

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

## PySpark Tutorial

[PySpark Tutorial For Beginners \(https://sparkbyexamples.com/pyspark-tutorial/\)](https://sparkbyexamples.com/pyspark-tutorial/)

[PySpark – Features \(https://sparkbyexamples.com/pyspark-tutorial/#features\)](https://sparkbyexamples.com/pyspark-tutorial/#features)

[PySpark – Advantages \(https://sparkbyexamples.com/pyspark-tutorial/#advantages\)](https://sparkbyexamples.com/pyspark-tutorial/#advantages)

[PySpark – Modules & Packages \(https://sparkbyexamples.com/pyspark-tutorial/#modules-packages\)](https://sparkbyexamples.com/pyspark-tutorial/#modules-packages)

[PySpark – Cluster Managers \(https://sparkbyexamples.com/pyspark-tutorial/#cluster-manager\)](https://sparkbyexamples.com/pyspark-tutorial/#cluster-manager)

[PySpark – Install on Windows \(https://sparkbyexamples.com/pyspark-tutorial/#pyspark-installation\)](https://sparkbyexamples.com/pyspark-tutorial/#pyspark-installation)

[PySpark – Web/Application UI \(https://sparkbyexamples.com/spark/spark-web-ui-understanding/\)](https://sparkbyexamples.com/spark/spark-web-ui-understanding/)

[PySpark – SparkSession \(https://sparkbyexamples.com/pyspark/pyspark-what-is-sparksession/\)](https://sparkbyexamples.com/pyspark/pyspark-what-is-sparksession/)

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

[PySpark – Parallelize \(https://sparkbyexamples.com/pyspark/pyspark-parallelize-create-rdd/\)](https://sparkbyexamples.com/pyspark/pyspark-parallelize-create-rdd/)

[PySpark – repartition\(\) vs coalesce\(\) \(https://sparkbyexamples.com/pyspark/pyspark-repartition-vs-coalesce/\)](https://sparkbyexamples.com/pyspark/pyspark-repartition-vs-coalesce/)

[PySpark – Broadcast Variables \(https://sparkbyexamples.com/pyspark/pyspark-broadcast-variables/\)](https://sparkbyexamples.com/pyspark/pyspark-broadcast-variables/)

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

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

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

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

[FAQ's In PySpark select\(\) function is used to select single, multiple, column by index, all columns from the list and the nested columns from a DataFrame, PySpark select\(\) is a transformation function hence it returns a new DataFrame with the selected columns.](https://sparkbyexamples.com/)

[More \(https://sparkbyexamples.com/\)](https://sparkbyexamples.com/)

## PySpark Select Columns From DataFrame



- [Select a Single & Multiple Columns from PySpark](#)
- [Select All Columns From List](#)
- [Select Columns By Index](#)
- [Select a Nested Column](#)
- [Other Ways to Select Columns](#)

## #1 Screen Recorder

Make Amazing Videos  
Screencasts. Shop Onl  
Now for the Latest...

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/)

First, let's [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/).

```
import pyspark
from pyspark.sql import SparkSession

spark = SparkSession.builder.appName("PySpark").getOrCreate()
data = [
    ("James", "Smith", "USA", "NY"),
    ("Michael", "Rose", "USA", "NY"),
    ("Robert", "Williams", "USA", "NY"),
    ("Maria", "Jones", "USA", "FL")
]
columns = ["firstname", "lastname", "country", "state"]
df = spark.createDataFrame(data, columns)
df.show(truncate=False)
```

## 1. Select Single & Multiple Columns From PySpark

You can select the single or multiple columns of the DataFrame by passing the column names you wanted to select to the `select()` function. Since DataFrame is immutable, this creates a new DataFrame with selected columns. `show()` function is used to show the DataFrame contents.

