

```
In [58]: import findspark
findspark.init()
```

```
In [59]: import pyspark
import random
```

```
In [69]: from pyspark.sql import SparkSession
from pyspark import SparkContext, SparkConf

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

sc = spark.sparkContext
```

```
In [85]: df = spark.read.csv('airports.csv',inferSchema=True,header=True)
groupBy=df.select('country','city').groupBy('country').count()
groupBy.toPandas()
groupBy.show()

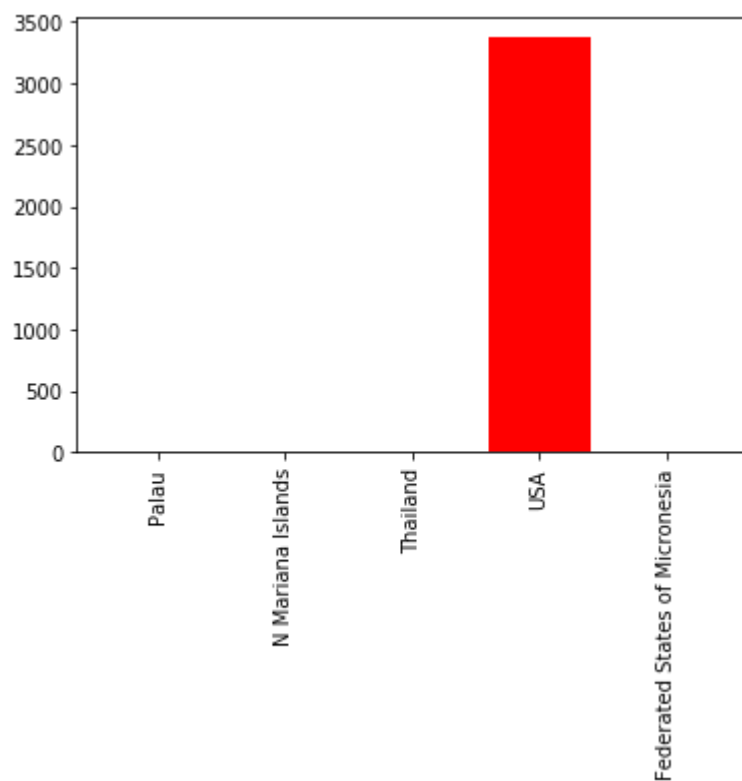
#some_df=df.select('country','city').groupBy('country')
#type(some_df)
pandas_df = groupBy.toPandas()

#df.show()
```

```
+-----+-----+
|          country|count|
+-----+-----+
|          Palau|    1|
| N Mariana Islands|    1|
|          Thailand|    1|
|          USA|  3372|
|Federated States ...|    1|
+-----+-----+
```

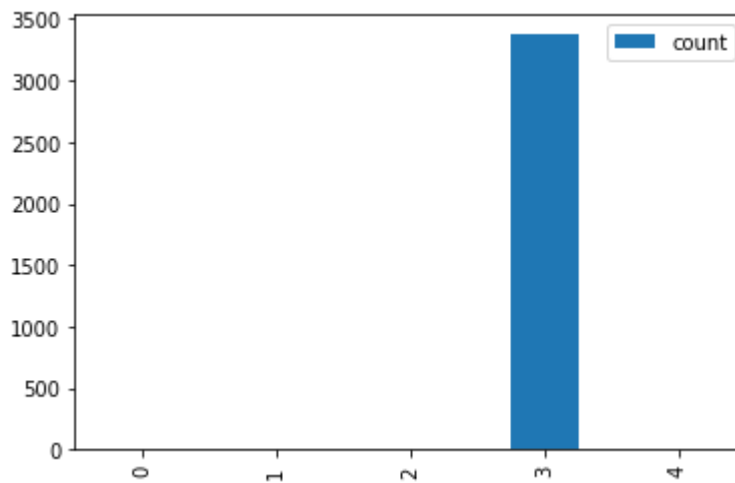
```
In [96]: import matplotlib.pyplot as plt

plt.bar(pandas_df['country'], pandas_df['count'],color='red')
plt.xticks(rotation='vertical')
plt.show()
```



```
In [97]: pandas_df.plot(kind='bar')
```

```
Out[97]: <matplotlib.axes._subplots.AxesSubplot at 0xad23940>
```



```
In [61]: data = [1, 2, 3, 4, 5]
distData2 = sc.parallelize(data).collect()
print(distData2)
```

```
[1, 2, 3, 4, 5]
```

```
In [62]: #df = spark.sql('show databases')
df = spark.sql("select 'alaa' union all select 'tt'")
df.show()
```

```
+-----+
|alaa|
+-----+
|alaa|
|  tt|
+-----+
```

```
In [68]: if sc==sc:
          sc.stop()
if spark==spark:
          spark.stop()
```

```
In [ ]:
```

