Bonus Assignment

Spark config is not provided because Databricks community edition doesn't have option. But for demonstration purpose it is given below:

import pyspark from pyspark.sql import SparkSession

EXE_MEMORY="2G" DRIVER_MEMORY="8G" spark = SparkSession.builder.appName("AWS").config("spark.executor.memory", EXE_MEMORY).config("spark.executor.cores", "2").config("spark.driver.memory", DRIVER_MEMORY).getOrCreate()

In [1]:

```
from pyspark.sql import *
```

In [2]:

```
import os
os.getcwd()
```

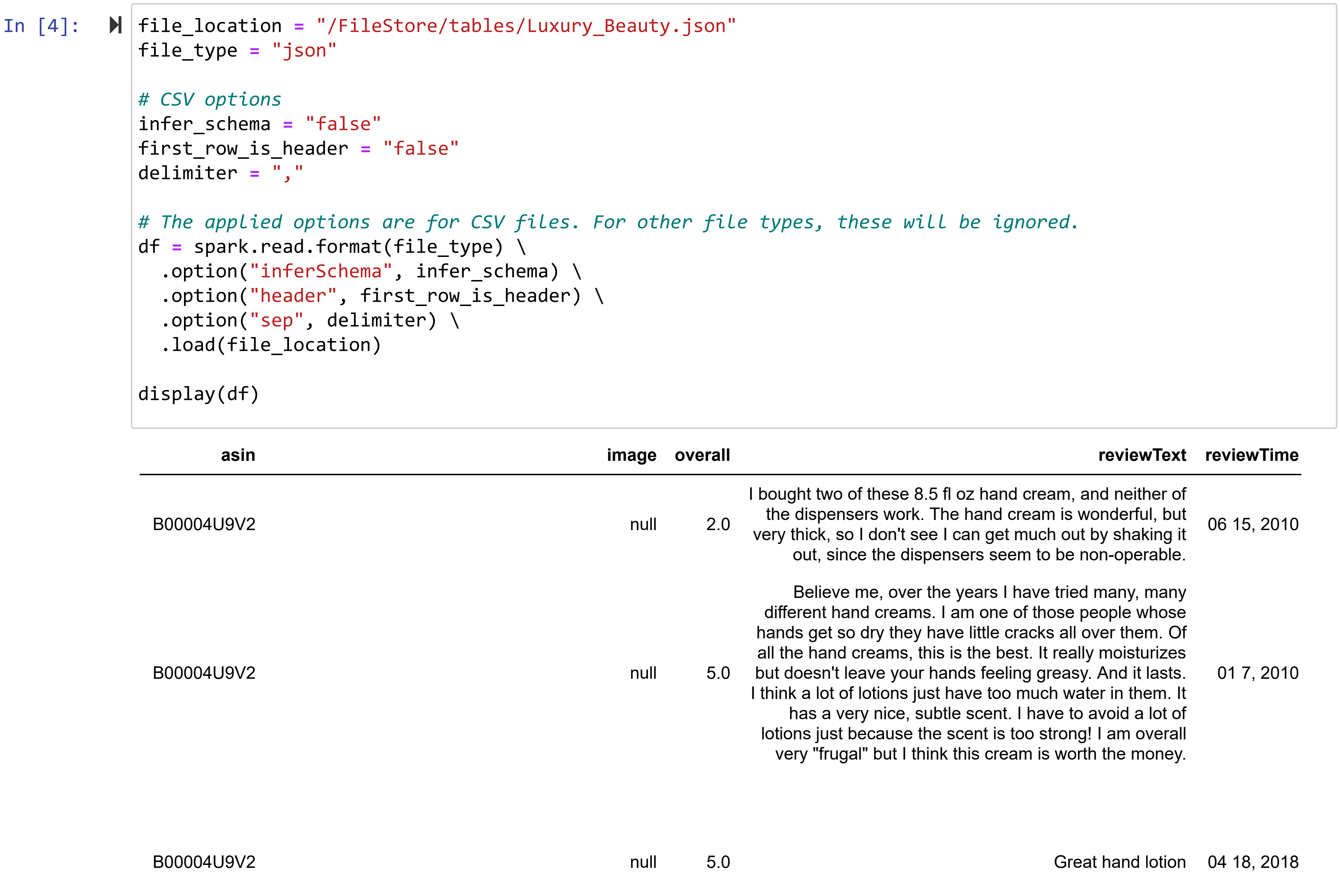
```
Out[2]: '/databricks/driver'
```

In [3]:

```
%fs ls
```

path	name	size
dbfs:/FileStore/	FileStore/	0
dbfs:/databricks-datasets/	databricks-datasets/	0
dbfs:/databricks-results/	databricks-results/	0
dbfs:/ml/	ml/	0
dbfs:/tmp/	tmp/	0
dbfs:/user/	user/	0

Dataframe of Amazon Dataset i.e., Luxury_Beauty in json format.



In [4]:

```
file_location = "/FileStore/tables/Luxury_Beauty.json"
file_type = "json"

# CSV options
infer_schema = "false"
first_row_is_header = "false"
delimiter = ","

# The applied options are for CSV files. For other file types, these will be ignored.
df = spark.read.format(file_type) \
  .option("inferSchema", infer_schema) \
  .option("header", first_row_is_header) \
  .option("sep", delimiter) \
  .load(file_location)

display(df)
```

asin	image	overall	reviewText	reviewTime
B00004U9V2	null	2.0	I bought two of these 8.5 fl oz hand cream, and neither of the dispensers work. The hand cream is wonderful, but very thick, so I don't see I can get much out by shaking it out, since the dispensers seem to be non-operable.	06 15, 2010
B00004U9V2	null	5.0	Believe me, over the years I have tried many, many different hand creams. I am one of those people whose hands get so dry they have little cracks all over them. Of all the hand creams, this is the best. It really moisturizes but doesn't leave your hands feeling greasy. And it lasts. I think a lot of lotions just have too much water in them. It has a very nice, subtle scent. I have to avoid a lot of lotions just because the scent is too strong! I am overall very "frugal" but I think this cream is worth the money.	01 7, 2010
B00004U9V2	null	5.0	Great hand lotion	04 18, 2018

Dataframe of Amazon Dataset i.e., Magazine_Subscriptions in json format.

In [5]:

```
file_location = "/FileStore/tables/Magazine_Subscriptions.json"
file_type = "json"

# CSV options
infer_schema = "false"
first_row_is_header = "false"
delimiter = ","

# The applied options are for CSV files. For other file types, these will be ignored.
df1 = spark.read.format(file_type) \
  .option("inferSchema", infer_schema) \
  .option("header", first_row_is_header) \
  .option("sep", delimiter) \
  .load(file_location)

display(df1)
```

asin	image	overall	reviewText	re
B00005N7P0	null	5.0	for computer enthusiast, MaxPC is a welcome sight in your mailbox. i can remember for years savorying every page of "boot" (as it was called in beginning) as i was (and still am) obcessed with PC's. Anyone, from advanced users - to beginners looking for knowledge - can profit from every issue of MaxPC. the icing on the cake is the subscription that comes with a CD-ROM as it is packed with demos, utilities, and other useful apps (very helpful for those not blessed with broadband connections). Until I discovered the community of hardware enthusiast web sites, MaxPC, formerly "boot", was my only really informative source for computing news and articles. To this day, i consider my subscription to it worth more than 10 subscriptions to most other computing mags. I can't wait until they merge with DVD media and maybe end up offering more info on Divx codecs, encoding your own movies, and best bang for the buck audio and video equipment. Try a few issues (with CD)and you may get hooked...	
B00005N7P0	null	5.0	Thank god this is not a Ziff Davis publication. MaxPC will actually tell you if a product is bad. They will print just what they think about something; no sugar coating. I would compare their style to Car and Driver. Technical, but they know how to have a good time.	1

