Electric Fuel Generation and Consumption

GROUP 3

Tomoki Kyotani (002986248)
Sai Bhavana Atluri (001594085)
Jignasuben Rohitbhai Vekariya (002981797)
Bhavya Maheshwari (002116459)

Agenda

1. Introduction / Questions to answer
2. Software Requirements
3. Data Sources
4. Data Model
5. Data Flow
6. Data Preprocessing
7. Data Generation
8. Initial Load, Incremental Load, and SCD
9. OLAP Cube
10. Visualizations
11. Answers to Questions

1. Introduction

Nowadays the population is increasing day by day, use of energy sources and pollution is also increasing rapidly which affects weather as well. We U.S.decided to build a warehouse with respective fields for more effective analysis on how energy source generation and consumption is affected by population, pollution, and weather in the

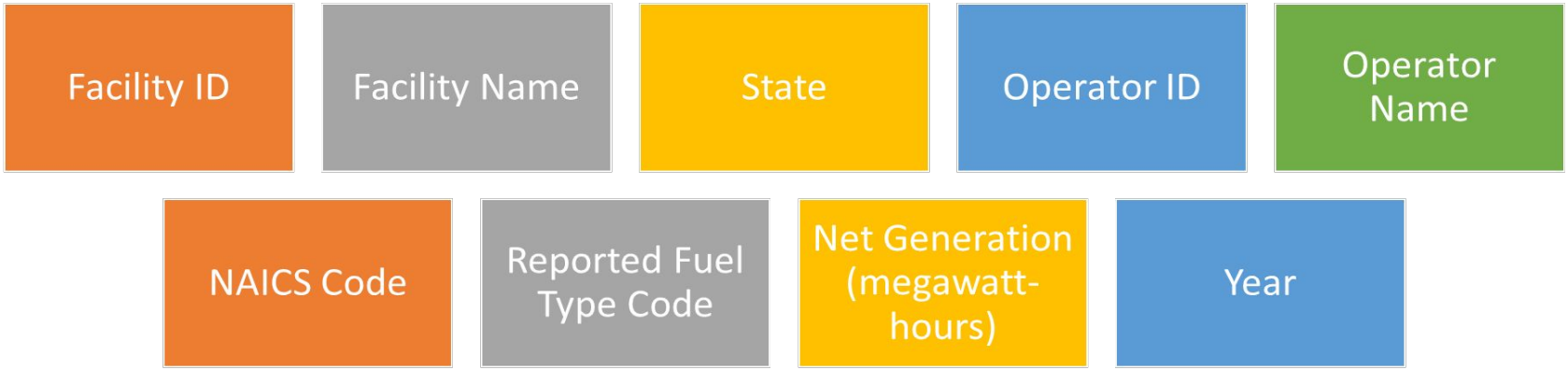
2. Software Requirement

- SSIS : ETL & Data Warehouse
- SSAS: OLAP Cube
- Tableau: Visualization

3. Data Sources

Electric Power Data

Data Source: https://www.eia.gov/electricity/data/eia923/



Air Pollution

Data Source: https://www.kaggle.com/datasets/sogun3/uspollution
scraped from the United States Environmental Protection Agency

State	Date Local	NO2 Mean	NO2 Air Quality Index	O3 Mean
O3 Air Quality Index	SO2 Mean	SO2 Air Quality Index	CO Mean	CO Air Quality Index

Population

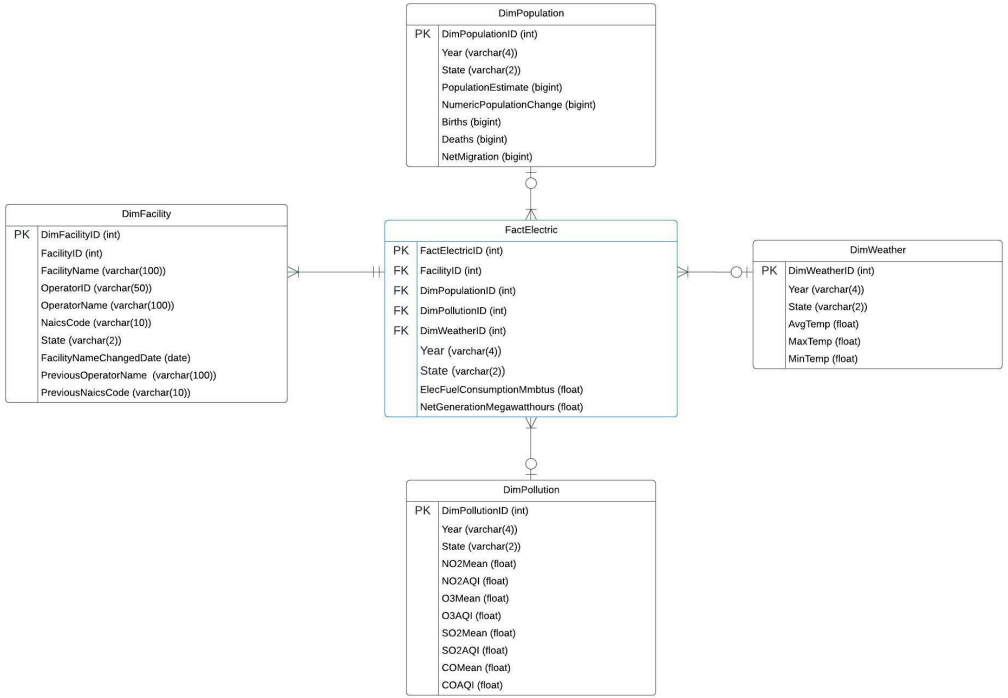
Data Source: https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-total.html#par_textimage_500989927

Name	POPESTIMATE201X	NPOPCHG201X
BIRTHS201X	DEATHS201X	NETMIG201X

Weather

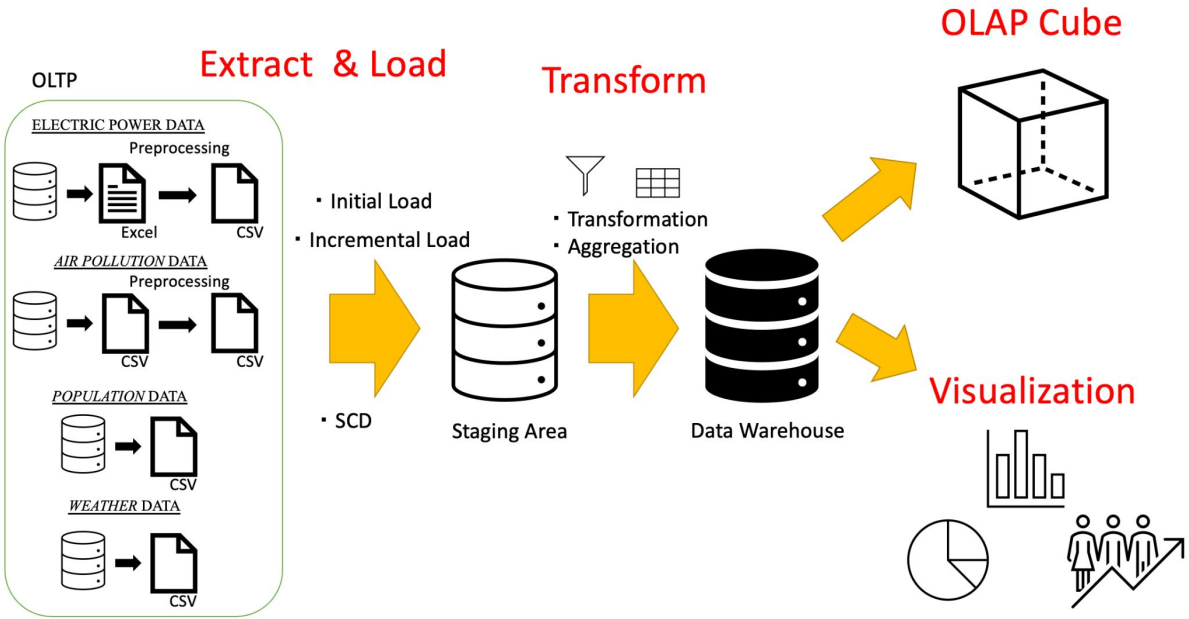
Data Source: https://data.world/mattwinter225/2015-usa-weather-avg-max-min

4. Data Model



1. In our data warehouse, the fact table is “FactElectric” which stores facts about electric power consumption and generation.
2. The dimension tables are “DimFacility”, “DimPopulation”, “DimPollution”, and “DimWeather”.
3. We choose state and year as our dimensions for the OLAP Cube and NO2Mean, NO2AQI, O3Mean, O3AQi, COMean, COAQL, SO2Mean, SO2AQI, AvgTemp, MaxTemp, MinTemp, ElecFuelConsumptionMmbtus, NetGenerationMegawatthours etc as our measures. This was all done to see the trends in the data we used.

