Assignment

January 9, 2022

0.0.1 Pyspark Assignment

Group 13

0.0.2 ETL using Pyspark

Extract

- Read input from different datasource - json,csv,parquet
- from data structures - dataframe, array, others

Transform

- Data preprocessing/cleaning
- RDD - transformation - map,filter
- Dataframe fuction - groupby,filter and so on

Load

- Store as file
- store to DB

0.1 Entertainment - netflix shows analytics

i. Extract: Load the data

- Read data as pandas dataframe and then create spark dataframe and create a table view "netflix" as spark SQL

ii. Transform: Exploratory data analysis using spark sql queries

- Unique showId count
- GroupBy type,release_year and count of showId
- Update column duration values as 90 min to 90 and 2 seasons to 2 and others
- groupby type and avg durations

iii. Load: Save analysis report

`- save as tables - partitionby type`

[1]:
```
# import required libraries
import pandas as pd
# importing pyspark
```

```
import pyspark
# importing all from pyspark.sql.function
from pyspark.sql.functions import *
from pyspark import SparkConf, SparkContext, SQLContext
from pyspark.sql.types import DoubleType, IntegerType, DateType
```

[2]:
```
# import spark session
conf = SparkConf().setAppName("test").setMaster("local")
sc = SparkContext(conf=conf)
sqlContext = SQLContext(sc)
print(sc)
print(sqlContext)
```

```
<SparkContext master=local appName=test>
<pyspark.sql.context.SQLContext object at 0x7f57c510b1f0>
```

0.1.1 i. Load the data

read netflix data as pandas datframe

[3]:
```
netflix= pd.read_csv("./netflix_titles.csv")
```

[4]:
```
netflix.head()
```

[4]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban…	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi…	NaN	September 24, 2021	2021	TV-MA	1 Season	
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K…	India	September 24, 2021	2021	TV-MA	2 Seasons	

```
2  Crime TV Shows, International TV Shows, TV Act…
3                            Docuseries, Reality TV
4  International TV Shows, Romantic TV Shows, TV …

                                          description
0  As her father nears the end of his life, filmm…
1  After crossing paths at a party, a Cape Town t…
2  To protect his family from a powerful drug lor…
3  Feuds, flirtations and toilet talk go down amo…
4  In a city of coaching centers known to train I…
```

[5]:
```
netflix.shape
```

[5]:
```
(8807, 12)
```

[6]:
```
netflix['show_id'].unique
```

[6]:
```
<bound method Series.unique of 0          s1
1          s2
2          s3
3          s4
4          s5
        …
8802    s8803
8803    s8804
8804    s8805
8805    s8806
8806    s8807
Name: show_id, Length: 8807, dtype: object>
```

Convert Pandas to PySpark (Spark) DataFrame

[7]:
```
# netflix_spark = sqlContext.createDataFrame(netflix)
```

In particular some columns (for example event_dt_num) in your data have missing values which pushes Pandas to represent them as mixed types (string for not missing, NaN for missing values).

If you're in doubt it is better to read all data as strings and cast afterwards. If you have access to code book you should always provide schema to avoid problems and reduce overall cost.

Finally passing data from the driver is anti-pattern. You should be able to read this data directly using csv format (Spark 2.0.0+) or spark-csv library (Spark 1.6 and below)

[8]:
```
## Created spark datframe for netflix data
netflix_spark_df = (sqlContext.read.format("csv").options(header="true")
    .load("./netflix_titles.csv"))
```

[9]:
```
netflix_spark_df
```

[9]:
```
DataFrame[show_id: string, type: string, title: string, director: string, cast:
string, country: string, date_added: string, release_year: string, rating:
string, duration: string, listed_in: string, description: string]
```

[10]:
```
## print schema of netflix data
netflix_spark_df.printSchema()
```

```
root
 |-- show_id: string (nullable = true)
 |-- type: string (nullable = true)
 |-- title: string (nullable = true)
 |-- director: string (nullable = true)
 |-- cast: string (nullable = true)
 |-- country: string (nullable = true)
 |-- date_added: string (nullable = true)
 |-- release_year: string (nullable = true)
 |-- rating: string (nullable = true)
 |-- duration: string (nullable = true)
 |-- listed_in: string (nullable = true)
 |-- description: string (nullable = true)
```

