

# Project

In this Project, you will bring together many of the tools and techniques that you have learned throughout this course into a final project. You can choose from many different paths to get to the solution.

## Business scenario

You work for a training organization that recently developed an introductory course about machine learning (ML). The course includes more than 40 videos that cover a broad range of ML topics. You have been asked to create an application that will students can use to quickly locate and view video content by searching for topics and key phrases.

You have downloaded all of the videos to an Amazon Simple Storage Service (Amazon S3) bucket. Your assignment is to produce a dashboard that meets your supervisor's requirements.

## Project steps

To complete this project, you will follow these steps:

1. [Viewing the video files](#)
2. [Transcribing the videos](#)
3. [Normalizing the text](#)
4. [Extracting key phrases and topics](#)
5. [Creating the dashboard](#)

## Useful information

The following cell contains some information that might be useful as you complete this project.

```
In [1]: bucket = "c56161a93943013396553t1w744137092661-labbucket-rn642jaq01e9"  
job_data_access_role = 'arn:aws:iam::744137092661:role/service-role/c56161a93943013'
```

### 1. Viewing the video files

[\(Go to top\)](#)

The source video files are located in the following shared Amazon Simple Storage Service (Amazon S3) bucket.

In [2]: !aws s3 ls s3://aws-tc-largeobjects/CUR-TF-200-ACMNL-1/video/

```
2021-04-26 20:17:33 410925369 Mod01_Course_Overview.mp4
2021-04-26 20:10:02 39576695 Mod02_Intro.mp4
2021-04-26 20:31:23 302994828 Mod02_Sect01.mp4
2021-04-26 20:17:33 416563881 Mod02_Sect02.mp4
2021-04-26 20:17:33 318685583 Mod02_Sect03.mp4
2021-04-26 20:17:33 255877251 Mod02_Sect04.mp4
2021-04-26 20:23:51 99988046 Mod02_Sect05.mp4
2021-04-26 20:24:54 50700224 Mod02_WrapUp.mp4
2021-04-26 20:26:27 60627667 Mod03_Intro.mp4
2021-04-26 20:26:28 272229844 Mod03_Sect01.mp4
2021-04-26 20:27:06 309127124 Mod03_Sect02_part1.mp4
2021-04-26 20:27:06 195635527 Mod03_Sect02_part2.mp4
2021-04-26 20:28:03 123924818 Mod03_Sect02_part3.mp4
2021-04-26 20:31:28 171681915 Mod03_Sect03_part1.mp4
2021-04-26 20:32:07 285200083 Mod03_Sect03_part2.mp4
2021-04-26 20:33:17 105470345 Mod03_Sect03_part3.mp4
2021-04-26 20:35:10 157185651 Mod03_Sect04_part1.mp4
2021-04-26 20:36:27 187435635 Mod03_Sect04_part2.mp4
2021-04-26 20:36:40 280720369 Mod03_Sect04_part3.mp4
2021-04-26 20:40:01 443479313 Mod03_Sect05.mp4
2021-04-26 20:40:08 234182186 Mod03_Sect06.mp4
2021-04-26 20:40:33 207718047 Mod03_Sect07_part1.mp4
2021-04-26 20:42:07 125592110 Mod03_Sect07_part2.mp4
2021-04-26 20:45:10 508500301 Mod03_Sect07_part3.mp4
2021-04-26 20:46:16 320126756 Mod03_Sect08.mp4
2021-04-26 20:46:43 41839508 Mod03_WrapUp.mp4
2021-04-26 20:46:55 34148489 Mod04_Intro.mp4
2021-04-26 20:48:24 84959465 Mod04_Sect01.mp4
2021-04-26 20:48:25 345182970 Mod04_Sect02_part1.mp4
2021-04-26 20:51:34 218661651 Mod04_Sect02_part2.mp4
2021-04-26 20:53:32 430140637 Mod04_Sect02_part3.mp4
2021-04-26 20:56:03 22036605 Mod04_WrapUp.mp4
2021-04-26 20:57:18 49187118 Mod05_Intro.mp4
2021-04-26 20:58:19 245798071 Mod05_Sect01_ver2.mp4
2021-04-26 20:58:50 233314835 Mod05_Sect02_part1_ver2.mp4
2021-04-26 20:59:14 348545306 Mod05_Sect02_part2.mp4
2021-04-26 20:59:17 239142711 Mod05_Sect03_part1.mp4
2021-04-26 21:06:04 267533559 Mod05_Sect03_part2.mp4
2021-04-26 21:06:06 212502220 Mod05_Sect03_part3.mp4
2021-04-26 21:06:48 206317022 Mod05_Sect03_part4_ver2.mp4
2021-04-26 21:06:48 60361230 Mod05_WrapUp_ver2.mp4
2021-04-26 21:09:14 35397860 Mod06_Intro.mp4
2021-04-26 21:09:24 845633599 Mod06_Sect01.mp4
2021-04-26 21:10:47 326126684 Mod06_Sect02.mp4
2021-04-26 21:12:26 19790740 Mod06_WrapUp.mp4
2021-04-26 21:12:56 131249036 Mod07_Sect01.mp4
```

