```
!pip install nltk
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk) (1.3.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk) (2023.12.25)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk) (4.66.2)
```

```
!pip install bertopic
```

```
Requirement already satisfied: bertopic in /usr/local/lib/python3.10/dist-packages (0.16.0)
Requirement already satisfied: numpy>=1.20.0 in /usr/local/lib/python3.10/dist-packages (from bertopic) (1.25.2)
Requirement already satisfied: hdbscan>=0.8.29 in /usr/local/lib/python3.10/dist-packages (from bertopic) (0.8.33)
Requirement already satisfied: umap-learn>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from bertopic) (0.5.5)
Requirement already satisfied: pandas>=1.1.5 in /usr/local/lib/python3.10/dist-packages (from bertopic) (1.5.3)
Requirement already satisfied: scikit-learn>=0.22.2.post1 in /usr/local/lib/python3.10/dist-packages (from bertopic) (1.2.2)
Requirement already satisfied: tqdm>=4.41.1 in /usr/local/lib/python3.10/dist-packages (from bertopic) (4.66.2)
Requirement already satisfied: sentence-transformers>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from bertopic) (2.6.1)
Requirement already satisfied: plotly>=4.7.0 in /usr/local/lib/python3.10/dist-packages (from bertopic) (5.15.0)
Requirement already satisfied: cython<3,>=0.27 in /usr/local/lib/python3.10/dist-packages (from hdbscan>=0.8.29->bertopic) (0.29.37
Requirement already satisfied: scipy>=1.0 in /usr/local/lib/python3.10/dist-packages (from hdbscan>=0.8.29->bertopic) (1.11.4)
Requirement already satisfied: joblib>=1.0 in /usr/local/lib/python3.10/dist-packages (from hdbscan>=0.8.29->bertopic) (1.3.2)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.1.5->bertopic) (2.8
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.1.5->bertopic) (2023.4)
Requirement already satisfied: tenacity>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from plotly>=4.7.0->bertopic) (8.2.3)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from plotly>=4.7.0->bertopic) (24.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn>=0.22.2.post1->ber
Requirement already satisfied: transformers<5.0.0,>=4.32.0 in /usr/local/lib/python3.10/dist-packages (from sentence-transformers>=0
Requirement already satisfied: torch>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from sentence-transformers>=0.4.1->bertopic
Requirement already satisfied: huggingface-hub>=0.15.1 in /usr/local/lib/python3.10/dist-packages (from sentence-transformers>=0.4.1
Requirement already satisfied: Pillow in /usr/local/lib/python3.10/dist-packages (from sentence-transformers>=0.4.1->bertopic) (9.4
Requirement already satisfied: numba>=0.51.2 in /usr/local/lib/python3.10/dist-packages (from umap-learn>=0.5.0->bertopic) (0.58.1)
Requirement already satisfied: pynndescent>=0.5 in /usr/local/lib/python3.10/dist-packages (from umap-learn>=0.5.0->bertopic) (0.5.1
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-transform
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-t
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-transform
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-transf
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->
Requirement already satisfied: llvmlite<0.42,>=0.41.0dev0 in /usr/local/lib/python3.10/dist-packages (from numba>=0.51.2->umap-learn
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.1->pandas>=1.1.5->bert
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence-transformers>=0.4.1->b
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence-transformers>=0.4.1
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence-transformers>=0.4.1->
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sent
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->se
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sent
Requirement already satisfied: nvidia-cudnn-cu12==8.9.2.26 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence
Requirement already satisfied: nvidia-cublas-cu12==12.1.3.1 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence
Requirement already satisfied: nvidia-cufft-cu12==11.0.2.54 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence
Requirement already satisfied: nvidia-curand-cu12==10.3.2.106 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->senter
Requirement already satisfied: nvidia-cusolver-cu12==11.4.5.107 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sent
Requirement already satisfied: nvidia-cusparse-cu12==12.1.0.106 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sent
Requirement already satisfied: nvidia-nccl-cu12==2.19.3 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence-tra