5. Data Flow

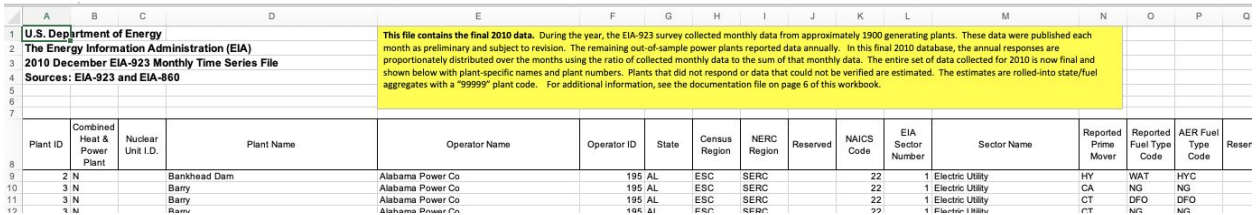


- 2010-2016 for the Initial Load and 2017 for the Incremental Load
- SCD is on the Facility dimension (DimFacility)
  - FacilityName, OperatorID, OperatorName, NaicsCode

6. Data Preprocessing

Electric Power Data

- Remove the first 5~7 rows
- Remove the aggregated records (at state/fuel type level) for the FacilityName="State-Fuel Level Increment" & FacilityID=99999
- Rename column names
- Replace "." values with an appropriate year for the Year column
- Done in Python



U.S. Department of Energy
The Energy Information Administration (EIA)
2010 December EIA-923 Monthly Time Series File
Sources: EIA-923 and EIA-860

This file contains the final 2010 data. During the year, the EIA-923 survey collected monthly data from approximately 1900 generating plants. These data were published each month as preliminary and subject to revision. The remaining out-of-sample power plants reported data annually. In this final 2010 database, the annual responses are proportionately distributed over the months using the ratio of collected monthly data to the sum of that monthly data. The entire set of data collected for 2010 is now final and shown below with plant-specific names and plant numbers. Plants that did not respond or data that could not be verified are estimated. The estimates are rolled-into state/fuel aggregates with a "99999" plant code. For additional information, see the documentation file on page 6 of this workbook.

Plant ID	Combined Heat & Power Plant	Nuclear Unit I.D.	Plant Name	Operator Name	Operator ID	State	Census Region	NERC Region	Reserved	NAICS Code	EIA Sector Number	Sector Name	Reported Prime Mover	Reported Fuel Type Code	AER Fuel Type Code	Reser
2	N		Bankhead Dam	Alabama Power Co	195	AL	ESC	SERC		22	1	Electric Utility	HY	WAT	HYC	
3	N		Barry	Alabama Power Co	195	AL	ESC	SERC		22	1	Electric Utility	CA	NG	NG	
3	N		Barry	Alabama Power Co	195	AL	ESC	SERC		22	1	Electric Utility	CT	DFO	DFO	

6. Data Preprocessing

Air Pollution Data

- Put double quotes around non-numeric values/columns
  - Issue: Columns are separated by a comma, and values are not quoted in the original data files. However, some records contain a comma in their Address column.
- Done in Python
- Specify a double quote “ for the Text Qualifier in Flat File Source in SSIS

7. Data Generation

Initial Load - Weather Data

- Found only the 2015 data for all states in the U.S.
- Generate 2010-2014 and 2016 data for initial load and 2017 data for incremental load
- Add random values between -3 and 3 to the AvgTemp, MaxTemp, and MinTemp from 2015 data

7. Data Generation

2017 Incremental Load

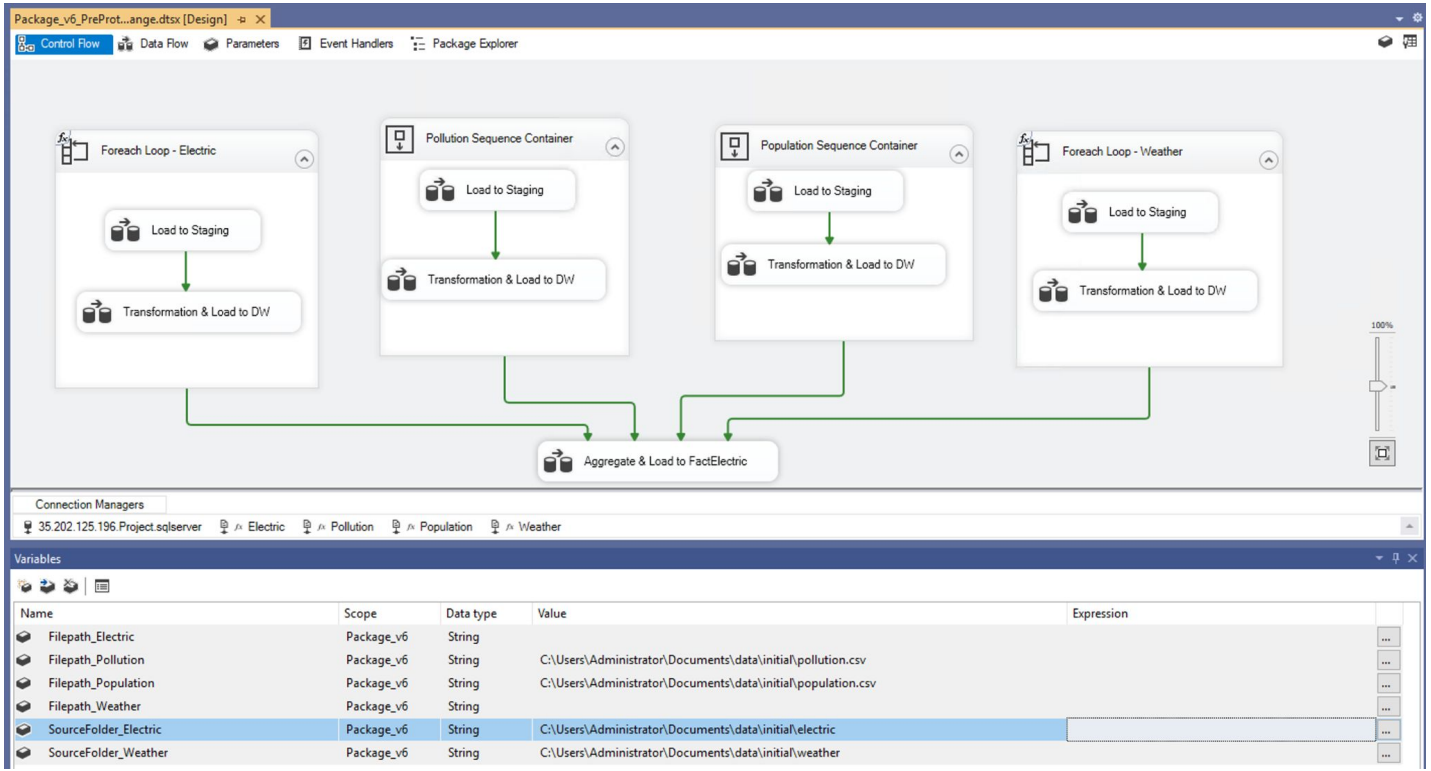
- Weather Data
- Air Pollution Data
  - Add random values between 0 and 1 to NO2Mean, NO2AQI, O3Mean, O3AQI, SO2Mean, SO2AQI, COMean, COAQI from 2016 data
- 2017 Electric Power Data and 2017 Population Data are provided from the original data sources.

