+ Code + Text

Task

Analyze student event feedback from "/content/Student_Satisfaction_Survey.csv" to uncover satisfaction trends and suggest improvements using survey data.

Load data

```
import pandas as pd
df = pd.read_csv('/content/Student_Satisfaction_Survey.csv', encoding='latin1')
display(df.head())
display(df.info())
```

	SN	Total Feedback Given	Total Configured	Questions	Weightage 1	Weightage 2	Weightage 3	Weightage 4	Weightage 5	Average/ Percentage	Course Name	Basic C
0	1	1	12	How much of the syllabus was covered in the cl...	0	0	1	0	0	3.00 / 60.00	FY B.VOC FOOD TECHNOLOGY	B.VOC F TECHNOI
1	2	1	12	How well did the teachers prepare for the clas...	0	0	0	0	1	5.00 / 100.00	FY B.VOC FOOD TECHNOLOGY	B.VOC F TECHNOI
2	3	1	12	How well were the teachers able to communicate?	0	0	0	0	1	5.00 / 100.00	FY B.VOC FOOD TECHNOLOGY	B.VOC F TECHNOI
3	4	1	12	The teacher□s approach to teaching can best be...	0	0	1	0	0	3.00 / 60.00	FY B.VOC FOOD TECHNOLOGY	B.VOC F TECHNOI
4	5	1	12	Fairness of the internal evaluation process by...	0	0	0	1	0	4.00 / 80.00	FY B.VOC FOOD TECHNOLOGY	B.VOC F TECHNOI

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 580 entries, 0 to 579
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   SN                    580 non-null    int64
 1   Total Feedback Given  580 non-null    int64
 2   Total Configured      580 non-null    int64
 3   Questions             580 non-null    object
 4   Weightage 1           580 non-null    int64
 5   Weightage 2           580 non-null    int64
 6   Weightage 3           580 non-null    int64
 7   Weightage 4           580 non-null    int64
 8   Weightage 5           580 non-null    int64
 9   Average/ Percentage   580 non-null    object
 10  Course Name           580 non-null    object
 11  Basic Course          580 non-null    object
dtypes: int64(8), object(4)
```

Data cleaning

```
print("Missing values before cleaning:")
print(df.isnull().sum())
print("\nData types before cleaning:")
print(df.dtypes)

df['Average/ Percentage'] = df['Average/ Percentage'].astype(str).str.split('/').str[0].str.split('%').str[0]
df['Average/ Percentage'] = pd.to_numeric(df['Average/ Percentage'], errors='coerce')

df_cleaned = df.drop(columns=['SN', 'Total Feedback Given', 'Total Configured'])

display(df_cleaned.head())
display(df_cleaned.info())
```

```
Missing values before cleaning:
SN                      0
Total Feedback Given    0
Total Configured        0
Questions               0
Weightage 1             0
Weightage 2             0
Weightage 3             0
Weightage 4             0
Weightage 5             0
Average/ Percentage     0
Course Name             0
Basic Course            0
dtype: int64

Data types before cleaning:
SN                       int64
Total Feedback Given     int64
Total Configured         int64
Questions               object
Weightage 1              int64
Weightage 2              int64
Weightage 3              int64
Weightage 4              int64
Weightage 5              int64
Average/ Percentage     object
Course Name             object
Basic Course            object
dtype: object
```

	Questions	Weightage 1	Weightage 2	Weightage 3	Weightage 4	Weightage 5	Average/ Percentage	Course Name	Basic Course
0	How much of the syllabus was covered in the cl...	0	0	1	0	0	3.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY
1	How well did the teachers prepare for the clas...	0	0	0	0	1	5.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY
2	How well were the teachers able to communicate?	0	0	0	0	1	5.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY
3	The teacher□s approach to teaching can best be...	0	0	1	0	0	3.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY
4	Fairness of the internal evaluation process by...	0	0	0	1	0	4.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 580 entries, 0 to 579
Data columns (total 9 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Questions            580 non-null    object
 1   Weightage 1          580 non-null    int64
 2   Weightage 2          580 non-null    int64
 3   Weightage 3          580 non-null    int64
 4   Weightage 4          580 non-null    int64
 5   Weightage 5          580 non-null    int64
```