Requirement already satisfied: nvidia-nvtx-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence-t
Requirement already satisfied: triton==2.2.0 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence-transformers>=
Requirement already satisfied: nvidia-nvjitlink-cu12 in /usr/local/lib/python3.10/dist-packages (from nvidia-cusolver-cu12==11.4.5.1
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers<5.0.0,>=4.32.0->sente
Requirement already satisfied: tokenizers<0.19,>=0.14 in /usr/local/lib/python3.10/dist-packages (from transformers<5.0.0,>=4.32.0->
Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from transformers<5.0.0,>=4.32.0->sent
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2->torch>=1.11.0->sentence-tran
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub>=
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub>=0.15.1->sent
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub>=0.15.1
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub>=0.15.1
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-packages (from sympy->torch>=1.11.0->sentence-transfor
```

```
!pip install bertopic[visualization]
```

```
Requirement already satisfied: bertopic[visualization] in /usr/local/lib/python3.10/dist-packages (0.16.0)
WARNING: bertopic 0.16.0 does not provide the extra 'visualization'
Requirement already satisfied: numpy>=1.20.0 in /usr/local/lib/python3.10/dist-packages (from bertopic[visualization]) (1.25.2)
Requirement already satisfied: hdbscan>=0.8.29 in /usr/local/lib/python3.10/dist-packages (from bertopic[visualization]) (0.8.33)
Requirement already satisfied: umap-learn>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from bertopic[visualization]) (0.5.5
Requirement already satisfied: pandas>=1.1.5 in /usr/local/lib/python3.10/dist-packages (from bertopic[visualization]) (1.5.3)
Requirement already satisfied: scikit-learn>=0.22.2.post1 in /usr/local/lib/python3.10/dist-packages (from bertopic[visualization
Requirement already satisfied: tqdm>=4.41.1 in /usr/local/lib/python3.10/dist-packages (from bertopic[visualization]) (4.66.2)
Requirement already satisfied: sentence-transformers>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from bertopic[visualizati
Requirement already satisfied: plotly>=4.7.0 in /usr/local/lib/python3.10/dist-packages (from bertopic[visualization]) (5.15.0)
Requirement already satisfied: cython<3,>=0.27 in /usr/local/lib/python3.10/dist-packages (from hdbscan>=0.8.29->bertopic[visuali
Requirement already satisfied: scipy>=1.0 in /usr/local/lib/python3.10/dist-packages (from hdbscan>=0.8.29->bertopic[visualizatio
Requirement already satisfied: joblib>=1.0 in /usr/local/lib/python3.10/dist-packages (from hdbscan>=0.8.29->bertopic[visualizati
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.1.5->bertopic[vi
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.1.5->bertopic[visualizatio
Requirement already satisfied: tenacity>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from plotly>=4.7.0->bertopic[visualiza
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from plotly>=4.7.0->bertopic[visualization])
```

```
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn>=0.22.2.post1->
Requirement already satisfied: transformers<5.0.0,>=4.32.0 in /usr/local/lib/python3.10/dist-packages (from sentence-transformers
Requirement already satisfied: torch>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from sentence-transformers>=0.4.1->berto
Requirement already satisfied: huggingface-hub>=0.15.1 in /usr/local/lib/python3.10/dist-packages (from sentence-transformers>=0.
Requirement already satisfied: Pillow in /usr/local/lib/python3.10/dist-packages (from sentence-transformers>=0.4.1->bertopic[vis
Requirement already satisfied: numba>=0.51.2 in /usr/local/lib/python3.10/dist-packages (from umap-learn>=0.5.0->bertopic[visuali
Requirement already satisfied: pynndescent>=0.5 in /usr/local/lib/python3.10/dist-packages (from umap-learn>=0.5.0->bertopic[visu
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-transf
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentenc
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-transf
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-tra
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.
Requirement already satisfied: llvmlite<0.42,>=0.41.0dev0 in /usr/local/lib/python3.10/dist-packages (from numba>=0.51.2->umap-le
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.1->pandas>=1.1.5->b
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence-transformers>=0.4.1
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence-transformers>=0.
