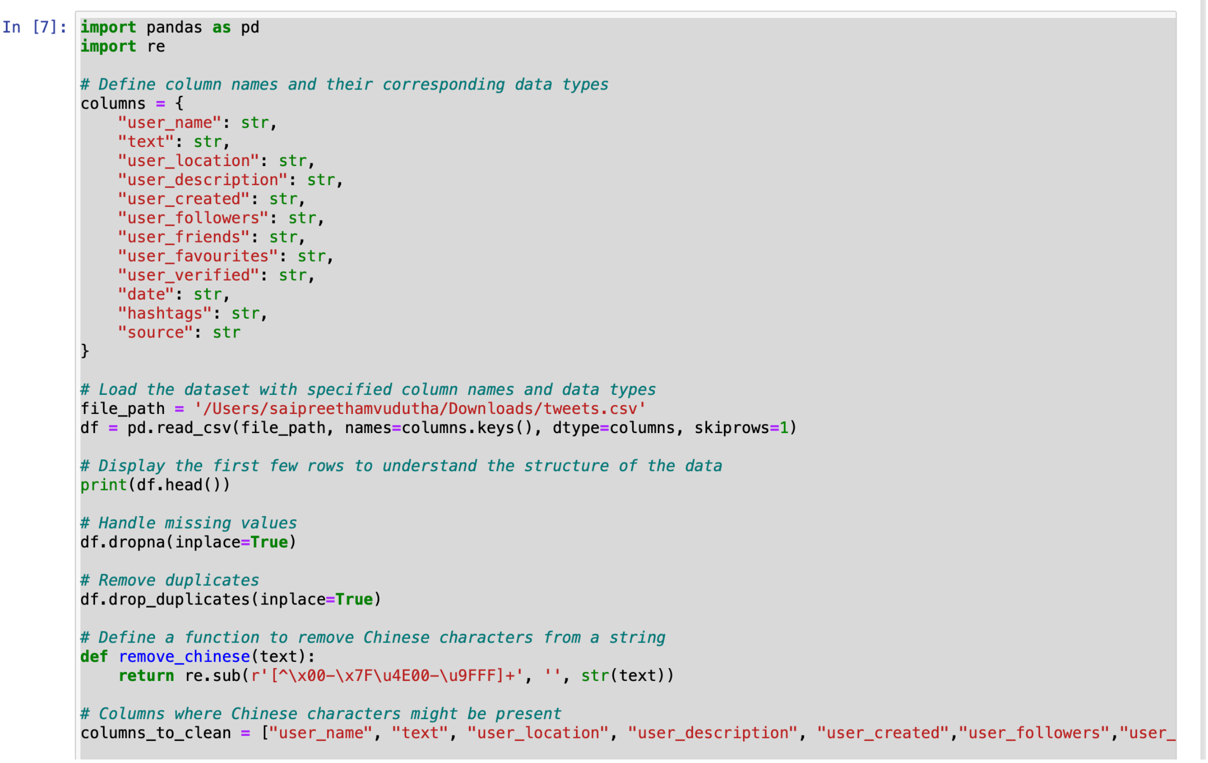
Data Set:

Large Language Models: the tweets https://www.kaggle.com/datasets/konradb/chatgpt-the-tweets/versions/169?resource=download

I have used the dataset from Kaggle, The dataset is about the Large Language Model: Tweets. Published on December 20, 2022.
This data set has a collection of tweets with the hashtag #chatgpt : discussions about the chatgpt language model, sharing experiences with using chatgpt, or asking for help with chatgpt-related issues. The tweets could also include links to articles or websites related to chatgpt, as well as images, videos, or other media. Overall, a collection of tweets with the hashtag #chatgpt would provide a glimpse into the online conversation surrounding chatgpt."

