databricks Literacy_in Haryana_by_district_2011

(https://databricks.com)

1

```python
%python

data_df=spark.read.csv.option("header","true").option("inferSchema",True)\
.load("dbfs:/FileStore/sample data/Literacy_in_Haryana_by_Districts_2011.csv")
display(data_df)
```

Table

	A^B_C District	1^2_3 Literate and Educated persons - Total - Total	1^2_3 Literate and Educated persons - Total - Males	1
1	Ambala	818025	461288	
2	Panchkula	405318	229862	
3	Yamunanagar	832457	474793	
4	Kurukshetra	646893	370828	
5	Kaithal	646529	385782	
6	Karnal	978013	564516	
7	Panipat	786599	464136	
8	Sonipat	998316	589881	
9	Rohtak	747582	436025	
10	Jhajjar	676091	399480	
11	Faridabad	1272739	734940	
12	Palwal	599796	379696	
13	Gurgaon	1111116	638666	
14	Mewat	454897	308435	

22 rows

2

```python
%python
from pyspark.sql.functions import current_date
data_df = data_df.withColumn("date", current_date())
print("Updated DataFrame with 'date' column:")
display(data_df)
```

Updated DataFrame with 'date' column:

Table

	A^B_C District	1^2_3 Literate and Educated persons - Total - Total	1^2_3 Literate and Educated persons - Total - Males	1
1	Ambala	818025	461288	
2	Panchkula	405318	229862	
3	Yamunanagar	832457	474793	
4	Kurukshetra	646893	370828	
5	Kaithal	646529	385782	
6	Karnal	978013	564516	
7	Panipat	786599	464136	
8	Sonipat	998316	589881	
9	Rohtak	747582	436025	
10	Jhajjar	676091	399480	
11	Faridabad	1272739	734940	
12	Palwal	599796	379696	
13	Gurgaon	1111116	638666	
14	Mewat	454897	308435	

22 rows

3

```python
%python
from pyspark.sql.functions import col

data_df = data_df. withColumn("URBAN FEMALE %", (col("Literate and Educated persons - Urban - Females") / col("Lit
Educated persons - Urban - Total")) * 100) \
        .withColumn("URBAN MALE %", (col("Literate and Educated persons - Urban - Males") / col("Literate and Educ
        - Urban - Total")) * 100)

display(data_df)
```

Table

4

```python
%python
from pyspark.sql.functions import lit, to_date, year, month, dayofmonth, col, when

start_date = lit('2011-01-01')
end_date = lit('2024-12-16')

data_df = data_df.withColumn("YEARS", year(to_date(end_date)) - year(to_date(start_date))) \
            .withColumn("MONTHS", month(to_date(end_date)) - month(to_date(start_date))) \
            .withColumn("DAYS", dayofmonth(to_date(end_date)) - dayofmonth(to_date(start_date))) \
            .withColumn("YEARS", col("YEARS") - (col("MONTHS") < 0).cast("int")) \
            .withColumn("MONTHS", (col("MONTHS") + 12) % 12) \
            .withColumn("DAYS", when(col("DAYS") < 0, col("DAYS") + 30).otherwise(col("DAYS"))) \
            .withColumn("MONTHS", when(col("DAYS") < 0, col("MONTHS") - 1).otherwise(col("MONTHS")))

display(data_df)
```

Table

5

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

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