Report on how ethical and legal issues have shaped public perception of PricewaterHouseCoopers (PwC) on Twitter.

By Ece E., Ilnaza S., Kyrylo S., Neilya B., Vasilii B.

Table of Contents:

1. Objective

The primary objective of this project is to analyze social media posts related to PricewaterhouseCoopers (PwC) to determine public sentiment, identify trends, and uncover key insights. This analysis aims to help PwC understand customer perceptions, improve their marketing strategies, and address any potential issues proactively.

2. Approach

To achieve this goal, we focused on collecting and analyzing Twitter posts mentioning PwC. Twitter was chosen due to its widespread use and the availability of comprehensive data through its API. The project involved the following steps:

2.1 Data Collection:

- **API Setup**: We utilized the Tweepy library to interact with the Twitter API. Authentication was handled using OAuth, which required API credentials (API key, API secret key, access token, and access token secret).
- **Defining Parameters**: We set specific parameters for the API requests, including keywords related to PwC, the date range for the data collection, and the maximum number of tweets to retrieve.
- **Making API Calls**: Using Tweepy, we made API calls to fetch tweets mentioning PwC within the defined parameters.
- **Handling Responses**: The collected tweets were parsed and stored in a structured format for further analysis.

2.2 Data Processing and Storage:

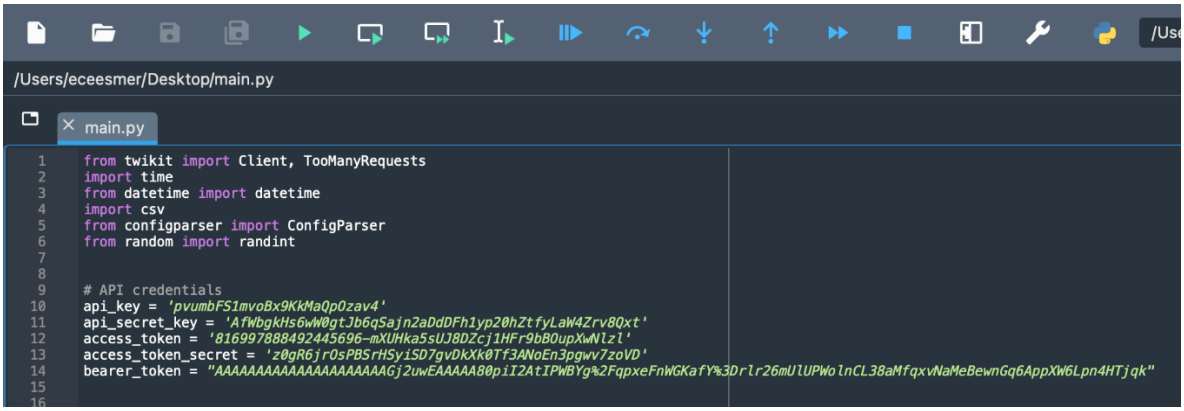
- **Data Structure**: The tweet data was stored in a JSON file, which included relevant information such as tweet text, user information, and timestamps.
- **Preprocessing**: The stored data was preprocessed to remove any irrelevant information, handle missing values, and normalize the text for analysis.

2.3 Sentiment Analysis and Insights:

- **Sentiment Analysis**: We employed natural language processing (NLP) techniques to analyze the sentiment of the tweets. This involved using sentiment analysis libraries to classify tweets as positive, negative, or neutral and gave it a score of +1 (positive), -1 (negative) and 0 (neutral)
- **Trend Identification**: We identified trending topics and keywords associated with PwC by analyzing the frequency and context of mentions.
- **Key Insights**: The analysis provided insights into customer perceptions, recurring themes, and potential areas of concern or interest for PwC.

3.Codes

3.1 Main Code



```python

MINIMUM_TWEETS = 200
QUERY = 'PwC lang:en until:2018-11-30 since:2018-11-01'


def get_tweets(tweets):
    if tweets is None:
        # * get tweets
        print(f'{datetime.now()} - Getting tweets...')
        tweets = client.search_tweet(QUERY, product='Top')
    else:
        wait_time = randint(5, 10)
        print(f'{datetime.now()} - Getting next tweets after {wait_time} seconds ...')
        time.sleep(wait_time)
        tweets = tweets.next()

    return tweets


# * login credentials
config = ConfigParser()
config.read('config.ini')
username = config['X']['username']
email = config['X']['email']
password = config['X']['password']

# * create a csv file
with open('old_tweets.csv', 'w', newline='') as file:
    writer = csv.writer(file)
    writer.writerow(['Tweet_count', 'Username', 'Text', 'Created At', 'Retweets', 'Likes'])

# * authenticate to X.com
# ! 1) use the login credentials. 2) use cookies.
client = Client(language='en-US')
# client.login(auth_info_1=username, auth_info_2=email, password=password)
# client.save_cookies('cookies.json')

client.load_cookies('cookies.json')

tweet_count = 0
tweets = None
```

```python
while tweet_count < MINIMUM_TWEETS:

    try:
        tweets = get_tweets(tweets)
    except TooManyRequests as e:
        rate_limit_reset = datetime.fromtimestamp(e.rate_limit_reset)
        print(f'{datetime.now()} - Rate limit reached. Waiting until {rate_limit_reset}')
        wait_time = rate_limit_reset - datetime.now()
        time.sleep(wait_time.total_seconds())
        continue

    if not tweets:
        print(f'{datetime.now()} - No more tweets found')
        break

    for tweet in tweets:
        tweet_count += 1
        tweet_data = [tweet_count, tweet.user.name, tweet.text, tweet.created_at, tweet.retweet_count,
                      tweet.favorite_count]

        with open('old_tweets.csv', 'a', newline='', encoding='utf-8') as file:
            writer = csv.writer(file)
            writer.writerow(tweet_data)

    print(f'{datetime.now()} - Got {tweet_count} tweets')

print(f'{datetime.now()} - Done! Got {tweet_count} tweets found')
```