## 2. Transcribing the videos

[\(Go to top\)](#)

Use this section to implement your solution to transcribe the videos.

```
In [3]: # Write your answer/code here
# Install libraries
!pip install wget
!pip install moviepy
!pip install SpeechRecognition
!pip install textblob
!pip install gensim

# Import Libraries
import wget
import moviepy.editor as mp
import speech_recognition as sr
import os
import pandas as pd

vid_files = !aws s3 ls s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/
vid_files = [file[31:] for file in vid_files]
vid_files
```

```
Requirement already satisfied: wget in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (3.2)
Requirement already satisfied: moviepy in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (1.0.3)
Requirement already satisfied: decorator<5.0,>=4.0.2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from moviepy) (4.4.2)
Requirement already satisfied: tqdm<5.0,>=4.11.2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from moviepy) (4.66.2)
Requirement already satisfied: requests<3.0,>=2.8.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from moviepy) (2.31.0)
Requirement already satisfied: proglol<1.0.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from moviepy) (0.1.10)
Requirement already satisfied: numpy>=1.17.3 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from moviepy) (1.22.4)
Requirement already satisfied: imageio<3.0,>=2.5 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from moviepy) (2.34.0)
Requirement already satisfied: imageio-ffmpeg>=0.2.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from moviepy) (0.4.9)
Requirement already satisfied: pillow>=8.3.2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from imageio<3.0,>=2.5->moviepy) (10.2.0)
Requirement already satisfied: setuptools in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from imageio-ffmpeg>=0.2.0->moviepy) (69.1.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from requests<3.0,>=2.8.1->moviepy) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from requests<3.0,>=2.8.1->moviepy) (3.6)
Requirement already satisfied: urllib3<3,>=1.21.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from requests<3.0,>=2.8.1->moviepy) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from requests<3.0,>=2.8.1->moviepy) (2024.2.2)
Requirement already satisfied: SpeechRecognition in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (3.10.3)
Requirement already satisfied: requests>=2.26.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from SpeechRecognition) (2.31.0)
Requirement already satisfied: typing-extensions in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from SpeechRecognition) (4.9.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from requests>=2.26.0->SpeechRecognition) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from requests>=2.26.0->SpeechRecognition) (3.6)
Requirement already satisfied: urllib3<3,>=1.21.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from requests>=2.26.0->SpeechRecognition) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from requests>=2.26.0->SpeechRecognition) (2024.2.2)
Requirement already satisfied: textblob in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (0.18.0.post0)
Requirement already satisfied: nltk>=3.8 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from textblob) (3.8.1)
Requirement already satisfied: click in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from nltk>=3.8->textblob) (8.1.7)
```

```
Requirement already satisfied: joblib in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from nltk>=3.8->textblob) (1.3.2)
Requirement already satisfied: regex>=2021.8.3 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from nltk>=3.8->textblob) (2023.12.25)
Requirement already satisfied: tqdm in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from nltk>=3.8->textblob) (4.66.2)
Requirement already satisfied: gensim in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (4.3.2)
Requirement already satisfied: numpy>=1.18.5 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from gensim) (1.22.4)
Requirement already satisfied: scipy>=1.7.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from gensim) (1.12.0)
Requirement already satisfied: smart-open>=1.8.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from gensim) (7.0.4)
Requirement already satisfied: wrapt in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from smart-open>=1.8.1->gensim) (1.16.0)
```

```
Out[3]: ['Mod01_Course Overview.mp4',
 'Mod02_Intro.mp4',
 'Mod02_Sect01.mp4',
 'Mod02_Sect02.mp4',
 'Mod02_Sect03.mp4',
 'Mod02_Sect04.mp4',
 'Mod02_Sect05.mp4',
 'Mod02_WrapUp.mp4',
 'Mod03_Intro.mp4',
 'Mod03_Sect01.mp4',
 'Mod03_Sect02_part1.mp4',
 'Mod03_Sect02_part2.mp4',
 'Mod03_Sect02_part3.mp4',
 'Mod03_Sect03_part1.mp4',
 'Mod03_Sect03_part2.mp4',
 'Mod03_Sect03_part3.mp4',
 'Mod03_Sect04_part1.mp4',
 'Mod03_Sect04_part2.mp4',
 'Mod03_Sect04_part3.mp4',
 'Mod03_Sect05.mp4',
 'Mod03_Sect06.mp4',
 'Mod03_Sect07_part1.mp4',
 'Mod03_Sect07_part2.mp4',
 'Mod03_Sect07_part3.mp4',
 'Mod03_Sect08.mp4',
 'Mod03_WrapUp.mp4',
 'Mod04_Intro.mp4',
 'Mod04_Sect01.mp4',
 'Mod04_Sect02_part1.mp4',
 'Mod04_Sect02_part2.mp4',
 'Mod04_Sect02_part3.mp4',
 'Mod04_WrapUp.mp4',
 'Mod05_Intro.mp4',
 'Mod05_Sect01_ver2.mp4',
 'Mod05_Sect02_part1_ver2.mp4',
 'Mod05_Sect02_part2.mp4',
 'Mod05_Sect03_part1.mp4',
 'Mod05_Sect03_part2.mp4',
 'Mod05_Sect03_part3.mp4',
 'Mod05_Sect03_part4_ver2.mp4',
 'Mod05_WrapUp_ver2.mp4',
 'Mod06_Intro.mp4',
 'Mod06_Sect01.mp4',
 'Mod06_Sect02.mp4',
 'Mod06_WrapUp.mp4',
 'Mod07_Sect01.mp4']
```

```
In [4]: extracted = []

for file in vid_files:

    url = 'http://aws-tc-largeobjects.s3.amazonaws.com/CUR-TF-200-ACMNL-1/video/'
    file = wget.download(url)
    video = mp.VideoFileClip(file)

    aud = video.audio
```

```
aud.write_audiofile(file + '.wav')

r = sr.Recognizer()

with sr.AudioFile(file + '.wav') as source:
    data = r.record(source, duration = 90)

text = r.recognize_google(data)

extracted.append(text)

os.system('rm *.wav *.mp4')
```