Below are ways to select single, multiple or all columns.

```
df.select("firstname", "lastname")
df.select(df.firstname, df.lastname)
df.select(df["firstname"], df["lastname"])

#By using col() function
from pyspark.sql.functions import col
df.select(col("firstname"), col("lastname"))

#Select columns by regular expression
df.select(df.colRegex("`^.*name`"))
```

[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\(\).](#)

## 2. Select All Columns

### From List

Sometimes you may need to select all DataFrame columns from a Python list. In the below example, we have all columns in the columns list object.



```
# Select All columns from List
df.select(*columns).show()

# Select All columns
df.select([col for col in df.columns]).show()
df.select("*").show()
```

## 3. Select Columns by Index

Using a python list features, you can select the columns by index.

```
#Selects first 3 columns and to
df.select(df.columns[:3]).show()

#Selects columns 2 to 4 and to
df.select(df.columns[2:4]).show()
```

## 4. Select Nested Struct Columns from PySpark

If you have a nested struct (StructType) column on PySpark DataFrame, you need to use an explicit column qualifier in order to select. If you are new to

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

PySpark and you have not learned StructType yet, I would recommend skipping the rest of the section or first [Understand PySpark StructType  
\(https://sparkbyexamples.com/pyspark/pyspark-structtype-and-structfield/\)](https://sparkbyexamples.com/pyspark/pyspark-structtype-and-structfield/) before you proceed.

First, let's create a new DataFrame with a struct type.

```
data = [
    ("James", None, "Smith")
    ("Anna", "Rose", ""), "NY"
    ("Julia", "", "Williams")
    ("Maria", "Anne", "Jones")
    ("Jen", "Mary", "Brown")
    ("Mike", "Mary", "William")
]

from pyspark.sql.types import StructType, StructField
schema = StructType([
    StructField('name', StringType, True),
    StructField('firstname', StringType, True),
    StructField('middlename', StringType, True),
    StructField('lastname', StringType, True),
    StructField('state', StringType, True),
    StructField('gender', StringType, True)
])

df2 = spark.createDataFrame(data, schema)
df2.printSchema()
df2.show(truncate=False) # show all columns
```

Yields below schema output. If you notice the column name is a struct type which consists of columns firstname, middlename, 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/>).

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

```
-----+-----+
name      |state|gender|
-----+-----+
[James,, Smith] |OH   |M      |
[Anna, Rose, ]  |NY   |F      |
[Julia, , Williams] |OH   |F      |
[Maria, Anne, Jones] |NY   |M      |
[Jen, Mary, Brown] |NY   |M      |
[Mike, Mary, Williams]|OH   |M      |
-----+-----+
```

, let's select struct column.

```
f2.select("name").show(truncate=false)
```

; returns struct column name as is.

```
-----+-----+
name      |
-----+-----+
[James,, Smith] |
[Anna, Rose, ]  |
[Julia, , Williams] |
[Maria, Anne, Jones] |
[Jen, Mary, Brown] |
[Mike, Mary, Williams]|
-----+-----+
```



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

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

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

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

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

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

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

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

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

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

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

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

In order to select the specific column from a nested struct, you need to explicitly qualify the nested struct column name.

```
df2.select("name.firstname", "name.lastname")
```

This outputs firstname and lastname from the name struct column.

```
+-----+-----+
|firstname|lastname|
+-----+-----+
|James    |Smith    |
|Anna     |         |
|Julia    |Williams |
|Maria    |Jones    |
|Jen      |Brown    |
|Mike     |Williams |
+-----+-----+
```

In order to get all columns from struct column.

```
df2.select("name.*").show(truncate=false)
```

This yields below output.

```
+-----+-----+-----+
|firstname|middlename|lastname|
+-----+-----+-----+
|James    |null      |Smith    |
|Anna     |Rose      |         |
|Julia    |          |Williams |
|Maria    |Anne      |Jones    |
|Jen      |Mary      |Brown    |
|Mike     |Mary      |Williams |
+-----+-----+-----+
```

## 5. Complete Example

[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](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](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](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](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>).

```
import pyspark
from pyspark.sql import SparkSession

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

data = [
    ("James", "Smith", "USA", "NY"),
    ("Michael", "Rose", "USA", "NY"),
    ("Robert", "Williams", "USA", "CA"),
    ("Maria", "Jones", "USA", "FL")
]

columns = ["firstname", "lastname", "state", "city"]
df = spark.createDataFrame(data, columns)
df.show(truncate=False)

df.select("firstname").show()

df.select("firstname", "lastname").show()

# Using Dataframe object name
df.select(df.firstname, df.lastname).show()

# Using col function
from pyspark.sql.functions import col
df.select(col("firstname"), col("lastname")).show()

data = [
    (("James", None, "Smith"), ("Anna", "Rose", ""), "NY"),
    (("Julia", "", "Williams"), ("Maria", "Anne", "Jones"), "CA"),
    (("Jen", "Mary", "Brown"), ("Mike", "Mary", "Williams"), "FL")
]

from pyspark.sql.types import StructType, StructField
schema = StructType([
    StructField('name', StructType([
        StructField('firstname', StringType),
        StructField('middlename', StringType),
        StructField('lastname', StringType)
    ])),
    StructField('state', StringType),
    StructField('gender', StringType)
])

df2 = spark.createDataFrame(data, schema)
df2.printSchema()
df2.show(truncate=False) # show all columns

df2.select("name").show(truncate=False)
df2.select("name.firstname", "name.lastname").show(truncate=False)
df2.select("name.*").show(truncate=False)
```

