Welcome To Colab ⟳ Cannot save changes
File Edit View Insert Runtime Tools Help

🔍 Commands + Code ▾ + Text ▷ Run all ▾ Copy to Drive

Table of contents

+ Section

```python
import pandas as pd
import numpy as np

# Define a list of sample SaaS customer support tickets with intentional artifacts
tickets = [
    "I have a billing issue!!! Can someone help me? <br> My subscription was charged twice. Please refund ASAP!!",
    "login error... can't access my account with the password reset link.   pls fix it fast!!!",
    "Feature Request: Would love to see a dark mode option in the dashboard. #FeatureRequest #UI",
    "The app is very slow today. Is the server down? \n\n Checking from London office.",
    "Why was I charged 50 dollars instead of 30? Check my invoice #12345. Typos in billing details found.",
    "HELP! I cannot reset my passwrd. It keeps sayin 'error 404' when i click the link provided in the email...",
    "Just wanted to say the new update is grate! Keep up the good work. <p>Best regards, John.</p>",
    "I need to upgrde my plan to Enterprise but the paymant button is not workin.   @support",
    "Is there any API for integrateing with Slack? Need it for team notifications. $$$$",
    "Urgent: Account locked after 3 unsuccessful login attempts. I need access for a meeting in 10 minutes!!!!"
]

# Create a Pandas DataFrame from the list
df_tickets = pd.DataFrame(tickets, columns=['ticket_text'])

# Print the first few rows to inspect the raw synthetic data
print("First 5 rows of the synthetic SaaS ticket dataset:")
display(df_tickets.head())
```

First 5 rows of the synthetic SaaS ticket dataset:

	ticket_text
0	I have a billing issue!!! Can someone help me?...
1	login error... can't access my account with th...
2	Feature Request: Would love to see a dark mode...
3	The app is very slow today. Is the server down...
4	Why was I charged 50 dollars instead of 30? Ch...

```python
import re
from textblob import TextBlob

def clean_text(text):
```

⟨⟩ Variables ▣ Terminal ✦ ✓ 8:12 PM 🖳 Python 3

```python
import re
from textblob import TextBlob

def clean_text(text):
    # Remove HTML tags using regex
    text = re.sub(r'<.*?>', '', text)
    # Strip out non-alphanumeric special characters except for spaces and basic punctuation
    text = re.sub(r'[^a-zA-Z0-9\s\.\!\?]', '', text)
    # Normalize inconsistent whitespace
    text = re.sub(r'\s+', ' ', text).strip()
    return text

# 1. Apply noise removal
df_tickets['intermediate_clean'] = df_tickets['ticket_text'].apply(clean_text)

# 2. Perform spelling correction using TextBlob
# Note: .correct() can be slow but works for small synthetic datasets
df_tickets['cleaned_text'] = df_tickets['intermediate_clean'].apply(lambda x: str(TextBlob(x).correct()))

# Drop the intermediate column and display results
df_tickets.drop(columns=['intermediate_clean'], inplace=True)

print("Inspection of cleaned and spell-corrected text:")
display(df_tickets[['ticket_text', 'cleaned_text']].head())
```

Inspection of cleaned and spell-corrected text:

	ticket_text	cleaned_text
0	I have a billing issue!!! Can someone help me?...	I have a filling issue!!! An someone help me? ...
1	login error... can't access my account with th...	login error... can access my account with the ...
2	Feature Request: Would love to see a dark mode...	Feature Request Would love to see a dark mode ...
3	The app is very slow today. Is the server down...	The pp is very slow today. Is the server down?...
4	Why was I charged 50 dollars instead of 30? Ch...	Why was I charged 50 dollars instead of 30? Ch...

```python
import spacy

# load the small English model
```

Welcome To Colab ⚠ Cannot save changes

File Edit View Insert Runtime Tools Help

Q Commands + Code ▾ + Text ▷ Run all ▾ Copy to Drive

```python
import spacy

# Load the small English model
# In Colab, en_core_web_sm is typically pre-installed.
# If not, you can run: !python -m spacy download en_core_web_sm
nlp = spacy.load('en_core_web_sm')

def preprocess_nlp(text):
    # Process the text with spaCy
    doc = nlp(text)
    # Tokenize, lemmatize, and remove stop-words and punctuation
    tokens = [token.lemma_.lower() for token in doc if not token.is_stop and not token.is_punct]
    return " ".join(tokens)

# Apply the processing function to the cleaned_text column
df_tickets['processed_text'] = df_tickets['cleaned_text'].apply(preprocess_nlp)

print("Inspection of Lemmatization and Stop-word Removal:")
display(df_tickets[['cleaned_text', 'processed_text']].head())
```

