**Individual Contribution Report - Person 1**

I worked mainly on this project developing database schema in SQL Server Management Studio (SSMS) and designing and implementing the ETL (Extract, Transform, Load) process using SQL Server Integration Services (SSIS).

I’ve built the basic structure of the data warehouse, and designed the Key Performance Indicators\_Fact table as well as the dimension tables such as Time\_Dimension, Location\_Dimension and Indicator\_Dimension that make it possible. I wrote the SQL scripts to make these tables in SSMS and set them up in a star schema so they would be optimized for analytical querying.

I was sent to work on ETL packages in SSIS. The list went on:

* Collecting raw data from multiple sources (EPA, NYC Open Data).
* Data manipulation: data cleaning; standardizing formats; dealing with nulls; data type conversions.
* And putting the clean data into the correct fact and dimension tables in your warehouse. To handle the complexity of transformation, I set up data flow tasks containing lookup transformations, derived columns and data conversion components.

And I added more error handling so that these packages will not fail if they hit unexpected data, like the little bombs they are. I also wrote data validation queries in SSMS to maintain referential integrity and to check for duplicate or missing foreign key mappings.

Dimensional modeling, SQL scripting and ETL process automation are all skills this work had enhanced. Challenges included dealing with different formats in a variety of data sets and optimizing the ETL — extract, transform and load — process so it would move data efficiently, even when it was immense in size.

That experience greatly enhanced my technical skills and my ability to architect scalable data solutions for business intelligence.