

Spark by {Examples} (https://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/)  | -30% | -50% | -30% |
| 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/)                                       |  |

shop.adidas.co.in

# PySpark withColumnRenamed to Rename Column on DataFrame

Use PySpark withColumnRenamed() to rename a DataFrame column, we often need to rename one column or multiple (or all) columns on PySpark DataFrame, you can do this in several ways. When columns are nested it becomes complicated.

Since DataFrame's are an immutable collection, you can't rename or update a column instead when using withColumnRenamed() it creates a new DataFrame with updated column names, In this PySpark article, I will cover different ways to rename columns with several use cases like rename nested column, all columns, selected multiple columns with Python/PySpark examples.

Setu Nutrition

Setu Nutrition

56% OFF

Setu Lean Lite

₹479  
₹1,100

Checkout the latest products

OPEN

[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/).

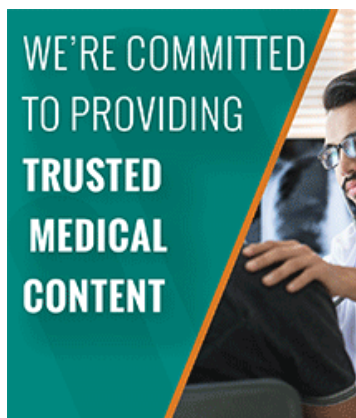
Refer to this page, If you are looking for a Spark with Scala example  
(https://sparkbyexamples.com/spark/rename-a-column-on-spark-dataframes/).

1. [PySpark withColumnRenamed – To rename DataFrame column name](#)
2. [PySpark withColumnRenamed – To rename multiple columns](#)
3. [Using StructType – To rename nested column on PySpark DataFrame](#)
4. [Using Select – To rename nested columns](#)
5. [Using withColumn – To rename nested columns](#)
6. [Using col\(\) function – To Dynamically rename all or multiple columns](#)
7. [Using toDF\(\) – To rename all or multiple columns](#)

First, let's create our data set to work with.

```
dataDF = [ (('James', '', 'Smith'),  
            (('Michael', 'Rose', ''), '2000-01-01'),  
            (('Robert', '', 'Williams'), '1990-01-01'),  
            (('Maria', 'Anne', 'Jones'), '1990-01-01'),  
            (('Jen', 'Mary', 'Brown'), '1980-01-01') ]
```

Our base schema with nested structure.



[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\(\).](#)

```
from pyspark.sql.types import StructType, StructField, StringType

schema = StructType([
    StructField('name', StringType, True),
    StructField('firstname', StringType, True),
    StructField('middlename', StringType, True),
    StructField('lastname', StringType, True),
    StructField('dob', StringType, True),
    StructField('gender', StringType, True),
    StructField('salary', IntegerType, True)
])
```

Let's create the DataFrame by using parallelize and provide the above schema.

```
from pyspark.sql import SparkSession

spark = SparkSession.builder.appName('Example').getOrCreate()

df = spark.createDataFrame(data, schema)
df.printSchema()
```

Below is our schema structure. I am not printing data here as it is not necessary for our examples. This schema has a nested structure.

```
root
 |-- name: struct (nullable = true)
 |   |-- firstname: string (nullable = true)
 |   |-- middlename: string (nullable = true)
 |   |-- lastname: string (nullable = true)
 |-- dob: string (nullable = true)
 |-- gender: string (nullable = true)
 |-- salary: integer (nullable = true)
```

## 1. PySpark withColumnRenamed – To rename DataFrame column name

PySpark has a `withColumnRenamed()` function on `DataFrame` to change a column name. This is the most straightforward approach; this function takes two parameters; the first is your existing column name and the second is the new column name you wish for.

**PySpark withColumnRenamed() Syntax:**

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

```
withColumnRenamed(existingName,
```

existingName – The existing column name you want to change



newName – New name of the column

Returns a new DataFrame with a column renamed.

