

Overview

This notebook will show you how to create and query a table or DataFrame that you uploaded to DBFS. [DBFS](#) is a Databricks File System that allows you to store data for querying inside of Databricks. This notebook assumes that you have a file already inside of DBFS that you would like to read from.

This notebook is written in **Python** so the default cell type is Python. However, you can use different languages by using the `%LANGUAGE` syntax. Python, Scala, SQL, and R are all supported.

```
# File location and type
file_location = "/FileStore/tables/weather_data_weather_data.csv"
file_type = "csv"

# CSV options
infer_schema = "true"
first_row_is_header = "true"
delimiter = ","

# The applied options are for CSV files. For other file types, these
will be ignored.
df = spark.read.format(file_type) \
    .option("inferSchema", infer_schema) \
    .option("header", first_row_is_header) \
    .option("sep", delimiter) \
    .load(file_location)

display(df)

# Rename columns for better readability
df_renamed = df \
    .withColumnRenamed("city", "City") \
    .withColumnRenamed("wind.gust", "Wind_Gust") \
    .withColumnRenamed("rain.1h", "Rain_1H") \
    .withColumnRenamed("weather[0].main", "Weather_Main") \
    .withColumnRenamed("weather[0].description",
"Weather_Description") \
    .withColumnRenamed("temp", "Temperature") \
    .withColumnRenamed("humidity", "Humidity") \
    .withColumnRenamed("pressure", "Pressure")

# Show the renamed columns
df_renamed.show()
```

```
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
```

```

+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+
|          _id|      City|coord.lon|coord.lat|weather[0].id|
weather[1].id|Weather_Main|weather[1].main|Weather_Description|
weather[1].description|weather[0].icon|weather[1].icon|    base|
main.temp|main.feels_like|main.temp_min|main.temp_max|main.pressure|
main.humidity|main.sea_level|main.grnd_level|visibility|wind.speed|
wind.deg|clouds.all|          dt|sys.type| sys.id|sys.country|
sys.sunrise|sys.sunset|timezone|    id|          name|cod|Wind_Gust|
Rain_1H|
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+
+-----+
|673f04fa59f9e5711...|New Delhi| 77.2311| 28.6128|          721|
null|          Haze|          null|          haze|
null|          50.0|          null|stations| 25.09|
24.59|          25.09|          25.09|          1014|          36|
1014|          988|          1500|          2.06|          280|          0|
1732175963|          1| 9165|          IN| 1732151925|1732190132|
19800|1261481| New Delhi|200|          null|          null|
|673f04fa59f9e5711...| Mumbai| 72.8479| 19.0144|          800|
null|          Clear|          null|          clear sky|
null|          1.0|          null|stations| 30.0|
32.0|          30.0|          30.0|          1013|          70|
null|          null|          6000|          5.1|          250|          10|
1732175963|          1| 9166|          IN| 1732150000|1732187000|
19800|1275339| Mumbai|200|          null|          null|
|673f420f644efe42f...|          null| 77.2311| 28.6128|          721|
null|          Haze|          null|          haze|
null|          50.0|          null|stations| 25.09|
24.59|          25.09|          25.09|          1014|          36|
1014|          988|          1500|          2.06|          280|          0|
1732175963|          1| 9165|          IN| 1732151925|1732190132|
19800|1261481| New Delhi|200|          null|          null|
|673f420f644efe42f...|          null| 72.8479| 19.0144|          711|
null|          Smoke|          null|          smoke|
null|          50.0|          null|stations| 30.99|
29.63|          30.99|          31.94|          1011|          29|
1011|          1011|          3000|          3.09|          110|          17|
1732175700|          1| 9052|          IN| 1732151981|1732192181|
19800|1275339| Mumbai|200|          null|          null|
|673f420f644efe42f...|          null| 88.3697| 22.5697|          721|

