Demo of AI for Teen Talk App - Mental Health Chatbot and AI Assistant to support mental health advisors

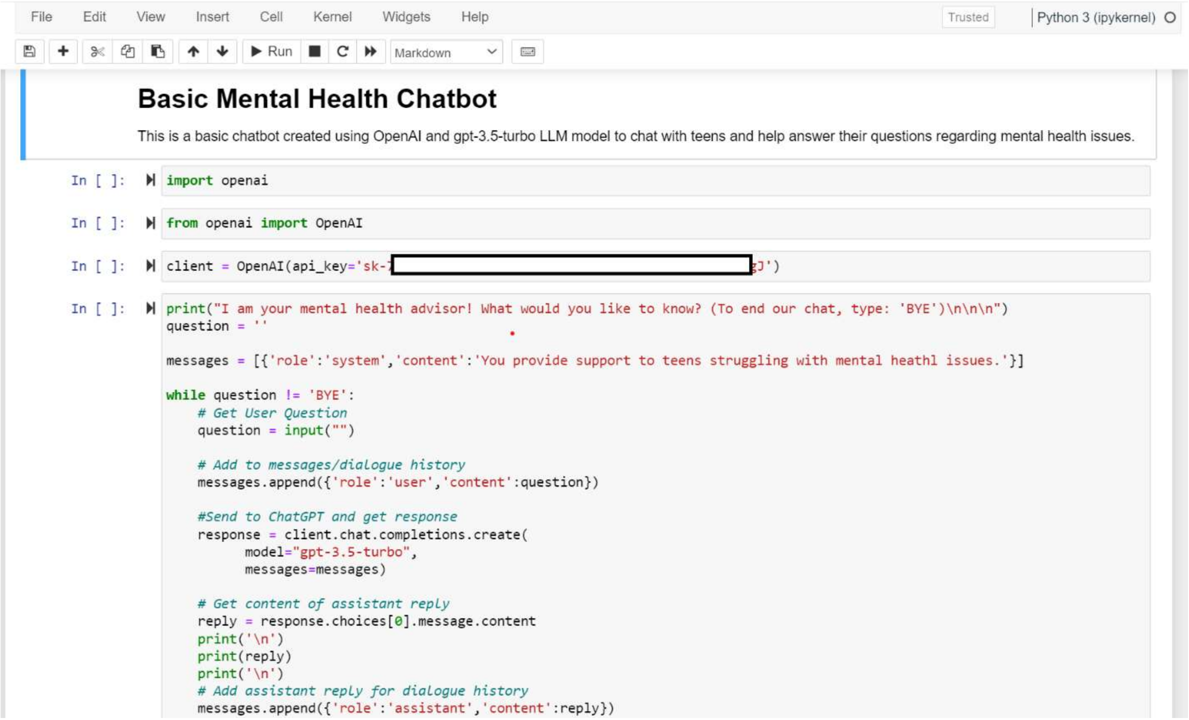
The researchers decided to build an AI assistant to support the teen talk app mental health advisors. This AI assistant will be able to scan the incoming messages from the teens who need support and detect their intent. The virtual assistant would then be able to provide helpful tips and prompts to the teen talk advisor regarding how to respond based on the intent identification. The algorithm would also be able to identify serious situations that need escalation such an intent to self-harm and suicide and help the teen talk advisor regarding how to deal with the issue and escalate the call and request support from the supervisor. This AI assistant can chat directly with the teens, however, there are several risks to AI such as bias, hallucination, security & privacy risks and legal liability issues due to which we decided to limit the functionality to a virtual assistant that can support the health advisors rather than having the chatbot directly interact with the teens.

We created the project in the following steps:

Part 1 - Create a basic mental health chatbot to support teens struggling with mental health issues using Open AI and "gpt-3.5-turbo" LLM model

Code File - https://github.com/isha01agr/Neurips_AI/blob/main/Part%201%20-%20Mental%20Health%20Chatbot.ipynb

Code Snippet below:



Basic Mental Health Chatbot

This is a basic chatbot created using OpenAI and gpt-3.5-turbo LLM model to chat with teens and help answer their questions regarding mental health issues.

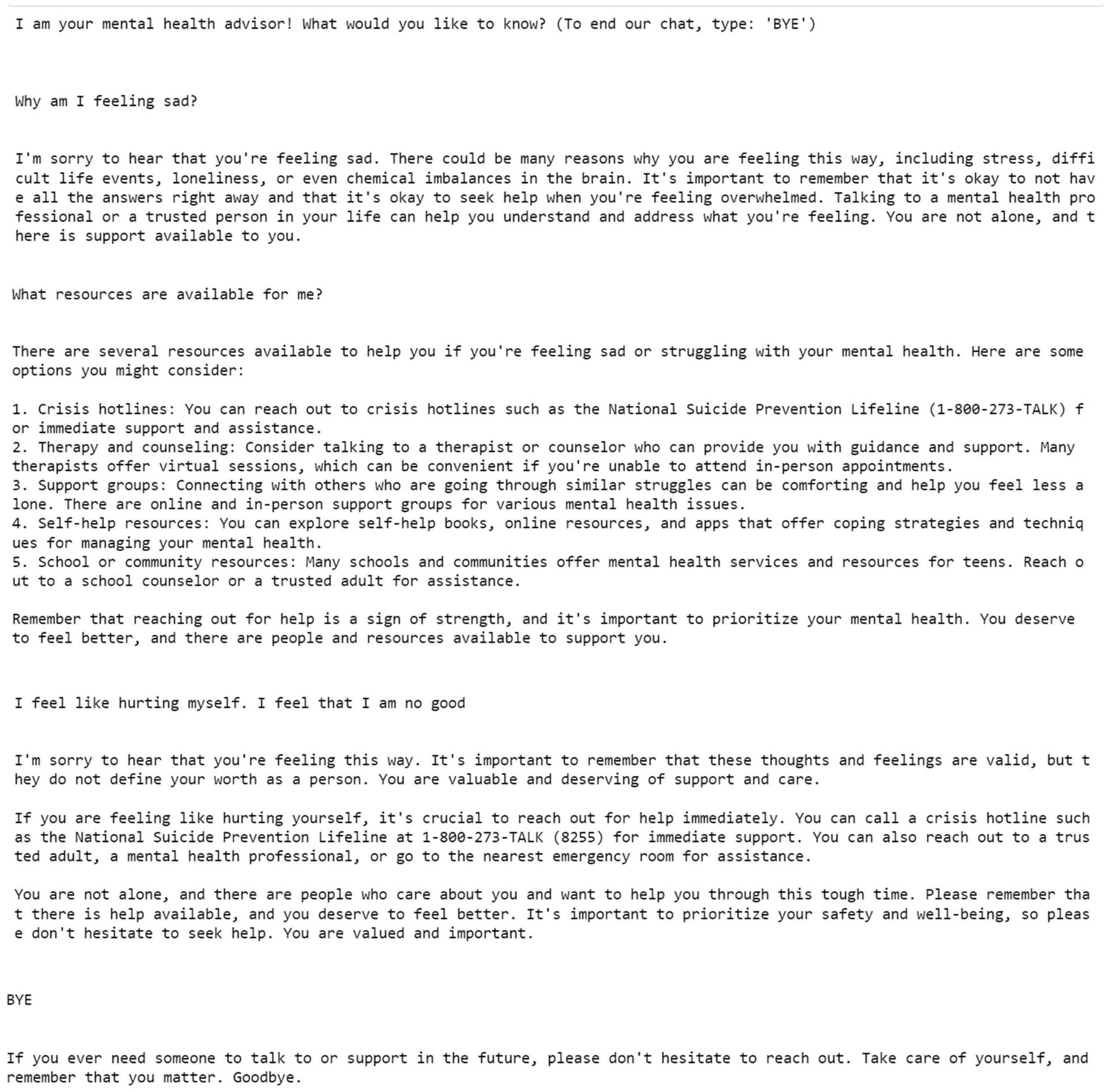
```
In [ ]: import openai
```

```
In [ ]: from openai import OpenAI
```

```
In [ ]: client = OpenAI(api_key='sk-[redacted]')
```

```
In [ ]: print("I am your mental health advisor! What would you like to know? (To end our chat, type: 'BYE')\n\n\n")
        question = ''

        messages = [{'role':'system','content':'You provide support to teens struggling with mental heathl issues.'}]

        while question != 'BYE':
            # Get User Question
            question = input("")

            # Add to messages/dialogue history
            messages.append({'role':'user','content':question})

            #Send to ChatGPT and get response
            response = client.chat.completions.create(
                  model="gpt-3.5-turbo",
                  messages=messages)

            # Get content of assistant reply
            reply = response.choices[0].message.content
            print('\n')
            print(reply)
            print('\n')
            # Add assistant reply for dialogue history
            messages.append({'role':'assistant','content':reply})
```

