In [ ]:
```
import pip
def import_or_install(package):
    try:
        __import__(package)
    except ImportError:
        pip.main(['install', package])
```

In [ ]:
```
import_or_install("datasets")
import_or_install("pyspark.sql")

import warnings

warnings.filterwarnings("ignore")
```

In [ ]:
```
from datasets import load_dataset
from pyspark.sql import SparkSession
dataset = load_dataset("jxm/the_office_lines")

warnings.filterwarnings("ignore")
```

```
Using custom data configuration jxm--the_office_lines-b0e578f6723864f0
Found cached dataset parquet (/Users/hakancangunerli/.cache/huggingface/datase
ts/jxm___parquet/jxm--the_office_lines-b0e578f6723864f0/0.0.0/2a3b91fbd88a2c90
d1dbbb32b460cf621d31bd5b05b934492fdef7d8d6f236ec)
  0%|          | 0/3 [00:00<?, ?it/s]
```

In [ ]:
```
spark = SparkSession.builder.appName("LoadDataset").getOrCreate()

dataset = load_dataset("jxm/the_office_lines")

df_test, df_train, df_valid = spark.createDataFrame(dataset["test"]), spark.cre

warnings.filterwarnings("ignore")
```

```
Using custom data configuration jxm--the_office_lines-b0e578f6723864f0
Found cached dataset parquet (/Users/hakancangunerli/.cache/huggingface/datase
ts/jxm___parquet/jxm--the_office_lines-b0e578f6723864f0/0.0.0/2a3b91fbd88a2c90
d1dbbb32b460cf621d31bd5b05b934492fdef7d8d6f236ec)
  0%|          | 0/3 [00:00<?, ?it/s]
```

In [ ]:
```
df_test.show()

# df = spark.createDataFrame(dataset["train"])
```

```
+-------+-------+-----+--------------------+-----+------+-------+
|deleted|episode|   id|           line_text|scene|season|speaker|
+-------+-------+-----+--------------------+-----+------+-------+
|  false|     13|13403|[lurking by the b...|   12|     3|   Andy|
|  false|     20|28990| Okay, striker, huh?|   31|     5| Chares|
|  false|      1|45805|Did you guys figu...|   27|     8| Angela|
|  false|     15|50014|Stanley! Wake up!...|    4|     8| Dwight|
|  false|      1|30938|               Yeah!|   19|     6|Michael|
|  false|      7|47629|           The temp?|   10|     8|    Jim|
|  false|      9| 5030|Alright.  Oscar, ...|   21|     2|Michael|
|  false|     10|55483|               Yeah.|   28|     9| Darryl|
|  false|     22|37674|Donna, come... Ah...|   19|     6| Dwight|
|  false|      4| 3216|I would take The ...|   23|     2| Angela|
|  false|      8|40513|   Which one's Glee?|   19|     7|Phyllis|
|  false|     24|52770|Do not bring Shak...|   39|     8|   Andy|
|  false|     15|13836|                I...|    6|     3| Dwight|
|  false|      5| 1326|Michael, look. [D...|   10|     1| Dwight|
|  false|     23|38066|               Whoa!|   28|     6|  Shane|
|  false|      1|38471|       Give me that.|   15|     7| Dwight|
|  false|     12|21670|But what do you d...|   25|     4|Stanley|
|  false|      2|53286|Just, you sure Cl...|   45|     9|   Andy|
|  false|      1|17013|Was she talkin' b...|   10|     4| Dwight|
|  false|     23|16811|               Yeah?|   63|     3|  Karen|
+-------+-------+-----+--------------------+-----+------+-------+
only showing top 20 rows
```

In [ ]:
```
df_train.show()
```

```
+-------+-------+-----+--------------------+-----+------+--------+
|deleted|episode|   id|           line_text|scene|season| speaker|
+-------+-------+-----+--------------------+-----+------+--------+
|  false|      2|18025|I like it a lot. ...|   60|     4|    Ryan|
|  false|     11|26382|Go away, Tuna! I'...|   31|     5|    Andy|
|  false|     12|42015|Learn to cook for...|   14|     7|     Pam|
|  false|     19|36621|     Hey, what's up?|    7|     6|  Darryl|
|  false|     26|30561|        Hey Charles.|   15|     5|     Pam|
|  false|     12|56088|Oh, no. I took a ...|   44|     9|     Pam|
|  false|      9| 4983|Question: on the ...|   19|     2|  Dwight|
|  false|     10|20930|You're asking wha...|   29|     4|Margaret|
|  false|      8|47927|Unless you're goi...|   16|     8|    Andy|
|  false|      2|23433|No.  Something I ...|   29|     5|Meredith|
|  false|     19|44012|Alright. This is ...|   22|     7|Deangelo|
|  false|     14|22474|[hands Meredith h...|   75|     4|    Toby|
|  false|     20|51534|What? No! No, I j...|   16|     8|     Pam|
|  false|      5|24487|[holding camera] ...|   50|     5|  Darryl|
|  false|     23|52372|       No, I do not.|   21|     8|    Andy|
|  false|     13|27065|When you're a kid...|   67|     5|     Pam|
|  false|      6|54386|Please sir! Spare...|   37|     9|  Dwight|
|  false|     23|45063|Dwight, are you c...|   20|     7|     Pam|
|  false|     15|35250|I think you [to M...|   21|     6|      Jo|
|  false|     12|34403|Yes, we are unvei...|    3|     6|  Dwight|
+-------+-------+-----+--------------------+-----+------+--------+
only showing top 20 rows
```

In [ ]:
```
df_valid.show()
```

deleted	episode	id	line_text	scene	season	speaker
false	5	3735	No, you have the ...	32	2	Creed
false	1	117	Go ahead.	24	1	Dwight
false	3	31515	You know, there's...	16	6	Michael
false	23	52391	No. I���m a rogue.	27	8	Andy
false	21	15765	I'll check the web.	2	3	Andy
false	5	39788	Can I talk to you...	31	7	Dwight
false	16	35882	Okay. So, I'm ju...	59	6	Pam
false	22	44919	I'm no MJ. I can...	33	7	Deangelo
false	21	15766	[on the phone] Th...	2	3	Jim
false	4	32203	Heard you might h...	71	6	Tom
false	7	54657	Molly. I am not T...	32	9	Jan
false	16	7558	Yeah, Happy Valen...	41	2	Michael
false	17	28085	Yes it is, but it...	26	5	Michael
false	1	217	Yeah.	39	1	Pam
false	4	31941	To waiting.	31	6	Jim
false	17	43302	[peering into con...	5	7	Michael
true	3	825	We stole Dwight's...	43	1	Pam
false	1	45737	Isn���t it amazin...	11	8	Angela
false	18	7921	No thanks.	15	2	Abby
false	20	15375	This day is banan...	9	3	Kelly

only showing top 20 rows