Sentiment analysis

```
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer

nltk.download('vader_lexicon')

analyzer = SentimentIntensityAnalyzer()

df_cleaned['sentiment_score'] = df_cleaned['Questions'].apply(lambda x: analyzer.polarity_scores(x)['compound'])

display(df_cleaned.head())
```

```
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
```

	Questions	Weightage 1	Weightage 2	Weightage 3	Weightage 4	Weightage 5	Average/ Percentage	Course Name	Basic Course	sentiment_score
0	How much of the syllabus was covered in the cl...	0	0	1	0	0	3.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY	0.0000
1	How well did the teachers prepare for the clas...	0	0	0	0	1	5.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY	0.2732
2	How well were the teachers able to communicate?	0	0	0	0	1	5.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY	0.2732
3	The teacher□s approach to teaching can best be...	0	0	1	0	0	3.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY	0.6369

```
from nltk.sentiment.vader import SentimentIntensityAnalyzer
import nltk

try:
    nltk.data.find('sentiment/vader_lexicon.zip')
except LookupError:
    nltk.download('vader_lexicon')


analyzer = SentimentIntensityAnalyzer()

df_cleaned['sentiment_score'] = df_cleaned['Questions'].apply(lambda x: analyzer.polarity_scores(x)['compound'])

display(df_cleaned.head())
```

	Questions	Weightage 1	Weightage 2	Weightage 3	Weightage 4	Weightage 5	Average/ Percentage	Course Name	Basic Course	sentiment_score
0	How much of the syllabus was covered in the cl...	0	0	1	0	0	3.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY	0.0000
1	How well did the teachers prepare for the clas...	0	0	0	0	1	5.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY	0.2732
2	How well were the teachers able to communicate?	0	0	0	0	1	5.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY	0.2732
3	The teacher□s approach to teaching can best be...	0	0	1	0	0	3.0	FY B.VOC FOOD TECHNOLOGY	B.VOC FOOD TECHNOLOGY	0.6369

Identify satisfaction trends

```
print(df_cleaned.columns)
```

```
Index(['Questions', 'Weightage 1', 'Weightage 2', 'Weightage 3', 'Weightage 4',
       'Weightage 5', 'Average/ Percentage', 'Course Name ', 'Basic Course',
       'sentiment_score'],
      dtype='object')
```

```
df_cleaned.rename(columns={'Course Name ': 'Course Name'}, inplace=True)

overall_average_sentiment = df_cleaned['sentiment_score'].mean()
print(f"Overall Average Sentiment Score: {overall_average_sentiment:.4f}")

course_sentiment = df_cleaned.groupby('Course Name')['sentiment_score'].mean().sort_values(ascending=False)
print("\nAverage Sentiment Score by Course Name:")
display(course_sentiment)
```

```
basic_course_sentiment = df_cleaned.groupby('Basic Course')['sentiment_score'].mean().sort_values(ascending=False)
print("\nAverage Sentiment Score by Basic Course:")
display(basic_course_sentiment)

print("\nSentiment Score Distribution (Descriptive Statistics):")
display(df_cleaned['sentiment_score'].describe())

plt.figure(figsize=(10, 6))
sns.histplot(df_cleaned['sentiment_score'], kde=True, bins=20)
plt.title('Distribution of Sentiment Scores')
plt.xlabel('Sentiment Score')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()

question_sentiment = df_cleaned.groupby('Questions')['sentiment_score'].mean().sort_values()

print("\nQuestions with Lowest Average Sentiment Scores:")
display(question_sentiment.head())

print("\nQuestions with Highest Average Sentiment Scores:")
display(question_sentiment.tail())
```

```
Overall Average Sentiment Score: 0.2873

Average Sentiment Score by Course Name:
```