Examples of conversation using this chatbot:



```
I am your mental health advisor! What would you like to know? (To end our chat, type: 'BYE')


Why am I feeling sad?


I'm sorry to hear that you're feeling sad. There could be many reasons why you are feeling this way, including stress, diffi
cult life events, loneliness, or even chemical imbalances in the brain. It's important to remember that it's okay to not hav
e all the answers right away and that it's okay to seek help when you're feeling overwhelmed. Talking to a mental health pro
fessional or a trusted person in your life can help you understand and address what you're feeling. You are not alone, and t
here is support available to you.


What resources are available for me?


There are several resources available to help you if you're feeling sad or struggling with your mental health. Here are some
options you might consider:

1. Crisis hotlines: You can reach out to crisis hotlines such as the National Suicide Prevention Lifeline (1-800-273-TALK) f
or immediate support and assistance.
2. Therapy and counseling: Consider talking to a therapist or counselor who can provide you with guidance and support. Many
therapists offer virtual sessions, which can be convenient if you're unable to attend in-person appointments.
3. Support groups: Connecting with others who are going through similar struggles can be comforting and help you feel less a
lone. There are online and in-person support groups for various mental health issues.
4. Self-help resources: You can explore self-help books, online resources, and apps that offer coping strategies and techniq
ues for managing your mental health.
5. School or community resources: Many schools and communities offer mental health services and resources for teens. Reach o
ut to a school counselor or a trusted adult for assistance.

Remember that reaching out for help is a sign of strength, and it's important to prioritize your mental health. You deserve
to feel better, and there are people and resources available to support you.


I feel like hurting myself. I feel that I am no good


I'm sorry to hear that you're feeling this way. It's important to remember that these thoughts and feelings are valid, but t
hey do not define your worth as a person. You are valuable and deserving of support and care.

If you are feeling like hurting yourself, it's crucial to reach out for help immediately. You can call a crisis hotline such
as the National Suicide Prevention Lifeline at 1-800-273-TALK (8255) for immediate support. You can also reach out to a trus
ted adult, a mental health professional, or go to the nearest emergency room for assistance.

You are not alone, and there are people who care about you and want to help you through this tough time. Please remember tha
t there is help available, and you deserve to feel better. It's important to prioritize your safety and well-being, so pleas
e don't hesitate to seek help. You are valued and important.


BYE


If you ever need someone to talk to or support in the future, please don't hesitate to reach out. Take care of yourself, and
remember that you matter. Goodbye.
```

This is a basic chatbot leveraging Chat GPT LLM. It is not advisable to use this chatbot for actual live use as the GPT model has **risks such AI bias and hallucination**. Providing inaccurate information to the teens may cause irreparable harm and expose the organization to legal risks. Also, due to several AI risks, it may be advisable to create an AI Assistant that would provide helpful resources to the Teen talk Advisors rather than directly interacting with the teens. In the next part, we created an AI assistant that uses the organization’s resource files for knowledge base to find correct answers and support the teen talk mental health advisors while chatting with the teens at risk.

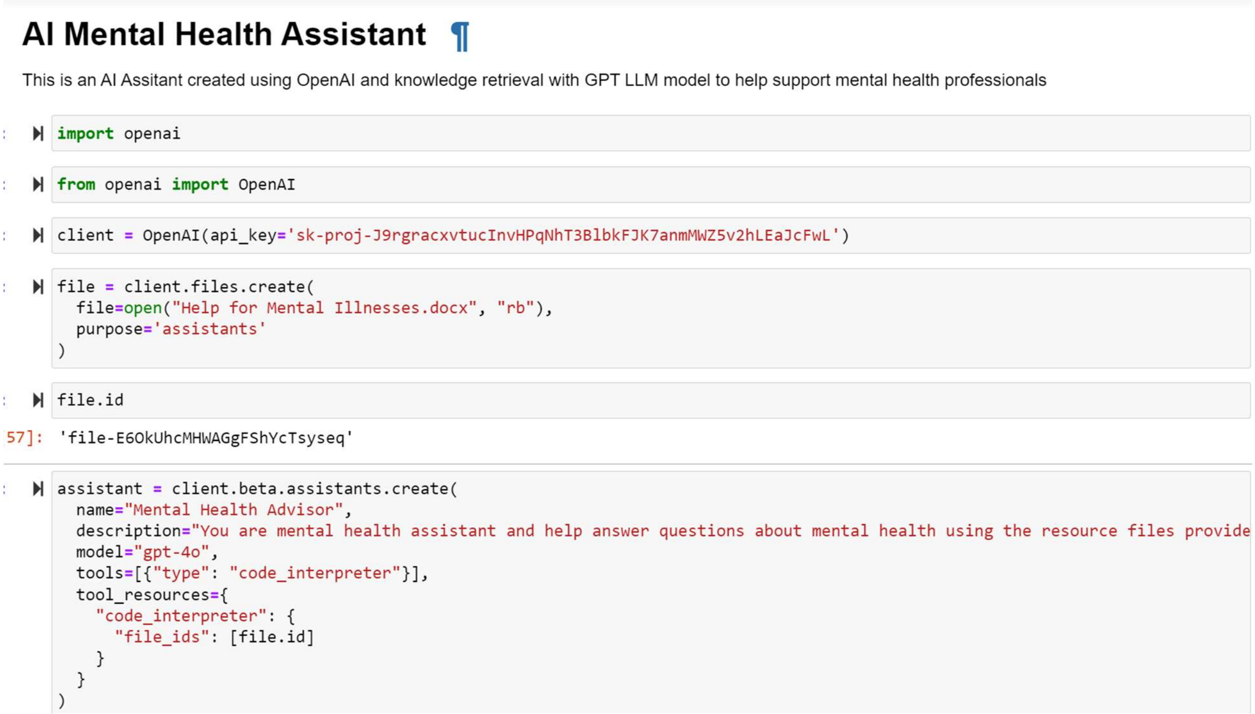
Part 2 - Create the AI Assistant to help support Mental Health professionals

The researchers built an AI assistant using Open AI “gpt-4o” model with code_interpreter tool for knowledge retrieval. We created a sample file – “Help for Mental Illnesses.docx” based on the mental illness data from - https://www.nimh.nih.gov/health/find-help. This model would require the AI assistant to refer to verified organization approved data to answer the user’s questions regarding mental illness and mitigate the risks associated with AI Bias and hallucination that Chat GPT models may introduce.