3.2 The sentiment analysis of old tweets

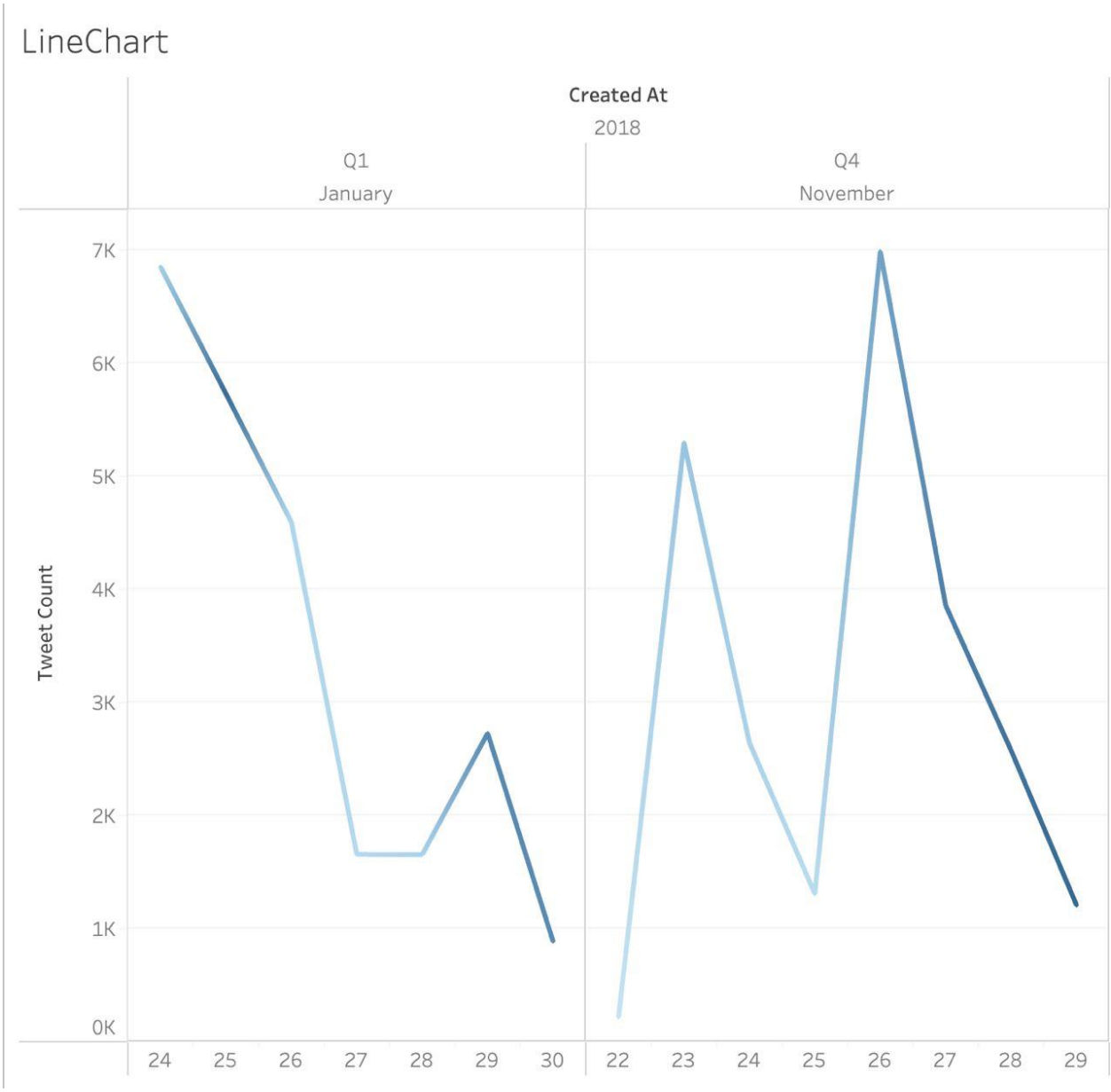
/Users/eceesmer/Desktop/sentiment_analysis_old_tweets.py

sentiment_analysis_old_tweets.py

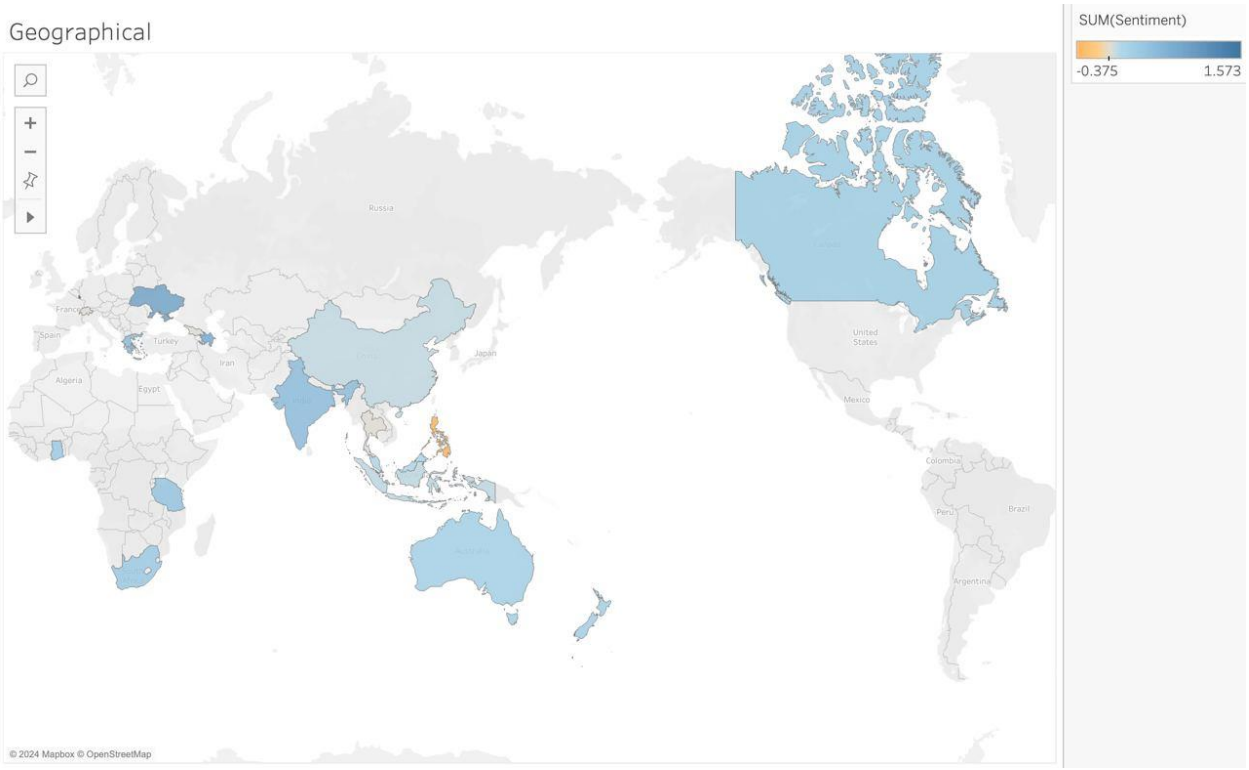
```python
import pandas as pd
import psycopg2
from textblob import TextBlob

# Database connection parameters
db_params = {
    'dbname': 'dataTwitter',
    'user': 'postgres',
    'password': 'postgres',
    'host': 'localhost',
    'port': '5433'
}

# Connect to the PostgreSQL database
conn = psycopg2.connect(**db_params)
cur = conn.cursor()

# Fetch data from the database
cur.execute("SELECT tweet_count, username, text, created_at, retweets, likes FROM oldtweets")
rows = cur.fetchall()

# Load data into a pandas DataFrame
columns = ['tweet_count', 'username', 'text', 'created_at', 'retweets', 'likes']
df = pd.DataFrame(rows, columns=columns)

# Perform sentiment analysis
def get_sentiment(text):
    return TextBlob(text).sentiment.polarity

df['sentiment'] = df['text'].apply(get_sentiment)

# Update the database with sentiment analysis results
update_query = """
    UPDATE oldtweets
    SET sentiment = %s
    WHERE tweet_count = %s
"""

for index, row in df.iterrows():
    cur.execute(update_query, (row['sentiment'], row['tweet_count']))

# Commit changes and close the connection
conn.commit()
cur.close()
conn.close()
```

