

```
In [1]: import findspark
findspark.init()
```

```
In [2]: from pyspark.sql import SparkSession
spark=SparkSession.builder.master("local[1]").appName("SparkByExamples.com").getOrCreate()
sparkContext=spark.sparkContext
```

```
In [3]: dataset = spark.read.csv('../data/carwood.csv',inferSchema = True, header = True)
```

```
In [4]: dataset.show(5)
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|      f1|      f2|      f3|      f4|      f5|      f6|      f7|      f8|      f9|     f10|     f11|     f12|     f13|     f14|     f15|     f16|
f17|    f18|    f19|    f20|    f21|    f22|    f23|    f24|    f25|    f26|    f27|    f28|    f29|    f30|    f31|    f32|    f
33|    f34|    f35|    f36|    f37|    f38|    f39|    f40|    f41|    f42|    f43|    f44|    f45|    f46|    f47|    f48|    f4
9|    f50|    f51|    f52|    f53|    f54|    f55|    f56|    f57|    f58|    f59|    f60|    f61|    f62|    f63|    f64|    f65
|    f66|    f67|label|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|170.39|167.28|143.44|124.67|139.01|125.83|144.33|151.26|175.51|171.31| 161.9|146.92| 141.8|140.91| 132.8|128.48|
170.83|161.06|169.61|168.07|154.88|149.33|152.91|152.61|137.09|169.59|183.01|180.44|164.55|152.63|157.77|165.07|1
49.52|142.73|173.85|182.23|174.58|163.28|164.01| 169.2|148.48|134.31|174.07|191.33|157.51|168.66|156.89|170.86|16
2.24|184.84|167.02|123.67|140.54|153.69|147.57|144.65|162.24|172.96|169.67|157.51|161.06|133.23|124.41|138.44|142
.93|137.13|134.44|    0|
|169.75|190.96|175.53|138.27|137.47|139.23|133.23|130.25|147.73|163.93|167.36|171.52|155.54|139.34|151.95| 149.3|
173.37| 141.0|153.57|128.45|159.93|165.33|147.94|143.74|140.88|182.53|184.08|148.09|125.26|139.67| 138.7|132.86|
141.5| 145.4|164.58|170.71|127.83|133.99| 141.2|152.87|142.36|148.26|162.54|156.55|153.39|137.99|129.64|137.59|15
5.39|147.92|152.14|162.48|168.72|161.14|147.87|141.02|155.39|139.58|141.58|153.39| 141.0|148.43|168.12| 169.9|165
.64|166.86|137.69|    0|
|153.69|153.68|144.02|158.73|178.87|157.04|152.92|147.52|142.87|165.26|160.39|137.86|149.62|153.43| 152.6|162.85|
146.35|167.11|134.27|126.81|136.28| 158.4|171.58| 161.6|162.91|143.14|129.73|125.97|151.09|177.89| 169.2|160.65|1
56.86|135.66|126.92|131.22|154.37|158.59|158.52|155.25|154.33|130.18|127.66|148.94|155.37|163.53| 139.7|143.79|14
1.61|166.88| 164.6|149.58|139.56|154.74|173.01|155.18|141.61|155.19|170.51|155.37|167.11|146.89|141.01|159.43|169
.68|163.24|165.17|    0|
|131.69|151.56|151.05| 134.0|151.18|175.53|171.34|159.77|151.95| 146.1|148.53|140.28|138.16|145.44| 150.4|158.18|
163.8|152.43|171.49| 150.2|131.28|157.18|157.04|151.13|151.66|143.32|157.23|152.91|134.75|154.65|171.65| 160.3|15
7.73|143.67|145.87|151.73|147.09|151.21|157.96| 148.5| 156.6|147.45|153.18|156.58|157.83|147.98|143.47|142.83|138
.08|147.91|148.05|145.66| 156.5|167.52|151.33|129.51|138.08|164.25|155.82|157.83|152.43|150.82|146.58|128.85|140.
