```
!pip install pyspark
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting pyspark
  Downloading pyspark-3.3.2.tar.gz (281.4 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 281.4/281.4 MB 4.4 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Collecting py4j==0.10.9.5
  Downloading py4j-0.10.9.5-py2.py3-none-any.whl (199 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 199.7/199.7 KB 19.8 MB/s eta 0:00:00
Building wheels for collected packages: pyspark
  Building wheel for pyspark (setup.py) ... done
  Created wheel for pyspark: filename=pyspark-3.3.2-py2.py3-none-any.whl size=281824028 sha256=2687462789de81d518409dc5df
  Stored in directory: /root/.cache/pip/wheels/6c/e3/9b/0525ce8a69478916513509d43693511463c6468db0de237c86
Successfully built pyspark
Installing collected packages: py4j, pyspark
  Attempting uninstall: py4j
    Found existing installation: py4j 0.10.9.7
    Uninstalling py4j-0.10.9.7:
      Successfully uninstalled py4j-0.10.9.7
Successfully installed py4j-0.10.9.5 pyspark-3.3.2
```

```
#Importing SparkSession from pysparkSQL
from pyspark.sql import SparkSession
from pyspark.ml.feature import Imputer
```

```
#Creating an app
spark = SparkSession.builder.appName('Project').getOrCreate()
```

```
data_frame = spark.read.format("com.dtabricks.spark.csv")\
.option("mode", "DROPMALFORMED").option("header" ,True)\
.option("inferschema", True).csv("/content/Chennai_1990_2022.csv")
```

```
data_frame.filter("tavg is null").show()
```

```
+----------+----+----+----+----+
|      time|tavg|tmin|tmax|prcp|
+----------+----+----+----+----+
|27-09-1990|null|null|null|null|
|28-09-1990|null|null|null|null|
|20-10-1990|null|null|null|null|
|21-10-1990|null|null|null|null|
|22-10-1990|null|null|null|null|
|24-10-1990|null|null|null|null|
|02-11-1990|null|null|null|null|
|04-11-1990|null|null|null|null|
|05-11-1990|null|null|null|null|
|20-11-1990|null|null|null|null|
|01-12-1990|null|null|null|null|
|02-12-1990|null|null|null|null|
|14-12-1990|null|null|null|null|
|15-12-1990|null|null|null|null|
|21-12-1990|null|null|null|null|
|27-12-1990|null|null|null|null|
|31-01-1991|null|null|null|null|
|03-02-1991|null|null|null|null|
|26-02-1991|null|null|null|null|
|24-03-1991|null|null|null|null|
+----------+----+----+----+----+
only showing top 20 rows
```

```
imputer = Imputer(
    inputCol = data_frame.columns[1],
    outputCol = "tavg"
)
data_frame = imputer.fit(data_frame).transform(data_frame)
```

```
data_frame.filter("tavg is null").show()
```

```
+----+----+----+----+----+
|time|tavg|tmin|tmax|prcp|
+----+----+----+----+----+
+----+----+----+----+----+
```

```
data_frame.write.option("header",True).csv("/content/Chennai_Temperature_Processed")
```

✓ 0s completed at 17:20 ● ✕