

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

## PySpark Tutorial

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

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

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

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

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

[PySpark – Install on Windows](https://sparkbyexamples.com/py-spark-tutorial/#pyspark-installation)  
(https://sparkbyexamples.com/py-spark-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/py-spark/pyspark-what-is-sparksession/)  
(https://sparkbyexamples.com/py-spark/pyspark-what-is-sparksession/)

[PySpark – RDD](https://sparkbyexamples.com/py-spark-rdd)  
(https://sparkbyexamples.com/py-spark-rdd)

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

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

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

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

## PySpark DataFrame

[PySpark – Create a DataFrame](https://sparkbyexamples.com/py-spark/create-a-dataframe/)  
(https://sparkbyexamples.com/py-spark/create-a-dataframe/)

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

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

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

[Kafka \\_tutorial/\)](https://sparkbyexamples.com/apache-kafka-tutorial/)  
[- Prabha](https://sparkbyexamples.com/author/prabha/)  
[PySpark tutorials with examples/\)](https://sparkbyexamples.com/category/pyspark/)

[FAQ's In PySpark the substring\(\) function is](https://sparkbyexamples.com/pyspark-questions/)

[More ▾ https://sparkbyexamples.com/pyspark-questions/\)](https://sparkbyexamples.com/pyspark-questions/)  
used to extract the substring from a DataFrame string column by providing the position and length of the string you wanted to extract.



## Pyspark – Get substring() from a column

In PySpark, the `substring()` function is used to extract the substring from a DataFrame string column by providing the position and length of the string you wanted to extract.

Succeed online

in.godaddy.com

In this tutorial, I have explained with an example of getting substring of a column using `substring()` from `pyspark.sql.functions` and using `substr()` from `pyspark.sql.Column` type.

## Using SQL function substring()

[spark/different-ways-to-create-dataframe-in-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/>)

[PySpark – Convert 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/>)

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

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

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

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

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

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

[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-sort/>)

Using the `substring()` function of `pyspark.sql.functions` module we can extract a substring or slice of a string from the DataFrame column by providing the position and length of the string you wanted to slice.

```
substring(str, pos, len)
```

Note: Please note that the position is not zero based, but 1 based index.

Below is an example of Pyspark `substring()` using `withColumn()`  
(<https://sparkbyexamples.com/pyspark/pyspark-withcolumn/>).



```
data = [(1, "20200828"), (2, "20180525")]
columns = ["id", "date"]
df = spark.createDataFrame(data, columns)
df.withColumn('year', substring('date', 1, 4))
   .withColumn('month', substring('date', 5, 2))
   .withColumn('day', substring('date', 7, 2))
df.printSchema()
df.show(truncate=False)
```

In above example, we have created a DataFrame with two columns, id and date. Here date is in the form "year month day". Here I have used `substring()` on date column to return sub strings of date as year, month, day respectively. Below is the output.

```
+---+-----+-----+-----+---+
|id |date   |year|month|day|
+---+-----+-----+-----+---+
|1  |20200828|2020|08  |28 |
|2  |20180525|2018|05  |25 |
+---+-----+-----+-----+---+
```



[spark/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\(\).](#)  
(<https://sparkbyexamples.com/pyspark/pyspark-sampling-example/>)

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

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

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

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

## 2. Using substring() with select()

In Pyspark we can get substring() of a column using [select](#) (<https://sparkbyexamples.com/pyspark/select-columns-from-pyspark-dataframe/>).  
Above example can be written as below.

```
df.select('date', substring('date',
                             substring('date',
                             substring('date',
```

## 3.Using substring() with selectExpr()

Sample example using **selectExpr** to get sub string of column(date) as year,month,day. Below is the code that gives same output as above.

```
df.selectExpr('date', 'substring(da
                        'substring(date,
                        'substring(date, 7
```

## 4. Using substr() from Column type

Below is the example of getting substring using `substr()` function from `pyspark.sql.Column` type in Pyspark.

```
df3=df.withColumn('year', col('date')
        .withColumn('month',col('date').s
        .withColumn('day', col('date').su
```

The above example gives output same as the above mentioned examples.

PRICE DROP	PRICE DROP
PRICE DROP	PRICE DROP
PRICE DROP	PRICE DROP
PRICE DROP	PRICE DF

Come Find Some Nice  
Khankudi

[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/](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/](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/](https://sparkbyexamples.com/pyspark/pyspark-read-csv-file-into-dataframe/)).

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

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

## PySpark Built-In Functions

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

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

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

Succeed online

in.godaddy.com

## Complete Example of PySpark substring()

```
import pyspark
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, substring

spark=SparkSession.builder.appName('PySpark substring()').getOrCreate()

data = [(1, "20200828"), (2, "20180525")]
columns=["id", "date"]
df=spark.createDataFrame(data, columns)

#Using SQL function substring()
df.withColumn('year', substring(col('date'), 1, 4)) \
  .withColumn('month', substring(col('date'), 5, 2)) \
  .withColumn('day', substring(col('date'), 7, 2)) \
  .printSchema()
df.show(truncate=False)

#Using select
df1=df.select('date', substring('date', 1, 4) as year,
              substring('date', 5, 2) as month,
              substring('date', 7, 2) as day)

#Using with selectExpr
df2=df.selectExpr('date', 'substring(date, 1, 4) as year,
                  'substring(date, 5, 2) as month,
                  'substring(date, 7, 2) as day')

#Using substr from Column type
df3=df.withColumn('year', col('date').substr(1, 4)) \
      .withColumn('month', col('date').substr(5, 2)) \
      .withColumn('day', col('date').substr(7, 2))

df3.show()
```

## Conclusion

In this session, we have learned different ways of getting substring of a column in PySpark DataFarme. I hope you liked it! Keep practicing. And do comment in the comment section for any kind of questions!!

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

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

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

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

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


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


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


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


[PySpark – collect\\_list\(\)](#)  
(<https://sparkbyexamples.com/pyspark>)


## Share this:


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


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

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

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

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

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

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

## TAGS: [SELECTEXPR](#)

([HTTPS://SPARKBYEXAMPLES.COM/TAG/SELECTEXPR/](https://sparkbyexamples.com/tag/selectexpr/)),  
[SUBSTR\(\)](#),  
([HTTPS://SPARKBYEXAMPLES.COM/TAG/SUBSTR/](https://sparkbyexamples.com/tag/substr/)),  
[SUBSTRING\(\)](#),  
([HTTPS://SPARKBYEXAMPLES.COM/TAG/SUBSTRING/](https://sparkbyexamples.com/tag/substring/))

## THIS POST HAS 2 COMMENTS



**Giri**

27 NOV 2020 [REPLY](#)

Good work



**NNK**

28 NOV 2020 [REPLY](#)

Thanks Giri.

## Leave a Reply

[ark/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/>).

---

[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>).

---

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

---

[PySpark – sum\(\),.avg\(\)](#)  
(<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>).

---

[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>).

---

[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](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](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](https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#schema_of_json))

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/h2o-sparkling-water/>)

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

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/))

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  
(<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/>)

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

(https: //www. //www. (https: (https: facebo linkedi (https: //twitte ok.co n.com/ //githu r.com/ m/spar in/n- b.com/ sparkb kbyex nk- spark-yexam ample b860a examp ples) s/) 8193/) les/)

