Satisfaction_Analysis_Plots.pdf) for numerical columns after cleaning provide visual insights into their distributions.



- Average_Score and Percentage_Value distributions appear to be somewhat left-skewed, suggesting that a majority of scores lean towards the higher end of the scale, which is positive.
- The Weightage columns show varying distributions, with Weightage 1 and Weightage 2 heavily skewed towards zero, indicating fewer instances of very low ratings. Weightage 4 and Weightage 5 show broader distributions, suggesting a wider range of responses for higher satisfaction levels.

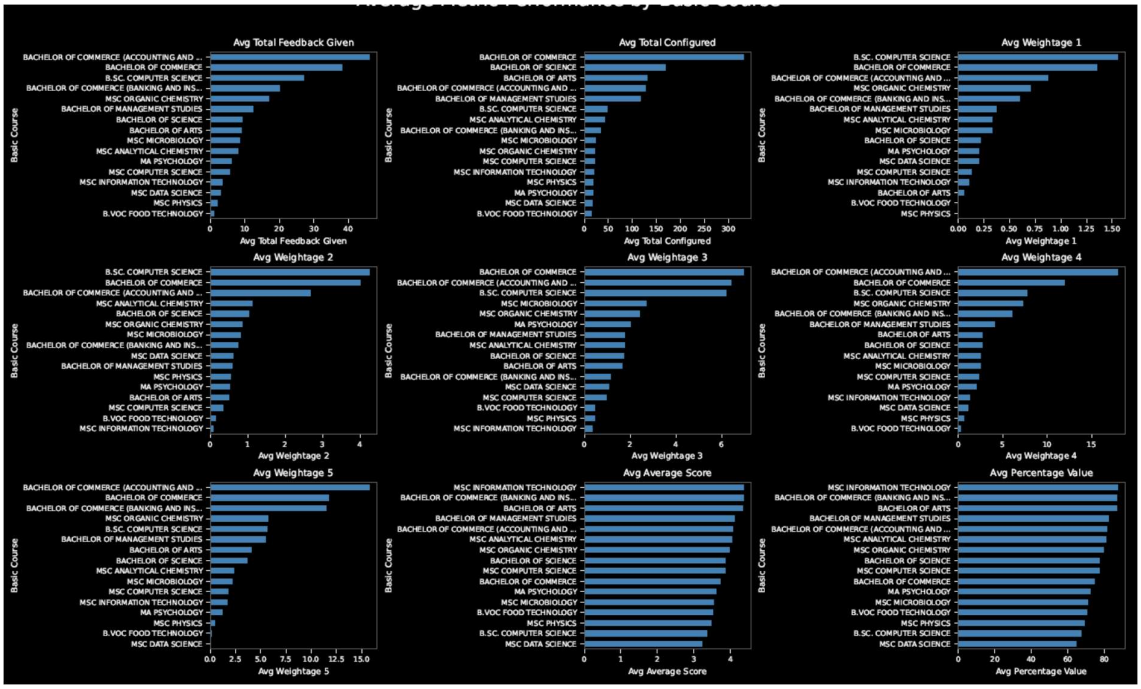
5. Performance Analysis by Course

- **5.1 Overall Performance by Basic Course:** The analysis of Average_Score by Basic Course provides an overall satisfaction ranking.

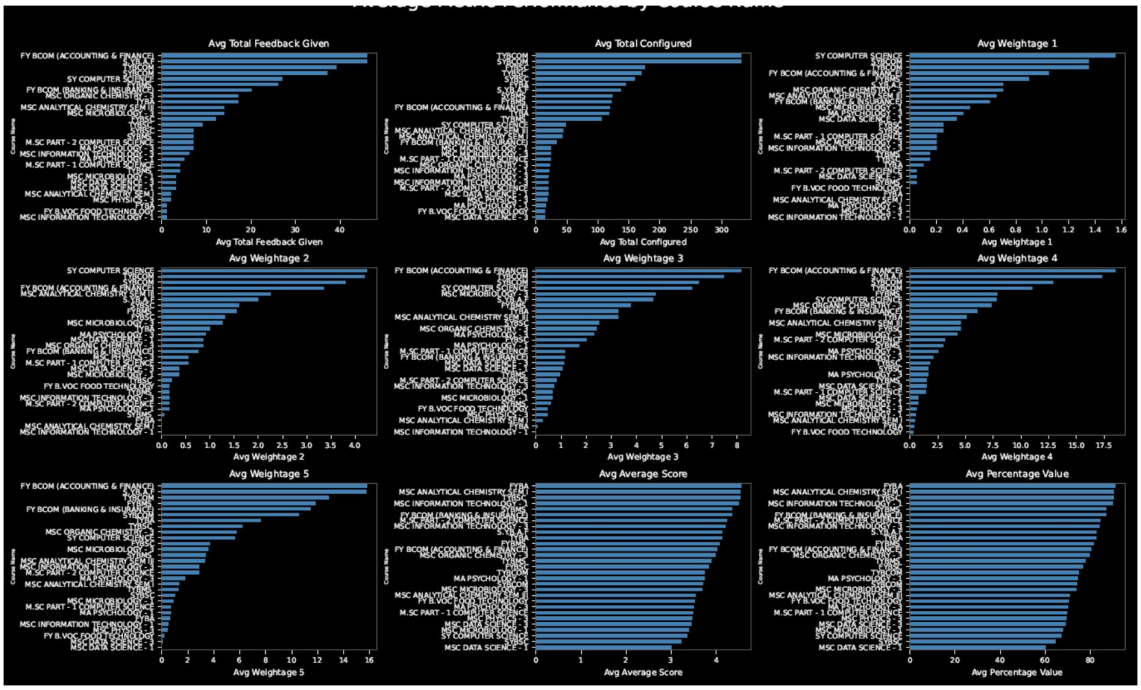


The "Overall Average Metric Performance by Basic Course (Average Score)" plot on Page 4 of Student_Satisfaction_Analysis_Plots.pdf visually confirms these rankings, with MSC Information Technology and Bachelor of Commerce (Banking and Insurance) showing the highest average scores.

- **5.2 Detailed Performance by Basic Course (All Metrics):** Detailed analysis across all metrics by Basic Course reveals nuanced performance.



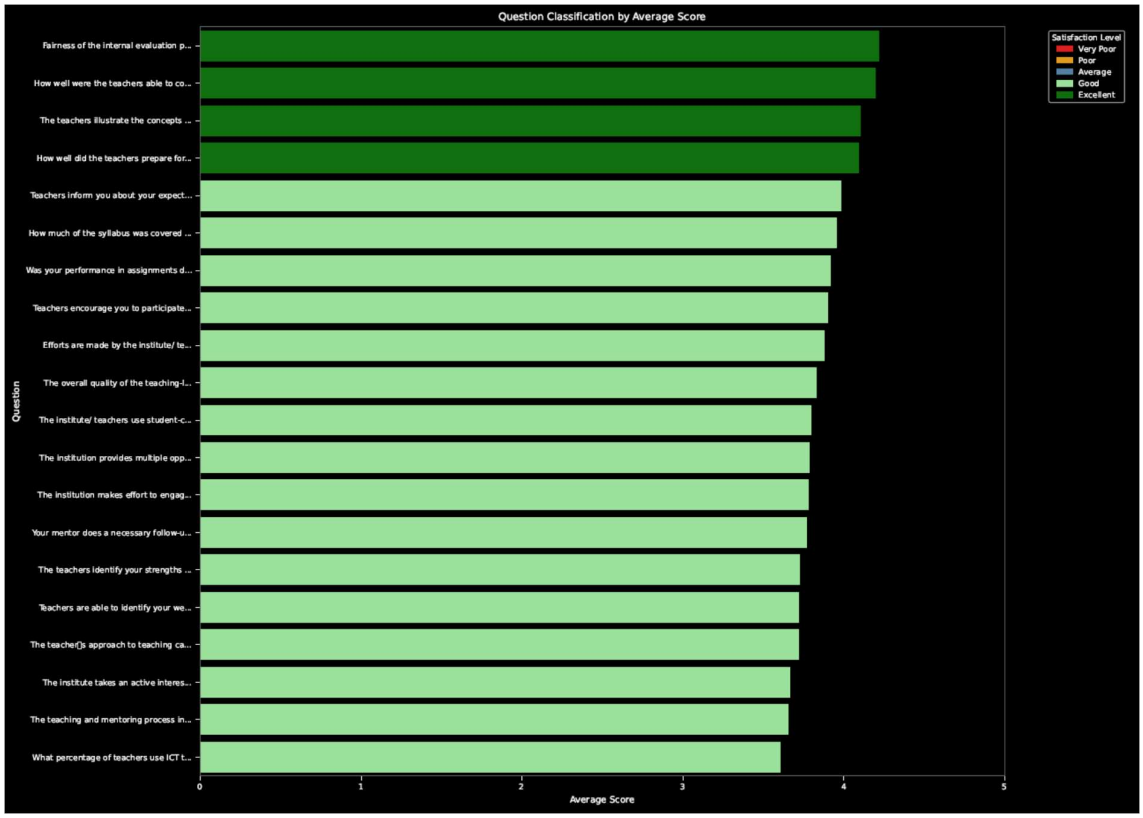
- For instance, while B.SC. COMPUTER SCIENCE has a lower average score, it shows high average values for Total Feedback Given and Total Configured, suggesting high engagement but perhaps specific areas of dissatisfaction. Conversely, courses like MSC INFORMATION TECHNOLOGY consistently perform well across Average_Score and Percentage_Value. These detailed plots are available on Page 5 of Student_Satisfaction_Analysis_Plots.pdf.
- **5.3 Performance by Specific Course Name:** Similar trends are observed when looking at individual Course Names. For example, 'MSC ANALYTICAL CHEMISTRY SEM I' and 'MSC INFORMATION TECHNOLOGY - 1' show very high average scores (4.5250 and 4.5000 respectively), indicating strong satisfaction within these specific programs. Conversely, 'MSC DATA SCIENCE - 1' and 'MSC DATA SCIENCE - 3' have lower average scores (2.9960 and 3.4460), aligning with the overall lower performance of the MSC DATA SCIENCE basic course.