Type of Data

In [6]:

```
type(df)
```

```
Out[4]: pyspark.sql.dataframe.DataFrame
```

Top 20 values of Dataframe df

In [7]:

```
df.show(20)
```

```
+----------+-----+-------+--------------------+-----------+--------------+-----------------+--------------------+--------------------+--------------+--------+----+
asin|image|overall| reviewText| reviewTime| reviewerID| reviewerName| style| summary|unixReviewTime|verified|vote| +----------+-----
+-------+--------------------+-----------+--------------+-----------------+--------------------+--------------------+--------------+--------+----+ B00004U9V2|
null| 2.0|I bought two of t...|06 15, 2010|A1Q6MUU0B2ZDQG| D. Poston| null|dispensers don't ...| 1276560000| true| 3| B00004U9V2|
null| 5.0|Believe me, over ...| 01 7, 2010|A3HO2SQDCZIE9S| chandra| null|Best hand cream e...| 1262822400| true| 14|
B00004U9V2| null| 5.0| Great hand lotion|04 18, 2018|A2EM03F99X3RJZ| Maureen G|[,,,,,,,,,, 3.5...| Five Stars| 1524009600|
true|null| B00004U9V2| null| 5.0|This is the best ...|04 18, 2018| A3Z74TDRGD0HU| Terry K|[,,,,,,,,,, 3.5...| Five Stars| 1524009600|
true|null| B00004U9V2| null| 5.0|The best non- oil...|04 17, 2018|A2UXFNW9RTL4VM| Patricia Wood|[,,,,,,,,,, 3.5...|I always have a
b...| 1523923200| true|null| B00004U9V2| null| 5.0|Ive used this lot...|04 14, 2018| AXX5G4LFF12R6| Ralla|[,,,,,,,,,, 250...|Ive used
this lot...| 1523664000| true|null| B00004U9V2| null| 5.0|Works great for d...|04 11, 2018| A7GUKMOJT2NR6| Lydia Speight|[,,,,,,,,,,
3.5...| Five Stars| 1523404800| true|null| B00004U9V2| null| 5.0|The best hand cre...|04 11, 2018|A3FU4L59BHA9FY| Allen Semer|
[,,,,,,,,,, 3.5...| Made in the USA| 1523404800| true|null| B00004U9V2| null| 5.0|LOVE THIS SCENT!!...| 04 7,
2018|A1AMNMIPQMXH9M| Vets park|[,,,,,,,,,, 3.5...|Moistens and smel...| 1523059200| true|null| B00004U9V2| null| 5.0|Its a great
moist...| 04 6, 2018|A3DMBDTA8VGWSX| Cynthia P. Irving|[,,,,,,,,,, 3.5...| Five Stars| 1522972800| true|null| B00004U9V2| null|
5.0|This hand cream i...| 04 5, 2018|A160DTI3H7VHLQ| CB|[,,,,,,,,,, 0.9...| Five Stars| 1522886400| true|null| B00004U9V2| null| 5.0|I
am a healthcare...| 04 5, 2018|A1H41DKPDPVA0R| Donna Butler RN|[,,,,,,,,,, 250...|Best hand therapy...| 1522886400| true|null|
B00004U9V2| null| 5.0|have used on and ...| 04 5, 2018| A2BDI7THUMJ8V| Teresa K. L.|[,,,,,,,,,, 250...|Product is good f...|
1522886400| true|null| B00004U9V2| null| 5.0| Great hand cream| 04 3, 2018| AM7EBP5TRX7AC|Glenn B. Guilbault|[,,,,,,,,,, 250...|
Five Stars| 1522713600| true|null| B00004U9V2| null| 5.0|This is my favori...| 04 2, 2018|A31FOVCS3WTWPT| Pam|[,,,,,,,,,, 3.5...|My
Favorite Lotio...| 1522627200| false|null| B00004U9V2| null| 4.0|Soothing! Love th...|03 30, 2018| AXUU8F9EM6U3E| Amazon
Customer|[,,,,,,,,,, 0.9...| Love the way it| 1522368000| true|null| B00004U9V2| null| 5.0|My wife loves the...|03 29,
2018|A24B46V78ATNRP| Michael Konrad|[,,,,,,,,,, 250...| Wonderful| 1522281600| true|null| B00004U9V2| null| 5.0|I always loved
th...|03 27, 2018| ABUBKML2EONCG| Lotte Hersey|[,,,,,,,,,, 3.5...| Five Stars| 1522108800| true|null| B00004U9V2| null|
5.0|Absolutely great....|03 26, 2018|A2UA6E1RVG3C1I| Ginny|[,,,,,,,,,, 0.9...| Absolutely great!| 1522022400| true|null| B00004U9V2|
null| 1.0|SOOOO not worth t...|03 23, 2018|A1TRMJHEDGX0HF| soulsurfer|[,,,,,,,,,, 0.9...| Disappointed.| 1521763200| true|null| +----
------+-----+-------+--------------------+-----------+--------------+-----------------+--------------------+--------------------+--------------+--------+----+ only
showing top 20 rows
```

Schema of Dataframe df

In [8]:

```
df.printSchema()
```

```
root -- asin: string (nullable = true) -- image: array (nullable = true) |-- element: string (containsNull = true) -- overall: double (nullable
= true) -- reviewText: string (nullable = true) -- reviewTime: string (nullable = true) -- reviewerID: string (nullable = true) --
reviewerName: string (nullable = true) -- style: struct (nullable = true) |-- Color:: string (nullable = true) |-- Design:: string (nullable =
true) |-- Flavor Name:: string (nullable = true) |-- Flavor:: string (nullable = true) |-- Format:: string (nullable = true) |-- Item Package
Quantity:: string (nullable = true) |-- Package Quantity:: string (nullable = true) |-- Package Type:: string (nullable = true) |-- SCENT::
string (nullable = true) |-- Scent Name:: string (nullable = true) |-- Scent:: string (nullable = true) |-- Size:: string (nullable = true) |--
Style Name:: string (nullable = true) |-- Style:: string (nullable = true) -- summary: string (nullable = true) -- unixReviewTime: long
(nullable = true) -- verified: boolean (nullable = true) -- vote: string (nullable = true)
```

RDD made of same data

In [9]:

```
rdd=sc.textFile("/FileStore/tables/Luxury_Beauty.json")
```

RDD count is 574628 of Luxury_Beauty data.

In [10]:

```
rdd.count()
```

```
Out[10]: 574628
```

In [11]:

```
df3 = spark.createDataFrame(rdd).toDF( "asin","image", "overall", "reviewText","reviewTime", "reviewerID","r
```

```
--------------------------------------------------------------------------- TypeError Traceback (most recent call last) <command-
4379487006160354> in <module> ----> 1 df3 = spark.createDataFrame(rdd).toDF( "asin","image", "overall",
"reviewText","reviewTime", "reviewerID","reviewerName","style", "summary", "unixReviewTime","verified", "vote")
/databricks/spark/python/pyspark/sql/session.py in createDataFrame(self, data, schema, samplingRatio, verifySchema) 813 else: 814
if isinstance(data, RDD): --> 815 rdd, schema = self._createFromRDD(data.map(prepare), schema, samplingRatio) 816 else: 817
rdd, schema = self._createFromLocal(map(prepare, data), schema) /databricks/spark/python/pyspark/sql/session.py in
_createFromRDD(self, rdd, schema, samplingRatio) 397 """ 398 if schema is None or isinstance(schema, (list, tuple)): --> 399 struct
= self._inferSchema(rdd, samplingRatio, names=schema) 400 converter = _create_converter(struct) 401 rdd = rdd.map(converter)
/databricks/spark/python/pyspark/sql/session.py in _inferSchema(self, rdd, samplingRatio, names) 377 378 if samplingRatio is None:
--> 379 schema = _infer_schema(first, names=names) 380 if _has_nulltype(schema): 381 for row in rdd.take(100)[1:]:
/databricks/spark/python/pyspark/sql/types.py in _infer_schema(row, names) 1060 1061 else: -> 1062 raise TypeError("Can not infer
schema for type: %s" % type(row)) 1063 1064 fields = [StructField(k, _infer_type(v), True) for k, v in items] TypeError: Can not infer
schema for type: <class 'str'>
```

