

```
In [1]: import pandas as pd
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
import string
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
```

```
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\asmaa\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\asmaa\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data] C:\Users\asmaa\AppData\Roaming\nltk_data...
[nltk_data] Package wordnet is already up-to-date!
```

```
Out[1]: True
```

explore the data

```
In [2]: # Load the first dataset (feedback data)
feedback_path = 'C:\\Users\\asmaa\\OneDrive\\Desktop\\americanuniversity\\Student_FeedB
feedback_df = pd.read_excel(feedback_path)
```

```
In [3]: feedback_df.isnull().sum()
```

```
Out[3]: AcademicYear      0
SemesterName      0
ParticipantID      0
DegreeLevel      0
Gender      0
Nationality      0
StudentProgram      0
CourseCode      0
CourseName      0
QuestionText      0
QuestionType      0
ParticipantResponse    12121
dtype: int64
```

```
In [4]: # Load the second dataset (course information)
course_info_path = 'C:\\Users\\asmaa\\OneDrive\\Desktop\\americanuniversity\\Course_INF
course_info_df = pd.read_excel(course_info_path)
```

```
In [5]: course_info_df.isnull().sum()
```

```
Out[5]: CourseCode      0
CourseName      0
College      0
DegreeName      0
```

```
FieldOfStudy    0
dtype: int64
```

```
In [6]: course_info_df.DegreeName.unique()
```

```
Out[6]: array(['Master Degree', 'Bachelor Degree'], dtype=object)
```

```
In [7]: course_info_df.College.unique()
```

```
Out[7]: array(['College of Media and Mass Communication',
               'College of Business Administration', 'College of Education',
               'Foundation Program', 'College of Security and Global Studies',
               'College of Computer Information Technology', 'College of Design',
               'College of Law', 'Training Courses'], dtype=object)
```

```
In [8]: course_info_df.FieldOfStudy.unique()
```

```
Out[8]: array(['UNSPECIFIED', 'BA', 'EDUCATION', 'IT', 'ARTS_DESIGN',
               'MASS_COMM_PR', 'LAW_SHARIA', 'HUMANITIES_SOCIAL_SC'], dtype=object)
```

```
In [9]: # Merge the datasets based on 'CourseCode'
merged_df = feedback_df.merge(course_info_df, on='CourseCode', how='left')
```

```
In [10]: # Save the merged_df DataFrame to an output file
output_filename = 'C:\\\\Users\\asmaa\\OneDrive\\Desktop\\americanuniversity\\merged_df.x
merged_df.to_excel(output_filename, index=False)
```

Data preprocessing

```
In [11]: #check null values
#check duplicate values
#check inconsistence data
```

```
In [12]: # Check for missing values in the dataframe
missing_values = merged_df.isnull().sum()
missing_values
```

```
Out[12]: AcademicYear          0
SemesterName                  0
ParticipantID                 0
DegreeLevel                   0
Gender                        0
Nationality                   0
StudentProgram                0
CourseCode                    0
CourseName_x                  0
QuestionText                  0
QuestionType                  0
ParticipantResponse    12121
CourseName_y            6541
College                 6541
DegreeName              6541
```

FieldOfStudy
dtype: int64

6541

Remove Duplicate Values

```
In [13]: # Check for duplicate values in the dataframe
duplicate_rows = merged_df.duplicated()

# Filter the DataFrame to show only the duplicate rows
duplicate_data = merged_df[duplicate_rows]
```

```
In [14]: merged_df.drop_duplicates(inplace=True)
```

drop Un necessary columns

```
In [15]: # column CourseName_y same CourseName_x
merged_df.drop(columns=['CourseName_y'], inplace=True)
```

check Data Value

```
In [16]: for column in merged_df.columns:
unique_values = merged_df[column].unique()
print(f"Unique values in column '{column}':\n{unique_values}\n")
```

Unique values in column 'AcademicYear':
[2021 2022 2023]

Unique values in column 'SemesterName':
['Fall 2021 - 2022' 'Fall 2022 - 2023' 'Spring 2022 - 2023'
'Spring 2021 - 2022' 'Summer I 2021 - 2022' 'Summer II 2021 - 2022'
'Summer I 2022 - 2023' 'Summer II 2022 - 2023' 'Fall 2023 - 2024']

Unique values in column 'ParticipantID':
[7040336900 6707335167 7373669733 ... 6390334967 5724000267 7041000167]

Unique values in column 'DegreeLevel':
['Bachelor' 'Master' 'Non-Degree']

Unique values in column 'Gender':
['Female' 'Male']

Unique values in column 'Nationality':
['Palestine' 'United Arab Emirates' 'Syria' 'Jordan'
'Kingdom of Saudi Arabia' 'Egypt' 'Lebanon' 'Nigeria' 'Yemen' 'Bahrain'
'Iraq' 'Pakistan' 'Somalia' 'Sudan' 'Bangladesh' 'Iran' 'Morocco'
'United States of America' 'India' 'Libya' 'Turkey' 'Afghanistan'
'Comoros' 'Dominican Republic' 'United Kingdom' 'Russia' 'Tunisia'
'Burkina Faso' 'Sweden' 'Kazakstan' 'Germany' 'Commonwealth of Dominica'
'Oman' 'Tanzania' 'Philippines' 'Angola' 'Congo Republic' 'Kenya'
'Republic of Tajikistan' 'Algeria' 'Canada' 'Italy' 'Armenia'
'Cote d'Ivoire - Ivory Coast' 'Ethiopia' 'Djibouti' 'Uzbekistan'
'St. Kitts and Nevis' 'Czech Republic' 'Antigua and Barbuda' 'Bulgaria'
'Brazil' 'China People's Republic' 'Serbia' 'Australia' 'Kuwait']

'Eriterea' 'Mauritania' 'Guinea']

Unique values in column 'StudentProgram':

['Bachelor of Business Administration'
'Bachelor of Arts in Media and Mass Communication'
'Bachelor of Science in Information Technology Management'
'Bachelor of Science in Design - Interior Design'
'Bachelor of Science in Design - Fashion Design'
'Master of Business Administration'
'Bachelor of Arts in Security and Strategic Studies' 'Bachelor of Law'
'Bachelor of Science in Computer Science'
'Bachelor of Science in Design - Graphic Design'
'Bachelor of Public Relations'
'Bachelor of Science in Design - Digital Animation'
'Bachelor of Education'
'Master of Arts in Security and Strategic Studies'
'Master of Arts in Diplomacy' 'Professional Master in Sports Law'
'Master of Sports Management'
'Master in Security Studies and Information Analysis'
'Master of Knowledge Management' 'Master in Arbitration'
'Master in Intellectual Property' 'Master in Criminal Sciences']

Unique values in column 'CourseCode':