Course Name	sentiment_score
FY B.VOC FOOD TECHNOLOGY	0.287335
FY BCOM (ACCOUNTING & FINANCE)	0.287335
FY BCOM (BANKING & INSURANCE)	0.287335
FYBA	0.287335
FYBMS	0.287335
FYBSC	0.287335
M.SC PART - 1 COMPUTER SCIENCE	0.287335
M.SC PART - 2 COMPUTER SCIENCE	0.287335
MA PSYCHOLOGY - 1	0.287335
MA PSYCHOLOGY - 3	0.287335
MSC ANALYTICAL CHEMISTRY SEM I	0.287335
MSC ANALYTICAL CHEMISTRY SEM III	0.287335
MSC DATA SCIENCE - 1	0.287335
MSC DATA SCIENCE - 3	0.287335
MSC INFORMATION TECHNOLOGY - 1	0.287335
MSC INFORMATION TECHNOLOGY - 3	0.287335
MSC MICROBIOLOGY - 1	0.287335
MSC MICROBIOLOGY - 3	0.287335
MSC ORGANIC CHEMISTRY - 3	0.287335
MSC PHYSICS - 3	0.287335
S.Y.B.A.F	0.287335
SY COMPUTER SCIENCE	0.287335
SYBCOM	0.287335
SYBMS	0.287335
SYBSC	0.287335
TYBA	0.287335
TYBCOM	0.287335
TYBMS	0.287335
TYBSC	0.287335

dtype: float64

Extract key themes and suggestions

```
import re
from nltk.corpus import stopwords
```

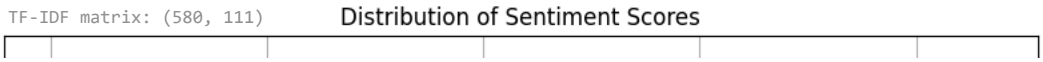
```
from nltk.tokenize import word_tokenize
from sklearn.feature_extraction.text import TfidfVectorizer
import nltk

try:
    nltk.data.find('tokenizers/punkt')
except LookupError:
    nltk.download('punkt')
try:
    nltk.data.find('corpora/stopwords')
except LookupError:
    nltk.download('stopwords')
try:
    nltk.data.find('tokenizers/punkt_tab')
except LookupError:
    nltk.download('punkt_tab')


stop_words = set(stopwords.words('english'))

def preprocess_text(text):
    text = text.lower()
    text = re.sub(r'[^\w\s]', '', text)
    tokens = word_tokenize(text)
    tokens = [word for word in tokens if word not in stop_words]
    return ' '.join(tokens)

df_cleaned['cleaned_questions'] = df_cleaned['Questions'].apply(preprocess_text)

display(df_cleaned[['Questions', 'cleaned_questions']].head())
```

Sentiment Score Distribution (Descriptive Statistics):

	sentiment_score	Questions	cleaned_questions
0		How much of the syllabus was covered in the cl...	much syllabus covered class
1		How well did the teachers prepare for the clas...	well teachers prepare classes
2		How well were the teachers able to communicate?	well teachers able communicate
3		The teacher□s approach to teaching can best be...	teachers approach teaching best described
4		Fairness of the internal evaluation process by...	fairness internal evaluation process teachers
mean	0.287335		
min	0.000000		
50%	0.273200		

```
tfidf_vectorizer = TfidfVectorizer(max_features=1000)
tfidf_matrix = tfidf_vectorizer.fit_transform(df_cleaned['cleaned_questions'])

print("Shape of TF-IDF matrix:", tfidf_matrix.shape)
```

```
Shape of TF-IDF matrix: (580, 111)
```




```
from sklearn.cluster import KMeans

num_clusters = 5
kmeans = KMeans(n_clusters=num_clusters, random_state=42, n_init=10)
clusters = kmeans.fit_predict(tfidf_matrix)

df_cleaned['cluster_label'] = clusters

print(f"\nTop terms per cluster (K={num_clusters}):")
order_centroids = kmeans.cluster_centers_.argsort()[:, ::-1]
terms = tfidf_vectorizer.get_feature_names_out()

for i in range(num_clusters):
    print(f"Cluster {i}:")
    top_terms = [terms[ind] for ind in order_centroids[i, :10]]
    print(f"  Top terms: {', '.join(top_terms)}")

    print("  Sample Questions:")
    sample_questions = df_cleaned[df_cleaned['cluster_label'] == i]['Questions'].sample(min(3, (df_cleaned['cluster_label'] == i).sum()),
    for q in sample_questions:
        print(f"    - {q}")
    print("-" * 30)

cluster_sentiment = df_cleaned.groupby('cluster_label')['sentiment_score'].mean().sort_values(ascending=False)
print("\nAverage Sentiment Score by Cluster:")
display(cluster_sentiment)
```

