Analysing Student Feedback Using Artificial Intelligence techniques in Python

1. Data Preprocessing:

I utilized the "xlrd" library, installed via "pip install xlrd," to extract data from an Excel spreadsheet.

The "ParticipantResponse" column presented a mixture of responses, including ratings and comments. To enhance clarity and facilitate sentiment analysis, I divided this column into two distinct columns: "Rating" and "Comments."

Additionally, several entries in the "ParticipantResponse" column contained null values. I addressed this issue by filling all null values with empty strings.

Notably, a substantial number of responses in the "Participant" column commenced with the word "No," such as 'No thanks,' 'No recommendations,' 'No,' 'No, everything is good,' and so forth. Recognizing that these comments beginning with "no" might introduce a negative bias in sentiment analysis, I diligently removed the word "no" from the beginning of each comment.

2. Sentiment Analysis

Sentiment Analysis was conducted using the two distinct columns: "Rating" and "Comments," as outlined previously.

For the "Rating" column, sentiments were discerned based on the provided ratings, ranging from "Strongly Agree" to "Strongly Disagree."

However, it's important to note that the sentiment analysis method was exclusively applied to the "Comments" column. Below, you will find an elaboration of the method employed for sentiment analysis:

2.1. VADER SentimentIntensityAnalyser

VADER (Valence Aware Dictionary and sentiment Reasoner) is a sentiment analysis tool designed to assess the sentiment or emotional tone expressed in text. It's particularly well-suited for analyzing text data from social media, customer reviews, and other forms of informal communication.

2.2. Scoring Words:

When you input a text or sentence into VADER, it breaks the text down into individual words and phrases. It then assigns sentiment scores to these words based on their presence in the lexicon.

2.3. **Sentiment Categories:** VADER classifies sentiment into four categories:

Positive: A compound score above a certain threshold below 0.0 (typically 0.05).

Negative: A compound score below a negative threshold above 0.0

Neutral: A deliberate compound score which is unique value for not available comments

Compound: A single value representing the overall sentiment of the text.

The overall sentiments were labelled to each record with analysing sentiment score and the rating columns.

2.4. **Limitations:**

While VADER is a useful tool, it has limitations. It works best for English text, may struggle with complex sentence structures, and is less accurate with formal language. Additionally, it's important to consider context and domain-specific nuances when interpreting results.

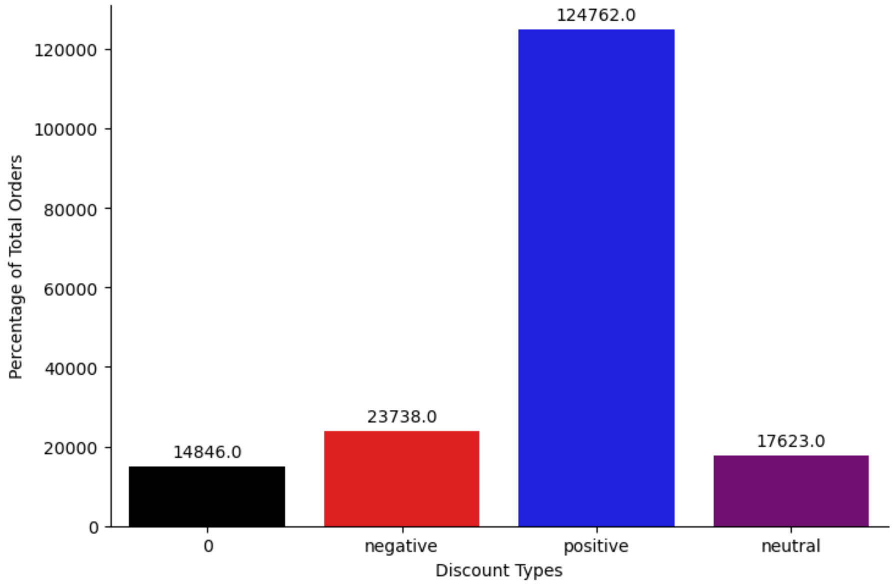
2.5. Sentiment Analysis result interpretation

After labeling the predicted sentiments using pandas skills, the overall sentiment results appear satisfactory when assessed visually.

It's worth highlighting that, as depicted in the provided example, the comment "nothing all good" was assigned a negative score intentionally left unchanged. This was done to reinforce the earlier assertion that negative words at the beginning of a sentence can indeed have a detrimental impact on the scoring outcome.

	UserComment	sentiment_score
1438	in a way the prof was so strict and following ...	-0.83
1565	very very hard course	-0.24
1733	stop teaching at the level of a master. and ma...	-0.30
1777	he doesn't explain well	-0.21
2635	nothing all good	-0.34
...	...	...
173838	the chairs and tables in the rows are uncomfor...	-0.38
175998	i have a point on the instructor that in the v...	-0.16
176382	the chairs and tables in the rows are uncomfor...	-0.38
177630	the dr are dependent on youtu.be and many fram...	-0.10
179833	the couse timing is very long what makes us bo...	-0.27