4. Visualization

4.1 The Sentiment Trends Over Time



Graph 1; Visualization of the frequency of tweets about PwC at various times



Map 1; Visualization of the countries of origin and frequency of tweets about PwC

4.2 Demographic Insights

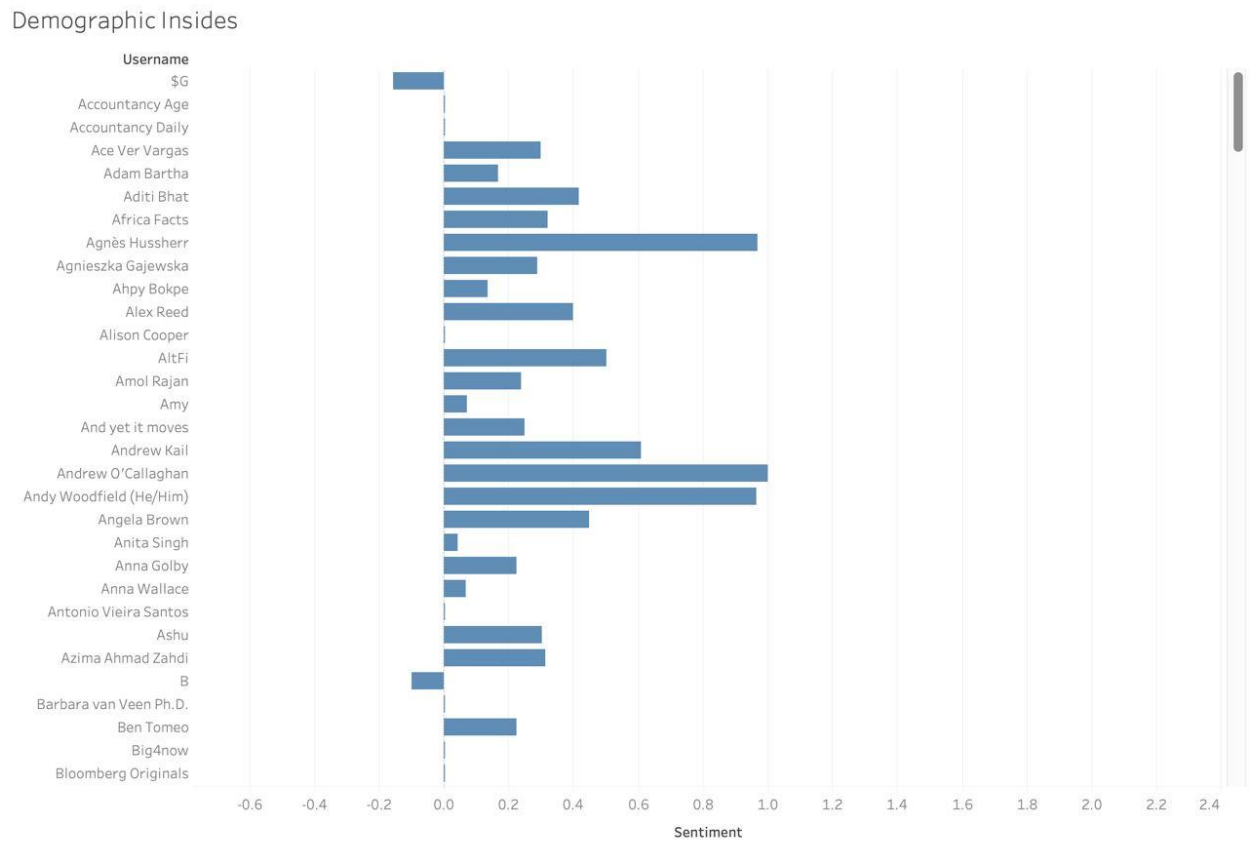


Table 1.1; Visualization of tweet sentiments

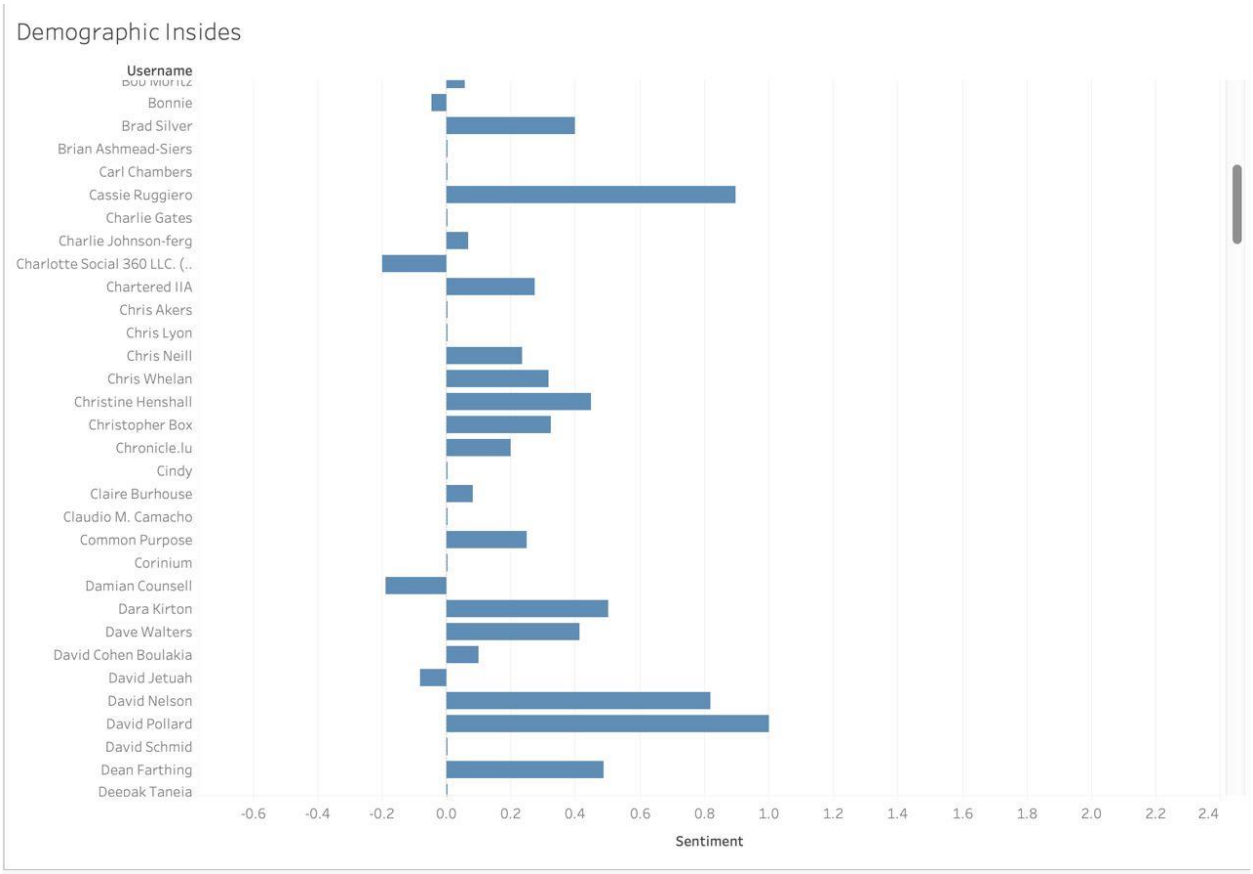


Table 1.2; Visualization of tweet sentiments

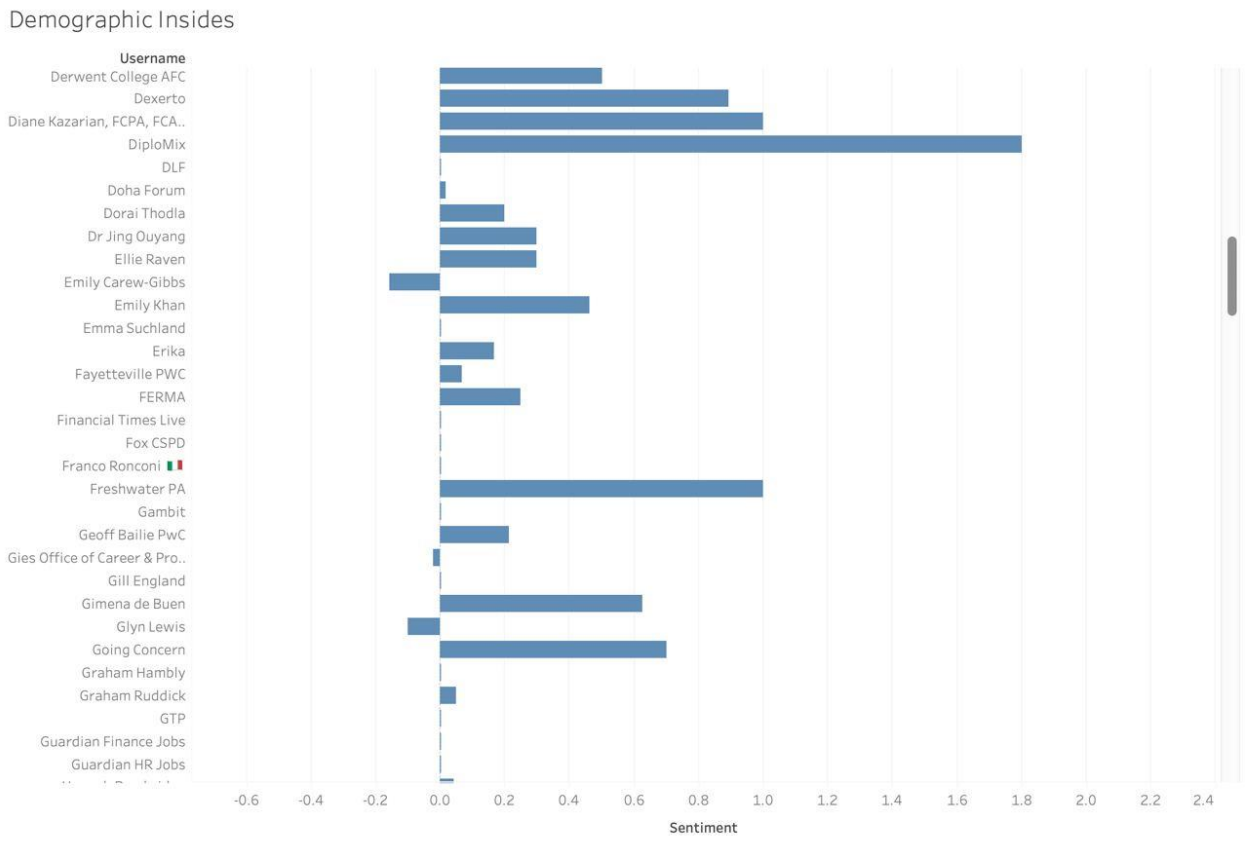


Table 1.3; Visualization of tweet sentiments

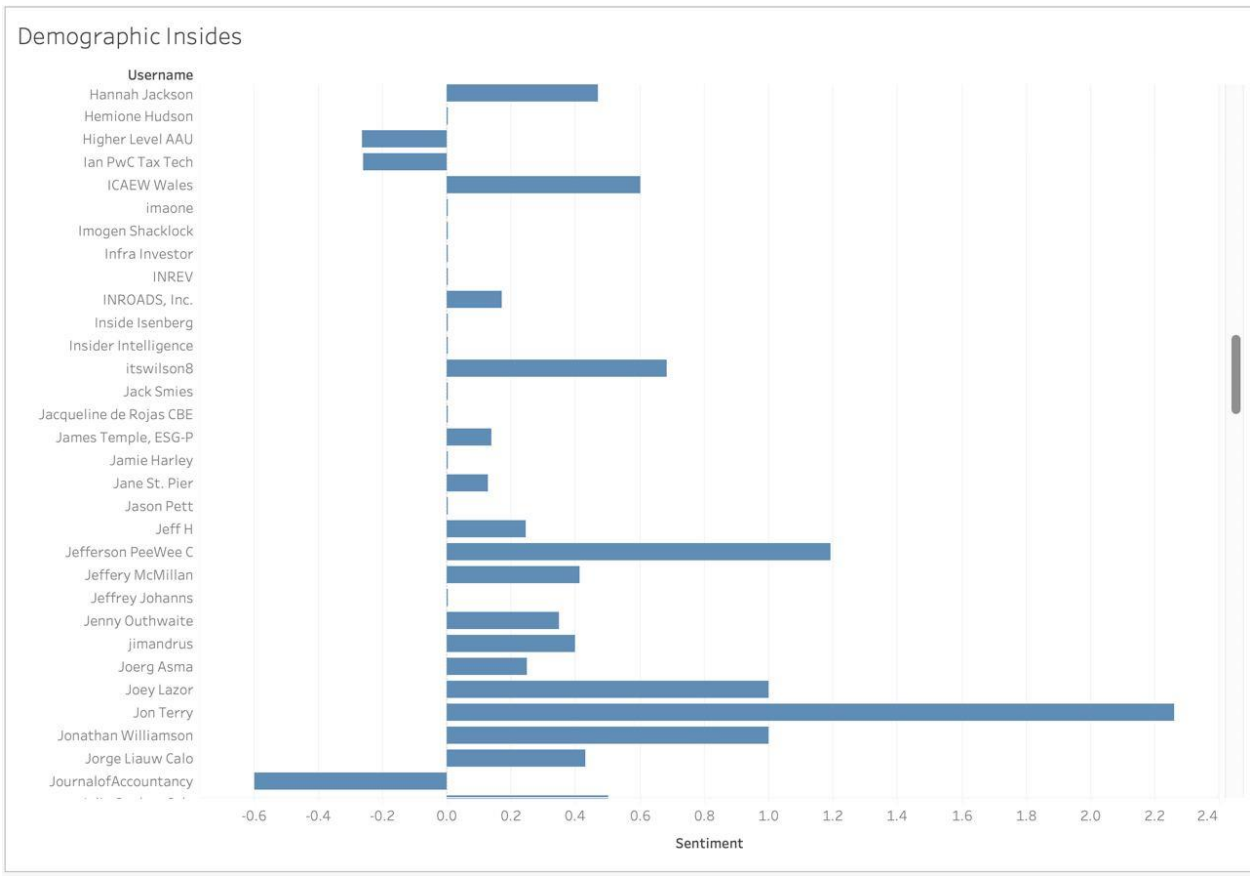


