

Spark by {Examples}(http://sparkbyexamples.com/)

PySpark Tutorial

PySpark Tutorial For Beginners

(https://sparkbyexamples.com/pyspark-tutorial/)

PySpark – Features

(https://sparkbyexamples.com/pyspark-tutorial/#features)

PySpark – Advantages

(https://sparkbyexamples.com/pyspark-tutorial/#advantages)

PySpark – Modules & Packages

(https://sparkbyexamples.com/pyspark-tutorial/#modules-packages)

PySpark – Cluster Managers

(https://sparkbyexamples.com/pyspark-tutorial/#cluster-manager)

PySpark – Install on Windows

(https://sparkbyexamples.com/pyspark-tutorial/#pyspark-installation)

PySpark – Web/Application UI

(https://sparkbyexamples.com/spark/spark-web-ui-understanding/)

PySpark – SparkSession

(https://sparkbyexamples.com/pyspark/pyspark-what-is-sparksession/)

PySpark – RDD

(https://sparkbyexamples.com/pyspark-rdd)

PySpark – Parallelize

(https://sparkbyexamples.com/pyspark/pyspark-parallelize-create-rdd/)

PySpark – repartition() vs coalesce()

(https://sparkbyexamples.com/pyspark/pyspark-repartition-vs-coalesce/)

PySpark – Broadcast Variables

(https://sparkbyexamples.com/pyspark/pyspark-broadcast-variables/)

PySpark (https://sparkbyexamples.com/pyspark-tutorial/)

Hive (https://sparkbyexamples.com/apache-hive-tutorial/)

HBase (https://sparkbyexamples.com/apache-hbase-tutorial/)

Kafka (https://sparkbyexamples.com/apache-kafka-tutorials-with-examples/)

FAQ's (https://sparkbyexamples.com/spark-questions/)

More (https://sparkbyexamples.com/)


POPULAR

Monogrammed Leather AirPods Cover - Tan

Now your Apple AirPods have a new home. Personalize it and flaunt it! Synthetic Leather. Perfect touch to this craft. Can be personalized with 2 initials Shipping- 4-6 working days Pa

The Messy Corner

PySpark map() Transformation

 **NNK**
(https://sparkbyexamples.com/author/admin/)
 **PySpark**
(https://sparkbyexamples.com/category/pyspark/)

PySpark map (map()) is an RDD transformation that is used to apply the transformation function (lambda) on every element of RDD/DataFrame and returns a new RDD. In this article, you will learn the syntax and usage of the RDD map() transformation with an example and how to use it with DataFrame.

Incredible Savings

Data Science Mercury Learn

Data Science-Amazing new deal everything U need to know re: data science

humblebundle.com

OPEN

RDD map() transformation is used to apply any complex operations like adding a column, updating a column, transforming the data e.t.c, the output



[PySpark – Accumulator
\(https://sparkbyexamples.com/pyspark/pyspark-accumulator-with-example/\)](https://sparkbyexamples.com/pyspark/pyspark-accumulator-with-example/)

PySpark DataFrame

[PySpark – Create a DataFrame
\(https://sparkbyexamples.com/pyspark/different-ways-to-create-dataframe-in-pyspark/\)](https://sparkbyexamples.com/pyspark/different-ways-to-create-dataframe-in-pyspark/)

[PySpark – Create an empty DataFrame
\(https://sparkbyexamples.com/pyspark/pyspark-create-an-empty-dataframe/\)](https://sparkbyexamples.com/pyspark/pyspark-create-an-empty-dataframe/)

[PySpark – Convert RDD to DataFrame
\(https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/\)](https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/)

[PySpark – Convert DataFrame to Pandas
\(https://sparkbyexamples.com/pyspark/convert-pyspark-dataframe-to-pandas/\)](https://sparkbyexamples.com/pyspark/convert-pyspark-dataframe-to-pandas/)

[PySpark – show\(\)
\(https://sparkbyexamples.com/pyspark/pyspark-show-display-dataframe-contents-in-table/\)](https://sparkbyexamples.com/pyspark/pyspark-show-display-dataframe-contents-in-table/)

[PySpark – StructType & StructField
\(https://sparkbyexamples.com/pyspark/pyspark-structtype-and-structfield/\)](https://sparkbyexamples.com/pyspark/pyspark-structtype-and-structfield/)

[PySpark – Row Class
\(https://sparkbyexamples.com/pyspark/pyspark-row-using-rdd-dataframe/\)](https://sparkbyexamples.com/pyspark/pyspark-row-using-rdd-dataframe/)