76|177.35|174.61|    0|
|162.85|158.88|132.27|138.41|143.98| 159.3|177.26|180.58|159.34|164.66|138.04|132.76|157.88|165.58|173.64| 163.5|
127.97|167.31|141.39|147.02|137.52|135.46|146.41|159.09|164.53|148.45|130.76|136.22|144.86|127.38|137.09|159.08|1
53.25|182.49|187.75|139.37|117.12|124.07| 134.6|144.85|132.64|170.87|188.26|173.32|135.74|127.15|131.69| 127.8|14
9.46|114.16|100.11|154.02| 175.3|175.46|144.39|142.47|149.46| 132.8|130.96|135.74|167.31|188.21|179.52| 146.2|153
.73|152.12|146.58|    0|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 5 rows
```

```
In [5]: print('Shape: Rows:', dataset.count(),',','Columns:',len(dataset.columns))
shape = [dataset.count(),len(dataset.columns)]
```

Shape: Rows: 2048 , Columns: 68

```
In [6]: dataset.printSchema()
```

```
root
|-- f1: double (nullable = true)
|-- f2: double (nullable = true)
|-- f3: double (nullable = true)
|-- f4: double (nullable = true)
|-- f5: double (nullable = true)
|-- f6: double (nullable = true)
|-- f7: double (nullable = true)
|-- f8: double (nullable = true)
|-- f9: double (nullable = true)
```

```

|-- f10: double (nullable = true)
|-- f11: double (nullable = true)
|-- f12: double (nullable = true)
|-- f13: double (nullable = true)
|-- f14: double (nullable = true)
|-- f15: double (nullable = true)
|-- f16: double (nullable = true)
|-- f17: double (nullable = true)
|-- f18: double (nullable = true)
|-- f19: double (nullable = true)
|-- f20: double (nullable = true)
|-- f21: double (nullable = true)
|-- f22: double (nullable = true)
|-- f23: double (nullable = true)
|-- f24: double (nullable = true)
|-- f25: double (nullable = true)
|-- f26: double (nullable = true)
|-- f27: double (nullable = true)
|-- f28: double (nullable = true)
|-- f29: double (nullable = true)
|-- f30: double (nullable = true)
|-- f31: double (nullable = true)
|-- f32: double (nullable = true)
|-- f33: double (nullable = true)
|-- f34: double (nullable = true)
|-- f35: double (nullable = true)
|-- f36: double (nullable = true)
|-- f37: double (nullable = true)
|-- f38: double (nullable = true)
|-- f39: double (nullable = true)
|-- f40: double (nullable = true)
|-- f41: double (nullable = true)
|-- f42: double (nullable = true)
|-- f43: double (nullable = true)
|-- f44: double (nullable = true)
|-- f45: double (nullable = true)
|-- f46: double (nullable = true)
|-- f47: double (nullable = true)
|-- f48: double (nullable = true)
|-- f49: double (nullable = true)
|-- f50: double (nullable = true)
|-- f51: double (nullable = true)
|-- f52: double (nullable = true)
|-- f53: double (nullable = true)
|-- f54: double (nullable = true)
|-- f55: double (nullable = true)
|-- f56: double (nullable = true)
|-- f57: double (nullable = true)
|-- f58: double (nullable = true)
|-- f59: double (nullable = true)
|-- f60: double (nullable = true)
|-- f61: double (nullable = true)
|-- f62: double (nullable = true)
|-- f63: double (nullable = true)
|-- f64: double (nullable = true)
|-- f65: double (nullable = true)
|-- f66: double (nullable = true)
|-- f67: double (nullable = true)
|-- label: integer (nullable = true)

```

In [7]: `dataset.dtypes`

```

Out[7]:
('f1', 'double'),
('f2', 'double'),
('f3', 'double'),
('f4', 'double'),
('f5', 'double'),
('f6', 'double'),
('f7', 'double'),
('f8', 'double'),
('f9', 'double'),
('f10', 'double'),
('f11', 'double'),
('f12', 'double'),
('f13', 'double'),
('f14', 'double'),
('f15', 'double'),
('f16', 'double'),
('f17', 'double'),
('f18', 'double'),
('f19', 'double'),
('f20', 'double'),
('f21', 'double'),
('f22', 'double'),

```

```
('f23', 'double'),
('f24', 'double'),
('f25', 'double'),
('f26', 'double'),
('f27', 'double'),
('f28', 'double'),
('f29', 'double'),
('f30', 'double'),
('f31', 'double'),
('f32', 'double'),
('f33', 'double'),
('f34', 'double'),
('f35', 'double'),
('f36', 'double'),
('f37', 'double'),
('f38', 'double'),
('f39', 'double'),
('f40', 'double'),
('f41', 'double'),
('f42', 'double'),
('f43', 'double'),
('f44', 'double'),
('f45', 'double'),
('f46', 'double'),
('f47', 'double'),
('f48', 'double'),
('f49', 'double'),
('f50', 'double'),
('f51', 'double'),
('f52', 'double'),
('f53', 'double'),
('f54', 'double'),
('f55', 'double'),
('f56', 'double'),
('f57', 'double'),
('f58', 'double'),
('f59', 'double'),
('f60', 'double'),
('f61', 'double'),
('f62', 'double'),
('f63', 'double'),
('f64', 'double'),
('f65', 'double'),
('f66', 'double'),
('f67', 'double'),
('label', 'int')]
```

```
in [8]: from pyspark.sql.functions import col, isnan, when, count
dataset.select([count(when(col(c).contains('None') | \
                        col(c).contains('NULL') | \
                        col(c).isNull() | \
                        isnan(c), c
                    )), alias(c)
                for c in dataset.columns])).show()
```

```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| f1| f2| f3| f4| f5| f6| f7| f8| f9|f10|f11|f12|f13|f14|f15|f16|f17|f18|f19|f20|f21|f22|f23|f24|f25|f26|f27|f28|
f29|f30|f31|f32|f33|f34|f35|f36|f37|f38|f39|f40|f41|f42|f43|f44|f45|f46|f47|f48|f49|f50|f51|f52|f53|f54|f55|f56|f
57|f58|f59|f60|f61|f62|f63|f64|f65|f66|f67|label|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

```
In [9]: dataset.summary().show()
```

A 7x10 grid of dashed lines. At each intersection of a horizontal and vertical dashed line, there is a '+' sign. There are 7 horizontal lines and 10 vertical lines, creating a grid of 6x9 small squares.