Code File – https://github.com/isha01agr/Neurips_AI/blob/main/Part%202%20-Virtual%20AI%20Assistant.ipynb

Data File – https://github.com/isha01agr/Neurips_AI/blob/main/Help%20for%20Mental%20Illnesses.docx

Code Snippet below:



AI Mental Health Assistant ¶

This is an AI Assitant created using OpenAI and knowledge retrieval with GPT LLM model to help support mental health professionals

```
import openai
```

```
from openai import OpenAI
```

```
client = OpenAI(api_key='sk-proj-J9rgracxvtucInvHPqNhT3BlbkFJK7anmMWZ5v2hLEaJcFwL')
```

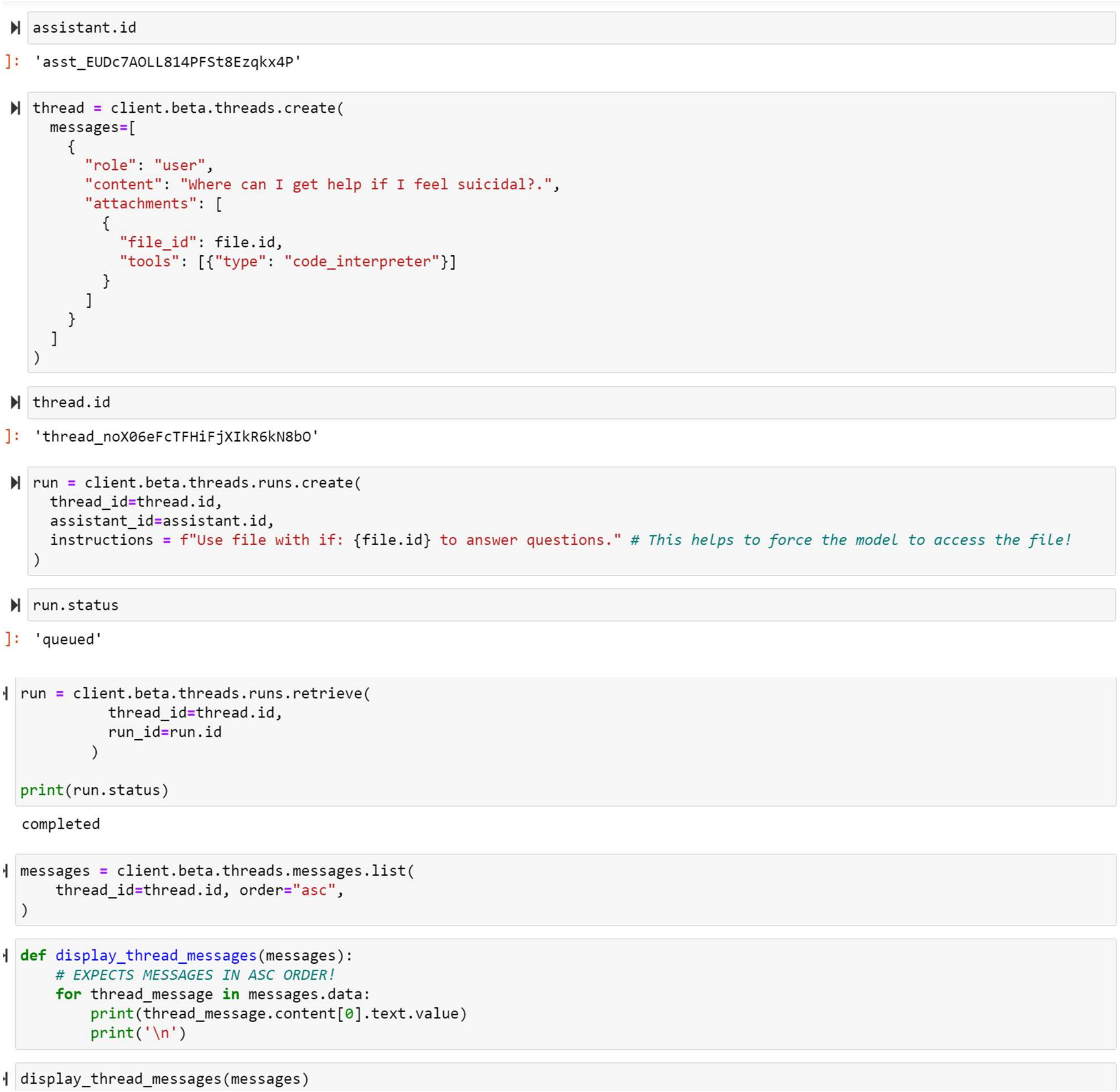
```
file = client.files.create(
  file=open("Help for Mental Illnesses.docx", "rb"),
  purpose='assistants'
)
```

```
file.id
```

```
57]: 'file-E6OkUhcMHWAGgFShYcTsyseq'
```

```
assistant = client.beta.assistants.create(
  name="Mental Health Advisor",
  description="You are mental health assistant and help answer questions about mental health using the resource files provide
  model="gpt-4o",
  tools=[{"type": "code_interpreter"}],
  tool_resources={
    "code_interpreter": {
      "file_ids": [file.id]
    }
  }
)
```

We uploaded the resource file to OpenAI Platform and created an “assistant” called “Mental Health Advisor” and provided code_interpreter tool to use the resource file that was provided. The next step was to create a thread and pass a user message – “Where can I get help if I feel suicidal? We created a run and waited for it to be “completed” to display the latest thread message.



```
assistant.id
```

```
]: 'asst_EUDc7AOLL814PFSt8Ezqkx4P'
```

```
thread = client.beta.threads.create(
  messages=[
    {
      "role": "user",
      "content": "Where can I get help if I feel suicidal?.",
      "attachments": [
        {
          "file_id": file.id,
          "tools": [{"type": "code_interpreter"}]
        }
      ]
    }
  ]
)
```

```
thread.id
```

```
]: 'thread_noX06eFcTFHiFjXIkR6kN8bO'
```

```
run = client.beta.threads.runs.create(
  thread_id=thread.id,
  assistant_id=assistant.id,
  instructions = f"Use file with if: {file.id} to answer questions." # This helps to force the model to access the file!
)
```

```
run.status
```

```
]: 'queued'
```

```
run = client.beta.threads.runs.retrieve(
          thread_id=thread.id,
          run_id=run.id
        )

print(run.status)
```

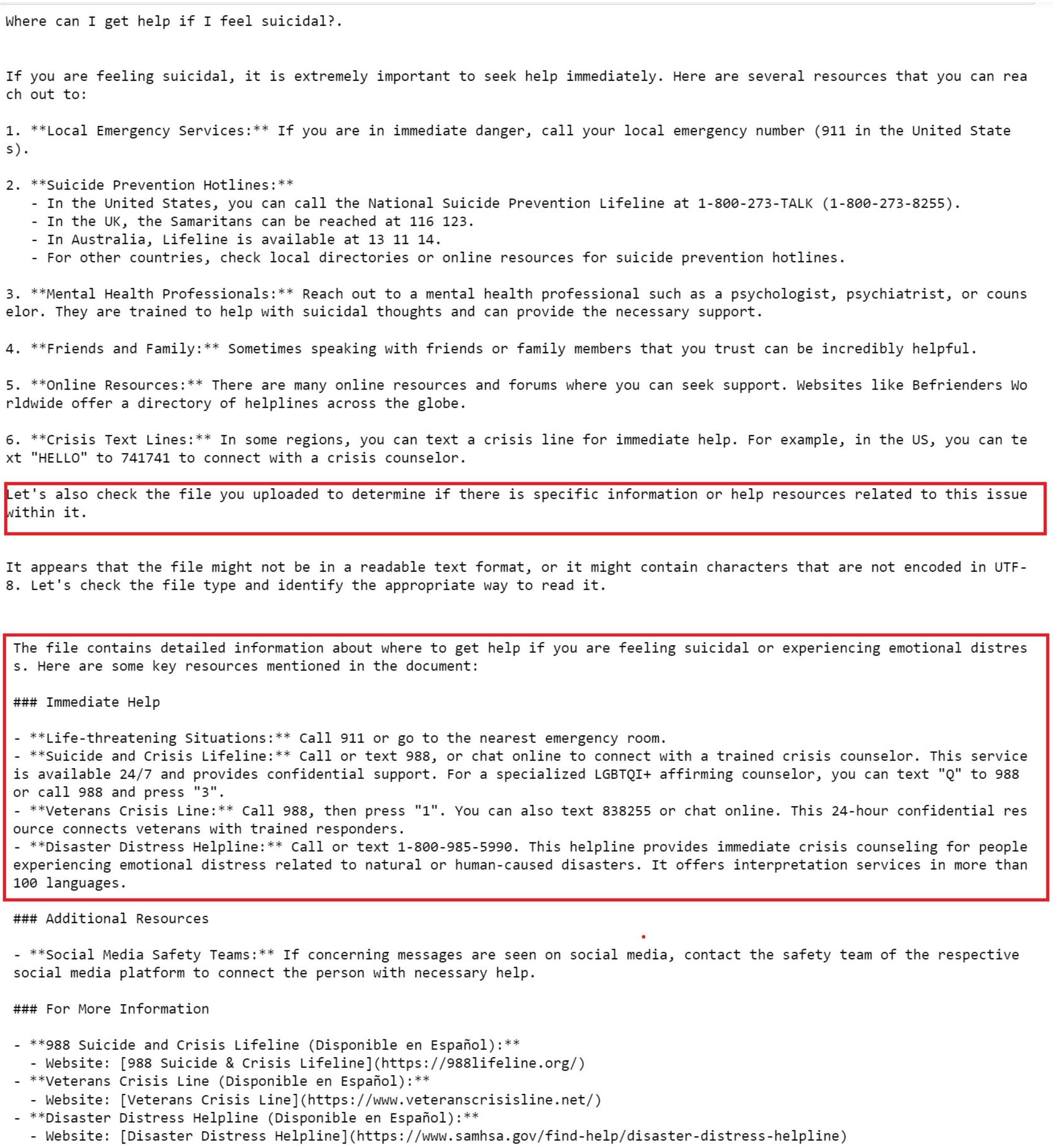
```
completed
```

```
messages = client.beta.threads.messages.list(
    thread_id=thread.id, order="asc",
)
```

```
def display_thread_messages(messages):
    # EXPECTS MESSAGES IN ASC ORDER!
    for thread_message in messages.data:
        print(thread_message.content[0].text.value)
        print('\n')
```

```
display_thread_messages(messages)
```