[PySpark – Column Class
\(https://sparkbyexamples.com/pyspark/pyspark-column-functions/\)](https://sparkbyexamples.com/pyspark/pyspark-column-functions/)

[PySpark – select\(\)
\(https://sparkbyexamples.com/pyspark/select-columns-from-pyspark-dataframe/\)](https://sparkbyexamples.com/pyspark/select-columns-from-pyspark-dataframe/)

[PySpark – collect\(\)
\(https://sparkbyexamples.com/pyspark/pyspark-collect/\)](https://sparkbyexamples.com/pyspark/pyspark-collect/)

[PySpark – withColumn\(\)
\(https://sparkbyexamples.com/pyspark/pyspark-withcolumn/\)](https://sparkbyexamples.com/pyspark/pyspark-withcolumn/)

of map transformations would always have the same number of records as input.

- **Note1:** DataFrame doesn't have map() transformation to use with DataFrame hence you need to DataFrame to RDD first.
- **Note2:** If you have a heavy initialization use PySpark mapPartitions() transformation instead of map(), as with mapPartitions() heavy initialization executes only once for each partition instead of every record.

Related: [Spark map\(\) vs mapPartitions\(\) Explained with Examples](#)
(https://sparkbyexamples.com/spark/spark-map-vs-mappartitions-transformation/)

First, let's create an RDD from the list.

```
from pyspark.sql import SparkSession
spark = SparkSession.builder.master("local[*]").appName("SparkByExamples.com").getOrCreate()

data = ["Project", "Gutenberg's", "in", "Wonderland", "Project", "Gutenberg's", "in", "Wonderland", "Project", "Gutenberg's"]

rdd = spark.sparkContext.parallelize(data)
```

map() Syntax

```
map(f, preservesPartitioning=False)
```

PySpark map() Example with RDD

In this PySpark map() example, we are adding a new element with value 1 for each element, the result of the RDD is PairRDDFunctions which contains key-value pairs, word of type String as Key and 1 of type Int as value.

[PySpark – withColumnRenamed\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-rename-dataframe-column/\)](#)

[PySpark – where\(\) & filter\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-where-filter/\)](#)

[PySpark – drop\(\) & dropDuplicates\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-distinct-to-drop-duplicates/\)](#)

[PySpark – orderBy\(\) and sort\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-orderby-and-sort-explained/\)](#)

[PySpark – groupBy\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-groupby-explained-with-example/\)](#)

[PySpark – join\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-join-explained-with-examples/\)](#)

[PySpark – union\(\) & unionAll\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-union-and-unionall/\)](#)

[PySpark – unionByName\(\).
\(https://sparkbyexamples.com/spark/spark-merge-two-dataframes-with-different-columns/\)](#)

[PySpark – UDF \(User Defined Function\).
\(https://sparkbyexamples.com/pyspark/pyspark-udf-user-defined-function/\)](#)

[PySpark – map\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-map-transformation/\)](#)

[PySpark – flatMap\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-flatmap-transformation/\)](#)

[pyspark – foreach\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-loop-iterate-through-rows-in-dataframe/#use-foreach-loop-through-dataframe\)](#)

[PySpark – sample\(\) vs sampleBy\(\).](#)

Incredible Savings

Data Science Mercury Learn

Data Science-Amazing new deal everything U need to know re: data science

[humblebundle.com](#)

OPEN

```
rdd2=rdd.map(lambda x: (x,1))
for element in rdd2.collect():
    print(element)
```

This yields below output.

```
('Project', 1)
('Gutenberg's', 1)
('Alice's', 1)
('Adventures', 1)
('in', 1)
('Wonderland', 1)
('Project', 1)
('Gutenberg's', 1)
('Adventures', 1)
('in', 1)
('Wonderland', 1)
('Project', 1)
('Gutenberg's', 1)
```

PySpark map() Example with DataFrame

PySpark DataFrame doesn't have map() transformation to apply the lambda function, when you wanted to apply the custom transformation, you need to convert the DataFrame to RDD and apply the map() transformation. Let's use another dataset to explain this.



The Ultimate **Keyword Research Checklist** –

Be Future Ready

Online Degrees

MBA, Data Science, Big Data, Artificial Intelligence, Cyber Security, among many others

[aimsenrich.com](#)

OPEN

[\(https://sparkbyexamples.com/pyspark/pyspark-sampling-example/\)](https://sparkbyexamples.com/pyspark/pyspark-sampling-example/)

[PySpark – fillna\(\) & fill\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-fillna-fill-replace-null-values/\)](https://sparkbyexamples.com/pyspark/pyspark-fillna-fill-replace-null-values/)

[PySpark – pivot\(\)_\(Row to Column\).
\(https://sparkbyexamples.com/pyspark/pyspark-pivot-and-unpivot-dataframe/\)](https://sparkbyexamples.com/pyspark/pyspark-pivot-and-unpivot-dataframe/)

[PySpark – partitionBy\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-partitionby-example/\)](https://sparkbyexamples.com/pyspark/pyspark-partitionby-example/)

[PySpark – ArrayType Column \(Array\).
\(https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-with-examples/\)](https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-with-examples/)

[PySpark – MapType \(Map/Dict\).
\(https://sparkbyexamples.com/pyspark/pyspark-maptype-dict-examples/\)](https://sparkbyexamples.com/pyspark/pyspark-maptype-dict-examples/)

PySpark SQL Functions

[PySpark – Aggregate Functions
\(https://sparkbyexamples.com/pyspark/pyspark-aggregate-functions/\)](https://sparkbyexamples.com/pyspark/pyspark-aggregate-functions/)

[PySpark – Window Functions
\(https://sparkbyexamples.com/pyspark/pyspark-window-functions/\)](https://sparkbyexamples.com/pyspark/pyspark-window-functions/)

[PySpark – Date and Timestamp Functions
\(https://sparkbyexamples.com/pyspark/pyspark-sql-date-and-timestamp-functions/\)](https://sparkbyexamples.com/pyspark/pyspark-sql-date-and-timestamp-functions/)

[PySpark – JSON Functions
\(https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/\)](https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/)

PySpark Datasources

[PySpark – Read & Write CSV File
\(https://sparkbyexamples.com/pyspark/pyspark-read-csv-file-into-dataframe/\)](https://sparkbyexamples.com/pyspark/pyspark-read-csv-file-into-dataframe/)

[PySpark – Read & Write Parquet File
\(https://sparkbyexamples.com/pyspark/pyspark-read-write-parquet-file/\)](https://sparkbyexamples.com/pyspark/pyspark-read-write-parquet-file/)

```
data = [('James','Smith','M',30),
        ('Anna','Rose','F',41),
        ('Robert','Williams','M',62),
        ]
```

```
columns = ["firstname","lastname","gender","salary"]
df = spark.createDataFrame(data,columns)
df.show()
```

```
+-----+-----+-----+-----+
|firstname|lastname|gender|salary|
+-----+-----+-----+-----+
|   James|   Smith|     M|    3000|
|   Anna|   Rose|     F|   41000|
| Robert|Williams|     M|   62000|
+-----+-----+-----+-----+
```

```
# Referring columns by index.
rdd2=df.rdd.map(lambda x:
                (x[0]+", "+x[1],x[2],x[3]*2)
                )
df2=rdd2.toDF(["name","gender","new_salary"])
df2.show()
```

```
+-----+-----+-----+-----+
|          name|gender|new_salary|
+-----+-----+-----+-----+
| James,Smith|     M|    60000|
| Anna,Rose|     F|  164000|
|Robert,Williams|     M|  248000|
+-----+-----+-----+-----+
```

Note that above I have used index to get the column values, alternatively, you can also refer to the DataFrame column names while iterating.

```
# Referring Column Names
rdd2=df.rdd.map(lambda x:
                (x["firstname"]+", "+x["lastname"]
                )
                )
```

Another alternative

```
# Referring Column Names
rdd2=df.rdd.map(lambda x:
                (x.firstname+", "+x.lastname
                )
                )
```

[pyspark/pyspark-read-and-write-parquet-file/](#)

[PySpark – Read & Write JSON file](#)
(<https://sparkbyexamples.com/pyspark/pyspark-read-json-file-into-dataframe/>)



PySpark Built-In Functions

[PySpark – when\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-when-otherwise/>)

[PySpark – expr\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-sql-expr-expression-function/>)

[PySpark – lit\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-lit-add-literal-constant/>)

[PySpark – split\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-convert-string-to-array-column/>)

[PySpark – concat_ws\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-convert-array-column-to-string-column/>)

[Pyspark – substring\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-substring-from-a-column/>)

[PySpark – translate\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-replace-column-values/#translate-replace-character-by-character>)

[PySpark – regexp_replace\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-replace-column->

You can also create a custom function to perform an operation. Below func1() function executes for every DataFrame row from the lambda function.



```
# By Calling function
def func1(x):
    firstName=x.firstname
    lastName=x.lastname
    name=firstName+","+lastName
    gender=x.gender.lower()
    salary=x.salary*2
    return (name,gender,salary)

rdd2=df.rdd.map(lambda x: func1
```

Complete PySpark map() example

Below is complete example of PySpark map() transformation.