Questions with Lowest Average Sentiment Scores:

Questions	sentiment_score
Fairness of the internal evaluation process by the teacher	0.0

```
Top terms per cluster (K=5):
Cluster 0:
  Top terms: identify, discussed, performance, assignments, class, much, syllabus, covered, weaknesses, help
  Sample Questions:
    - Teachers are able to identify your weaknesses and help you to overcome them.
    - How much of the syllabus was covered in the class?
    - Your mentor does a necessary follow-up with an assigned task to you.
------------------------------
Cluster 1:
  Top terms: learning, skills, opportunities, institute, learn, provides, multiple, grow, institution, internships
  Sample Questions:
    - Efforts are made by the institute/ teachers to inculcate soft skills, life skills and employability skills to make you ready
    - The institute takes an active interest in promoting internships, student exchange, field visit opportunities for students.
    - The institute/ teachers use student-centric methods, such as experiential learning, participative learning and problem-solvi
------------------------------
Cluster 2:
  Top terms: teaching, best, approach, described, methodology, process, enhance, cognitive, facilitates, growth
  Sample Questions:
    - The teaching and mentoring process in your institution facilitates you in cognitive, social and
emotional growth.
    - The teacher's approach to teaching can best be described as
    - What percentage of teachers use ICT tools such as LCD projectors, Multimedia, etc. while teaching?
------------------------------
Cluster 3:
  Top terms: well, teachers, outcomes, communicate, prepare, classes, ready, extracurricular, participate, activities
  Sample Questions:
    - Teachers encourage you to participate in extracurricular activities.
    - How well did the teachers prepare for the classes?
    - Teachers inform you about your expected competencies, course outcomes and program
outcomes.
------------------------------
Cluster 4:
  Top terms: process, quality, teachinglearning, fairness, evaluation, internal, overall, good, institute, makes
  Sample Questions:
    - The institution makes effort to engage students in the monitoring, review and continuous quality improvement of the teaching
    - Fairness of the internal evaluation process by the teachers.
    - The overall quality of the teaching-learning process in your institute is very good.
------------------------------

Average Sentiment Score by Cluster:
```

cluster_label	sentiment_score
1	0.402975
4	0.384133
2	0.376933
3	0.211400
0	0.158920