7. Data Generation

Slowly Changing Dimension (SCD)

- DimFacility
  - FacilityName : Type 2
  - OperatorID: Type 1
  - OperatorName & NaicsCode: Type 3
- There are already value changes on these columns as we load data from 2010 to 2017
- For the testing purpose, selected and modified 5 records from the 2017 Electric Power Data

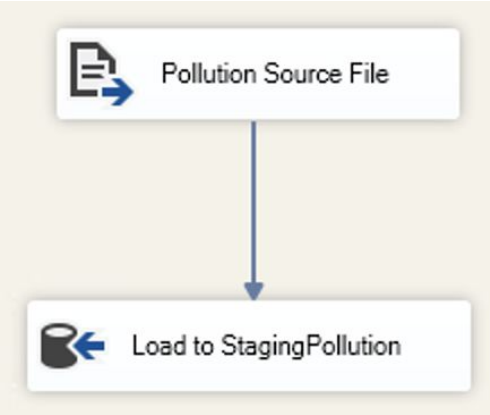
8. Initial Load, Incremental Load, and SCD



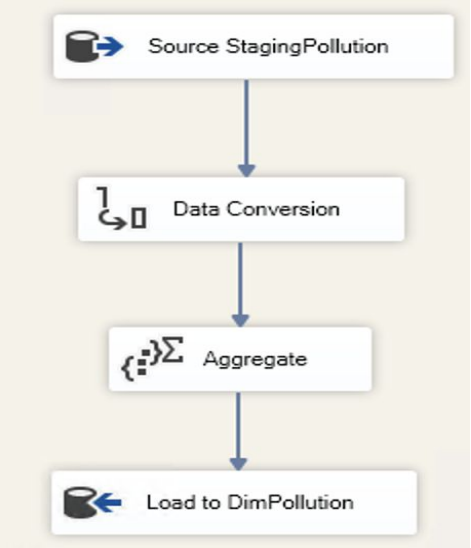
8. Initial Load, Incremental Load, and SCD

Air Pollution, Population, and Weather Data

"Load to Staging" Data Flow



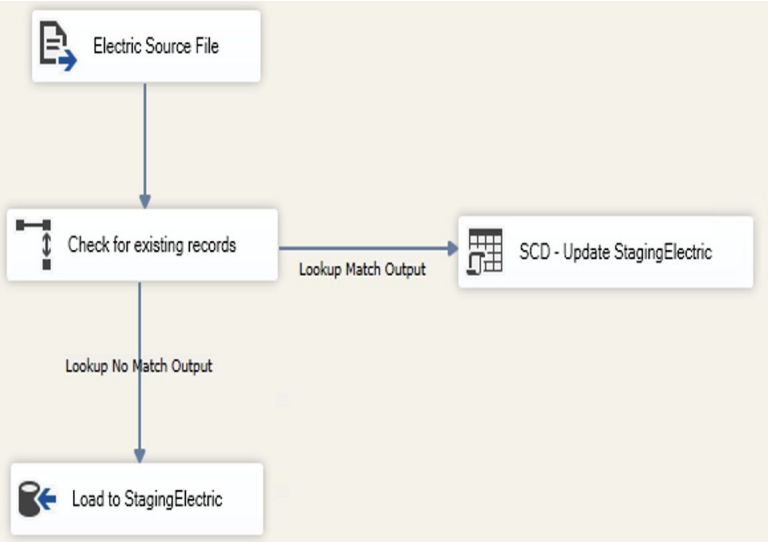
"Transformation & Load to DW" Data Flow



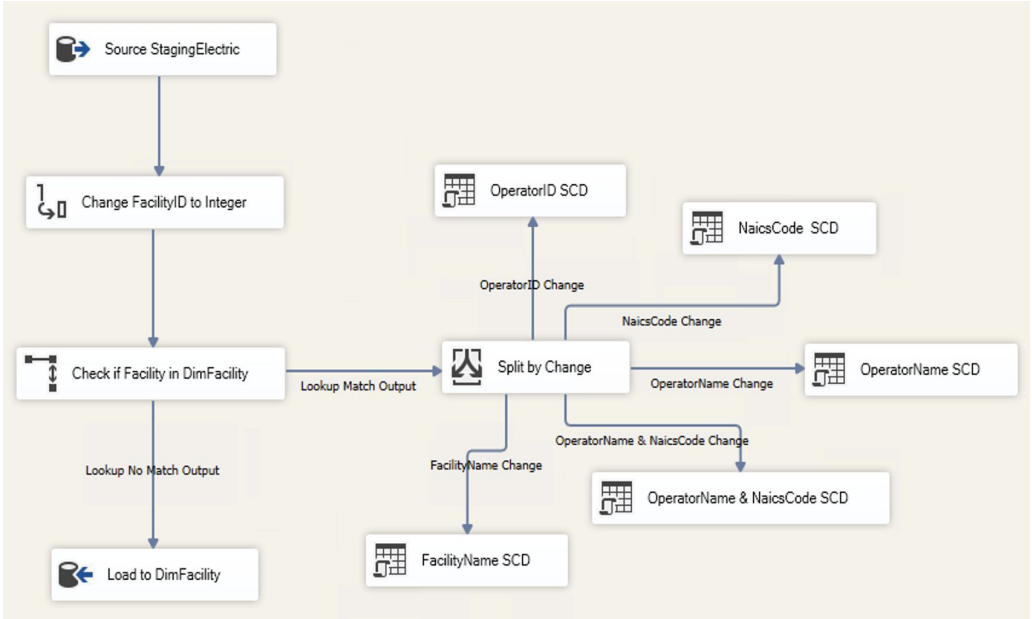
8. Initial Load, Incremental Load, and SCD