The output from the thread message is below. The highlighted sections indicate that the assistant referred to the file that was provided and used it to answer the user's question. This helps the model to rely on verified and approved information and answer the question correctly.

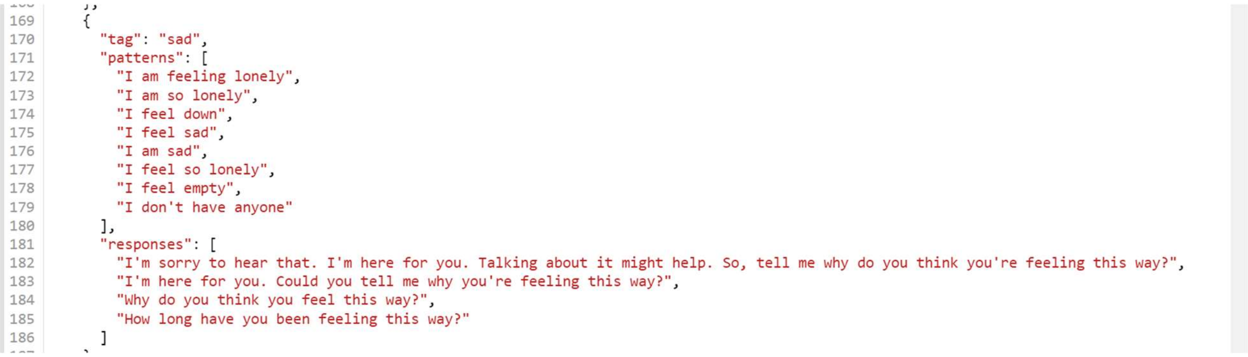


```
Where can I get help if I feel suicidal?.


If you are feeling suicidal, it is extremely important to seek help immediately. Here are several resources that you can rea
ch out to:

1. **Local Emergency Services:** If you are in immediate danger, call your local emergency number (911 in the United State
s).

2. **Suicide Prevention Hotlines:**
   - In the United States, you can call the National Suicide Prevention Lifeline at 1-800-273-TALK (1-800-273-8255).
   - In the UK, the Samaritans can be reached at 116 123.
   - In Australia, Lifeline is available at 13 11 14.
   - For other countries, check local directories or online resources for suicide prevention hotlines.

3. **Mental Health Professionals:** Reach out to a mental health professional such as a psychologist, psychiatrist, or couns
elor. They are trained to help with suicidal thoughts and can provide the necessary support.

4. **Friends and Family:** Sometimes speaking with friends or family members that you trust can be incredibly helpful.

5. **Online Resources:** There are many online resources and forums where you can seek support. Websites like Befrienders Wo
rldwide offer a directory of helplines across the globe.

6. **Crisis Text Lines:** In some regions, you can text a crisis line for immediate help. For example, in the US, you can te
xt "HELLO" to 741741 to connect with a crisis counselor.

Let's also check the file you uploaded to determine if there is specific information or help resources related to this issue
within it.


It appears that the file might not be in a readable text format, or it might contain characters that are not encoded in UTF-
8. Let's check the file type and identify the appropriate way to read it.


The file contains detailed information about where to get help if you are feeling suicidal or experiencing emotional distres
s. Here are some key resources mentioned in the document:

### Immediate Help

- **Life-threatening Situations:** Call 911 or go to the nearest emergency room.
- **Suicide and Crisis Lifeline:** Call or text 988, or chat online to connect with a trained crisis counselor. This service
is available 24/7 and provides confidential support. For a specialized LGBTQI+ affirming counselor, you can text "Q" to 988
or call 988 and press "3".
- **Veterans Crisis Line:** Call 988, then press "1". You can also text 838255 or chat online. This 24-hour confidential res
ource connects veterans with trained responders.
- **Disaster Distress Helpline:** Call or text 1-800-985-5990. This helpline provides immediate crisis counseling for people
experiencing emotional distress related to natural or human-caused disasters. It offers interpretation services in more than
100 languages.

### Additional Resources

- **Social Media Safety Teams:** If concerning messages are seen on social media, contact the safety team of the respective
social media platform to connect the person with necessary help.

### For More Information

- **988 Suicide and Crisis Lifeline (Disponible en Español):**
  - Website: [988 Suicide & Crisis Lifeline](https://988lifeline.org/)
- **Veterans Crisis Line (Disponible en Español):**
  - Website: [Veterans Crisis Line](https://www.veteranscrisisline.net/)
- **Disaster Distress Helpline (Disponible en Español):**
  - Website: [Disaster Distress Helpline](https://www.samhsa.gov/find-help/disaster-distress-helpline)
```

Part 3 - Create the intent Classification Model

Researchers reused the dataset from Kaggle.com to build the machine learning model for Teen Talk App - kaggle/input/mental-health-conversational-data/intents.json

This dataset is a set of past conversations related to mental health. It consists of tags, patterns and responses. Patterns are the messages that are sent by the users who need mental health support, and the responses are the replies sent by the health care advisor to answer the messages (patterns) received from the users. Tags are the "intents" behind the message sent by the users. For e.g. if the user sends a message such as - I am lonely or sad or empty etc., then the tag (intent) would be classified as "sad"