### Example

```
df.withColumnRenamed("dob", "DateOf
```

The above statement changes column “dob” to “DateOfBirth” on PySpark DataFrame. Note that withColumnRenamed function returns a new DataFrame and doesn’t modify the current DataFrame.

```
root
 |-- name: struct (nullable = true)
 |   |-- firstname: string (nullable = true)
 |   |-- middlename: string (nullable = true)
 |   |-- lastname: string (nullable = true)
 |-- DateOfBirth: string (nullable = true)
 |-- gender: string (nullable = true)
 |-- salary: integer (nullable = true)
```

## 2. PySpark withColumnRenamed – To rename multiple columns

To change multiple column names, we should chain withColumnRenamed functions as shown below. You can also

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

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

Free Laptop Bag Worth ₹1999\*

Skechers - WHC Road, Dharampeth, Nagpur

Get a Free Laptop Bag! Shop for ₹6999

Store T&C\*

Nagpur

STORE INFO



## PySpark Built-In Functions

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

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

store all columns to rename in a list and loop through to rename all columns, I will leave this to you to explore.

```
df2 = df.withColumnRenamed("dob", "fname")\
        .withColumnRenamed("salary", "lname")\
df2.printSchema()
```

creates a new DataFrame “df2” by renaming dob and salary columns.

## Using PySpark StructType – To rename a nested column in a dataframe

Renaming a column name on nested data is not straight forward and we can achieve this by creating a new schema with PySpark DataFrame columns using StructType. For more details visit <https://sparkbyexamples.com/spark/sql-structtype-on-dataframe/> and see it using cast function as shown below.

```
schema2 = StructType([
    StructField("fname",StringType()),
    StructField("middlename",StringType()),
    StructField("lname",StringType()),
    StructField("dob",StringType()),
    StructField("gender",StringType()),
    StructField("salary",IntegerType())
])

df = df.select(col("name").cast(schema2), col("dob"), col("gender"), col("salary"))
df.printSchema()
```

This statement renames firstname to fname and lastname to lname within name structure.

```
root
 |-- name: struct (nullable = true)
 |    |-- fname: string (nullable = true)
 |    |-- middlename: string (nullable = true)
 |    |-- lname: string (nullable = true)
 |-- dob: string (nullable = true)
 |-- gender: string (nullable = true)
 |-- salary: integer (nullable = true)
```

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

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

[PySpark – concat\\_ws\(\).  
\(https://sparkbyexamples.com/pyspark/pyspark-convert-array-column-to-string-column/\)](https://sparkbyexamples.com/pyspark/pyspark-convert-array-column-to-string-column/)

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

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

[PySpark – regexp\\_replace\(\).  
\(https://sparkbyexamples.com/pyspark/pyspark-replace-column-values/#regexp\\_replace-replace-string-columns\)](https://sparkbyexamples.com/pyspark/pyspark-replace-column-values/#regexp_replace-replace-string-columns)

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

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

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

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

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

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

## 4. Using Select – To rename nested elements.

Let's see another way to change nested columns by transposing the structure to flat.

```
from pyspark.sql.functions import col
df.select(col("name.firstname"),
          col("name.middlename").alias("lname"),
          col("name.lastname").alias("lname"),
          col("dob"), col("gender"), col("salary"))\
    .printSchema()
```

```
root
 |-- fname: string (nullable = true)
 |-- mname: string (nullable = true)
 |-- lname: string (nullable = true)
 |-- dob: string (nullable = true)
 |-- gender: string (nullable = true)
 |-- salary: integer (nullable = true)
```

## 5. Using PySpark DataFrame with Column – To rename nested columns

When you have nested columns on PySpark DataFrame and if you want to rename it, use withColumn on a data frame object to create a new column from an existing and we will need to drop the existing column. Below example creates a “fname” column from “name.firstname” and drops the “name” column

```
from pyspark.sql.functions import col
df4 = df.withColumn("fname", col("name.firstname"))\
        .withColumn("mname", col("name.middlename"))\
        .withColumn("lname", col("name.lastname"))\
        .drop("name")
df4.printSchema()
```



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

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

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

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

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

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

[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)

## 6. Using col() function – To Dynamically rename all or multiple columns

Another way to change all column names on Dataframe is to use `col()` function.

IN progress

## 7. Using toDF() – To change all columns in a PySpark DataFrame

When we have data in a flat structure (without nested) , use `toDF()` with a new schema to change all column names.

```
newColumns = ["newCol1", "newCol2"]  
df.toDF(*newColumns).printSchema()
```

### Source code

[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)

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

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

```
import pyspark
from pyspark.sql import SparkSession
from pyspark.sql.types import StructType, StructField, StringType, IntegerType
from pyspark.sql.functions import col, lit, concat_ws, explode, array, map, to_json, from_json
```

```
spark = SparkSession.builder.appName("PySpark").getOrCreate()
```

```
dataDF = [
    (('James', '', 'Smith'), '1991-08-02', 300000),
    (('Michael', 'Rose', ''), '2000-03-01', 200000),
    (('Robert', '', 'Williams'), '1990-05-10', 250000),
    (('Maria', 'Anne', 'Jones'), '1996-01-09', 180000),
    (('Jen', 'Mary', 'Brown'), '1980-09-17', 220000)
]
```

```
schema = StructType([
    StructField('name', StringType, True),
    StructField('first_name', StringType, True),
    StructField('middle_name', StringType, True),
    StructField('last_name', StringType, True),
    StructField('dob', StringType, True),
    StructField('gender', StringType, True),
    StructField('salary', IntegerType, True)
])
```

```
df = spark.createDataFrame(dataDF, schema)
df.printSchema()
```

```
# Example 1
```

```
df.withColumnRenamed("dob", "DateOfBirth")
```

```
# Example 2
```

```
df2 = df.withColumnRenamed("dob", "DateOfBirth") \
    .withColumnRenamed("salary", "AnnualSalary")
df2.printSchema()
```

```
# Example 3
```

```
schema2 = StructType([
    StructField("fname", StringType, True),
    StructField("middlename", StringType, True),
    StructField("lname", StringType, True)
])
```

```
df.select(col("name").cast(schema2),
    col("dob"),
    col("gender"),
    col("salary")) \
    .printSchema()
```

```
# Example 4
```

```
df.select(col("name.firstname"),
    col("name.middlename").alias("middle_name"),
    col("name.lastname").alias("last_name"),
    col("dob"), col("gender"), col("salary")) \
    .printSchema()
```

```
# Example 5
```



```

df4 = df.withColumn("fname", col("first_name"))
        .withColumn("mname", col("middle_name"))
        .withColumn("lname", col("last_name"))
        .drop("name")
df4.printSchema()

#Example 7
newColumns = ["newCol1", "newCol2"]
df.toDF(*newColumns).printSchema()

# Example 6
'''
not working
old_columns = Seq("dob", "gender")
new_columns = Seq("DateOfBirth", "Gender")
columnsList = old_columns.zip(new_columns)
df5 = df4.select(columnsList:_*
df5.printSchema()
'''

```

The complete code can be downloaded from [GitHub](https://github.com/sparkbyexamples/spark-examples/blob/master/spark-sql-examples/src/main/scala/com/sparkbyexamples/spark/dataframe/RenameColumnDataFrame.scala) (<https://github.com/sparkbyexamples/spark-examples/blob/master/spark-sql-examples/src/main/scala/com/sparkbyexamples/spark/dataframe/RenameColumnDataFrame.scala>)

## Conclusion:

This article explains different ways to rename all, a single, multiple, and nested columns on PySpark DataFrame.

I hope you like this article!!

Happy Learning.

