Sentiment Analysis

**Design and Implementation**

To create the GitHub actions workflow, I focused on automating the process of fetching, analysing, and reporting on the sentiment of news over time sourced from the CNN latest news. Here is how I did it:

1. First, I make a function called “extract\_titles\_descriptions” which will make a GET request from the CNN news website to fetch data on titles and descriptions of news in that day.
2. After fetching the data, the sentiment analysis begins. I use the package “sentiment” to analyse the sentiment of both the titles and descriptions of the news articles. The sentiment scores are then computed for each sentence, providing insight into the overall sentiment of the news stories.
3. Once the sentiment analysis is complete, I implemented code to calculate the summary statistics for both titles and descriptions. These statistics offer a concise overview of the sentiment distribution, including measures such as mean and sentiment score.
4. The summary statistics are then printed to the console to allow quick inspection of the sentiment. I then save the results to text files which contain the current date the sentiment was performed for later comparison between dates.
5. The GitHub action workflow executes these tasks on regular basis. By scheduling the workflow to run three times a week, it will make easy to receive timely updates on sentiment trend without manual intervention.

**Efficacy:**

I evaluated the effectiveness of this project based on several criteria. Firstly, the workflow efficiently retrieves the latest news data, conducts sentiment analysis on titles and descriptions, calculates summary statistics, and commits the results to the repository. This guarantees the stakeholders would have easy access to up-to-date information without any manual intervention. Furthermore, as this is set up to run multiple times a week, it ensures that frequent updates on sentiment trends are shown and minimizes unnecessary computational resources.

**Optimization:**

While designing the GitHub Actions workflow, it's essential to consider platform limitations and scalability challenges. GitHub imposes a workflow limit of 35 days, after which the workflow will be automatically cancelled. Additionally, there's a cap of 20 unique workflows per repository. These constraints pose potential limitations when scaling the sentiment analysis over an extended period or across multiple projects.

To address these limitations and optimize the workflow for long-term scalability, it is essential to explore alternative hosting options outside of GitHub. By decoupling the workflow from GitHub Actions and using cloud-based solutions such as AWS Lambda, Google Cloud Functions, or Azure Functions, we can overcome the constraints imposed by GitHub’s workflow limits.

**Analysis of results**

The GitHub workflow has been running for a full 2 weeks now. I will be showing the summary statistics over the course of a one of the weeks, this should offer good insight on the sentiment trends in the news titles and descriptions.

**Sentiment in Descriptions:**

* **April 13, 2024:**
  + The sentiment scores for descriptions ranged from -0.10781 to 0.17709.
  + The mean sentiment score was 0.03464, indicating a slightly positive sentiment overall.
* **April 15, 2024:**
  + Sentiment scores ranged from -0.052169 to 0.160745.
  + The mean sentiment score increased to 0.054288, suggesting a more positive sentiment compared to April 13.
* **April 17, 2024:**
  + Sentiment scores ranged from -0.13235 to 0.08527.
  + The mean sentiment score decreased slightly to -0.02354, indicating a shift towards a slightly negative sentiment compared to April 15.

The analysis reveals fluctuations in sentiment over the week, with sentiment turning slightly more positive on April 15 before trending slightly negative again by April 17. These fluctuations reflect changes in the news landscape and public sentiment during this period.

**Sentiment in Titles:**

* **April 13, 2024:**
  + Sentiment scores for titles ranged from -0.20000 to 0.27745.
  + The mean sentiment score was 0.03873, indicating a slightly positive sentiment overall.
* **April 15, 2024:**
  + Sentiment scores ranged from 0.00000 to 0.33264.
  + The mean sentiment scores notably increased to 0.16632, suggesting a more positive sentiment compared to April 13.
* **April 17, 2024:**
  + Sentiment scores ranged from -0.30619 to 0.19576.
  + The mean sentiment score decreased to -0.05521, indicating a shift towards a slightly negative sentiment compared to April 15.

Similarly to the sentiment in descriptions, sentiment in titles also exhibited fluctuations throughout the week. Notably, there was a significant increase in sentiment on April 15, followed by a decrease by April 17.

In conclusion, the analysis of sentiment summary statistics aligned closely with the observed news events during the week of April 13 to April 17. The fluctuations in sentiment corresponded to significant news topics and events reported during those dates. On April 13, sentiment rose slightly, coinciding with news coverage on food prices and developments related to Trump. April 15 saw a notable increase in sentiment, reflecting coverage of events involving Israel and Trump's legal trials. However, sentiment took a downturn on April 17, correlating with reports of a massive flooding in Dubai and escalating conflicts between countries such as Ukraine and Russia, as well as heightened tensions between Palestine and Israel. This alignment underscores the direct impact of news events on public sentiment, highlighting the importance of considering real-time news context when interpreting sentiment analysis results. Such insights can aid in understanding societal reactions, predicting market trends, and informing decision-making processes in various domains.