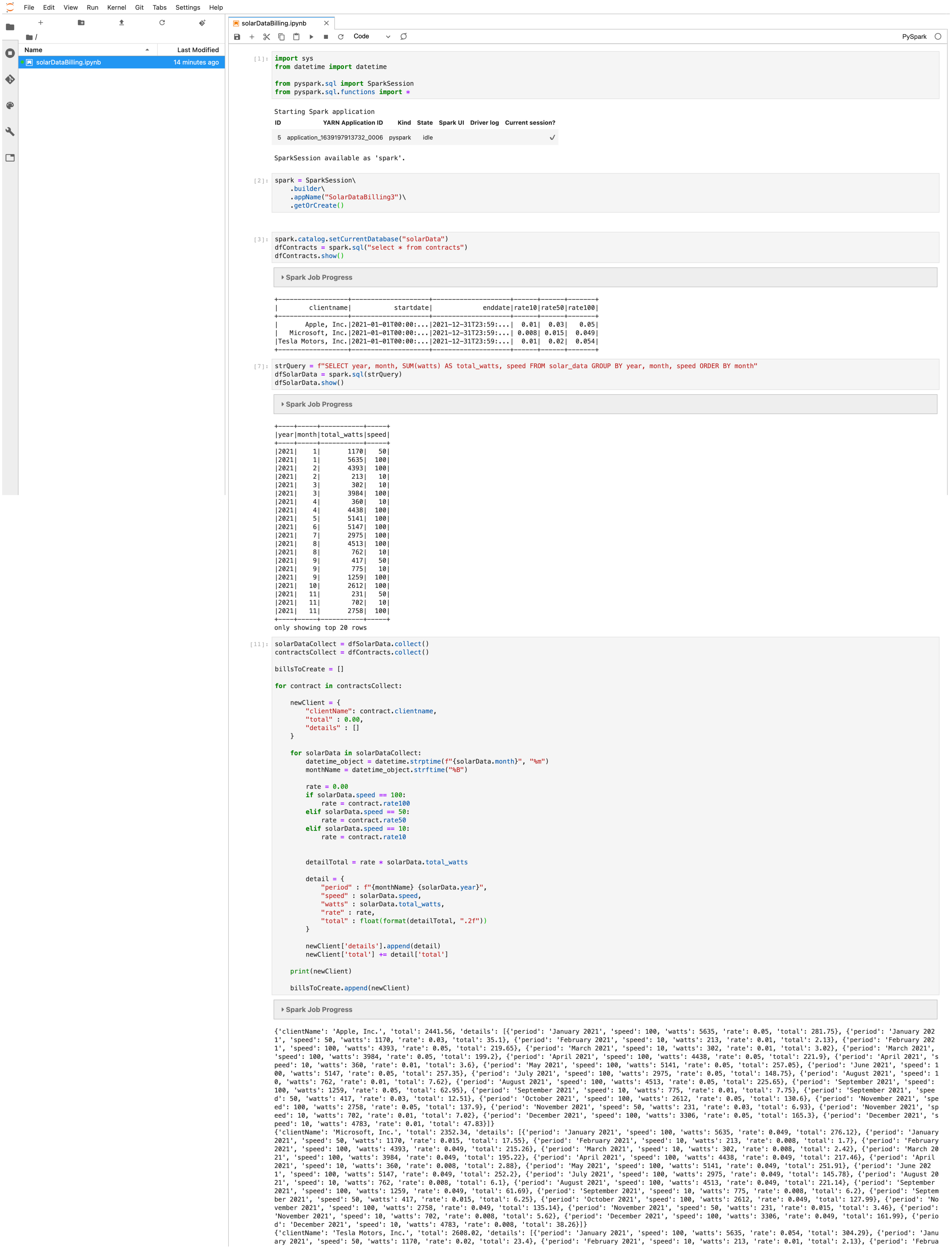


[1]:
```python
import sys
from datetime import datetime

from pyspark.sql import SparkSession
from pyspark.sql.functions import *
```

Starting Spark application

ID	YARN Application ID	Kind	State	Spark UI	Driver log	Current session?
5	application_1639197913732_0006	pyspark	idle			✓

SparkSession available as 'spark'.