```
def getDuplicateColumns(dataset, shape):
    z = []
    y = []
    for x in range(shape[1]):
        if(dataset.select(dataset.columns[x]).collect() not in y):
            y.append(dataset.select(dataset.columns[x]).collect())
        else:
            z.append('f'+str(x+1))
    return z
duplicateColumnNames = getDuplicateColumns(dataset, shape)
print('Duplicate Columns are as follows')
for col in duplicateColumnNames:
    print('Column name : ', col)
```

```
dataset = dataset.drop(*duplicateColumnNames)
print('Shape: Rows:', dataset.count(), ',', 'Columns:', len(dataset.columns))
```

```
dataset.show(10)
```

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

```
--+
| f1| f2| f3| f4| f5| f6| f7| f8| f9| f10| f11| f12| f13| f14| f15| f16|
|f17| f18| f19| f20| f21| f22| f23| f24| f25| f26| f27| f28| f29| f30| f31| f32| f
33| f34| f35| f36| f37| f38| f39| f40| f41| f42| f43| f44| f45| f46| f47| f48| f4
9| f50| f51| f52| f53| f54| f55| f56| f58| f59| f62| f63| f64| f65| f66| f67|label|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
--+
```

|        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 170.39 | 167.28 | 143.44 | 124.67 | 139.01 | 125.83 | 144.33 | 151.26 | 175.51 | 171.31 | 161.9  | 146.92 | 141.8  | 140.91 | 132.8  | 128.48 |
| 170.83 | 161.06 | 169.61 | 168.07 | 154.88 | 149.33 | 152.91 | 152.61 | 137.09 | 169.59 | 183.01 | 180.44 | 164.55 | 152.63 | 157.77 | 165.07 |
| 49.52  | 142.73 | 173.85 | 182.23 | 174.58 | 163.28 | 164.01 | 169.2  | 148.48 | 134.31 | 174.07 | 191.33 | 157.51 | 168.66 | 156.89 | 170.86 |
| 2.24   | 184.84 | 167.02 | 123.67 | 140.54 | 153.69 | 147.57 | 144.65 | 172.96 | 169.67 | 133.23 | 124.41 | 138.44 | 142.93 | 137.13 | 134.44 |
| 0      | 169.75 | 190.96 | 175.53 | 138.27 | 137.47 | 139.23 | 133.23 | 130.25 | 147.73 | 163.93 | 167.36 | 171.52 | 155.54 | 139.34 | 149.3  |
| 173.37 | 141.0  | 153.57 | 128.45 | 159.93 | 165.33 | 147.94 | 143.74 | 140.88 | 182.53 | 184.08 | 148.09 | 125.26 | 139.67 | 138.7  | 132.86 |
| 141.5  | 145.4  | 164.58 | 170.71 | 127.83 | 133.99 | 141.2  | 152.87 | 142.36 | 148.26 | 162.54 | 156.55 | 153.39 | 137.99 | 129.64 | 137.59 |
| 5.39   | 147.92 | 152.14 | 162.48 | 168.72 | 161.14 | 147.87 | 141.02 | 139.58 | 141.58 | 148.43 | 168.12 | 169.9  | 165.64 | 166.86 | 137.69 |
| 0      | 153.69 | 153.68 | 144.02 | 158.73 | 178.87 | 157.04 | 152.92 | 147.52 | 142.87 | 165.26 | 160.39 | 137.86 | 149.62 | 153.43 | 152.6  |
| 146.35 | 167.11 | 134.27 | 126.81 | 136.28 | 158.4  | 171.58 | 161.6  | 162.91 | 143.14 | 129.73 | 125.97 | 151.09 | 177.89 | 169.2  | 160.65 |
| 56.86  | 135.66 | 126.92 | 131.22 | 154.37 | 158.59 | 158.52 | 155.25 | 154.33 | 130.18 | 127.66 | 148.94 | 155.37 | 163.53 | 139.7  | 143.79 |
| 1.61   | 166.88 | 164.6  | 149.58 | 139.56 | 154.74 | 173.01 | 155.18 | 155.19 | 170.51 | 146.89 | 141.01 | 159.43 | 169.68 | 163.24 | 165.17 |
| 0      | 131.69 | 151.56 | 151.05 | 134.0  | 151.18 | 175.53 | 171.34 | 159.77 | 151.95 | 146.1  | 148.53 | 140.28 | 138.16 | 145.44 | 150.4  |
| 163.8  | 152.43 | 171.49 | 150.2  | 131.28 | 157.18 | 157.04 | 151.13 | 151.66 | 143.32 | 157.23 | 152.91 | 134.75 | 154.65 | 171.65 | 160.3  |
| 7.73   | 143.67 | 145.87 | 151.73 | 147.09 | 151.21 | 157.96 | 148.5  | 156.6  | 147.45 | 153.18 | 156.58 | 157.83 | 147.98 | 143.47 | 142.83 |
| .08    | 147.91 | 148.05 | 145.66 | 156.5  | 167.52 | 151.33 | 129.51 | 164.25 | 155.82 | 150.82 | 146.58 | 128.85 | 140.76 | 177.35 | 174.61 |