MoviePy - Writing audio in Mod01\_Course Overview.mp4.wav

MoviePy - Done.

MoviePy - Writing audio in Mod02\_Intro.mp4.wav

MoviePy - Done.

MoviePy - Writing audio in Mod02\_Sect01.mp4.wav

MoviePy - Done.

MoviePy - Writing audio in Mod02\_Sect02.mp4.wav

MoviePy - Done.

MoviePy - Writing audio in Mod02\_Sect03.mp4.wav

MoviePy - Done.

MoviePy - Writing audio in Mod02\_Sect04.mp4.wav

MoviePy - Done.

MoviePy - Writing audio in Mod02\_WrapUp.mp4.wav

MoviePy - Done.

MoviePy - Writing audio in Mod03\_Intro.mp4.wav

MoviePy - Done.

MoviePy - Writing audio in Mod03\_Sect01.mp4.wav

MoviePy - Done.

MoviePy - Writing audio in Mod03\_Sect02\_part1.mp4.wav

MoviePy - Done.

MoviePy - Writing audio in Mod03\_Sect02\_part2.mp4.wav

MoviePy - Done.

MoviePy - Writing audio in Mod03\_Sect02\_part3.mp4.wav

MoviePy - Done.

MoviePy - Writing audio in Mod03\_Sect03\_part1.mp4.wav

```
MoviePy - Done.  
MoviePy - Writing audio in Mod03_Sect03_part2.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod03_Sect03_part3.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod03_Sect04_part1.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod03_Sect04_part2.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod03_Sect04_part3.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod03_Sect05.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod03_Sect06.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod03_Sect07_part1.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod03_Sect07_part2.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod03_Sect07_part3.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod03_Sect08.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod03_WrapUp.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod04_Intro.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod04_Sect01.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod04_Sect02_part1.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod04_Sect02_part2.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod04_Sect02_part3.mp4.wav
```

```
MoviePy - Done.  
MoviePy - Writing audio in Mod04_WrapUp.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod05_Intro.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod05_Sect01_ver2.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod05_Sect02_part1_ver2.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod05_Sect02_part2.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod05_Sect03_part1.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod05_Sect03_part2.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod05_Sect03_part3.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod05_Sect03_part4_ver2.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod05_WrapUp_ver2.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod06_Intro.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod06_Sect01.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod06_Sect02.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod06_WrapUp.mp4.wav  
  
MoviePy - Done.  
MoviePy - Writing audio in Mod07_Sect01.mp4.wav  
  
MoviePy - Done.
```

### 3. Normalizing the text

[\(Go to top\)](#)

Use this section to perform any text normalization steps that are necessary for your solution.

```
In [5]: # Write your answer/code here  
  
df_videos = pd.DataFrame({'File': vid_files, 'Data': extracted})  
df_videos
```

Out[5]:

	File	Data
0	Mod01_Course Overview.mp4	hi and welcome to Amazon Academy machine learn...
1	Mod02_Intro.mp4	hi and welcome to module 2 of AWS Academy mach...
2	Mod02_Sect01.mp4	hi and welcome to section 1 in this section we...
3	Mod02_Sect02.mp4	hi and welcome back in this section we're goin...
4	Mod02_Sect03.mp4	I am welcome back this is Section 3 and we're ...
5	Mod02_Sect04.mp4	welcome back in this section we'll look at som...
6	Mod02_Sect05.mp4	hi welcome back this is Section 5 and we're go...
7	Mod02_WrapUp.mp4	it's now time to review the module here are th...
8	Mod03_Intro.mp4	welcome back to AWS Academy machine learning t...
9	Mod03_Sect01.mp4	hi and welcome back to module 3 this is Sectio...
10	Mod03_Sect02_part1.mp4	hi welcome back we're now going to look at a f...
11	Mod03_Sect02_part2.mp4	hi welcome back will continue exploring data c...
12	Mod03_Sect02_part3.mp4	hi welcome back will continue exploring data c...
13	Mod03_Sect03_part1.mp4	hi and welcome back this is Section 3 and we'r...
14	Mod03_Sect03_part2.mp4	hi welcome back we'll continue exploring how t...
15	Mod03_Sect03_part3.mp4	hi welcome back now we'll review how to find c...
16	Mod03_Sect04_part1.mp4	hi and welcome to section 4 in this section we...
17	Mod03_Sect04_part2.mp4	hi welcome back we'll continue exploring featu...
18	Mod03_Sect04_part3.mp4	hi welcome back we'll continue exploring featu...
19	Mod03_Sect05.mp4	hi welcome back to module 3 this is Section 5 ...
20	Mod03_Sect06.mp4	hi and welcome back this is section 6 and we'r...
21	Mod03_Sect07_part1.mp4	hi welcome back to module 3 in this section we...
22	Mod03_Sect07_part2.mp4	hi welcome back will continue exploring how to...
23	Mod03_Sect07_part3.mp4	hi welcome back we'll continue exploring how t...
24	Mod03_Sect08.mp4	hi and welcome back to module 3 this is Sectio...
25	Mod03_WrapUp.mp4	it's now time to review the module and wrap up...
26	Mod04_Intro.mp4	hi and welcome to module 4 of AWS Academy mach...
27	Mod04_Sect01.mp4	hi and welcome to section 1 we'll get started ...
28	Mod04_Sect02_part1.mp4	hi and welcome back this is section 2 and we'r...
29	Mod04_Sect02_part2.mp4	hi welcome back we'll continue exploring wrang...
30	Mod04_Sect02_part3.mp4	hi and welcome back in this section we'll look...
31	Mod04_WrapUp.mp4	hi welcome back it's now time to review the mo...
32	Mod05_Intro.mp4	welcome back to AWS Academy machine learning t...

	File	Data
33	Mod05_Sect01_ver2.mp4	hi welcome back this is Section 1 and we're go...
34	Mod05_Sect02_part1_ver2.mp4	welcome back in this section will explore imag...
35	Mod05_Sect02_part2.mp4	hi welcome back we'll continue exploring image...
36	Mod05_Sect03_part1.mp4	in this section we'll look at preparing custom...
37	Mod05_Sect03_part2.mp4	hi welcome back we'll continue exploring video...
38	Mod05_Sect03_part3.mp4	hi welcome back we'll continue exploring video...
39	Mod05_Sect03_part4_ver2.mp4	hi welcome back we'll continue exploring video...
40	Mod05_WrapUp_ver2.mp4	it's now time to summarize some of the main po...
41	Mod06_Intro.mp4	hi and welcome to module 6 of AWS Academy mach...
42	Mod06_Sect01.mp4	we'll get started by reviewing what natural la...
43	Mod06_Sect02.mp4	welcome back in this section will review 5 man...
44	Mod06_WrapUp.mp4	welcome back it's now time to review the modul...
45	Mod07_Sect01.mp4	welcome to module 7 course wrap up congratulat...

```
In [6]: # Import necessary Libraries
import re
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from textblob import TextBlob
from gensim.models import LdaModel
from gensim import corpora

nltk.download('punkt')
nltk.download('stopwords')
!python -m textblob.download_corpora

# Normalizing text using Stemming
def normalization(text):
    text = text.lower()
    text = re.sub(r'^\w\s', '', text)
    tokens = word_tokenize(text)
    stop_words = set(stopwords.words('english'))
    filtered_tokens = [word for word in tokens if word not in stop_words]

    lemmaObj = WordNetLemmatizer()
    lemmas = [lemmaObj.lemmatize(word) for word in filtered_tokens]

    return lemmas

df_videos['Norm Text'] = df_videos['Data'].apply(normalization)
df_videos
```

```
[nltk_data] Downloading package punkt to /home/ec2-user/nltk_data...
[nltk_data]  Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package stopwords to
[nltk_data]      /home/ec2-user/nltk_data...
[nltk_data]  Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package brown to /home/ec2-user/nltk_data...
[nltk_data]  Unzipping corpora/brown.zip.
[nltk_data] Downloading package punkt to /home/ec2-user/nltk_data...
[nltk_data]  Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to /home/ec2-user/nltk_data...
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]      /home/ec2-user/nltk_data...
[nltk_data]  Unzipping taggers/averaged_perceptron_tagger.zip.
[nltk_data] Downloading package conll2000 to
[nltk_data]      /home/ec2-user/nltk_data...
[nltk_data]  Unzipping corpora/conll2000.zip.
[nltk_data] Downloading package movie_reviews to
[nltk_data]      /home/ec2-user/nltk_data...
[nltk_data]  Unzipping corpora/movie_reviews.zip.
Finished.
```