Electric Power Data

"Load to Staging" Data Flow

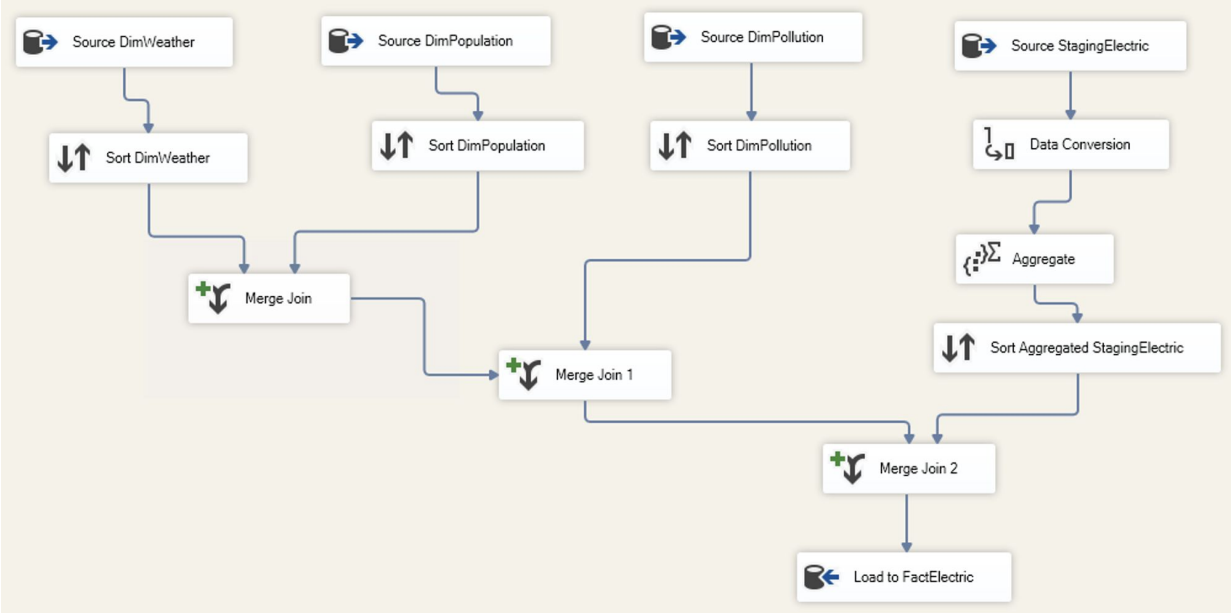


"Transformation & Load to DW" Data Flow



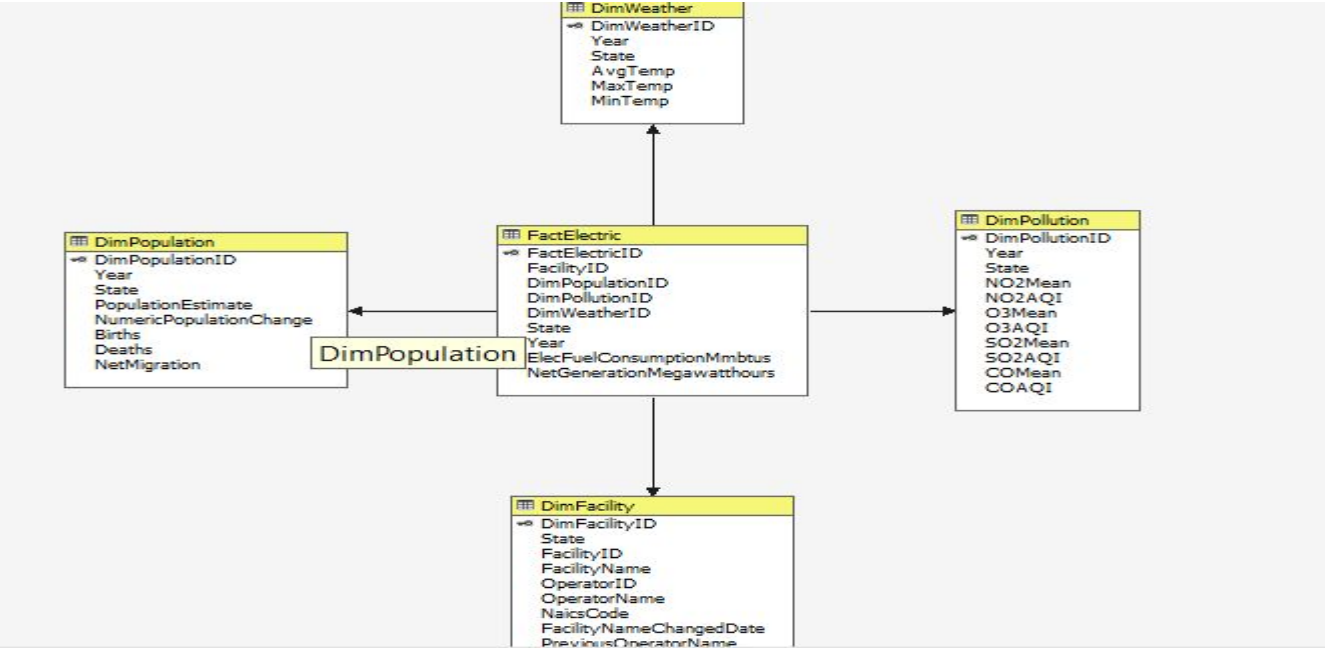
8. Initial Load, Incremental Load, and SCD

"Aggregate & Load to FactElectric" Data Flow



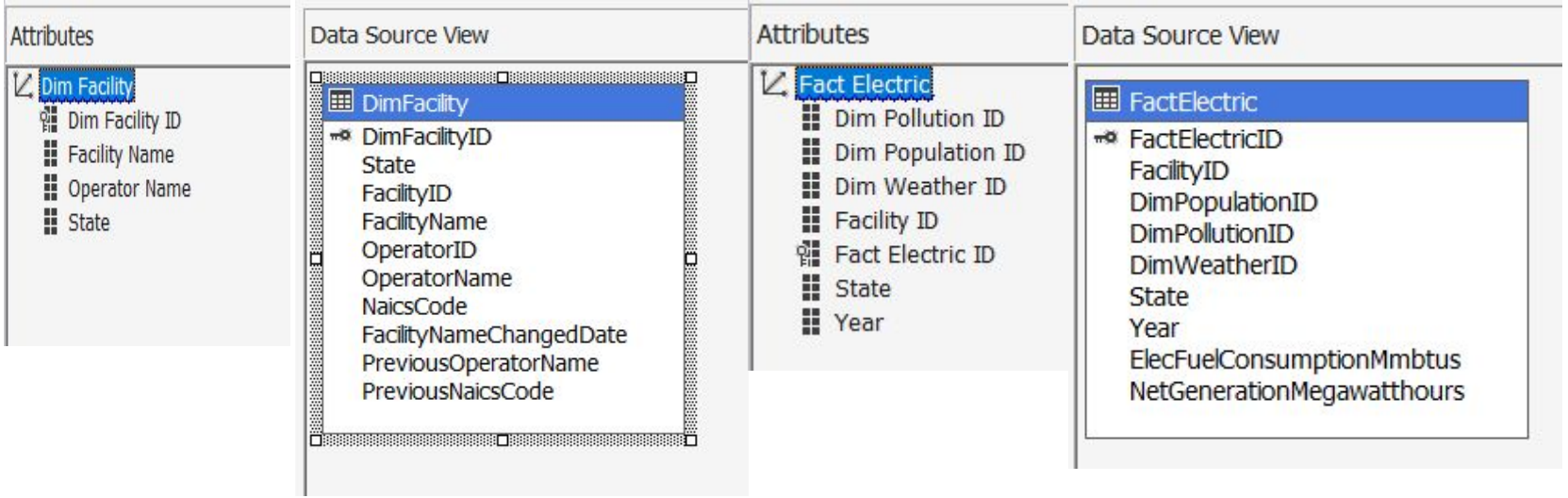
9. Storing the Data in OLAP System (SSAS)

Creating a Data Source and the Data Source View for the Cube.



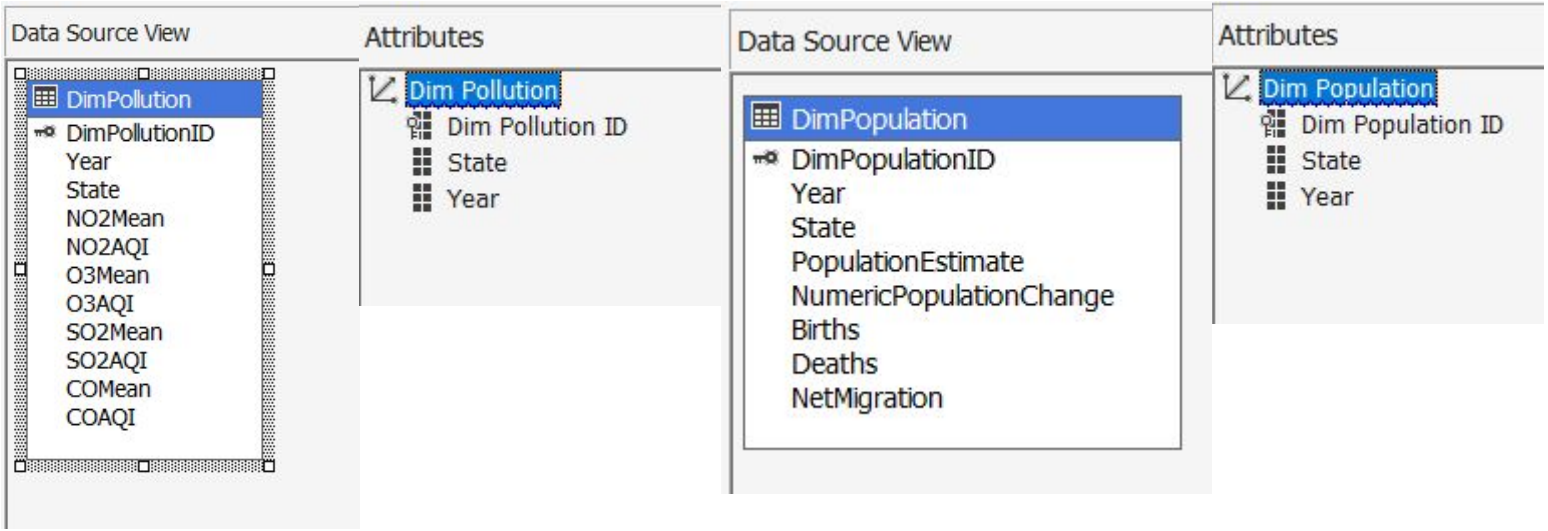
9. Storing the Data in OLAP System (SSAS)