['ACC 100' 'ACC 105' 'ACC 200' 'ACC 205' 'ACC 300' 'ACC 301' 'ACC 306'
'ACC 400' 'ACC 401' 'ACC 402' 'ACC 405' 'ACC 500' 'ACC 600' 'ARLAA 100'
'ARLAA 105' 'ARLNA 100' 'ASC 100L' 'ASC 101' 'ASC 101AR' 'ASC 102'
'ASC 102AR' 'ASC 104' 'ASC 105' 'ASC 105L' 'ASC 106' 'ASC 110'
'ASC 110AR' 'ASC 111' 'ASC 200' 'ASC 200AR' 'ASC 201' 'ASC 201AR'
'ASC 202' 'ASC 202AR' 'ASC 203' 'ASC 203AR' 'ASC 204' 'ASC 204AR'
'ASC 205' 'ASC 205AR' 'ASC 206' 'ASC 206AR' 'ASC 207' 'ASC 207AR'
'ASC 208' 'ASC 208AR' 'ASC 210' 'ASC 210AR' 'ASC 211' 'ASC 212' 'ASC 213'
'CEE 800' 'CIT 100L' 'CIT 200' 'CIT 201' 'CIT 202' 'CIT 205' 'CIT 208'
'CIT 212' 'CIT 214' 'CIT 300' 'CIT 301' 'CIT 303' 'CIT 305' 'CIT 307'
'CIT 310' 'CIT 400' 'CIT 402' 'CIT 405' 'CIT 407' 'CIT 409' 'CIT 410'
'CIT 411' 'CIT 413' 'CIT 415' 'CIT 90' 'CSC 110' 'CSC 111' 'CSC 200'
'CSC 201' 'CSC 203' 'CSC 206' 'CSC 210' 'CSC 212' 'CSC 215' 'CSC 301'
'CSC 302' 'CSC 303' 'CSC 304' 'CSC 306' 'CSC 310' 'CSC 315' 'CSC 320'
'CSC 330' 'CSC 400' 'CSC 401' 'CSC 411' 'CSC 417' 'CSCS 310' 'CSCS 400'
'CSCS 403' 'CSCS 404' 'CSCS 412' 'CSCS 413' 'CSCS 414' 'DES 100'
'DES 101' 'DES 102' 'DES 104' 'DES 121' 'DES 122' 'DES 210' 'DES 211'
'DES 212' 'DES 301' 'DES 414' 'DIA 120' 'DIA 214' 'DIA 222' 'DIA 223'
'DIA 224' 'DIA 303' 'DIA 305' 'DIA 312' 'DIA 313' 'DIA 314' 'DIA 322'
'DIA 323' 'DIA 325' 'DIA 404' 'DIA 411' 'DIP 600' 'DIP 607' 'DIP 609'
'ECO 100' 'ECO 100L' 'ECO 200' 'ECO 600' 'EDP 313' 'EDP 410' 'EDP 420'
'EDP 421' 'EDU 110' 'EDU 121AR' 'EDU 122' 'EDU 123' 'EDU 211' 'EDU 212'
'EDU 213' 'EDU 223' 'EDU 224AR' 'EDU 227' 'EDU 228' 'EMSAT' 'ENG 100AR'
'ENG 101' 'ENG 101A' 'ENG 102' 'ENG 105AR' 'ENG 200' 'ENG 201' 'ENG 301'
'ENG 99' 'FAD 100' 'FAD 101' 'FAD 104' 'FAD 105' 'FAD 110' 'FAD 121'
'FAD 122' 'FAD 210' 'FAD 211' 'FAD 212' 'FAD 215' 'FAD 300' 'FAD 305'
'FIN 200' 'FIN 300' 'FIN 301' 'FIN 302' 'FIN 400' 'FIN 401' 'FIN 402'
'FIN 404' 'FIN 600' 'FIN 601' 'FSD 202' 'FSD 211' 'FSD 222' 'FSD 223'
'FSD 224' 'FSD 301' 'FSD 306' 'FSD 307' 'FSD 311' 'FSD 312' 'FSD 314'
'FSD 321' 'FSD 322' 'FSD 402' 'FSD 404' 'FSD 410' 'FSD 412' 'FSD 421'
'FSD 422' 'GRD 213' 'GRD 221' 'GRD 222' 'GRD 223' 'GRD 224' 'GRD 300'
'GRD 303' 'GRD 311' 'GRD 312' 'GRD 321' 'GRD 322' 'GRD 401' 'GRD 402'
'GRD 411' 'GRD 421' 'HAP 200' 'HAP 200AR' 'HHM 200' 'HHM 300' 'HHM 302'
'HHM 304' 'HHM 305' 'HHM 400' 'HHM 401' 'HHM 402' 'HHM 404' 'HRM 205'
'HRM 300' 'HRM 302' 'HRM 303' 'HRM 304' 'HRM 400' 'HRM 401' 'HRM 402'
'HRM 404' 'HRM 405' 'IELTS' 'IMC 300' 'IMC 302' 'IMC 310' 'IMC 400'
'IMC 402' 'IMC 405' 'IND 122' 'IND 212' 'IND 214' 'IND 221' 'IND 222'
'IND 224' 'IND 225' 'IND 226' 'IND 303' 'IND 306' 'IND 310' 'IND 311'
'IND 313' 'IND 314' 'IND 316' 'IND 317' 'IND 321' 'IND 322' 'IND 323'
'IND 324' 'IND 325' 'IND 326' 'IND 327' 'IND 402' 'IND 403' 'IND 406'
'IND 411' 'IND 412' 'IND 413' 'IND 415' 'IND 421' 'INTERMED' 'INTERN 101'
'INV 200' 'IRE 300' 'IRE 300AR' 'IRE 301' 'IRE 302' 'IRE 310' 'IRE 400']