Table 1.4; Visualization of tweet sentiments

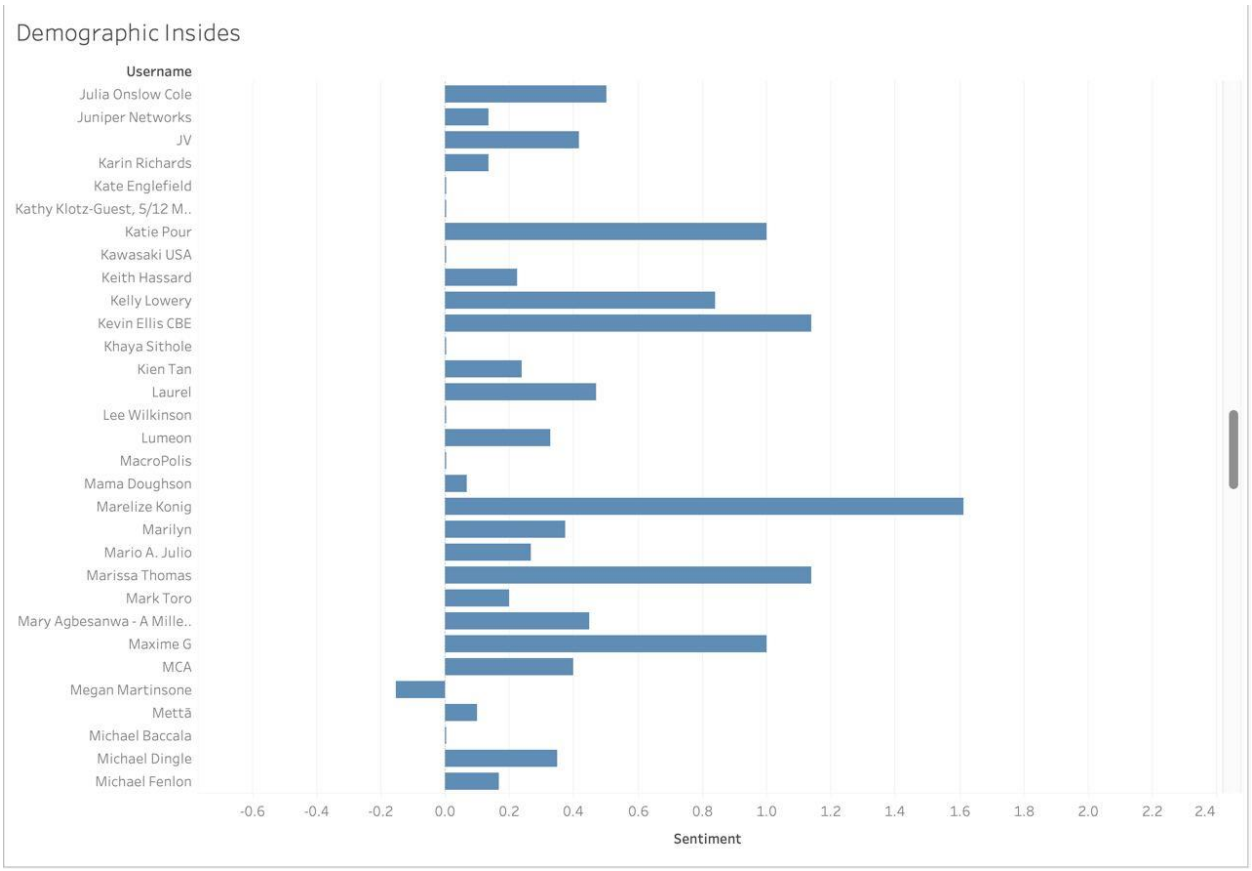


Table 1.5; Visualization of tweet sentiments

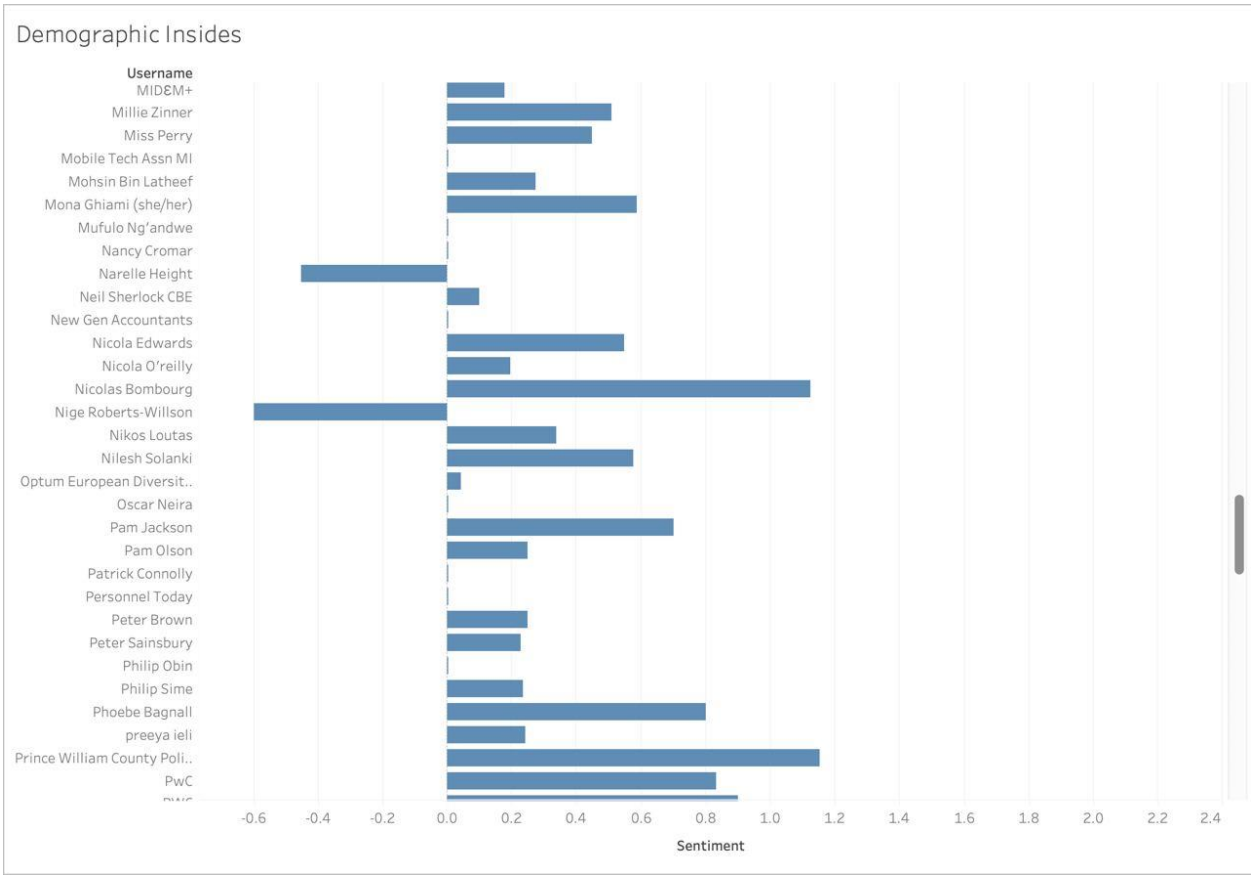


Table 1.6; Visualization of tweet sentiments

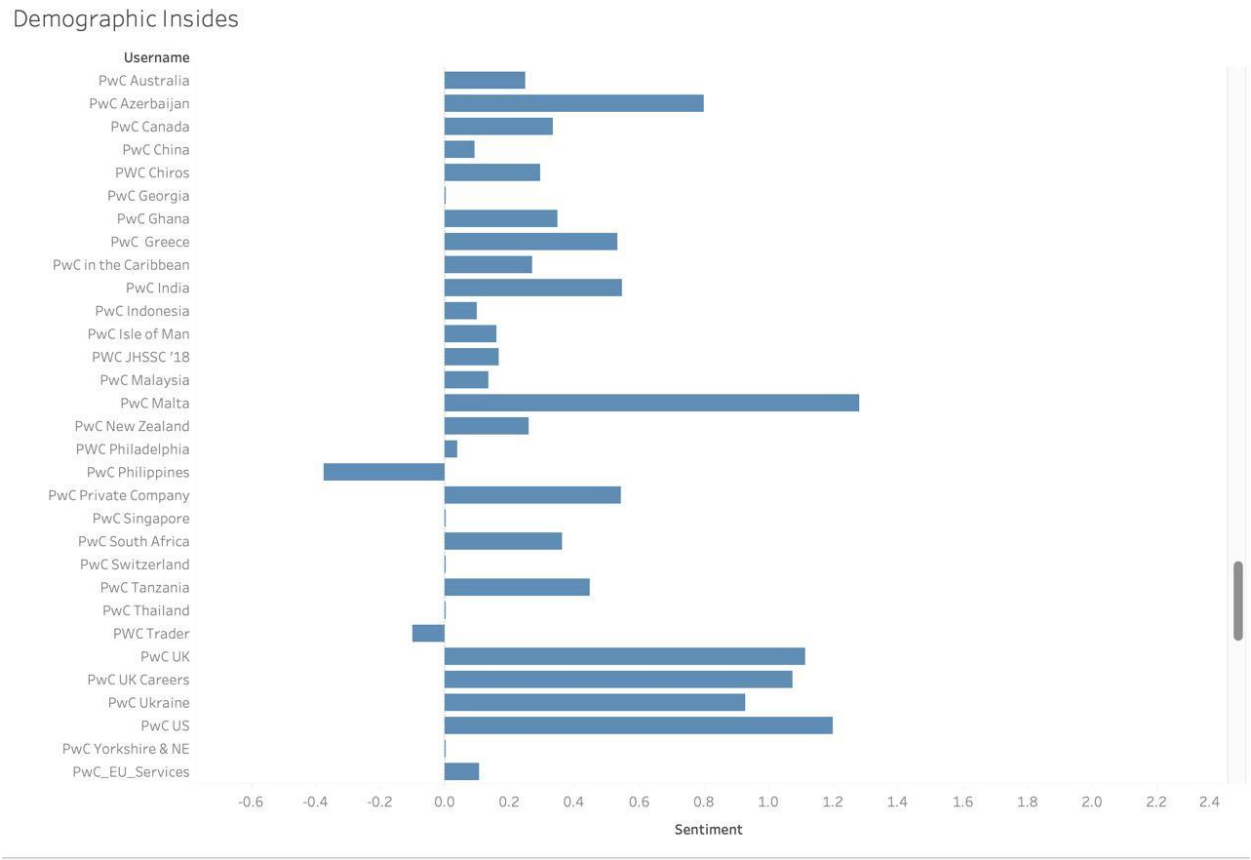


Table 1.7; Visualization of tweet sentiments

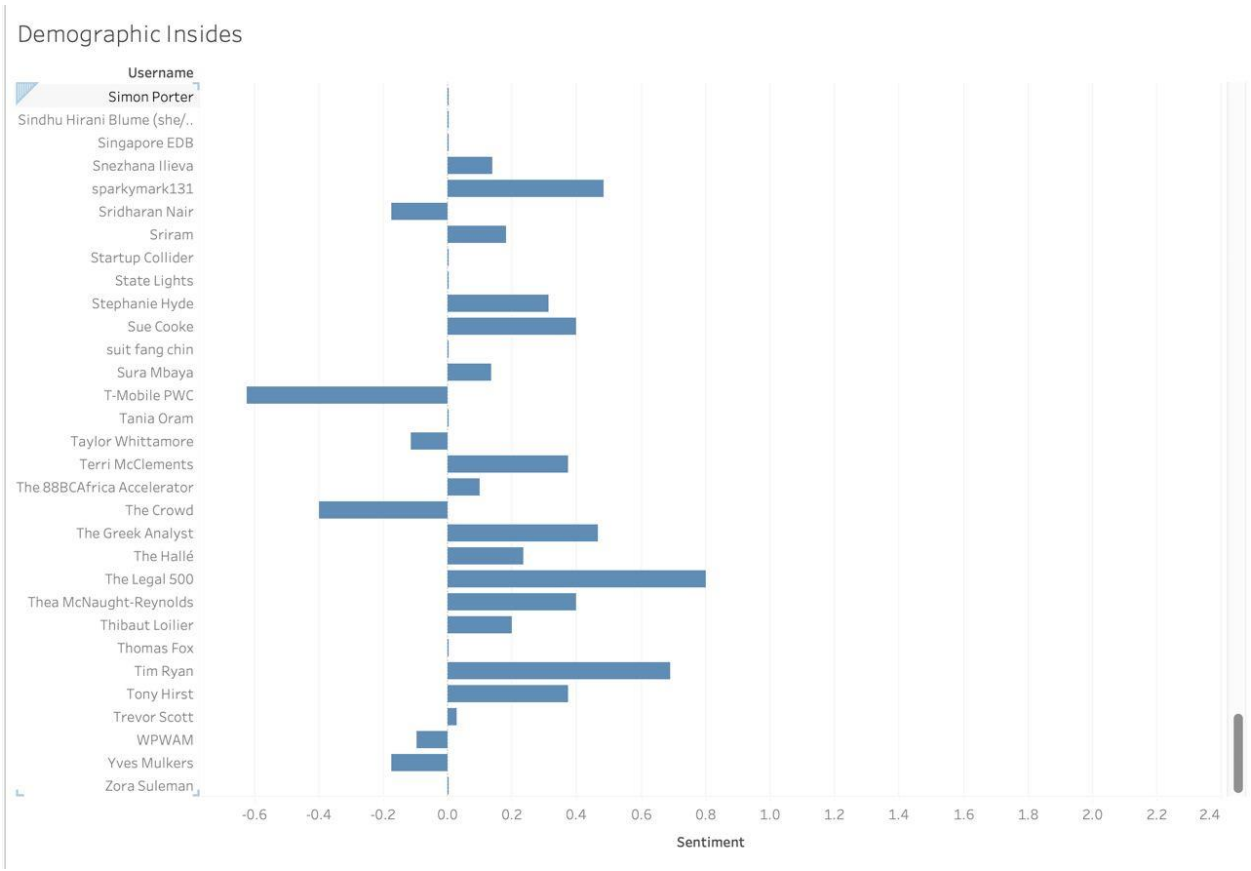
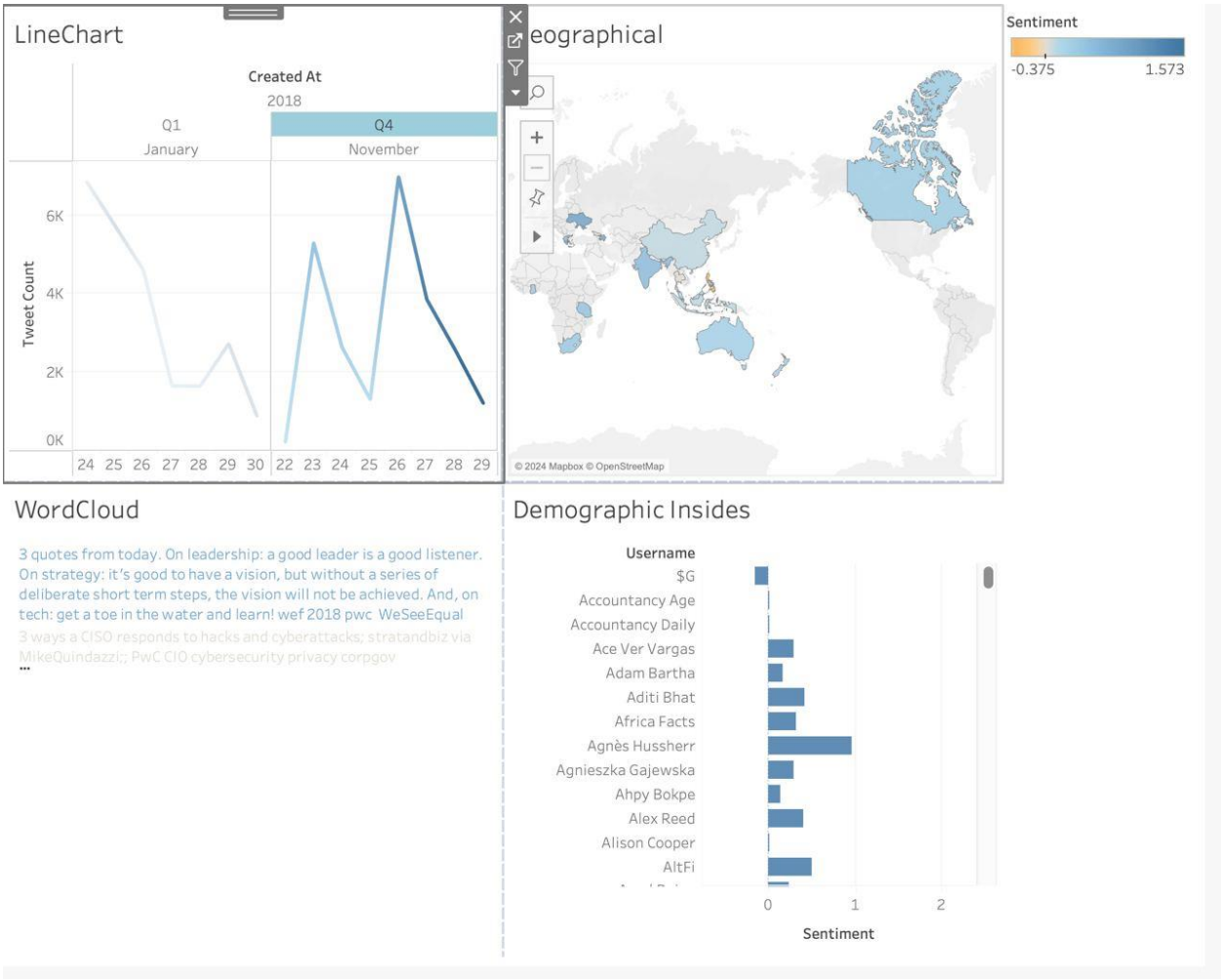


Table 1.8; Visualization of tweet sentiments

4.3 Dashboard



Picture 1.1; Depiction of the usage of Tableau for data visualisation