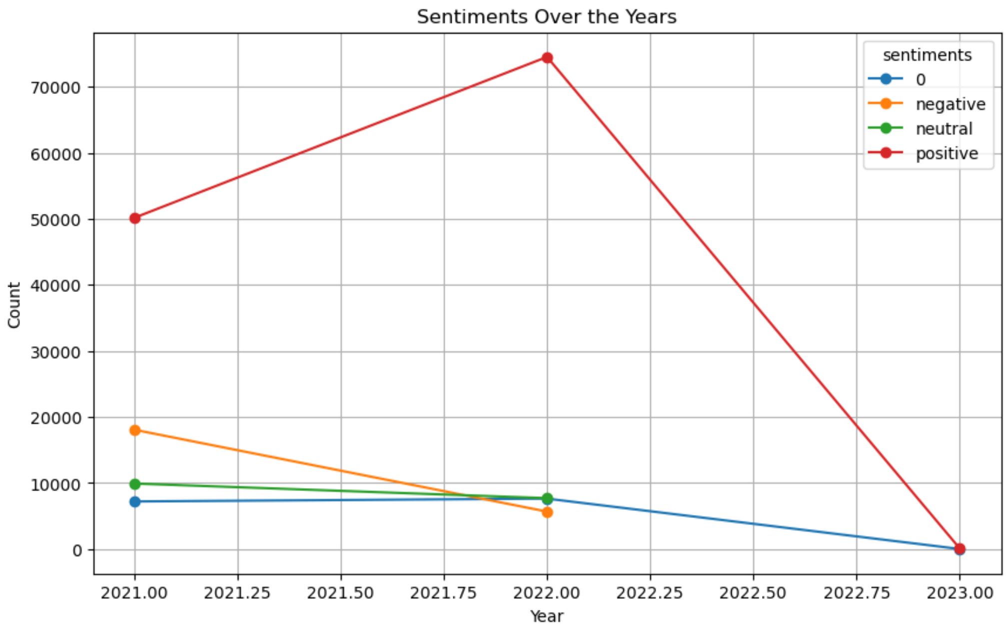
2.6. Sentiment Analysis with data Visualisation



You can see from the above visual analysis, the approx.
Positive sentiments - 70% of total feedback
Negative sentiments – 13% of total feedback
Neutral sentiments – 10% of total feedback
No response – 7%

Graph below shows the number of all types of sentiment.

A significant drop in positive sentiments from the beginning of 2022.



3. Topic Modelling - Latent Dirichlet Allocation (LDA):

LDA is used as a probabilistic model to identify topics within a corpus of documents. It operates on the premise that documents are mixtures of topics, and topics are mixtures of words. The generative process behind LDA involves the Dirichlet distribution. LDA is an unsupervised learning technique, meaning it doesn't require labeled data. It identifies topics solely based on the patterns of word co-occurrence within the text.

After implementing the model there were **1954** words stored in vocabulary. There are some random words as per below: -

Random Words out of 1954 total collected words.

محمد الأستاذ حياتي طرح process وقد lost والحضارات لاتوجد theory	remotely adobe lost explained lab tamara implementations wants ويتم واسلوب	health انا 3d ease examples تحياتي كلاس فهي based يسر	تعرف uncomfortable اثناء معها تعلمت نو errors عدد recommendation تعلمت

3.1. Top words per topic:

To identify these top words, LDA assigns probabilities to each word in the vocabulary for each topic. The words with the highest probabilities within a topic are considered the most representative. The "top_word_indices" are simply the positions or indices of these top words within the vocabulary.

List of top 15 words for 4 Topics:

```
THE TOP 15 WORDS FOR TOPIC #0
['lot', 'courses', 'professor', 'clear', 'instructor', 'thanks', 'class', 'real
ly', 'amazing', 'students', 'doctor', 'best', 'great', 'dr', 'course']


THE TOP 15 WORDS FOR TOPIC #1
['الوسائل', 'survey', 'par', 'hello', 'shahad', 'thankyou', 'لاشي', 'noooo', 'بشده
اوافق'', 'kk', 'thankz', 'nope', 'strongly', 'disagree']


THE TOP 15 WORDS FOR TOPIC #2
['ان' ,'الطالب' ,'دكتور' ,'بشكل' ,'افضل' ,'ممتاز', 'perfect', 'الدكتو' ,'كل' ,'جدا' ,'المساق
في' ,'من' ,'على' ,'المادة' ,'ر']


THE TOP 15 WORDS FOR TOPIC #3
['كلا', 'suggestion', 'lessons', 'dr', 'الماده' ,'توجد', 'does', 'لايوجد', 'understan
d', 'recommendations', 'thank', 'شكرا' ,'يوجد', 'good', 'لا']


THE TOP 15 WORDS FOR TOPIC #4
['going', 'fine', 'right', 'prepare', 'prior', 'far', 'consumed', 'showing', 'm
odern', 'number', 'statements', 'survey', 'nth', 'strongly', 'agree']
```

4. Conclusion:

As you are aware, I am currently enrolled in a Data Science bootcamp and concurrently hold a part-time job. While this project has immense potential for further exploration, I have managed to showcase the core skills necessary for the task.

There is ample room for expanding this analysis, including the following possibilities:

Deeper Sentiment Analysis: A more comprehensive sentiment analysis could provide valuable insights. Examining overall sentiments for each course, for instance, would enhance our understanding of the course content's reception.

Enhanced Model Optimization: The model's performance can be further refined through hyperparameter tuning. This step could potentially result in even more accurate and meaningful topic modeling results.

Evaluation Metrics: Utilizing evaluation metrics like coherence and perplexity would allow for a more robust assessment of the quality of the topics generated by the Latent Dirichlet Allocation (LDA) model.

4.1. **Recommendations**:

There are numerous insights that can be gleaned from the feedback. However, one notable observation suggests taking action by offering courses in both English and Arabic to accommodate students who may struggle with the Arabic language.

Despite time constraints, this project has illuminated the capabilities of LDA and sentiment analysis in uncovering valuable information within text data. As time permits, I look forward to exploring these avenues for improvement and expanding the scope of this analysis.