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence-transformers>=0.4.
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->s
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0-
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->s
Requirement already satisfied: nvidia-cudnn-cu12==8.9.2.26 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->senten
Requirement already satisfied: nvidia-cublas-cu12==12.1.3.1 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sente
Requirement already satisfied: nvidia-cufft-cu12==11.0.2.54 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sente
Requirement already satisfied: nvidia-curand-cu12==10.3.2.106 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sen
Requirement already satisfied: nvidia-cusolver-cu12==11.4.5.107 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->s
Requirement already satisfied: nvidia-cusparse-cu12==12.1.0.106 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->s
Requirement already satisfied: nvidia-nccl-cu12==2.19.3 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence-
Requirement already satisfied: nvidia-nvtx-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentenc
Requirement already satisfied: triton==2.2.0 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11.0->sentence-transformer
Requirement already satisfied: nvidia-nvjitlink-cu12 in /usr/local/lib/python3.10/dist-packages (from nvidia-cusolver-cu12==11.4.
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers<5.0.0,>=4.32.0->se
Requirement already satisfied: tokenizers<0.19,>=0.14 in /usr/local/lib/python3.10/dist-packages (from transformers<5.0.0,>=4.32.
Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from transformers<5.0.0,>=4.32.0->s
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2->torch>=1.11.0->sentence-t
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hu
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub>=0.15.1->s
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub>=0.1
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub>=0.1
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-packages (from sympy->torch>=1.11.0->sentence-trans
```

```
from bertopic import BERTopic
import nltk
import os
import pandas as pd
from tqdm.notebook import tqdm
import matplotlib.pyplot as plt
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.preprocessing import LabelEncoder
from nltk.stem import PorterStemmer
from nltk.stem import WordNetLemmatizer
from nltk import word_tokenize
import numpy as np
from gensim.models import Nmf
from gensim.models.ldamodel import LdaModel
from gensim.models import LsiModel
from gensim.models import HdpModel
from gensim.models import LdaMulticore
from gensim.models import EnsembleLda
from gensim.models.ldamulticore import LdaMulticore
from gensim.models.coherencemodel import CoherenceModel
import gensim
import gensim.corpora as corpora
from gensim.models import Phrases
from gensim.utils import simple_preprocess
from gensim.corpora.dictionary import Dictionary
from gensim.utils import tokenize
from nltk.corpus import stopwords
from collections import Counter
nltk.download('stopwords')
nltk.download('wordnet')
import re
import json
import string
from bs4 import BeautifulSoup
from gensim.parsing.preprocessing import (
    strip_non_alphanum,
    split_alphanum,
    strip_short,
    strip_numeric
)
from collections import Counter
from wordcloud import WordCloud, STOPWORDS
from PIL import Image
```

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

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

```
df=pd.read_csv("latest_ticket_data.csv")
```

```
df.head()
```

	Description	Category
0	hi since recruiter lead permission approve req...	Application
1	re expire days hi ask help update passwords co...	Application
2	verification warning hi has got attached pleas...	Application
3	please dear looks blacklisted receiving mails ...	Application
4	dear modules report report cost thank much req...	Application

```
#Assign nan in place of blanks in the body column
df[df.loc[:, 'Description'] == ''] = np.nan
```

```
# Check if blank values still exist
df[df.loc[:, 'Description'] == '']
```

	Description	Category

```
df.shape
```

```
(3000, 2)
```

```
#Remove all rows where body column is nan
df = df[~df['Description'].isnull()]
```

```
df.shape
```

```
(3000, 2)
```

```
# Convert body column to string for performing text operations
df['Description'] = df['Description'].astype(str)
```

```
stop_words = stopwords.words('english')
custom_stop_words = ['hi', 'since', 'please', 'best', 'regards', 'thank', 'thanks', 'hello', 'sent', 'great', 'dear', 'help', 'kind']
time_words = ['january', 'february', 'march', 'april', 'may', 'june', 'july', 'august', 'september', 'october', 'november', 'december',

def remove_stop_words(text):
    pattern = re.compile(r'\b(' + r'|'.join(stop_words) + r')\b\s*')
    text = pattern.sub('', text)
    return text

def remove_custom_words(text):
    pattern = re.compile(r'\b(' + r'|'.join(custom_stop_words) + r')\b\s*')
    text = pattern.sub('', text)
    return text

def remove_time_words(text):
    pattern = re.compile(r'\b(' + r'|'.join(time_words) + r')\b\s*')
    text = pattern.sub('', text)
    return text

df['Description'] = df["Description"].map(lambda x: remove_stop_words(x))
df['Description'] = df["Description"].map(lambda x: remove_custom_words(x))
df['Description'] = df["Description"].map(lambda x: remove_time_words(x))
```

```
def parse_html(text, parser="html.parser"):
    soup = BeautifulSoup(text, parser)
    soup = remove_html_tags(soup)
    text = remove_multiple_space(soup.get_text()).strip()
    return text
```

```
def parse_html_v2(text, loop=2, parser="html.parser"):
```

```
    if not text:
        text = ""
    # some contents still have html code after first parse
    # loop solved problem
    for _ in range(loop):
        soup = BeautifulSoup(text, parser)
        text = soup.get_text()
    text = remove_multiple_space(text)
    return text


def remove_links_content(text):
    text = re.sub(r"http\S+", "", text)
    return text


def remove_emails(text):
    return re.sub('\S*@\S*\s?', '', text)  # noqa


def remove_punctuation(text):
    """https://stackoverflow.com/a/37221663"""
    table = str.maketrans({key: None for key in string.punctuation})
    return text.translate(table)


def remove_special_tags(text):
    """Remove html tags from a string"""
    clean = re.compile('{.*?}')
    return re.sub(clean, '', text)


def preprocess_text(text):
    text = parse_html_v2(text)
    text = text.lower()
    text = remove_links_content(text)
    text = remove_emails(text)
    text = remove_special_tags(text)  # remove content between {}
    text = remove_punctuation(text)  # remove all puntuations
    text = split_alphanum(text)  # add space between word and numeric
    text = strip_numeric(text)  # remove digits
    text = strip_non_alphanum(text)  # remove non-alphabetic characters
    text = strip_short(text, minsize=2)  # remove word with length  <  minsize
    text = remove_multiple_space(text).strip()  # remove space and strip
    #text = tokenize(text)
    return text


def remove_multiple_space(text):
    return re.sub("\s\s+", " ", text)  # noqa


def remove_html_tags(soup,
                     tags=["script", "style"],
                     get_text=False):
    for tag in tags:
        for sample in soup.find_all(tag):
            sample.replaceWith('')

    if get_text:
        return soup.get_text()
    return soup

# Convert body column to string for performing text operations
df['Description'] = df['Description'].astype(str)
df['Description'] = df["Description"].map(lambda x: preprocess_text(x))


lemmatizer = WordNetLemmatizer()
df['Lemma_Description'] = df["Description"].map(lambda x: lemmatizer.lemmatize(x))


# create model
bert_model = BERTopic(verbose=True)
#convert to list
docs = df.Lemma_Description.to_list()


#bert_model.fit_transform(docs)
topics, probabilities = bert_model.fit_transform(docs)
```

```
2024-03-31 09:01:21,702 - BERTopic - Embedding - Transforming documents to embeddings.
