

Analysis of Electricity Demand, Consumption, and Electric Vehicle Trends Across Regions from 2016 to 2021

Processing Big Data for Analytics Applications

Fall 2023

Data ingestion

```
scala> :load merging.scala
Loading merging.scala...
import org.apache.spark.sql.{DataFrame, SparkSession}
import org.apache.spark.sql.functions._
import org.apache.spark.sql.types.DoubleType
import org.apache.spark.rdd.RDD
import scala.math.BigDecimal.RoundingMode
23/12/03 18:08:42 WARN org.apache.spark.sql.SparkSession$Builder: Using an existing SparkSession; some spark core conf
spark: org.apache.spark.sql.SparkSession = org.apache.spark.sql.SparkSession@164500f9
import spark.implicits._
consumptionDataPath: String = /user/svn9705_nyu_edu/final/consumption_agg_data.csv
pricingDataPath: String = /user/ar7165_nyu_edu/FinalProject/pricing_agg_data.csv
vehicleDataPath: String = /user/ssl4758_nyu_edu/finalproject/vehicle_agg_data.csv
consumptionDF: org.apache.spark.sql.DataFrame = [Year: int, Region: string ... 5 more fields]
pricingDF: org.apache.spark.sql.DataFrame = [region: string, year: int ... 9 more fields]
vehicleDF: org.apache.spark.sql.DataFrame = [Region: string, Transaction Year: int ... 5 more fields]
mergedDF: org.apache.spark.sql.DataFrame = [Year: int, Region: string ... 23 more fields]
finalDF: org.apache.spark.sql.DataFrame = [Region: string, Year: int ... 4 more fields]
+-----+-----+-----+-----+-----+-----+-----+-----+
|Region|Year|avg(Demand (MW)_imputed)|avg(Net Generation (MW)_imputed)|avg(avg_rate)|EV_adoption_count|
+-----+-----+-----+-----+-----+-----+-----+
|CENT|2017|14700.73864155251|15195.3553652968|0.10532125489613052|1|
|CENT|2018|11934.07603514916|13910.658206950031|0.1058389390687911|1|
|CENT|2019|15419.285901826484|15996.742865296803|0.10641573310413417|2|
|CENT|2020|14951.561703096539|15440.561987704918|0.11017308891367866|4|
|CENT|2021|15303.561358447489|15814.35462328767|0.10866549826449182|4|
|CAR|2016|5812.18759559083|4203.058374351534|0.09138635481392009|5|
|CAR|2017|5900.619970285446|4515.4114900284685|0.09179452985090777|2|
|CAR|2018|6158.935819940604|4920.673570020897|0.0895505580019013|8|
|CAR|2019|5822.201837440155|4326.143542968146|0.09311702366986217|12|
|CAR|2020|5637.939961021429|4038.1385568255655|0.09095822300196607|10|
|CAR|2021|5808.285419295379|4148.494855794077|0.09036248933553533|24|
|TEN|2016|18269.314063657068|18501.26278577736|0.16835413952230638|1|
|TEN|2017|6069.9083783015285|6597.4763485591075|0.13494546208863265|3|
|TEN|2018|18118.753628749215|18813.45669255206|0.13025406267386225|2|
|TEN|2019|18064.151703945216|18370.295091800836|0.16222832938662293|3|
|TEN|2020|17440.703020531048|17375.86328129425|0.202728483430992|2|
|MIDW|2016|22566.054309174506|19281.491903040776|0.07709587447523539|9|
|MIDW|2017|22404.803609360497|18776.489441237525|0.08008182379981417|14|
|MIDW|2018|23193.260565674012|20432.476635323485|0.07904086382655084|18|
|MIDW|2019|22666.916511232743|19425.992349202654|0.07940301919602817|29|
+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows
```

Merged all the datasets [electricity consumption, electricity pricing, and EV adoption] based on the region and the year for combined analysis

Analytics

Printing the schema of the merged dataset, and the number of rows and columns

```

svn9705_nyu_edu@nyu-dataproc-m:~$ spark-shell --deploy-mode client
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/12/03 20:27:20 INFO org.apache.spark.SparkEnv: Registering MapOutputTracker
23/12/03 20:27:20 INFO org.apache.spark.SparkEnv: Registering BlockManagerMaster
23/12/03 20:27:20 INFO org.apache.spark.SparkEnv: Registering BlockManagerMasterHeartbeat
23/12/03 20:27:20 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoordinator
Spark context Web UI available at http://nyu-dataproc-m.c.hpc-dataproc-19b8.internal:44039
Spark context available as 'sc' (master = yarn, app id = application_1691775874963_37625).
Spark session available as 'spark'.
Welcome to

  ____  __
 / ___/  /
/ /   / __ \
/ /   / /_/ /
/_/   /____/

version 3.1.2

Using Scala version 2.12.14 (OpenJDK 64-Bit Server VM, Java 1.8.0_322)
Type in expressions to have them evaluated.
Type :help for more information.

scala> :load merg
merge_analysis.scala   merging.scala   mergedData.csv

scala> :load merge_analysis.scala
Loading merge_analysis.scala...
import org.apache.spark.sql.{SparkSession, DataFrame}
import org.apache.spark.sql.functions._
import org.apache.spark.sql.types._
import org.apache.spark.sql.types.DoubleType
data: org.apache.spark.sql.DataFrame = [Region: string, Year: int ... 4 more fields]
root
 |-- Region: string (nullable = true)
 |-- Year: integer (nullable = true)
 |-- avg(Demand (MW) _imputed): double (nullable = true)
 |-- avg(Net Generation (MW) _imputed): double (nullable = true)
 |-- avg(avg_rate): double (nullable = true)
 |-- EV_adoption_count: integer (nullable = true)

numRows: Long = 75
numCols: Int = 6

```

Top 10 regions having the highest energy consumption across the years

```

top_demand_regions: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [Region: string, Year: int ... 1 more field]
Top 10 regions with the highest average demand:
+-----+-----+-----+
|Region|Year|avg(Demand (MW) _imputed)|
+-----+-----+-----+
| MIDA|2021| 559615.7164383562|
| MIDA|2019| 91390.09310470104|
| MIDA|2020| 87351.92742148535|
| MIDA|2018| 48981.24502844427|
| MIDA|2016| 47488.22854386769|
| MIDA|2017| 45923.939645578954|
| TEX|2021| 44808.24657534246|
| TEX|2019| 43797.57305936073|
| TEX|2020| 43376.729735883426|
| TEX|2018| 42636.09022024867|
+-----+-----+-----+

```

Top 10 regions having the highest energy pricing across the years

```

top_price_regions: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [Region: string, Year: int ... 1 more field]
Top 10 regions with the highest average pricing:
+-----+-----+-----+
|Region|Year|      avg(avg_rate) |
+-----+-----+-----+
| TEN|2020| 0.202728483430992|
| CAL|2021| 0.1737567761221979|
| TEN|2016|0.16835413952230638|
| TEN|2019|0.16222832938662293|
| CAL|2020|0.15483141272907833|
| CAL|2019| 0.1428420371686846|
| CAL|2018|0.13734635540976045|
| TEN|2017|0.13494546208863265|
| CAL|2017| 0.1313309686759356|
| TEN|2018|0.13025406267386225|
+-----+-----+-----+

```

Top 10 regions having the highest ev adoption across the years

```
top_ev_regions: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [Region: string, Year: int ... 1 more field]
Top 10 regions with the highest average EV adoption:
+-----+-----+
|Region|Year|EV_adoption_count|
+-----+-----+
| NW|2021|121898|
| NW|2020|87955|
| NW|2019|76818|
| NW|2018|61737|
| NW|2017|42946|
| NW|2016|31209|
| CAL|2021|104|
| CAL|2020|99|
| CAL|2019|88|
| CAL|2018|63|
+-----+-----+
```

Aggregating data all across the years and finding a general correlation trend between electricity pricing, consumption and ev adoption

```
analytic_result: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [Region: string, avg_demand: double ... 2 more fields]
+-----+-----+-----+-----+
|Region|avg_demand|avg_price|avg_ev_adoption|
+-----+-----+-----+-----+
| TEN|15592.566159036814|0.15970209542048325|2.2|
| CAL|5671.9745510560315|0.14356732095574945|67.33333333333333|
| CENT|14461.844728014437|0.10728290284944526|2.4|
| NE|13717.087565598371|0.09194794762250541|19.0|
| CAR|5856.695100595641|0.0911948631123488|10.166666666666666|
| FLA|3162.8395274984805|0.09055669163481436|19.2|
| SW|4844.460060817105|0.0904126529058501|41.166666666666664|
| NY|17796.549856318252|0.0892468881410048|7.5|
| MIDW|22417.079304592484|0.0799310772538213|24.333333333333332|
| TEX|42489.95071156693|0.07840273718155237|25.833333333333332|
| SE|12519.14516814301|0.07516693940268164|43.166666666666664|
| MIDA|146791.85836373892|0.07128225734640305|18.333333333333332|
| NW|3738.5249890611576|0.06826041980828358|70427.16666666667|
+-----+-----+-----+-----+

+-----+-----+-----+
|Corr_demand_price|Corr_demand_ev|Corr_price_ev|
+-----+-----+-----+
|-0.29023682960290725|-0.156707261829343|-0.29552165749044085|
+-----+-----+-----+
```