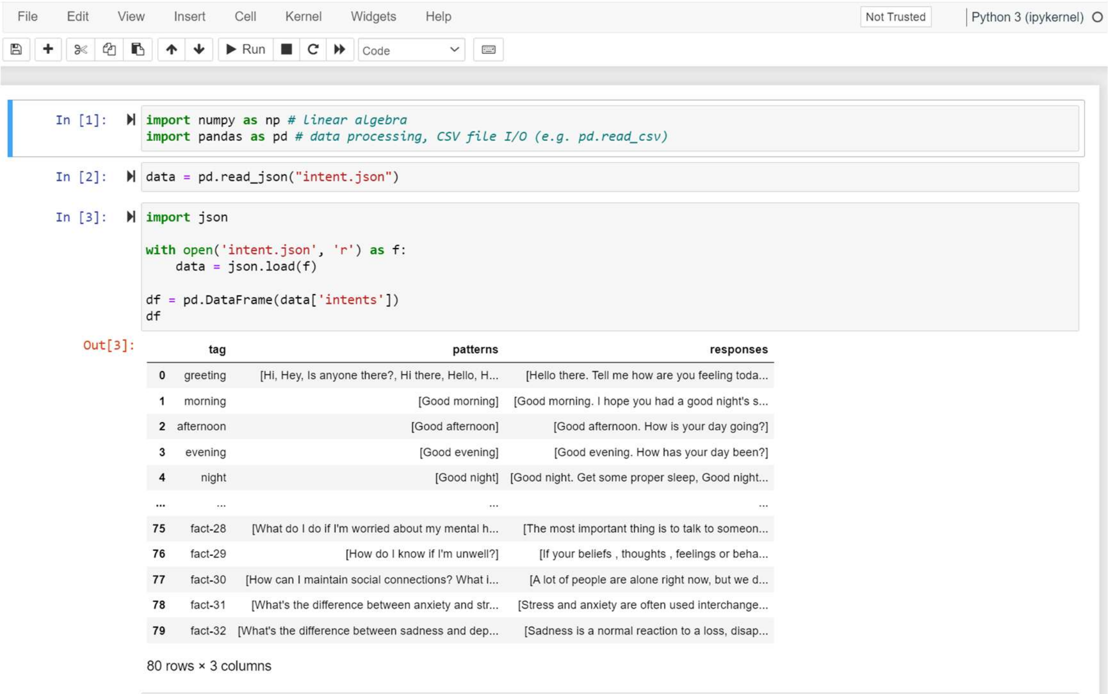
```
    {
      "tag": "sad",
      "patterns": [
        "I am feeling lonely",
        "I am so lonely",
        "I feel down",
        "I feel sad",
        "I am sad",
        "I feel so lonely",
        "I feel empty",
        "I don't have anyone"
      ],
      "responses": [
        "I'm sorry to hear that. I'm here for you. Talking about it might help. So, tell me why do you think you're feeling this way?",
        "I'm here for you. Could you tell me why you're feeling this way?",
        "Why do you think you feel this way?",
        "How long have you been feeling this way?"
      ]
```

Code File – https://github.com/isha01agr/Neurips_AI/blob/main/Part%203%20-%20Mental_Health_intent_detection.ipynb

Data File - https://github.com/isha01agr/Neurips_AI/blob/main/intent.json

Code Snippet below:

We imported the "intent.json" file downloaded from Kaggle.com and create a pandas data frame in Python Jupyter notebook.



```python
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
```

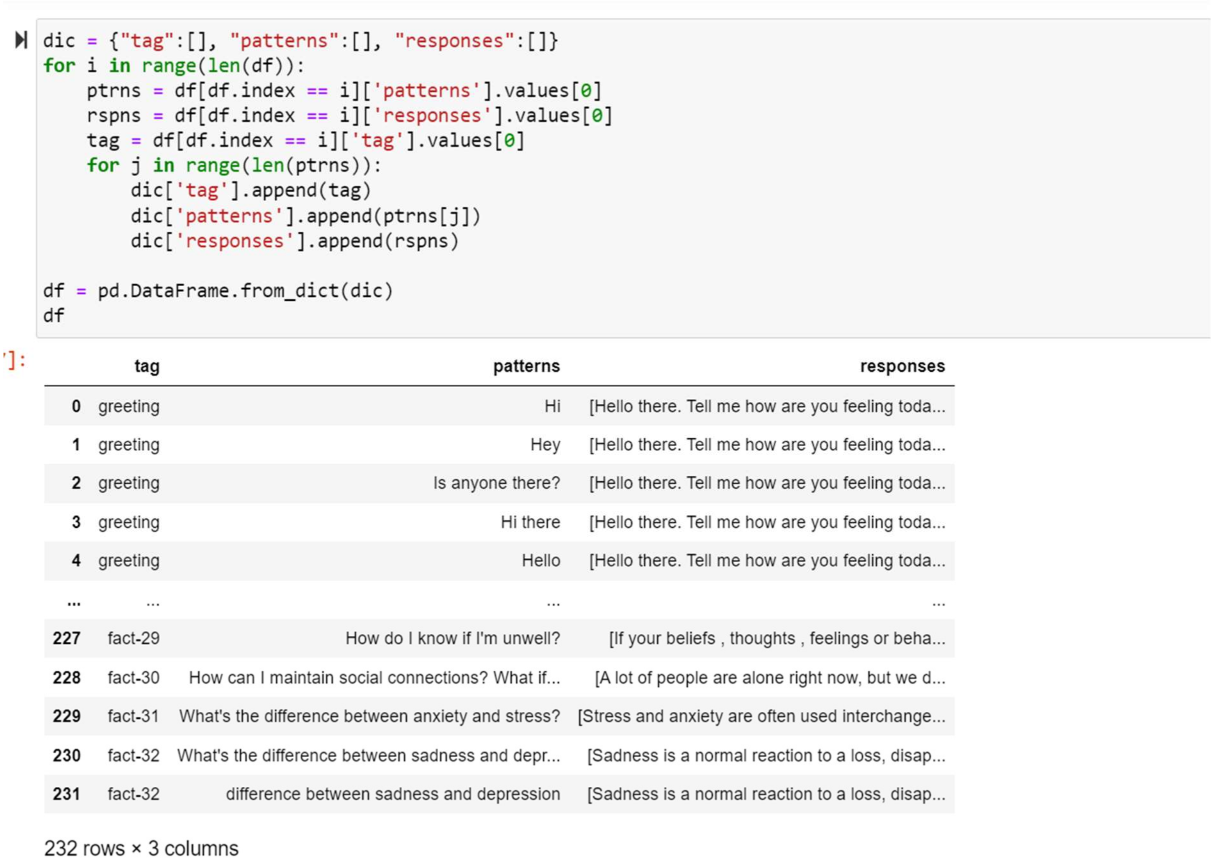
```python
data = pd.read_json("intent.json")
```

```python
import json

with open('intent.json', 'r') as f:
    data = json.load(f)

df = pd.DataFrame(data['intents'])
df
```

	tag	patterns	responses
0	greeting	[Hi, Hey, Is anyone there?, Hi there, Hello, H...	[Hello there. Tell me how are you feeling toda...
1	morning	[Good morning]	[Good morning. I hope you had a good night's s...
2	afternoon	[Good afternoon]	[Good afternoon. How is your day going?]
3	evening	[Good evening]	[Good evening. How has your day been?]
4	night	[Good night]	[Good night. Get some proper sleep, Good night...
...	...	...	...
75	fact-28	[What do I do if I'm worried about my mental h...	[The most important thing is to talk to someon...
76	fact-29	[How do I know if I'm unwell?]	[If your beliefs , thoughts , feelings or beha...
77	fact-30	[How can I maintain social connections? What i...	[A lot of people are alone right now, but we d...
78	fact-31	[What's the difference between anxiety and str...	[Stress and anxiety are often used interchange...
79	fact-32	[What's the difference between sadness and dep...	[Sadness is a normal reaction to a loss, disap...