[11]:
```
## Show data
netflix_spark_df.show()
```

```
+-------+-------+--------------------+--------------------+--------------------+
--------------------+-----------------+------------+------+---------+----------
----------+--------------------+
|show_id|   type|               title|            director|                cast|
country|        date_added|release_year|rating| duration|           listed_in|
description|
+-------+-------+--------------------+--------------------+--------------------+
--------------------+-----------------+------------+------+---------+----------
----------+--------------------+
|     s1|  Movie|Dick Johnson Is Dead|     Kirsten Johnson|                null|
United States|September 25, 2021|        2020| PG-13|   90 min|
Documentaries|As her father nea…|
|     s2|TV Show|       Blood & Water|                null|Ama Qamata, Khosi…|
South Africa|September 24, 2021|        2021| TV-MA|2 Seasons|International TV
…|After crossing pa…|
|     s3|TV Show|           Ganglands|     Julien Leclercq|Sami Bouajila, Tr…|
null|September 24, 2021|        2021| TV-MA| 1 Season|Crime TV Shows, I…|To
protect his fa…|
|     s4|TV Show|Jailbirds New Orl…|                null|                null|
null|September 24, 2021|        2021| TV-MA| 1 Season|Docuseries,
Reali…|Feuds, flirtation…|
|     s5|TV Show|        Kota Factory|                null|Mayur More, Jiten…|
India|September 24, 2021|        2021| TV-MA|2 Seasons|International TV …|In a
```

```
city of coac…|
|     s6|TV Show|       Midnight Mass|       Mike Flanagan|Kate Siegel, Zach…|
null|September 24, 2021|        2021| TV-MA| 1 Season|TV Dramas, TV Hor…|The
arrival of a …|
|     s7|  Movie|My Little Pony: A…|Robert Cullen, Jo…|Vanessa Hudgens, …|
null|September 24, 2021|        2021|    PG|   91 min|Children &
Family…|Equestria's divid…|
|     s8|  Movie|             Sankofa|        Haile Gerima|Kofi Ghanaba,
Oya…|United States, Gh…|September 24, 2021|        1993| TV-MA|  125
min|Dramas, Independe…|On a photo shoot …|
|     s9|TV Show|The Great British…|     Andy Devonshire|Mel Giedroyc, Sue…|
United Kingdom|September 24, 2021|        2021| TV-14|9 Seasons|British TV
Shows,…|A talented batch …|
|    s10|  Movie|        The Starling|      Theodore Melfi|Melissa McCarthy,…|
United States|September 24, 2021|        2021| PG-13|  104 min|    Comedies,
Dramas|A woman adjusting…|
|    s11|TV Show|Vendetta: Truth, …|                null|                null|
null|September 24, 2021|        2021| TV-MA| 1 Season|Crime TV Shows,
D…|"Sicily boasts a …|
|    s12|TV Show|    Bangkok Breaking|   Kongkiat Komesiri|Sukollawat Kanaro…|
null|September 23, 2021|        2021| TV-MA| 1 Season|Crime TV Shows,
I…|Struggling to ear…|
|    s13|  Movie|        Je Suis Karl| Christian Schwochow|Luna Wedler,
Jann…|Germany, Czech Re…|September 23, 2021|        2021| TV-MA|  127
min|Dramas, Internati…|After most of her…|
|    s14|  Movie|Confessions of an…|       Bruno Garotti|Klara Castanho, L…|
null|September 22, 2021|        2021| TV-PG|   91 min|Children & Family…|When
the clever b…|
|    s15|TV Show|Crime Stories: In…|                null|                null|
null|September 22, 2021|        2021| TV-MA| 1 Season|British TV
Shows,…|Cameras following…|
|    s16|TV Show|   Dear White People|                null|Logan Browning, B…|
United States|September 22, 2021|        2021| TV-MA|4 Seasons|TV Comedies, TV
D…|"Students of colo…|
|    s17|  Movie|Europe's Most Dan…|Pedro de Echave G…|                null|
null|September 22, 2021|        2020| TV-MA|   67 min|Documentaries,
In…|Declassified docu…|
|    s18|TV Show|     Falsa identidad|                null|Luis Ernesto Fran…|
Mexico|September 22, 2021|        2020| TV-MA|2 Seasons|Crime TV Shows,
S…|Strangers Diego a…|
|    s19|  Movie|           Intrusion|          Adam Salky|Freida Pinto, Log…|
null|September 22, 2021|        2021| TV-14|   94 min|           Thrillers|After
a deadly ho…|
|    s20|TV Show|              Jaguar|                null|Blanca Suárez, Iv…|
null|September 22, 2021|        2021| TV-MA| 1 Season|International TV …|In
the 1960s, a H…|
+-------+-------+--------------------+--------------------+--------------------+
--------------------+-----------------+------------+------+---------+----------
```

```
----------+--------------------+
only showing top 20 rows
```