Dataframe df3 from rdd

In [12]:

```
df3=rdd.map(lambda x: (x, )).toDF()
```

In [13]:

```
df3.show(3,truncate=False)
```

```
+----------------------------------------------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------+ _1 | +---------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------+ {"overall": 2.0, "vote": "3",
"verified": true, "reviewTime": "06 15, 2010", "reviewerID": "A1Q6MUU0B2ZDQG", "asin": "B00004U9V2", "reviewerName": "D.
Poston", "reviewText": "I bought two of these 8.5 fl oz hand cream, and neither of the dispensers work. The hand cream is wonderful,
but very thick, so I don't see I can get much out by shaking it out, since the dispensers seem to be non-operable.", "summary":
"dispensers don't work", "unixReviewTime": 1276560000} | {"overall": 5.0, "vote": "14", "verified": true, "reviewTime": "01 7, 2010",
"reviewerID": "A3HO2SQDCZIE9S", "asin": "B00004U9V2", "reviewerName": "chandra", "reviewText": "Believe me, over the years I
have tried many, many different hand creams. I am one of those people whose hands get so dry they have little cracks all over
them.\n\nOf all the hand creams, this is the best. It really moisturizes but doesn't leave your hands feeling greasy. And it lasts. I think
a lot of lotions just have too much water in them. It has a very nice, subtle scent. I have to avoid a lot of lotions just because the scent
is too strong!\n\nI am overall very \"frugal\" but I think this cream is worth the money.", "summary": "Best hand cream ever.",
"unixReviewTime": 1262822400}| {"overall": 5.0, "verified": true, "reviewTime": "04 18, 2018", "reviewerID": "A2EM03F99X3RJZ",
"asin": "B00004U9V2", "style": {"Size:": " 3.5 oz."}, "reviewerName": "Maureen G", "reviewText": "Great hand lotion", "summary": "Five
Stars", "unixReviewTime": 1524009600} | +------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------+ only showing top 3 rows
```

Again rdd from dataframe

In [14]:

```
rdd2=df3.rdd.map(list)
```

In [15]:

```
rdd2.take(3)
```

```
Out[67]: [['{"overall": 2.0, "vote": "3", "verified": true, "reviewTime": "06 15, 2010", "reviewerID": "A1Q6MUU0B2ZDQG", "asin":
"B00004U9V2", "reviewerName": "D. Poston", "reviewText": "I bought two of these 8.5 fl oz hand cream, and neither of the
dispensers work. The hand cream is wonderful, but very thick, so I don\'t see I can get much out by shaking it out, since the
dispensers seem to be non-operable.", "summary": "dispensers don\'t work", "unixReviewTime": 1276560000}'], ['{"overall": 5.0,
"vote": "14", "verified": true, "reviewTime": "01 7, 2010", "reviewerID": "A3HO2SQDCZIE9S", "asin": "B00004U9V2", "reviewerName":
"chandra", "reviewText": "Believe me, over the years I have tried many, many different hand creams. I am one of those people whose
hands get so dry they have little cracks all over them.\\n\\nOf all the hand creams, this is the best. It really moisturizes but doesn\'t
leave your hands feeling greasy. And it lasts. I think a lot of lotions just have too much water in them. It has a very nice, subtle scent.
I have to avoid a lot of lotions just because the scent is too strong!\\n\\nI am overall very \\"frugal\\" but I think this cream is worth the
money.", "summary": "Best hand cream ever.", "unixReviewTime": 1262822400}'], ['{"overall": 5.0, "verified": true, "reviewTime": "04
18, 2018", "reviewerID": "A2EM03F99X3RJZ", "asin": "B00004U9V2", "style": {"Size:": " 3.5 oz."}, "reviewerName": "Maureen G",
"reviewText": "Great hand lotion", "summary": "Five Stars", "unixReviewTime": 1524009600}']]
```

Pandas Dataframe from Pyspark Dataframe

In [16]:

```
panddf=df.toPandas()
```

In [17]:

```
panddf.head()
```

	asin	image	overall	reviewText	reviewTime	reviewerID	reviewerName	style	summary	unixReviewTime	ve
0	B00004U9V2	None	2.0	I bought two of these 8.5 fl oz hand cream, an...	06 15, 2010	A1Q6MUU0B2ZDQG	D. Poston	None	dispensers don't work	1276560000	
1	B00004U9V2	None	5.0	Believe me, over the years I have tried many, ...	01 7, 2010	A3HO2SQDCZIE9S	chandra	None	Best hand cream ever.	1262822400	
2	B00004U9V2	None	5.0	Great hand lotion	04 18, 2018	A2EM03F99X3RJZ	Maureen G	{'Color:': None, 'Design:': None, 'Flavor Name...	Five Stars	1524009600	
3	B00004U9V2	None	5.0	This is the best for the severely dry skin on ...	04 18, 2018	A3Z74TDRGD0HU	Terry K	{'Color:': None, 'Design:': None, 'Flavor Name...	Five Stars	1524009600	
4	B00004U9V2	None	5.0	The best non- oily hand cream ever. It heals o...	04 17, 2018	A2UXFNW9RTL4VM	Patricia Wood	{'Color:': None, 'Design:': None, 'Flavor Name...	I always have a backup ready.	1523923200	

In [18]:

```
%python
df.createOrReplaceTempView("df")
```

Reading top 10 rows from Pyspark Dataframe initiating SQL library

In [19]:

```
from pyspark.sql.types import *
df.take(10)
```

```
Out[94]: [Row(asin='B00004U9V2', image=None, overall=2.0, reviewText="I bought two of these 8.5 fl oz hand cream, and neither of
the dispensers work. The hand cream is wonderful, but very thick, so I don't see I can get much out by shaking it out, since the
dispensers seem to be non-operable.", reviewTime='06 15, 2010', reviewerID='A1Q6MUU0B2ZDQG', reviewerName='D. Poston',
style=None, summary="dispensers don't work", unixReviewTime=1276560000, verified=True, vote='3'), Row(asin='B00004U9V2',
image=None, overall=5.0, reviewText='Believe me, over the years I have tried many, many different hand creams. I am one of those
people whose hands get so dry they have little cracks all over them.\n\nOf all the hand creams, this is the best. It really moisturizes
but doesn\'t leave your hands feeling greasy. And it lasts. I think a lot of lotions just have too much water in them. It has a very nice,
subtle scent. I have to avoid a lot of lotions just because the scent is too strong!\n\nI am overall very "frugal" but I think this cream is
worth the money.', reviewTime='01 7, 2010', reviewerID='A3HO2SQDCZIE9S', reviewerName='chandra', style=None,
summary='Best hand cream ever.', unixReviewTime=1262822400, verified=True, vote='14'), Row(asin='B00004U9V2', image=None,
overall=5.0, reviewText='Great hand lotion', reviewTime='04 18, 2018', reviewerID='A2EM03F99X3RJZ', reviewerName='Maureen G',
style=Row(Color:=None, Design:=None, Flavor Name:=None, Flavor:=None, Format:=None, Item Package Quantity:=None,
Package Quantity:=None, Package Type:=None, SCENT:=None, Scent Name:=None, Scent:=None, Size:=' 3.5 oz.', Style
Name:=None, Style:=None), summary='Five Stars', unixReviewTime=1524009600, verified=True, vote=None),
Row(asin='B00004U9V2', image=None, overall=5.0, reviewText='This is the best for the severely dry skin on my hands',
reviewTime='04 18, 2018', reviewerID='A3Z74TDRGD0HU', reviewerName='Terry K', style=Row(Color:=None, Design:=None, Flavor
Name:=None, Flavor:=None, Format:=None, Item Package Quantity:=None, Package Quantity:=None, Package Type:=None,
SCENT:=None, Scent Name:=None, Scent:=None, Size:=' 3.5 oz.', Style Name:=None, Style:=None), summary='Five Stars',
unixReviewTime=1524009600, verified=True, vote=None), Row(asin='B00004U9V2', image=None, overall=5.0, reviewText='The best
non- oily hand cream ever. It heals overnight.', reviewTime='04 17, 2018', reviewerID='A2UXFNW9RTL4VM', reviewerName='Patricia
Wood', style=Row(Color:=None, Design:=None, Flavor Name:=None, Flavor:=None, Format:=None, Item Package Quantity:=None,
Package Quantity:=None, Package Type:=None, SCENT:=None, Scent Name:=None, Scent:=None, Size:=' 3.5 oz.', Style
Name:=None, Style:=None), summary='I always have a backup ready.', unixReviewTime=1523923200, verified=True, vote=None),
Row(asin='B00004U9V2', image=None, overall=5.0, reviewText="Ive used this lotion for many years. I try others occasionally and
always come back to Gardners. Please don't change a thing.", reviewTime='04 14, 2018', reviewerID='AXX5G4LFF12R6',
reviewerName='Ralla', style=Row(Color:=None, Design:=None, Flavor Name:=None, Flavor:=None, Format:=None, Item Package
Quantity:=None, Package Quantity:=None, Package Type:=None, SCENT:=None, Scent Name:=None, Scent:=None, Size:=' 250 g',
Style Name:=None, Style:=None), summary='Ive used this lotion for many years. I try ...', unixReviewTime=1523664000,
verified=True, vote=None), Row(asin='B00004U9V2', image=None, overall=5.0, reviewText='Works great for dry hands.',
reviewTime='04 11, 2018', reviewerID='A7GUKMOJT2NR6', reviewerName='Lydia Speight', style=Row(Color:=None, Design:=None,
Flavor Name:=None, Flavor:=None, Format:=None, Item Package Quantity:=None, Package Quantity:=None, Package Type:=None,
SCENT:=None, Scent Name:=None, Scent:=None, Size:=' 3.5 oz.', Style Name:=None, Style:=None), summary='Five Stars',
unixReviewTime=1523404800, verified=True, vote=None), Row(asin='B00004U9V2', image=None, overall=5.0, reviewText='The best
hand cream ever.', reviewTime='04 11, 2018', reviewerID='A3FU4L59BHA9FY', reviewerName='Allen Semer',
style=Row(Color:=None, Design:=None, Flavor Name:=None, Flavor:=None, Format:=None, Item Package Quantity:=None,
Package Quantity:=None, Package Type:=None, SCENT:=None, Scent Name:=None, Scent:=None, Size:=' 3.5 oz.', Style
Name:=None, Style:=None), summary='Made in the USA', unixReviewTime=1523404800, verified=True, vote=None),
Row(asin='B00004U9V2', image=None, overall=5.0, reviewText='LOVE THIS SCENT!! But Crabtree and Evelyn make so many.
Washes off easily too!!', reviewTime='04 7, 2018', reviewerID='A1AMNMIPQMXH9M', reviewerName='Vets park',
style=Row(Color:=None, Design:=None, Flavor Name:=None, Flavor:=None, Format:=None, Item Package Quantity:=None,
Package Quantity:=None, Package Type:=None, SCENT:=None, Scent Name:=None, Scent:=None, Size:=' 3.5 oz.', Style
Name:=None, Style:=None), summary='Moistens and smells good', unixReviewTime=1523059200, verified=True, vote=None),
Row(asin='B00004U9V2', image=None, overall=5.0, reviewText='Its a great moisturizer especially for gardners', reviewTime='04 6,
2018', reviewerID='A3DMBDTA8VGWSX', reviewerName='Cynthia P. Irving', style=Row(Color:=None, Design:=None, Flavor
Name:=None, Flavor:=None, Format:=None, Item Package Quantity:=None, Package Quantity:=None, Package Type:=None,
SCENT:=None, Scent Name:=None, Scent:=None, Size:=' 3.5 oz.', Style Name:=None, Style:=None), summary='Five Stars',
unixReviewTime=1522972800, verified=True, vote=None)]
```