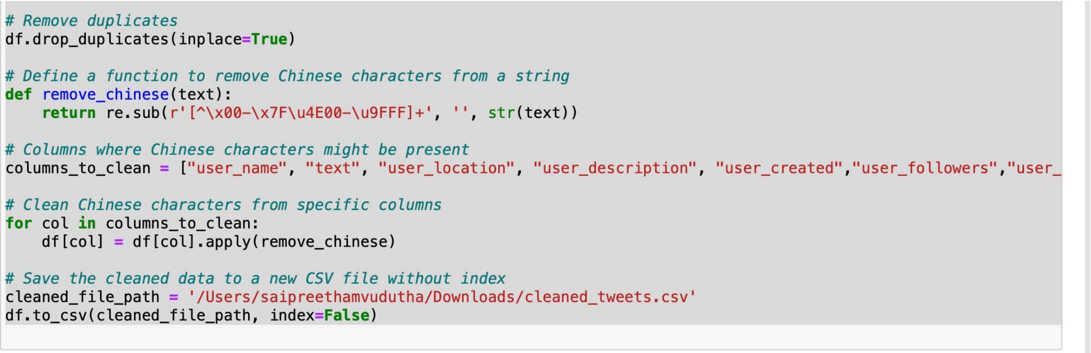
Data cleaning:

I have used python to clean the data using jupyter note book.

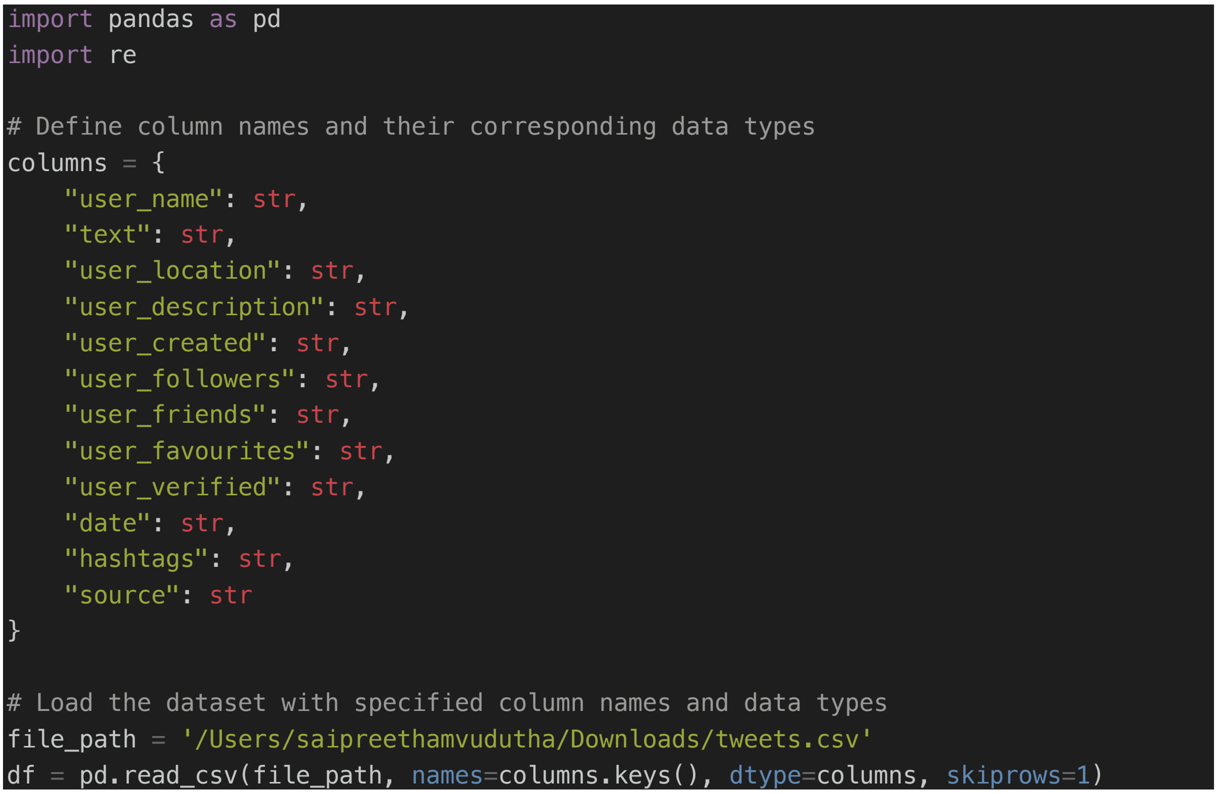


In [7]:

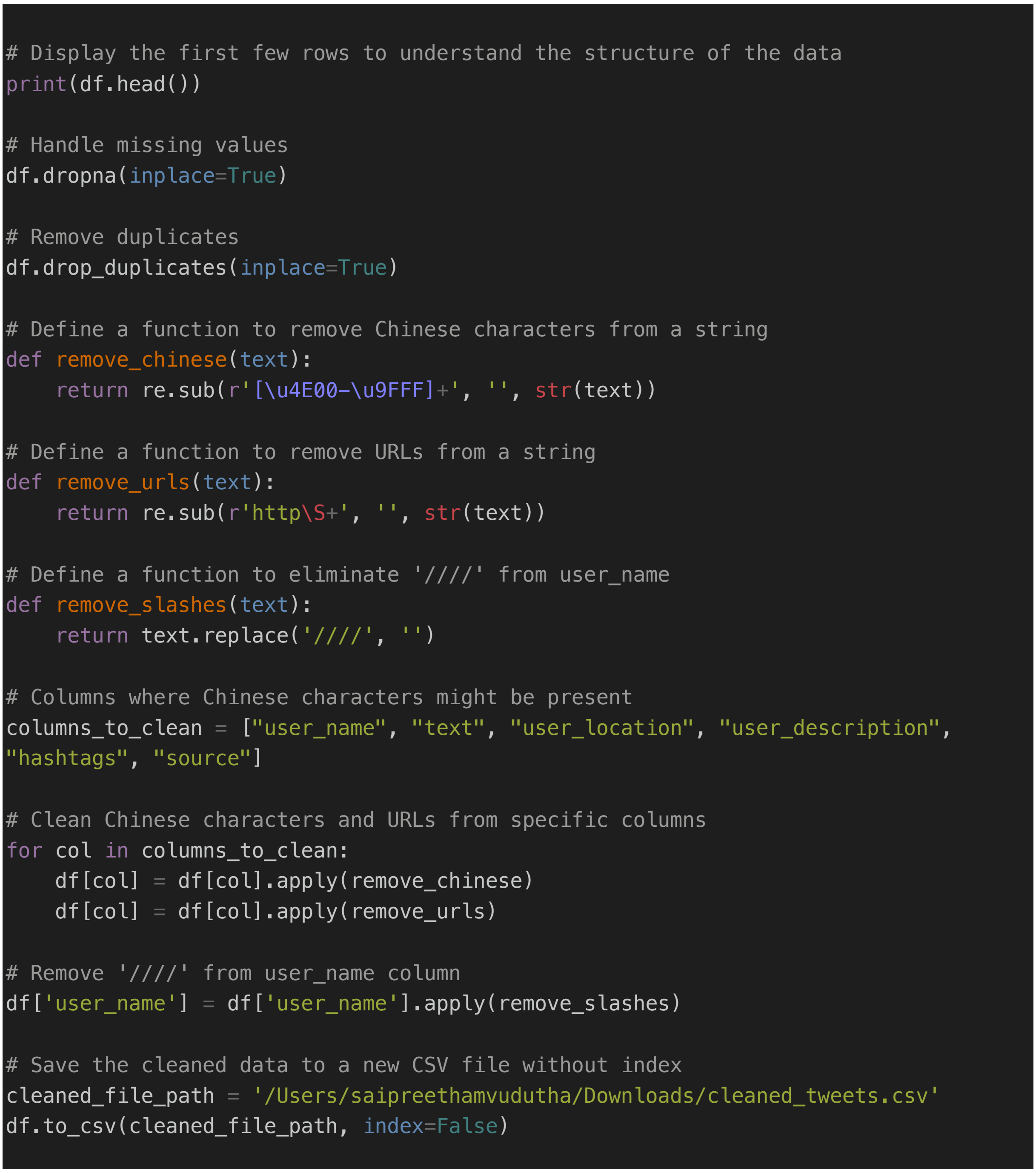
```
import pandas as pd
import re

# Define column names and their corresponding data types
columns = {
    "user_name": str,
    "text": str,
    "user_location": str,
    "user_description": str,
    "user_created": str,
    "user_followers": str,
    "user_friends": str,
    "user_favourites": str,
    "user_verified": str,
    "date": str,
    "hashtags": str,
    "source": str
}

# Load the dataset with specified column names and data types
file_path = '/Users/saipreethamvudutha/Downloads/tweets.csv'
df = pd.read_csv(file_path, names=columns.keys(), dtype=columns, skiprows=1)

# Display the first few rows to understand the structure of the data
print(df.head())

# Handle missing values
df.dropna(inplace=True)

# Remove duplicates
df.drop_duplicates(inplace=True)

# Define a function to remove Chinese characters from a string
def remove_chinese(text):
    return re.sub(r'[^\x00-\x7F\u4E00-\u9FFF]+', '', str(text))

# Columns where Chinese characters might be present
columns_to_clean = ["user_name", "text", "user_location", "user_description", "user_created","user_followers","user_
```



```
# Remove duplicates
df.drop_duplicates(inplace=True)

# Define a function to remove Chinese characters from a string
def remove_chinese(text):
    return re.sub(r'[^\x00-\x7F\u4E00-\u9FFF]+', '', str(text))