```

null	Haze	null	haze
null	50.0	null stations	27.97
28.16	27.97	27.97	1011 47
1011	1010	3500	2.57 320 40
1732175149	1 9114	IN	1732148606 1732188104
19800 1275004	Kolkata 200	null	null
673f420f644efe42f...	null	80.2785	13.0878 721
null	Haze	null	haze
null	50.0	null stations	30.09
34.54	28.94	30.59	1011 67
1011	1010	4000	4.63 60 75
1732175698	2 2093220	IN	1732149641 1732190953
19800 1264527	Chennai 200	null	null
673f420f644efe42f...	null	77.6033	12.9762 721
null	Haze	null	haze
null	50.0	null stations	25.92
26.11	24.9	26.32	1011 59
1011	914	5000	1.03 100 40
1732175599	2 2017753	IN	1732150273 1732191606
19800 1277333	Bengaluru 200	null	null
673f420f644efe42f...	null	78.4744	17.3753 721
null	Haze	null	haze
null	50.0	null stations	28.23
27.6	27.73	28.23	1012 36
1012	949	5000	2.57 100 79
1732175879	1 9214	IN	1732150473 1732190988
19800 1269843	Hyderabad 200	null	null
673f420f644efe42f...	null	75.8167	26.9167 721
null	Haze	null	haze
null	50.0	null stations	24.62
23.84	24.62	24.62	1014 27
1014	966	3000	3.09 90 0
1732175444	1 9170	IN	1732152077 1732190660
19800 1269515	Jaipur 200	null	null
673f420f644efe42f...	null	77.4	23.2667 800
null	Clear	null	clear sky
null	1.0	null stations	25.44
24.61	25.44	25.44	1013 22
1013	956	10000	2.12 5 0
1732175962	null	null	IN 1732151311 1732190666
19800 1275841	Bhopal 200	2.18	null
673f420f644efe42f...	null	80.9167	26.85 711
null	Smoke	null	smoke
null	50.0	null stations	23.99
23.65	23.99	23.99	1013 46
1013	999	1500	1.54 350 0
1732175309	1 9176	IN	1732150845 1732189443
19800 1264733	Lucknow 200	null	null
673f420f644efe42f...	null	85.1167	25.6 711

null	Smoke	null	smoke
null	50.0	null stations	26.96
27.63	26.96	26.96	1012 54
1012	1006	2200	2.06 320 0
1732175674	1 9129	IN	1732149702 1732188570
19800 1260086	Patna 200	null	null
673f420f644efe42f...	null	76.9167	8.4833 501
701	Rain	Mist	moderate rain
mist	10.0	50.0 stations	27.99
32.54	27.99	27.99	1008 83
1008	1008	4000	3.09 360 100
1732175968	1 9226	IN	1732150035 1732192174
19800 1254163	Trivandrum 200	null	2.2
673f420f644efe42f...	null	76.7933	30.7343 800
null	Clear	null	clear sky
null	1.0	null stations	25.25
24.41	25.25	25.25	1014 22
1014	974	10000	2.81 302 0
1732175813	null	null	IN 1732152274 1732189993
19800 1274746	Chandigarh 200	3.44	null
673f420f644efe42f...	null	85.8333	20.2333 721
null	Haze	null	haze
null	50.0	null stations	28.12
27.91	28.12	28.32	1011 42
1011	1007	5000	1.54 40 20
1732175969	1 9113	IN	1732148982 1732188946
19800 1275817	Bhubaneswar 200	null	null
673f420f644efe42f...	null	85.3333	23.35 721
null	Haze	null	haze
null	50.0	null stations	23.06
22.47	23.06	23.06	1013 40
1013	938	4000	2.57 70 0
1732175812	1 9131	IN	1732149415 1732188753
19800 1258526	Ranchi 200	null	null
673f420f644efe42f...	null	78.0437	30.3256 800
null	Clear	null	clear sky
null	1.0	null stations	22.19
21.4	22.19	22.19	1014 36
1014	919	10000	2.88 234 0
1732175814	1 9162	IN	1732151926 1732189741
19800 1273313	Dehradun 200	1.79	null
673f420f644efe42f...	null	91.8831	25.5689 802
null	Clouds	null	scattered clouds
null	3.0	null stations	21.02
20.12	21.02	21.02	1013 36
1013	853	6000	3.09 140 40
1732175970	1 9111	IN	1732148074 1732186950
19800 1256523	Shillong 200	null	null
673f420f644efe42f...	null	88.6167	27.3333 803