Creating the dimensions for DimFacility and Fact Electric.



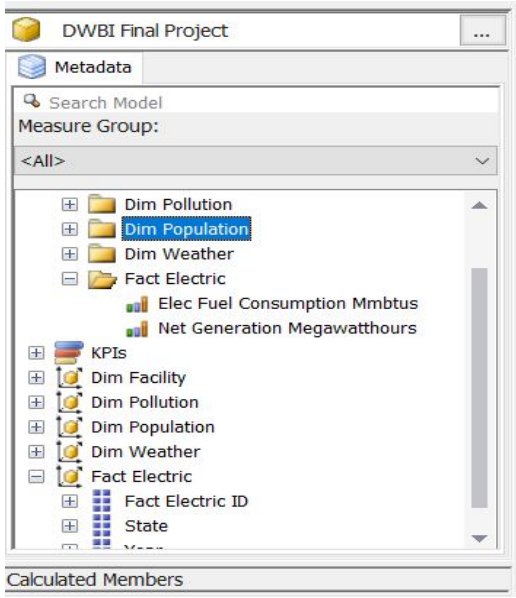
9. Storing the Data in OLAP System (SSAS)

Creating the dimensions for DimPollution and DimPopulation.



9. Storing the Data in OLAP System (SSAS)

Building, processing and deploying the Cube with the dimensions created.

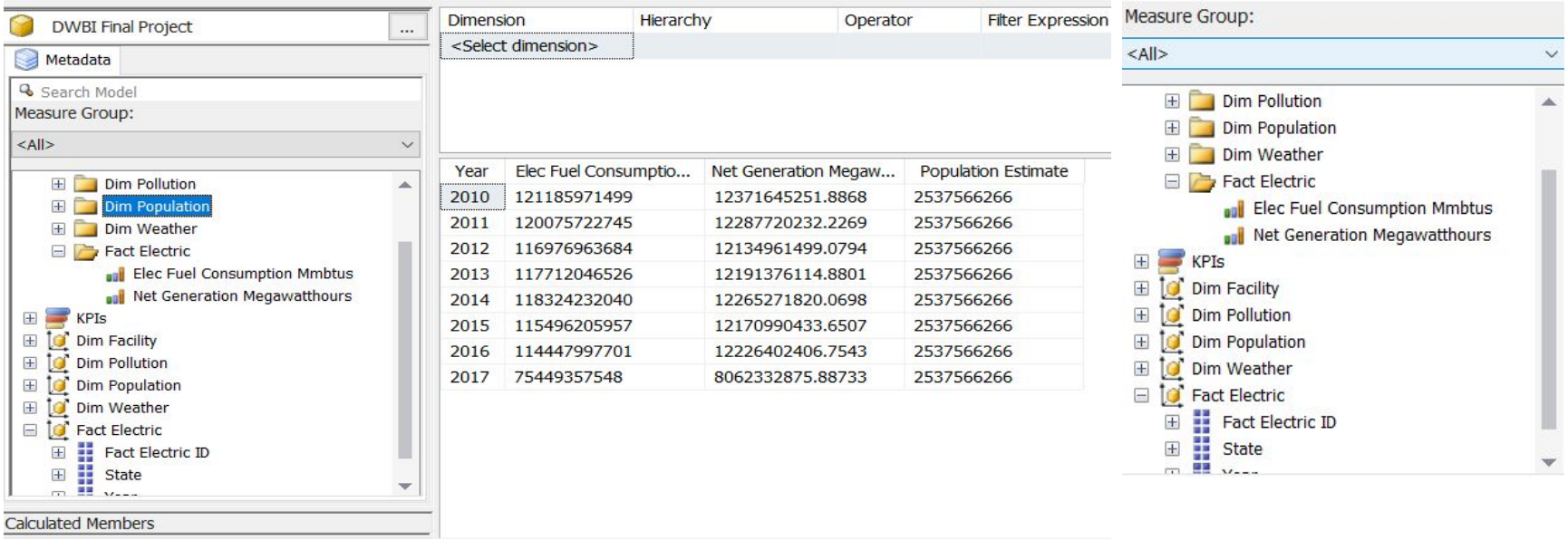


Dimension	Hierarchy	Operator	Filter Expression
<Select dimension>			

Year	Elec Fuel Consumptio...	Net Generation Megaw...	Population Estimate
2010	121185971499	12371645251.8868	2537566266
2011	120075722745	12287720232.2269	2537566266
2012	116976963684	12134961499.0794	2537566266
2013	117712046526	12191376114.8801	2537566266
2014	118324232040	12265271820.0698	2537566266
2015	115496205957	12170990433.6507	2537566266
2016	114447997701	12226402406.7543	2537566266
2017	75449357548	8062332875.88733	2537566266

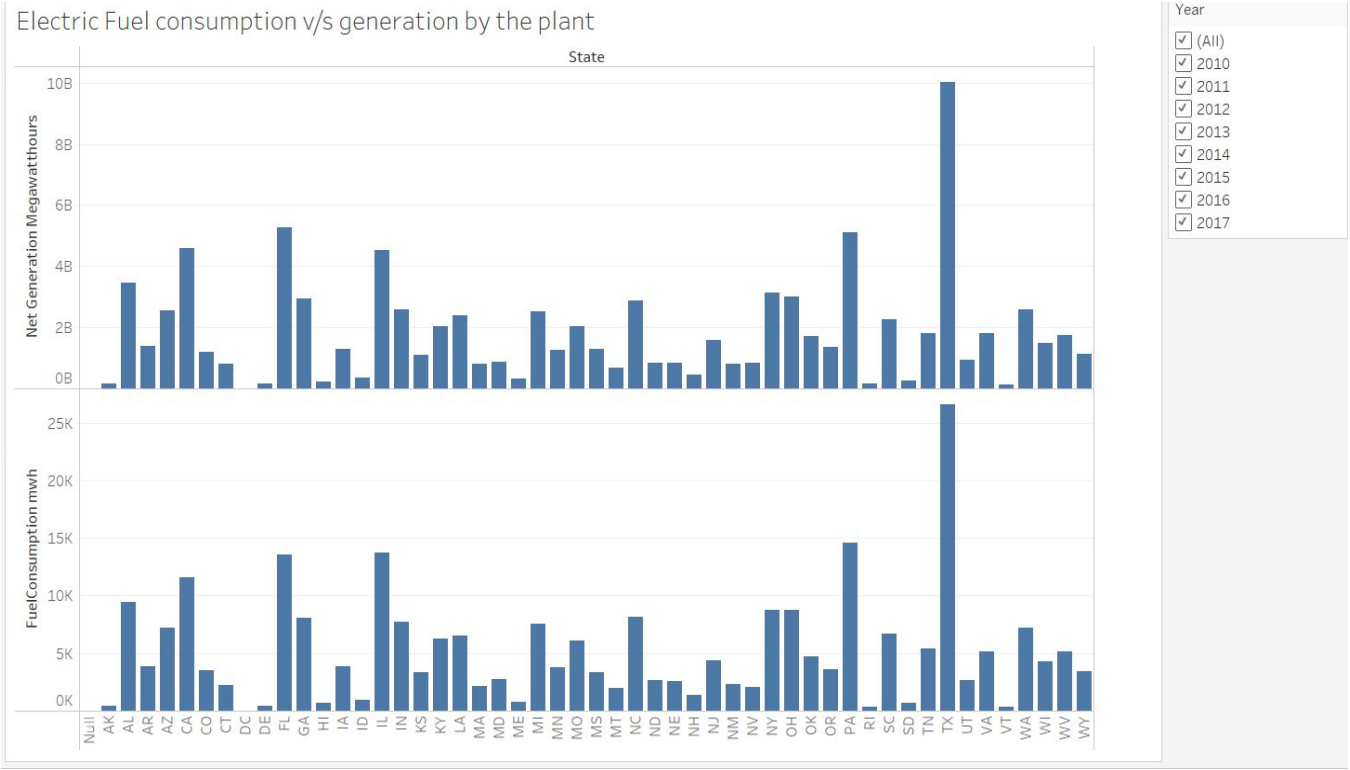
9. Storing the Data in OLAP System (SSAS)