- These detailed plots are available on Page 6 of Student_Satisfaction_Analysis_Plots.pdf.

6. Question-Level Analysis and Classification

- **6.1 Average Scores per Question:** The survey questions were analyzed based on their average Average_Score. Top-scoring questions (indicating highest satisfaction):



- **6.2 Satisfaction Level Classification:** Questions were classified into satisfaction levels based on their Average_Score (on a 1-5 scale) using the following bins:
    - 0-1: Very Poor
    - 1-2: Poor
    - 2-3: Average
    - 3-4: Good
    - 4-5: Excellent

Questions classified as "Excellent" (Average Score 4-5):

- Fairness of the internal evaluation process by the teachers.
- How well were the teachers able to communicate?
- The teachers illustrate the concepts through examples and applications.
- How well did the teachers prepare for the classes?

Questions classified as "Good" (Average Score 3-4):

- Teachers inform you about your expected competencies, course outcomes, and programme outcomes.

- How much of the syllabus was covered in the class?
- Was your performance in assignments discussed with you?
- Teachers encourage you to participate in extracurricular activities.
- Efforts are made by the institute/ teachers to inculcate soft skills, life skills and employability skills to make you ready for the world of work.
- The overall quality of the teaching-learning process in your institute.
- The institute/ teachers use student-centric methods, such as experiential learning, participative learning and problem-solving methodologies for enhancing learning experiences.
- The institution provides multiple opportunities to learn and grow.
- The institution makes effort to engage students in the monitoring, review and continuous quality improvement of the teaching-learning process.
- Your mentor does a necessary follow-up with an assigned student at regular intervals for academic and personal guidance.
- The teachers identify your strengths and encourage you to provide the proper level of challenges.
- Teachers are able to identify your weaknesses and help you to overcome them.
- The teacher's approach to teaching can best be described as...
- The institute takes an active interest in promoting internship, student exchange, field visit opportunities for students.
- The teaching and mentoring process in your institution facilitates your overall development.
- What percentage of teachers use ICT tools such as LCD Projector, Multimedia, etc., while teaching?

