Generate | a slider using jupyter widgets | Close

```
# Install required libraries
!pip install -q transformers datasets nltk

# Import libraries
import pandas as pd
import nltk
import re
from transformers import pipeline
nltk.download('punkt')
nltk.download('stopwords')

# Sample mock social media data (can be replaced with live data later)
data = {
    'username': ['user1', 'user2', 'user3'],
    'text': [
        "I love this product! It's absolutely amazing 😍",
        "I'm so frustrated with the service. Never again!",
        "Feeling kinda meh about today... not bad, not great."
    ]
}

# Create a DataFrame
df = pd.DataFrame(data)

# Function to clean text
def clean_text(text):
    text = re.sub(r'http\S+', '', text)  # remove URLs
    text = re.sub(r'@\w+', '', text)     # remove mentions
    text = re.sub(r'#\w+', '', text)     # remove hashtags
    text = re.sub(r'[^\w\s]', '', text)  # remove punctuation
    text = text.lower().strip()
    return text

df['cleaned_text'] = df['text'].apply(clean_text)

# Load a sentiment analysis pipeline
sentiment_analyzer = pipeline("text-classification", model="nateraw/bert-base-uncased-emotion")

# Apply sentiment analysis
df['emotion_result'] = df['cleaned_text'].apply(lambda x: sentiment_analyzer(x)[0])

# Extract emotion label and score
df['emotion'] = df['emotion_result'].apply(lambda x: x['label'])
df['confidence'] = df['emotion_result'].apply(lambda x: x['score'])

# Display results
df = df.drop(columns='emotion_result')
df
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as secre
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
  warnings.warn(
```

config.json: 100% 768/768 [00:00<00:00, 67.1kB/s]

pytorch_model.bin: 100% 438M/438M [00:03<00:00, 161MB/s]

tokenizer_config.json: 100% 252/252 [00:00<00:00, 23.2kB/s]

model.safetensors: 60% 262M/438M [00:01<00:01, 157MB/s]

vocab.txt: 100% 232k/232k [00:00<00:00, 4.21MB/s]

special_tokens_map.json: 100% 112/112 [00:00<00:00, 5.15kB/s]

```
Device set to use cpu
```

	username	text	cleaned_text	emotion	confidence
0	user1	I love this product! It's absolutely amazing 😍	i love this product its absolutely amazing	joy	0.677191
1	user2	I'm so frustrated with the service. Never again!	im so frustrated with the service never again	anger	0.990045
2	user3	Feeling kinda meh about today... not bad, not ...	feeling kinda meh about today not bad not great	joy	0.945631

Next steps: Generate code with df View recommended plots New interactive sheet