[illegible]

```
df_cleaned = df_renamed.drop("city", "weather[1].id",
                              "weather[1].main", "weather[1].description", "weather[1].icon",
                              "wind_gust", "rain_1h")
```

```
display(df_cleaned)
```

```
df_filtered = df_cleaned \
    .withColumnRenamed("name", "location") \
    .withColumnRenamed("main.temp", "temperature") \
    .withColumnRenamed("main.humidity", "humidity") \
    .withColumnRenamed("wind.speed", "wind_speed") \
    .withColumnRenamed("air quality column", "air_quality")
```

```
df_filtered.show()
```

```

|_id|coord.lon|coord.lat|weather[0].id|Weather_Main|
Weather_Description|weather[0].icon|base|temperature|
main.feels_like|main.temp_min|main.temp_max|main.pressure|humidity|
main.sea_level|main.grnd_level|visibility|wind_speed|wind.deg|
clouds.all|dt|sys.type|sys.id|sys.country|sys.sunrise|
sys.sunset|timezone|id|location|cod|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
|673f04fa59f9e5711...|77.2311|28.6128|721|Haze|
haze|50.0|stations|25.09|24.59|
25.09|25.09|1014|36|1014|
988|1500|2.06|280|0|1732175963|1|
9165|IN|1732151925|1732190132|19800|1261481|New Delhi|
200|
|673f04fa59f9e5711...|72.8479|19.0144|800|Clear|
clear sky|1.0|stations|30.0|32.0|
30.0|30.0|1013|70|null|
null|6000|5.1|250|10|1732175963|1|
9166|IN|1732150000|1732187000|19800|1275339|Mumbai|
200|
|673f420f644efe42f...|77.2311|28.6128|721|Haze|
haze|50.0|stations|25.09|24.59|
25.09|25.09|1014|36|1014|
988|1500|2.06|280|0|1732175963|1|
9165|IN|1732151925|1732190132|19800|1261481|New Delhi|
200|
|673f420f644efe42f...|72.8479|19.0144|711|Smoke|
smoke|50.0|stations|30.99|29.63|
30.99|31.94|1011|29|1011|
1011|3000|3.09|110|17|1732175700|1|
9052|IN|1732151981|1732192181|19800|1275339|Mumbai|
200|
|673f420f644efe42f...|88.3697|22.5697|721|Haze|
haze|50.0|stations|27.97|28.16|
27.97|27.97|1011|47|1011|
1010|3500|2.57|320|40|1732175149|1|
9114|IN|1732148606|1732188104|19800|1275004|Kolkata|
200|
|673f420f644efe42f...|80.2785|13.0878|721|Haze|
haze|50.0|stations|30.09|34.54|
28.94|30.59|1011|67|1011|
1010|4000|4.63|60|75|1732175698|2|
2093220|IN|1732149641|1732190953|19800|1264527|
Chennai|200|
|673f420f644efe42f...|77.6033|12.9762|721|Haze|