Out[6]:

	File	Data	Norm Text
0	Mod01_Course Overview.mp4	hi and welcome to Amazon Academy machine learn...	[hi, welcome, amazon, academy, machine, learni...
1	Mod02_Intro.mp4	hi and welcome to module 2 of AWS Academy mach...	[hi, welcome, module, 2, aws, academy, machine...
2	Mod02_Sect01.mp4	hi and welcome to section 1 in this section we...	[hi, welcome, section, 1, section, going, talk...
3	Mod02_Sect02.mp4	hi and welcome back in this section we're goin...	[hi, welcome, back, section, going, look, type...
4	Mod02_Sect03.mp4	I am welcome back this is Section 3 and we're ...	[welcome, back, section, 3, going, give, quick...
5	Mod02_Sect04.mp4	welcome back in this section we'll look at som...	[welcome, back, section, well, look, tool, you...
6	Mod02_Sect05.mp4	hi welcome back this is Section 5 and we're go...	[hi, welcome, back, section, 5, going, discus...
7	Mod02_WrapUp.mp4	it's now time to review the module here are th...	[time, review, module, main, takeaway, module,...
8	Mod03_Intro.mp4	welcome back to AWS Academy machine learning t...	[welcome, back, aws, academy, machine, learnin...
9	Mod03_Sect01.mp4	hi and welcome back to module 3 this is Sectio...	[hi, welcome, back, module, 3, section, 1, goi...
10	Mod03_Sect02_part1.mp4	hi welcome back we're now going to look at a f...	[hi, welcome, back, going, look, way, collect...
11	Mod03_Sect02_part2.mp4	hi welcome back will continue exploring data c...	[hi, welcome, back, continue, exploring, data,...
12	Mod03_Sect02_part3.mp4	hi welcome back will continue exploring data c...	[hi, welcome, back, continue, exploring, data,...
13	Mod03_Sect03_part1.mp4	hi and welcome back this is Section 3 and we'r...	[hi, welcome, back, section, 3, going, cover, ...
14	Mod03_Sect03_part2.mp4	hi welcome back we'll continue exploring how t...	[hi, welcome, back, well, continue, exploring,...
15	Mod03_Sect03_part3.mp4	hi welcome back now we'll review how to find c...	[hi, welcome, back, well, review, find, correl...
16	Mod03_Sect04_part1.mp4	hi and welcome to section 4 in this section we...	[hi, welcome, section, 4, section, going, look...
17	Mod03_Sect04_part2.mp4	hi welcome back we'll continue exploring featu...	[hi, welcome, back, well, continue, exploring,...
18	Mod03_Sect04_part3.mp4	hi welcome back we'll continue exploring featu...	[hi, welcome, back, well, continue, exploring,...
19	Mod03_Sect05.mp4	hi welcome back to module 3 this is Section 5 ...	[hi, welcome, back, module, 3, section, 5, tra...
20	Mod03_Sect06.mp4	hi and welcome back this is section 6 and we'r...	[hi, welcome, back, section, 6, going, look, h...

		File	Data	Norm Text
21		Mod03_Sect07_part1.mp4	hi welcome back to module 3 in this section we...	[hi, welcome, back, module, 3, section, well, ...]
22		Mod03_Sect07_part2.mp4	hi welcome back will continue exploring how to...	[hi, welcome, back, continue, exploring, evalua...
23		Mod03_Sect07_part3.mp4	hi welcome back we'll continue exploring how t...	[hi, welcome, back, well, continue, exploring,...]
24		Mod03_Sect08.mp4	hi and welcome back to module 3 this is Sectio...	[hi, welcome, back, module, 3, section, 8, sec...]
25		Mod03_WrapUp.mp4	it's now time to review the module and wrap up...	[time, review, module, wrap, knowledge, check,...]
26		Mod04_Intro.mp4	hi and welcome to module 4 of AWS Academy mach...	[hi, welcome, module, 4, aws, academy, machine...]
27		Mod04_Sect01.mp4	hi and welcome to section 1 we'll get started ...	[hi, welcome, section, 1, well, get, started, ...]
28		Mod04_Sect02_part1.mp4	hi and welcome back this is section 2 and we'r...	[hi, welcome, back, section, 2, going, focus, ...]
29		Mod04_Sect02_part2.mp4	hi welcome back we'll continue exploring wrang...	[hi, welcome, back, well, continue, exploring,...]
30		Mod04_Sect02_part3.mp4	hi and welcome back in this section we'll look...	[hi, welcome, back, section, well, look, use, ...]
31		Mod04_WrapUp.mp4	hi welcome back it's now time to review the mo...	[hi, welcome, back, time, review, module, wrap...]
32		Mod05_Intro.mp4	welcome back to AWS Academy machine learning t...	[welcome, back, aws, academy, machine, learnin...]
33		Mod05_Sect01_ver2.mp4	hi welcome back this is Section 1 and we're go...	[hi, welcome, back, section, 1, going, introdu...]
34		Mod05_Sect02_part1_ver2.mp4	welcome back in this section will explore imag...	[welcome, back, section, explore, image, analy...]
35		Mod05_Sect02_part2.mp4	hi welcome back we'll continue exploring image...	[hi, welcome, back, well, continue, exploring,...]
36		Mod05_Sect03_part1.mp4	in this section we'll look at preparing custom...	[section, well, look, preparing, custom, data,...]
37		Mod05_Sect03_part2.mp4	hi welcome back we'll continue exploring video...	[hi, welcome, back, well, continue, exploring,...]
38		Mod05_Sect03_part3.mp4	hi welcome back we'll continue exploring video...	[hi, welcome, back, well, continue, exploring,...]
39		Mod05_Sect03_part4_ver2.mp4	hi welcome back we'll continue exploring video...	[hi, welcome, back, well, continue, exploring,...]
40		Mod05_WrapUp_ver2.mp4	it's now time to summarize some of the main po...	[time, summarize, main, point, module, module,...]
41		Mod06_Intro.mp4	hi and welcome to module 6 of AWS Academy mach...	[hi, welcome, module, 6, aws, academy, machine...]

