

HuggingFace dataset

The [Hugging Face Hub](#) is home to over 5,000 [datasets](#) in more than 100 languages that can be used for a broad range of tasks across NLP, Computer Vision, and Audio. They used for a diverse range of tasks such as translation, automatic speech recognition, and image classification.

This notebook shows how to load [Hugging Face Hub](#) datasets to LangChain.

```
from langchain.document_loaders import HuggingFaceDatasetLoader
```

```
dataset_name = "imdb"  
page_content_column = "text"
```



```
loader = HuggingFaceDatasetLoader(dataset_name, page_content_column)
```

```
data = loader.load()
```

```
data[:15]
```

```
[Document(page_content='I rented I AM CURIOUS-YELLOW from my video store because of all the controversy  
that surrounded it when it was first released in 1967. I also heard that at first it was seized by U.S.
```

customs if it ever tried to enter this country, therefore being a fan of films considered "controversial" I really had to see this for myself.

The plot is centered around a young Swedish drama student named Lena who wants to learn everything she can about life. In particular she wants to focus her attentions to making some sort of documentary on what the average Swede thought about certain political issues such as the Vietnam War and race issues in the United States. In between asking politicians and ordinary denizens of Stockholm about their opinions on politics, she has sex with her drama teacher, classmates, and married men.

What kills me about I AM CURIOUS-YELLOW is that 40 years ago, this was considered pornographic. Really, the sex and nudity scenes are few and far between, even then it\'s not shot like some cheaply made porno. While my countrymen mind find it shocking, in reality sex and nudity are a major staple in Swedish cinema. Even Ingmar Bergman, arguably their answer to good old boy John Ford, had sex scenes in his films.

I do commend the filmmakers for the fact that any sex shown in the film is shown for artistic purposes rather than just to shock people and make money to be shown in pornographic theaters in America. I AM CURIOUS-YELLOW is a good film for anyone wanting to study the meat and potatoes (no pun intended) of Swedish cinema. But really, this film doesn\'t have much of a plot.', metadata={'label': 0}),

Document(page_content="\"I Am Curious: Yellow\" is a risible and pretentious steaming pile. It doesn\'t matter what one\'s political views are because this film can hardly be taken seriously on any level. As for the claim that frontal male nudity is an automatic NC-17, that isn\'t true. I\'ve seen R-rated films with male nudity. Granted, they only offer some fleeting views, but where are the R-rated films with gaping vulvas and flapping labia? Nowhere, because they don\'t exist. The same goes for those crappy cable shows: schlongs swinging in the breeze but not a clitoris in sight. And those pretentious indie movies like The Brown Bunny, in which we\'re treated to the site of Vincent Gallo\'s throbbing johnson, but not a trace of pink visible on Chloe Sevigny. Before crying (or implying) \"double-standard\" in matters of nudity, the mentally obtuse should take into account one unavoidably obvious anatomical difference between men and women: there are no genitals on display when actresses appears nude, and the same cannot be said for a man. In fact, you generally won\'t see female genitals in an American film in anything short of porn or explicit erotica. This alleged double-standard is less a double standard than an admittedly depressing ability to come to terms culturally with the insides of women\'s bodies.', metadata={'label': 0}),

Document(page_content="\"If only to avoid making this type of film in the future. This film is interesting as an experiment but tells no cogent story.

One might feel virtuous for sitting thru it because it touches on so many IMPORTANT issues but it does so without any discernable motive. The viewer comes away with no new perspectives (unless one comes up with one while one\'s mind wanders, as it will invariably do during this pointless film).

One might better spend one\'s time staring out a window at a tree growing.

\", metadata={'label': 0}),

```
Document(page_content="This film was probably inspired by Godard's Masculin, féminin and I urge you to see that film instead.<br /><br />The film has two strong elements and those are, (1) the realistic acting (2) the impressive, undeservedly good, photo. Apart from that, what strikes me most is the endless stream of silliness. Lena Nyman has to be most annoying actress in the world. She acts so stupid and with all the nudity in this film,...it's unattractive. Comparing to Godard's film, intellectuality has been replaced with stupidity. Without going too far on this subject, I would say that follows from the difference in ideals between the French and the Swedish society.<br /><br />A movie of its time, and place. 2/10.", metadata={'label': 0}),
```

```
Document(page_content='Oh, brother...after hearing about this ridiculous film for umpteen years all I can think of is that old Peggy Lee song..<br /><br />"Is that all there is??" ...I was just an early teen when this smoked fish hit the U.S. I was too young to get in the theater (although I did manage to sneak into "Goodbye Columbus"). Then a screening at a local film museum beckoned - Finally I could see this film, except now I was as old as my parents were when they schlepped to see it!!<br /><br />The ONLY reason this film was not condemned to the anonymous sands of time was because of the obscenity case sparked by its U.S. release. MILLIONS of people flocked to this stinker, thinking they were going to see a sex film...Instead, they got lots of closeups of gnarly, repulsive Swedes, on-street interviews in bland shopping malls, asinine political pretension...and feeble who-cares simulated sex scenes with saggy, pale actors.<br /><br />Cultural icon, holy grail, historic artifact..whatever this thing was, shred it, burn it, then stuff the ashes in a lead box!<br /><br />Elite esthetes still scrape to find value in its boring pseudo revolutionary political spewings..But if it weren\'t for the censorship scandal, it would have been ignored, then forgotten.<br /><br />Instead, the "I Am Blank, Blank" rhythmed title was repeated endlessly for years as a titilation for porno films (I am Curious, Lavender - for gay films, I Am Curious, Black - for blaxploitation films, etc..) and every ten years or so the thing rises from the dead, to be viewed by a new generation of suckers who want to see that "naughty sex film" that "revolutionized the film industry"...<br /><br />Yeesh, avoid like the plague..Or if you MUST see it - rent the video and fast forward to the "dirty" parts, just to get it over with.<br /><br />', metadata={'label': 0}),
```

```
Document(page_content="I would put this at the top of my list of films in the category of unwatchable trash! There are films that are bad, but the worst kind are the ones that are unwatchable but you are suppose to like them because they are supposed to be good for you! The sex sequences, so shocking in its day, couldn't even arouse a rabbit. The so called controversial politics is strictly high school sophomore amateur night Marxism. The film is self-consciously arty in the worst sense of the term. The photography is in a harsh grainy black and white. Some scenes are out of focus or taken from the wrong angle. Even the sound is bad! And some people call this art?<br /><br />", metadata={'label': 0}),
```

```
Document(page_content="Whoever wrote the screenplay for this movie obviously never consulted any books about Lucille Ball, especially her autobiography. I've never seen so many mistakes in a biopic, ranging from her early years in Celoron and Jamestown to her later years with Desi. I could write a whole list of factual errors, but it would go on for pages. In all, I believe that Lucille Ball is one of those inimitable people who simply cannot be portrayed by anyone other than themselves. If I were Lucie Arnaz and Desi, Jr., I would be irate at how many mistakes were made in this film. The filmmakers tried hard, but the movie seems awfully sloppy to me.", metadata={'label': 0})),
```

```
Document(page_content='When I first saw a glimpse of this movie, I quickly noticed the actress who was playing the role of Lucille Ball. Rachel York\'s portrayal of Lucy is absolutely awful. Lucille Ball was an astounding comedian with incredible talent. To think about a legend like Lucille Ball being portrayed the way she was in the movie is horrendous. I cannot believe out of all the actresses in the world who could play a much better Lucy, the producers decided to get Rachel York. She might be a good actress in other roles but to play the role of Lucille Ball is tough. It is pretty hard to find someone who could resemble Lucille Ball, but they could at least find someone a bit similar in looks and talent. If you noticed York\'s portrayal of Lucy in episodes of I Love Lucy like the chocolate factory or vitavetavegamin, nothing is similar in any way-her expression, voice, or movement.<br /><br />To top it all off, Danny Pino playing Desi Arnaz is horrible. Pino does not qualify to play as Ricky. He\'s small and skinny, his accent is unreal, and once again, his acting is unbelievable. Although Fred and Ethel were not similar either, they were not as bad as the characters of Lucy and Ricky.<br /><br />Overall, extremely horrible casting and the story is badly told. If people want to understand the real life situation of Lucille Ball, I suggest watching A&E Biography of Lucy and Desi, read the book from Lucille Ball herself, or PBS\' American Masters: Finding Lucy. If you want to see a docudrama, "Before the Laughter" would be a better choice. The casting of Lucille Ball and Desi Arnaz in "Before the Laughter" is much better compared to this. At least, a similar aspect is shown rather than nothing.', metadata={'label': 0})),
```

```
Document(page_content='Who are these "They"- the actors? the filmmakers? Certainly couldn\'t be the audience- this is among the most air-puffed productions in existence. It\'s the kind of movie that looks like it was a lot of fun to shoot\x97 TOO much fun, nobody is getting any actual work done, and that almost always makes for a movie that\'s no fun to watch.<br /><br />Ritter dons glasses so as to hammer home his character\'s status as a sort of doppelganger of the bespectacled Bogdanovich; the scenes with the breezy Ms. Stratten are sweet, but have an embarrassing, look-guys-I\'m-dating-the-prom-queen feel to them. Ben Gazzara sports his usual cat\'s-got-canary grin in a futile attempt to elevate the meager plot, which requires him to pursue Audrey Hepburn with all the interest of a narcoleptic at an insomnia clinic. In the meantime, the budding couple\'s respective children (nepotism alert: Bogdanovich\'s daughters) spew cute and pick up some
```

fairly disturbing pointers on 'love' while observing their parents. (Ms. Hepburn, drawing on her dignity, manages to rise above the proceedings- but she has the monumental challenge of playing herself, ostensibly.) Everybody looks great, but so what? It's a movie and we can expect that much, if that's what you're looking for you'd be better off picking up a copy of Vogue.

Oh- and it has to be mentioned that Colleen Camp thoroughly annoys, even apart from her singing, which, while competent, is wholly unconvincing... the country and western numbers are woefully mismatched with the standards on the soundtrack. Surely this is NOT what Gershwin (who wrote the song from which the movie's title is derived) had in mind; his stage musicals of the 20's may have been slight, but at least they were long on charm. "They All Laughed" tries to coast on its good intentions, but nobody- least of all Peter Bogdanovich - has the good sense to put on the brakes.

Due in no small part to the tragic death of Dorothy Stratten, this movie has a special place in the heart of Mr. Bogdanovich- he even bought it back from its producers, then distributed it on his own and went bankrupt when it didn't prove popular. His rise and fall is among the more sympathetic and tragic of Hollywood stories, so there's no joy in criticizing the film... there _is_ real emotional investment in Ms. Stratten's scenes. But "Laughed" is a faint echo of "The Last Picture Show", "Paper Moon" or "What's Up, Doc"- following "Daisy Miller" and "At Long Last Love", it was a thundering confirmation of the phase from which P.B. has never emerged.

All in all, though, the movie is harmless, only a waste of rental. I want to watch people having a good time, I'll go to the park on a sunny day. For filmic expressions of joy and love, I'll stick to Ernest Lubitsch and Jaques Demy...', metadata={'label': 0}),

Document(page_content="This is said to be a personal film for Peter Bogdonavitch. He based it on his life but changed things around to fit the characters, who are detectives. These detectives date beautiful models and have no problem getting them. Sounds more like a millionaire playboy filmmaker than a detective, doesn't it? This entire movie was written by Peter, and it shows how out of touch with real people he was. You're supposed to write what you know, and he did that, indeed. And leaves the audience bored and confused, and jealous, for that matter. This is a curio for people who want to see Dorothy Stratten, who was murdered right after filming. But Patti Hanson, who would, in real life, marry Keith Richards, was also a model, like Stratten, but is a lot better and has a more ample part. In fact, Stratten's part seemed forced; added. She doesn't have a lot to do with the story, which is pretty convoluted to begin with. All in all, every character in this film is somebody that very few people can relate with, unless you're millionaire from Manhattan with beautiful supermodels at your beckon call. For the rest of us, it's an irritating snore fest. That's what happens when you're out of touch. You entertain your few friends with inside jokes, and bore all the rest.", metadata={'label': 0}),

Document(page_content='It was great to see some of my favorite stars of 30 years ago including John

Ritter, Ben Gazzarra and Audrey Hepburn. They looked quite wonderful. But that was it. They were not given any characters or good lines to work with. I neither understood or cared what the characters were doing.

Some of the smaller female roles were fine, Patty Henson and Colleen Camp were quite competent and confident in their small sidekick parts. They showed some talent and it is sad they didn't go on to star in more and better films. Sadly, I didn't think Dorothy Stratten got a chance to act in this her only important film role.

The film appears to have some fans, and I was very open-minded when I started watching it. I am a big Peter Bogdanovich fan and I enjoyed his last movie, "Cat's Meow" and all his early ones from "Targets" to "Nickleodeon". So, it really surprised me that I was barely able to keep awake watching this one.

It is ironic that this movie is about a detective agency where the detectives and clients get romantically involved with each other. Five years later, Bogdanovich's ex-girlfriend, Cybil Shepherd had a hit television series called "Moonlighting" stealing the story idea from Bogdanovich. Of course, there was a great difference in that the series relied on tons of witty dialogue, while this tries to make do with slapstick and a few screwball lines.

Bottom line: It ain't no "Paper Moon" and only a very pale version of "What's Up, Doc".', metadata={'label': 0})),

Document(page_content="I can't believe that those praising this movie herein aren't thinking of some other film. I was prepared for the possibility that this would be awful, but the script (or lack thereof) makes for a film that's also pointless. On the plus side, the general level of craft on the part of the actors and technical crew is quite competent, but when you've got a sow's ear to work with you can't make a silk purse. Ben G fans should stick with just about any other movie he's been in. Dorothy S fans should stick to Galaxina. Peter B fans should stick to Last Picture Show and Target. Fans of cheap laughs at the expense of those who seem to be asking for it should stick to Peter B's amazingly awful book, Killing of the Unicorn.", metadata={'label': 0})),

Document(page_content='Never cast models and Playboy bunnies in your films! Bob Fosse's "Star 80" about Dorothy Stratten, of whom Bogdanovich was obsessed enough to have married her SISTER after her murder at the hands of her low-life husband, is a zillion times more interesting than Dorothy herself on the silver screen. Patty Hansen is no actress either..I expected to see some sort of lost masterpiece a la Orson Welles but instead got Audrey Hepburn cavorting in jeans and a god-awful "poodlesque" hair-do....Very disappointing...."Paper Moon" and "The Last Picture Show" I could watch again and again. This clunker I could barely sit through once. This movie was reputedly not released because of the brouhaha surrounding Ms. Stratten's tawdry death; I think the real reason was because it was so bad!', metadata={'label': 0})),

Document(page_content="Its not the cast. A finer group of actors, you could not find. Its not the setting. The director is in love with New York City, and by the end of the film, so are we all! Woody Allen could not improve upon what Bogdonovich has done here. If you are going to fall in love, or find love,

```
Manhattan is the place to go. No, the problem with the movie is the script. There is none. The actors fall in love at first sight, words are unnecessary. In the director's own experience in Hollywood that is what happens when they go to work on the set. It is reality to him, and his peers, but it is a fantasy to most of us in the real world. So, in the end, the movie is hollow, and shallow, and message-less.", metadata={'label': 0}},
```

```
Document(page_content='Today I found "They All Laughed" on VHS on sale in a rental. It was a really old and very used VHS, I had no information about this movie, but I liked the references listed on its cover: the names of Peter Bogdanovich, Audrey Hepburn, John Ritter and specially Dorothy Stratten attracted me, the price was very low and I decided to risk and buy it. I searched IMDb, and the User Rating of 6.0 was an excellent reference. I looked in "Mick Martin & Marsha Porter Video & DVD Guide 2003" and \x96 wow \x96 four stars! So, I decided that I could not waste more time and immediately see it. Indeed, I have just finished watching "They All Laughed" and I found it a very boring overrated movie. The characters are badly developed, and I spent lots of minutes to understand their roles in the story. The plot is supposed to be funny (private eyes who fall in love for the women they are chasing), but I have not laughed along the whole story. The coincidences, in a huge city like New York, are ridiculous. Ben Gazarra as an attractive and very seductive man, with the women falling for him as if her were a Brad Pitt, Antonio Banderas or George Clooney, is quite ridiculous. In the end, the greater attractions certainly are the presence of the Playboy centerfold and playmate of the year Dorothy Stratten, murdered by her husband pretty after the release of this movie, and whose life was showed in "Star 80" and "Death of a Centerfold: The Dorothy Stratten Story"; the amazing beauty of the sexy Patti Hansen, the future Mrs. Keith Richards; the always wonderful, even being fifty-two years old, Audrey Hepburn; and the song "Amigo", from Roberto Carlos. Although I do not like him, Roberto Carlos has been the most popular Brazilian singer since the end of the 60\'s and is called by his fans as "The King". I will keep this movie in my collection only because of these attractions (manly Dorothy Stratten). My vote is four.<br /><br />Title (Brazil): "Muito Riso e Muita Alegria" ("Many Laughs and Lots of Happiness")', metadata={'label': 0})]
```

Example

In this example, we use data from a dataset to answer a question

```
from langchain.indexes import VectorstoreIndexCreator
```

```
from langchain.document_loaders.hugging_face_dataset import HuggingFaceDatasetLoader
```

```
dataset_name = "tweet_eval"  
page_content_column = "text"  
name = "stance_climate"
```

```
loader = HuggingFaceDatasetLoader(dataset_name, page_content_column, name)
```

```
index = VectorstoreIndexCreator().from_loaders([loader])
```

Found cached dataset tweet_eval

0%| | 0/3 [00:00<?, ?it/s]

Using embedded DuckDB without persistence: data will be transient

```
query = "What are the most used hashtag?"  
result = index.query(query)
```

```
result
```


' The most used hashtags in this context are #UKClimate2015, #Sustainability, #TakeDownTheFlag, #LoveWins, #CSOTA, #ClimateSummitoftheAmericas, #SM, and #SocialMedia.'