Our Cube!!.

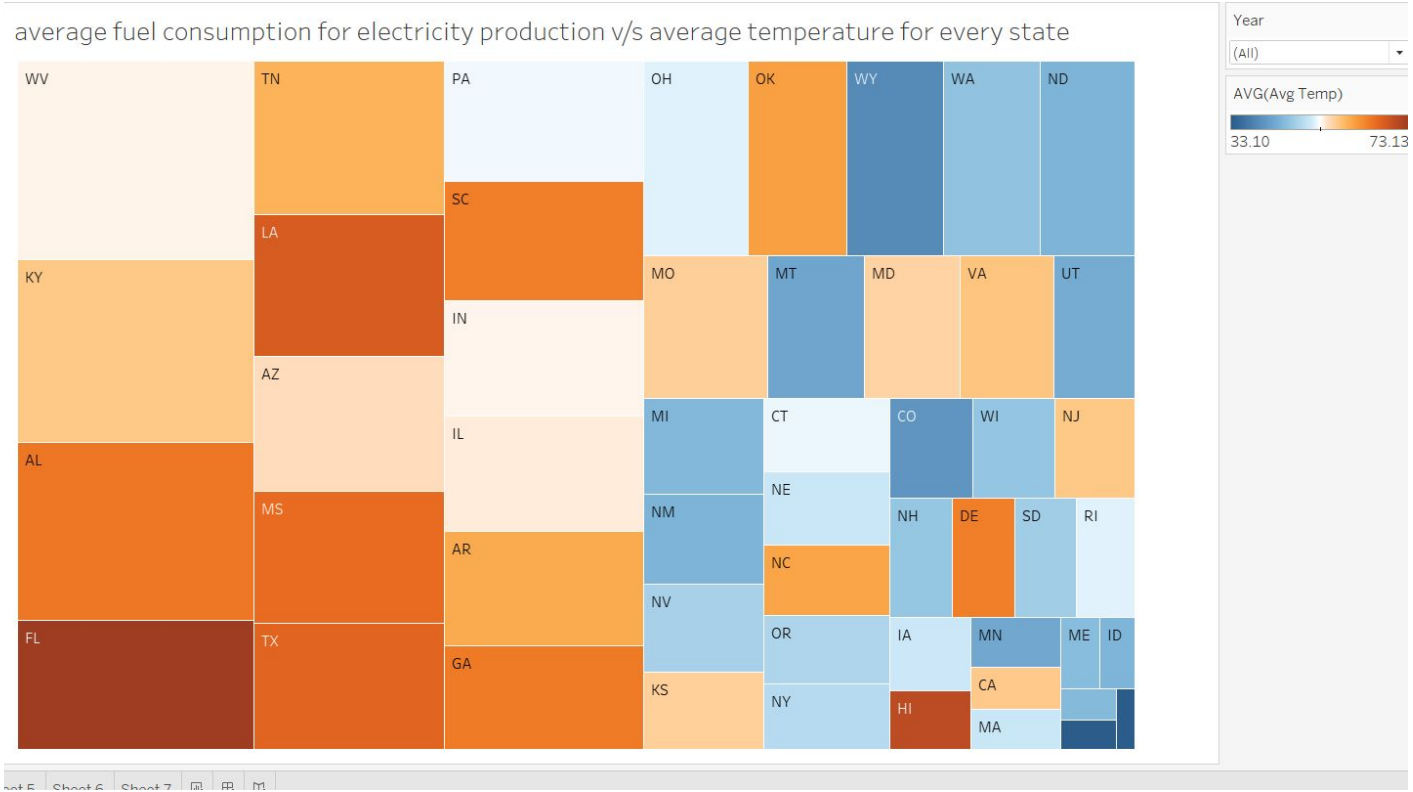


10. Visualization

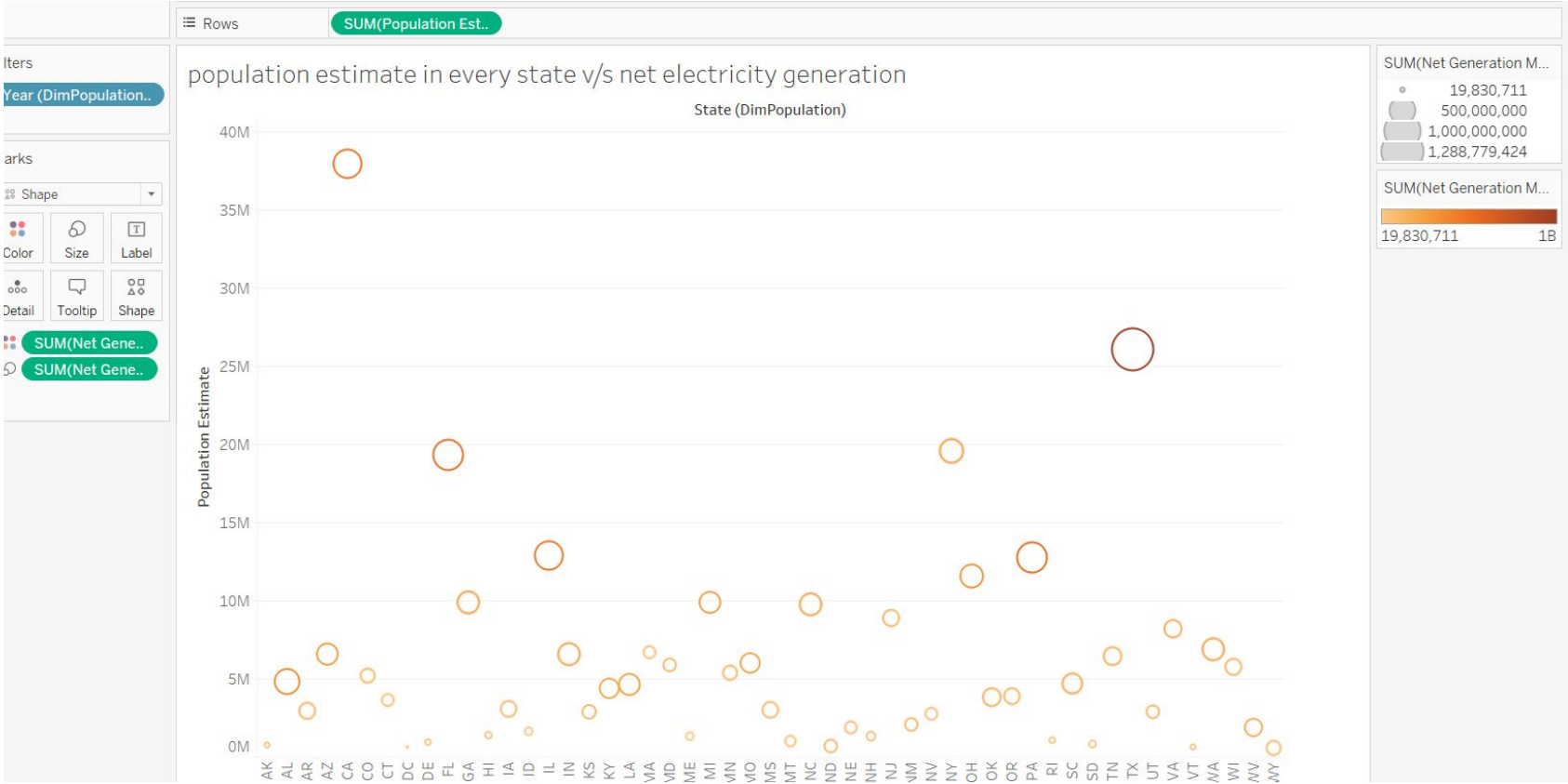
1.Electricity generation v/s Consumption