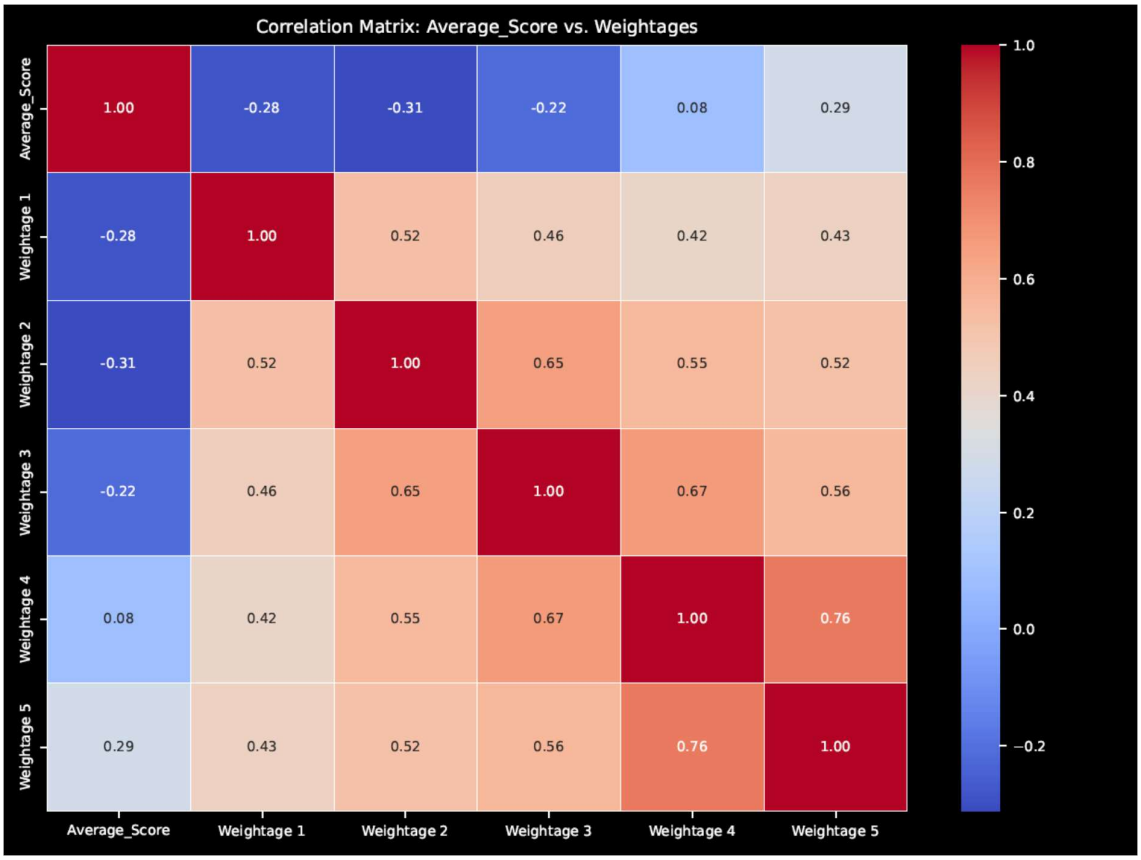
The "Question Classification by Average Score" plot on Page 7 of Student_Satisfaction_Analysis_Plots.pdf provides a visual representation of these classifications, clearly showing the distribution of questions across satisfaction levels. All questions fell into the "Good" or "Excellent" categories, indicating a generally high level of student satisfaction.

7. Correlation and Regression Analysis

7.1 Correlation Analysis Results

This analysis aimed to understand the linear relationships between the student satisfaction Weightage metrics and the overall Average_Score.

The correlation matrix: (see Page 8 of Student_Satisfaction_Analysis_Plots.pdf)



- **Key Findings:** The correlation matrix revealed varying degrees of relationships between the individual weightage factors and the Average_Score.
    - Weightage 5 (Best rating) showed the strongest positive correlation with Average_Score (r = 0.29), indicating that higher ratings in the top satisfaction category are associated with higher overall average scores.
    - Conversely, Weightage 1 (Lowest grade on scale) and Weightage 2 (better than lowest but below average rating) exhibited negative correlations with Average_Score (r = -0.28 and r = -0.31 respectively), suggesting that higher instances of very low ratings are associated with lower overall average scores.
    - The 'Correlation of Weightage Metrics with Average_Score' bar plot (also on Page 8 of Student_Satisfaction_Analysis_Plots.pdf) visually confirms these relationships, highlighting the relative strength and direction of each weightage's correlation with the average score.

7.2 Multiple Linear Regression Results

A multiple linear regression model was constructed to predict the Average_Score (dependent variable, Y) based on the five Weightage metrics (independent variables, X). This analysis helps quantify the individual impact of each weightage category on the overall average satisfaction.

- **Model Summary:** The full regression results summary is presented on Page 9 of Student_Satisfaction_Analysis_Plots.pdf.