	File	Data	Norm Text
42	Mod06_Sect01.mp4	we'll get started by reviewing what natural la...	[well, get, started, reviewing, natural, langu...
43	Mod06_Sect02.mp4	welcome back in this section will review 5 man...	[welcome, back, section, review, 5, manage, ma...
44	Mod06_WrapUp.mp4	welcome back it's now time to review the modul...	[welcome, back, time, review, module, wrap, su...
45	Mod07_Sect01.mp4	welcome to module 7 course wrap up congratulat...	[welcome, module, 7, course, wrap, congratulat...

## 4. Extracting key phrases and topics

([Go to top](#))

Use this section to extract the key phrases and topics from the videos.

```
In [7]: # Write your answer/code here

dict = corpora.Dictionary(df_videos['Norm Text'])
corpus = [dict.doc2bow(text) for text in df_videos['Norm Text']]
mod = LdaModel(corpus, num_topics=2, id2word=dict)
topics = mod.print_topics(num_words = 3)
top_nms = [', '.join(re.findall(r'"([^\"]*)"', topic[1])) for topic in topics]

def key_phrase_ext(text):
    blob = TextBlob(' '.join(text))
    noun_phrases = blob.noun_phrases
    noun_phrases = [phrase.lower() for phrase in noun_phrases]
    return noun_phrases[:10]

def topic_ext(text):
    doc_bow = dict.doc2bow(text)
    doc_topics = mod.get_document_topics(doc_bow)
    srted_tops = sorted(doc_topics, key = lambda x: x[1], reverse = True)
    primary_tops = [top_nms[topic[0]] for topic in srted_tops[:10]]
    return primary_tops

df_videos['Key Phrases'] = df_videos['Norm Text'].apply(key_phrase_ext)
df_videos['Topics'] = df_videos['Norm Text'].apply(topic_ext)
df_videos
```

Out[7]:

	File	Data	Norm Text	Key Phrases	Topics
0	Mod01_Course Overview.mp4	hi and welcome to Amazon Academy machine learn...	[hi, welcome, amazon, academy, machine, learni...	[amazon academy machine learning foundation mo...	[data, learning, machine]
1	Mod02_Intro.mp4	hi and welcome to module 2 of AWS Academy mach...	[hi, welcome, module, 2, aws, academy, machine...	[aws academy machine learning module, introduc...	[data, learning, machine]
2	Mod02_Sect01.mp4	hi and welcome to section 1 in this section we...	[hi, welcome, section, 1, section, going, talk...	[machine learning course introduction machine ...	[data, learning, machine]
3	Mod02_Sect02.mp4	hi and welcome back in this section we're goin...	[hi, welcome, back, section, going, look, type...	[type business problem machine learning, machi...	[data, learning, machine]
4	Mod02_Sect03.mp4	I am welcome back this is Section 3 and we're ...	[welcome, back, section, 3, going, give, quick...	[quick highlevel overview machine learning ter...	[data, learning, machine, data, model, set]
5	Mod02_Sect04.mp4	welcome back in this section we'll look at som...	[welcome, back, section, well, look, tool, you...	[tool youll, rest course, list isnt exhaustive...	[data, learning, machine]
6	Mod02_Sect05.mp4	hi welcome back this is Section 5 and we're go...	[hi, welcome, back, section, 5, going, discus,...	[discuss challenge machine learning youll, chal...	[data, learning, machine]
7	Mod02_WrapUp.mp4	it's now time to review the module here are th...	[time, review, module, main, takeaway, module,...	[time review module, main takeaway module, mac...	[data, learning, machine]
8	Mod03_Intro.mp4	welcome back to AWS Academy machine learning t...	[welcome, back, aws, academy, machine, learnin...	[aws academy machine learning module, entire m...	[data, learning, machine]
9	Mod03_Sect01.mp4	hi and welcome back to module 3 this is Sectio...	[hi, welcome, back, module, 3, section, 1, goi...	[guidance formulate business problem, reminder...	[data, learning, machine]
10	Mod03_Sect02_part1.mp4	hi welcome back we're now going to look at a f...	[hi, welcome, back, going, look, way, collect,...	[data section, explore technique challenge, da...	[data, learning, machine]