modules.json: 100%                                                  349/349 [00:00<00:00, 13.7kB/s]
config_sentence_transformers.json: 100%                             116/116 [00:00<00:00, 4.32kB/s]
README.md: 100%                                                     10.7k/10.7k [00:00<00:00, 210kB/s]
sentence_bert_config.json: 100%                                     53.0/53.0 [00:00<00:00, 1.62kB/s]
config.json: 100%                                                   612/612 [00:00<00:00, 20.9kB/s]
model.safetensors: 100%                                             90.9M/90.9M [00:00<00:00, 209MB/s]
tokenizer_config.json: 100%                                         350/350 [00:00<00:00, 22.0kB/s]
vocab.txt: 100%                                                     232k/232k [00:00<00:00, 586kB/s]
tokenizer.json: 100%                                                466k/466k [00:00<00:00, 1.19MB/s]
special_tokens_map.json: 100%                                       112/112 [00:00<00:00, 1.54kB/s]
1_Pooling/config.json: 100%                                         190/190 [00:00<00:00, 4.04kB/s]
Batches: 100%                                                       94/94 [01:11<00:00,  5.59it/s]
2024-03-31 09:02:43,229 - BERTopic - Embedding - Completed ✓
2024-03-31 09:02:43,232 - BERTopic - Dimensionality - Fitting the dimensionality reduction algorithm
2024-03-31 09:03:06,568 - BERTopic - Dimensionality - Completed ✓
2024-03-31 09:03:06,570 - BERTopic - Cluster - Start clustering the reduced embeddings
2024-03-31 09:03:06,776 - BERTopic - Cluster - Completed ✓