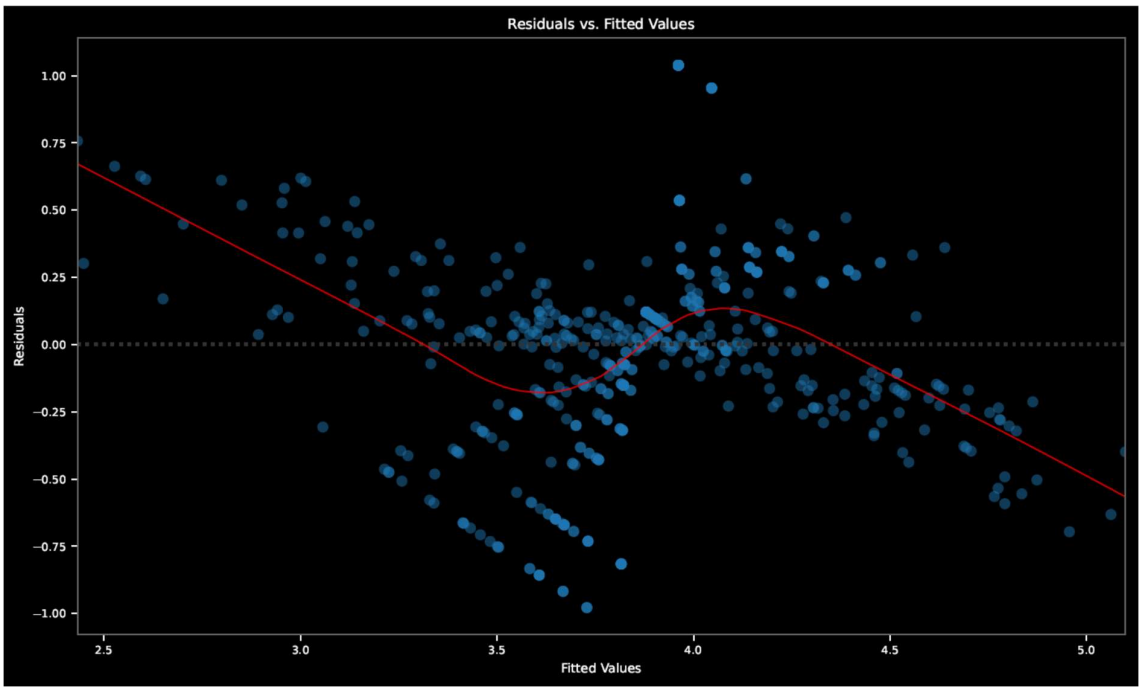
```
Multiple Linear Regression Results Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:          Average_Score   R-squared:                       0.482
Model:                            OLS   Adj. R-squared:                  0.477
Method:                 Least Squares   F-statistic:                     106.6
Date:                Sun, 20 Jul 2025   Prob (F-statistic):           1.64e-79
Time:                        14:11:45   Log-Likelihood:                -313.16
No. Observations:                 580   AIC:                             638.3
Df Residuals:                     574   BIC:                             664.5
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.8762      0.025    155.784      0.000       3.827       3.925
Weightage 1   -0.2518      0.031     -8.053      0.000      -0.313      -0.190
Weightage 2   -0.1480      0.016     -9.006      0.000      -0.180      -0.116
Weightage 3   -0.0605      0.010     -5.977      0.000      -0.080      -0.041
Weightage 4    0.0030      0.006      0.509      0.611      -0.008       0.014
Weightage 5    0.0847      0.005     15.740      0.000       0.074       0.095
==============================================================================
Omnibus:                       14.007   Durbin-Watson:                   1.066
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               16.711
Skew:                           0.282   Prob(JB):                     0.000235
Kurtosis:                       3.611   Cond. No.                         18.3
==============================================================================
```

  - **Overall Model Fit:** The model achieved an **R-squared of 0.482** (Adjusted R-squared = 0.477), indicating that approximately 48.2% of the variance in Average_Score can be explained by the combined influence of the Weightage metrics.
  - **Model Significance:** The model is statistically significant, as indicated by a **F-statistic of 106.6** and a **Prob (F-statistic) of 1.64e-79** (which is effectively 0). This low p-value suggests that the independent variables collectively have a significant linear relationship with the Average_Score.