		File	Data	Norm Text	Key Phrases	Topics
11		Mod03_Sect02_part2.mp4	hi welcome back will continue exploring data c...	[hi, welcome, back, continue, exploring, data,...	[data collection, extract transform load data ...	[data, learning, machine, data, model, set]
12		Mod03_Sect02_part3.mp4	hi welcome back will continue exploring data c...	[hi, welcome, back, continue, exploring, data,...	[data collection, security data, public real d...	[data, model, set]
13		Mod03_Sect03_part1.mp4	hi and welcome back this is Section 3 and we'r...	[hi, welcome, back, section, 3, going, cover, ...	[data section, different data format type, vis...	[data, learning, machine]
14		Mod03_Sect03_part2.mp4	hi welcome back we'll continue exploring how t...	[hi, welcome, back, well, continue, exploring,...	[describe data data, readable format, descript...	[data, learning, machine, data, model, set]
15		Mod03_Sect03_part3.mp4	hi welcome back now we'll review how to find c...	[hi, welcome, back, well, review, find, correl...	[correlation data, quantify linear relationshi...	[data, model, set]
16		Mod03_Sect04_part1.mp4	hi and welcome to section 4 in this section we...	[hi, welcome, section, 4, section, going, look...	[impactful thing, machine learning model, extr...	[data, learning, machine]
17		Mod03_Sect04_part2.mp4	hi welcome back we'll continue exploring featu...	[hi, welcome, back, well, continue, exploring,...	[clean data, string data numerical data youll ...	[data, model, set]
18		Mod03_Sect04_part3.mp4	hi welcome back we'll continue exploring featu...	[hi, welcome, back, well, continue, exploring,...	[work outlier, clean data, outlier exists outl...	[data, model, set]
19		Mod03_Sect05.mp4	hi welcome back to module 3 this is Section 5 ...	[hi, welcome, back, module, 3, section, 5, tra...	[training section, select model train data, po...	[data, learning, machine]
20		Mod03_Sect06.mp4	hi and welcome back this is section 6 and we'r...	[hi, welcome, back, section, 6, going, look, h...	[model section, deploy train model, applicatio...	[data, model, set, data, learning, machine]

		File	Data	Norm Text	Key Phrases	Topics
21		Mod03_Sect07_part1.mp4	hi welcome back to module 3 in this section we...	[hi, welcome, back, module, 3, section, well, ...]	[model success, result point youve, model time...]	[data, model, set]
22		Mod03_Sect07_part2.mp4	hi welcome back will continue exploring how to...	[hi, welcome, back, continue, exploring, evalu...]	[model diagram show confusion matrix, differen...]	[data, model, set]
23		Mod03_Sect07_part3.mp4	hi welcome back we'll continue exploring how t...	[hi, welcome, back, well, continue, exploring,...]	[model classification model, return probabilit...]	[data, model, set]
24		Mod03_Sect08.mp4	hi and welcome back to module 3 this is Sectio...	[hi, welcome, back, module, 3, section, 8, sec...]	[tune model hyperparameters, model performance...]	[data, learning, machine, data, model, set]
25		Mod03_WrapUp.mp4	it's now time to review the module and wrap up...	[time, review, module, wrap, knowledge, check,...]	[time review module wrap knowledge check modul...]	[data, learning, machine, data, model, set]
26		Mod04_Intro.mp4	hi and welcome to module 4 of AWS Academy mach...	[hi, welcome, module, 4, aws, academy, machine...]	[aws academy machine learning module, time ser...]	[data, learning, machine, data, model, set]
27		Mod04_Sect01.mp4	hi and welcome to section 1 we'll get started ...	[hi, welcome, section, 1, well, get, started, ...]	[important area machine learning, future outco...]	[data, learning, machine, data, model, set]
28		Mod04_Sect02_part1.mp4	hi and welcome back this is section 2 and we'r...	[hi, welcome, back, section, 2, going, focus, ...]	[focus processing time series data, different ...]	[data, model, set]
29		Mod04_Sect02_part2.mp4	hi welcome back we'll continue exploring wrang...	[hi, welcome, back, well, continue, exploring,...]	[time series data seasonality data kind, obser...]	[data, model, set]
30		Mod04_Sect02_part3.mp4	hi and welcome back in this section we'll look...	[hi, welcome, back, section, well, look, use, ...]	[predictor generate, machine learning developm...]	[data, learning, machine, data, model, set]

