Project Overview & Goal

❑**Title:** Sentiment Analysis of Media Articles on USA-Canada Tariffs

❑**Goal:**

- Analyze sentiment in news media articles related to USA-Canada tariffs using NLP techniques and visualize insights using Power BI.

Tools & Libraries:

- Python, Transformers, RoBERTa, Newspaper3k, Seaborn, Matplotlib, WordCloud, Power BI

Project Architecture

1. Mount Google Drive

2. Install & Import Required Libraries
(transformers, newspaper3k, BeautifulSoup, wordcloud, seaborn, matplotlib, pandas)

3. Extract Article Text
- Use newspaper3k for basic scraping
- Use BeautifulSoup for fallback parsing

4. Language Detection
- Detect and filter out non-English articles using langdetect or similar libraries

5. Classify Sentiment
- Load RoBERTa model (cardiffnlp/twitter-roberta-base-sentiment)
- Tokenize and predict
- Map outputs to sentiment labels

6. Real-time or Scheduled Scraping
- Automate scraping with scheduling tools like cron or APScheduler

7. Keyword Tagging
- Tag articles with keywords (e.g., 'tariff', 'trade war') for advanced filtering

8. Store Output in Google Drive
- Save intermediate & final files
- Ensure GDrive is mounted

9. Visualize Sentiment Results
- Use Seaborn/Matplotlib
- Plot distribution, word clouds

10. Export Final Data as CSV
- Save to Google Drive
- Load into Power BI using OneDrive sync

Objective & Methodology

❑ **Objective:**

- Understand media tone (positive/negative/neutral) toward tariff-related news.
- Compare sentiment between Canadian and US outlets.

❑ Methodology:

- Scrape & clean articles → Apply RoBERTa sentiment classifier → Store outputs → Visualize → Generate dashboard.

Code and Libraries Used

❑**NLP Model:**transformers for RoBERTa sentiment classification

```
from transformers import pipeline

# Load sentiment analysis pipeline with RoBERTa
sentiment_pipeline = pipeline("sentiment-analysis", model="cardiffnlp/twitter-roberta-base-se

# Predict sentiment
result = sentiment_pipeline("Canada’s trade deal boosts domestic markets.")
print(result)
```

❑**Scraping:**newspaper3k, BeautifulSoup, requests

Scraping:newspaper3k, BeautifulSoup, requests

```
from newspaper import Article
import requests
from bs4 import BeautifulSoup

# Example using newspaper3k
url = 'https://example.com/article'
article = Article(url)
article.download()
article.parse()
text = article.text

# Fallback with requests + BeautifulSoup
response = requests.get(url)
soup = BeautifulSoup(response.content, 'html.parser')
text_alt = soup.get_text()
```

Data Handling:pandas, os

```
import pandas as pd
import os

# Load URLs and metadata
df = pd.read_csv('article_links.csv')

# Store results
df['sentiment'] = df['text'].apply(lambda x: sentiment_pipeline(x)[0]['label'])
df.to_csv('SentimentResults.csv', index=False)
```

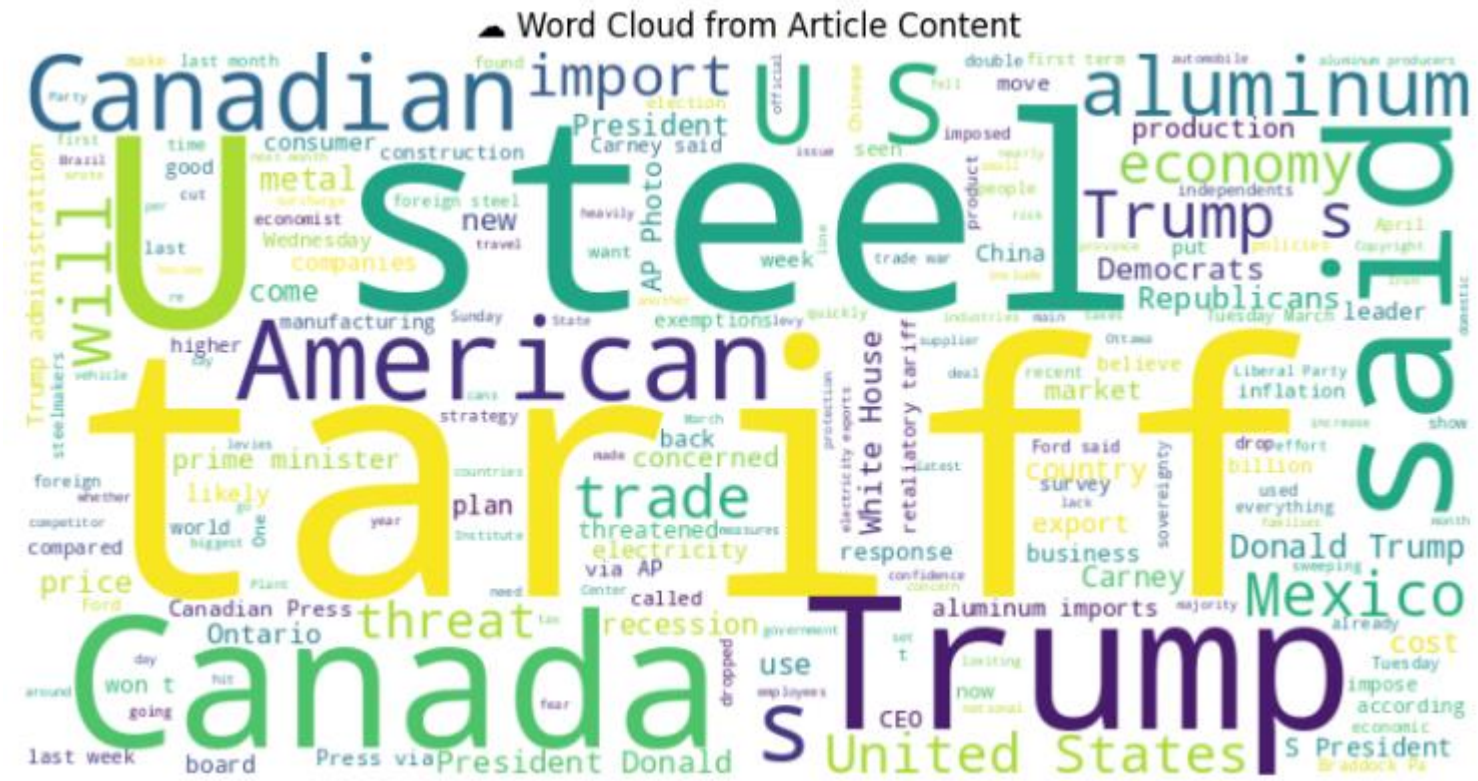
Visualization (matplotlib, seaborn, wordcloud)

```
import seaborn as sns
import matplotlib.pyplot as plt
from wordcloud import WordCloud

# Sentiment distribution
sns.countplot(x='sentiment', data=df)
plt.title('Sentiment Distribution')
plt.show()

# Word cloud
text_corpus = ' '.join(df['text'].tolist())
wordcloud = WordCloud(width=800, height=400).generate(text_corpus)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

Storage & Reporting:Google Drive + Power BI

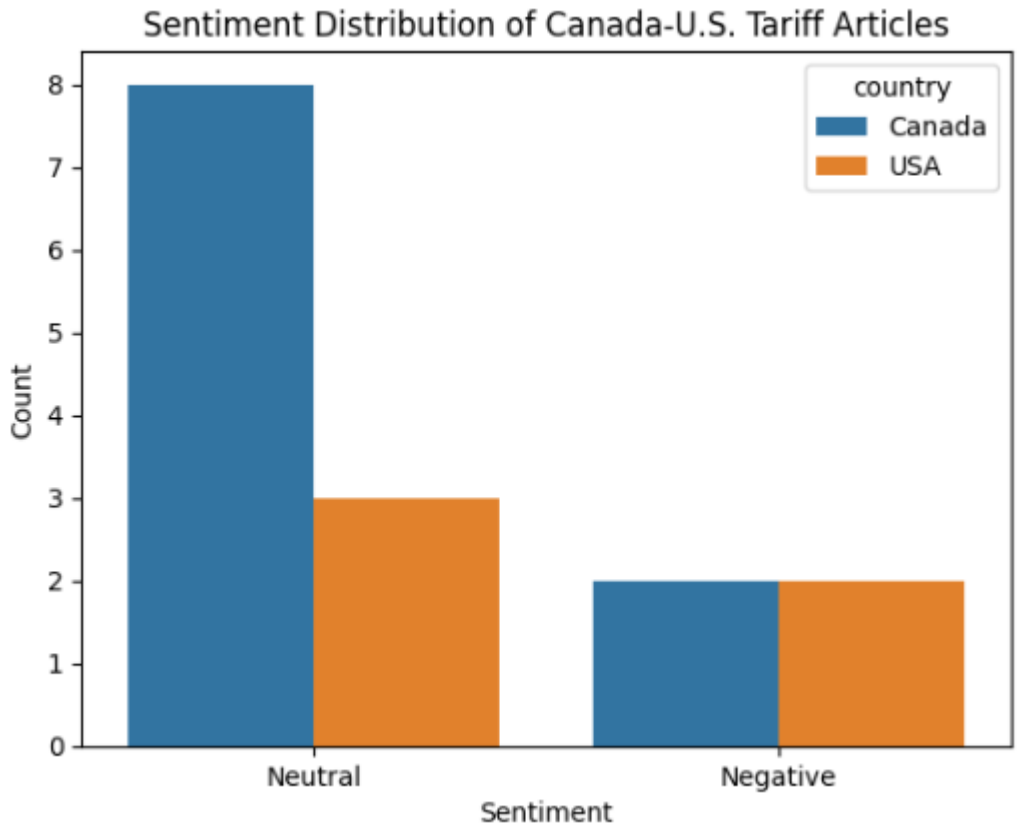
```
# Mount Google Drive in Colab
from google.colab import drive
drive.mount('/content/drive')


# Save results to Drive
output_path = '/content/drive/MyDrive/SentimentAnalysis/SentimentResults.csv'
df.to_csv(output_path, index=False)
```

Word Cloud from Article Content



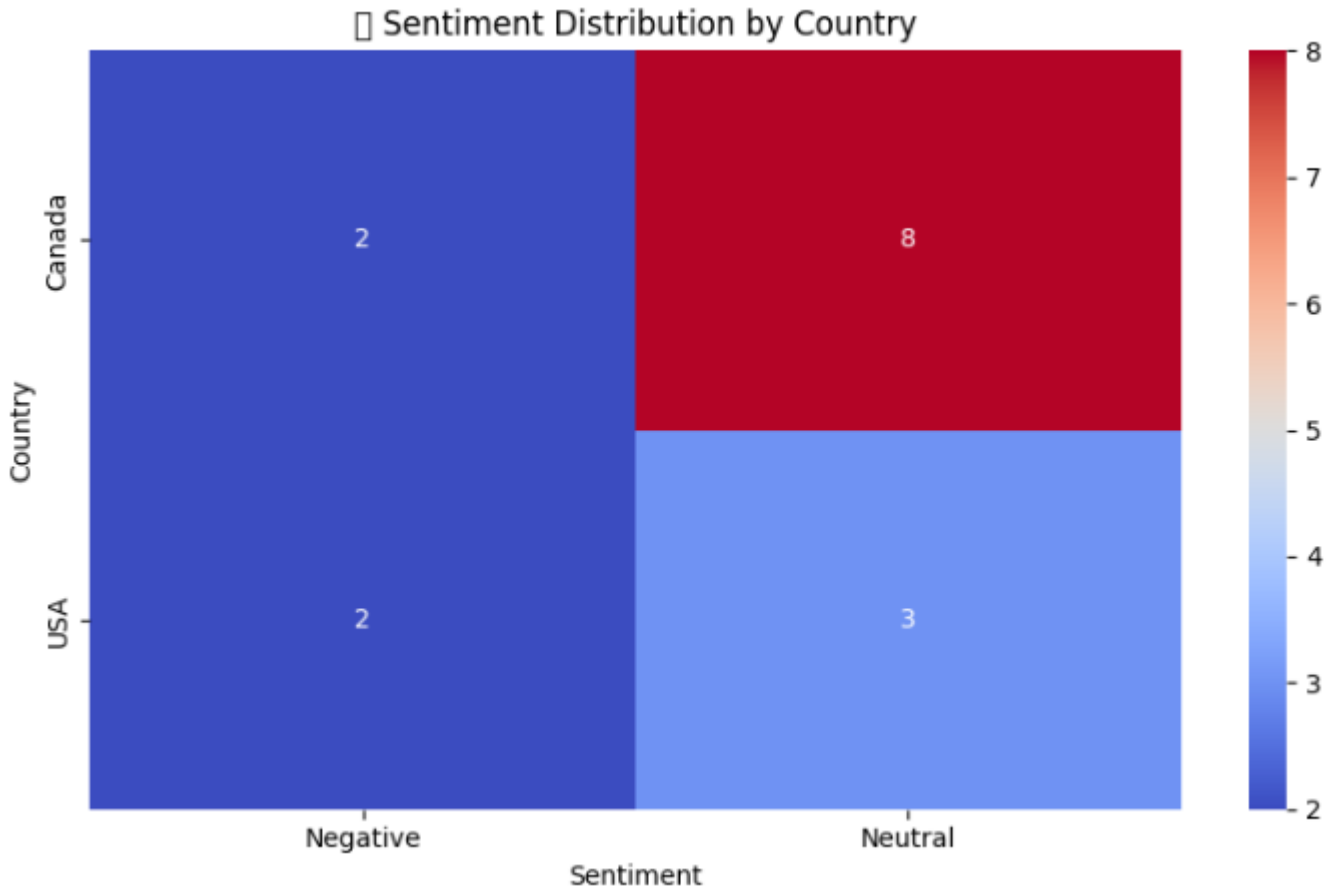
Word cloud highlights key terms from articles on Canada-U.S. tariffs. Dominant words like *tariff*, *steel*, *Canada*, *Trump*, and *aluminum* reflect the central themes of trade tension and economic impact.

Sentiment Breakdown of Canada-U.S. Tariff News Coverage



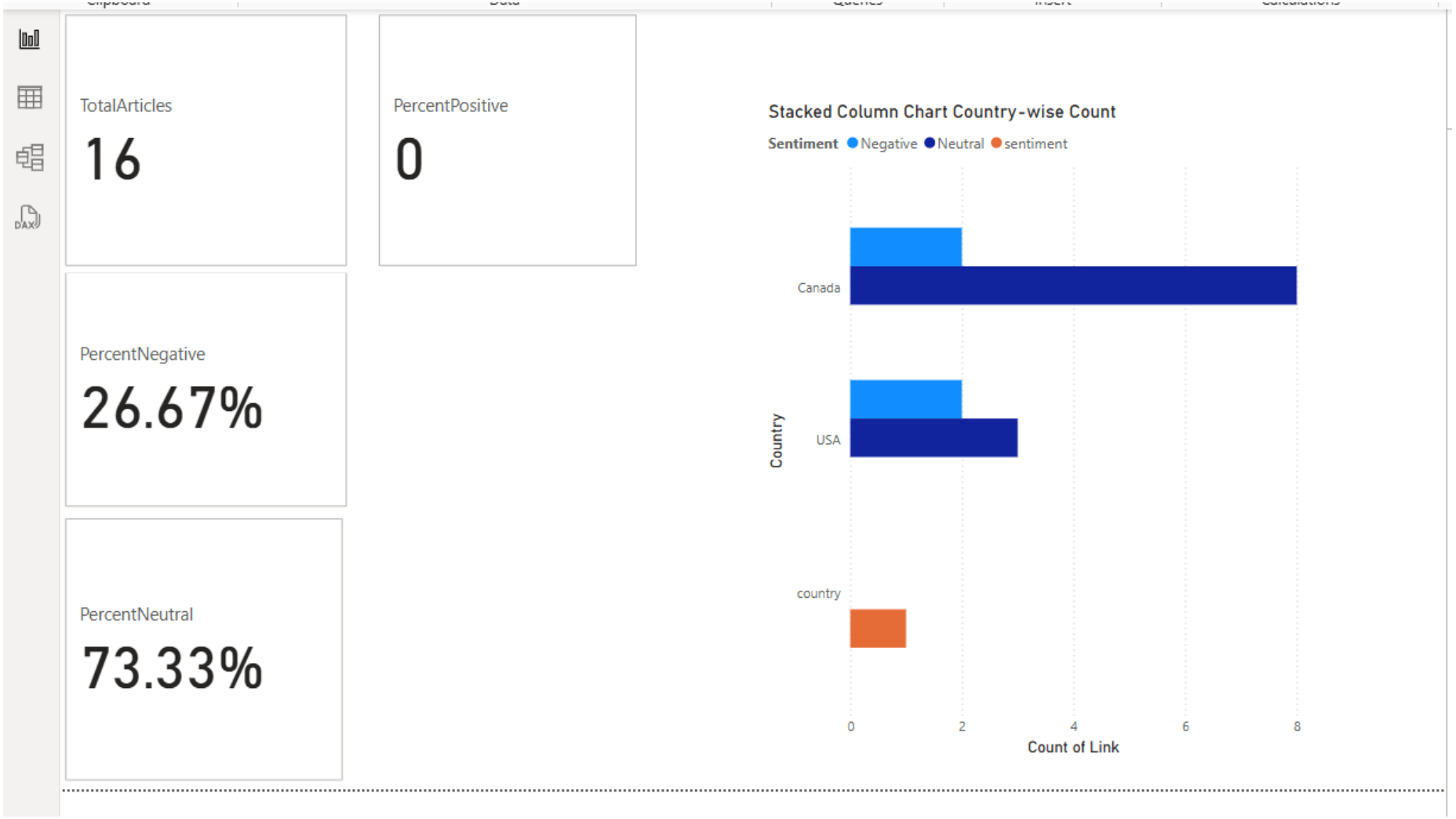
Canada has more neutral articles, while both countries show equal negative sentiment in tariff-related coverage."

Sentiment Distribution by Country on Tariff Articles



Canada shows mostly neutral coverage, while the U.S. has a more balanced mix of neutral and negative sentiment.

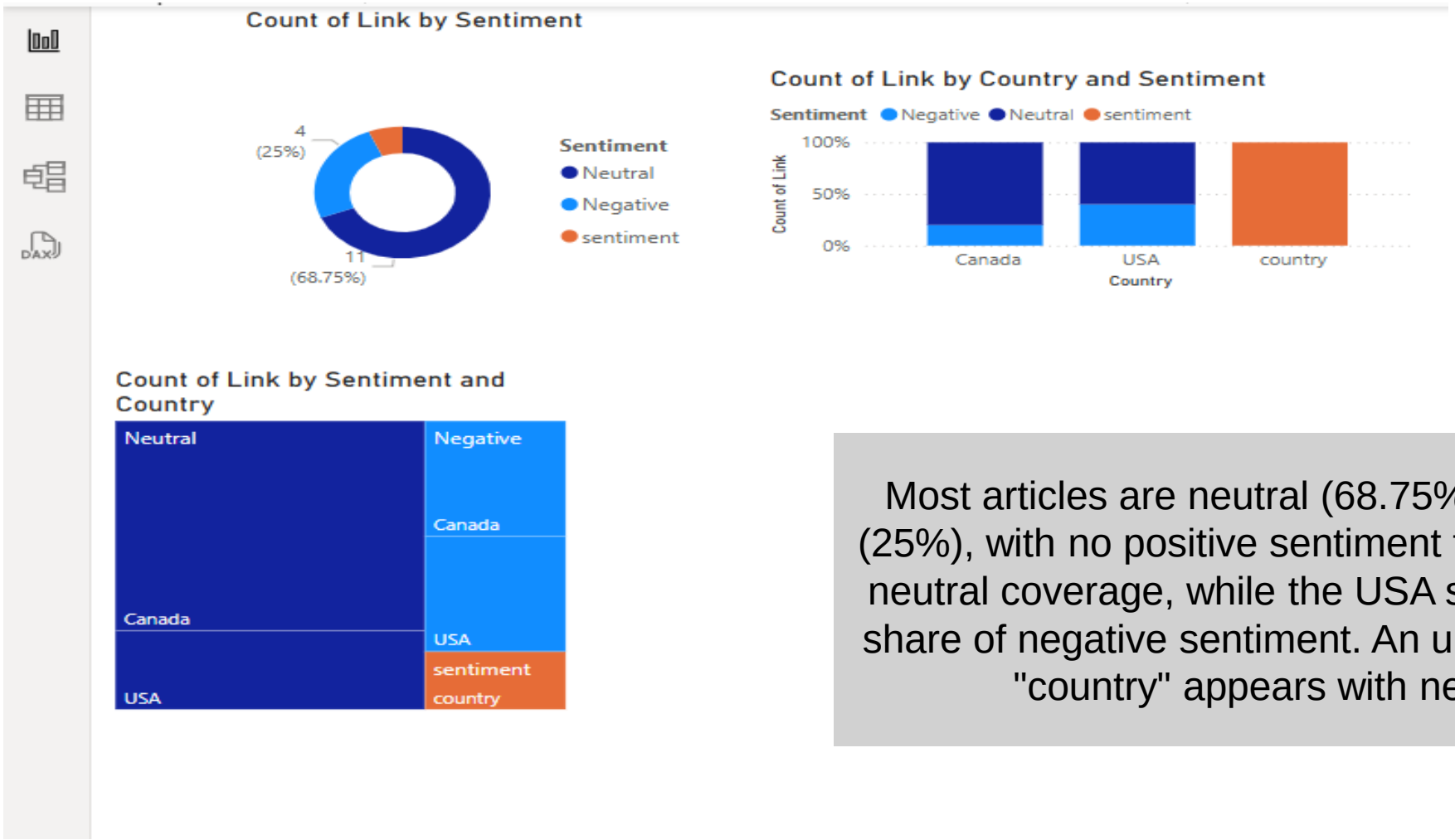
Power BI Dashboard Screenshots



Title: Sentiment Dashboard Overview of Canada-U.S. Tariff News

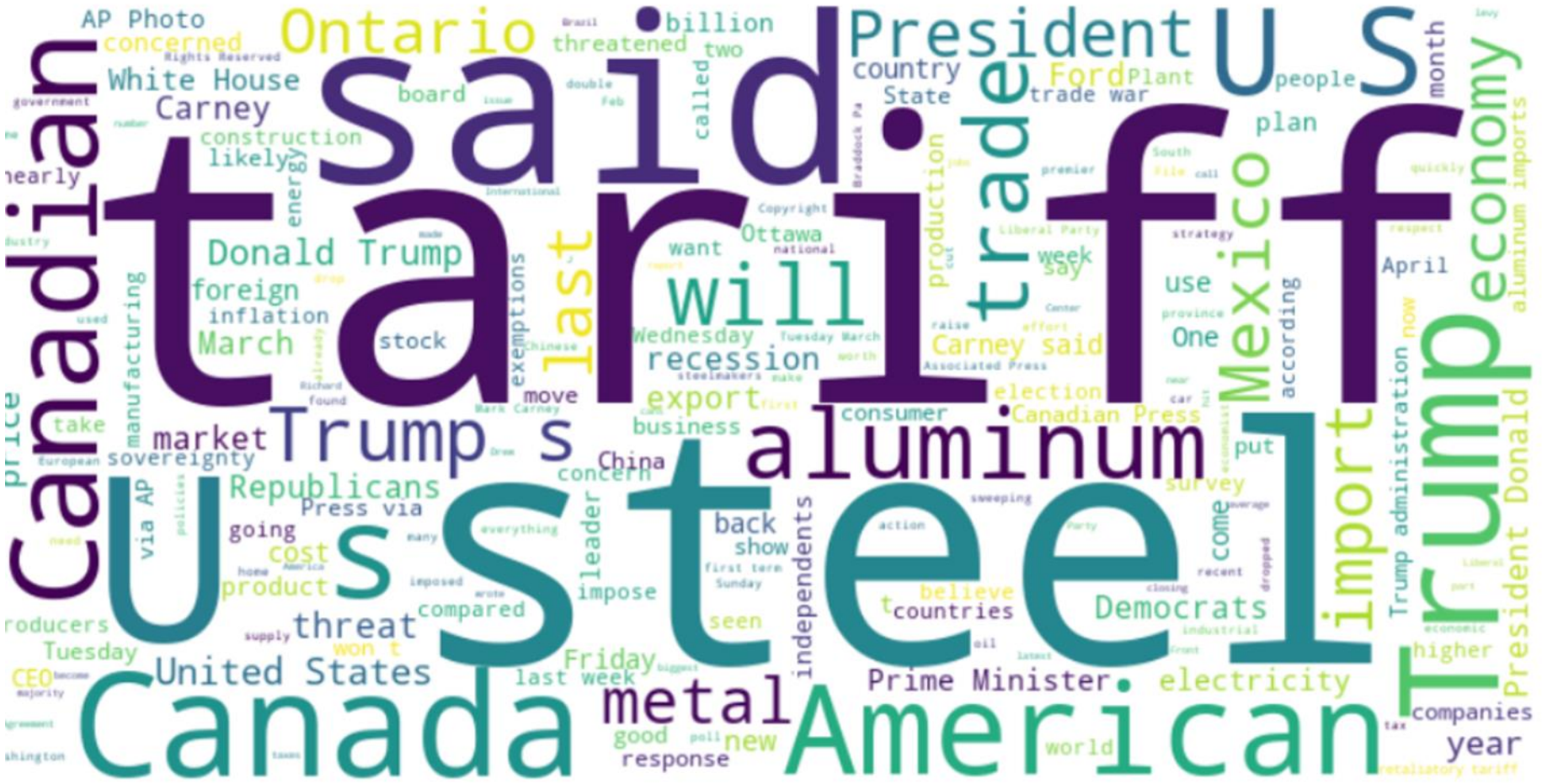
Description:
Out of 16 analyzed articles, 73.33% are neutral and 26.67% negative. No positive sentiment was detected. Canada had more article coverage, predominantly neutral in tone.

Sentiment Breakdown of Tariff-Related Articles by Country



Most articles are neutral (68.75%), followed by negative (25%), with no positive sentiment found. Canada has more neutral coverage, while the USA shows a relatively higher share of negative sentiment. An unidentified group labeled "country" appears with neutral sentiment.

Word Cloud & Sentiment Distribution



Word cloud showing key terms from Canadian articles on U.S. tariffs—top words include **"tariff," "steel," "Canada,"** and **"Trump."**

Results

- Out of 15 analyzed articles, sentiment distribution was:
  - 73.33% Neutral (11 articles)
  - 26.67% Negative (4 articles)
  - 0% Positive
- By country:
  - Canada: 8 neutral, 2 negative
  - USA: 3 neutral, 2 negative
- A word cloud highlighted frequent terms like tariff, Canada, Trump, and steel, emphasizing trade-related discourse.

Conclusions

- Neutral reporting dominates, suggesting balanced media coverage.
- No positive sentiment appeared in any article.
- Canada’s coverage is more neutral, while the USA shows slightly more polarity.
- One entry was excluded due to data labeling issues.

References:

- Hugging Face RoBERTa: https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment
- Newspaper3k Docs: https://newspaper.readthedocs.io
- Power BI: https://powerbi.microsoft.com
- Global News, CBC, CNN, Fox, etc. (for article sources)

- GitHub: https://github.com/ronaldkalani/TariffSentimentAnalysis.git