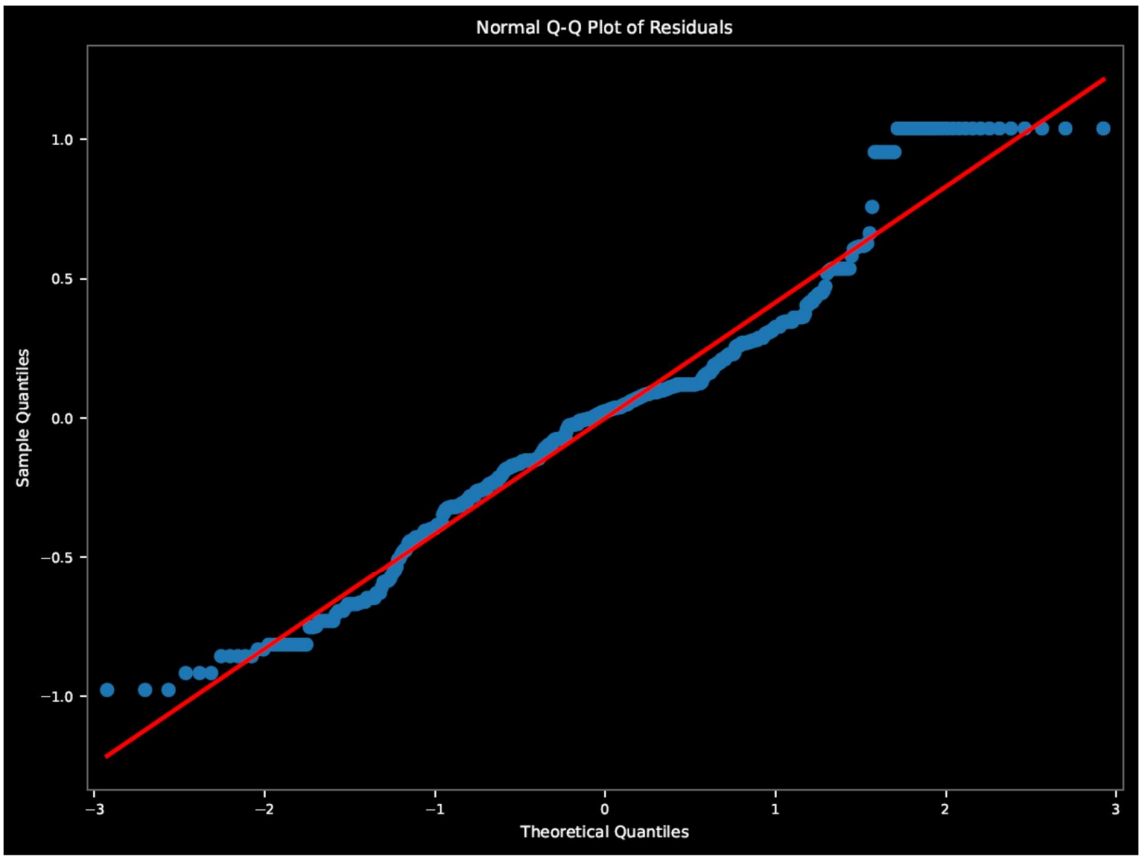
- **Coefficients and Significance:**
  - **Weightage 5 (Coefficient: 0.0847, p < 0.001):** This metric is a highly significant positive predictor. For every one-unit increase in Weightage 5 (i.e., more students giving the highest rating), the Average_Score is expected to increase by approximately 0.0847 units, holding other weightages constant. This reinforces the importance of achieving excellent ratings.

- **Weightage 1 (Coefficient: -0.2518, p < 0.001):** This metric is a highly significant negative predictor. For every one-unit increase in Weightage 1 (i.e., more students giving the lowest rating), the Average_Score is expected to decrease by approximately 0.2518 units. This highlights the strong negative impact of very low satisfaction.
- **Weightage 2 (Coefficient: -0.1481, p < 0.001):** Similar to Weightage 1, this is a significant negative predictor, indicating that higher instances of below-average ratings also negatively impact the overall score.
- **Weightage 3 (Coefficient: -0.0605, p = 0.010):** This metric is a statistically significant negative predictor, though with a smaller coefficient than Weightage 1 and 2.
- **Weightage 4 (Coefficient: 0.0038, p = 0.509):** This metric was **not statistically significant** in predicting Average_Score when controlling for other weightages. Its coefficient is close to zero, and its p-value is greater than 0.05.
- **Intercept (Constant: 3.8762, p < 0.001):** This represents the predicted Average_Score when all Weightage variables are zero.