[values/#regex_replace-replace-string-columns\).](#)

[PySpark – overlay\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-replace-column-values/#overlay-function\)](#)

[PySpark – to_timestamp\(\).
\(https://sparkbyexamples.com/spark/pyspark-to_timestamp-convert-string-to-timestamp-type/\)](#)

[PySpark – to_date\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-to_date-convert-timestamp-to-date/\)](#)

[PySpark – date_format\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-date_format-convert-date-to-string-format/\)](#)

[PySpark – datediff\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/#datediff\)](#)

[PySpark – months_between\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/#months_between\(\)\)](#)

[PySpark – explode\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-explode-nested-array-into-rows/\)](#)

[PySpark – array_contains\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-with-examples/#array_contains\)](#)

[PySpark – array\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-with-examples/#array\)](#)

[PySpark – collect_list\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-aggregate-functions/#collect-list\)](#)

[PySpark – collect_set\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-aggregate-functions/#collect-set\)](#)

[PySpark – create_map\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-convert-dataframe-columns-to-maptype-dict/\)](#)

```
from pyspark.sql import SparkSession
spark = SparkSession.builder.appName('PySpark SQL').getOrCreate()
```

```
data = ["Project",
        "Gutenberg's",
        "Alice's",
        "Adventures",
        "in",
        "Wonderland",
        "Project",
        "Gutenberg's",
        "Adventures",
        "in",
        "Wonderland",
        "Project",
        "Gutenberg's"]
```

```
rdd=spark.sparkContext.parallelize(data)
```

```
rdd2=rdd.map(lambda x: (x,1))
for element in rdd2.collect():
    print(element)
```

```
data = [('James','Smith','M',30),
        ('Anna','Rose','F',41),
        ('Robert','Williams','M',62),
        ]
```

```
columns = ["firstname","lastname","gender","salary"]
df = spark.createDataFrame(data,columns)
df.show()
```

```
rdd2=df.rdd.map(lambda x:
    (x[0]+" "+x[1],x[2],x[3]*2)
)
df2=rdd2.toDF(["name","gender","salary"])
df2.show()
```

#Referring Column Names

```
rdd2=df.rdd.map(lambda x:
    (x["firstname"]+" "+x["lastname"],x["gender"],x["salary"]*2)
)
```

#Referring Column Names

```
rdd2=df.rdd.map(lambda x:
    (x.firstname+" "+x.lastname,x.gender,x.salary*2)
)
```

```
def func1(x):
    firstName=x.firstname
    lastName=x.lastname
    name=firstName+" "+lastName
    gender=x.gender.lower()
    salary=x.salary*2
    return (name,gender,salary)
```


[PySpark – map_keys\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-maptype-dict-examples/#map_keys\)](https://sparkbyexamples.com/pyspark/pyspark-maptype-dict-examples/#map_keys)

[PySpark – map_values\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-maptype-dict-examples/#map_values\)](https://sparkbyexamples.com/pyspark/pyspark-maptype-dict-examples/#map_values)

[PySpark – struct\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-structtype-and-structfield/#update-struct-function\)](https://sparkbyexamples.com/pyspark/pyspark-structtype-and-structfield/#update-struct-function)

[PySpark – countDistinct\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-count-distinct-from-dataframe/\)](https://sparkbyexamples.com/pyspark/pyspark-count-distinct-from-dataframe/)

[PySpark – sum\(\).avg\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-dataframe-groupby-and-sort-by-descending-order/\)](https://sparkbyexamples.com/pyspark/pyspark-dataframe-groupby-and-sort-by-descending-order/)

[PySpark – row_number\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-window-functions/#row_number\)](https://sparkbyexamples.com/pyspark/pyspark-window-functions/#row_number)

[PySpark – rank\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-window-functions/#rank\)](https://sparkbyexamples.com/pyspark/pyspark-window-functions/#rank)

[PySpark – dense_rank\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-window-functions/#dense_rank\)](https://sparkbyexamples.com/pyspark/pyspark-window-functions/#dense_rank)

[PySpark – percent_rank\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-window-functions/#percent_rank\)](https://sparkbyexamples.com/pyspark/pyspark-window-functions/#percent_rank)

[PySpark – typedLit\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-lit-add-literal-constant/#typedlit\)](https://sparkbyexamples.com/pyspark/pyspark-lit-add-literal-constant/#typedlit)

[PySpark – from_json\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#from_json\)](https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#from_json)

[PySpark – to_json\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#to_json\)](https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#to_json)

[PySpark – json_tuple\(\).
\(https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#json_tuple\)](https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#json_tuple)

```
rdd2=df.rdd.map(lambda x: func1
```

In conclusion, you have learned how to apply a `map()` transformation on every element of PySpark RDD and learned it returns the same number of elements as input RDD. This is one of the differences between `map()` and `flatMap()` (<https://sparkbyexamples.com/pyspark/pyspark-rdd-flatmap-transformation/>) transformations. And you have also learned how to use `map()` on DataFrame by converting DataFrame to RDD.