dtype: float64

```
Questions with Highest Average Sentiment Scores:
```

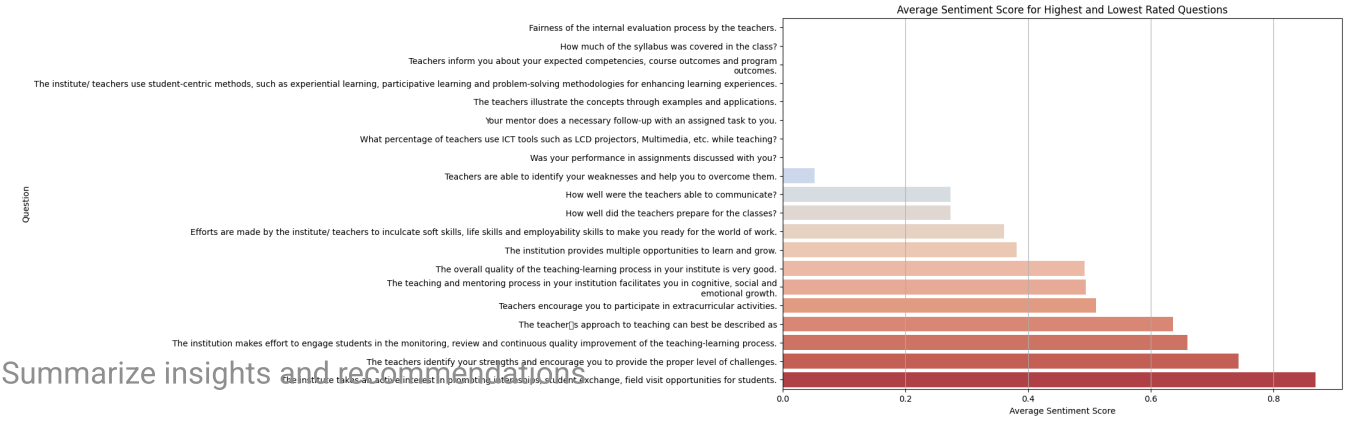
Questions	sentiment_score
Fairness of the internal evaluation process by the teachers.	0.0
How much of the syllabus was covered in the class?	0.0
Teachers inform you about your expected competencies, course outcomes and program outcomes.	0.0
The institute/ teachers use student-centric methods, such as experiential learning, participative learning and problem-solving methodologies for enhancing learning experiences.	0.0
The teachers illustrate the concepts through examples and applications.	0.0
Teachers encourage you to participate in extracurricular activities.	0.5106
The teacher's approach to teaching can best be described as	0.6369
The institution makes effort to engage students in the monitoring, review and continuous quality improvement of the teaching-learning process.	0.6597
The teachers identify your strengths and encourage you to provide the proper level of challenges.	0.7430
The institute takes an active interest in promoting internships, student exchange, field visit opportunities for students.	0.8689

Visualize findings

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

plt.figure(figsize=(10, 6))
sns.barplot(x=cluster_sentiment.index, y=cluster_sentiment.values, palette='viridis')
plt.title('Average Sentiment Score by Cluster')
plt.xlabel('Cluster Label')
plt.ylabel('Average Sentiment Score')
plt.xticks(rotation=0)
plt.grid(axis='y')
plt.show()

top_10_questions = question_sentiment.tail(10)
bottom_10_questions = question_sentiment.head(10)
combined_questions = pd.concat([bottom_10_questions, top_10_questions])

plt.figure(figsize=(12, 8))
sns.barplot(x=combined_questions.values, y=combined_questions.index, palette='coolwarm')
plt.title('Average Sentiment Score for Highest and Lowest Rated Questions')
plt.xlabel('Average Sentiment Score')
plt.ylabel('Question')
plt.grid(axis='x')
plt.show()
```

```
/tmp/ipython-input-42-2510660229.py:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `

  sns.barplot(x=cluster_sentiment.index, y=cluster_sentiment.values, palette='viridis')
```


```
/tmp/ipython-input-42-2510660229.py:18: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `

  sns.barplot(x=combined_questions.values, y=combined_questions.index, palette='coolwarm')
/usr/local/lib/python3.11/dist-packages/IPython/core/pylabtools.py:151: UserWarning: Glyph 146 (\x92) missing from font(s) DejaVu
  fig.canvas.print_figure(bytes_io, **kw)
```



Summarize insights and recommendations

```
print("--- Overall Sentiment ---")
print(f"The overall average sentiment score is: {df_cleaned['sentiment_score'].mean():.4f}")
print("This indicates a generally positive, but not overwhelmingly enthusiastic, sentiment among students.")


print("\n--- Sentiment by Course and Basic Course ---")
print("Average sentiment scores are relatively consistent across different Course Names and Basic Courses.")
print("This suggests that satisfaction levels, based on these survey questions, do not significantly vary by academic program.")


print("\n--- Sentiment Distribution ---")
print("The distribution of sentiment scores shows a range from negative/neutral (0.0) to highly positive (0.8689).")
print("A notable number of responses are neutral or slightly positive, with fewer strongly positive responses.")


print("\n--- Sentiment by Cluster and Thematic Areas ---")
print("Analyzing sentiment by cluster, related to the identified themes:")
display(cluster_sentiment)
print("\nCluster 1 (highest sentiment) relates to the institute's efforts in providing learning opportunities, skill development, inte
print("Cluster 4 and 2 also show relatively high sentiment, relating to overall quality, evaluation fairness, teacher approach, and me
print("Cluster 3 and 0 show lower sentiment, related to teacher communication, preparation, informing about outcomes, teacher support,


print("\n--- Sentiment for Highest and Lowest Rated Questions ---")
print("Questions with the highest sentiment highlight areas of strength:")
```