| 0      | 162.85 | 158.88 | 132.27 | 138.41 | 143.98 | 159.3  | 177.26 | 180.58 | 159.34 | 164.66 | 138.04 | 132.76 | 157.88 | 165.58 | 173.64 |
| 163.5  | 127.97 | 167.31 | 141.39 | 147.02 | 137.52 | 135.46 | 146.41 | 159.09 | 164.53 | 148.45 | 130.76 | 136.22 | 144.86 | 127.38 | 137.09 |
| 159.08 | 53.25  | 182.49 | 187.75 | 139.37 | 117.12 | 124.07 | 134.6  | 144.85 | 132.64 | 170.87 | 188.26 | 173.32 | 135.74 | 127.15 | 131.69 |
| 127.8  | 149.46 | 114.16 | 100.11 | 154.02 | 175.3  | 175.46 | 144.39 | 142.47 | 132.8  | 130.96 | 188.21 | 179.52 | 146.2  | 153.73 | 152.12 |
| 146.58 | 0      | 132.05 | 149.12 | 165.08 | 170.62 | 162.19 | 157.1  | 145.86 | 149.52 | 162.84 | 149.5  | 138.86 | 140.41 | 156.82 | 171.41 |
| 158.94 | 153.78 | 176.64 | 137.61 | 169.6  | 149.55 | 138.69 | 142.72 | 172.72 | 172.08 | 156.74 | 154.95 | 155.39 | 152.42 | 146.35 | 141.2  |
| 150.94 | 154.63 | 150.24 | 140.62 | 144.13 | 159.41 | 148.12 | 149.68 | 139.02 | 121.29 | 132.76 | 161.71 | 157.6  | 165.57 | 142.59 | 151.45 |
| 153.89 | 152.43 | 153.07 | 139.58 | 131.01 | 155.6  | 157.89 | 162.75 | 153.9  | 145.37 | 147.36 | 140.65 | 157.63 | 152.16 | 140.43 | 142.32 |
| 142.06 | 154.87 | 0      | 153.59 | 142.25 | 157.33 | 156.08 | 149.33 | 162.97 | 150.25 | 146.47 | 145.99 | 137.82 | 152.9  | 161.64 | 150.23 |
| 170.7  | 185.03 | 174.97 | 126.8  | 154.72 | 140.58 | 158.21 | 163.89 | 159.1  | 178.79 | 181.28 | 188.27 | 105.94 | 127.84 | 152.78 | 161.62 |
| 163.97 | 179.15 | 180.57 | 179.18 | 127.12 | 121.63 | 142.36 | 149.33 | 158.32 | 181.72 | 184.42 | 175.92 | 141.72 | 121.99 | 130.39 | 174.9  |
| 124.85 | 137.45 | 164.72 | 144.44 | 165.39 | 157.7  | 162.63 | 154.26 | 159.03 | 143.8  | 125.68 | 159.71 | 149.75 | 162.83 | 162.37 | 162.75 |
| 168.6  | 170.81 | 168.1  | 0      | 167.68 | 153.49 | 149.19 | 148.71 | 166.03 | 167.04 | 153.06 | 157.48 | 133.57 | 143.66 | 167.27 | 172.45 |
| 179.8  | 169.15 | 150.5  | 152.77 | 127.77 | 147.58 | 127.63 | 143.08 | 167.35 | 176.35 | 165.35 | 168.15 | 168.1  | 129.51 | 129.52 | 142.58 |
| 161.59 | 167.7  | 162.09 | 167.67 | 170.75 | 125.04 | 126.4  | 161.1  | 164.27 | 155.41 | 160.66 | 172.89 | 177.24 | 157.48 | 143.26 | 166.09 |
| 145.76 | 167.91 | 157.71 | 161.01 | 155.85 | 169.13 | 159.15 | 160.73 | 156.99 | 152.19 | 168.29 | 171.96 | 150.72 | 154.01 | 145.01 | 164.87 |
| 157.2  | 147.07 | 162.98 | 167.99 | 0      | 136.48 | 130.02 | 131.72 | 152.04 | 163.03 | 172.93 | 170.11 | 165.2  | 166.41 | 120.67 | 119.02 |
| 135.76 | 147.52 | 164.59 | 176.47 | 168.61 | 166.5  | 121.19 | 133.17 | 111.33 | 114.23 | 127.57 | 155.83 | 178.35 | 186.52 | 168.26 | 149.86 |
| 131.02 | 140.33 | 135.15 | 145.21 | 164.04 | 173.8  | 153.96 | 122.81 | 125.75 | 156.86 | 164.41 | 151.03 | 154.61 | 163.85 | 151.98 | 120.14 |
| 131.17 | 142.04 | 161.46 | 158.05 | 143.62 | 142.59 | 146.69 | 167.86 | 144.87 | 119.45 | 140.61 | 173.71 | 158.99 | 145.14 | 162.0  | 124.21 |
| 145.11 | 142.15 | 148.38 | 142.86 | 154.15 | 0      | 145.96 | 140.31 | 126.34 | 113.12 | 118.66 | 140.33 | 139.9  | 139.51 | 168.7  | 149.54 |
| 149.38 | 147.88 | 129.45 | 143.81 | 131.18 | 120.76 | 172.49 | 125.42 | 152.75 | 148.41 | 153.47 | 139.24 | 138.86 | 143.44 | 117.8  | 156.93 |
| 145.75 | 133.26 | 146.65 | 162.72 | 156.35 | 131.51 | 36.35  | 151.53 | 143.73 | 127.52 | 140.33 | 168.8  | 163.12 | 136.13 | 150.21 | 148.94 |
| 173.1  | 158.82 | 121.89 | 137.21 | 154.26 | 160.58 | 158.22 | 142.27 | 141.34 | 148.79 | 165.43 | 159.23 | 133.41 | 127.64 | 171.77 | 152.2  |
| 146.03 | 142.27 | 127.53 | 141.77 | 160.67 | 159.55 | 0      | 145.96 | 140.31 | 126.34 | 113.12 | 118.66 | 140.33 | 139.9  | 139.51 | 168.7  |