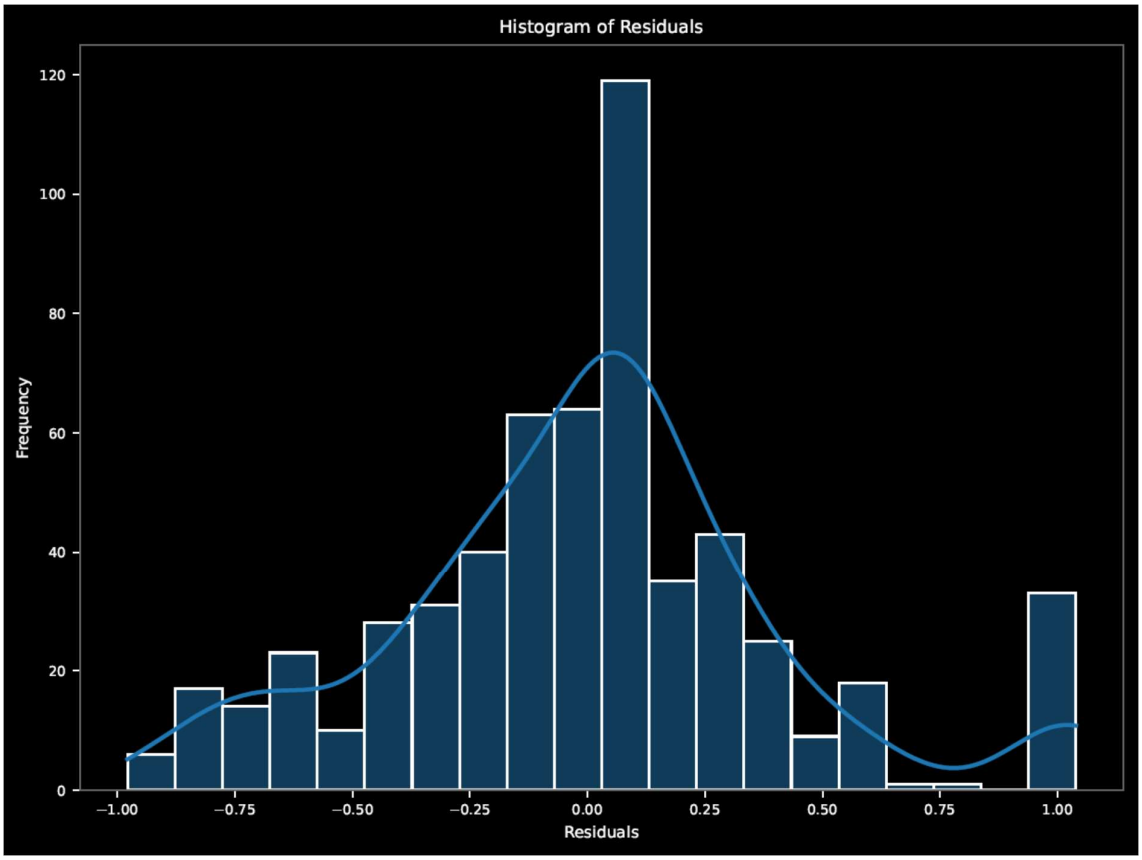
- **Model Diagnostics (Residual Plots):** To assess the model's assumptions, diagnostic plots were generated (see Page 10 of Student_Satisfaction_Analysis_Plots.pdf).
  - **Residuals vs. Fitted Values:**



- The plot shows a relatively random scatter of residuals around the zero line, suggesting that the assumptions of linearity and homoscedasticity (constant variance of residuals) are reasonably met.
- **Normal Q-Q Plot:**



- The Q-Q plot of residuals indicates that the residuals generally follow a normal distribution, as most points lie close to the 45-degree line.
- **Histogram of Residuals:**



- The histogram of residuals appears approximately bell-shaped, further supporting the normality assumption.

8. Key Findings and Recommendations

- **8.1 Key Findings:**
  - **Overall High Satisfaction:** The student satisfaction levels are generally high, with all survey questions falling into the "Good" or "Excellent" categories based on average scores.
  - **Strengths in Teaching and Evaluation:** Students are highly satisfied with the fairness of internal evaluation, teacher communication, and teachers' ability to illustrate concepts and prepare for classes. These are significant strengths of the institution.
  - **Varied Course Performance:** While overall satisfaction is high, there are noticeable differences in satisfaction levels across Basic Courses and specific Course Names. MSC Information Technology and Bachelor of Commerce (Banking and Insurance) appear to be top performers, while MSC Data

Science and B.SC. Computer Science show relatively lower (though still "Good") average satisfaction.

- **Opportunities for Enhancement:** Areas with slightly lower average scores, such as the use of ICT tools by teachers, promotion of internships/exchange programs, and syllabus coverage, present opportunities for targeted improvements.
- **Engagement Metrics:** Some courses with lower average satisfaction scores still show high Total Feedback Given and Total Configured values, suggesting that students in these programs are actively engaged in providing feedback, which is a positive sign for continuous improvement efforts.
- **Lack of Sentiment Analysis:** The current analysis focused solely on quantitative ratings. Qualitative feedback (comments) was not analyzed for sentiment, which could provide deeper insights into the "why" behind the ratings.
- **Impact of Rating Categories:** The regression analysis quantitatively confirms that the highest Weightage 5 ratings significantly boost overall satisfaction, while Weightage 1 and Weightage 2 ratings have a substantial negative impact. Weightage 4 was not a significant predictor in this model.
- **Predictive Power:** The Weightage metrics collectively explain a significant portion (48.2%) of the variation in the Average_Score, demonstrating their strong influence on overall student satisfaction.