Overall counts by using sql groupby function

In [20]:

```
dfs=df.groupby("overall").count().show()
```

```
+-------+------+ overall| count| +-------+------+ 1.0| 50501| 4.0| 70481| 3.0| 41988| 2.0| 29506| 5.0|382152| +-------+------+
```

Minimum of overall variable

In [21]:

```
from pyspark.sql import functions as F
df.agg(F.min(df.overall)).collect()
```

```
Out[103]: [Row(min(overall)=1.0)]
```

Maximum of overall variable

```
In [22]:  from pyspark.sql import functions as F
          df.agg(F.max(df.overall)).collect()
```

```
Out[104]: [Row(max(overall)=5.0)]
```

Where function used in Dataframe with asin variable i.e., B00004U9V2

```
In [23]:  df.where(df.asin == "B00004U9V2").collect()
```

```
Out[107]: [Row(asin='B00004U9V2', image=None, overall=2.0, reviewText="I bought two of these 8.5 fl oz hand cream, and neithe
the dispensers work. The hand cream is wonderful, but very thick, so I don't see I can get much out by shaking it out, since the
dispensers seem to be non-operable.", reviewTime='06 15, 2010', reviewerID='A1Q6MUU0B2ZDQG', reviewerName='D. Poston',
style=None, summary="dispensers don't work", unixReviewTime=1276560000, verified=True, vote='3'), Row(asin='B00004U9V2',
image=None, overall=5.0, reviewText='Believe me, over the years I have tried many, many different hand creams. I am one of thos
people whose hands get so dry they have little cracks all over them.\n\nOf all the hand creams, this is the best. It really moisturizes
but doesn\'t leave your hands feeling greasy. And it lasts. I think a lot of lotions just have too much water in them. It has a very nice
subtle scent. I have to avoid a lot of lotions just because the scent is too strong!\n\nI am overall very "frugal" but I think this cream i
worth the money.', reviewTime='01 7, 2010', reviewerID='A3HO2SQDCZIE9S', reviewerName='chandra', style=None, summary='B
hand cream ever.', unixReviewTime=1262822400, verified=True, vote='14'), Row(asin='B00004U9V2', image=None, overall=5.0,
reviewText='Great hand lotion', reviewTime='04 18, 2018', reviewerID='A2EM03F99X3RJZ', reviewerName='Maureen G',
style=Row(Color:=None, Design:=None, Flavor Name:=None, Flavor:=None, Format:=None, Item Package Quantity:=None, Packa
Quantity:=None, Package Type:=None, SCENT:=None, Scent Name:=None, Scent:=None, Size:=' 3.5 oz.', Style Name:=None,
Style:=None), summary='Five Stars', unixReviewTime=1524009600, verified=True, vote=None), Row(asin='B00004U9V2',
image=None, overall=5.0, reviewText='This is the best for the severely dry skin on my hands', reviewTime='04 18, 2018',
```

Groupby function

```
In [24]:  df.groupBy().avg().collect()
```

```
Out[108]: [Row(avg(overall)=4.22562248968028, avg(unixReviewTime)=1447112948.9492333)]
```

Orderby function used in reviewTime variable in descending order.

```
In [25]:  df.orderBy(df.reviewTime.desc()).collect()
```

```
Out[110]: [Row(asin='B000142FVW', image=None, overall=1.0, reviewText='One star is just for the color. Needs lot of multiple
coating as the polish runs like a water. Disappointed!!!', reviewTime='12 9, 2017', reviewerID='A2C6XNIMDSUKK4',
reviewerName='Jeeva Priya', style=Row(Color:=' Life Gave Me Lemons', Design:=None, Flavor Name:=None, Flavor:=None,
Format:=None, Item Package Quantity:=None, Package Quantity:=None, Package Type:=None, SCENT:=None, Scent
Name:=None, Scent:=None, Size:=None, Style Name:=None, Style:=None), summary='Thin liquid',
unixReviewTime=1512777600, verified=True, vote=None), Row(asin='B000142FVW', image=['https://images-na.ssl-images-
amazon.com/images/I/71dwMJ8GpNL._SY88.jpg', 'https://images-na.ssl-images-
amazon.com/images/I/617sTpJMNEL._SY88.jpg', 'https://images-na.ssl-images-
amazon.com/images/I/71v5ziEZsSL._SY88.jpg'], overall=4.0, reviewText='Came well wrapped! Please excuse the condition of
my nails lol Cant wait to paint my toes!', reviewTime='12 9, 2017', reviewerID='A2NLCS7QPZCUV8', reviewerName='T C',
style=Row(Color:=' Barefoot in Barcelona', Design:=None, Flavor Name:=None, Flavor:=None, Format:=None, Item Package
Quantity:=None, Package Quantity:=None, Package Type:=None, SCENT:=None, Scent Name:=None, Scent:=None,
Size:=None, Style Name:=None, Style:=None), summary='Nice color', unixReviewTime=1512777600, verified=True, vote=None),
Row(asin='B000142FVW', image=None, overall=1.0, reviewText='Been using Tutti Frutti Tonga for years. This is NOT Tutti Frutti
Tonga. What I received was a very bright, brassy pink polish. Buyer Beware!', reviewTime='12 9, 2017',
reviewerID='A2GV9H5E76ELID', reviewerName='angela milana', style=Row(Color:=' Tutti Frutti Tonga', Design:=None, Flavor
Name:=None, Flavor:=None, Format:=None, Item Package Quantity:=None, Package Quantity:=None, Package Type:=None
```

Inner join used in two dataframes to join.

