CREATING CHATBOT USING PYTHON

STEPS FOR LOADING AND PREPROCESSING DATASET

1. Import necessary libraries:

```
import pandas as pd
```

2. Load the dataset:

Depending on your data format (e.g., CSV, JSON), use the appropriate Pandas function to load the data into a DataFrame.

```
df = pd.read_csv('your_dataset.csv')
```

3. Preprocess the data:

Preprocessing steps may include removing duplicates, handling missing values, and any other specific data cleaning required for your dataset.

Example: Remove duplicates

```
df.drop_duplicates(inplace=True)
```

Example: Handle missing values

```
df.dropna(inplace=True)
```

4. Prepare data for training:

Depending on your chatbot approach (e.g., rule-based, machine learning), format the data in a way suitable for your training method. This may involve separating questions and answers or creating training pairs.

5. Tokenize the text:

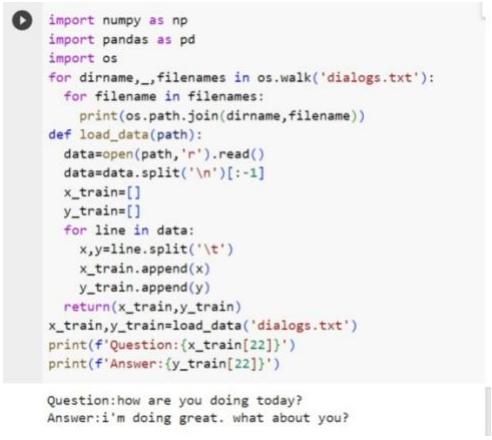
Tokenize the text data into words or subwords for further processing. Libraries like NLTK or spaCy can be helpful for this step.

Once you have preprocessed and tokenized the data, you can proceed to build and train your chatbot using appropriate models and techniques based on your specific requirements.

CODE

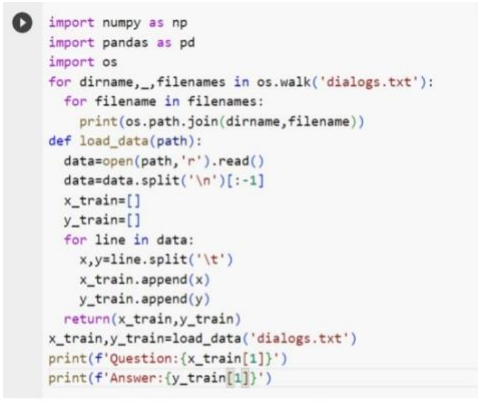
```
import numpy as np
import pandas as pd
import os
for dirname,_,filenames in os.walk('dialogs.txt'):
   for filename in filenames:
   print(os.path.join(dirname,filename))
def load_data(path):
  data=open(path,'r').read()
```

```
  data=data.split('\n')[:-1]
  x_train=[]
  y_train=[]
  for line in data:
    x,y=line.split('\t')
    x_train.append(x)
    y_train.append(y)
 return(x_train,y_train)
x_train,y_train=load_data('dialogs.txt')
print(f'Question:{x_train[1]}')
print(f'Answer:{y_train[1]}')
```



```
import numpy as np
import pandas as pd
import os
for dirname,_,filenames in os.walk('dialogs.txt'):
  for filename in filenames:
    print(os.path.join(dirname,filename))
def load_data(path):
  data=open(path,'r').read()
  data=data.split('\n')[:-1]
  x_train=[]
  y_train=[]
  for line in data:
    x,y=line.split('\t')
    x_train.append(x)
    y_train.append(y)
  return(x_train,y_train)
x_train,y_train=load_data('dialogs.txt')
print(f'Question:{x_train[22]}')
print(f'Answer:{y_train[22]}')
```

```
Question:how are you doing today?
Answer:i'm doing great. what about you?
```



```
import numpy as np
import pandas as pd
import os
for dirname,_,filenames in os.walk('dialogs.txt'):
  for filename in filenames:
    print(os.path.join(dirname,filename))
def load_data(path):
  data=open(path,'r').read()
  data=data.split('\n')[:-1]
  x_train=[]
  y_train=[]
  for line in data:
    x,y=line.split('\t')
    x_train.append(x)
    y_train.append(y)
  return(x_train,y_train)
x_train,y_train=load_data('dialogs.txt')
print(f'Question:{x_train[1]}')
print(f'Answer:{y_train[1]}')
```

```
Question:i'm fine. how about yourself?
Answer:i'm pretty good. thanks for asking.
```