80 rows × 3 columns

Create a dictionary using the data frame



```python
dic = {"tag":[], "patterns":[], "responses":[]}
for i in range(len(df)):
    ptrns = df[df.index == i]['patterns'].values[0]
    rspns = df[df.index == i]['responses'].values[0]
    tag = df[df.index == i]['tag'].values[0]
    for j in range(len(ptrns)):
        dic['tag'].append(tag)
        dic['patterns'].append(ptrns[j])
        dic['responses'].append(rspns)

df = pd.DataFrame.from_dict(dic)
df
```

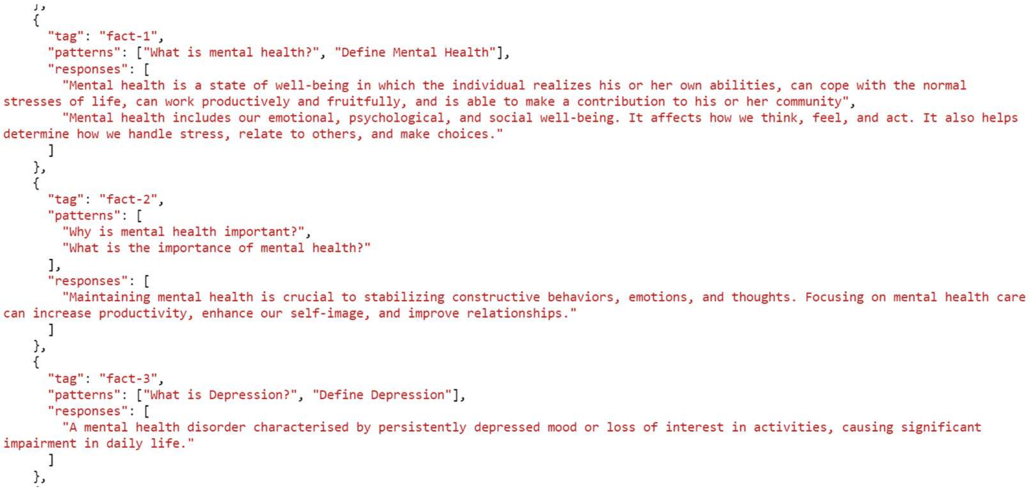
	tag	patterns	responses
0	greeting	Hi	[Hello there. Tell me how are you feeling toda...
1	greeting	Hey	[Hello there. Tell me how are you feeling toda...
2	greeting	Is anyone there?	[Hello there. Tell me how are you feeling toda...
3	greeting	Hi there	[Hello there. Tell me how are you feeling toda...
4	greeting	Hello	[Hello there. Tell me how are you feeling toda...
...	...	...	...
227	fact-29	How do I know if I'm unwell?	[If your beliefs , thoughts , feelings or beha...
228	fact-30	How can I maintain social connections? What if...	[A lot of people are alone right now, but we d...
229	fact-31	What's the difference between anxiety and stress?	[Stress and anxiety are often used interchange...
230	fact-32	What's the difference between sadness and depr...	[Sadness is a normal reaction to a loss, disap...
231	fact-32	difference between sadness and depression	[Sadness is a normal reaction to a loss, disap...

232 rows × 3 columns

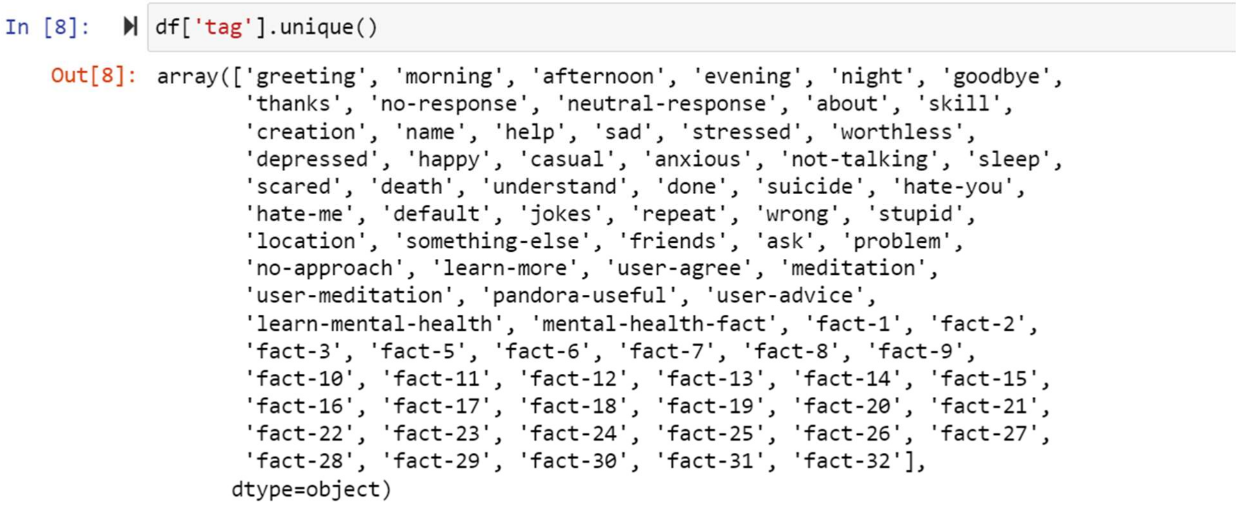
Data Visualization

The data consists of the following tags (intents) which include greetings (good morning, afternoon, bye etc.), emotional state (depressed, suicide, anxious etc.) and mental health facts.

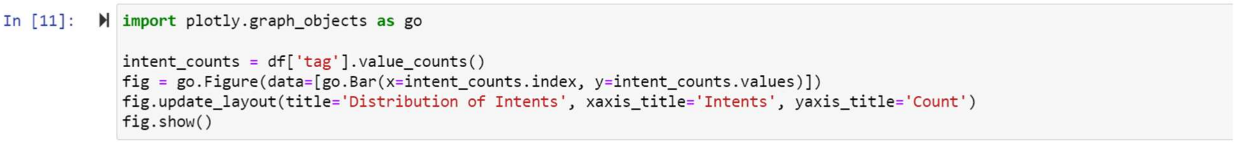
Facts are intentions where the users ask specific question about mental health such as – “What is mental health?”, “What is depression?” etc.



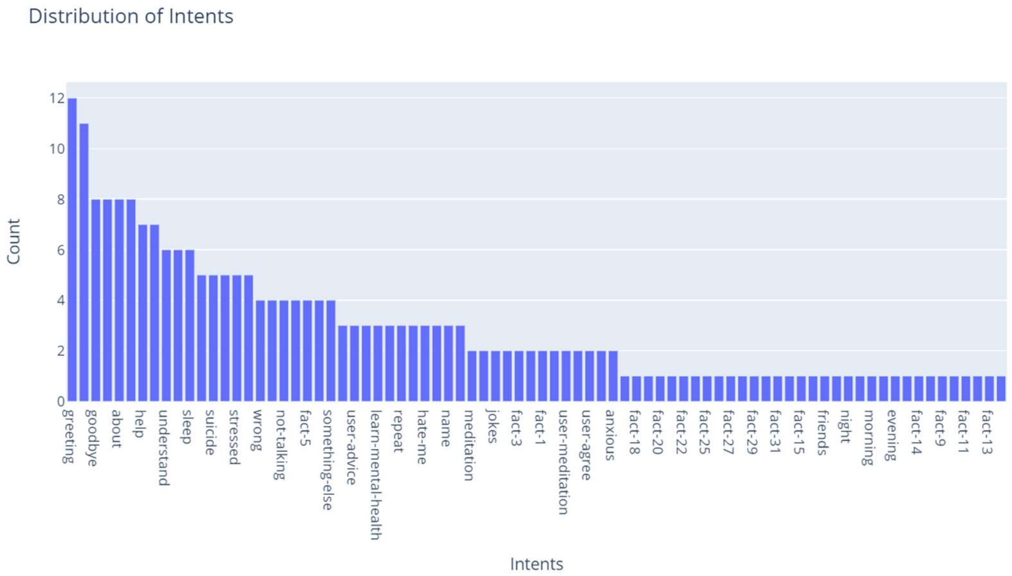
```
    },
    {
      "tag": "fact-1",
      "patterns": ["What is mental health?", "Define Mental Health"],
      "responses": [
        "Mental health is a state of well-being in which the individual realizes his or her own abilities, can cope with the normal
stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community",
        "Mental health includes our emotional, psychological, and social well-being. It affects how we think, feel, and act. It also helps
determine how we handle stress, relate to others, and make choices."
      ]
    },
    {
      "tag": "fact-2",
      "patterns": [
        "Why is mental health important?",
        "What is the importance of mental health?"
      ],
      "responses": [
        "Maintaining mental health is crucial to stabilizing constructive behaviors, emotions, and thoughts. Focusing on mental health care
can increase productivity, enhance our self-image, and improve relationships."
      ]
    },
    {
      "tag": "fact-3",
      "patterns": ["What is Depression?", "Define Depression"],
      "responses": [
        "A mental health disorder characterised by persistently depressed mood or loss of interest in activities, causing significant
impairment in daily life."
      ]
    },
```