[2]:
```python
spark = SparkSession\
    .builder\
    .appName("SolarDataBilling3")\
    .getOrCreate()
```

[3]:
```python
spark.catalog.setCurrentDatabase("solarData")
dfContracts = spark.sql("select * from contracts")
dfContracts.show()
```

▸ Spark Job Progress

```
+------------------+--------------------+--------------------+------+------+-------+
|        clientname|           startdate|             enddate|rate10|rate50|rate100|
+------------------+--------------------+--------------------+------+------+-------+
|       Apple, Inc.|2021-01-01T00:00:...|2021-12-31T23:59:...|  0.01|  0.03|   0.05|
|   Microsoft, Inc.|2021-01-01T00:00:...|2021-12-31T23:59:...| 0.008| 0.015|  0.049|
|Tesla Motors, Inc.|2021-01-01T00:00:...|2021-12-31T23:59:...|  0.01|  0.02|  0.054|
+------------------+--------------------+--------------------+------+------+-------+
```

[7]:
```python
strQuery = f"SELECT year, month, SUM(watts) AS total_watts, speed FROM solar_data GROUP BY year, month, speed ORDER BY month"
dfSolarData = spark.sql(strQuery)
dfSolarData.show()
```

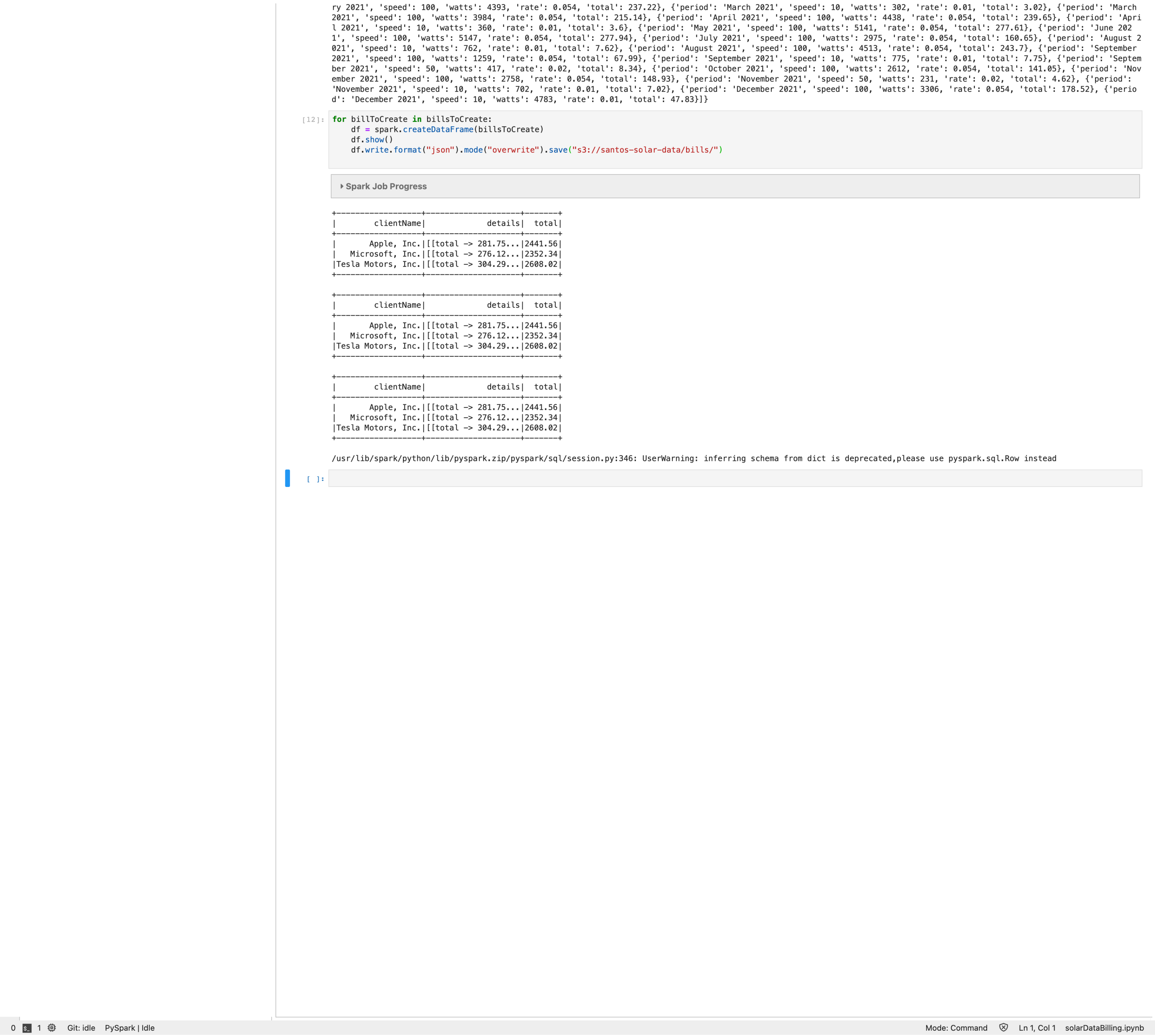
▸ Spark Job Progress

```
+----+-----+-----------+-----+
|year|month|total_watts|speed|
+----+-----+-----------+-----+
|2021|    1|       1170|   50|
|2021|    1|       5635|  100|
|2021|    2|       4393|  100|
|2021|    2|        213|   10|
|2021|    3|        302|   10|
|2021|    3|       3984|  100|
|2021|    4|        360|   10|
|2021|    4|       4438|  100|
|2021|    5|       5141|  100|
|2021|    6|       5147|  100|
|2021|    7|       2975|  100|
|2021|    8|       4513|  100|
|2021|    8|        762|   10|
|2021|    9|        417|   50|
|2021|    9|        775|   10|
|2021|    9|       1259|  100|
|2021|   10|       2612|  100|
|2021|   11|        231|   50|
|2021|   11|        702|   10|
|2021|   11|       2758|  100|
+----+-----+-----------+-----+
only showing top 20 rows
```

