ENERGY DEMAND ANALYSIS IN SPAIN

Claire Ndofor, Wes McNall, Scott Howard, Shelby Mohar

GOALS AND OBJECTIVES

MOTIVATION

- Forecasting in energy markets is one exceedingly helpful tool in making the transition to a renewable-based electrical infrastructure

- Our goal is to demonstrate this by leveraging Big Data analysis tools on a dataset that consists of energy usage and weather data for five large cities in Spain

SIGNIFICANCE

- Predict energy usage to increase efficiency of electrical production
- Predict energy price
- Locate areas that would benefit from renewable energies

OBJECTIVES

- Predict energy usage based on the weather
- Predict energy prices by:
  - Time of day
  - Day of the week
  - Time of year
- Analyze the factors that affect the fluctuations in energy usage, as well as the sources of energy

FEATURES

- dt_iso (datetime index localized to CET)
- generation biomass (in MW)
- generation fossil brown coal/lignite (in MW)
- generation fossil coal-derived gas (in MW)
- generation fossil gas (in MW)
- generation fossil hard coal (in MW)
- generation fossil oil (in MW)
- generation fossil oil shale (in MW)
- generation fossil peat (in MW)
- generation geothermal (in MW)
- city_name
- temp (in kelvin)
- temp_min (in kelvin)
- temp_max (in kelvin)
- pressure (in hPa)
- humidity (in %)
- wind_speed (in m/s)
- wind_deg (wind direction)
- rain_1h (rain in last hour in mm)
- ect...

STORYTELLING – CHAPTER 1

WHO?

- This data has the ability to impact every community that uses an electrical grid. Not only is it advantageous at the individual level to be able to predict the cost of an electric bill, but it is also extremely helpful to be able to predict energy usage at a macro level as communities across the globe begin to make the transition to renewable energies in response to climate change.

WHAT?

- We seek to analyze the patterns of weather and energy uses and use them to predict energy demand and price. While the short-term results have the potential to improve 24-hour and hour-by-hour predictions, this work also has the potential to help identify areas that would most benefit from the introduction of renewable energy.

WHEN? / WHERE?

- This data was collected between 2015 - 2019.
- The data is collected from the five largest cities in Spain: Madrid, Barcelona, Valencia, Seville, and Bilbao.

WHY?

- By improving forecasting, we can increase the efficiency of a power grid and help reduce the usage of peak demand on power plants, which are generally less efficient than their counterparts. As stated in the motivation, forecasting in energy markets is an exceedingly helpful tool in making the transition to a renewable-based electrical infrastructure.

CURRENT WORK

HIVE

- Load data into Hive using Cloudera
- Timestamp conversion
- Table merging
- Querying

city_name	temp	temp_min	temp_max
Barcelona	289.84594528884395	288.59273345304103	291.0194106526
Bilbao	286.3761614905216	284.914581781618	288.0341095718221
Madrid	288.0581355640967	286.8221485281031	289.1525875339992
Seville	293.1028779257481	291.1815026537708	295.959789556759
Valencia	290.77789969745476	290.21955934550954	291.3519373417158