```
display(top_10_questions)
print("\nThese include the institute's active interest in internships, teachers identifying strengths, engaging students in quality im
print("\nQuestions with the lowest sentiment highlight areas needing improvement:")
display(bottom_10_questions)
print("\nThese include fairness of internal evaluation, syllabus coverage, teachers informing about outcomes, use of student-centric m


print("\n--- Actionable Insights and Recommendations ---")
print("Based on the analysis, here are some actionable insights and recommendations:")

print("\nAreas of Strength to Leverage:")
print("- Internships and Opportunities: The high sentiment around internships, student exchange, and field visits suggests these are h
print("- Teacher Engagement and Support (Identifying Strengths): Teachers identifying strengths is perceived positively. Recommendatio
print("- Overall Quality and Teacher Approach: The generally positive sentiment towards overall quality and teacher approach/mentoring

print("\nAreas for Improvement:")
print("- Fairness of Internal Evaluation and Syllabus Coverage: These questions received the lowest sentiment. Recommendation: Review
print("- Teacher Communication and Preparation: Lower sentiment regarding communication and preparation for classes. Recommendation: P
print("- Student-Centric Methods and Concept Illustration: Low sentiment suggests these teaching methods may not be consistently appli
print("- Mentor Follow-up and Assignment Discussion: Low sentiment in these areas. Recommendation: Implement guidelines or training fo
print("- Use of ICT Tools: Low sentiment regarding the percentage of teachers using ICT tools. Recommendation: Invest in and promote t
print("- Informing about Outcomes: Low sentiment regarding teachers informing about expected competencies, course outcomes, and progra

print("\nGeneral Recommendations:")
print("- Gather More Detailed Feedback: The current 'Questions' column is limited. Implement mechanisms for collecting more detailed,
print("- Follow-up on Low-Sentiment Areas: Conduct targeted surveys or focus groups in areas identified with low sentiment to understa
```

```
--- Overall Sentiment ---
The overall average sentiment score is: 0.2873
This indicates a generally positive, but not overwhelmingly enthusiastic, sentiment among students.

--- Sentiment by Course and Basic Course ---
Average sentiment scores are relatively consistent across different Course Names and Basic Courses.
This suggests that satisfaction levels, based on these survey questions, do not significantly vary by academic program.

--- Sentiment Distribution ---
The distribution of sentiment scores shows a range from negative/neutral (0.0) to highly positive (0.8689).
A notable number of responses are neutral or slightly positive, with fewer strongly positive responses.

--- Sentiment by Cluster and Thematic Areas ---
Analyzing sentiment by cluster, related to the identified themes:
```

cluster_label	sentiment_score
1	0.402975
4	0.384133
2	0.376933
3	0.211400
0	0.158920

dtype: float64

```
Cluster 1 (highest sentiment) relates to the institute's efforts in providing learning opportunities, skill development, internshi
Cluster 4 and 2 also show relatively high sentiment, relating to overall quality, evaluation fairness, teacher approach, and mento
Cluster 3 and 0 show lower sentiment, related to teacher communication, preparation, informing about outcomes, teacher support, sy

--- Sentiment for Highest and Lowest Rated Questions ---
Questions with the highest sentiment highlight areas of strength:
```

Questions	sentiment_score
How well did the teachers prepare for the classes?	0.2732
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.	0.3612
The institution provides multiple opportunities to learn and grow.	0.3818
The overall quality of the teaching-learning process in your institute is very good.	0.4927
The teaching and mentoring process in your institution facilitates you in cognitive, social and\nemotional growth.	0.4939
Teachers encourage you to participate in extracurricular activities.	0.5106
The teacher□s approach to teaching can best be described as	0.6369
The institution makes effort to engage students in the monitoring, review and continuous quality improvement of the teaching-learning process.	0.6597