[11]:
```python
solarDataCollect = dfSolarData.collect()
contractsCollect = dfContracts.collect()

billsToCreate = []

for contract in contractsCollect:

    newClient = {
        "clientName": contract.clientname,
        "total" : 0.00,
        "details" : []
    }

    for solarData in solarDataCollect:
        datetime_object = datetime.strptime(f"{solarData.month}", "%m")
        monthName = datetime_object.strftime("%B")

        rate = 0.00
        if solarData.speed == 100:
            rate = contract.rate100
        elif solarData.speed == 50:
            rate = contract.rate50
        elif solarData.speed == 10:
            rate = contract.rate10


        detailTotal = rate * solarData.total_watts

        detail = {
            "period" : f"{monthName} {solarData.year}",
            "speed" : solarData.speed,
            "watts" : solarData.total_watts,
            "rate" : rate,
            "total" : float(format(detailTotal, ".2f"))
        }

        newClient['details'].append(detail)
        newClient['total'] += detail['total']

    print(newClient)

    billsToCreate.append(newClient)
```

▸ Spark Job Progress

```
{'clientName': 'Apple, Inc.', 'total': 2441.56, 'details': [{'period': 'January 2021', 'speed': 100, 'watts': 5635, 'rate': 0.05, 'total': 281.75}, {'period': 'January 202
1', 'speed': 50, 'watts': 1170, 'rate': 0.03, 'total': 35.1}, {'period': 'February 2021', 'speed': 10, 'watts': 213, 'rate': 0.01, 'total': 2.13}, {'period': 'February 202
1', 'speed': 100, 'watts': 4393, 'rate': 0.05, 'total': 219.65}, {'period': 'March 2021', 'speed': 10, 'watts': 302, 'rate': 0.01, 'total': 3.02}, {'period': 'March 2021',
'speed': 100, 'watts': 3984, 'rate': 0.05, 'total': 199.2}, {'period': 'April 2021', 'speed': 100, 'watts': 4438, 'rate': 0.05, 'total': 221.9}, {'period': 'April 2021', 's
peed': 10, 'watts': 360, 'rate': 0.01, 'total': 3.6}, {'period': 'May 2021', 'speed': 100, 'watts': 5141, 'rate': 0.05, 'total': 257.05}, {'period': 'June 2021', 'speed': 1
00, 'watts': 5147, 'rate': 0.05, 'total': 257.35}, {'period': 'July 2021', 'speed': 100, 'watts': 2975, 'rate': 0.05, 'total': 148.75}, {'period': 'August 2021', 'speed': 1
0, 'watts': 762, 'rate': 0.01, 'total': 7.62}, {'period': 'August 2021', 'speed': 100, 'watts': 4513, 'rate': 0.05, 'total': 225.65}, {'period': 'September 2021', 'speed':
100, 'watts': 1259, 'rate': 0.05, 'total': 62.95}, {'period': 'September 2021', 'speed': 10, 'watts': 775, 'rate': 0.01, 'total': 7.75}, {'period': 'September 2021', 'spee
d': 50, 'watts': 417, 'rate': 0.03, 'total': 12.51}, {'period': 'October 2021', 'speed': 100, 'watts': 2612, 'rate': 0.05, 'total': 130.6}, {'period': 'November 2021', 'spe
ed': 100, 'watts': 2758, 'rate': 0.05, 'total': 137.9}, {'period': 'November 2021', 'speed': 50, 'watts': 231, 'rate': 0.03, 'total': 6.93}, {'period': 'November 2021', 'sp
eed': 10, 'watts': 702, 'rate': 0.01, 'total': 7.02}, {'period': 'December 2021', 'speed': 100, 'watts': 3306, 'rate': 0.05, 'total': 165.3}, {'period': 'December 2021', 's
peed': 10, 'watts': 4783, 'rate': 0.01, 'total': 47.83}]}
{'clientName': 'Microsoft, Inc.', 'total': 2352.34, 'details': [{'period': 'January 2021', 'speed': 100, 'watts': 5635, 'rate': 0.049, 'total': 276.12}, {'period': 'January
2021', 'speed': 50, 'watts': 1170, 'rate': 0.015, 'total': 17.55}, {'period': 'February 2021', 'speed': 10, 'watts': 213, 'rate': 0.008, 'total': 1.7}, {'period': 'February
2021', 'speed': 100, 'watts': 4393, 'rate': 0.049, 'total': 215.26}, {'period': 'March 2021', 'speed': 10, 'watts': 302, 'rate': 0.008, 'total': 2.42}, {'period': 'March 20