'IRE 402' 'IRE 405' 'IRM 200' 'IRM 300' 'IRM 302' 'IRM 304' 'IRM 305'
 'IRM 400' 'IRM 401' 'IRM 402' 'IRM 404' 'IRM 405' 'ITM 306' 'ITM 310'
 'ITM 330' 'ITM 400' 'ITM 404' 'ITM 411' 'KMI 600' 'KMI 601' 'KMI 602'
 'KMI 603' 'KMI 605' 'KMI 608' 'KMI 610' 'KMI 612' 'KMI 613' 'KMI 620'
 'KMI 700' 'LAR 600' 'LAR 601' 'LAR 602' 'LAR 603' 'LAR 604' 'LAR 605'
 'LAR 606' 'LAR 607' 'LAR 630' 'LAW 100' 'LAW 102' 'LAW 103' 'LAW 104'
 'LAW 105' 'LAW 200' 'LAW 201' 'LAW 202' 'LAW 203' 'LAW 204' 'LAW 205'
 'LAW 206' 'LAW 207' 'LAW 209' 'LAW 211' 'LAW 213' 'LAW 300' 'LAW 301'
 'LAW 302' 'LAW 303' 'LAW 304' 'LAW 305' 'LAW 306' 'LAW 308' 'LAW 309'
 'LAW 310' 'LAW 312' 'LAW 315' 'LAW 400' 'LAW 401' 'LAW 402' 'LAW 403'
 'LAW 404' 'LAW 405' 'LAW 407' 'LAW 408' 'LAW 409' 'LAW 411' 'LAW 413'
 'LAW 414' 'LAW 415' 'LAW 416' 'LAW 418' 'LAW 419' 'LAW 420' 'LAW 501'
 'LAW 512' 'LAW 513' 'LAW 514' 'LAW 516' 'LAW 521' 'LAW 526' 'LAW 540'
 'LAW 541' 'LAW 542' 'LAW 543' 'LAW 550' 'LAW 551' 'LAW 552' 'LAW 601'
 'LAW 602' 'LAW 603' 'LAW 604' 'LAW 605' 'LAW 606' 'LAW 607' 'LAW 608'
 'LAW 609' 'LAW 610' 'LAW 614' 'LAW 615' 'LAW 617' 'LAW 618' 'LAW 621'
 'LCS 600' 'LCS 601' 'LCS 602' 'LCS 603' 'LCS 604' 'LCS 605' 'LCS 606'
 'LCS 607' 'LCS 611' 'LCS 612' 'LCS 614' 'LCS 630' 'LIP 505' 'LIP 506'
 'LIP 507' 'LIP 508' 'LIP 600' 'LIP 601' 'LIP 602' 'LIP 603' 'LIP 604'
 'LIP 605' 'LIP 606' 'LIP 609' 'LIP 610' 'LIP 612' 'LIP 613' 'LIP 630'
 'LSC 200' 'LSC 300' 'LSC 302' 'LSC 305' 'LSC 400' 'LSC 402' 'LSC 403'
 'LSC 405' 'MAD 500' 'MAD 501' 'MAD 506' 'MAD 600' 'MAD 601' 'MAD 602'
 'MAD 603' 'MAD 606' 'MAD 607' 'MAD 608' 'MAD 610' 'MAD 618' 'MAD 621'
 'MASI 504' 'MASI 600' 'MASI 601' 'MASI 602' 'MASI 603' 'MASI 604'
 'MASI 605' 'MASI 606' 'MASI 610' 'MASI 615' 'MASI 621' 'MAT 100'
 'MAT 100AR' 'MAT 101' 'MAT 105' 'MAT 110' 'MAT 200' 'MAT 210' 'MAT 300'
 'MAT 90' 'MGT 100' 'MGT 100AR' 'MGT 200' 'MGT 202' 'MGT 205' 'MGT 300'
 'MGT 301' 'MGT 302' 'MGT 303' 'MGT 304' 'MGT 305' 'MGT 306' 'MGT 400'
 'MGT 402' 'MGT 403' 'MGT 404' 'MGT 405' 'MGT 406' 'MGT 408' 'MGT 410'
 'MGT 500' 'MGT 501' 'MGT 505' 'MGT 506' 'MGT 507' 'MGT 600' 'MGT 601'
 'MGT 602' 'MGT 604' 'MGT 605' 'MGT 606' 'MGT 608' 'MGT 615' 'MGT 616'
 'MKT 200' 'MKT 200AR' 'MKT 202' 'MKT 203' 'MKT 301' 'MKT 302' 'MKT 304'
 'MKT 305' 'MKT 401' 'MKT 404' 'MKT 405' 'MKT 406' 'MKT 407' 'MKT 500'
 'MKT 600' 'MKT 604' 'MMC 200' 'MMC 200AR' 'MMC 201' 'MMC 206' 'MMC 208'
 'MMC 213' 'MMC 302' 'MMC 303' 'MMC 304' 'MMC 305' 'MMC 306' 'MMC 315'
 'MMC 316' 'MMC 400' 'MMC 401' 'MMC 403' 'MMC 404' 'MMC 412' 'MSM 604'
 'MSM 605' 'MSM 608' 'MSM 611' 'MSS 505' 'MSS 600' 'MSS 601' 'MSS 602'
 'MSS 603' 'MSS 604' 'MSS 605' 'MSS 606' 'MSS 607' 'MSS 608' 'MSS 609'
 'MSS 620' 'MSS 630' 'PHY 100' 'PHY 200' 'PHY 90' 'PRE 300' 'PRE 302'
 'PRE 310' 'PRE 400' 'PRE 402' 'PRE 405' 'PRP 222' 'PRP 225' 'PRP 226'
 'PRP 227' 'PRP 228' 'PRP 310' 'PRP 321' 'PRP 331' 'PRP 332' 'PRP 333'
 'PRP 334' 'PRP 425' 'PRP 426' 'PRP 431' 'PRP 432' 'PRP 435' 'PRP 444'
 'PRP 446' 'RMT 600' 'RMT 600L' 'RMT 601' 'RTV 300' 'RTV 302' 'RTV 310'
 'RTV 401' 'RTV 403' 'RTV 405' 'SED 300' 'SQA 500' 'SS 300' 'SS 301'
 'SS 302' 'SS 303' 'SS 304' 'SS 305' 'SS 306' 'SS 307' 'SS 308' 'SS 309'
 'SS 310' 'SS 311' 'SS 312' 'SS 401' 'SS 402' 'SS 403' 'SS 404' 'SS 410'
 'SS 415' 'SSCM 400' 'SSCM 401' 'SSCM 402' 'SSCM 403' 'SSCM 404'
 'SSDM 400' 'SSDM 401' 'SSDM 402' 'SSDM 403' 'SSDM 404' 'SSEM 400'
 'SSEM 401' 'SSEM 402' 'SSEM 403' 'SSEM 404' 'SSRM 400' 'SSRM 401'
 'SSRM 402' 'SSRM 403' 'SSRM 404' 'TAX 600' 'TOEFL' 'TOL 200']

Unique values in column 'CourseName_x':

['Principles of Accounting' 'Managerial Accounting'
 'Intermediate Accounting I' 'Intermediate Accounting II'
 'Cost Accounting' 'Accounting Information Systems'
 'Planning, Budgeting and Control' 'Auditing' 'Advanced Accounting'
 'Advanced Cost Accounting' 'International Accounting'
 'Accounting for Managers' 'Strategic Cost and Managerial Accounting'
 'Communication Skills in Arabic' 'Arabic Composition'
 'Arabic for Non Arab Speakers' 'مهارات الدراسة والتعلم' 'University Life'
 'الحياة الجامعية' 'Introduction to Anthropology'
 'مقدمة في علم تطور الاجناس' 'Introduction to Sociology' 'Islamic Culture'
 'الثقافة الإسلامية' 'Islamic Civilization'
 'Introduction to Social Sciences' 'مقدمة في العلوم الاجتماعية']