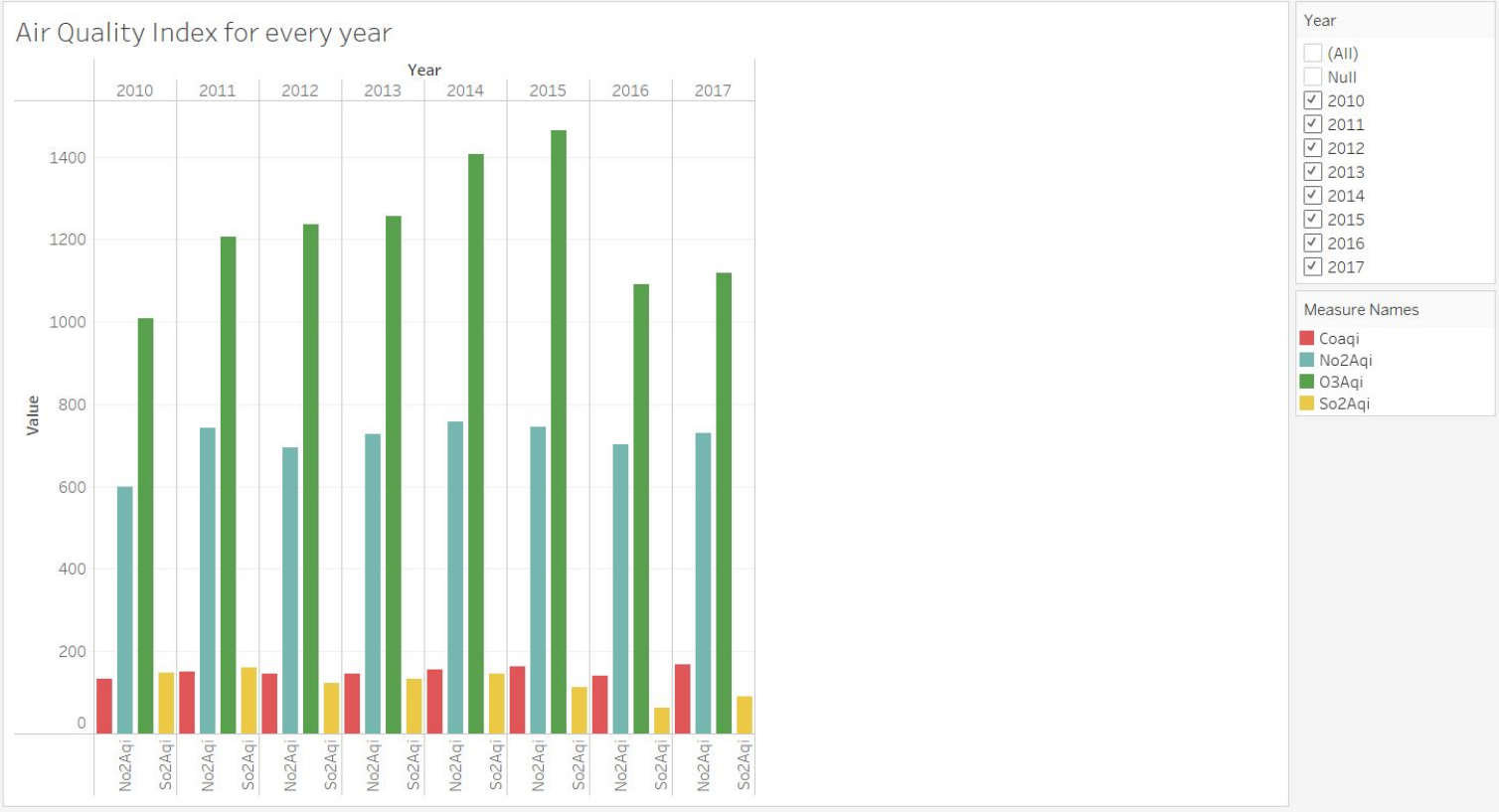
2. Fuel Consumption for electricity generation VS Temperature