21', 'speed': 100, 'watts': 3984, 'rate': 0.049, 'total': 195.22}, {'period': 'April 2021', 'speed': 100, 'watts': 4438, 'rate': 0.049, 'total': 217.46}, {'period': 'April
2021', 'speed': 10, 'watts': 360, 'rate': 0.008, 'total': 2.88}, {'period': 'May 2021', 'speed': 100, 'watts': 5141, 'rate': 0.049, 'total': 251.91}, {'period': 'June 202
1', 'speed': 100, 'watts': 5147, 'rate': 0.049, 'total': 252.2}, {'period': 'July 2021', 'speed': 100, 'watts': 2975, 'rate': 0.049, 'total': 145.78}, {'period': 'August 20
21', 'speed': 10, 'watts': 762, 'rate': 0.008, 'total': 6.1}, {'period': 'August 2021', 'speed': 100, 'watts': 4513, 'rate': 0.049, 'total': 221.14}, {'period': 'September
2021', 'speed': 100, 'watts': 1259, 'rate': 0.049, 'total': 61.69}, {'period': 'September 2021', 'speed': 10, 'watts': 775, 'rate': 0.008, 'total': 6.2}, {'period': 'Septem
ber 2021', 'speed': 50, 'watts': 417, 'rate': 0.015, 'total': 6.25}, {'period': 'October 2021', 'speed': 100, 'watts': 2612, 'rate': 0.049, 'total': 127.99}, {'period': 'No
vember 2021', 'speed': 100, 'watts': 2758, 'rate': 0.049, 'total': 135.14}, {'period': 'November 2021', 'speed': 50, 'watts': 231, 'rate': 0.015, 'total': 3.46}, {'period':
'November 2021', 'speed': 10, 'watts': 702, 'rate': 0.008, 'total': 5.62}, {'period': 'December 2021', 'speed': 100, 'watts': 3306, 'rate': 0.049, 'total': 161.99}, {'perio
d': 'December 2021', 'speed': 10, 'watts': 4783, 'rate': 0.008, 'total': 38.26}]}
{'clientName': 'Tesla Motors, Inc.', 'total': 2608.02, 'details': [{'period': 'January 2021', 'speed': 100, 'watts': 5635, 'rate': 0.054, 'total': 304.29}, {'period': 'Janu
ary 2021', 'speed': 50, 'watts': 1170, 'rate': 0.02, 'total': 23.4}, {'period': 'February 2021', 'speed': 10, 'watts': 213, 'rate': 0.01, 'total': 2.13}, {'period': 'Februa
```



```
ry 2021', 'speed': 100, 'watts': 4393, 'rate': 0.054, 'total': 237.22}, {'period': 'March 2021', 'speed': 10, 'watts': 302, 'rate': 0.01, 'total': 3.02}, {'period': 'March
2021', 'speed': 100, 'watts': 3984, 'rate': 0.054, 'total': 215.14}, {'period': 'April 2021', 'speed': 100, 'watts': 4438, 'rate': 0.054, 'total': 239.65}, {'period': 'Apri
l 2021', 'speed': 10, 'watts': 360, 'rate': 0.01, 'total': 3.6}, {'period': 'May 2021', 'speed': 100, 'watts': 5141, 'rate': 0.054, 'total': 277.61}, {'period': 'June 202
1', 'speed': 100, 'watts': 5147, 'rate': 0.054, 'total': 277.94}, {'period': 'July 2021', 'speed': 100, 'watts': 2975, 'rate': 0.054, 'total': 160.65}, {'period': 'August 2
021', 'speed': 10, 'watts': 762, 'rate': 0.01, 'total': 7.62}, {'period': 'August 2021', 'speed': 100, 'watts': 4513, 'rate': 0.054, 'total': 243.7}, {'period': 'September
2021', 'speed': 100, 'watts': 1259, 'rate': 0.054, 'total': 67.99}, {'period': 'September 2021', 'speed': 10, 'watts': 775, 'rate': 0.01, 'total': 7.75}, {'period': 'Septem
ber 2021', 'speed': 50, 'watts': 417, 'rate': 0.02, 'total': 8.34}, {'period': 'October 2021', 'speed': 100, 'watts': 2612, 'rate': 0.054, 'total': 141.05}, {'period': 'Nov
ember 2021', 'speed': 100, 'watts': 2758, 'rate': 0.054, 'total': 148.93}, {'period': 'November 2021', 'speed': 50, 'watts': 231, 'rate': 0.02, 'total': 4.62}, {'period':
'November 2021', 'speed': 10, 'watts': 702, 'rate': 0.01, 'total': 7.02}, {'period': 'December 2021', 'speed': 100, 'watts': 3306, 'rate': 0.054, 'total': 178.52}, {'perio
d': 'December 2021', 'speed': 10, 'watts': 4783, 'rate': 0.01, 'total': 47.83}]}
```