'Women and Politics' 'General Psychology' 'علم النفس العام'
 'Introduction to Philosophy' 'مقدمة في الفلسفة'
 'Introduction to Political Science' 'مقدمة العلوم السياسية'
 'World History' 'تاريخ العالم' 'Middle East History' 'تاريخ الشرق الأوسط'
 'General Statistics' 'الإحصاء العام' 'Introduction to Geography'
 'مبادئ البحث العلمي' 'Introduction to Research' 'مقدمة في علم الجغرافيا'
 'UAE and GCC Society' 'مجتمع الإمارات ودول مجلس التعاون'
 'Natural Science' 'العلوم الطبيعية' 'Environmental Science'
 'Sustainable Energy' 'Human Rights' 'Comprehensive Exit Exam'
 'المدخل إلى الحاسب الآلي' 'Network Essentials'
 'Introduction to Programming' 'System Analysis and Design'
 'Website Design' 'Network Security' 'Database Systems'
 'Multimedia Technology' 'Introduction to Operating Systems'
 'Computer Threats and Risks' 'Internship' 'Computer Ethics'
 'Server Operating System and Network Management'
 'E-Business and E-Commerce Systems' 'Capstone Graduation Project'
 'Software Engineering' 'Mobile Technology' 'Advanced Topics in CIT'
 'Human Computer Interaction' 'Intrusion Forensics'
 'Decision Support Systems' 'Computer Preparatory'
 'Introduction to Computer System' 'Object Oriented Programming'
 'Digital Logic and Design' 'Discrete Mathematics'
 'Data Structures and Algorithms' 'Theory of Computation'
 'Computer Architecture and Organization' 'Numerical Analysis'
 'Modern Programming Languages' 'Capstone Seminar'
 'Artificial Intelligence' 'Reverse Engineering'
 'Introduction to Computer Forensics' 'Distributed Systems'
 'Computer Crime and Investigation Techniques' 'Color Theory and Practice'
 'Basic Design' 'Fundamentals of Graphic Illustration and Digital Imaging'
 'Contemporary Arab Art' 'Drawing I' 'Art History I' 'Design Thinking'
 'Art History II' '3D Design' 'Collaborative Design Studio'
 'Introduction to Animation' '2D Animation' 'History of Animation'
 'Surface Modeling' 'Fundamentals of 3D Animation' 'Motion Graphics'
 'Introduction to Digital Video Production' 'Storyboarding'
 'Character Modeling' 'Character Rigging and Setup' 'Character Animation'
 'Animation Production' 'Negotiation and Conflict Resolution'
 'International Political Economy' 'Theories of International Relations'
 'Microeconomics' 'مبادئ علم الاقتصاد' 'Macroeconomics'
 'Managerial Economics' 'Learning and Behavioral Problems in Education'
 'Counseling Skills' 'Practicum in School Counseling'
 'Foundation of Education'
 'Introduction to Developmental Psychology (Arabic)'
 'Introduction to Educational Psychology'
 'Introduction to Inclusive Education'
 'Research Methods and Statistics in Education'
 'Principles and Methods of Teaching'
 'Teaching in a Multicultural Environment'
 'Education of Exceptional Children'
 'Collaboration in Schools - Working with Teams in the Teaching/Learning Process (Arabic)'
 'Professional Ethics in Education' 'Universal Design for Learning'
 'EmSAT English Preparatory' 'اللغة الإنجليزية' 'Academic Writing'
 'Academic Writing 1' 'Academic Writing 2'
 'اللغة الانجليزية لأغراض خاصة ومهارات الإتصال' 'English Literature'
 'Business Communication' 'American Literature' 'Writing Preparatory'
 'Drawing II' 'Financial Management' 'Corporate Finance & Investment'
 'Money and Banking' 'Islamic Banking and Finance'
 'Management of Financial Institutions'
 'Investment and Financial Strategies'
 'Security Analysis and Portfolio Management'
 'International Financial Management' 'Advanced Financial Management'
 'Islamic Economics & Finance' 'History of Costume'
 'Introduction to the Fashion Industry'
 'Introduction to Textiles, Materials and Processes'
 'Garment Construction' 'Fashion Illustration' 'Fashion Studio I'

'Fashion Management' 'Fashion Studio II' 'Patternmaking'
 'CAD for Fashion Design' 'Fashion Design by Draping'
 'Special Topics in Design' 'Typography I' 'Life Drawing'
 'History of Graphic Design' 'Graphic Design I' 'Typography II'
 'Information Design' 'Communication Design' 'Graphic Design II'
 'Graphic Design III' 'Desktop Publishing' 'Happiness Studies'
 'دراسات في السعادة' 'Introduction to Health Care Industry'
 'Healthcare Law and Regulatory'
 'Economics of Hospital and Healthcare Sector'
 'Ethics and Fraud in Healthcare' 'Healthcare Management'
 'Hospital and Healthcare Information Technology'
 'Clinic Services Management' 'Hospital Support Services Management'
 'Hospital Strategic Management' 'UAE Labour Law and Relation'
 'Human Resource Management' 'Human Resource Information System'
 'HR Analytics' 'HRM in the Public Sector'
 'Staffing and Techniques for Employee Selection'
 'Employees Training and Development'
 'Compensation and Benefits Management' 'Special Topics in HR'
 'Strategic Human Resource Management' 'IELTS Preparatory Course'
 'Principles of Advertising' 'Global Media and Marketing'
 'Digital Media Management' 'Media Planning' 'Drafting Lab'
 'ID Principles' 'Studio I – Introduction to Interior Design' 'CAD'
 'History of Interior Design' 'Textiles for Interior Design'
 'Materials and Finishes' 'Studio II – Residential'
 'Building Systems and Codes' 'Materials and Resources' 'Advanced CAD'
 'Studio III – Office Design' 'Building Information Modeling'
 'Studio III – Commercial' 'Lighting for Interior Design'
 'Furniture Design' 'Time Based Media for Interior Design'
 'Interior Design Profession' 'Studio IV – Commercial' 'Detailing'
 'ELI – Intermediate Course' 'Pre-Internship'
 'Innovation and Entrepreneurship'
 'Introduction to International Relations' 'المدخل إلى العلاقات الدولية'
 'Writing for Diplomacy' 'Introduction to Diplomacy'
 'Public International Law' 'Intercultural Communication'
 'Introduction to Insurance and Risk Management' 'Insurance Law'
 'Life and Health Insurance'
 'Personal Property-Liability Insurance and Risk Mgt'
 'Risk Analysis and Evaluation'
 'Insurance Operations, Regulations, and Statutory Accounting'
 'Commercial Property Risk Management and Insurance'
 'Commercial Casualty Risk Mgt & Insurance'
 'Alternative Risk Financing and Capital Markets'
 'Risk Management Strategy' 'Database Administration'
 'IT Project Management'
 'Information and Knowledge Management in the 21st Century'
 'Knowledge Assessment and Evaluation'
 'Knowledge Discovery and Findability' 'Collaboration and Social Networks'
 'Organizational Culture and Impacting Change' 'Research Methodology'
 'Enterprise Wide Application and Portfolio Project Management'
 'Knowledge Management Systems and Implementation\xa0'
 'Foundations of Document and Record Management' 'Capstone Project'
 'Thesis' 'اتفاق التحكيم' 'إجراءات التحكيم'
 'القانون الواجب التطبيق على التحكيم وحكم المحكم'
 'حكم التحكيم (إصداره، بطلانه و تنفيذه)' 'دور القضاء في التحكيم'
 'التحكيم في عقود المقاولات (باللغة الانجليزية)'
 'الإتفاقيات الدولية و مؤسسات و مراكز التحكيم'
 'التحكيم في منازعات الإستثمار (باللغة الانجليزية)' 'الرسالة'
 'المدخل إلى علم القانون' 'القانون الدستوري والنظم السياسية'
 'المدخل إلى الشريعة الإسلامية' 'المدخل الى البحث العلمي'
 'دراسات قانونية في اللغة الإنكليزية' 'قانون العقوبات/القسم العام (1)'
 'مصادر الالتزام (1)'
 'قانون العقوبات القسم العام (2) (النظرية العامة للجزاء الجنائي)'
 'مبادئ القانون التجاري' 'الأحوال الشخصية 1 (أحكام الزواج والطلاق)'
 'مصادر الالتزام (2)' 'القانون الدولي العام باللغة الإنجليزية'

'القانون الإداري' 'علم الاجرام والعقاب'
'المالية العامة والتشريعات الاقتصادية'
'حقوق الإنسان في المواثيق الدولية والشريعة الإسلامية'
'الشركات التجارية و الإفلاس' 'أحكام الالتزام'
'قانون العقوبات/ القسم الخاص (1) جرائم الأشخاص والأموال والجرائم المضرة بالمصلحة العامة'
'قانون العقوبات/ القسم الخاص (2) (الجرائم المستحدثة)'
'قانون الإجراءات المدنية' 'الأحوال الشخصية 2 (الوصايا والميراث والوقف)
'العقود المسماة' 'الحقوق العينية' 'القانون البحري و الجوي'
'الأعمال المصرفية والأوراق التجارية' 'أحكام الإثبات' 'الملكية الفكرية'
'قانون الإجراءات الجزائية' 'القانون الدولي الخاص (1)'
'القانون الدولي الخاص (2)' 'علم أصول الفقه'
'التحكيم في المنازعات المدنية والتجارية' 'قانون حماية البيئة'
'قانون التنفيذ' 'عقود التجارة الدولية (باللغة الانجليزية)'
'قانون العمل والضمان الاجتماعي' 'التدريب العملي (الداخلي و الخارجي)'
'قانون حماية المستهلك' 'الاقتصاد الدولي (باللغة الانجليزية)'
'المنظمات الدولية (باللغة الانجليزية)'
'بحث التخرج موعد طرح المساق: الفصلان الأخيران قبل التخرج'
'فن الخطابة والمرافعة' 'النقود والبنوك باللغة الانجليزية'
'مصادر الإلتزام(1)' 'قانون العقوبات' 'أحكام الإثبات' 'مصادر الإلتزام 1'
'قانون الاجراءات المدنية' 'المدخل لعلم القانون' 'مصادر الإلتزام 2'
'International Organization' 'مبادئ القانون التجاري' 'قانون العمل'
'المدخل لدراسة القانون الرياضي' 'Commercial Law' 'Constitutional Law'
'طرق فض المنازعات الرياضية (50٪ باللغة الإنكليزية)'
'النظام القانوني للاعتراف الرياضي' 'التنظيم الرياضي' 'التدريب العملي'
'مصطلحات قانونية رياضية باللغة الإنجليزية' 'المسؤولية القانونية الرياضية'
'العقود التجارية في المجال الرياضي' 'النظام القانوني للمنشطات الرياضية'
'موضوعات قانونية رياضية معاصرة (التحكيم- أمن الملاعب....)'
'الإدارة الرياضية' 'المواثيق الرياضية' 'الرعاية الرياضية' 'مشروع البحث'
'النظام القانوني للمؤسسات والمنظمات الرياضية' 'قانون الإجراءات الجنائية'
'التشريع الجنائي الإسلامي'
'حقوق الإنسان في القانون الجنائي (50٪ باللغة الإنكليزية)'
'القانون الدولي الجنائي (50٪ باللغة الإنكليزية)' 'علم الإجرام'
'علم العقاب مع التعمق' 'بدائل الدعوى الجنائية'
'التشريعات الجنائية الخاصة' 'الإثبات الجنائي'
'علماء الإجرام والعقاب (50٪ باللغة الإنكليزية)' 'مصادر الإلتزام(2)'
'القانون الدولي الخاص (تنازع القوانين)' 'قانون العقوبات (القسم العام 1)'
'براءة الاختراع و الرسوم و التصميمات الصناعية'
'العلامات والاسماء التجارية والمؤشرات الجغرافية (25٪ باللغة الانجليزية)'
'حقوق المؤلف و الحقوق المجاورة'
'حماية حقوق الملكية الفكرية عبر الوسائط الرقمية'
'تنازع القوانين في مسائل الملكية الفكرية'
'انفاذ وتسوية منازعات الملكية الفكرية (25٪ باللغة الانجليزية)'
'الملكية الفكرية في المجالين الطبي و الصيدلي'
'حقوق الملكية الفكرية و قانون المنافسة'
'حماية المعلومات غير مفصح عنها و الأسرار التجارية' 'حقوق المؤلف'
'Supply Chain Operations' 'الحقوق المجاورة'
'Service Operations Management' 'Transportation & Physical Distribution'
'Purchasing and Supply Chain Management'
'Inventory and Material Management'
'Warehouse Management and Material Handling'
'SCLM Technologies and Information Systems'
'Global Logistics and Supply Chain Management'
'International Relation since 1945' 'The World Today'
'Remedial Foreign Policy Analysis' 'The Art of Diplomatic Negotiation'
'Diplomacy in the Gulf State' 'Research and Writing in Diplomacy'
'Diplomacy and Security' 'Regional and International Organization'
'Economic Diplomacy' 'History of Global Politics'
'Remedial Leadership in Security, Strategy and Intelligence'
'Intelligence and Security' 'Issues in Intelligence'
'Intelligence Analysis' 'Security Strategic Works and Intelligence'
'Contemporary Conflicts and Security Chanllenges'
'Research Methods in Intelligence and Security Studies'

'Homeland Security' 'Critical Infrastructure Protection'
 'Leadership and Security, Strategy and Intelligence'
 'General Mathematics' 'رياضيات عامة' 'Math of Business' 'Calculus'
 'Calculus I' 'Calculus II' 'Linear Algebra' 'Calculus III'
 'Mathematics Preparatory' 'Principles of Management' 'مبادئ الإدارة'
 'Management Information System' 'Business Statistics'
 'Organizational Behavior' 'Production and Operations Management'
 'Business Law'
 'Sustainable Business Practices in the GCC and MENA Region (formerly Business Practice in UAE and GCC)'
 'Business Analytics and Quantitative Analysis (formerly Quantitative Analysis)'
 'Small Business Management' 'Negotiation Skills'
 'Managing in the Service Environment' 'International Business'
 'Project Management' 'Entrepreneurship Management'
 'Organization Theory and Design' 'Strategic Management'
 'Total Quality Management' 'Leadership and Organizational Creativity'
 'Computer Applications in Management'
 'Management and Organizational Behavior' 'Managing and Leading People'
 'Statistical Tools and Techniques'
 'Strategic Management : Theory and Practice'
 'Leadership and Change Management' 'Management of Information Technology'
 'Seminar in Contemporary Topics in Management' 'Organization Development'
 'Principles of Marketing' 'مبادئ التسويق' 'Consumer Behavior'
 'Brand Management' 'Advertising and Promotion'
 'Marketing in the Digital Environment'
 'Introduction to Marketing Analytics'
 'Integrated Marketing Communications' 'Marketing Research'
 'Special Topics in Marketing' 'Strategic Marketing'
 'Sustainability and Responsible Marketing' 'Global Marketing'
 'Marketing Concepts' 'Marketing and International Business'
 'Digital Marketing and Social Media' 'Introduction to Mass Communication'
 'Introduction to Social Media' 'Fundamentals of News'
 'Public Opinion and Propaganda' 'وسائل الاتصال في دولة الإمارات'
 'Writing for the Mass Media' 'Organizational Communication'
 'Introduction to Multimedia Production' 'مناهج البحث العلمي' 'فن الإقناع'
 'Mass Communication Theory and Research' 'Photojournalism'
 'Editing for the Mass Media' 'Media Law and Ethics' 'الرأي العام'
 'Sport Facilities and Events Management' 'Sports Law and Ethics'
 'Sport Consumer Behavior' 'Sport Tourism'
 'Remedial Terrorism and Organized Crime'
 'International Security and Strategic Studies'
 'Crisis Management Communication'
 'Research in Security and Strategic Studies'
 'Sustainability, Growth and National Security'
 'Special Topics in the Middle East Security and Strategy'
 'Diplomacy and the Use of Forces'
 'Contemporary Issues in Foreign Politics'
 'International Law and Security' 'Physics' 'Physics 2'
 'Physics Preparatory' 'Public Speaking' 'Writing in Public Relation'
 'Media Strategies and Implementation' 'مبادئ العلاقات العامة'
 'فنون التحرير الصحفي (عملي)' 'العلاقات العامة والإعلام الجديد'
 'البروتوكول والإتيكيت: (عملي)' 'التسويق الاجتماعي'
 'Capstone Seminar (Arabic)' 'الاتصال التنظيمي'
 'العلاقات العامة الاستراتيجية' 'الكتابة للعلاقات العامة (عملي)'
 'التصوير الرقمي: (عملي)' 'اتصال إدارة الأزمات' 'التدريب العملي الميداني'
 'إنتاج المواد المصورة والمسموعة والمرئية (تطبيقي عملي)'
 'العلاقات العامة الدولية' 'تشريعات الإعلام وأخلاقياته'
 'الاتصالات التسويقية المتكاملة'
 "' موضوعات خاصة في العلاقات العامة : "يتناول الموضوعات والاتجاهات الجديدة في مجال العلاقات العامة"
 'Business Research Methods' 'مشروع التخرج في العلاقات العامة'
 'Research Methods in Security and Global Studies'
 'Writing for Radio and TV' 'Producing and Planning'
 'Broadcast and Digital Media Advertising'

```
'Television and Digital Video Production'
'Diagnosis and Assessment in Special Education (Arabic) (Practical)_x000D_\n'
'Statistics and Quantitative Analysis'
'Theoretical Basis of Security & Strategic Studies'
'Basic of Social Science Research' 'Introduction to Strategic Thought'
'International Law' 'Security and Environmental Politics'
'Geopolitics in International Relations' 'Comparative Politics'
'Weapons of Mass Destruction' 'National Security and Arms Race'
'Communication and Diplomacy'
'Special Topics in Security and Strategic Studies'
'International Conflict Management' 'Terrorism'
'Defense Policies of Major Powers' 'The Emerging World Powers'
'Crisis Management and Business Continuation'
'Intergovernmental Crisis and Security Management'
'Crisis Psychology and Mental Health' 'Disaster Management'
'Catastrophe Sciences' 'Economics of Natural Disasters'
'Incident Response and Disaster Recovery' 'Emergency Management'
'Community Preparedness' 'Response and Recovery'
'Prevention and Planning' 'Risk Management' 'Risk and Mitigation'
'Enterprise Risk Management' 'Risk Analysis'
'Contemporary Topics in Taxation' 'TOEFL Preparatory Course'
'Tolerance and Diversity']
```

Unique values in column 'QuestionText':

```
['CLO : Evaluate financial and managerial accounting both from internal and external per
spectives.'
'The coursework and assignments are challenging'
'The course is relevant to my program of study' ...
'CLO : analyze the relationship between diversity and tolerance in a way that enables t
hem to create initiatives that promote tolerance in various settings'
'CLO : compare and contrast the different debates on the importance of tolerance at the
individual and group levels'
'CLO : describe the fundamental concepts of diversity and tolerance in various context
s']
```

Unique values in column 'QuestionType':

```
['Rating' 'User Comment']
```

Unique values in column 'ParticipantResponse':

```
['Disagree' 'Strongly Disagree' 'Strongly Agree' ... '.']
'Dr. Dalal was so helpful and understanding'
'A great course for those who needs to learn how to be tolerant and diverse in specific
areas. It will teach you a lot about the people around you.']
```

Unique values in column 'College':

```
['College of Business Administration' 'College of Education' nan
'College of Security and Global Studies'
'College of Computer Information Technology' 'College of Design'
'College of Law' 'College of Media and Mass Communication'
'Training Courses']
```

Unique values in column 'DegreeName':

```
['Bachelor Degree' 'Master Degree' nan]
```

Unique values in column 'FieldOfStudy':

```
['BA' 'UNSPECIFIED' 'EDUCATION' nan 'IT' 'ARTS_DESIGN' 'MASS_COMM_PR'
'LAW_SHARIA' 'HUMANITIES_SOCIAL_SC']
```

handle Null values in ParticipantResponse

```
In [17]: #handle Null values in ParticipantResponse
merged_df['ParticipantResponse'].fillna('No Response', inplace=True)
```

handle Null values in College

```
In [18]: # Define a function to extract the word after "of" in Student Program
def extract_word_after_of_or_in(program):
    if pd.notnull(program):
        if 'of' in program:
            return program.split('of', 1)[1].strip()
        elif 'in' in program:
            return program.split('in', 1)[1].strip()
        return ''
merged_df.loc[merged_df['College'].isnull(), 'College'] = 'College of ' + merged_df.loc
```

impute null values in DegreeName

```
In [19]: # Define a function to extract the first word from Student Program
def extract_first_word(program):
    if pd.notnull(program):
        # Split the program string into words
        words = program.split()
        if len(words) > 1 and words[0].lower() == 'professional':
            # If the first word is 'Professional', return the second word
            return words[1] + ' Degree'
        else:
            # Otherwise, return the first word with 'Degree'
            return words[0] + ' Degree'
    return ''

# Apply the function to extract the first word# Apply the function to fill null values
merged_df.loc[merged_df['DegreeName'].isnull(), 'DegreeName'] = merged_df.loc[merged_df
```

```
In [20]: merged_df.College.unique()
```

```
Out[20]: array(['College of Business Administration', 'College of Education',
                'College of Law', 'College of Public Relations',
                'College of Security and Global Studies',
                'College of Computer Information Technology',
                'College of Science in Computer Science',
                'College of Science in Information Technology Management',
                'College of Arts in Media and Mass Communication',
                'College of Design',
                'College of Arts in Security and Strategic Studies',
                'College of Science in Design - Graphic Design',
                'College of Science in Design - Interior Design',
                'College of Science in Design - Digital Animation',
                'College of Science in Design - Fashion Design',
                'College of Media and Mass Communication', 'Training Courses',
                'College of Knowledge Management', 'College of Arbitration',
```

```

'College of Professional Master in Sports Law',
'College of Arts in Diplomacy',
'College of Security Studies and Information Analysis'],
dtype=object)

```

```

In [21]: merged_df.FieldOfStudy.unique()

```

```

Out[21]: array(['BA', 'UNSPECIFIED', 'EDUCATION', nan, 'IT', 'ARTS_DESIGN',
'MASS_COMM_PR', 'LAW_SHARIA', 'HUMANITIES_SOCIAL_SC'], dtype=object)

```

impute null values in FieldOfStudy

```

In [22]: #Create a dictionary mapping between 'College' and 'FieldOfStudy'
college_to_field_of_study = {
    'College of Business Administration': 'BA',
    'College of Education': 'EDUCATION',
    'College of Law': 'LAW_SHARIA',
    'College of Public Relations': 'Public Relations',
    'College of Security and Global Studies': 'HUMANITIES_SOCIAL_SC',
    'College of Computer Information Technology': 'IT',
    'College of Science in Computer Science': 'Computer Science',
    'College of Science in Information Technology Management': 'Information Technology',
    'College of Arts in Media and Mass Communication': 'MASS_COMM_PR',
    'College of Design': 'ARTS_DESIGN',
    'College of Arts in Security and Strategic Studies': 'Security and Strategic Studies',
    'College of Science in Design - Graphic Design': 'Graphic Design',
    'College of Science in Design - Interior Design': 'Interior Design',
    'College of Science in Design - Digital Animation': 'Digital Animation',
    'College of Science in Design - Fashion Design': 'Fashion Design',
    'College of Media and Mass Communication': 'MASS_COMM_PRzh',
    'Training Courses': 'Training Courses',
    'College of Knowledge Management': 'Knowledge Management',
    'College of Arbitration': 'Arbitration',
    'College of Professional Master in Sports Law': 'Sports Law',
    'College of Arts in Diplomacy': 'Diplomacy',
    'College of Security Studies and Information Analysis': 'Security Studies and Information Analysis'
}

# Step 3: Use the dictionary to impute null values in 'FieldOfStudy'
# Iterate through the DataFrame
for index, row in merged_df.iterrows():
    college = row['College']
    field_of_study = row['FieldOfStudy']

    # Check if 'FieldOfStudy' is null
    if pd.isnull(field_of_study):
        # Look up the corresponding 'FieldOfStudy' value in the dictionary using 'College'
        if college in college_to_field_of_study:
            merged_df.at[index, 'FieldOfStudy'] = college_to_field_of_study[college]