In [26]:

```
inner_join = df.join(df1, df.reviewerID == df1.reviewerID)
inner_join.show()
```

```
+----------+-----+-------+--------------------+-----------+--------------+--------------------+--------------------+--------------------+--------------+--------+-
---+----------+-----+-------+--------------------+-----------+--------------+--------------------+-----------------+--------------------+--------------+--------+-
---+ asin|image|overall| reviewText| reviewTime| reviewerID| reviewerName| style| summary|unixReviewTime|verified|vote|
asin|image|overall| reviewText| reviewTime| reviewerID| reviewerName| style| summary|unixReviewTime|verified|vote| +----------+-
----+-------+--------------------+-----------+--------------+--------------------+--------------------+--------------------+--------------+--------+----+----------
+-----+-------+--------------------+-----------+--------------+--------------------+-----------------+--------------------+--------------+--------+----+
B00J5KDCO2| null| 4.0|For a physical bl...|11 15, 2014|A1TZQTQEMUTJN5| Heather Straub| null|Good for a physic...|
1416009600| true|null|B0000CCY9P| null| 5.0|Ever get tired of...|12 26, 2012|A1TZQTQEMUTJN5| Heather Straub| null|Finally, a
fitnes...| 1356480000| true| 3| B00SWYOBBU| null| 3.0|These are nice be...| 02 9, 2018|A1TZQTQEMUTJN5| Heather Straub|[
Point Dume,,,,,,...|Maybe for someone...| 1518134400| true|null|B0000CCY9P| null| 5.0|Ever get tired of...|12 26,
2012|A1TZQTQEMUTJN5| Heather Straub| null|Finally, a fitnes...| 1356480000| true| 3| B005CVGJFM| null| 5.0|Have used it for
...|10 28, 2017|A207C606CY0AMC| EDDA LAINO|[,,,, Health and...| Revitalash serum| 1509148800| true|null|B00005N7QW|
null| 2.0|It is not what I ...|12 28, 2012|A207C606CY0AMC| EDDA LAINO|[ Print Magazine]| House Beauriful| 1356652800|
true|null| B005CVGJFM| null| 5.0|Have used it for ...|10 28, 2017|A207C606CY0AMC| EDDA LAINO|[,,,, Health and...| Revitalash
serum| 1509148800| true|null|B000IOEJ8W| null| 1.0|I was so disappin...|12 28, 2012|A207C606CY0AMC| EDDA LAINO| null|
TV Guide| 1356652800| true|null| B01EI6N6AC| null| 2.0|I have ordered th...|05 27, 2018| A27GOL7EP1SRX|Rob From Bellport
NY| null|I have ordered th...| 1527379200| true|null|B00015UYBO| null| 1.0|I don't know why ...| 09 4, 2014|
A27GOL7EP1SRX|Rob From Bellport NY|[ Print Magazine]|I don't know why ...| 1409788800| false|null| B000PJ2R0C| null|
5.0|Works really well...|05 26, 2015|A296O33PHOSXUJ| kendel| null| I love it!!| 1432598400| true|null|B00006KTDQ| null| 5.0|I
bought this for...| 10 6, 2014|A296O33PHOSXUJ| kendel| null| Great magazine| 1412553600| true|null| B00OOVKJW0| null| 5.0|I
bought my first...|04 25, 2015|A29BZBM1A50GO8| KWarner|[,,,,,,,,,,, 8.4...|This is the real ...| 1429920000| true| 9|B000063XJP|
null| 5.0|I have subscribed...|05 11, 2014|A29BZBM1A50GO8| KWarner| null| Threads is great.| 1399766400| true| 3|
B00OOVKJW0| null| 5.0|I bought my first...|04 25, 2015|A29BZBM1A50GO8| KWarner|[,,,,,,,,,,, 8.4...|This is the real ...|
1429920000| true| 9|B000IJ7RQ8| null| 5.0|The best place to...| 06 4, 2013|A29BZBM1A50GO8| KWarner| null| I love the
magazine| 1370304000| true|null| B00OOVKJW0| null| 5.0|I bought my first...|04 25, 2015|A29BZBM1A50GO8| KWarner|[,,,,,,,,,,,
8.4...|This is the real ...| 1429920000| true| 9|B0037STB02| null| 5.0|The best place to...| 06 4, 2013|A29BZBM1A50GO8|
KWarner|[ Print Magazine]| I love the magazine| 1370304000| true|null| B001T0HHDS| null| 5.0|My hair is smooth...|01 28,
2018|A2HEI8AD5FLER1| JuJuBunny| null|Leaves hair silky...| 1517097600| true|null|B00005N7SL| null| 4.0|While I still lov...|12
17, 2015|A2HEI8AD5FLER1| JuJuBunny|[ Print Magazine]|While I still lov...| 1450310400| true| 2| B000142FVW| null| 5.0|Great
polish and ...| 05 5, 2016|A31NK65672UL9N| JOYOUS1941|[ Color So Hot It...| OPI| 1462406400| true|null|B00079RO7G| null|
5.0|Great magazine. L...|10 19, 2014|A31NK65672UL9N| JOYOUS1941|[ Print Magazine]| Oorah| 1413676800| false|null|
B0009EXM52| null| 5.0|I always buy OPI....|10 30, 2011|A31NK65672UL9N| JOYOUS1941| null| Nail polish| 1319932800| true|
7|B00079RO7G| null| 5.0|Great magazine. L...|10 19, 2014|A31NK65672UL9N| JOYOUS1941|[ Print Magazine]| Oorah|
1413676800| false|null| B000NGI4QI| null| 5.0|OPI is always my ...|07 16, 2011|A31NK65672UL9N| JOYOUS1941|[,,,,,,,,,, Orig...|
OPI remover| 1310774400| true| 5|B00079RO7G| null| 5.0|Great magazine. L...|10 19, 2014|A31NK65672UL9N| JOYOUS1941|[
Print Magazine]| Oorah| 1413676800| false|null| B000NG80GM| null| 5.0|Great polish and ...| 05 5, 2016|A31NK65672UL9N|
JOYOUS1941|[ Color So Hot It...| OPI| 1462406400| true|null|B00079RO7G| null| 5.0|Great magazine. L...|10 19,
2014|A31NK65672UL9N| JOYOUS1941|[ Print Magazine]| Oorah| 1413676800| false|null| B000BX1Z00| null| 4.0|Initially, this
p...| 03 5, 2011|A3DKP8M0GSP8UK| SAM| null| Smooths my hair.| 1299283200| true| 5|B00077B7M6| null| 5.0|The Economist is
...|03 26, 2010|A3DKP8M0GSP8UK| SAM|[ Kindle Edition]|A Great Global Fi...| 1269561600| false| 2| B002CML1XE| null|
5.0|Derm gave me seve...|10 10, 2017|A3DKP8M0GSP8UK| SAM| null| Nice- get it.| 1507593600| true|null|B00077B7M6| null|
5.0|The Economist is ...|03 26, 2010|A3DKP8M0GSP8UK| SAM|[ Kindle Edition]|A Great Global Fi...| 1269561600| false| 2|
B004QA77JW| null| 5.0|Tastes amazing. M...| 07 4, 2017|A3DKP8M0GSP8UK| SAM|[,,,,,,,,,, 3.8...| Love it.| 1499126400|
true|null|B00077B7M6| null| 5.0|The Economist is ...|03 26, 2010|A3DKP8M0GSP8UK| SAM|[ Kindle Edition]|A Great Global Fi...|
1269561600| false| 2| B00512SFDQ| null| 5.0|amazing. I receiv...|09 11, 2016|A3DKP8M0GSP8UK| SAM| null|Amazing, great,
f...| 1473552000| false|null|B00077B7M6| null| 5.0|The Economist is ...|03 26, 2010|A3DKP8M0GSP8UK| SAM|[ Kindle Edition]|A
Great Global Fi...| 1269561600| false| 2| B00535OW6A| null| 4.0|This brush head i...|05 21, 2013|A3DKP8M0GSP8UK| SAM|
[,,,,,,,,,, Sin...| Great brush head| 1369094400| true|null|B00077B7M6| null| 5.0|The Economist is ...|03 26,
2010|A3DKP8M0GSP8UK| SAM|[ Kindle Edition]|A Great Global Fi...| 1269561600| false| 2| B00CNE2GOO| null| 1.0|Not only
does my ...|02 25, 2014|A3DKP8M0GSP8UK| SAM| null| Does not work| 1393286400| true| 5|B00077B7M6| null| 5.0|The
Economist is ...|03 26, 2010|A3DKP8M0GSP8UK| SAM|[ Kindle Edition]|A Great Global Fi...| 1269561600| false| 2| +----------+----
-+-------+--------------------+-----------+-------------+-------------------+-------------------+-------------------+-------------+--------+----+----------+-
----+-------+--------------------+-----------+-------------+-------------------+-----------------+-------------------+-------------+--------+----+ only
showing top 20 rows
```