Share this:



(<https://sparkbyexamples.com/pyspark/pyspark-rename-dataframe-column/?share=facebook&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-rename-dataframe-column/?share=reddit&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-rename-dataframe-column/?share=pinterest&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-rename-dataframe-column/?share=tumblr&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-rename-dataframe-column/?share=pocket&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-rename-dataframe-column/?share=linkedin&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-rename-dataframe-column/?share=twitter&nb=1>)

**TAGS: [SPARK.CREATEDATAFRAME](#)**

**([HTTPS://SPARKBYEXAMPLES.COM/TAG/CREATEDATAFRAME/](https://sparkbyexamples.com/tag/createdataframe/)), [STRUCTTYPE](#)**

**([HTTPS://SPARKBYEXAMPLES.COM/TAG/STRUCTTYPE/](https://sparkbyexamples.com/tag/structtype/)), [TODF\(\)](#)**

**([HTTPS://SPARKBYEXAMPLES.COM/TAG/TODF/](https://sparkbyexamples.com/tag/todf/)), [WITHCOLUMN](#)**

**([HTTPS://SPARKBYEXAMPLES.COM/TAG/WITHCOLUMN/](https://sparkbyexamples.com/tag/withcolumn/)), [WITHCOLUMNRENAMED](#)**

**([HTTPS://SPARKBYEXAMPLES.COM/TAG/WITHCOLUMNRENAMED/](https://sparkbyexamples.com/tag/withcolumnrenamed/))**



**NNK**

**([Https://Sparkbyexamples.Com/Author/Admin/](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/>)

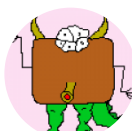
#### ➤ THIS POST HAS 14 COMMENTS



**Anonymous**

10 APR 2021 [REPLY](#)

Really helped me. Thanks a lot



**Anonymous**

24 FEB 2021 [REPLY](#)

Veri good tutorials , keep it up



**NNK** 24 FEB 2021 [REPLY](#)

Thanks for liking  
PySpark  
withColumnRenamed  
Example

---



**GonzaloGF**

30 JAN 2021 [REPLY](#)

hi NNK I would like to tell  
you that you have done a  
great job with this  
website, I thank you from  
the bottom of my heart  
because I am learning a  
lot and in a simple way,  
since you explain  
everything very well.

---



**NNK** 31 JAN 2021 [REPLY](#)

Thank you for your  
comments. It means a  
lot to me and  
motivates me to write  
more.

---



**Avula Sreenu**

27 NOV 2020 [REPLY](#)

Thank you so much bro,  
the way you have to show  
simple easy to  
understandable tutorial

---



**NNK** 27 NOV 2020 [REPLY](#)

Thanks Sreenu. Glad  
you like the articles.

---



**Janusz**

12 NOV 2020 [REPLY](#)

NNK, Example 7 is working when you use asterisk:  
newColumns =  
["newCol1","newCol2","newCol3","newCol4"]  
new\_df =  
df.toDF(\*newColumns).printSchema()  
Please update the article 😊



**NNK** 12 NOV 2020 [REPLY](#)

Thanks Janusz. I've added this statement.  
appreciate your comment. Hope articles are helpful.



**Anonymous**

19 OCT 2020 [REPLY](#)

Thank you so much



**Anonymous**

9 SEP 2020 [REPLY](#)

Thank you so much for the entire article. You have explained it in such simple words that it is so easy to understand. Keep up the good work! 😊



**NNK** 9 SEP 2020 [REPLY](#)

Thank you for your wonderful words.  
Happy helping the Spark community.



**Manjunath**

18 AUG 2020 [REPLY](#)

Hi NNK(I dont have your full name, I m sorry)  
Thanks for your postings,

I appreciate them heart  
fully.  
Learning Pyspark on my  
own.



**Anonymous**

18 AUG 2020 [REPLY](#)

You are simply Awesome.

[Leave a Reply](#)

#### Categories

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

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

#### Recent Posts

Spark regexp\_replace() – Replace  
String Value  
([https://sparkbyexamples.com/spark/spark-regexp\\_replace-replace-string-value/](https://sparkbyexamples.com/spark/spark-regexp_replace-replace-string-value/))

#### 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 ..

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

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

ples)    s/)    8193/)    les/)