| 149.54 | 149.38 | 147.88 | 129.45 | 143.81 | 131.18 | 120.76 | 172.49 | 125.42 | 152.75 | 148.41 | 153.47 | 139.24 | 138.86 | 143.44 | 117.8  |
| 156.93 | 145.75 | 133.26 | 146.65 | 162.72 | 156.35 | 131.51 | 36.35  | 151.53 | 143.73 | 127.52 | 140.33 | 168.8  | 163.12 | 136.13 | 150.21 |
| 148.94 | 173.1  | 158.82 | 121.89 | 137.21 | 154.26 | 160.58 | 158.22 | 142.27 | 141.34 | 148.79 | 165.43 | 159.23 | 133.41 | 127.64 | 171.77 |
| 152.2  | 146.03 | 142.27 | 127.53 | 141.77 | 160.67 | 159.55 | 0      | 145.96 | 140.31 | 126.34 | 113.12 | 118.66 | 140.33 | 139.9  | 139.51 |
| 168.7  | 149.54 | 149.38 | 147.88 | 129.45 | 143.81 | 131.18 | 120.76 | 172.49 | 125.42 | 152.75 | 148.41 | 153.47 | 139.24 | 138.86 | 143.44 |
| 117.8  | 156.93 | 145.75 | 133.26 | 146.65 | 162.72 | 156.35 | 131.51 | 36.35  | 151.53 | 143.73 | 127.52 | 140.33 | 168.8  | 163.12 | 136.13 |
| 150.21 | 148.94 | 173.1  | 158.82 | 121.89 | 137.21 | 154.26 | 160.58 | 158.22 | 142.27 | 141.34 | 148.79 | 165.43 | 159.23 | 133.41 | 127.64 |
| 171.77 | 152.2  | 146.03 | 142.27 | 127.53 | 141.77 | 160.67 | 159.55 | 0      | 145.96 | 140.31 | 126.34 | 113.12 | 118.66 | 140.33 | 139.9  |
| 139.51 | 168.7  | 149.54 | 149.38 | 147.88 | 129.45 | 143.81 | 131.18 | 120.76 | 172.49 | 125.42 | 152.75 | 148.41 | 153.47 | 139.24 | 138.86 |
| 143.44 | 117.8  | 156.93 | 145.75 | 133.26 | 146.65 | 162.72 | 156.35 | 131.51 | 36.35  | 151.53 | 143.73 | 127.52 | 140.33 | 168.8  | 163.12 |
| 136.13 | 150.21 | 148.94 | 173.1  | 158.82 | 121.89 | 137.21 | 154.26 | 160.58 | 158.22 | 142.27 | 141.34 | 148.79 | 165.43 | 159.23 | 133.41 |
| 127.64 | 171.77 | 152.2  | 146.03 | 142.27 | 127.53 | 141.77 | 160.67 | 159.55 | 0      | 145.96 | 140.31 | 126.34 | 113.12 | 118.66 | 140.33 |
| 139.9  | 139.51 | 168.7  | 149.54 | 149.38 | 147.88 | 129.45 | 143.81 | 131.18 | 120.76 | 172.49 | 125.42 | 152.75 | 148.41 | 153.47 | 139.24 |
| 138.86 | 143.44 | 117.8  | 156.93 | 145.75 | 133.26 | 146.65 | 162.72 | 156.35 | 131.51 | 36.35  | 151.53 | 143.73 | 127.52 | 140.33 | 168.8  |
| 163.12 | 136.13 | 150.21 | 148.94 | 173.1  | 158.82 | 121.89 | 137.21 | 154.26 | 160.58 | 158.22 | 142.27 | 141.34 | 148.79 | 165.43 | 159.23 |
| 133.41 | 127.64 | 171.77 | 152.2  | 146.03 | 142.27 | 127.53 | 141.77 | 160.67 | 159.55 | 0      | 145.96 | 140.31 | 126.34 | 113.12 | 118.66 |
| 140.33 | 139.9  | 139.51 | 168.7  | 149.54 | 149.38 | 147.88 | 129.45 | 143.81 | 131.18 | 120.76 | 172.49 | 125.42 | 152.75 | 148.41 | 153.47 |
| 139.24 | 138.86 | 143.44 | 117.8  | 156.93 | 145.75 | 133.26 | 146.65 | 162.72 | 156.35 | 131.51 | 36.35  | 151.53 | 143.73 | 127.52 | 140.33 |
| 168.8  | 163.12 | 136.13 | 150.21 | 148.94 | 173.1  | 158.82 | 121.89 | 137.21 | 154.26 | 160.58 | 158.22 | 142.27 | 141.34 | 148.79 | 165.43 |
| 159.23 | 133.41 | 127.64 | 171.77 | 152.2  | 146.03 | 142.27 | 127.53 | 141.77 | 160.67 | 159.55 | 0      | 145.96 | 140.31 | 126.34 | 113.12 |
| 118.66 | 140.33 | 139.9  | 139.51 | 168.7  | 149.54 | 149.38 | 147.88 | 129.45 | 143.81 | 131.18 | 120.76 | 172.49 | 125.42 | 152.75 | 148.41 |
| 153.47 | 139.24 | 138.86 | 143.44 | 117.8  | 156.93 | 145.75 | 133.26 | 146.65 | 162.72 | 156.35 | 131.51 | 36.35  | 151.53 | 143.73 | 127.52 |
| 140.33 | 168.8  | 163.12 | 136.13 | 150.21 | 148.94 | 173.1  | 158.82 | 121.89 | 137.21 | 154.26 | 160.58 | 158.22 | 142.27 | 141.34 | 148.79 |
| 165.43 | 159.23 | 133.41 | 127.64 | 171.77 | 152.2  | 146.03 | 142.27 | 127.53 | 141.77 | 160.67 | 159.55 | 0      | 145.96 | 140.31 | 126.34 |