---

[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)

---

This example is also available at [PySpark github project  
\(https://github.com/spark-examples/pyspark-examples/blob/master/pyspark-select-columns.py\)](https://github.com/spark-examples/pyspark-examples/blob/master/pyspark-select-columns.py).

## 6. Conclusion

In this article, you have learned select() is a transformation function of the DataFrame and is used to select single, multiple columns, select all columns from the list, select by index, and finally select nested struct columns, you have also learned how to select nested elements from the DataFrame.

## Related Articles

- [How to Replace Column Values in PySpark DataFrame  
\(https://sparkbyexamples.com/pyspark/pyspark-replace-column-values/\)](https://sparkbyexamples.com/pyspark/pyspark-replace-column-values/)
- [How to Retrieve DataType & Column Names of PySpark DataFrame  
\(https://sparkbyexamples.com/pyspark/pyspark-find-datatype-column-names-of-dataframe/\)](https://sparkbyexamples.com/pyspark/pyspark-find-datatype-column-names-of-dataframe/)

Happy Learning !!

---

Share this:



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



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



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



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



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



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





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

**TAGS:** [SELECT\(\)](#).

([HTTPS://SPARKBYEXAMPLES.COM/TAG/SELECT/](https://sparkbyexamples.com/tag/select/)).

, [STRUCT](#)

([HTTPS://SPARKBYEXAMPLES.COM/TAG/STRUCT/](https://sparkbyexamples.com/tag/struct/)).

, [STRUCTTYPE](#)

([HTTPS://SPARKBYEXAMPLES.COM/TAG/STRUCTTYPE/](https://sparkbyexamples.com/tag/structtype/)).



**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 5 COMMENTS



**Shaun Ryan**

(<https://semanticmi.wordpress.com/>).

2 APR 2021 [REPLY](#)

df["firstname"] returns a column object of firstname. So in effect is equivalent to col("firstname")



**Nid**

20 FEB 2021 [REPLY](#)

How to select first N column in a data frame and make it into another data frame ?

I have a DF with 180 columns and I want to create another DF with first 100 column with out implicitly mention the column name



**NNK**

14 MAR 2021 [REPLY](#)

Can you try below?

```
df.select(df.columns[:100]).show(3)
```





**lex**

23 DEC 2020 [REPLY](#)

not sure if this an apache  
spark thing or just a  
databricks thing but  
`select(df["firstname"])`  
works also



**NNK** 25 DEC 2020 [REPLY](#)

You are right. You can  
also use  
`select(df["firstname"])`

---

**Leave a Reply**

<div>Apache Hadoop</div> <div>(<a href="https://sparkbyexamples.com/category/hadoop/">https://sparkbyexamples.com/category/hadoop/</a>)</div>
<div>Apache Spark</div> <div>(<a href="https://sparkbyexamples.com/category/spark/">https://sparkbyexamples.com/category/spark/</a>)</div>
<div>Apache Spark Streaming</div> <div>(<a href="https://sparkbyexamples.com/category/spark/apache-spark-streaming/">https://sparkbyexamples.com/category/spark/apache-spark-streaming/</a>)</div>
<div>Apache Kafka</div> <div>(<a href="https://sparkbyexamples.com/category/kafka/">https://sparkbyexamples.com/category/kafka/</a>)</div>
<div>Apache HBase</div> <div>(<a href="https://sparkbyexamples.com/category/hbase/">https://sparkbyexamples.com/category/hbase/</a>)</div>
<div>Apache Cassandra</div> <div>(<a href="https://sparkbyexamples.com/category/cassandra/">https://sparkbyexamples.com/category/cassandra/</a>)</div>
<div>Snowflake Database</div> <div>(<a href="https://sparkbyexamples.com/category/snowflake/">https://sparkbyexamples.com/category/snowflake/</a>)</div>
<div>H2O Sparkling Water</div> <div>(<a href="https://sparkbyexamples.com/category/h2o-sparkling-water/">https://sparkbyexamples.com/category/h2o-sparkling-water/</a>)</div>
<div>PySpark</div> <div>(<a href="https://sparkbyexamples.com/category/pyspark/">https://sparkbyexamples.com/category/pyspark/</a>)</div>