0.1.2 Register netflix spark dataframe as table

[12]:
```
sqlContext.registerDataFrameAsTable(netflix_spark_df, "netflix_table")
```

[13]:
```
## Using query show table
sqlContext.sql("SELECT * FROM netflix_table ")
```

[13]:
```
DataFrame[show_id: string, type: string, title: string, director: string, cast:
 string, country: string, date_added: string, release_year: string, rating:
 string, duration: string, listed_in: string, description: string]
```

[14]:
```
## Using query show spark netflix data
netflix_raw_data = sqlContext.sql("SELECT * FROM netflix_table ")
netflix_raw_data.show()
```

```
+-------+-------+--------------------+--------------------+--------------------+
--------------------+-----------------+-----------+------+---------+----------
----------+--------------------+
|show_id|   type|               title|            director|                cast|
country|        date_added|release_year|rating| duration|           listed_in|
description|
+-------+-------+--------------------+--------------------+--------------------+
--------------------+-----------------+-----------+------+---------+----------
----------+--------------------+
|     s1|  Movie|Dick Johnson Is Dead|     Kirsten Johnson|                null|
United States|September 25, 2021|        2020| PG-13|   90 min|
Documentaries|As her father nea…|
|     s2|TV Show|       Blood & Water|                null|Ama Qamata, Khosi…|
South Africa|September 24, 2021|        2021| TV-MA|2 Seasons|International TV
…|After crossing pa…|
|     s3|TV Show|           Ganglands|     Julien Leclercq|Sami Bouajila, Tr…|
null|September 24, 2021|        2021| TV-MA| 1 Season|Crime TV Shows, I…|To
protect his fa…|
|     s4|TV Show|Jailbirds New Orl…|                null|                null|
null|September 24, 2021|        2021| TV-MA| 1 Season|Docuseries,
Reali…|Feuds, flirtation…|
|     s5|TV Show|        Kota Factory|                null|Mayur More, Jiten…|
India|September 24, 2021|        2021| TV-MA|2 Seasons|International TV …|In a
city of coac…|
|     s6|TV Show|       Midnight Mass|       Mike Flanagan|Kate Siegel, Zach…|
null|September 24, 2021|        2021| TV-MA| 1 Season|TV Dramas, TV Hor…|The
arrival of a …|
|     s7|  Movie|My Little Pony: A…|Robert Cullen, Jo…|Vanessa Hudgens, …|
null|September 24, 2021|        2021|    PG|   91 min|Children &
```

```
Family…|Equestria's divid…|
|     s8|  Movie|             Sankofa|        Haile Gerima|Kofi Ghanaba,
Oya…|United States, Gh…|September 24, 2021|        1993| TV-MA|  125
min|Dramas, Independe…|On a photo shoot …|
|     s9|TV Show|The Great British…|     Andy Devonshire|Mel Giedroyc, Sue…|
United Kingdom|September 24, 2021|        2021| TV-14|9 Seasons|British TV
Shows,…|A talented batch …|
|    s10|  Movie|        The Starling|      Theodore Melfi|Melissa McCarthy,…|
United States|September 24, 2021|        2021| PG-13|  104 min|    Comedies,
Dramas|A woman adjusting…|
|    s11|TV Show|Vendetta: Truth, …|                null|                null|
null|September 24, 2021|        2021| TV-MA| 1 Season|Crime TV Shows,
D…|"Sicily boasts a …|
|    s12|TV Show|    Bangkok Breaking|   Kongkiat Komesiri|Sukollawat Kanaro…|
null|September 23, 2021|        2021| TV-MA| 1 Season|Crime TV Shows,
I…|Struggling to ear…|
|    s13|  Movie|        Je Suis Karl| Christian Schwochow|Luna Wedler,
Jann…|Germany, Czech Re…|September 23, 2021|        2021| TV-MA|  127
min|Dramas, Internati…|After most of her…|
|    s14|  Movie|Confessions of an…|       Bruno Garotti|Klara Castanho, L…|
null|September 22, 2021|        2021| TV-PG|   91 min|Children & Family…|When
the clever b…|
|    s15|TV Show|Crime Stories: In…|                null|                null|
null|September 22, 2021|        2021| TV-MA| 1 Season|British TV
Shows,…|Cameras following…|
|    s16|TV Show|   Dear White People|                null|Logan Browning, B…|
United States|September 22, 2021|        2021| TV-MA|4 Seasons|TV Comedies, TV
D…|"Students of colo…|
|    s17|  Movie|Europe's Most Dan…|Pedro de Echave G…|                null|
null|September 22, 2021|        2020| TV-MA|   67 min|Documentaries,
In…|Declassified docu…|
|    s18|TV Show|     Falsa identidad|                null|Luis Ernesto Fran…|
Mexico|September 22, 2021|        2020| TV-MA|2 Seasons|Crime TV Shows,
S…|Strangers Diego a…|
|    s19|  Movie|           Intrusion|          Adam Salky|Freida Pinto, Log…|
null|September 22, 2021|        2021| TV-14|   94 min|           Thrillers|After
a deadly ho…|
|    s20|TV Show|              Jaguar|                null|Blanca Suárez, Iv…|
null|September 22, 2021|        2021| TV-MA| 1 Season|International TV …|In
the 1960s, a H…|
+-------+-------+--------------------+--------------------+--------------------+
-------------------+-----------------+------------+------+---------+----------
----------+--------------------+
only showing top 20 rows
```