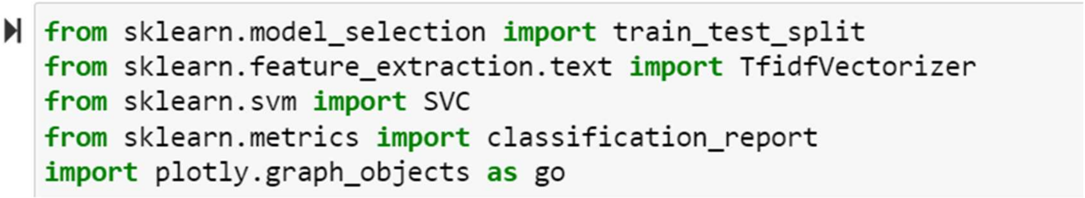
```
In [8]:  df['tag'].unique()

Out[8]: array(['greeting', 'morning', 'afternoon', 'evening', 'night', 'goodbye',
               'thanks', 'no-response', 'neutral-response', 'about', 'skill',
               'creation', 'name', 'help', 'sad', 'stressed', 'worthless',
               'depressed', 'happy', 'casual', 'anxious', 'not-talking', 'sleep',
               'scared', 'death', 'understand', 'done', 'suicide', 'hate-you',
               'hate-me', 'default', 'jokes', 'repeat', 'wrong', 'stupid',
               'location', 'something-else', 'friends', 'ask', 'problem',
               'no-approach', 'learn-more', 'user-agree', 'meditation',
               'user-meditation', 'pandora-useful', 'user-advice',
               'learn-mental-health', 'mental-health-fact', 'fact-1', 'fact-2',
               'fact-3', 'fact-5', 'fact-6', 'fact-7', 'fact-8', 'fact-9',
               'fact-10', 'fact-11', 'fact-12', 'fact-13', 'fact-14', 'fact-15',
               'fact-16', 'fact-17', 'fact-18', 'fact-19', 'fact-20', 'fact-21',
               'fact-22', 'fact-23', 'fact-24', 'fact-25', 'fact-26', 'fact-27',
               'fact-28', 'fact-29', 'fact-30', 'fact-31', 'fact-32'],
              dtype=object)
```



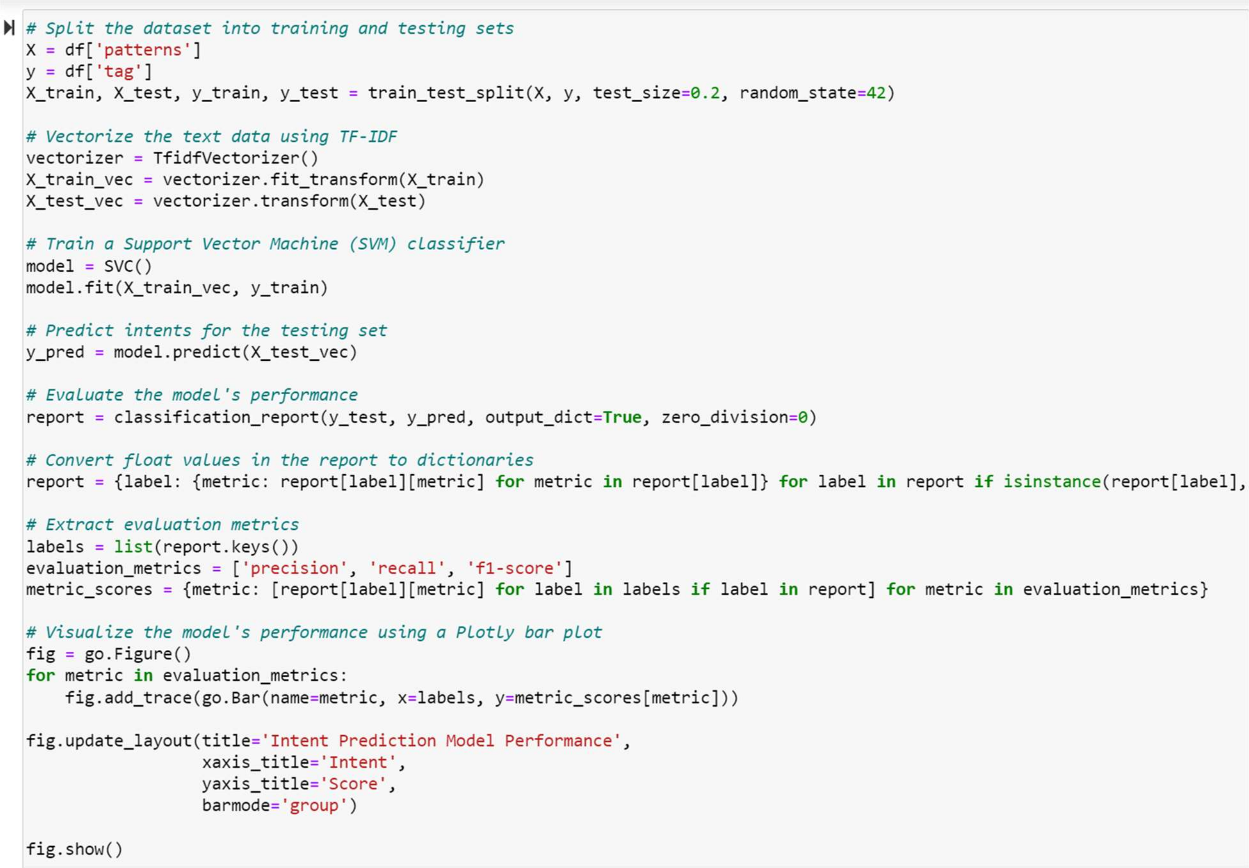
```
In [11]:  import plotly.graph_objects as go

          intent_counts = df['tag'].value_counts()
          fig = go.Figure(data=[go.Bar(x=intent_counts.index, y=intent_counts.values)])
          fig.update_layout(title='Distribution of Intents', xaxis_title='Intents', yaxis_title='Count')
          fig.show()
```