Left Join of Two Dataframes

In [27]:

```
left_join = df.join(df1, df.reviewerID == df1.reviewerID, how='left')
left_join.show()
```

```
+----------+-----+-------+--------------------+-----------+--------------+--------------------+--------------------+--------------------+--------------+--------+-
---+----+-----+-------+----------+----------+----------+------------+-----+-------+--------------+--------+----+ asin|image|overall| reviewText|
reviewTime| reviewerID| reviewerName| style|
summary|unixReviewTime|verified|vote|asin|image|overall|reviewText|reviewTime|reviewerID|reviewerName|style|summary|unixRe
+----------+-----+-------+--------------------+-----------+--------------+--------------------+--------------------+--------------------+--------------+--------+-
---+----+-----+-------+----------+----------+----------+------------+-----+-------+--------------+--------+----+ B001AO0WCG| null| 5.0|I've been
using t...|10 18, 2010|A100JSLU0DKS1Z| Lisa J.| null| love this stuff!| 1287360000| true|null|null| null| null| null| null| null| null|
null| null| null| null|null| B014Q1KWOW| null| 5.0|I just reordered....|06 27, 2016|A103B4C6BVOXDK| Amazonaholic| null|Great
Product. I ...| 1466985600| true| 6|null| null| null| null| null| null| null| null| null| null| null|null| B0009OAI40| null| 4.0|Green Tea
smell i...|08 20, 2014|A107JQ0F2BTP2I| Valerie H.| null|Green Tea smells ...| 1408492800| true|null|null| null| null| null| null| null|
null| null| null| null| null|null| B00J8NYYSE| null| 3.0|Not as a soft hol...|03 11, 2015|A109AU0BTFR94F| ARM| null| Three Stars|
1426032000| true|null|null| null| null| null| null| null| null| null| null| null| null|null| B000QU75H0| null| 5.0|Smells sooo good....|02
12, 2018|A10CLR1IQPN4B9| Kat|[,,,,,,,,,,, 16 ...|Smells great and ...| 1518393600| true|null|null| null| null| null| null| null| null| null|
null| null| null|null| B01257WUW6| null| 5.0|Thrilled to find ...| 07 8, 2018|A10CLR1IQPN4B9| Kat|[ Medium,,,,,,,,,,...| No break
outs| 1531008000| true|null|null| null| null| null| null| null| null| null| null| null| null|null| B0001435D4| null| 5.0|Love using Drip D...|
05 2, 2013|A10DOT5X6SXF2H| Killing Time| null| One of my staples| 1367452800| true|null|null| null| null| null| null| null| null|
null| null| null| null|null| B000PZ8E62| null| 3.0|It's ok, but didn...| 05 2, 2013|A10DOT5X6SXF2H| Killing Time| null| It's ok|
1367452800| true| 3|null| null| null| null| null| null| null| null| null| null| null|null| B00EOBAY3Y| null| 5.0|Fast shipping and...|09 21,
2016|A10E1H27GLD0MJ| Christine Mason|[,,,,,,,,,, Suma...| Five Stars| 1474416000| true|null|null| null| null| null| null| null| null|
null| null| null| null|null| B006RBR7NO| null| 5.0| I love it|03 22, 2017|A10ELMTUWXTEHB|maria constanza c...|[,,,,,,,,,,, 16....|
Five Stars| 1490140800| true| 2|null| null| null| null| null| null| null| null| null| null| null|null| B005CVGJFM| null| 5.0|Ive been using
th...| 03 2, 2017|A10FTXMY87DKQ3| Amazon Customer|[,,,, Health and...| Just WOW!!!| 1488412800| false| 4|null| null| null| null|
null| null| null| null| null| null| null|null| B0168Y5646| null| 5.0| My wife's favorite|04 10, 2016|A10IIGVW3NGBH9| Oneshot| null|
Five Stars| 1460246400| true|null|null| null| null| null| null| null| null| null| null| null| null|null| B000SX3380| null| 5.0|I noticed my
frek...|07 19, 2017|A10KTUJZP16YQ3| RockaPants|[ Classic,,,,,,,,,,...| So good!| 1500422400| true| 2|null| null| null| null| null| null|
null| null| null| null| null|null| B00BKJ6YD2| null| 1.0|Product arrived d...|03 23, 2015|A10KTUJZP16YQ3| RockaPants|[,,,,,,,, Rose
T...|Product arrived d...| 1427068800| true|null|null| null| null| null| null| null| null| null| null| null| null|null| B0038COKQ2| null|
5.0|I've used this pr...|01 27, 2016|A10MRNVP7L7TFC| GMP|[,,,,,,,,,,,, N...|It is great on se...| 1453852800| false|null|null| null|
null| null| null| null| null| null| null| null| null| null|null| B000OYHN2K| null| 5.0|This is perfect f...|09 21, 2015|A10Q1QIHVXKT8M|
Teresa W| null| Great Hair| 1442793600| true|null|null| null| null| null| null| null| null| null| null| null| null|null| B00BY55J7Y| null|
5.0|Love JIN Soon Nai...|06 18, 2018|A10UMUD4URRKED| Tracy P.|[ Poppy Blue,,,,,,...|The color looks e...| 1529280000|
true|null|null| null| null| null| null| null| null| null| null| null| null|null| B002D48QI4| null| 5.0|Used for years pe...| 07 7,
2016|A10ZU4TS912PWS| Virginia Halman| null| Five Stars| 1467849600| true|null|null| null| null| null| null| null| null| null| null|
null| null|null| B001965UU4| null| 4.0|May work well for...|01 31, 2017|A111HV8S5912NU| Marlene|[,,,,,,,,,,, 4.4...|May work well
for...| 1485820800| true|null|null| null| null| null| null| null| null| null| null| null| null|null| B00STG63IU| null| 2.0|I bought this and...|
12 4, 2017|A114XTMCISHBPL| Brandt King| null|I bought this and...| 1512345600| true|null|null| null| null| null| null| null| null| null|
null| null| null|null| +----------+-----+-------+--------------------+-----------+-------------+-------------------+-------------------+-------------------+--
-----------+--------+----+----+-----+-------+----------+----------+----------+------------+-----+-------+-------------+--------+----+ only showing top
20 rows
```

Right join of two dataframes df and df1.

In [28]:

```python
right_join = df.join(df1, df.reviewerID == df1.reviewerID, how='right')
right_join.show()
```

```
+----+-----+-------+----------+----------+----------+------------+-----+-------+--------------+--------+----+----------+-----+-------+--------------------+---------
--+--------------+--------------------+-----------------+--------------------+--------------+--------+----+
asin|image|overall|reviewText|reviewTime|reviewerID|reviewerName|style|summary|unixReviewTime|verified|vote| asin|image|overall|
reviewText| reviewTime| reviewerID| reviewerName| style| summary|unixReviewTime|verified|vote| +----+-----+-------+----------+----------
+----------+------------+-----+-------+--------------+--------+----+----------+-----+-------+--------------------+-----------+--------------+--------------------+---
--------------+--------------------+--------------+--------+----+ null| null| null| null| null| null| null| null| null| null| null|null|B00005N7QN| null|
5.0|Receiving Bazaar ...| 04 4, 2015|A10RG9737676YY| Linda|[ Print Magazine]|... a year at suc...| 1428105600| true|null| null| null|
null| null| null| null| null| null| null| null| null|null|B0000AWD92| null| 5.0| My son love it|01 18, 2017|A11S3TYJ8EO3ZP| Amazon
Customer| null| Five Stars| 1484697600| true|null| null| null| null| null| null| null| null| null| null| null|null|B000IOE9Y6| null| 5.0|
very good|01 18, 2017|A11S3TYJ8EO3ZP| Amazon Customer| null| Five Stars| 1484697600| true|null| null| null| null| null| null| null|
null| null| null| null| null|null|B00005N7R0| null| 5.0|The pictures are ...|11 16, 2008| A12LH2100CKQO|God is a refuge f...| null|it helps
to dream...| 1226793600| true| 4| null| null| null| null| null| null| null| null| null| null| null|null|B00006LBP6| null| 5.0|Great magazine.
V...|06 18, 2016|A14DNTKBHOAYDL| Steven Lovett| null| Five Stars| 1466208000| true|null| null| null| null| null| null| null| null| null|
null| null| null|null|B00O9K26HC| null| 5.0|I love this magaz...|03 13, 2015|A14LNI1UTNBD73| Shane Fogarty| null| Five Stars|
1426204800| true|null| null| null| null| null| null| null| null| null| null| null| null|null|B00005N7RA| null| 3.0|I had no problems...|12 28,
2011|A14PHHZNLPGP8J| readyoga|[ Print Magazine]|No problem with s...| 1325030400| true| 3| null| null| null| null| null| null| null|
null| null| null| null|null|B000XBBZ9Q| null| 5.0| Grandma love it!|06 12, 2016|A150SDVSXZ8JZA| Joel Paez| null| Great magazine|
1465689600| true|null| null| null| null| null| null| null| null| null| null| null| null|null|B000ILVRWQ| null| 5.0|Bought this for m...|12 21,
2016|A152ZS82PP81BR|Bill from Bakersf...| null|Bought this for m...| 1482278400| true|null| null| null| null| null| null| null| null| null|
null| null| null|null|B00005N7QW| null| 5.0|Beautiful quality...|07 10, 2013|A15HPE00GOC3WO| CarlaVS|[ Kindle Edition]| Love it
indeed!| 1373414400| true|null| null| null| null| null| null| null| null| null| null| null| null|null|B00HG1BOWO| null| 5.0|Relatively low co...|
10 5, 2014|A15NF13UMC12AZ| Bobo|[ Print Magazine]| Good stuff!| 1412467200| true|null| null| null| null| null| null| null| null| null|
null| null| null|null|B00007MHIY| null| 5.0|Like a child at t...|04 10, 2007|A15S6U3CBETB6D| Anna Sochocky| null|Gorgeous and
info...| 1176163200| false|null| null| null| null| null| null| null| null| null| null| null| null|null|B00005NIND| null| 5.0|I always look for...|01
29, 2014|A162HO9OAQR5FC| reesenana|[ Kindle Edition]| Always yummy!| 1390953600| false|null| null| null| null| null| null| null| null|
null| null| null| null|null|B00005N7Q1| null| 5.0|I bought this for...|08 21, 2014|A167DMHQREZS7N| Shirley J. Twogood|[ Print
Magazine]| Five Stars| 1408579200| true|null| null| null| null| null| null| null| null| null| null| null| null|null|B00005N7T3| null| 4.0|I only
read this ...|08 29, 2014|A16BU96ESFYNF7| David M|[ Kindle Edition]| Four Stars| 1409270400| true|null| null| null| null| null| null|
null| null| null| null| null| null|null|B000HOJOZ6| null| 5.0|An excellent, in-...|02 20, 2015|A16FR674KWT5KL| Bradley|[ Kindle Edition]|
Deep reads| 1424390400| false| 3| null| null| null| null| null| null| null| null| null| null| null|null|B015PRLQNM| null| 5.0| Great
magazine|03 24, 2018|A17F3O4JFR6GWJ|Venkatarman Gokul...| null| Five Stars| 1521849600| true|null| null| null| null| null| null|
null| null| null| null| null| null|null|B00005NIPX| null| 5.0|There's only one ...|04 28, 2008|A17UT84AR48XZI| Karen Harrington|[ Print
Magazine]|The great writing...| 1209340800| false| 2| null| null| null| null| null| null| null| null| null| null| null|null|B000063XJL| null|
4.0|The magazine is s...| 05 1, 2018|A18DL04OJHDUFY| SC Mom| null|The magazine is s...| 1525132800| true|null| null| null| null|
null| null| null| null| null| null| null| null|null|B00005UF1T| null| 5.0|If you want to sh...|05 12, 2002|A18OR8GCQFO7H6| C. Isner| null|
sheer beauty| 1021161600| false| 8| +----+-----+-------+----------+----------+----------+------------+-----+-------+-------------+--------+----+---------
-+-----+-------+--------------------+-----------+--------------+--------------------+----------------+-------------------+-------------+--------+----+ only
showing top 20 rows
```

Distinct function used in df with count.

Distinct function with variable reviewerID. In total 574628 counts only 416174 are with distinct reviewerID.

In [29]:
```
df.distinct().count()
```

```
Out[124]: 540046
```

In [30]:
```
df.select('reviewerID').distinct().show(5)
df.select('reviewerID').distinct().count()
```

```
+--------------+ reviewerID| +--------------+ A3CIZTN0CSIJN| A35YXEDATMIJ9S| A3U2344Q25MYW7| A8279Z6EBQI6|
A2AVMDTFW79N2N| +--------------+ only showing top 5 rows Out[128]: 416174
```

User defined function say_hello used with reviewerName and created new variable "greetings".

In [31]:

```python
from pyspark.sql.types import StringType
from pyspark.sql.functions import udf, col
def say_hello(name : str) -> str:
     return f"Hello {name}"

say_hello_udf = udf(lambda name: say_hello(name), StringType())

df.withColumn("greetings", say_hello_udf(col("reviewerName"))).show()
```

```
+----------+-----+-------+--------------------+-----------+--------------+------------------+--------------------+--------------------+--------------+--------+----+---
-----------------+ asin|image|overall| reviewText| reviewTime| reviewerID| reviewerName| style| summary|unixReviewTime|verified|vote|
greetings| +----------+-----+-------+--------------------+-----------+--------------+------------------+--------------------+--------------------+--------------+----
----+----+--------------------+ B00004U9V2| null| 2.0|I bought two of t...|06 15, 2010|A1Q6MUU0B2ZDQG| D. Poston| null|dispensers
don't ...| 1276560000| true| 3| Hello D. Poston| B00004U9V2| null| 5.0|Believe me, over ...| 01 7, 2010|A3HO2SQDCZIE9S| chandra|
null|Best hand cream e...| 1262822400| true| 14| Hello chandra| B00004U9V2| null| 5.0| Great hand lotion|04 18,
2018|A2EM03F99X3RJZ| Maureen G|[,,,,,,,,,,, 3.5...| Five Stars| 1524009600| true|null| Hello Maureen G| B00004U9V2| null| 5.0|This
is the best ...|04 18, 2018| A3Z74TDRGD0HU| Terry K|[,,,,,,,,,,, 3.5...| Five Stars| 1524009600| true|null| Hello Terry K| B00004U9V2|
null| 5.0|The best non- oil...|04 17, 2018|A2UXFNW9RTL4VM| Patricia Wood|[,,,,,,,,,,, 3.5...|I always have a b...| 1523923200|
true|null| Hello Patricia Wood| B00004U9V2| null| 5.0|Ive used this lot...|04 14, 2018| AXX5G4LFF12R6| Ralla|[,,,,,,,,,,, 250...|Ive used
this lot...| 1523664000| true|null| Hello Ralla| B00004U9V2| null| 5.0|Works great for d...|04 11, 2018| A7GUKMOJT2NR6| Lydia
Speight|[,,,,,,,,,,, 3.5...| Five Stars| 1523404800| true|null| Hello Lydia Speight| B00004U9V2| null| 5.0|The best hand cre...|04 11,
2018|A3FU4L59BHA9FY| Allen Semer|[,,,,,,,,,,, 3.5...| Made in the USA| 1523404800| true|null| Hello Allen Semer| B00004U9V2| null|
5.0|LOVE THIS SCENT!!...| 04 7, 2018|A1AMNMIPQMXH9M| Vets park|[,,,,,,,,,,, 3.5...|Moistens and smel...| 1523059200| true|null|
Hello Vets park| B00004U9V2| null| 5.0|Its a great moist...| 04 6, 2018|A3DMBDTA8VGWSX| Cynthia P. Irving|[,,,,,,,,,,, 3.5...| Five
Stars| 1522972800| true|null|Hello Cynthia P. ...| B00004U9V2| null| 5.0|This hand cream i...| 04 5, 2018|A160DTI3H7VHLQ| CB|
[,,,,,,,,,,, 0.9...| Five Stars| 1522886400| true|null| Hello CB| B00004U9V2| null| 5.0|I am a healthcare...| 04 5,
2018|A1H41DKPDPVA0R| Donna Butler RN|[,,,,,,,,,,, 250...|Best hand therapy...| 1522886400| true|null|Hello Donna Butle...|
B00004U9V2| null| 5.0|have used on and ...| 04 5, 2018| A2BDI7THUMJ8V| Teresa K. L.|[,,,,,,,,,,, 250...|Product is good f...|
1522886400| true|null| Hello Teresa K. L.| B00004U9V2| null| 5.0| Great hand cream| 04 3, 2018| AM7EBP5TRX7AC|Glenn B.
Guilbault|[,,,,,,,,,,, 250...| Five Stars| 1522713600| true|null|Hello Glenn B. Gu...| B00004U9V2| null| 5.0|This is my favori...| 04 2,
2018|A31FOVCS3WTWPT| Pam|[,,,,,,,,,,, 3.5...|My Favorite Lotio...| 1522627200| false|null| Hello Pam| B00004U9V2| null|
4.0|Soothing! Love th...|03 30, 2018| AXUU8F9EM6U3E| Amazon Customer|[,,,,,,,,,,, 0.9...| Love the way it| 1522368000|
true|null|Hello Amazon Cust...| B00004U9V2| null| 5.0|My wife loves the...|03 29, 2018|A24B46V78ATNRP| Michael Konrad|[,,,,,,,,,,,
250...| Wonderful| 1522281600| true|null|Hello Michael Konrad| B00004U9V2| null| 5.0|I always loved th...|03 27, 2018|
ABUBKML2EONCG| Lotte Hersey|[,,,,,,,,,,, 3.5...| Five Stars| 1522108800| true|null| Hello Lotte Hersey| B00004U9V2| null|
5.0|Absolutely great....|03 26, 2018|A2UA6E1RVG3C1I| Ginny|[,,,,,,,,,,, 0.9...| Absolutely great!| 1522022400| true|null| Hello Ginny|
B00004U9V2| null| 1.0|SOOOO not worth t...|03 23, 2018|A1TRMJHEDGX0HF| soulsurfer|[,,,,,,,,,,, 0.9...| Disappointed.|
1521763200| true|null| Hello soulsurfer| +----------+-----+-------+--------------------+-----------+--------------+------------------+--------------------+---
-----------------+--------------+--------+----+--------------------+ only showing top 20 rows
```