ii.Transform: Exploratory data analysis using spark sql queries

- Unique showId count

- GroupBy type,release_year and count of showId
- Update column duration values as 90 min to 90 and 2 seasons to 2 and others
- groupby type and avg durations

[15]:
```
## distinct count using spark sql queries
unique_id_count = sqlContext.sql("select distinct cnt_id from (select␣
 ↪count(show_id) as cnt_id from netflix_table) netflix_table;")
unique_id_count.show()
```

```
+------+
|cnt_id|
+------+
|  8809|
+------+
```

[16]:
```
unique_id = sqlContext.sql("select distinct count(show_id) as cnt_id from␣
 ↪netflix_table")
unique_id.show()
```

```
+------+
|cnt_id|
+------+
|  8809|
+------+
```

[17]:
```
#distinct count through spark dataframe
netflix_spark_df.distinct().count()
```

[17]: 8809

- GroupBy type,release_year and count of showId using spark dataframe

[18]:
```
netflix_spark_df.groupBy("type").count().show()
```

```
+-------------+-----+
|         type|count|
+-------------+-----+
|         null|    1|
|      TV Show| 2676|
|        Movie| 6131|
|William Wyler|    1|
+-------------+-----+
```

[19]:
```
#GroupBy on multiple columns
netflix_spark_df.groupBy("type","release_year").count().show()
```

```
+-------+--------------+-----+
|   type|  release_year|count|
+-------+--------------+-----+
|  Movie|  June 12, 2021|    1|
|  Movie|          1963|    1|
|TV Show|          1981|    1|
|  Movie|          1971|    5|
|TV Show|          1972|    1|
|TV Show|          1988|    2|
|TV Show|  Nse Ikpe-Etim|    1|
|  Movie|          1956|    2|
|  Movie| Charles Rocket|    1|
|  Movie|          1997|   33|
|  Movie|          2015|  397|
|  Movie|          1969|    2|
|  Movie|          2010|  153|
|  Movie|          1993|   24|
|  Movie|          1977|    6|
|TV Show|          2020|  436|
|TV Show|          1997|    4|
|  Movie|          2016|  657|
|  Movie|          1992|   20|
|TV Show|          1945|    1|
+-------+--------------+-----+
only showing top 20 rows
```