	File	Data	Norm Text	Key Phrases	Topics
31	Mod04_WrapUp.mp4	hi welcome back it's now time to review the mo...	[hi, welcome, back, time, review, module, wrap...]	[time review module wrap module, describe busi...	[data, learning, machine, data, model, set]
32	Mod05_Intro.mp4	welcome back to AWS Academy machine learning t...	[welcome, back, aws, academy, machine, learnin...]	[aws academy machine learning module, great to...]	[data, learning, machine, data, model, set]
33	Mod05_Sect01_ver2.mp4	hi welcome back this is Section 1 and we're go...	[hi, welcome, back, section, 1, going, introdu...]	[introduce computer vision computer vision, ex...]	[data, learning, machine, data, model, set]
34	Mod05_Sect02_part1_ver2.mp4	welcome back in this section will explore imag...	[welcome, back, section, explore, image, analy...]	[section explore image analysis detail part, v...]	[data, learning, machine, data, model, set]
35	Mod05_Sect02_part2.mp4	hi welcome back we'll continue exploring image...	[hi, welcome, back, well, continue, exploring,...]	[image analysis, facial detection facial detec...]	[data, model, set]
36	Mod05_Sect03_part1.mp4	in this section we'll look at preparing custom...	[section, well, look, preparing, custom, data,...]	[custom data, computer vision detect custom ob...]	[data, model, set, data, learning, machine]
37	Mod05_Sect03_part2.mp4	hi welcome back we'll continue exploring video...	[hi, welcome, back, well, continue, exploring,...]	[video analysis, training data, information th...]	[data, model, set, data, learning, machine]
38	Mod05_Sect03_part3.mp4	hi welcome back we'll continue exploring video...	[hi, welcome, back, well, continue, exploring,...]	[video analysis, test data, final step train m...]	[data, model, set]
39	Mod05_Sect03_part4_ver2.mp4	hi welcome back we'll continue exploring video...	[hi, welcome, back, well, continue, exploring,...]	[video analysis, model general, quality model,...]	[data, model, set]
40	Mod05_WrapUp_ver2.mp4	it's now time to summarize some of the main po...	[time, summarize, main, point, module, module,...]	[time summarize, main point module module, des...]	[data, learning, machine, data, model, set]

	File	Data	Norm Text	Key Phrases	Topics
41	Mod06_Intro.mp4	hi and welcome to module 6 of AWS Academy machine...	[hi, welcome, module, 6, aws, academy, machine...]	[aws academy machine learning introduction, na...]	[data, learning, machine]
42	Mod06_Sect01.mp4	we'll get started by reviewing what natural la...	[well, get, started, reviewing, natural, langu...]	[natural language processing, natural language...]	[data, learning, machine, data, model, set]
43	Mod06_Sect02.mp4	welcome back in this section will review 5 man...	[welcome, back, section, review, 5, manage, ma...]	[section review, manage machine learning servi...]	[data, learning, machine]
44	Mod06_WrapUp.mp4	welcome back it's now time to review the modul...	[welcome, back, time, review, module, wrap, su...]	[time review module wrap summary module, descr...]	[data, learning, machine, data, model, set]
45	Mod07_Sect01.mp4	welcome to module 7 course wrap up congratulat...	[welcome, module, 7, course, wrap, congratulat...]	[course wrap congratulation, aws academy machi...]	[data, learning, machine]

## 5. Creating the dashboard

([Go to top](#))

Use this section to create the dashboard for your solution.

```
In [8]: # Write your answer/code here

from ipywidgets import interact, widgets

def find_vid(keyword):
    if not keyword:
        return []

    relevant_vid = df_videos.copy()
    keyword = keyword.lower()
    relevant_vid['Key Phrase Match'] = relevant_vid['Key Phrases'].apply(
        lambda x: any(keyword in phrase.lower() for phrase in x))
    relevant_vid['Topic Match'] = relevant_vid['Topics'].apply(
        lambda x: any(keyword in topic.lower() for topic in x))
    relevant_vid = relevant_vid[relevant_vid['Key Phrase Match'] | relevant_vid['Topic Match']]
    return relevant_vid['File'].tolist()

search_input = widgets.Text(placeholder='Type Keyword')

def gen_output(keyword):
```

```
relevant_vid = find_vid(keyword)
if relevant_vid:
    print("Recommended:")
    for file_name in relevant_vid:
        print(file_name)
else:
    print("No recommendations.")

interact(gen_output, keyword = search_input)

interactive(children=(Text(value='', description='keyword', placeholder='Type Keyword'), Output()), _dom_class...
Out[8]: <function __main__.gen_output(keyword)>
```

In [ ]: