```
In [1]:  !pip install spacy nltk pandas textblob matplotlib seaborn
         !python -m spacy download en_core_web_sm
```

```
Requirement already satisfied: spacy in c:\users\sugen\anaconda3\lib\site-packa
ges (3.8.11)
Requirement already satisfied: nltk in c:\users\sugen\anaconda3\lib\site-packag
es (3.9.1)
Requirement already satisfied: pandas in c:\users\sugen\anaconda3\lib\site-pack
ages (2.2.2)
Requirement already satisfied: textblob in c:\users\sugen\anaconda3\lib\site-pa
ckages (0.19.0)
Requirement already satisfied: matplotlib in c:\users\sugen\anaconda3\lib\site-
packages (3.9.2)
Requirement already satisfied: seaborn in c:\users\sugen\anaconda3\lib\site-pac
kages (0.13.2)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in c:\users\sugen\an
aconda3\lib\site-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in c:\users\sugen\an
aconda3\lib\site-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\sugen\anac
onda3\lib\site-packages (from spacy) (1.0.15)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\sugen\anaconda3\
lib\site-packages (from spacy) (2.0.13)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\sugen\anaconda
3\lib\site-packages (from spacy) (3.0.12)
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in c:\users\sugen\anaconda3\
lib\site-packages (from spacy) (8.3.10)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in c:\users\sugen\anaconda
3\lib\site-packages (from spacy) (1.1.3)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in c:\users\sugen\anaconda3\
lib\site-packages (from spacy) (2.5.2)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in c:\users\sugen\anacon
da3\lib\site-packages (from spacy) (2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.4.2 in c:\users\sugen\anaconda
3\lib\site-packages (from spacy) (0.4.3)
Requirement already satisfied: typer-slim<1.0.0,>=0.3.0 in c:\users\sugen\anaco
nda3\lib\site-packages (from spacy) (0.21.1)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\sugen\anaconda3\
lib\site-packages (from spacy) (4.66.5)
Requirement already satisfied: numpy>=1.19.0 in c:\users\sugen\anaconda3\lib\si
te-packages (from spacy) (1.26.4)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\sugen\anacon
da3\lib\site-packages (from spacy) (2.32.3)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in c:\user
s\sugen\anaconda3\lib\site-packages (from spacy) (2.8.2)
Requirement already satisfied: jinja2 in c:\users\sugen\anaconda3\lib\site-pack
ages (from spacy) (3.1.4)
Requirement already satisfied: setuptools in c:\users\sugen\anaconda3\lib\site-
packages (from spacy) (75.1.0)
Requirement already satisfied: packaging>=20.0 in c:\users\sugen\anaconda3\lib\
site-packages (from spacy) (24.1)
Requirement already satisfied: click in c:\users\sugen\anaconda3\lib\site-packa
ges (from nltk) (8.1.7)
Requirement already satisfied: joblib in c:\users\sugen\anaconda3\lib\site-pack
ages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in c:\users\sugen\anaconda3\lib\
site-packages (from nltk) (2024.9.11)
```

```
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\sugen\anacond
a3\lib\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\sugen\anaconda3\lib\sit
e-packages (from pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\sugen\anaconda3\lib\s
ite-packages (from pandas) (2023.3)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\sugen\anaconda3\li
b\site-packages (from matplotlib) (1.2.0)
Requirement already satisfied: cycler>=0.10 in c:\users\sugen\anaconda3\lib\sit
e-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\sugen\anaconda3\li
b\site-packages (from matplotlib) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\sugen\anaconda3\li