- GroupBy type,release_year and count of showId using spark sql queries

[20]:
```
grp_data= sqlContext.sql("select type, release_year, count(show_id) as cnt_id␣
 ↪from netflix_table GROUP BY type, release_year")
grp_data.show()
```

```
+-------+--------------+------+
|   type|  release_year|cnt_id|
+-------+--------------+------+
|  Movie|  June 12, 2021|     1|
|  Movie|          1963|     1|
|TV Show|          1981|     1|
|  Movie|          1971|     5|
|TV Show|          1972|     1|
|TV Show|          1988|     2|
|TV Show|  Nse Ikpe-Etim|     1|
|  Movie|          1956|     2|
|  Movie| Charles Rocket|     1|
|  Movie|          1997|    33|
|  Movie|          2015|   397|
|  Movie|          1969|     2|
```

```
|  Movie|           2010|   153|
|  Movie|           1993|    24|
|  Movie|           1977|     6|
|TV Show|           2020|   436|
|TV Show|           1997|     4|
|  Movie|           2016|   657|
|  Movie|           1992|    20|
|TV Show|           1945|     1|
+-------+---------------+------+
only showing top 20 rows
```

- Update column duration values as 90 min to 90 and 2 seasons to 2 and others ref link

[21]:
```
# split() function defining parameters
split_cols = pyspark.sql.functions.split(netflix_spark_df['duration'], ' ')
```

[22]:
```
netflix_spark_df = netflix_spark_df.withColumn('upd_duration', split_cols.
↪getItem(0))
```

[23]:
```
netflix_spark_df.show()
```

```
+-------+-------+--------------------+--------------------+--------------------+
--------------------+-----------------+------------+------+---------+----------
----------+--------------------+------------+
|show_id|   type|               title|            director|                cast|
country|        date_added|release_year|rating| duration|           listed_in|
description|upd_duration|
+-------+-------+--------------------+--------------------+--------------------+
--------------------+-----------------+------------+------+---------+----------
----------+--------------------+------------+
|     s1|  Movie|Dick Johnson Is Dead|     Kirsten Johnson|                null|
United States|September 25, 2021|        2020| PG-13|   90 min|
Documentaries|As her father nea…|          90|
|     s2|TV Show|       Blood & Water|                null|Ama Qamata, Khosi…|
South Africa|September 24, 2021|        2021| TV-MA|2 Seasons|International TV
…|After crossing pa…|           2|
|     s3|TV Show|           Ganglands|     Julien Leclercq|Sami Bouajila, Tr…|
null|September 24, 2021|        2021| TV-MA| 1 Season|Crime TV Shows, I…|To
protect his fa…|           1|
|     s4|TV Show|Jailbirds New Orl…|                null|                null|
null|September 24, 2021|        2021| TV-MA| 1 Season|Docuseries,
Reali…|Feuds, flirtation…|           1|
|     s5|TV Show|        Kota Factory|                null|Mayur More, Jiten…|
India|September 24, 2021|        2021| TV-MA|2 Seasons|International TV …|In a
city of coac…|           2|
|     s6|TV Show|       Midnight Mass|       Mike Flanagan|Kate Siegel, Zach…|
```

```
null|September 24, 2021|        2021| TV-MA| 1 Season|TV Dramas, TV Hor…|The
arrival of a …|           1|
|     s7|  Movie|My Little Pony: A…|Robert Cullen, Jo…|Vanessa Hudgens, …|
null|September 24, 2021|        2021|    PG|   91 min|Children &
Family…|Equestria's divid…|          91|
|     s8|  Movie|             Sankofa|        Haile Gerima|Kofi Ghanaba,
Oya…|United States, Gh…|September 24, 2021|        1993| TV-MA|  125
min|Dramas, Independe…|On a photo shoot …|         125|
|     s9|TV Show|The Great British…|     Andy Devonshire|Mel Giedroyc, Sue…|
United Kingdom|September 24, 2021|        2021| TV-14|9 Seasons|British TV
Shows,…|A talented batch …|           9|
|    s10|  Movie|        The Starling|      Theodore Melfi|Melissa McCarthy,…|
United States|September 24, 2021|        2021| PG-13|  104 min|    Comedies,
Dramas|A woman adjusting…|         104|
|    s11|TV Show|Vendetta: Truth, …|                null|                null|
null|September 24, 2021|        2021| TV-MA| 1 Season|Crime TV Shows,
D…|"Sicily boasts a …|           1|
|    s12|TV Show|    Bangkok Breaking|   Kongkiat Komesiri|Sukollawat Kanaro…|
null|September 23, 2021|        2021| TV-MA| 1 Season|Crime TV Shows,
I…|Struggling to ear…|           1|
|    s13|  Movie|        Je Suis Karl| Christian Schwochow|Luna Wedler,
Jann…|Germany, Czech Re…|September 23, 2021|        2021| TV-MA|  127
min|Dramas, Internati…|After most of her…|         127|
|    s14|  Movie|Confessions of an…|       Bruno Garotti|Klara Castanho, L…|
null|September 22, 2021|        2021| TV-PG|   91 min|Children & Family…|When
the clever b…|          91|
|    s15|TV Show|Crime Stories: In…|                null|                null|
null|September 22, 2021|        2021| TV-MA| 1 Season|British TV
Shows,…|Cameras following…|           1|
|    s16|TV Show|   Dear White People|                null|Logan Browning, B…|
United States|September 22, 2021|        2021| TV-MA|4 Seasons|TV Comedies, TV
D…|"Students of colo…|           4|
|    s17|  Movie|Europe's Most Dan…|Pedro de Echave G…|                null|
null|September 22, 2021|        2020| TV-MA|   67 min|Documentaries,
In…|Declassified docu…|          67|
|    s18|TV Show|     Falsa identidad|                null|Luis Ernesto Fran…|
Mexico|September 22, 2021|        2020| TV-MA|2 Seasons|Crime TV Shows,
S…|Strangers Diego a…|           2|
|    s19|  Movie|           Intrusion|          Adam Salky|Freida Pinto, Log…|
null|September 22, 2021|        2021| TV-14|   94 min|           Thrillers|After
a deadly ho…|          94|
|    s20|TV Show|              Jaguar|                null|Blanca Suárez, Iv…|
null|September 22, 2021|        2021| TV-MA| 1 Season|International TV …|In
the 1960s, a H…|           1|
+-------+-------+--------------------+--------------------+--------------------+
-------------------+-----------------+------------+------+---------+----------
----------+--------------------+------------+
only showing top 20 rows
```

[31]:
```
# show upd_duration column values
netflix_spark_df.select("upd_duration")
```

[31]:
```
DataFrame[upd_duration: int]
```

[32]:
```
netflix_spark_df.select("upd_duration").show()
```

```
+------------+
|upd_duration|
+------------+
|          90|
|           2|
|           1|
|           1|
|           2|
|           1|
|          91|
|         125|
|           9|
|         104|
|           1|
|           1|
|         127|
|          91|
|           1|
|           4|
|          67|
|           2|
|          94|
|           1|
+------------+
only showing top 20 rows
```

[33]:
```
# update_dur = sqlContext.sql("select dbo.GetNumericValue(duration) AS␣
 ↪updated_duration from netflix_table")
# update_dur.show()
```

- groupby type and avg durations ref link

[27]:
```
netflix_spark_df = netflix_spark_df.withColumn('upd_duration',␣
 ↪col('upd_duration').cast(IntegerType()))
```

[34]:
```
#GroupBy on column
netflix_spark_df.groupBy("type").avg('upd_duration').show()
```

```
+-------------+------------------+
|         type| avg(upd_duration)|
+-------------+------------------+
|         null|              null|
|      TV Show|1.7654320987654322|
|        Movie| 99.88907068062828|
|William Wyler|              null|
+-------------+------------------+
```

iii. Load: Save analysis report

- save as tables - partitionby type ref link

[35]:
```
# save spark dataframe as table
netflix_spark_df.write.partitionBy('type').saveAsTable('netflix_trans_table')
```

[30]:
```
#save transformed spark datframe as partitionBy() column 'test'
netflix_spark_df.write.option("header",True) \
        .partitionBy("type") \
        .mode("overwrite") \
        .csv("./output")
```