4.4 WorldCloud

Picture 1.2; Depiction of the usage of WorldCloud for data summarization, highlighting and identification of a large set of data to identify and locate tweets regarding PwC

5. Conclusion

Our analysis of Twitter posts mentioning PwC provides key insights into public sentiment and trends.

5.1 Key Findings

Sentiment Distribution:

Tweets about PwC are a mix of positive, negative, and neutral sentiments, with notable variations across different countries. Such as PwC Philippines having strong negative tweets, in contrast to PwC Ukraine having a vast amount of positive tweets.

Trends and Topics:

Identified recurring themes and spikes in tweet activity, correlating with specific events, such as the heightened amounts in tweets during January 1st, which correlates with the day when Donald Trump signed into law.

Demographic Insights:

Highlighted influential users and their sentiments, showing who is discussing PwC and their general views.

5.2 Implications for PwC

Addressing Negative Sentiment:

Targeted strategies to mitigate negative perceptions and improve customer satisfaction, such as responding promptly to complaints, providing transparent updates on issues, and offering solutions or compensation when necessary.

Leveraging Positive Sentiment:

Amplify positive feedback and engage with supportive users to strengthen brand loyalty.

Monitoring and Adaptation:

Ongoing social media monitoring to stay updated on public perception and adjust strategies as needed.

5.3 Recommendations

Enhanced Social Media Engagement:

Increase interactions with users and utilize influencers to broaden reach.

Proactive Communication:

Address common concerns and highlight PwC's ethical and legal efforts.

Data-Driven Decision Making:

Use insights for strategic planning and invest in advanced analytics for future analyses.

Overall, by understanding public perception on Twitter, PwC can improve its reputation, customer relations, and strategic positioning in the industry as long as they increase their interactions with users, address common concerns efficiently and overall create strategic planning that is well suited for future analyses.