```

```

In [23]: # Check for missing values in the dataframe
missing_values = merged_df.isnull().sum()
missing_values

```

```

Out[23]: AcademicYear      0
SemesterName              0

```

```

ParticipantID      0
DegreeLevel        0
Gender             0
Nationality         0
StudentProgram     0
CourseCode         0
CourseName_x       0
QuestionText       0
QuestionType       0
ParticipantResponse 0
College            0
DegreeName         0
FieldOfStudy       0
dtype: int64

```

```

In [25]: # Save the merged_df DataFrame to an output file
outputprocessing_filename = 'C:\\Users\\asmaa\\OneDrive\\Desktop\\americanuniversity\\m
merged_df.to_excel(outputprocessing_filename, index=False)

```

```

In [26]: merged_df.ParticipantResponse.head(20)

```

```

Out[26]: 0          Disagree
1    Strongly Disagree
2    Strongly Disagree
3    Strongly Disagree
4    Strongly Disagree
5    Strongly Disagree
6    Strongly Disagree
7          Disagree
8    Strongly Agree
9    Strongly Agree
10   Strongly Disagree
11   Strongly Disagree
12   Strongly Disagree
13         No Response
14   Strongly Agree
15   Strongly Agree
16   Strongly Agree
17          Agree
18          Agree
19          Agree
Name: ParticipantResponse, dtype: object

```

text preprocessing for ParticipantResponse

```

In [27]: # Function for text preprocessing for ParticipantResponse
def preprocess_text(text):
    # Convert text to lowercase
    text = text.lower()

    # Remove punctuation and numbers
    translator = str.maketrans('', '', string.punctuation + string.digits)
    text = text.translate(translator)

    # Tokenization (splitting text into words)
    tokens = word_tokenize(text)

    # Remove stop words

```

```

stop_words = set(stopwords.words('english'))
tokens = [word for word in tokens if word not in stop_words]

# Lemmatization
lemmatizer = WordNetLemmatizer()
tokens = [lemmatizer.lemmatize(word) for word in tokens]

# Rejoin tokens into a single string
cleaned_text = ' '.join(tokens)

return cleaned_text

# Apply preprocessing to the DataFrame
merged_df['ParticipantResponse'] = merged_df['ParticipantResponse'].apply(preprocess_te

# Display the cleaned data
print(merged_df['ParticipantResponse'])

```

```

0          disagree
1    strongly disagree
2    strongly disagree
3    strongly disagree
4    strongly disagree
...
180964      response
180965          agree
180966    strongly agree
180967    strongly agree
180968    strongly agree
Name: ParticipantResponse, Length: 179746, dtype: object

```

```

In [28]: merged_df['CourseName_x'] = merged_df['CourseName_x'].apply(preprocess_text)
merged_df['CourseName_x']

```

```

Out[28]: 0      principle accounting
1      principle accounting
2      principle accounting
3      principle accounting
4      principle accounting
...
180964    tolerance diversity
180965    tolerance diversity
180966    tolerance diversity
180967    tolerance diversity
180968    tolerance diversity
Name: CourseName_x, Length: 179746, dtype: object

```

```

In [29]: merged_df['QuestionText'] = merged_df['QuestionText'].apply(preprocess_text)
merged_df['QuestionText']

```

```

Out[29]: 0      clo evaluate financial managerial accounting i...
1      coursework assignment challenging
2      course relevant program study
3      course learning material available portal
4      faculty us additional resource textbook enhanc...
...
180964    recommendation suggestion course could improve...
180965    clo analyze relationship diversity tolerance w...
180966    faculty discussed course syllabus beginning se...
180967    coursework assignment challenging

```

```
180968                                course relevant program study
Name: QuestionText, Length: 179746, dtype: object
```

```
In [30]: merged_df.columns
```

```
Out[30]: Index(['AcademicYear', 'SemesterName', 'ParticipantID', 'DegreeLevel',
               'Gender', 'Nationality', 'StudentProgram', 'CourseCode', 'CourseName_x',
               'QuestionText', 'QuestionType', 'ParticipantResponse', 'College',
               'DegreeName', 'FieldOfStudy'],
              dtype='object')
```

Sentiment Analysis

```
In [31]: from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
import matplotlib.pyplot as plt

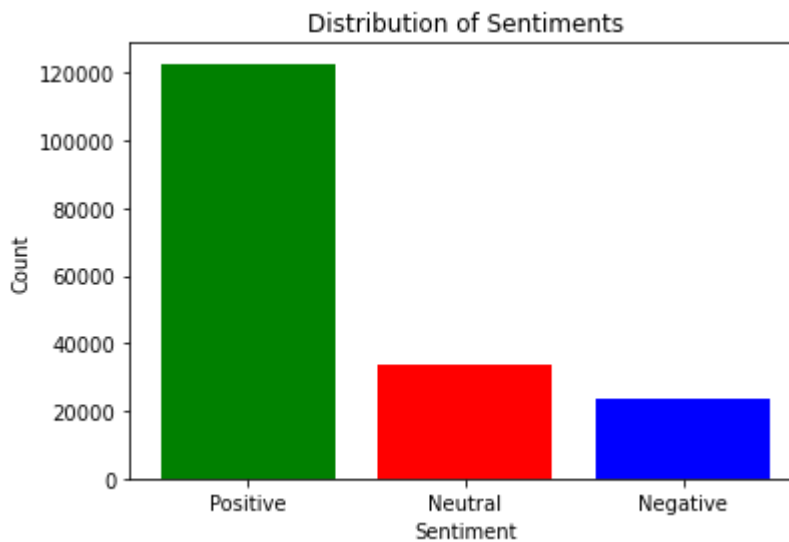
# Create a SentimentIntensityAnalyzer object
analyzer = SentimentIntensityAnalyzer()

# Perform sentiment analysis and store the results in a new column
merged_df['Sentiment'] = merged_df['ParticipantResponse'].apply(lambda x: analyzer.pola

# Categorize sentiment into positive, negative, or neutral
def categorize_sentiment(score):
    if score > 0.05:
        return 'Positive'
    elif score < -0.05:
        return 'Negative'
    else:
        return 'Neutral'

merged_df['SentimentLabel'] = merged_df['Sentiment'].apply(categorize_sentiment)

# Create a bar plot to show the distribution of sentiments
sentiment_counts = merged_df['SentimentLabel'].value_counts()
plt.bar(sentiment_counts.index, sentiment_counts.values, color=['green', 'red', 'blue'])
plt.xlabel('Sentiment')
plt.ylabel('Count')
plt.title('Distribution of Sentiments')
plt.show()
```



In [32]: `pip install gensim`

Requirement already satisfied: gensim in c:\users\asmaa\anaconda3\lib\site-packages (4.3.2)
 Requirement already satisfied: smart-open>=1.8.1 in c:\users\asmaa\anaconda3\lib\site-packages (from gensim) (6.3.0)
 Requirement already satisfied: numpy>=1.18.5 in c:\users\asmaa\anaconda3\lib\site-packages (from gensim) (1.19.5)
 Requirement already satisfied: scipy>=1.7.0 in c:\users\asmaa\anaconda3\lib\site-packages (from gensim) (1.10.1)
 Note: you may need to restart the kernel to use updated packages.