b\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: pillow>=8 in c:\users\sugen\anaconda3\lib\site-p
ackages (from matplotlib) (10.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\sugen\anaconda3\li
b\site-packages (from matplotlib) (3.1.2)
Requirement already satisfied: annotated-types>=0.4.0 in c:\users\sugen\anacond
a3\lib\site-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy) (0.6.0)
Requirement already satisfied: pydantic-core==2.20.1 in c:\users\sugen\anaconda
3\lib\site-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy) (2.20.1)
Requirement already satisfied: typing-extensions>=4.6.1 in c:\users\sugen\anaco
nda3\lib\site-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy) (4.1
1.0)
Requirement already satisfied: six>=1.5 in c:\users\sugen\anaconda3\lib\site-pa
ckages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\sugen\anaco
nda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in c:\users\sugen\anaconda3\lib\sit
e-packages (from requests<3.0.0,>=2.13.0->spacy) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\sugen\anaconda3\l
ib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\sugen\anaconda3\l
ib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (2024.8.30)
Requirement already satisfied: blis<1.4.0,>=1.3.0 in c:\users\sugen\anaconda3\l
ib\site-packages (from thinc<8.4.0,>=8.3.4->spacy) (1.3.3)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in c:\users\sugen\anaco
nda3\lib\site-packages (from thinc<8.4.0,>=8.3.4->spacy) (0.1.5)
Requirement already satisfied: colorama in c:\users\sugen\anaconda3\lib\site-pa
ckages (from tqdm<5.0.0,>=4.38.0->spacy) (0.4.6)
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in c:\users\sugen\ana
conda3\lib\site-packages (from weasel<0.5.0,>=0.4.2->spacy) (0.23.0)
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in c:\users\sugen\anaco
nda3\lib\site-packages (from weasel<0.5.0,>=0.4.2->spacy) (5.2.1)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\sugen\anaconda3\lib\
site-packages (from jinja2->spacy) (2.1.3)
Collecting en-core-web-sm==3.8.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en_co
re_web_sm-3.8.0/en_core_web_sm-3.8.0-py3-none-any.whl (12.8 MB)
     ------------------------------------- 0.0/12.8 MB ? eta -:--:--
     ------------------------------------- 0.0/12.8 MB ? eta -:--:--
      ------------------------------------ 0.3/12.8 MB ? eta -:--:--
     - ----------------------------------- 0.5/12.8 MB 1.5 MB/s eta 0:00:09
```

```
   --- ----------------------------------- 1.0/12.8 MB 1.6 MB/s eta 0:00:08
   ---- ---------------------------------- 1.6/12.8 MB 1.8 MB/s eta 0:00:07
   ------- ------------------------------- 2.4/12.8 MB 2.2 MB/s eta 0:00:05
   --------- ----------------------------- 2.9/12.8 MB 2.4 MB/s eta 0:00:05
   ---------- ---------------------------- 3.7/12.8 MB 2.5 MB/s eta 0:00:04
   ----------- --------------------------- 4.5/12.8 MB 2.7 MB/s eta 0:00:04
   --------------- ----------------------- 5.8/12.8 MB 3.0 MB/s eta 0:00:03
   ------------------ -------------------- 6.8/12.8 MB 3.3 MB/s eta 0:00:02
   ------------------------ -------------- 8.1/12.8 MB 3.5 MB/s eta 0:00:02
   ---------------------------- ---------- 9.7/12.8 MB 3.8 MB/s eta 0:00:01
   --------------------------------- ----- 11.0/12.8 MB 4.0 MB/s eta 0:00:01
   ----------------------------------- --- 11.8/12.8 MB 4.1 MB/s eta 0:00:01
   -------------------------------------- -- 12.1/12.8 MB 4.0 MB/s eta 0:00:01
   -------------------------------------- - 12.3/12.8 MB 3.8 MB/s eta 0:00:01
   ---------------------------------------- 12.6/12.8 MB 3.6 MB/s eta 0:00:01
   ---------------------------------------- 12.8/12.8 MB 3.5 MB/s eta 0:00:00
[+] Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
```