One Hot Encoder

It maps a column of label indices to a column of binary vectors, with at most a single one-value. As there are no continuous variable therefore used it in summary variable.

In [32]:

```python
from pyspark.ml.feature import OneHotEncoder, StringIndexer
stringIndexer = StringIndexer(inputCol="overall", outputCol="overallIndex")
model = stringIndexer.fit(df)
indexed = model.transform(df)

encoder = OneHotEncoder(inputCol="overallIndex", outputCol="overallVec")
encoded = encoder.transform(indexed)
encoded.show()
```

```
+----------+-----+-------+-------------------+-----------+-------------+-----------------+-------------------+-------------------+-------------+--------+---
-+------------+-------------+ asin|image|overall| reviewText| reviewTime| reviewerID| reviewerName| style|
summary|unixReviewTime|verified|vote|overallIndex| overallVec| +----------+-----+-------+-------------------+-----------+-------------+------
------------+-------------------+-------------------+-------------+--------+----+------------+-------------+ B00004U9V2| null| 2.0|I bought two of
t...|06 15, 2010|A1Q6MUU0B2ZDQG| D. Poston| null|dispensers don't ...| 1276560000| true| 3| 4.0| (4,[],[])| B00004U9V2| null|
5.0|Believe me, over ...| 01 7, 2010|A3HO2SQDCZIE9S| chandra| null|Best hand cream e...| 1262822400| true| 14| 0.0|(4,[0],
[1.0])| B00004U9V2| null| 5.0| Great hand lotion|04 18, 2018|A2EM03F99X3RJZ| Maureen G|[,,,,,,,,,,, 3.5...| Five Stars|
1524009600| true|null| 0.0|(4,[0],[1.0])| B00004U9V2| null| 5.0|This is the best ...|04 18, 2018| A3Z74TDRGD0HU| Terry K|
[,,,,,,,,,,, 3.5...| Five Stars| 1524009600| true|null| 0.0|(4,[0],[1.0])| B00004U9V2| null| 5.0|The best non- oil...|04 17,
2018|A2UXFNW9RTL4VM| Patricia Wood|[,,,,,,,,,,, 3.5...|I always have a b...| 1523923200| true|null| 0.0|(4,[0],[1.0])|
B00004U9V2| null| 5.0|Ive used this lot...|04 14, 2018| AXX5G4LFF12R6| Ralla|[,,,,,,,,,,, 250...|Ive used this lot...| 1523664000|
true|null| 0.0|(4,[0],[1.0])| B00004U9V2| null| 5.0|Works great for d...|04 11, 2018| A7GUKMOJT2NR6| Lydia Speight|[,,,,,,,,,,,
3.5...| Five Stars| 1523404800| true|null| 0.0|(4,[0],[1.0])| B00004U9V2| null| 5.0|The best hand cre...|04 11,
2018|A3FU4L59BHA9FY| Allen Semer|[,,,,,,,,,,, 3.5...| Made in the USA| 1523404800| true|null| 0.0|(4,[0],[1.0])| B00004U9V2|
null| 5.0|LOVE THIS SCENT!!...| 04 7, 2018|A1AMNMIPQMXH9M| Vets park|[,,,,,,,,,,, 3.5...|Moistens and smel...| 1523059200|
true|null| 0.0|(4,[0],[1.0])| B00004U9V2| null| 5.0|Its a great moist...| 04 6, 2018|A3DMBDTA8VGWSX| Cynthia P. Irving|[,,,,,,,,,,,
3.5...| Five Stars| 1522972800| true|null| 0.0|(4,[0],[1.0])| B00004U9V2| null| 5.0|This hand cream i...| 04 5,
2018|A160DTI3H7VHLQ| CB|[,,,,,,,,,,, 0.9...| Five Stars| 1522886400| true|null| 0.0|(4,[0],[1.0])| B00004U9V2| null| 5.0|I am a
healthcare...| 04 5, 2018|A1H41DKPDPVA0R| Donna Butler RN|[,,,,,,,,,,, 250...|Best hand therapy...| 1522886400| true|null| 0.0|
(4,[0],[1.0])| B00004U9V2| null| 5.0|have used on and ...| 04 5, 2018| A2BDI7THUMJ8V| Teresa K. L.|[,,,,,,,,,,, 250...|Product is
good f...| 1522886400| true|null| 0.0|(4,[0],[1.0])| B00004U9V2| null| 5.0| Great hand cream| 04 3, 2018|
AM7EBP5TRX7AC|Glenn B. Guilbault|[,,,,,,,,,,, 250...| Five Stars| 1522713600| true|null| 0.0|(4,[0],[1.0])| B00004U9V2| null|
5.0|This is my favori...| 04 2, 2018|A31FOVCS3WTWPT| Pam|[,,,,,,,,,,, 3.5...|My Favorite Lotio...| 1522627200| false|null| 0.0|(4,
[0],[1.0])| B00004U9V2| null| 4.0|Soothing! Love th...|03 30, 2018| AXUU8F9EM6U3E| Amazon Customer|[,,,,,,,,,,, 0.9...| Love the
way it| 1522368000| true|null| 1.0|(4,[1],[1.0])| B00004U9V2| null| 5.0|My wife loves the...|03 29, 2018|A24B46V78ATNRP|
Michael Konrad|[,,,,,,,,,,, 250...| Wonderful| 1522281600| true|null| 0.0|(4,[0],[1.0])| B00004U9V2| null| 5.0|I always loved th...|03
27, 2018| ABUBKML2EONCG| Lotte Hersey|[,,,,,,,,,,, 3.5...| Five Stars| 1522108800| true|null| 0.0|(4,[0],[1.0])| B00004U9V2| null|
5.0|Absolutely great....|03 26, 2018|A2UA6E1RVG3C1I| Ginny|[,,,,,,,,,,, 0.9...| Absolutely great!| 1522022400| true|null| 0.0|(4,[0],
[1.0])| B00004U9V2| null| 1.0|SOOOO not worth t...|03 23, 2018|A1TRMJHEDGX0HF| soulsurfer|[,,,,,,,,,,, 0.9...| Disappointed.|
1521763200| true|null| 2.0|(4,[2],[1.0])| +----------+-----+-------+--------------------+-----------+--------------+------------------+--------------------
+--------------------+--------------+--------+----+------------+-------------+ only showing top 20 rows
```