Inspection of Lemmatization and Stop-word Removal:

	cleaned_text	processed_text
0	I have a filling issue!!! An someone help me? ...	filling issue help subscription charge twice t...
1	login error... can access my account with the ...	login error access account password rest link ...
2	Feature Request Would love to see a dark mode ...	feature request love dark mode option dashboar...
3	The pp is very slow today. Is the server down?...	pp slow today server check london office
4	Why was I charged 50 dollars instead of 30? Ch ...	charge 50 dollar instead 30 check voice 12345 ...

```python
def extract_entities(text):
    # Process the text with the existing spaCy 'nlp' object
    doc = nlp(text)
    # Return a list of tuples (entity text, label)
    return [(ent.text, ent.label_) for ent in doc.ents]

# Apply NER to the 'cleaned_text' column which retains case and structure
df_tickets['entities'] = df_tickets['cleaned_text'].apply(extract_entities)
```

{} Variables   ⌨ Terminal   ✓ 8:12 PM   Python 3

Welcome To Colab — Cannot save changes
File Edit View Insert Runtime Tools Help

Q Commands + Code ▾ + Text ▷ Run all ▾ Copy to Drive

RAM / Disk

Table of contents

+ Section

```python
def extract_entities(text):
    # Process the text with the existing spaCy 'nlp' object
    doc = nlp(text)
    # Return a list of tuples (entity text, label)
    return [(ent.text, ent.label_) for ent in doc.ents]

# Apply NER to the 'cleaned_text' column which retains case and structure
df_tickets['entities'] = df_tickets['cleaned_text'].apply(extract_entities)

print("Inspection of Named Entity Recognition (NER) results:")
display(df_tickets[['cleaned_text', 'entities']])
```

Inspection of Named Entity Recognition (NER) results:

	cleaned_text	entities
0	I have a filling issue!!! An someone help me? ...	[]
1	login error... can access my account with the ...	[]
2	Feature Request Would love to see a dark mode ...	[]
3	The pp is very slow today. Is the server down?...	[(today, DATE), (London, GPE)]
4	Why was I charged 50 dollars instead of 30? Ch...	[(50 dollars, MONEY), (30, CARDINAL), (12345, ...
5	HELP! I cannot rest my password. It keeps sayi...	[(404, CARDINAL)]
6	Must wanted to say the new update is grate! De...	[(John, PERSON)]
7	I need to upgrade my plan to Enterprise but th...	[]
8	Is there any API for integrating with Black? N...	[(API, ORG)]
9	Urgent Account locked after 3 unsuccessful log...	[(3, CARDINAL), (10 minutes, TIME)]

```python
def label_support_tier(text):
    # Convert text to lowercase for keyword matching
    text = text.lower()

    # Define keyword-based rules for categorization
    if any(keyword in text for keyword in ['charge', 'refund', 'dollar', 'payment', 'billing', 'invoice']):
        return 'Billing'
    elif any(keyword in text for keyword in ['login', 'error', 'server', 'api', 'access', 'password']):
        return 'Technical'
```

{} Variables ▭ Terminal

✓ 8:12 PM Python 3

```python
def label_support_tier(text):
    # Convert text to lowercase for keyword matching
    text = text.lower()

    # Define keyword-based rules for categorization
    if any(keyword in text for keyword in ['charge', 'refund', 'dollar', 'payment', 'billing', 'invoice']):
        return 'Billing'
    elif any(keyword in text for keyword in ['login', 'error', 'server', 'api', 'access', 'password']):
        return 'Technical'
    elif any(keyword in text for keyword in ['feature', 'request', 'dashboard', 'ui']):
        return 'Feature Request'
    else:
        return 'General Inquiry'

# Apply the labeling function to the 'processed_text' column
df_tickets['support_tier'] = df_tickets['processed_text'].apply(label_support_tier)

print("Inspection of Rule-Based Annotation results:")
display(df_tickets[['processed_text', 'support_tier']])
```