In [2]:
```python
import re
import pandas as pd
import nltk
import spacy
import matplotlib.pyplot as plt
import seaborn as sns

from nltk.corpus import stopwords
from textblob import TextBlob
```

In [3]:
```python
nlp = spacy.load("en_core_web_sm")
stop_words = set(stopwords.words("english"))
```

In [4]:
```python
df = pd.read_csv("customer_support_tickets.csv")
df.head()
```

Out[4]:

	Ticket ID	Customer Name	Customer Email	Customer Age	Customer Gender	Produc Purchase
0	1	Marisa Obrien	carrollallison@example.com	32	Other	GoPro Her
1	2	Jessica Rios	clarkeashley@example.com	42	Female	LG Sma T
2	3	Christopher Robbins	gonzalestracy@example.com	48	Other	Dell XP
3	4	Christina Dillon	bradleyolson@example.org	27	Female	Microso Offic
4	5	Alexander Carroll	bradleymark@example.com	67	Female	Autodes AutoCA

In [5]:
```python
df.columns = df.columns.str.lower()
df = df.rename(columns={
    "ticket description": "raw_text",
    "ticket type": "category"
})

df = df[["raw_text", "category"]]
df.head()
```

Out[5]:

	raw_text	category
0	I'm having an issue with the {product_purchase...	Technical issue
1	I'm having an issue with the {product_purchase...	Technical issue
2	I'm facing a problem with my {product_purchase...	Technical issue
3	I'm having an issue with the {product_purchase...	Billing inquiry
4	I'm having an issue with the {product_purchase...	Billing inquiry

In [6]:
```python
print("Dataset Shape:", df.shape)
print("\nMissing Values:\n")
print(df.isnull().sum())
```

```
Dataset Shape: (8469, 2)

Missing Values:

raw_text    0
category    0
dtype: int64
```

In [7]:
```python
def clean_text(text):
    text = str(text).lower()
    text = re.sub(r"http\S+|www\S+", "", text)
    text = re.sub(r"[^a-zA-Z\s]", "", text)
    text = re.sub(r"\s+", " ", text).strip()
    return text
```

In [8]:
```python
def preprocess_text(text):
    cleaned = clean_text(text)
    doc = nlp(cleaned)

    tokens = [
        token.lemma_
        for token in doc
        if token.text not in stop_words and not token.is_punct
    ]

    sentiment = TextBlob(cleaned).sentiment.polarity
    return " ".join(tokens), sentiment
```

In [10]:
```python
# Take a small representative sample (VERY IMPORTANT)
df_sample = df.sample(300, random_state=42).reset_index(drop=True)

df_sample.shape
```

Out[10]: (300, 2)

In [11]:
```python
import spacy

# Disable slow components
nlp = spacy.load(
    "en_core_web_sm",
    disable=["ner", "parser", "textcat"]
)
```

In [12]:
```python
def fast_preprocess(text):
    text = str(text).lower()
    text = re.sub(r"http\S+|www\S+", "", text)
    text = re.sub(r"[^a-zA-Z\s]", "", text)
    text = re.sub(r"\s+", " ", text).strip()

    doc = nlp(text)
    return " ".join(
        token.lemma_
```

```
        for token in doc
        if token.text not in stop_words and not token.is_punct
    )
```

In [13]:
```
df_sample["processed_text"] = df_sample["raw_text"].apply(fast_preprocess)
df_sample.head()
```

Out[13]:

	raw_text	category	processed_text
0	I'm having an issue with the {product_purchase...	Refund request	issue productpurchased please assist use xdade...
1	I'm having trouble connecting my {product_purc...	Product inquiry	trouble connect productpurchased home wifi net...
2	I'm having an issue with the {product_purchase...	Billing inquiry	issue productpurchased please assist please gi...
3	I'm having an issue with the {product_purchase...	Billing inquiry	issue productpurchased please assist check com...
4	I'm having an issue with the {product_purchase...	Refund request	issue productpurchased please assist would lik...

In [14]:
```
df_sample["sentiment"] = df_sample["raw_text"].apply(
    lambda x: TextBlob(str(x)).sentiment.polarity
)

df_sample.head()
```

Out[14]:

	raw_text	category	processed_text	sentiment
0	I'm having an issue with the {product_purchase...	Refund request	issue productpurchased please assist use xdade...	0.000000
1	I'm having trouble connecting my {product_purc...	Product inquiry	trouble connect productpurchased home wifi net...	0.122917
2	I'm having an issue with the {product_purchase...	Billing inquiry	issue productpurchased please assist please gi...	0.500000
3	I'm having an issue with the {product_purchase...	Billing inquiry	issue productpurchased please assist check com...	-0.156250
4	I'm having an issue with the {product_purchase...	Refund request	issue productpurchased please assist would lik...	0.066667

In [15]:
```
nlp_ner = spacy.load("en_core_web_sm")

def extract_entities(text):
    doc = nlp_ner(str(text))
    return [(ent.text, ent.label_) for ent in doc.ents]
```