5 rows selected (35.883 seconds)

```
+--------------+---------+----------------------+--+
| x.city_name  | x.year  |         temp         |
+--------------+---------+----------------------+--+
|  Barcelona   | 2015    | 289.8164228233073    |
|  Barcelona   | 2016    | 289.9148484257806    |
|  Barcelona   | 2017    | 289.7989752190377    |
|  Barcelona   | 2018    | 289.85331927296517   |
| Bilbao       | 2015    | 286.3070560449736    |
| Bilbao       | 2016    | 286.51417122701736   |
| Bilbao       | 2017    | 286.2761192779457    |
| Bilbao       | 2018    | 286.40737739985383   |
| Madrid       | 2015    | 288.14126680611423   |
| Madrid       | 2016    | 288.0550533660679    |
| Madrid       | 2017    | 288.6491765270033    |
| Madrid       | 2018    | 287.413208712464     |
| Seville      | 2015    | 293.9569442655581    |
| Seville      | 2016    | 294.4828176623545    |
| Seville      | 2017    | 293.36447039732354   |
| Seville      | 2018    | 290.6301364922095    |
| Valencia     | 2015    | 290.03898830741883   |
| Valencia     | 2016    | 290.45791672346394   |
| Valencia     | 2017    | 291.2427910806921    |
| Valencia     | 2018    | 291.3730365067371    |
+--------------+---------+----------------------+--+
20 rows selected (36.517 seconds)
0: jdbc:hive2://>
```

```
SELECT x.city_name, x.year,
AVG(x.temp) as temp

FROM (

SELECT city_name, YEAR(time)
AS year, temp

FROM joined

) x

GROUP BY x.city_name,
x.year;
```

```
+--------------+---------+---------------------+--+
| x.city_name  | x.year  |        temp         |
+--------------+---------+---------------------+--+
|  Barcelona   | 2015    | 289.8164228233073   |
|  Barcelona   | 2016    | 289.9148484257806   |
|  Barcelona   | 2017    | 289.7989752190377   |
|  Barcelona   | 2018    | 289.85331927296517  |
| Bilbao       | 2015    | 286.3070560449736   |
| Bilbao       | 2016    | 286.51417122701736  |
| Bilbao       | 2017    | 286.2761192779457   |
| Bilbao       | 2018    | 286.40737739985383  |
| Madrid       | 2015    | 288.14126680611423  |
| Madrid       | 2016    | 288.0550533660679   |
| Madrid       | 2017    | 288.6491765270033   |
| Madrid       | 2018    | 287.413208712464    |
| Seville      | 2015    | 293.9569442655581   |
| Seville      | 2016    | 294.4828176623545   |
| Seville      | 2017    | 293.36447039732354  |
| Seville      | 2018    | 290.6301364922095   |
| Valencia     | 2015    | 290.03898830741883  |
| Valencia     | 2016    | 290.45791672346394  |
| Valencia     | 2017    | 291.2427910806921   |
| Valencia     | 2018    | 291.3730365067371   |
+--------------+---------+---------------------+--+
20 rows selected (36.517 seconds)
0: jdbc:hive2://>
```

LIMITATIONS

- Solr / Lucene - great for indexing, searching full-text data in a high volume setting
- Our dataset contains numerical data rather than text
- Other tools fit our use case better

FUTURE WORK

FUTURE WORK

- MapReduce – sophisticated query/analysis with custom MapReduce .jar
- Cassandra – build table in Cassandra and run query
- Sqoop – transfer data from HDFS to MySQL and run query

```
mysql> SHOW TABLES;
+--------------------------+
| Tables_in_group_project |
+--------------------------+
| energy                   |
+--------------------------+
1 row in set (0.00 sec)

mysql> DESCRIBE energy;
+-------------------------------------------+-----------+------+-----+--------
-----------+-------------------------------+
| Field                                     | Type      | Null | Key | Default
          | Extra                         |
+-------------------------------------------+-----------+------+-----+--------
-----------+-------------------------------+
| time                                      | timestamp | NO   |     | CURRENT
_TIMESTAMP | on update CURRENT_TIMESTAMP |
| generation_biomass                        | float     | YES  |     | NULL
           |                               |
| generation_fossil_brown_coal_lignite      | float     | YES  |     | NULL
           |                               |
| generation_fossil_coal_derived_gas        | float     | YES  |     | NULL
           |                               |
| generation_fossil_gas                     | float     | YES  |     | NULL
           |                               |
| generation_fossil_hard_coal               | float     | YES  |     | NULL
           |                               |
| generation_fossil_oil                     | float     | YES  |     | NULL
           |                               |
| generation_fossil_oil_shale               | float     | YES  |     | NULL
           |                               |
| generation_fossil_peat                    | float     | YES  |     | NULL
           |                               |
| generation_geothermal                     | float     | YES  |     | NULL
           |                               |
| generation_hydro_pumped_storage_aggregated | float     | YES  |     | NULL
           |                               |
| generation_hydro_pumped_storage_consumption | float     | YES  |     | NULL
           |                               |
| generation_hydro_run_of_river_and_poundage | float     | YES  |     | NULL
```