- **8.2 Recommendations:** Based on these findings, the following recommendations are proposed:
  - **Leverage Strengths:**
    - **Best Practices Sharing:** Identify and document the best practices employed by teachers in areas of high satisfaction (e.g., communication, concept illustration, fair evaluation). Share these practices across departments and courses to uplift overall teaching quality.
    - **Highlight Successes:** Publicize the high satisfaction in top-performing courses (e.g., MSC Information Technology, Bachelor of Commerce (Banking and Insurance)) to attract prospective students and reinforce positive perceptions.
  - **Targeted Improvements for "Good" Categories:**
    - **Enhance ICT Tool Usage:** Provide training and resources for teachers to increase their effective use of ICT tools (LCD Projectors,

Multimedia) in classes. This could involve workshops, access to modern equipment, and showcasing successful examples.

        - **Review Syllabus Coverage:** Investigate why "How much of the syllabus was covered in the class?" is not in the "Excellent" category. This might involve curriculum review, time management strategies for teachers, or clearer communication of course scope.
        - **Boost Extracurricular Engagement & Development:** Strengthen efforts to encourage student participation in extracurricular activities and to inculcate soft skills, life skills, and employability skills. This could involve more diverse offerings, better promotion, and integration into academic planning.
        - **Promote Experiential Learning:** Continue to emphasize and expand student-centric methods like experiential learning, participative learning, and problem-solving, as these are generally well-received.
    - **Address Course-Specific Areas:**
        - **Deep Dive for Lower-Scoring Courses:** Conduct more focused qualitative and quantitative analysis for courses like MSC Data Science and B.SC. Computer Science to understand the specific pain points contributing to relatively lower satisfaction scores. This could involve direct student interviews or focus groups.
        - **Curriculum and Pedagogical Review:** For these courses, review curriculum relevance, teaching methodologies, and faculty support to identify specific areas for intervention and improvement.
    - **Future Survey Enhancements:**
        - **Implement Sentiment Analysis:** Integrate Natural Language Processing (NLP) tools to analyze qualitative comments from the survey. This will provide a richer understanding of student sentiment (positive, neutral, negative) and uncover specific themes or issues not captured by numerical ratings. This is crucial for understanding the nuances of student feedback.
        - **Refine Questionnaires:** Periodically review and refine survey questions to ensure they are clear, comprehensive, and capture the most relevant aspects of the student experience.
- **Strategic Focus on Extreme Ratings:** Given the significant impact of Weightage 1, Weightage 2, and Weightage 5, efforts should be strategically focused on:

    - **Minimizing Low Ratings:** Implement targeted interventions to address specific issues contributing to Weightage 1 and Weightage 2 scores. This might involve direct feedback mechanisms for students giving low ratings or immediate support for identified problem areas.
    - **Maximizing High Ratings:** Continue and enhance initiatives that lead to Weightage 5 scores. Identify and replicate best practices from areas consistently receiving the highest ratings.
- **Data-Driven Improvement Cycles:** Utilize the regression model's insights to prioritize improvement efforts. For instance, addressing factors that contribute to Weightage 1 and Weightage 2 scores may yield a greater positive change in overall satisfaction than focusing on Weightage 4 if it remains non-significant.

9. Conclusion

This analysis of the student satisfaction survey data provides valuable insights into the strengths and areas for development within the institution's educational offerings. While overall satisfaction is commendable, particularly in areas of teaching quality and evaluation fairness, continuous improvement is essential. By acting on the recommendations, especially by leveraging existing strengths and addressing specific areas identified for enhancement, the institution can further elevate the student experience and foster a more engaging and effective learning environment. Incorporating sentiment analysis in future iterations will undoubtedly provide a more holistic understanding of student feedback.

10. Appendix

- **Student Satisfaction Analysis Plots:** Please refer to the accompanying document, Student_Satisfaction_Analysis_Plots.pdf, for all generated visualizations, including box plots, density plots, performance analyses by course and question classification, **correlation heatmaps, and multiple linear regression results summary table and diagnostic plots (residuals vs. fitted, Q-Q plot, histogram of residuals)**.