[12]:

```
for billToCreate in billsToCreate:
    df = spark.createDataFrame(billsToCreate)
    df.show()
    df.write.format("json").mode("overwrite").save("s3://santos-solar-data/bills/")
```

```
+------------------+--------------------+-------+
|        clientName|             details|  total|
+------------------+--------------------+-------+
|       Apple, Inc.|[[total -> 281.75...|2441.56|
|   Microsoft, Inc.|[[total -> 276.12...|2352.34|
|Tesla Motors, Inc.|[[total -> 304.29...|2608.02|
+------------------+--------------------+-------+

+------------------+--------------------+-------+
|        clientName|             details|  total|
+------------------+--------------------+-------+
|       Apple, Inc.|[[total -> 281.75...|2441.56|
|   Microsoft, Inc.|[[total -> 276.12...|2352.34|
|Tesla Motors, Inc.|[[total -> 304.29...|2608.02|
+------------------+--------------------+-------+

+------------------+--------------------+-------+
|        clientName|             details|  total|
+------------------+--------------------+-------+
|       Apple, Inc.|[[total -> 281.75...|2441.56|
|   Microsoft, Inc.|[[total -> 276.12...|2352.34|
|Tesla Motors, Inc.|[[total -> 304.29...|2608.02|
+------------------+--------------------+-------+

/usr/lib/spark/python/lib/pyspark.zip/pyspark/sql/session.py:346: UserWarning: inferring schema from dict is deprecated,please use pyspark.sql.Row instead
```

[ ]: