Select() function is used to select single, multiple columns by index, all columns from the list and the nested columns from a DataFrame.

from pyspark.sql import \*  
from pyspark.sql.functions import \*  
  
spark = SparkSession.builder.appName('select()').getOrCreate()  
  
myData = [(1,'Harsha','male',2000,'IT'),  
 (2,'Mokshit','male',3000,'HR'),  
 (3, 'Harika', 'female', 3500, 'payroll'),  
 (4, 'Kiran', 'male', 4000, 'HR'),  
 (5, 'venkat', 'male', 5000, 'IT'),  
 (6, 'Rama', 'female', 4500, 'payroll')  
 ]  
  
mySchema = ['id','name','gender','salary','dept']  
  
df = spark.createDataFrame(myData,mySchema)  
  
df.select('\*').show()  
df.select(df.id,df.name).show()  
df.select('id','name').show()  
df.select(df['id'],df['name']).show()  
df.select(col('id'),col('name')).show()  
df.select(['id','name']).show()

**Output:**

+---+-------+------+------+-------+

| id| name|gender|salary| dept|

+---+-------+------+------+-------+

| 1| Harsha| male| 2000| IT|

| 2|Mokshit| male| 3000| HR|

| 3| Harika|female| 3500|payroll|

| 4| Kiran| male| 4000| HR|

| 5| venkat| male| 5000| IT|

| 6| Rama|female| 4500|payroll|

+---+-------+------+------+-------+

+---+-------+

| id| name|

+---+-------+

| 1| Harsha|

| 2|Mokshit|

| 3| Harika|

| 4| Kiran|

| 5| venkat|

| 6| Rama|

+---+-------+

+---+-------+

| id| name|

+---+-------+

| 1| Harsha|

| 2|Mokshit|

| 3| Harika|

| 4| Kiran|

| 5| venkat|

| 6| Rama|

+---+-------+

+---+-------+

| id| name|

+---+-------+

| 1| Harsha|

| 2|Mokshit|

| 3| Harika|

| 4| Kiran|

| 5| venkat|

| 6| Rama|

+---+-------+

+---+-------+

| id| name|

+---+-------+

| 1| Harsha|

| 2|Mokshit|

| 3| Harika|

| 4| Kiran|

| 5| venkat|

| 6| Rama|

+---+-------+

+---+-------+

| id| name|

+---+-------+

| 1| Harsha|

| 2|Mokshit|

| 3| Harika|

| 4| Kiran|

| 5| venkat|

| 6| Rama|

+---+-------+