Happy Learning !!

Reference:

<https://spark.apache.org/docs/latest/api/python/pyspark.sql.html>
(<https://spark.apache.org/docs/latest/api/python/pyspark.sql.html>)

Share this:



(<https://sparkbyexamples.com/pyspark/pyspark-map-transformation/?share=facebook&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-map-transformation/?share=reddit&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-map-transformation/?share=pinterest&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-map-transformation/?share=tumblr&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-map-transformation/?share=pocket&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-map-transformation/?share=linkedin&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-map-transformation/?share=twitter&nb=1>)

TAGS: [MAP\(\)](#).
([HTTPS://SPARKBYEXAMPLES.COM/TAG/MAP-2/](https://sparkbyexamples.com/tag/map-2/))

[PySpark – get_json_object\(\)
\(https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#get_json_object\)](https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#get_json_object)

[PySpark – schema_of_json\(\)
\(https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#schema_of_json\)](https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#schema_of_json)



[\(https://sparkbyexamples.com/author/admin/\)](https://sparkbyexamples.com/author/admin/)

NNK

[\(https://Sparkbyexamples.Com/Author/Admin/\)](https://Sparkbyexamples.Com/Author/Admin/)

SparkByExamples.com is a Big Data and Spark examples community page, all examples are simple and easy to understand and well tested in our development environment [Read more ..](#)
(<https://sparkbyexamples.com/about-sparkbyexamples/>)

Leave a Reply

Enter your comment here...

Categories

Apache Hadoop
(<https://sparkbyexamples.com/category/hadoop/>)

Apache Spark
(<https://sparkbyexamples.com/category/spark/>)

Apache Spark Streaming
(<https://sparkbyexamples.com/category/spark/apache-spark-streaming/>)

Apache Kafka
(<https://sparkbyexamples.com/category/kafka/>)

Apache HBase
(<https://sparkbyexamples.com/category/hbase/>)

Apache Cassandra
(<https://sparkbyexamples.com/category/cassandra/>)

Snowflake Database
(<https://sparkbyexamples.com/category/snowflake/>)

H2O Sparkling Water
(<https://sparkbyexamples.com/category/>)

Recent Posts

Spark regexp_replace() – Replace String Value
(https://sparkbyexamples.com/spark/spark-regexp_replace-replace-string-value/)

How to Run a PySpark Script from Python?
(<https://sparkbyexamples.com/pyspark/run-pyspark-script-from-python-subprocess/>)

Spark SQL like() Using Wildcard Example
(<https://sparkbyexamples.com/spark/spark-sql-like-using-wildcard-example/>)

Spark isin() & IS NOT IN Operator Example
(<https://sparkbyexamples.com/spark/spark-isin-is-not-in-operator-example/>)

Spark – Get Size/Length of Array & Map Column
(<https://sparkbyexamples.com/spark/spark-get-size-length-of-array-map-column/>)

Spark Using Length/Size Of a DataFrame Column

About SparkByExamples.Com

SparkByExamples.com is a Big Data and Spark examples community page, all examples are simple and easy to understand, and well tested in our development environment [Read more ..](#)
(<https://sparkbyexamples.com/about-sparkbyexamples/>)

Follow Us



ry/h2o-sparkling-water/


PySpark
(https://sparkbyexamples.com/category/pyspark/)


(https://sparkbyexamples.com/spark/spark-using-length-size-of-a-dataframe-column/)

Spark rlike() Working with Regex Matching Examples
(https://sparkbyexamples.com/spark/spark-rlike-regex-matching-examples/)


Spark Check String Column Has Numeric Values
(https://sparkbyexamples.com/spark/spark-check-string-column-has-numeric-values/)

Spark Check Column Data Type is Integer or String
(https://sparkbyexamples.com/spark/spark-check-column-data-type-is-integer-or-string/)

 (https://www.facebook.com/sparkbyexamples/)

 /www.facebook.com/sparkbyexamples/ (https://www.linkedin.com/company/sparkbyexamples/)

 (https://github.com/sparkbyexamples/sparkbyexamples/blob/master/README.md)

 acebo (https://medium.com/sparkbyexamples)

 /twitter ok.co nkedin /github

 .com/s m/spar .com/i .com/s

 parkby kbyexa n/n-nk- park-examples/ b860a example-les)) 8193/) les/)