Inspection of Rule-Based Annotation results:

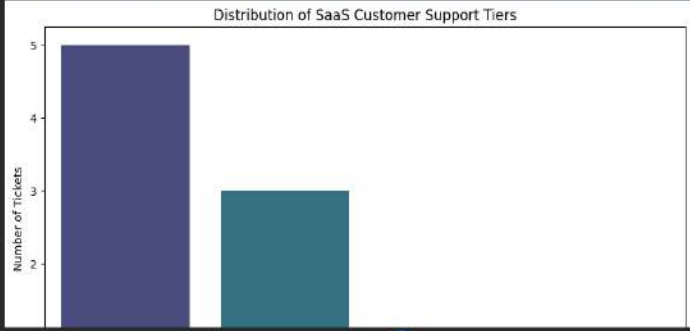|	processed_text	support_tier
0	filling issue help subscription charge twice r...	Billing
1	login error access account password rest link ...	Technical
2	feature request love dark mode option dashboar...	Feature Request
3	pp slow today server check london office	Technical
4	charge 50 dollar instead 30 check voice 12345 ...	Billing
5	help rest password keep say error 404 click li...	Technical
6	want new update grate deep good work west rega...	General Inquiry
7	need upgrade plan enterprise payment button wo...	Billing
8	api integrate black need team modification	Technical
9	urgent account lock 3 unsuccessful login attem...	Technical

```python
import matplotlib.pyplot as plt
import seaborn as sns

# 1. Visualize the distribution of Support Tiers
plt.figure(figsize=(10, 6))
sns.countplot(data=df_tickets, x='support_tier', palette='viridis', order=df_tickets['support_tier'].value_counts().index)
plt.title('Distribution of SaaS Customer Support Tiers')
plt.xlabel('Support Tier')
plt.ylabel('Number of Tickets')
plt.xticks(rotation=45)
plt.show()

# 2. Display the final processed dataset
print("Final Processed and Annotated Dataset:")
display(df_tickets[['ticket_text', 'processed_text', 'entities', 'support_tier']])
```
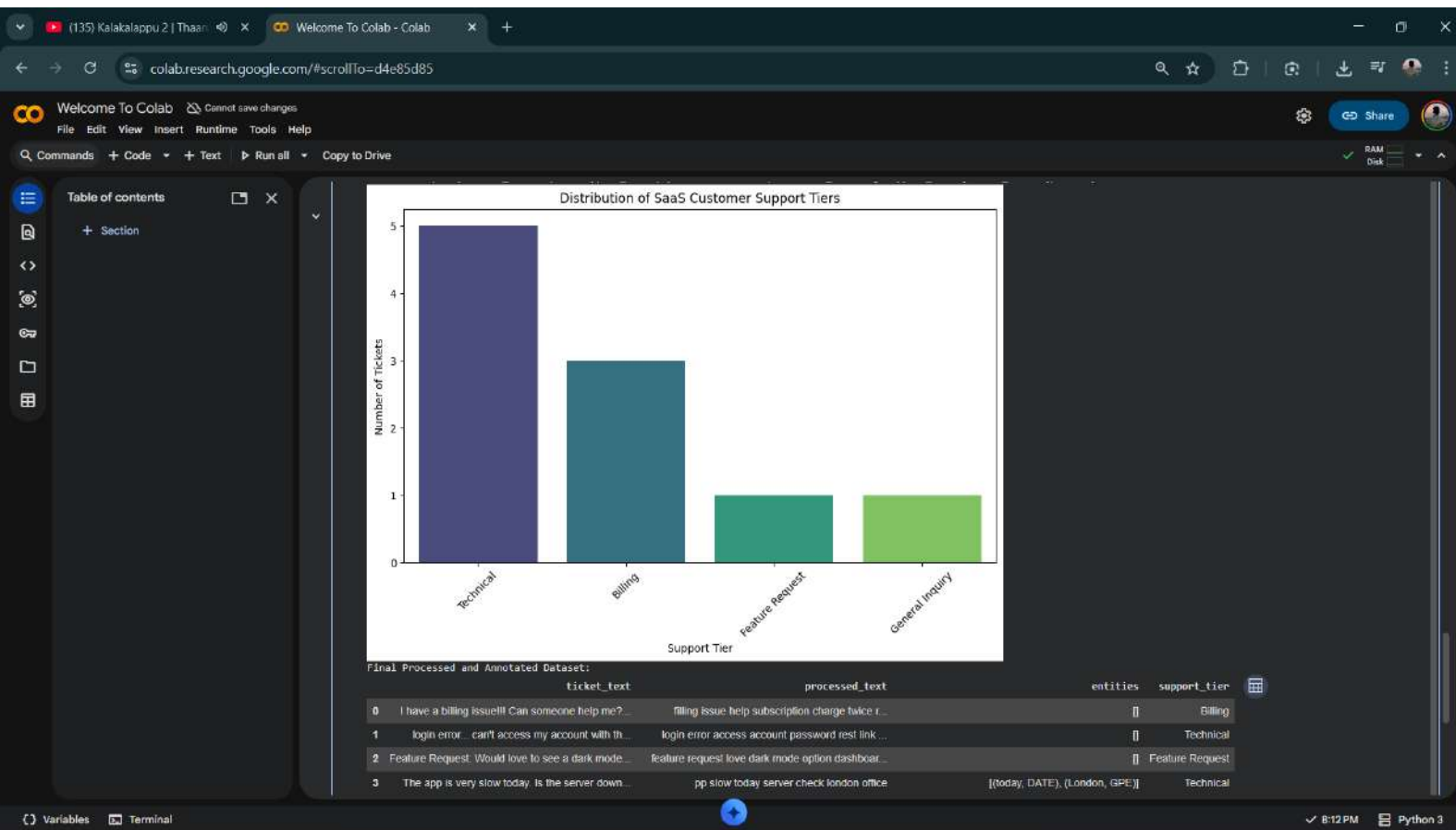
/tmp/ipython-input-2372884293.py:6: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.countplot(data=df_tickets, x='support_tier', palette='viridis', order=df_tickets['support_tier'].value_counts().index)