In [33]: `import gensim
 from gensim import corpora
 from gensim.models import LdaModel
 import pprint`

Topic Modeling:

In [34]: `participant_response_data = merged_df['ParticipantResponse'].values.tolist()

Tokenize the text and create a List of words for each document
 tokenized_participant_response = [text.split() for text in participant_response_data]

Create a dictionary and corpus for the "ParticipantResponse" column
 dictionary_participant_response = corpora.Dictionary(tokenized_participant_response)
 corpus_participant_response = [dictionary_participant_response.doc2bow(tokens) for token`

In [35]: `# Define the number of topics (you can adjust this)
 num_topics = 5

Train an LDA model for the "ParticipantResponse" column
 lda_model_participant_response = LdaModel(corpus_participant_response, num_topics=num_t`


```
In [36]: # Print topics and representative keywords for the "ParticipantResponse" column
print("Topics in ParticipantResponse:")
pprint.pprint(lda_model_participant_response.print_topics())
```

Topics in ParticipantResponse:

```
[(0,
 '0.656*"response" + 0.016*"على"*0.011 + "من"*0.014 + "في"*0.015 + "و" + '
 '0.008*"0.005 + "كل"*0.005 + "المادة"*thanks" + 0.004*"الدكتور" + '
 '0.004*"الطلاب"''),
 (1,
 '0.040*"course" + 0.024*"0.022 + "لا"*good" + 0.021*"dr" + 0.018*"great" + '
 '0.013*"best" + 0.013*"thank" + 0.011*"doctor" + 0.009*"instructor" + '
 '0.009*"u"''),
 (2,
 '0.937*"neutral" + 0.000*"0.000 + "لا يوجد"*0.000 + "ه"*nn" + 0.000*"n1" + '
 '0.000*"hahshshha" + 0.000*"hzhzhahz" + 0.000*"sss" + 0.000*"aa" + '
 '0.000*"gggg"''),
 (3,
 '0.566*"disagree" + 0.408*"strongly" + 0.000*"لا يوجد"*0.000 + "ه" + '
 '0.000*"nn" + 0.000*"n1" + 0.000*"sss" + 0.000*"hahshshha" + '
 '0.000*"hzhzhahz" + 0.000*"aa"''),
 (4,
 '0.573*"agree" + 0.422*"strongly" + 0.000*"nope" + 0.000*"ه" + '
 '0.000*"0.000 + "لا يوجد"*nn" + 0.000*"n1" + 0.000*"sss" + 0.000*"hahshshha" '
 '+ 0.000*"hzhzhahz"'')]
```

Opinion Mining:

```
In [37]: from textblob import TextBlob

# Define a function for opinion mining
def perform_opinion_mining(text):
    # Create a TextBlob object to analyze sentiment
    analysis = TextBlob(text)

    # Determine sentiment polarity (-1 for negative, 1 for positive)
    sentiment_polarity = analysis.sentiment.polarity

    # Categorize sentiment based on polarity
    if sentiment_polarity > 0:
        return 'Positive'
    elif sentiment_polarity < 0:
        return 'Negative'
    else:
        return 'Neutral'

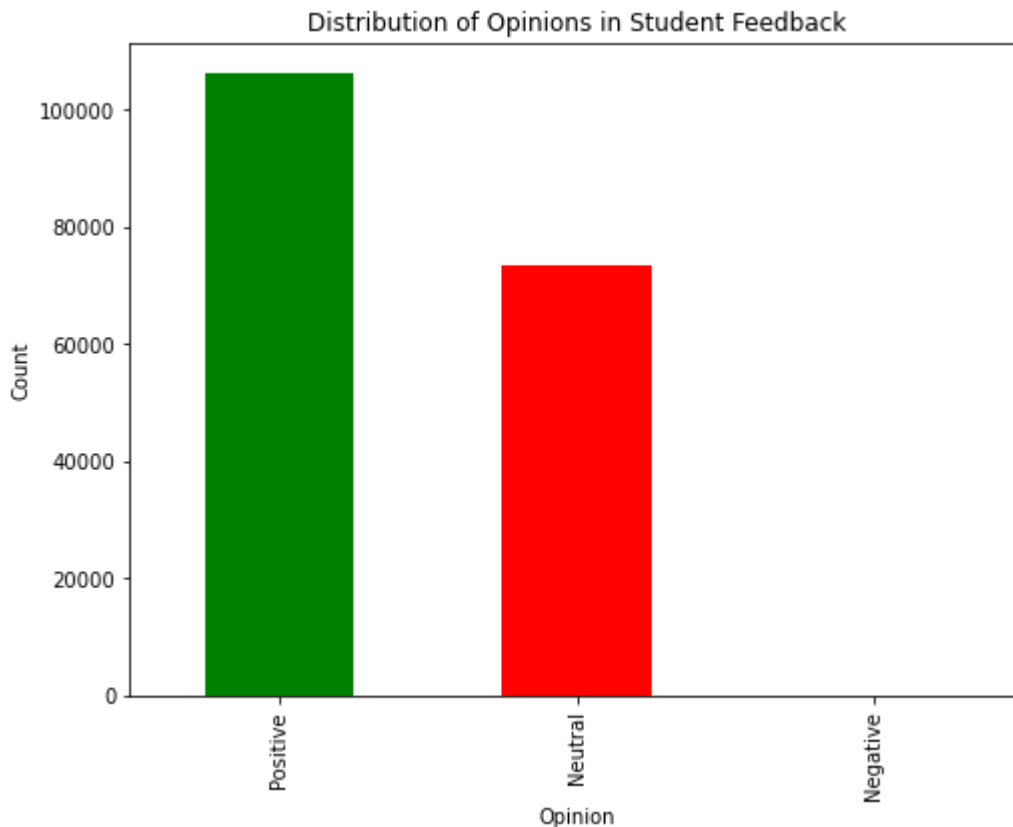
# Apply opinion mining to the 'ParticipantResponse' column and store the results in a new column
merged_df['Opinion'] = merged_df['ParticipantResponse'].apply(perform_opinion_mining)

# Visualize the distribution of opinions
opinion_counts = merged_df['Opinion'].value_counts()

import matplotlib.pyplot as plt

plt.figure(figsize=(8, 6))
opinion_counts.plot(kind='bar', color=['green', 'red', 'blue'])
plt.xlabel('Opinion')
plt.ylabel('Count')
```

```
plt.title('Distribution of Opinions in Student Feedback')  
plt.show()
```



Insights and Recommendations:

Positive (120,000): The majority of student feedback appears to be positive, indicating that many students have positive experiences with their courses. This is a significant strength and suggests that the institution is doing well in providing a positive learning environment.

Neutral (4,000): A portion of feedback is neutral, which may imply that some students have mixed feelings or are neutral about their experiences. Further investigation is needed to understand the reasons behind this sentiment and gather more detailed feedback.

Negative (2,000): Although a smaller portion, the negative feedback should not be ignored. It's crucial to analyze these comments carefully to identify specific issues or concerns raised by students. Addressing these concerns can lead to improvements in courses and student satisfaction.

Based on this sentiment distribution, it's recommended to:

Conduct in-depth analysis of the negative feedback to pinpoint specific issues and areas for improvement. Explore the neutral feedback to understand the reasons behind students' mixed sentiments. Continue to leverage the positive feedback as a valuable indicator of the institution's strengths in course delivery. These insights can guide actions and improvements in course offerings and overall student satisfaction.