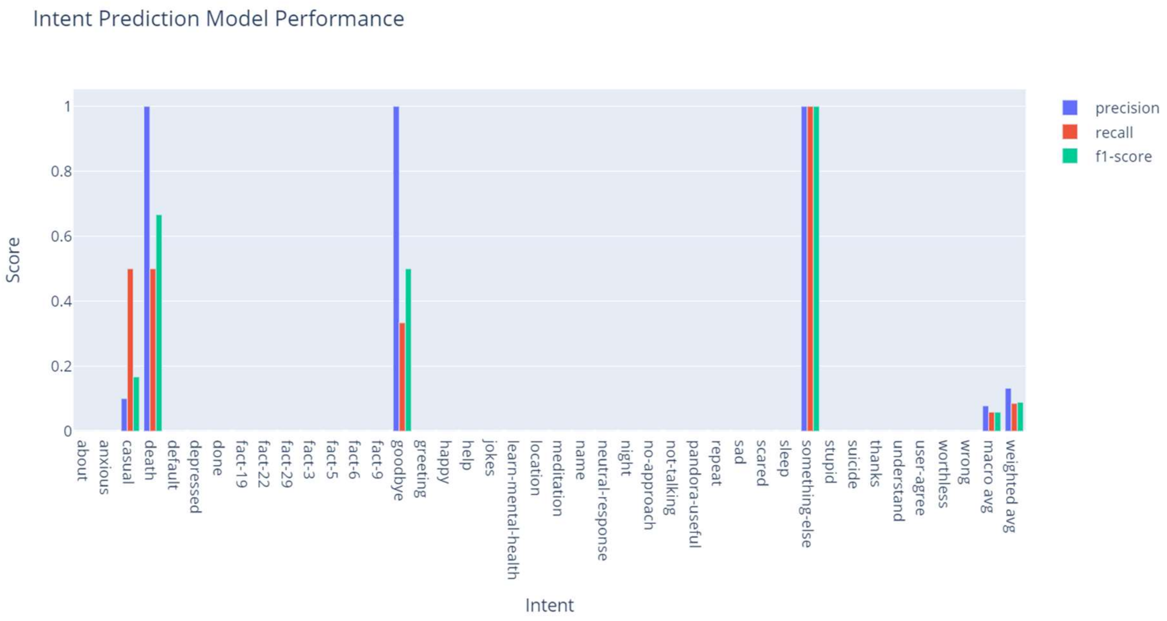
Finally, we created a Machine Learning model that can be trained using this dataset and predict users’ intent for future chats. The text data was vectorized using the TF-IDF vectorization method.



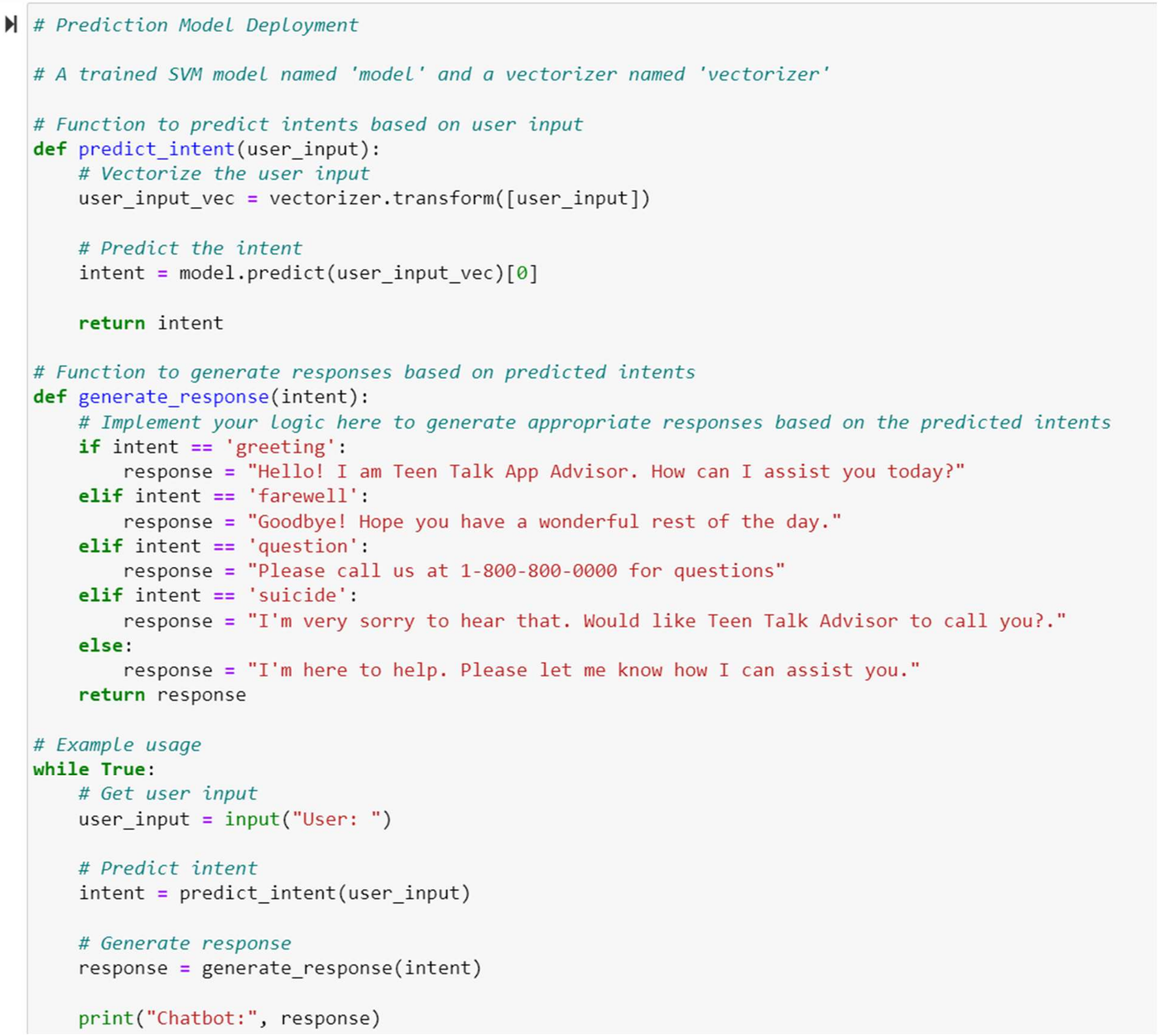
```
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import SVC
from sklearn.metrics import classification_report
import plotly.graph_objects as go
```



```
# Split the dataset into training and testing sets
X = df['patterns']
y = df['tag']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Vectorize the text data using TF-IDF
vectorizer = TfidfVectorizer()
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)

# Train a Support Vector Machine (SVM) classifier
model = SVC()
model.fit(X_train_vec, y_train)

# Predict intents for the testing set
y_pred = model.predict(X_test_vec)

# Evaluate the model's performance
report = classification_report(y_test, y_pred, output_dict=True, zero_division=0)

# Convert float values in the report to dictionaries
report = {label: {metric: report[label][metric] for metric in report[label]} for label in report if isinstance(report[label],

# Extract evaluation metrics
labels = list(report.keys())
evaluation_metrics = ['precision', 'recall', 'f1-score']
metric_scores = {metric: [report[label][metric] for label in labels if label in report] for metric in evaluation_metrics}

# Visualize the model's performance using a Plotly bar plot
fig = go.Figure()
for metric in evaluation_metrics:
    fig.add_trace(go.Bar(name=metric, x=labels, y=metric_scores[metric]))

fig.update_layout(title='Intent Prediction Model Performance',
                  xaxis_title='Intent',
                  yaxis_title='Score',
                  barmode='group')

fig.show()
```



We deployed the model and taught it how to generate responses for each intent that the ML model identifies.



```python
# Prediction Model Deployment

# A trained SVM model named 'model' and a vectorizer named 'vectorizer'

# Function to predict intents based on user input
def predict_intent(user_input):
    # Vectorize the user input
    user_input_vec = vectorizer.transform([user_input])

    # Predict the intent
    intent = model.predict(user_input_vec)[0]

    return intent

# Function to generate responses based on predicted intents
def generate_response(intent):
    # Implement your logic here to generate appropriate responses based on the predicted intents
    if intent == 'greeting':
        response = "Hello! I am Teen Talk App Advisor. How can I assist you today?"
    elif intent == 'farewell':
        response = "Goodbye! Hope you have a wonderful rest of the day."
    elif intent == 'question':
        response = "Please call us at 1-800-800-0000 for questions"
    elif intent == 'suicide':
        response = "I'm very sorry to hear that. Would like Teen Talk Advisor to call you?."
    else:
        response = "I'm here to help. Please let me know how I can assist you."
    return response

# Example usage
while True:
    # Get user input
    user_input = input("User: ")

    # Predict intent
    intent = predict_intent(user_input)

    # Generate response
    response = generate_response(intent)

    print("Chatbot:", response)
```

This is a basic code that identifies how to respond to a user when the intent is determined to be greeting, farewell, question or suicide. The code can be extended to include responses for all the intents in the model.