<div>Spark regexp_replace() – Replace String Value</div> <div>(<a href="https://sparkbyexamples.com/spark/spark-regexp_replace-replace-string-value/">https://sparkbyexamples.com/spark/spark-regexp_replace-replace-string-value/</a>)</div>
<div>How to Run a PySpark Script from Python?</div> <div>(<a href="https://sparkbyexamples.com/pyspark/run-pyspark-script-from-python-subprocess/">https://sparkbyexamples.com/pyspark/run-pyspark-script-from-python-subprocess/</a>)</div>
<div>Spark SQL like() Using Wildcard Example</div> <div>(<a href="https://sparkbyexamples.com/spark/spark-sql-like-using-wildcard-example/">https://sparkbyexamples.com/spark/spark-sql-like-using-wildcard-example/</a>)</div>
<div>Spark isin() &amp; IS NOT IN Operator Example</div> <div>(<a href="https://sparkbyexamples.com/spark/spark-isin-is-not-in-operator-example/">https://sparkbyexamples.com/spark/spark-isin-is-not-in-operator-example/</a>)</div>
<div>Spark – Get Size/Length of Array &amp; Map Column</div> <div>(<a href="https://sparkbyexamples.com/spark/spark-get-size-length-of-array-map-column/">https://sparkbyexamples.com/spark/spark-get-size-length-of-array-map-column/</a>)</div>
<div>Spark Using Length/Size Of a DataFrame Column</div> <div>(<a href="https://sparkbyexamples.com/spark/spark-using-length-size-of-a-dataframe-column/">https://sparkbyexamples.com/spark/spark-using-length-size-of-a-dataframe-column/</a>)</div>
<div>Spark rlike() Working with Regex Matching Examples</div> <div>(<a href="https://sparkbyexamples.com/spark/spark-rlike-regex-matching-examples/">https://sparkbyexamples.com/spark/spark-rlike-regex-matching-examples/</a>)</div>
<div>Spark Check String Column Has Numeric Values</div> <div>(<a href="https://sparkbyexamples.com/spark/spark-check-string-column-has-numeric-values/">https://sparkbyexamples.com/spark/spark-check-string-column-has-numeric-values/</a>)</div>
<div>Spark Check Column Data Type is Integer or String</div> <div>(<a href="https://sparkbyexamples.com/spark/spark-check-column-data-type-is-integer-or-string/">https://sparkbyexamples.com/spark/spark-check-column-data-type-is-integer-or-string/</a>)</div>

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

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

facebook

([https://www.](https://www.linkedin.com/company/sparkbyexamples/)

linkedin

([https://twitter.](https://twitter.com/sparkbyexamples/)

ok.co

([https://github](https://github.com/sparkbyexamples/sparkbyexamples)

n.com/

([https://medium](https://medium.com/sparkbyexamples)

m/spar

([https://www.](https://www.youtube.com/watch?v=b860a819378)

in/n-

([https://www.](https://www.blogger.com/2019/08/19/sparkbyexamples-819378.html)

b.com/

([https://www.](https://www.youtube.com/watch?v=b860a819378)

sparkb

([https://www.](https://www.youtube.com/watch?v=b860a819378)

kbyex

([https://www.](https://www.youtube.com/watch?v=b860a819378)

nk-

([https://www.](https://www.youtube.com/watch?v=b860a819378)

spark-

([https://www.](https://www.youtube.com/watch?v=b860a819378)

yexam

([https://www.](https://www.youtube.com/watch?v=b860a819378)

ample

([https://www.](https://www.youtube.com/watch?v=b860a819378)

b860a

([https://www.](https://www.youtube.com/watch?v=b860a819378)

examp

([https://www.](https://www.youtube.com/watch?v=b860a819378)

ples)

([https://www.](https://www.youtube.com/watch?v=b860a819378)

s/)

([https://www.](https://www.youtube.com/watch?v=b860a819378)

8193/)

([https://www.](https://www.youtube.com/watch?v=b860a819378)

les/)