2024-03-31 09:03:06,788 - BERTopic - Representation - Extracting topics from clusters using representation models.
2024-03-31 09:03:07,161 - BERTopic - Representation - Completed ✓
```

```
# Preprocess documents
cleaned_docs = bert_model._preprocess_text(docs)

# Extract vectorizer and tokenizer from BERTopic
vectorizer = bert_model.vectorizer_model
tokenizer = vectorizer.build_tokenizer()

# Extract features for Topic Coherence evaluation
words = vectorizer.get_feature_names_out()
tokens = [tokenizer(doc) for doc in cleaned_docs]
dictionary = corpora.Dictionary(tokens)
corpus = [dictionary.doc2bow(token) for token in tokens]
topic_words = [[words for words, _ in bert_model.get_topic(topic)]
               for topic in range(len(set(topics))-1)]


# Evaluate
coherence_model = CoherenceModel(topics=topic_words,
                                 texts=tokens,
                                 corpus=corpus,
                                 dictionary=dictionary,
                                 coherence='c_v')
coherence = coherence_model.get_coherence()
coherence
```

```
0.5108303713239383
```

```
# Evaluate
u_mass_coherence_model = CoherenceModel(topics=topic_words,
                                 texts=tokens,
                                 corpus=corpus,
                                 dictionary=dictionary,
                                 coherence='u_mass')
u_mass_coherence = u_mass_coherence_model.get_coherence()
u_mass_coherence
```

```
-6.442755937630485
```

```
# Evaluate
c_uci_coherence_model = CoherenceModel(topics=topic_words,
                                 texts=tokens,
                                 corpus=corpus,
                                 dictionary=dictionary,
                                 coherence='c_uci')
c_uci_coherence = c_uci_coherence_model.get_coherence()
c_uci_coherence
```

```
-2.620747508656331
```

```
# Evaluate
c_npmi_coherence_model = CoherenceModel(topics=topic_words,
```

```
                                texts=tokens,
                                corpus=corpus,
                                dictionary=dictionary,
                                coherence='c_npmi')
c_npmi_coherence = c_npmi_coherence_model.get_coherence()
c_npmi_coherence
```

```
0.063172328595613
```

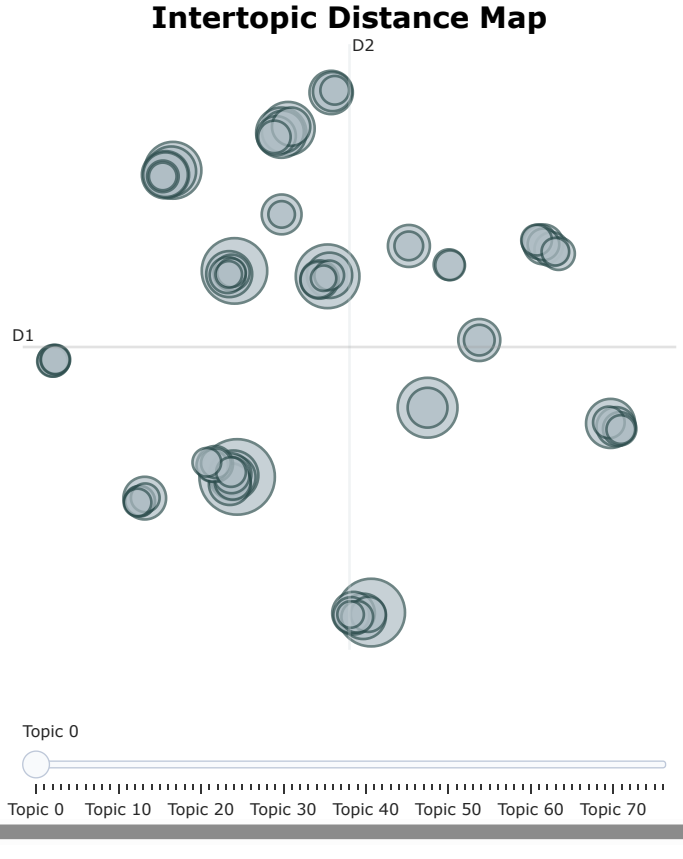
```
bert_model.get_topic_freq().head(11)
```

	Topic	Count
6	-1	963
24	0	97
25	1	78
47	2	73
23	3	69
56	4	61
14	5	55
34	6	48
4	7	43
27	8	42
0	9	42

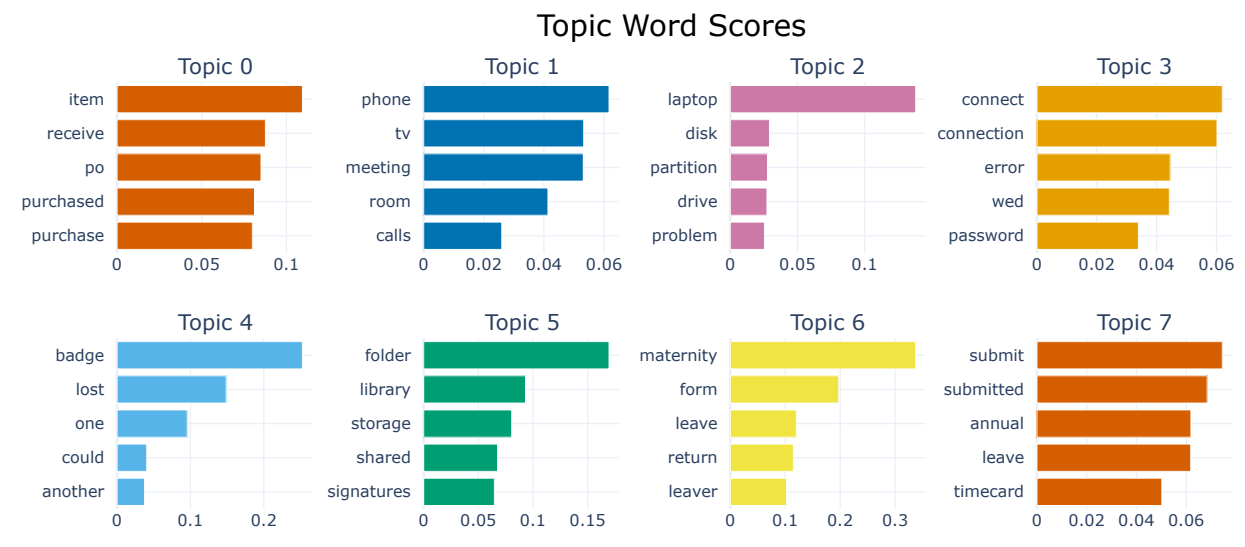
```
bert_model.get_topic(7)
```

```
[('submit', 0.0744794167125373),
 ('submitted', 0.06840681754853037),
 ('annual', 0.06187634278482901),
 ('leave', 0.061796572029398235),
 ('timecard', 0.050173290746146224),
 ('entered', 0.048411094710755935),
 ('absence', 0.042266296006762474),
 ('reports', 0.039900748999711112),
 ('action', 0.038729547333533725),
 ('accordingly', 0.038079113557479165)]
```

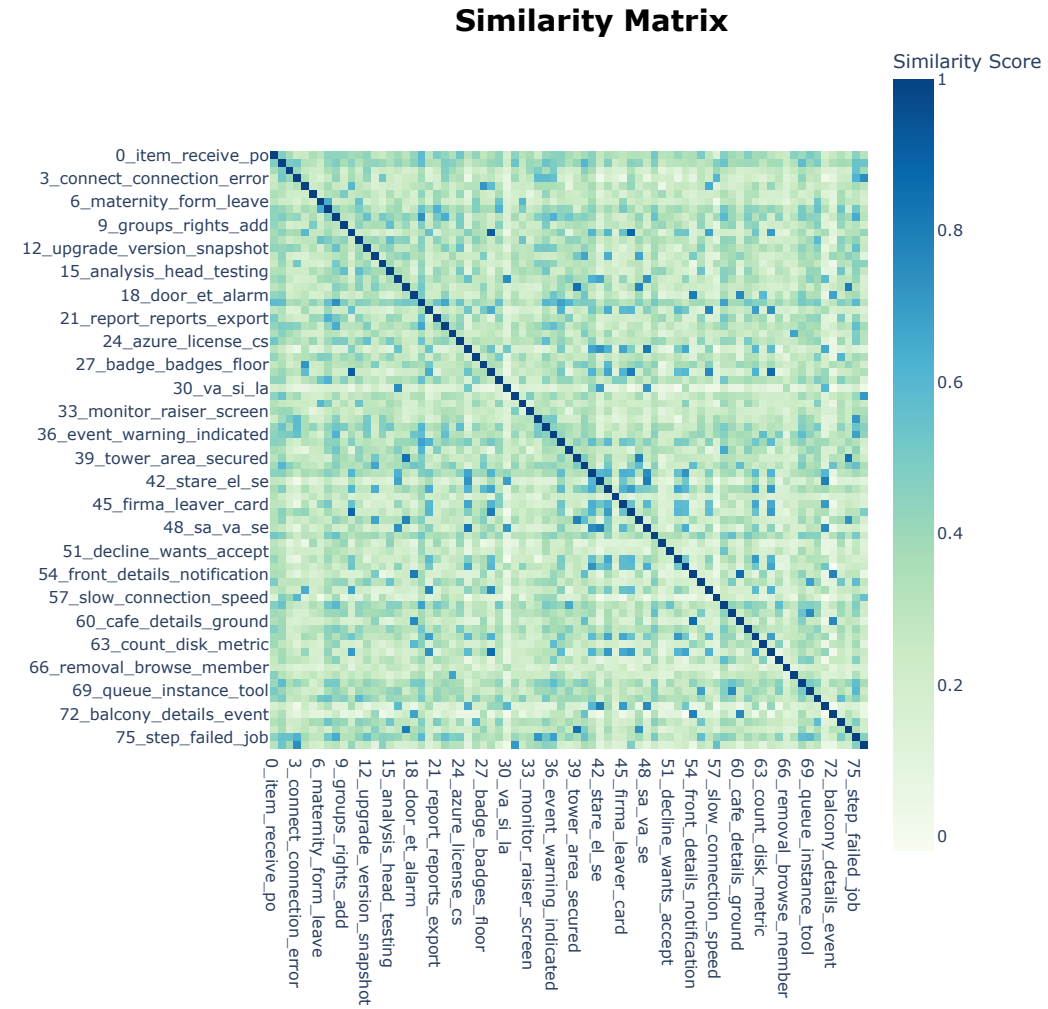
```
bert_model.visualize_topics()
```



```
bert_model.visualize_barchart()
```




```
bert_model.visualize_heatmap()
```




```
bert_model.save("latest_ticket_data_bert_model")
```

```
2024-03-31 09:03:18,219 - BERTopic - WARNING: When you use `pickle` to save/load a BERTopic model,please make sure that the environm
```

Start coding or generate with AI.