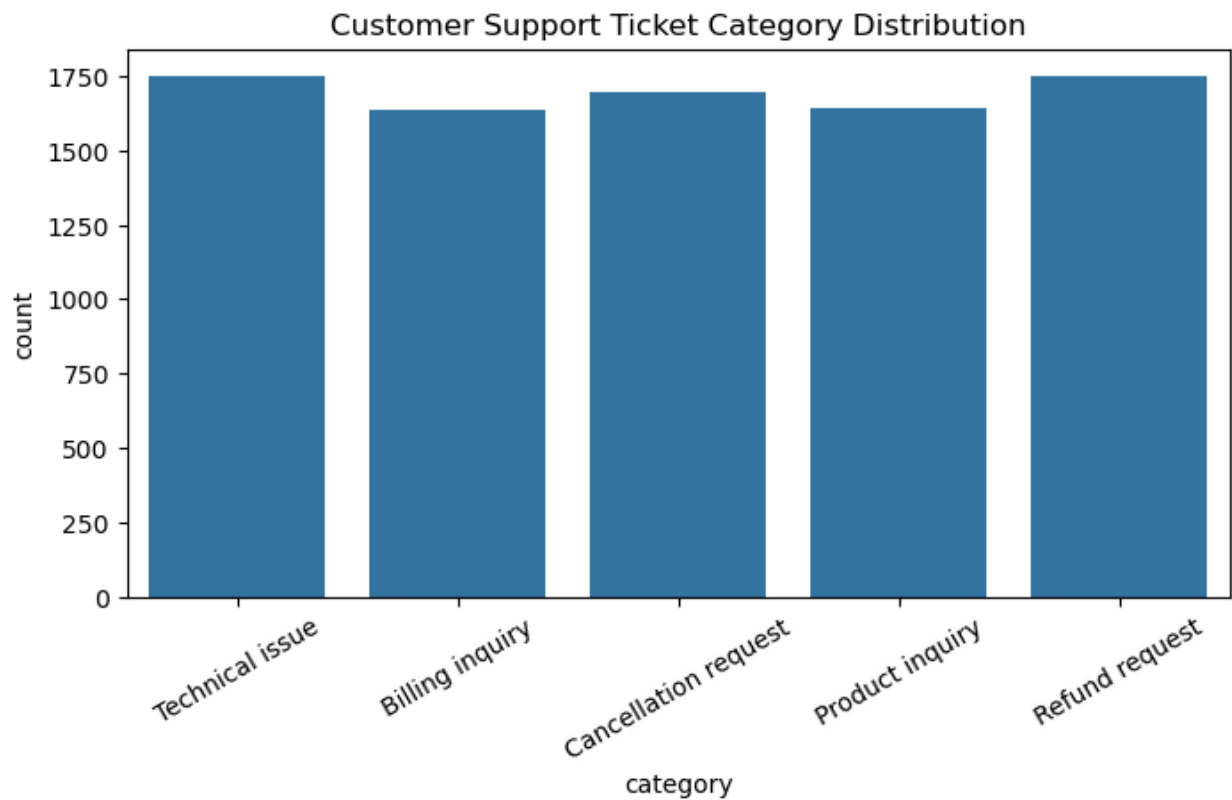In [16]:
```
df_sample["entities"] = None
```

```
df_sample.loc[:10, "entities"] = df_sample.loc[:10, "raw_text"].apply(extract_

df_sample.head(12)
```
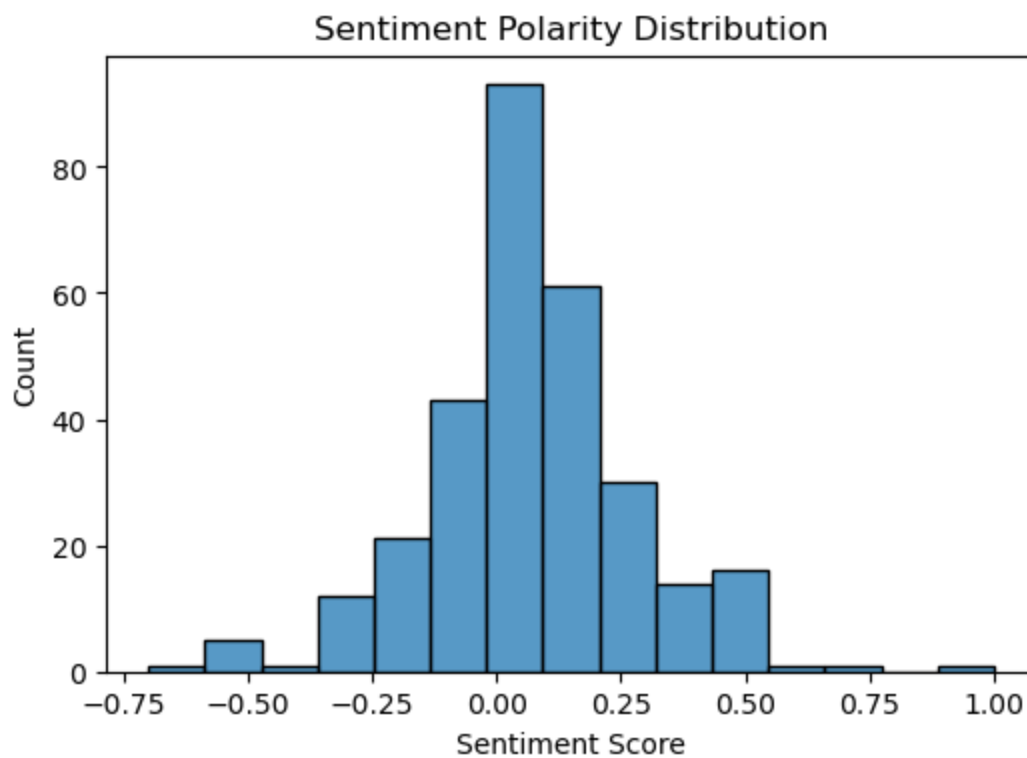
	raw_text	category	processed_text	sentiment	entities
0	I'm having an issue with the {product_purchase...	Refund request	issue productpurchased please assist use xdade...	0.000000	[]
1	I'm having trouble connecting my {product_purc...	Product inquiry	trouble connect productpurchased home wifi net...	0.122917	[]
2	I'm having an issue with the {product_purchase...	Billing inquiry	issue productpurchased please assist please gi...	0.500000	[]
3	I'm having an issue with the {product_purchase...	Billing inquiry	issue productpurchased please assist check com...	-0.156250	[(4, CARDINAL)]
4	I'm having an issue with the {product_purchase...	Refund request	issue productpurchased please assist would lik...	0.066667	[]
5	I've encountered a data loss issue with my {pr...	Refund request	encounter data loss issue productpurchase file...	0.150000	[]
6	I'm having an issue with the {product_purchase...	Product inquiry	issue productpurchased please assist notice pe...	-0.312500	[]
7	I'm having an issue with the {product_purchase...	Refund request	issue productpurchased please assist trouble p...	0.000000	[]
8	I'm encountering a software bug in the {produc...	Refund request	encounter software bug productpurchased whenev...	0.250000	[]
9	I'm having an issue with the {product_purchase...	Product inquiry	issue productpurchased please assist thank tha...	0.200000	[]
10	I'm having an issue with the {product_purchase...	Product inquiry	issue productpurchased please assist product s...	0.000000	[]
11	I've encountered a data loss issue with my {pr...	Cancellation request	encounter data loss issue productpurchase file...	0.208333	None

In [17]:
```python
plt.figure(figsize=(8,4))
sns.countplot(x="category", data=df)
plt.title("Customer Support Ticket Category Distribution")
plt.xticks(rotation=30)
plt.show()
```


In [19]:
```python
plt.figure(figsize=(6,4))
sns.histplot(df_sample["sentiment"], bins=15)
plt.title("Sentiment Polarity Distribution")
plt.xlabel("Sentiment Score")
plt.show()
```

Sentiment Polarity Distribution


```
In [20]:  sample_text = df.iloc[0]["raw_text"]

          processed, sentiment = preprocess_text(sample_text)
          entities = extract_entities(sample_text)

          print("ORIGINAL TICKET:\n", sample_text)
          print("\nPROCESSED TEXT:\n", processed)
          print("\nSENTIMENT SCORE:\n", sentiment)
          print("\nNAMED ENTITIES:\n", entities)
```

```
ORIGINAL TICKET:
 I'm having an issue with the {product_purchased}. Please assist.

Your billing zip code is: 71701.

We appreciate that you have requested a website address.

Please double check your email address. I've tried troubleshooting steps mentio
ned in the user manual, but the issue persists.

PROCESSED TEXT:
 issue productpurchased please assist billing zip code appreciate request websi
te address please double check email address try troubleshoot step mention user
manual issue persist

SENTIMENT SCORE:
 0.0

NAMED ENTITIES:
 [('71701', 'CARDINAL')]
```

```
In [21]: df.to_csv("processed_customer_support_tickets.csv", index=False)
         print("Processed dataset saved successfully!")
```

Processed dataset saved successfully!

```
In [ ]:
```