| 113.12 | 118.66 | 140.33 | 139.9  | 139.51 | 168.7  | 149.54 | 149.38 | 147.88 | 129.45 | 143.81 | 131.18 | 120.76 | 172.49 | 125.42 | 152.75 |
| 148.41 | 153.47 | 139.24 | 138.86 | 143.44 | 117.8  | 156.93 | 145.75 | 133.26 | 146.65 | 162.72 | 156.35 | 131.51 | 36.35  | 151.53 | 143.73 |
| 127.52 | 140.33 | 168.8  | 163.12 |        |        |        |        |        |        |        |        |        |        |        |        |

In [16]:

[illegible]

[illegible]

In [17]:

```
test_data_spark.show()
print(test_data_spark.count(), len(test_data_spark.columns))
```

| f17    | f18    | f19    | f20    | f21    | f22    | f23    | f24    | f25    | f26    | f27    | f28    | f29    | f30    | f31    | f32    | f      |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 33     | f34    | f35    | f36    | f37    | f38    | f39    | f40    | f41    | f42    | f43    | f44    | f45    | f46    | f47    | f48    | f4     |
| 9      | f50    | f51    | f52    | f53    | f54    | f55    | f56    | f58    | f59    | f62    | f63    | f64    | f65    | f66    | f67    | label  |
| 52.511 | 53.543 | 53.75  | 51.629 | 49.214 | 49.75  | 50.018 | 48.29  | 51.031 | 53.001 | 53.099 | 52.927 | 52.204 | 52.687 | 55.477 | 54.809 | 61.85  |
| 75.271 | 62.607 | 57.658 | 56.907 | 61.229 | 64.868 | 65.663 | 66.23  | 64.351 | 65.696 | 65.251 | 67.543 | 68.325 | 68.938 | 69.57  | 69.759 | 68.322 |
| 71.198 | 69.489 | 69.527 | 70.854 | 70.417 | 72.564 | 75.872 | 69.417 | 68.513 | 68.886 | 74.253 | 69.927 | 71.283 | 72.277 | 76.014 | 74.339 | 75.767 |
| 72.892 | 74.622 | 72.905 | 74.434 | 74.698 | 76.585 | 76.87  | 75.381 | 74.404 | 73.044 | 72.042 | 75.249 | 75.434 | 1      | 55.662 | 55.748 | 59.649 |
| 57.632 | 62.191 | 63.659 | 64.781 | 66.804 | 58.161 | 58.415 | 57.944 | 59.165 | 57.503 | 56.567 | 58.475 | 61.384 | 55.997 | 73.513 | 58.067 | 61.252 |
| 58.802 | 56.601 | 53.879 | 53.389 | 59.164 | 54.803 | 55.047 | 54.546 | 53.686 | 52.381 | 50.808 | 50.883 | 53.855 | 58.437 | 57.675 | 56.636 | 55.639 |
| 54.894 | 54.765 | 56.132 | 60.39  | 70.006 | 67.962 | 69.123 | 75.117 | 68.692 | 67.512 | 66.561 | 66.957 | 68.737 | 71.076 | 69.56  | 67.299 | 66.663 |
| 66.203 | 65.88  | 66.127 | 63.87  | 75.387 | 75.316 | 75.203 | 76.302 | 73.891 | 70.217 | 1      | 57.133 | 61.74  | 62.219 | 62.903 | 64.589 | 67.159 |
| 67.64  | 69.995 | 61.048 | 61.628 | 61.056 | 62.656 | 65.393 | 67.092 | 68.86  | 72.265 | 63.149 | 69.971 | 63.934 | 63.406 | 61.74  | 64.049 | 69.103 |
| 68.111 | 72.206 | 64.652 | 65.289 | 63.67  | 59.565 | 62.055 | 66.984 | 68.67  | 69.92  | 68.061 | 65.716 | 64.032 | 62.344 | 64.251 | 65.946 | 69.708 |
| 73.987 | 68.004 | 69.896 | 68.723 | 64.274 | 67.854 | 67.84  | 72.024 | 78.066 | 74.372 | 74.98  | 65.473 | 70.072 | 73.365 | 73.996 | 77.832 | 79.842 |
| 78.951 | 73.654 | 76.33  | 80.693 | 78.107 | 78.23  | 76.135 | 1      | 59.406 | 61.01  | 63.612 | 65.067 | 64.525 | 67.823 | 66.35  | 64.413 | 52.944 |
| 54.123 | 56.146 | 56.202 | 56.639 | 57.606 | 58.949 | 57.733 | 48.361 | 71.196 | 51.671 | 52.333 | 53.464 | 56.149 | 55.206 | 57.278 | 54.268 | 47.75  |
| 48.261 | 47.26  | 48.259 | 52.255 | 52.886 | 53.44  | 51.424 | 49.004 | 48.788 | 51.666 | 54.056 | 55.845 | 57.214 | 55.155 | 54.111 | 51.781 | 48.751 |
| 51.978 | 70.458 | 52.885 | 53.037 | 55.777 | 54.802 | 55.221 | 54.292 | 61.945 | 60.935 | 59.322 | 59.689 | 56.407 | 56.035 | 53.041 | 74.025 | 73.191 |
| 69.461 | 64.86  | 62.93  | 64.217 | 1      | 61.339 | 59.886 | 60.785 | 60.526 | 58.066 | 57.546 | 58.672 | 56.818 | 49.457 | 51.899 | 50.35  | 49.748 |
| 52.203 | 52.229 | 51.047 | 50.432 | 49.357 | 67.959 | 51.669 | 55.082 | 55.646 | 57.677 | 57.426 | 58.64  | 59.908 | 63.099 | 64.675 | 65.057 | 69.25  |
| 69.038 | 71.635 | 71.3   | 67.457 | 68.151 | 68.899 | 6      |        |        |        |        |        |        |        |        |        |        |



1.823|78.288|73.502|71.206|69.168|68.604|70.782|71.984|74.543| 77.77|75.679|74.789|76.361|71.133|75.559|76.955|84.071|79.424|78.765| 78.41|77.927|81.347|79.519|81.393|86.229|82.182|84.851|85.252|84.715|84.734|84.279|81.273|1|  
|61.734|64.408| 67.62|68.494|68.576|63.387|60.117|58.865|61.926|62.619|63.263|69.367|67.805|63.245|63.064|57.143|64.985|60.196|66.968|66.705|71.755|69.559|65.568|63.342|57.744|68.774|70.677|70.853|69.954|68.265|65.347|62.321|58.097|70.072|69.126|68.102|69.506|67.975|65.386|65.898|66.749|64.832|63.876|63.204|57.948|64.313|67.149|68.218| 64.12|67.782|67.399|61.561|65.854|66.147|66.546|65.883|64.148|65.317| 57.78|59.105|61.746|60.258|59.728|62.549|1|  
| 62.8|68.942|70.733| 72.27|74.104|70.765|70.433|73.389| 83.64|83.944|88.289|87.761|85.594|84.764|83.381| 84.21|88.539|91.711|88.984|89.777|89.711|90.207|88.429|89.882|90.201| 86.99|89.028|87.721|87.014|88.114|85.218|88.576|90.162|78.969|81.362|84.892|85.082|81.996|82.254|81.217|85.179|75.091|82.033|85.999|90.223|84.767|81.454|80.516|82.219|80.471|82.209|81.035|85.185|87.327|88.542|85.139|80.844|85.389|93.813|92.941|92.318|91.933|86.628|86.684|1|  
|63.384|65.347|67.293|65.818|68.409|71.619|73.612|70.232|62.096|64.167|64.131|65.126|71.325|80.822|79.995|77.892|70.436|82.982|69.777| 71.28|71.076|71.766|75.921|77.333|79.169|73.688|71.286|72.013|73.677|74.157|76.264|79.803|81.891|82.735| 80.01|76.215|74.958|72.729|73.849|77.971| 84.65| 80.91|77.776|78.133|76.992|78.504|77.196|77.043| 83.48|78.962|82.056|80.023| 81.33|80.101|82.565|82.684|80.189|81.945|84.102|83.754|87.363|86.611|87.597|85.507|1|  
|64.004|61.371|60.146|59.177|60.706|60.123|57.281|58.069|57.733|57.166|57.822|58.998| 56.99|54.461|53.486|54.778|57.327|63.081|55.753|55.711|56.272|55.188|51.364| 49.03|49.971|57.904|57.133| 55.74|54.123|54.801|49.798|48.762|48.337|56.042|52.771|50.755|51.671|51.949|51.734|51.406|53.961|50.275|50.685|50.488|63.982|48.373|48.407| 47.38|40.721|45.962|48.894|50.706|52.479|48.546|42.369|42.543|45.416|50.894|58.237|51.939|48.726|51.113|59.896|69.192|1|  
|64.079|64.647|66.216|67.547|68.126|71.193|70.889|72.895|79.797|80.737|81.601|79.491|80.121|81.976|83.985|85.811|83.994| 61.1|83.706|82.719|82.065|80.751|81.635|83.408|82.876|75.529|74.226|72.151|72.114|73.594|76.086| 76.51|76.756|70.059|71.302|70.478| 69.67|69.598|74.431|74.397|75.382|72.087|72.255|69.078|62.857|68.203|66.635|68.011|59.139|67.535|69.315|68.013| 67.84|67.167|62.397| 58.53|57.551|56.969|60.518|57.667|55.805|57.084|58.711|59.659|1|  
|65.167|64.757|65.674|67.307|67.538|67.053|65.567|66.037|63.803|61.757|61.783|64.155|63.872|63.204|62.067|60.701|54.086| 70.09|54.727|55.711|56.253|55.941|54.957|49.989|49.684|55.659| 55.67|59.671|58.714|59.671|60.998|57.243|56.947|65.998|65.711|66.909|64.344|66.057|65.651|64.784|64.126|65.303|66.915|65.894|68.197|65.027|63.913|66.385|70.938|70.987|72.104|65.018| 66.73|68.634|69.049|70.263|72.973|73.604| 72.03|72.787|72.937| 71.79|70.405| 70.09|1|  