# Columns where Chinese characters might be present
columns_to_clean = ["user_name", "text", "user_location", "user_description", "user_created","user_followers","user_

# Clean Chinese characters from specific columns
for col in columns_to_clean:
    df[col] = df[col].apply(remove_chinese)

# Save the cleaned data to a new CSV file without index
cleaned_file_path = '/Users/saipreethamvudutha/Downloads/cleaned_tweets.csv'
df.to_csv(cleaned_file_path, index=False)
```



```
import pandas as pd
import re

# Define column names and their corresponding data types
columns = {
    "user_name": str,
    "text": str,
    "user_location": str,
    "user_description": str,
    "user_created": str,
    "user_followers": str,
    "user_friends": str,
    "user_favourites": str,
    "user_verified": str,
    "date": str,
    "hashtags": str,
    "source": str
}

# Load the dataset with specified column names and data types
file_path = '/Users/saipreethamvudutha/Downloads/tweets.csv'
df = pd.read_csv(file_path, names=columns.keys(), dtype=columns, skiprows=1)
```



```python
# Display the first few rows to understand the structure of the data
print(df.head())

# Handle missing values
df.dropna(inplace=True)

# Remove duplicates
df.drop_duplicates(inplace=True)

# Define a function to remove Chinese characters from a string
def remove_chinese(text):
    return re.sub(r'[\u4E00-\u9FFF]+', '', str(text))

# Define a function to remove URLs from a string
def remove_urls(text):
    return re.sub(r'http\S+', '', str(text))

# Define a function to eliminate '////' from user_name
def remove_slashes(text):
    return text.replace('////', '')

# Columns where Chinese characters might be present
columns_to_clean = ["user_name", "text", "user_location", "user_description",
"hashtags", "source"]

# Clean Chinese characters and URLs from specific columns
for col in columns_to_clean:
    df[col] = df[col].apply(remove_chinese)
    df[col] = df[col].apply(remove_urls)

# Remove '////' from user_name column
df['user_name'] = df['user_name'].apply(remove_slashes)

# Save the cleaned data to a new CSV file without index
cleaned_file_path = '/Users/saipreethamvudutha/Downloads/cleaned_tweets.csv'
df.to_csv(cleaned_file_path, index=False)
```

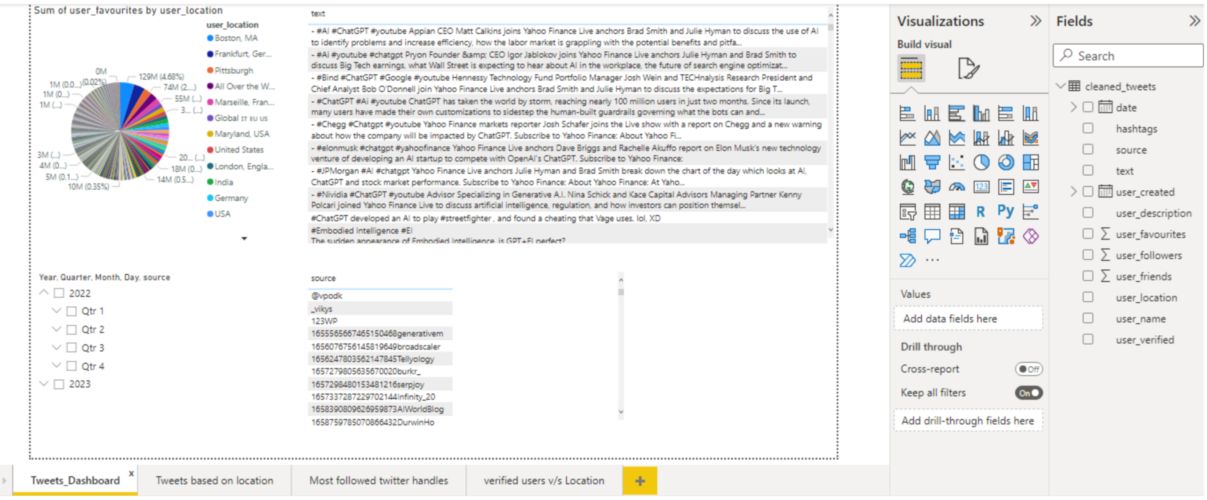
OUTPUT:

```
                                         username  \
0                                         Walee MENA
1                                            Dataiku
3                                    Lithium Systems
4                            Paramendra Kumar Bhagat

                                                text  \
0  #OpenAI has revealed its plan to launch #ChatG...
1  What are #LargeLanguageModels, how are they de...
2  apodecisionacious\natappear\nhe \ #Cha...
```

```
3  Business owner? Stay informed with the latest ...
4  ChatGPT: Motorbike For The Mind (13) #ChatGPT ...

                user_location  \
0                         UAE
1                New York, NY
2                Fayetteville
3  Based in Central Scotland
4                          NY

                                   user_description  \
0  OFFICIALLY IN MENA! We are the region's larges...
1  Dataiku is the only AI platform that connects ...
2  TG：https://t.co/C2kHIu7BKg 官网：https://t.co/56a...
...
1                           ['LargeLanguageModels']          HubSpot
2                           ['黑客', '合约', 'ChatGPT']  Twitter Web App
3                                               NaN   Hootsuite Inc.
4  ['ChatGPT', 'GPT4', 'AI', 'ArtificialIntellige...   Hootsuite Inc.
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...

POWER BI DASH BOARDS:



Sum of user_favourites by user_location

text

- #AI #ChatGPT #youtube Appian CEO Matt Calkins joins Yahoo Finance Live anchors Brad Smith and Julie Hyman to discuss the use of AI to identify problems and increase efficiency, how the labor market is grappling with the potential benefits and pitfa...
- #Ai #youtube #chatgpt Pryon Founder &amp; CEO Igor Jablokov joins Yahoo Finance Live anchors Julie Hyman and Brad Smith to discuss Big Tech earnings, what Wall Street is expecting to hear about AI in the workplace, the future of search engine optimizat...
- #Bind #ChatGPT #Google #youtube Hennessy Technology Fund Portfolio Manager Josh Wein and TECHnalysis Research President and Chief Analyst Bob O'Donnell join Yahoo Finance Live anchors Brad Smith and Julie Hyman to discuss the expectations for Big T...
- #ChatGPT #Ai #youtube ChatGPT has taken the world by storm, reaching nearly 100 million users in just two months. Since its launch, many users have made their own customizations to sidestep the human-built guardrails governing what the bots can and...
- #Chegg #Chatgpt #youtube Yahoo Finance markets reporter Josh Schafer joins the Live show with a report on Chegg and a new warning about how the company will be impacted by ChatGPT. Subscribe to Yahoo Finance: About Yahoo Fi...
- #elonmusk #chatgpt #yahoofinance Yahoo Finance Live anchors Dave Briggs and Rachelle Akuffo report on Elon Musk's new technology venture of developing an AI startup to compete with OpenAI's ChatGPT. Subscribe to Yahoo Finance:
- #JPMorgan #AI #chatgpt Yahoo Finance Live anchors Julie Hyman and Brad Smith break down the chart of the day which looks at AI, ChatGPT and stock market performance. Subscribe to Yahoo Finance: About Yahoo Finance: At Yaho...
- #Nividia #ChatGPT #youtube Advisor Specializing in Generative A.I.. Nina Schick and Kace Capital Advisors Managing Partner Kenny Polcari joined Yahoo Finance Live to discuss artificial intelligence, regulation, and how investors can position themsel...

#ChatGPT developed an AI to play #streetfighter , and found a cheating that Vage uses. lol, XD

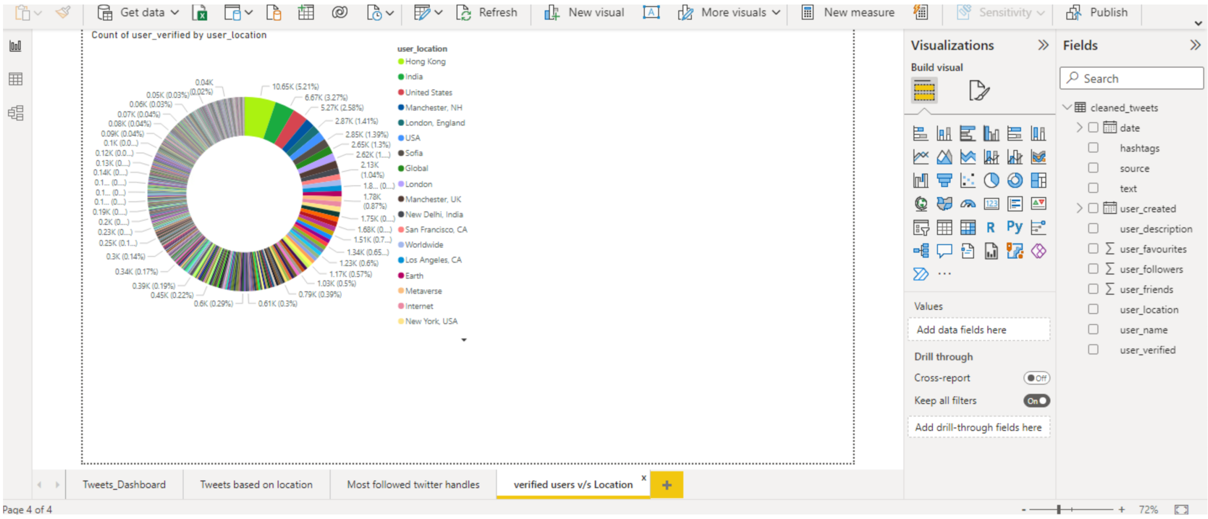
#Embodied Intelligence #EI
The sudden appearance of Embodied Intelligence, is GPT+EI perfect?

Year, Quarter, Month, Day, source

- □ 2022
  - □ Qtr 1
  - □ Qtr 2
  - □ Qtr 3
  - □ Qtr 4
- □ 2023

source

@vpodk
_vikys
123WP
1655565667465150468generativem
1656076756145819649broadscaler
1656247803562147845Tellyology
1657279805635670020burkr_
1657298480153481216serpjoy
1657337287229702144Infinity_20
1658390809626959873AIWorldBlog
1658759785070866432DurwinHo



Key Findings:

Hong Kong has the most numbers of verified twitter (now ‘x’) users.

Times Now is the most followed twitter handle.