```

haze	50.0	stations	25.92	26.11	
24.9	26.32	1011	59	1011	
914	5000	1.03	100	40 1732175599	2
2017753	IN	1732150273	1732191606	19800 1277333	
Bengaluru	200				
673f420f644efe42f...	78.4744	17.3753	721	Haze	
haze	50.0	stations	28.23	27.6	
27.73	28.23	1012	36	1012	
949	5000	2.57	100	79 1732175879	1
9214	IN	1732150473	1732190988	19800 1269843	Hyderabad
200					
673f420f644efe42f...	75.8167	26.9167	721	Haze	
haze	50.0	stations	24.62	23.84	
24.62	24.62	1014	27	1014	
966	3000	3.09	90	0 1732175444	1
9170	IN	1732152077	1732190660	19800 1269515	Jaipur
200					
673f420f644efe42f...	77.4	23.2667	800	Clear	
clear sky	1.0	stations	25.44	24.61	
25.44	25.44	1013	22	1013	
956	10000	2.12	5	0 1732175962	null
null	IN	1732151311	1732190666	19800 1275841	Bhopal
200					
673f420f644efe42f...	80.9167	26.85	711	Smoke	
smoke	50.0	stations	23.99	23.65	
23.99	23.99	1013	46	1013	
999	1500	1.54	350	0 1732175309	1
9176	IN	1732150845	1732189443	19800 1264733	Lucknow
200					
673f420f644efe42f...	85.1167	25.6	711	Smoke	
smoke	50.0	stations	26.96	27.63	
26.96	26.96	1012	54	1012	
1006	2200	2.06	320	0 1732175674	1
9129	IN	1732149702	1732188570	19800 1260086	Patna
200					
673f420f644efe42f...	76.9167	8.4833	501	Rain	
moderate rain	10.0	stations	27.99	32.54	
27.99	27.99	1008	83	1008	
1008	4000	3.09	360	100 1732175968	1
9226	IN	1732150035	1732192174	19800 1254163	Trivandrum
200					
673f420f644efe42f...	76.7933	30.7343	800	Clear	
clear sky	1.0	stations	25.25	24.41	
25.25	25.25	1014	22	1014	
974	10000	2.81	302	0 1732175813	null
null	IN	1732152274	1732189993	19800 1274746	Chandigarh
200					
673f420f644efe42f...	85.8333	20.2333	721	Haze	
haze	50.0	stations	28.12	27.91	

```

28.12|      28.32|      1011|      42|      1011|
1007|      5000|      1.54|      40|      20|1732175969|      1|
9113|      IN| 1732148982|1732188946|      19800|1275817|Bhubaneswar|
200|
|673f420f644efe42f...| 85.3333|      23.35|      721|      Haze|
haze|      50.0|stations|      23.06|      22.47|
23.06|      23.06|      1013|      40|      1013|
938|      4000|      2.57|      70|      0|1732175812|      1|
9131|      IN| 1732149415|1732188753|      19800|1258526|      Ranchi|
200|
|673f420f644efe42f...| 78.0437| 30.3256|      800|      Clear|
clear sky|      1.0|stations|      22.19|      21.4|
22.19|      22.19|      1014|      36|      1014|
919|      10000|      2.88|      234|      0|1732175814|      1|
9162|      IN| 1732151926|1732189741|      19800|1273313|      Dehradun|
200|
|673f420f644efe42f...| 91.8831| 25.5689|      802|      Clouds|
scattered clouds|      3.0|stations|      21.02|      20.12|
21.02|      21.02|      1013|      36|      1013|
853|      6000|      3.09|      140|      40|1732175970|      1|
9111|      IN| 1732148074|1732186950|      19800|1256523|      Shillong|
200|
|673f420f644efe42f...| 88.6167| 27.3333|      803|      Clouds|
broken clouds|      4.0|stations|      17.8|      17.25|
17.8|      17.8|      1015|      62|      1015|
817|      10000|      1.47|      214|      71|1732175970|      null|
null|      IN| 1732149049|1732187543|      19800|1271631|      Gangtok|
200|
|673f420f644efe42f...| 93.6167|      27.1|      800|      Clear|
clear sky|      1.0|stations|      24.54|      23.99|
24.54|      24.54|      1012|      36|      1012|
947|      10000|      1.68|      132|      0|1732175971|      null|
null|      IN| 1732147822|1732186369|      19800|1269655|      Itanagar|
200|

```

```

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

```

only showing top 20 rows

```
# Select relevant columns
```

```
df_filtered = df_filtered.select("location", "temperature",
                                "humidity", "wind_speed")
```

```
# Display the filtered data
```

```
df_filtered.show()
```


location	temperature	humidity	wind_speed
New Delhi	25.09	36	2.06
Mumbai	30.0	70	5.1
New Delhi	25.09	36	2.06
Mumbai	30.99	29	3.09
Kolkata	27.97	47	2.57
Chennai	30.09	67	4.63
Bengaluru	25.92	59	1.03
Hyderabad	28.23	36	2.57
Jaipur	24.62	27	3.09
Bhopal	25.44	22	2.12
Lucknow	23.99	46	1.54
Patna	26.96	54	2.06
Trivandrum	27.99	83	3.09
Chandigarh	25.25	22	2.81
Bhubaneswar	28.12	42	1.54
Ranchi	23.06	40	2.57
Dehradun	22.19	36	2.88
Shillong	21.02	36	3.09
Gangtok	17.8	62	1.47
Itanagar	24.54	36	1.68

only showing top 20 rows

```
display(df_filtered)
```