Distribution of SaaS Customer Support Tiers

Distribution of SaaS Customer Support Tiers

Final Processed and Annotated Dataset:

	ticket_text	processed_text	entities	support_tier
0	I have a billing issue!!! Can someone help me?...	billing issue help subscription charge twice r...	[]	Billing
1	login error... can't access my account with th...	login error access account password rest link ...	[]	Technical
2	Feature Request: Would love to see a dark mode...	feature request love dark mode option dashboar...	[]	Feature Request
3	The app is very slow today. Is the server down...	pp slow today server check london office	[(today, DATE), (London, GPE)]	Technical

Welcome To Colab — Cannot save changes
File Edit View Insert Runtime Tools Help

Commands + Code ▾ + Text ▷ Run all ▾ Copy to Drive

Table of contents

+ Section

Technical   Billing   Feature Request   General Inquiry

Support Tier

Final Processed and Annotated Dataset:

	ticket_text	processed_text	entities	support_tier
0	I have a billing issue!!! Can someone help me?...	filling issue help subscription charge twice r...	[]	Billing
1	login error... can't access my account with th...	login error access account password rest link ...	[]	Technical
2	Feature Request: Would love to see a dark mode...	feature request love dark mode option dashboar...	[]	Feature Request
3	The app is very slow today. Is the server down...	pp slow today server check london office	[(today, DATE), (London, GPE)]	Technical
4	Why was I charged 50 dollars instead of 30? Ch...	charge 50 dollar instead 30 check voice 12345 ...	[(50 dollars, MONEY), (30, CARDINAL), (12345, ...	Billing
5	HELP! I cannot reset my passwrd. It keeps sayi...	help rest password keep say error 404 click li...	[(404, CARDINAL)]	Technical
6	Just wanted to say the new update is grate! Ke...	want new update grate deep good work west rega...	[(John, PERSON)]	General Inquiry
7	I need to upgrde my plan to Enterprise but the...	need upgrade plan enterprise payment button wo...	[]	Billing
8	Is there any API for integrateing with Slack? ...	api integrate black need team modification	[(API, ORG)]	Technical
9	Urgent: Account locked after 3 unsuccessful lo...	urgent account lock 3 unsuccessful login attem...	[(3, CARDINAL), (10 minutes, TIME)]	Technical

Variables   Terminal   8:12 PM   Python 3