3. Population VS Net generation

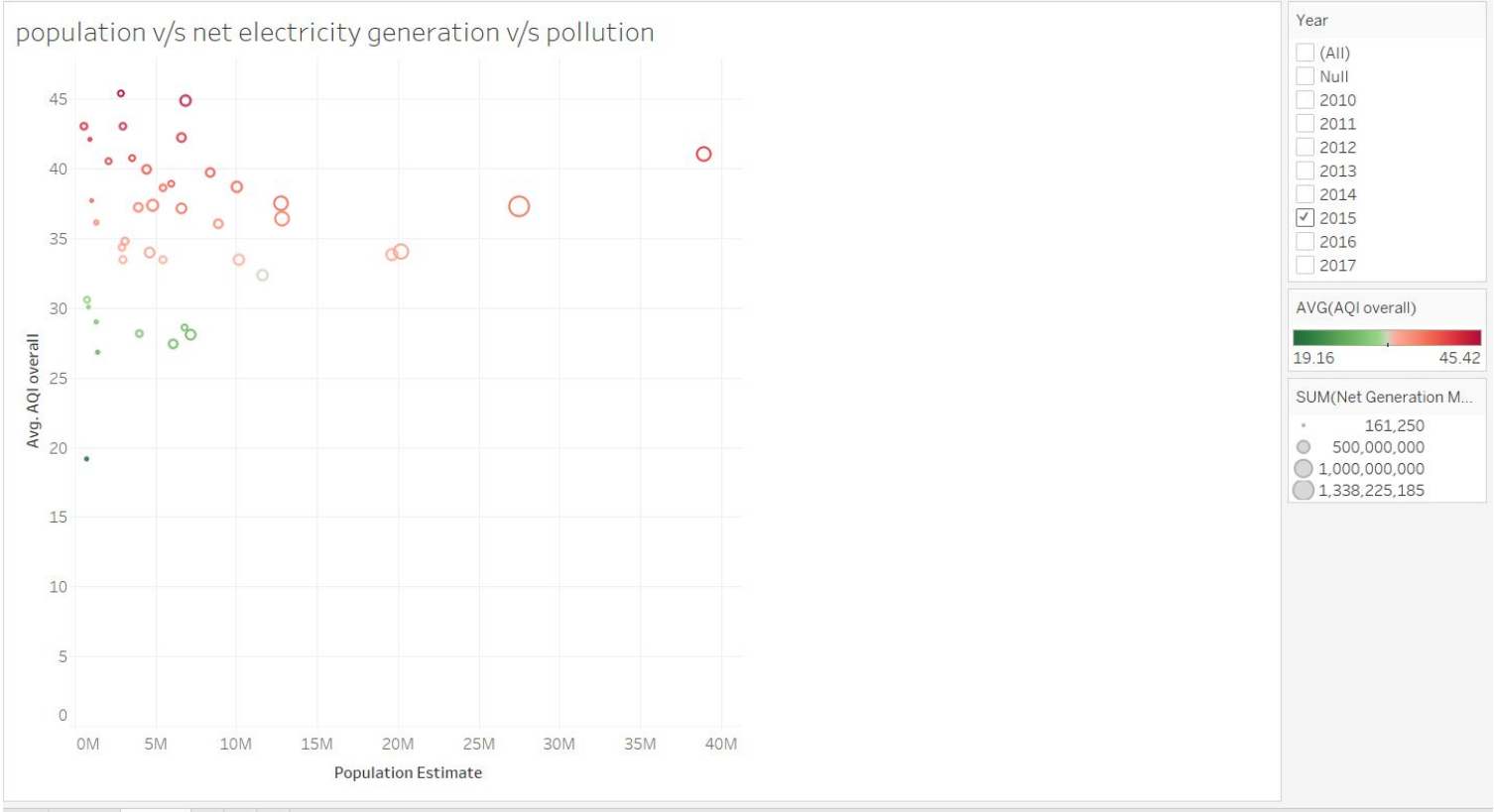


4. Air Quality Analysis

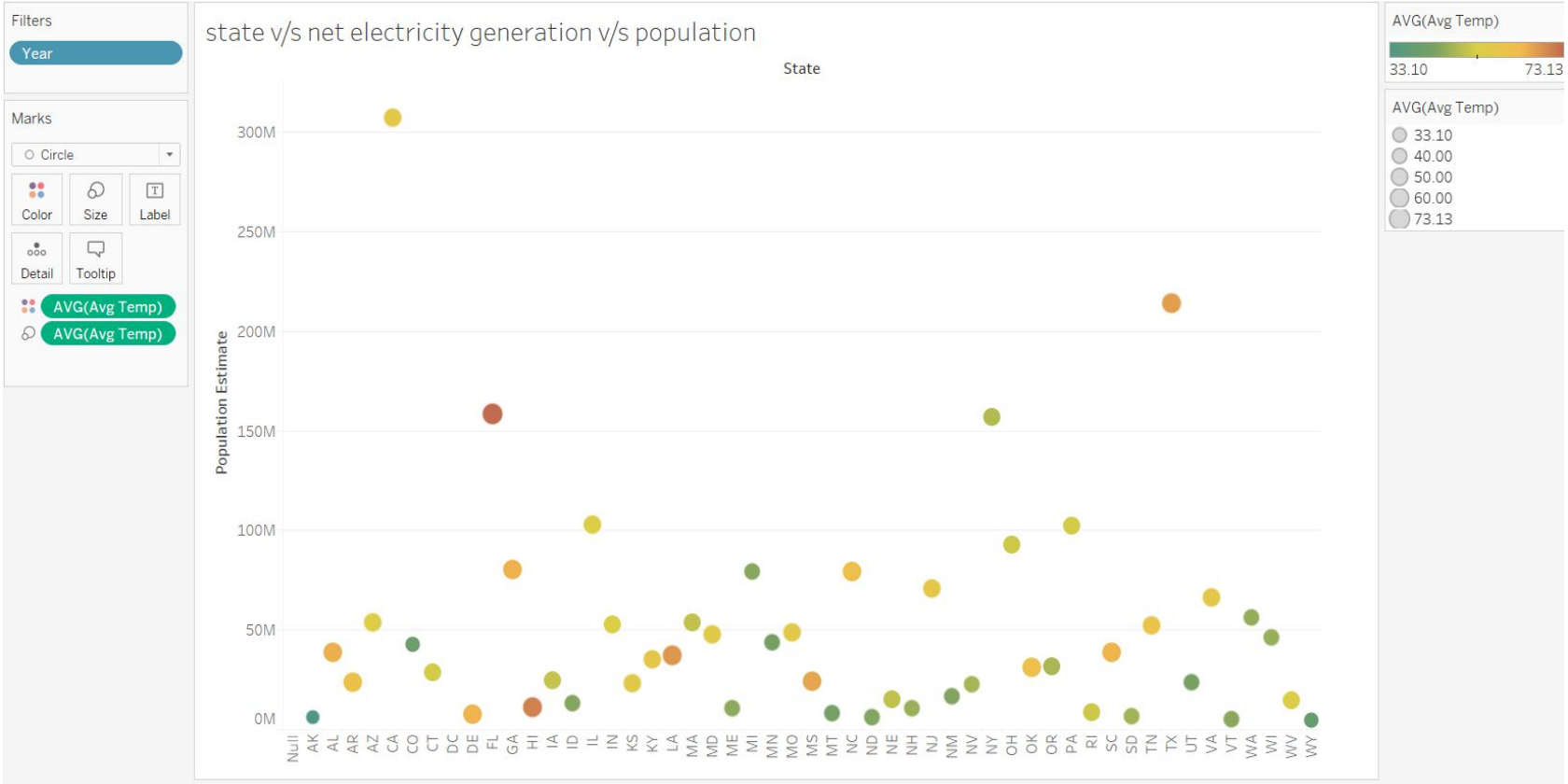


5. Pollution VS Temperature

6. Correlation between pollution, population and fuel consumption



7. Population V/S Electricity generation



11. Answers to Questions

1. What kind of relationship/impact does Electric Power generation and fuel consumption have with various pollutions?

2. What kind of relationship/impact does Population have on Electric Power generation and fuel consumption?

3. What kind of relationship/impact does Temperature have on Electric Power generation and fuel consumption?

ANSWERS TO THE 1st QUESTION

Pages
Columns
State (DimPolluti..
Rows
AVG(AQI overall)
AVG(AQI overall)
Filters
Year
Marks
All
AVG(AQI overall)
Circle
Color
Size
Label
Detail
Tooltip
AVG(Avg Temp)
SUM(Net Gene..
AVG(AQI overall...
pollution v/s temperature
State (DimPollution)
5
0
Null AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA MA MD ME MI MN MO NC ND NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA WA WI WY
SUM(Net Generation M...
36,946,282
5,000,000,000
10,000,000,000
15,000,000,000
18,581,513,593
AVG(Avg Temp)
33.11
73.13

ANSWERS TO THE 2nd and 3rd QUESTIONS USING THE OLAP SYSTEM

Year	Population Estimate	Elec Fuel Consumptio...	Net Generation Megaw...	Avg Temp
2010	309321666	121178154660	12371045677.8868	21254.29...
2011	311556874	120067705431	12287117610.3441	21254.29...
2012	313830990	116973398052	12134746138.7357	21254.29...
2013	315993715	117709292109	12191178558.6457	21254.29...
2014	318301008	118321878033	12265068983.601	21254.29...
2015	320635163	115492978923	12170829183.6482	21254.29...
2016	322941311	114444402549	12226172984.7506	21254.29...
2017	324985539	75447167722	8062199133.88733	21254.29...

State	Population Estimate	Elec Fuel Consumptio...	Net Generation Megaw...	Avg Temp
CO	2537566266	11993079303	1210830183.1181	21254.29...
CT	2537566266	7615868645	806278230.56193	21254.29...
DC	2537566266	33520221	2342362.93593025	21254.29...
DE	2537566266	1584590439	173247678.180069	21254.29...
FL	2537566266	46256822361	5267306397.47278	21254.29...
GA	2537566266	27399341989	2931683990.76633	21254.29...
HI	2537566266	2305952459	236329235.473618	21254.29...
IA	2537566266	13066696413	1300623992.73295	21254.29...
ID	2537566266	3237105122	351715897.376895	21254.29...
IL	2537566266	46917415217	4521351545.34155	21254.29...
IN	2537566266	26205147055	2577977248.3733	21254.29...
KS	2537566266	11543412467	1088540100.16482	21254.29...
KY	2537566266	21338032460	2033624844.71702	21254.29...
LA	2537566266	22225838125	2391195432.9133	21254.29...