|65.633|63.843|65.408|69.745|69.677|69.875|70.502|71.232|74.077|73.682|76.209|76.115|77.981|75.999|76.089|79.135|89.156| 89.47|88.558|87.601|88.414| 87.21|85.945|84.439|83.119|87.755|89.783|89.955|90.279|90.071|87.023|87.946|85.313|86.825|86.973|90.798|90.548|87.203|85.962|86.935| 88.8|87.296|90.186|90.252|91.621|91.376|89.711|90.623|92.517|92.165|92.129|91.461|92.571|92.721|92.549|89.763|92.657|91.187|93.039|92.911|91.024|90.382|90.695|88.833|1|  
|65.702|65.413|66.569|65.598|67.852|71.063|72.427|74.823|72.418|72.376|75.317|75.858|76.603|79.292|82.665|80.105|79.387|70.757|75.418|76.884| 78.24|80.164|81.032|80.033|77.909|83.779|80.933|80.266|79.458|81.804|81.338|80.768|78.97|84.087|82.095|81.737|79.633|78.894|81.903|82.843|79.933|78.134|78.567|77.273|69.958|77.029|73.568|75.323|75.358|78.001|76.637|78.885|78.514|77.515|73.736|74.793|78.011|72.605|71.006|70.483|67.373|65.674|67.011|64.704|1|  
|66.066|67.368|68.996|67.981|68.833|68.162|68.083|70.634|67.008|68.256|70.419| 70.9| 69.84|69.053| 70.83| 75.0|77.569|67.658|73.991|74.488|73.987|74.168|76.323|75.943|77.795|76.843|77.613|76.926|76.352|77.494| 79.37|79.948|78.915|70.186|71.746|69.712|71.907|74.017|73.749|76.151|77.112| 64.77|62.791|62.877|68.217|62.534|61.747|63.695|59.4|65.441|72.756|63.326|61.953|63.863|61.906|60.969|60.969| 63.25|66.274| 62.88|63.103|65.213|65.846|68.503|1|  
|66.332|67.544| 66.97|66.256|69.167|70.901|71.507|72.774|65.539| 65.29|66.936|67.123|65.431|69.128|65.794|68.684|73.743|79.139|74.568|73.559|71.583|67.124|69.934|66.887|66.722|75.969|76.185|75.642|74.251|72.706|74.903|72.765|74.245| 72.47|73.322|71.066|69.598|72.847|71.395|74.824|76.273|68.429|66.815|61.513|78.714|64.072|64.881|63.815|67.671| 66.41|65.823|70.721|67.837|66.719|66.627|66.931|71.666|73.283|80.394|80.475|79.361|80.487|79.457| 81.25|1|  
|66.362| 65.71|65.922| 65.88|67.347|68.625|68.396|68.196|67.612|65.567|67.179|66.945|66.089|68.467|68.008|65.114|66.437|73.126|64.024|67.265|67.749|68.029|67.207|67.755|64.241|62.248|61.405|60.142|61.103|62.101|62.975|64.871|62.191|62.705|63.932|62.068|59.869|58.374|57.176|57.767|58.628| 58.39|59.431|58.613|71.661|58.118|54.051|52.268|57.211|53.902|53.013| 61.46|62.185|58.503|58.001|54.885|58.222|57.286|70.096|67.247|67.147| 68.19|69.413|68.928|1|  
|67.043|64.236|64.901|66.698|66.004|65.497|64.173|66.898|64.414|62.164| 63.64|65.794| 67.31|66.437|65.341|66.626|61.831|67.304|61.762|62.646|64.092|64.891|64.289|63.813|62.153|60.932|61.378|56.887|61.044|60.898|61.744|60.349|58.853| 56.23|58.317|57.385|59.227| 56.48|54.039|53.381|52.242|51.004|54.138|53.644|69.031|55.148| 52.33|50.863|52.117|49.957|47.009|56.931|57.513|58.503|56.755|53.119|50.234|49.128| 65.73|60.126|57.557| 63.02|61.043|59.849|1|  
|67.619|71.722|68.699| 63.55|68.327|71.728|70.477|66.271|67.487|69.413|65.935|62.822|65.039|64.389|62.926|59.099|66.451|62.076| 66.67|62.559|65.498|68.929|68.732|62.933|58.314|62.813|64.365|64.349|64.067|68.941|68.549|64.399|66.035|63.525| 61.14|62.654|63.277|66.508|64.966|65.422|65.161|62.721| 61.42|64.224|58.693|64.506|66.204|65.946|69.847|65.378|63.663|62.876|61.043|62.656|64.513| 67.44|69.194|67.326|65.465|65.954|67.077|70.768|70.707|70.757|1|  
|67.621| 63.42|61.733|65.349|65.604|67.332|68.975|67.469|49.963|49.888| 51.72|53.802|57.909|56.762|57.195|54.658|42.081|52.888|44.227|48.473|49.147|50.945|53.396|51.451|54.071|50.606|50.445|48.604|49.952|52.624|52.358|53.741|56.487|54.186|53.891|50.315|48.892| 51.07|50.005|52.939|54.958|48.941|50.283|49.325|54.559|48.172|45.733|46.233|50.827| 49.03|51.718|41.958|43.798|46.778|48.064|46.665|52.378|52.907|57.107|58.256|58.715| 60.68| 59.21|60.371|1|

1|  
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
--+

only showing top 20 rows