CENTENNIAL COLLEGE

Exploring the Relationship Between Environmental Quality and Housing prices in Canada: An ETL study in data science

Huaye Zhan[1], Aziz Syed[1], Pei-Chi Tseng[1], Aniko Lee Dominic Espeleta[1]

[1] Artificial Intelligence - Software Engineering and Technology, Centennial College

ORACLE®

SQL Developer

Group 5 Team Members

1. Huaye Zhan (301324797)
2. Aziz Syed (301322427)
3. Pei-Chi Tseng (301341444)
4. Aniko Lee Dominic Espeleta (301294763)

Content

1. Project Overview
2. Datasets
3. ETL Process
4. How will it be used?
5. How will it be maintained?

Project Overview

Purpose: Analyze the impact of environmental factors, particularly toxic air emissions, on housing prices in Canada.

Approach: Integrate datasets from Environment and Climate Change Canada (ECCC) and Statistics Canada.

Goal: Provide valuable insights for real estate investors and environmental advocates.

Business Problem:

Real estate prices are influenced by a variety of factors, including location, economic conditions, and environmental quality. Meaning there is a growing need to understand:

How do environmental factors influence housing prices in Canada?

Data Sets

To solve this question, we intended to integrate 3 Datasets from Open Canada:

- New housing price index
- Greenhouse Gas Emissions
- Air pollutant emissions - Total volatile organic compound emissions

Data Sets - New housing price index

Dataset: New Housing Price Index

Provides percentage-based values of houses to track increases or decreases in value.

Offers detailed analysis of the value of just the land, just the house, or the combined value of both.

REF_DATE	GEO	DGUID	New hous	UOM	UOM_ID	SCALAR_F	SCALAR_ID	VECTOR	COORDIN	VALUE	STATUS	SYMBOL	TERMINAT	DECIMALS
Jun-00	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	57.8				1
Jul-00	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	57.9				1
Aug-00	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	58.1				1
Sep-00	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	58.2				1
Oct-00	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	58.3				1
Nov-00	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	58.5				1
Dec-00	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	58.6				1
Jan-01	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	58.7				1
Feb-01	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	58.8				1
Mar-01	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	59				1
Apr-01	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	59.2				1
May-01	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	59.3				1
Jun-01	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	59.5				1
Jul-01	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	59.6				1
Aug-01	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	59.7				1
Sep-01	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	59.9				1
Oct-01	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	59.9				1
Nov-01	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	60.1				1
Dec-01	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	60.2				1
Jan-02	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	60.4				1
Feb-02	Canada	2016A000(	Total (hou	Index, 201	347	units	0	v11195544	1.1	60.8				1

Data Dictionary

VARIABLE	DATA TYPE	DESCRIPTION
REF_DATE	DATE	Date of Reference
GEO	STRING	Location of Reference
DGUID	STRING	Unique identifier for geographical unit used for data
NEW HOUSING INDEX	STRING	Types of Housing
UOM	STRING	Unit of Measure
UOM_ID	INTEGER	ID associated to Unit of Measure
SCALAR_FACTOR	STRING	A factor used for scaling
SCALAR_ID	INTEGER	ID associated with scalar factor
VECTOR	STRING	Vector identifier representing group of data
COORDINATE	DOUBLE	Identifier related to the data series
VALUE	DOUBLE	Percentage depicting rise or drop of housing value
STATUS	STRING	Status of Data
SYMBOL		
TERMINATED	STRING	Represents if data is to be terminated
DECIMALS	DOUBLE	Decimal Places being used

Data Sets - Greenhouse Gas Emissions

Dataset: Greenhouse Gas Emissions

Provides total GHG emissions in megatonnes from 1990 to 2022.

Year	Total greenhouse gas emissions (megatonnes of carbon dioxide equivalent)
1990	608
1991	603
1992	622
1993	627
1994	649
1995	668
1996	689
1997	706
1998	712
1999	723
2000	748
2001	739
2002	745
2003	763
2004	767
2005	761
2006	757
2007	777

VARIABLE	DATA TYPE	DESCRIPTION
YEAR	DATE	Year of Reference
TOTAL GREENHOUSE GAS EMISSIONS	INTEGER	Amount of GHG emissions in megatonnes

Data Sets - Total volatile organic compound emissions

Dataset: Total Volatile Organic Compound (VOC) Emissions

This Data set gives us access to the total amount of volatile emissions, and also gives us the source of what makes the total. This data is from 1990 to 2022.

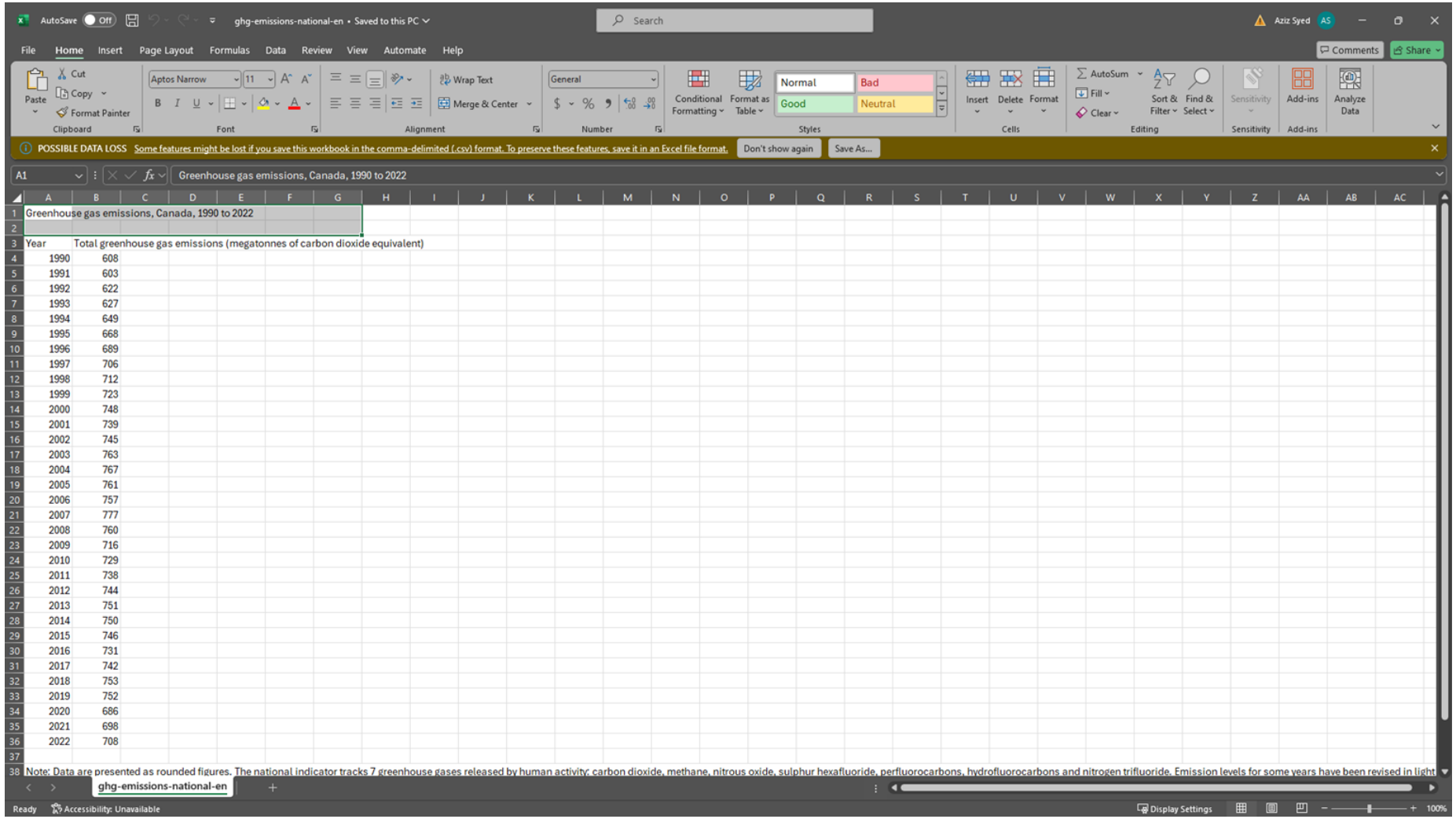
Year	Oil and gas	Paints and	Other sour	Off-road ve	Manufactu	Home firev	Agriculture	Transporta	Total emissions (emissions in kilotonnes)
1990	598.6	357.5	147.8	287.5	257	188.3	103.6	346.8	2287.2
1991	594.1	350	149	287.6	253.7	189.5	103.6	335.4	2262.9
1992	607.9	352.2	133.8	296.8	254.6	199.2	105.3	345.4	2295.2
1993	635.9	351.4	157	305.5	271.1	200	105.1	362.3	2388.3
1994	653.2	362.1	153.3	322.6	273	194.3	107.5	384.1	2450.1
1995	664.6	375.2	148.3	382.9	262.8	189.6	111.5	376.9	2511.7
1996	692.7	373	137.8	410.1	262.2	193.7	114.5	380.2	2564.3
1997	679.7	374.9	126	440.5	257.4	191.1	115.2	373.5	2558.3
1998	691.9	377.5	127.2	467.7	261.5	153.1	115.6	366.4	2560.9
1999	644.4	389.5	120.5	491.2	259.3	148.1	116	351.4	2520.3
2000	658.6	395.9	117.4	489.7	254.5	149.7	116.9	330.2	2512.8
2001	662.8	375.7	112.9	495.6	229.5	128.8	119.9	309.9	2435.1
2002	673.9	366.9	108.3	498	232.8	124.1	121	290.1	2415.2
2003	666.5	368.8	120.5	483.2	216	116.5	120.6	263.6	2355.7
2004	645.6	364.3	108.5	471.8	200.9	123.4	124.5	243.9	2282.9
2005	641.9	441	105.9	449.2	187.1	126.8	125.6	215.9	2293.3
2006	626.6	417.1	96.6	418.4	165.2	124	123	198.2	2169.2
2007	623.2	409.4	93.1	376	149.7	142.9	120.9	188.4	2103.7

VARIABLE	DATA TYPE	DESCRIPTION
YEAR	DATE	Year of Reference
OIL AND GAS	DOUBLE	Emissions from this source
PAINTS AND SOLVENTS	DOUBLE	Emissions from this source
OTHER SOURCES	DOUBLE	Emissions from this source
OFF-ROAD VEHICLES AND MOBILE EQUIPMENT	DOUBLE	Emissions from this source
MANUFACTURING	DOUBLE	Emissions from this source
HOME FIREWOOD BURNING	DOUBLE	Emissions from this source
AGRICULTURE	DOUBLE	Emissions from this source
TRANSPORTATION	DOUBLE	Emissions from this source
TOTAL EMISSIONS	DOUBLE	Total emissions for the specified year

ETL Process

ETL Process

Prepare the data for import by removing any rows or columns that don't contain relevant information.



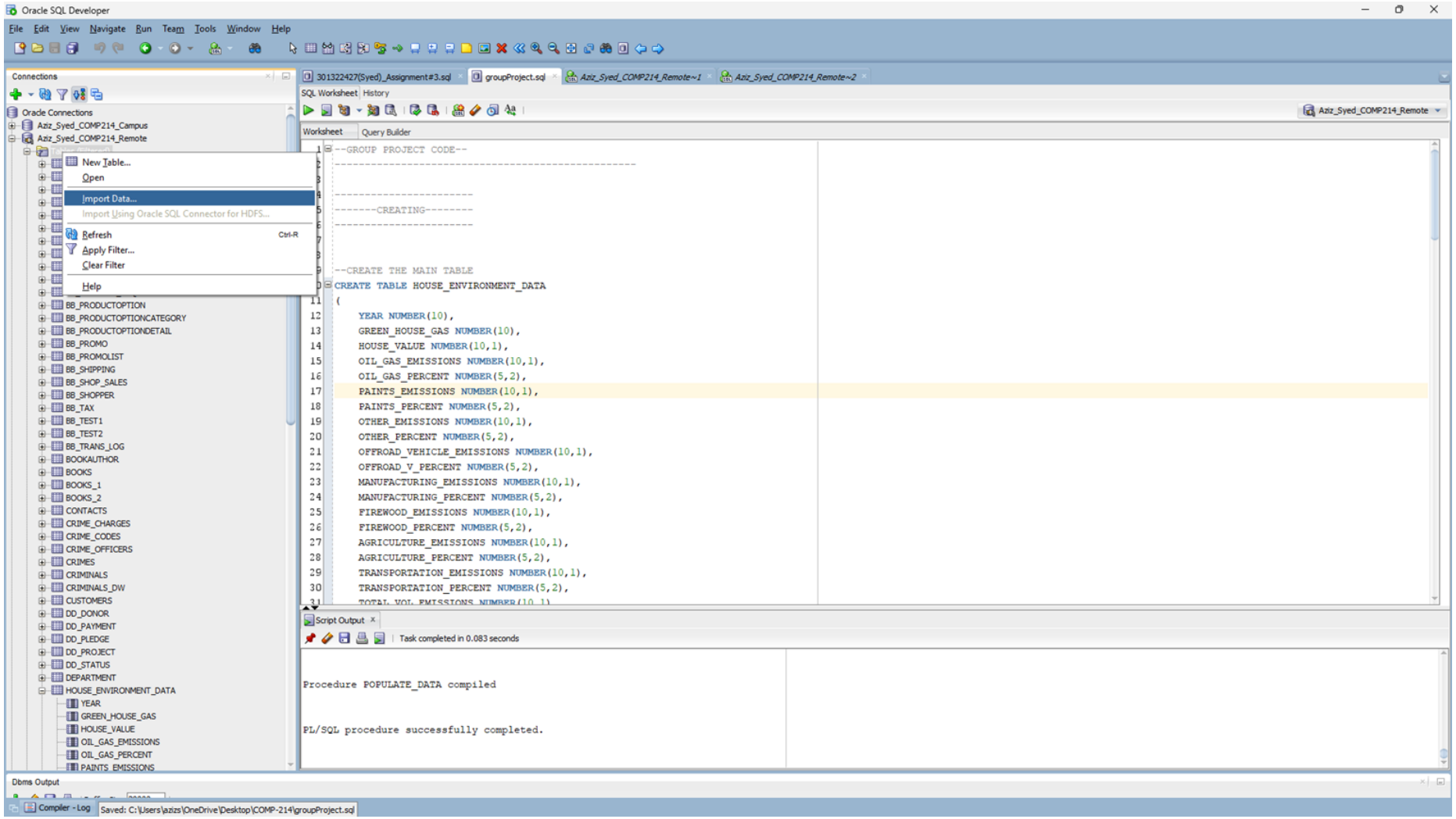
Greenhouse gas emissions, Canada, 1990 to 2022

Year	Total greenhouse gas emissions (megatonnes of carbon dioxide equivalent)
1990	608
1991	603
1992	622
1993	627
1994	649
1995	668
1996	689
1997	706
1998	712
1999	723
2000	748
2001	739
2002	745
2003	763
2004	767
2005	761
2006	757
2007	777
2008	760
2009	716
2010	729
2011	738
2012	744
2013	751
2014	750
2015	746
2016	731
2017	742
2018	753
2019	752
2020	686
2021	698
2022	708

Note: Data are presented as rounded figures. The national indicator tracks 7 greenhouse gases released by human activity: carbon dioxide, methane, nitrous oxide, sulphur hexafluoride, perfluorocarbons, hydrofluorocarbons and nitrogen trifluoride. Emission levels for some years have been revised in light

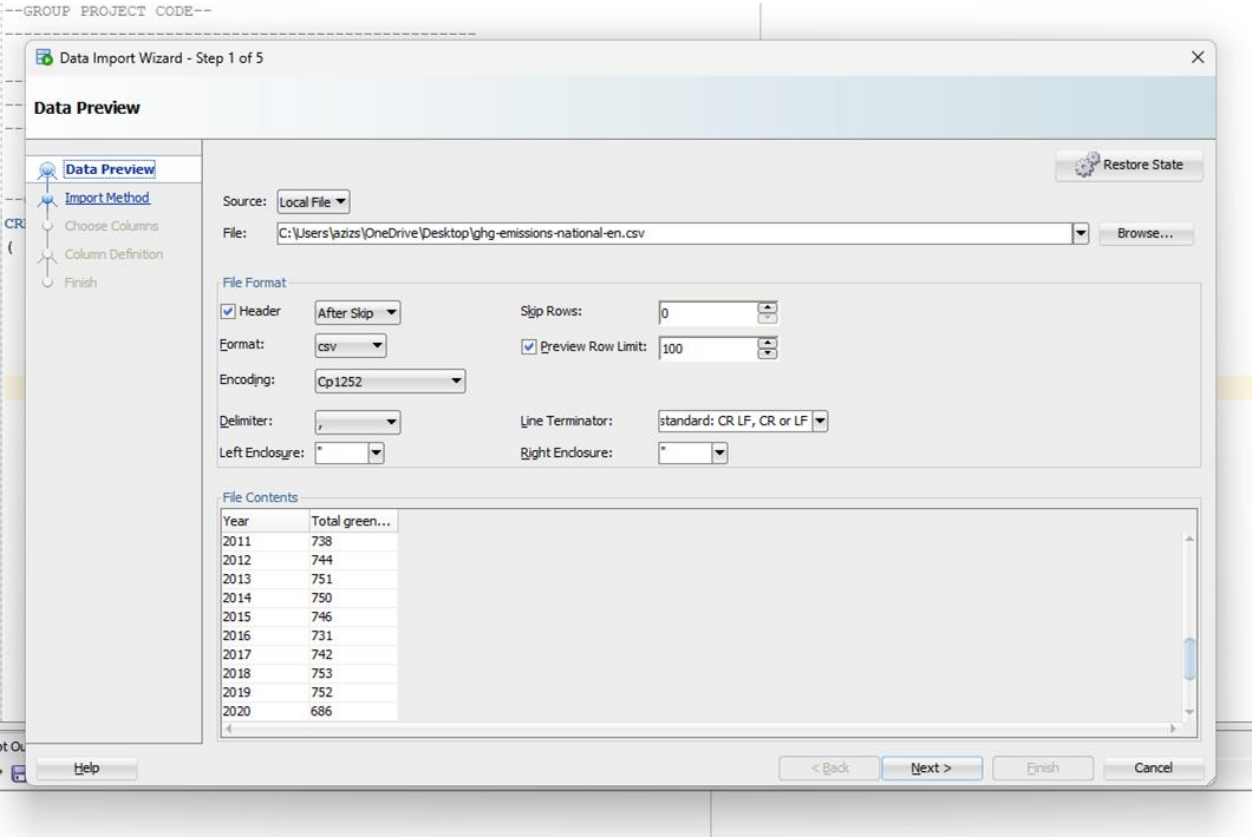
ETL Process

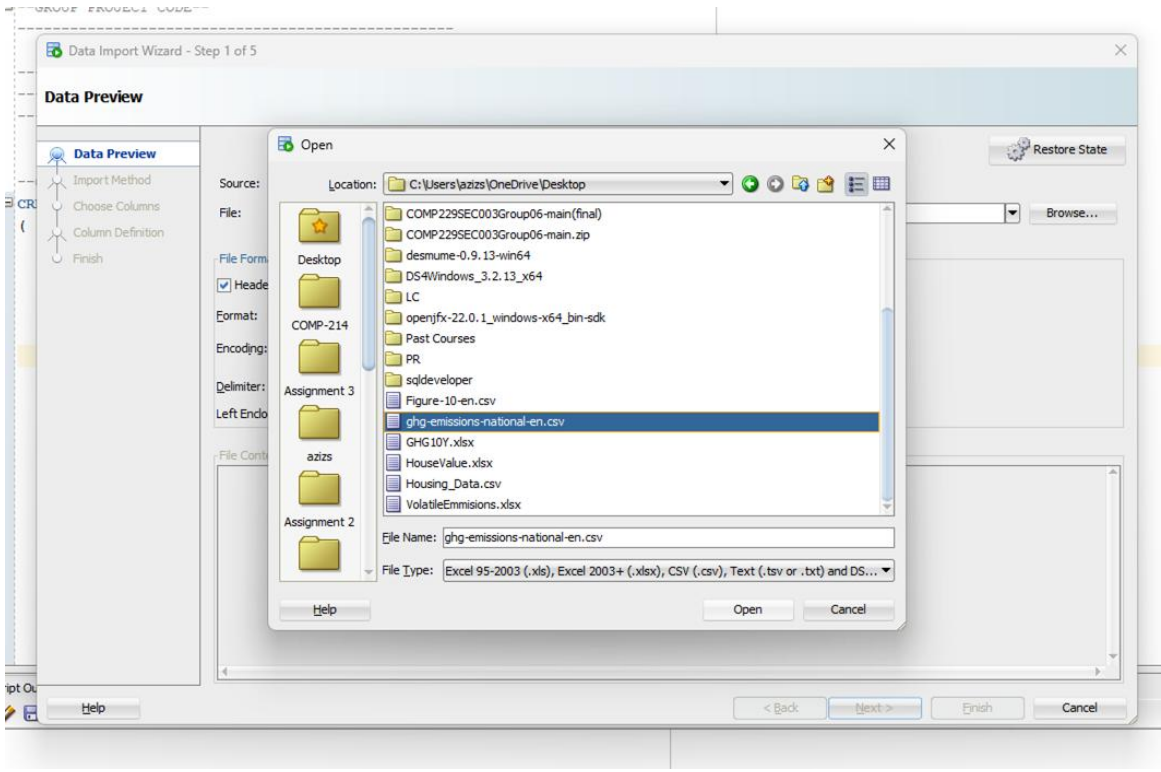
Start importing the data into the database using the import data…



ETL Process

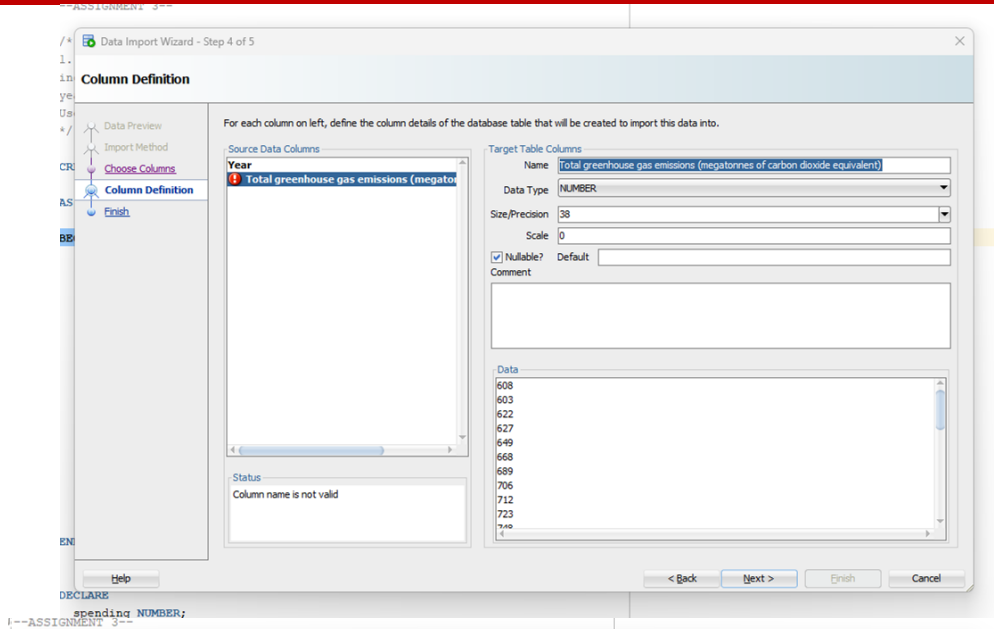
Select the downloaded data set.

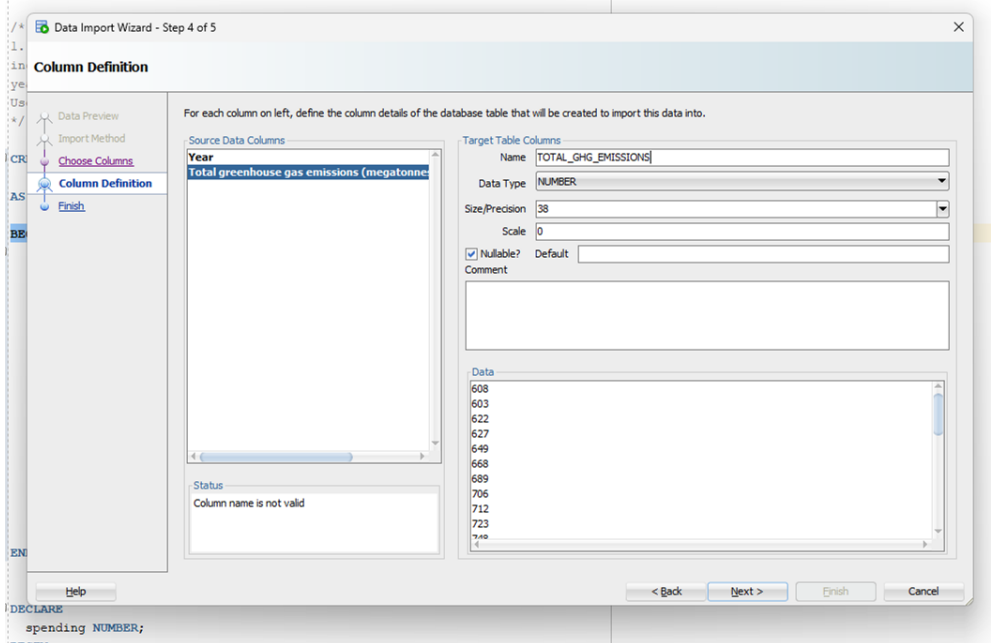




ETL Process

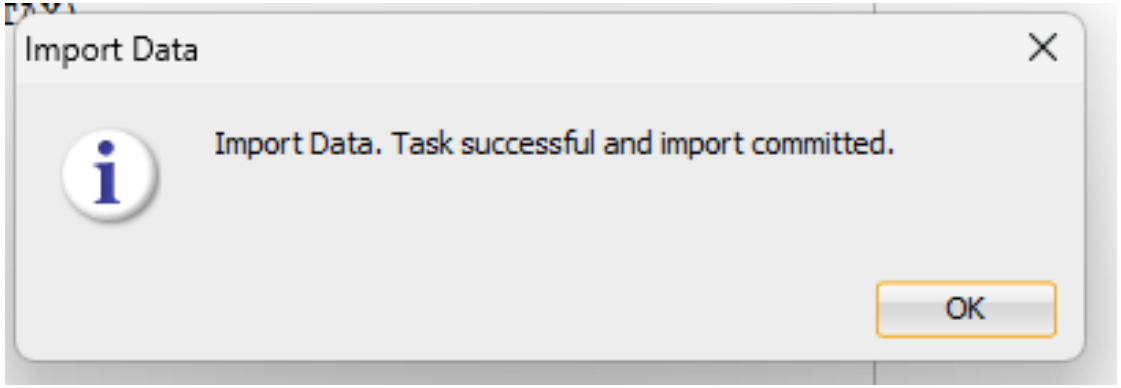
Remove Spacing from column names to fit SQL format.

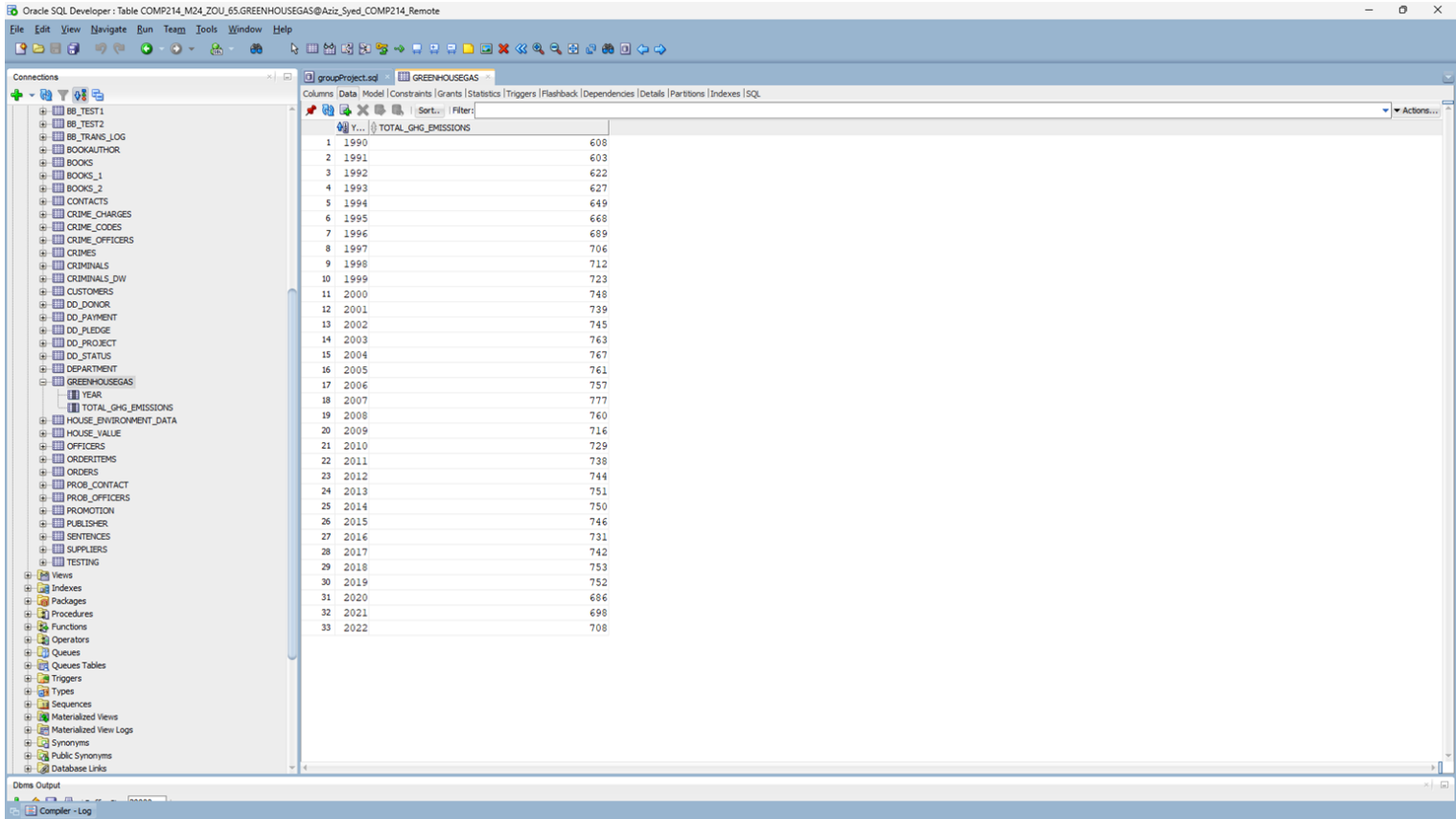




ETL Process

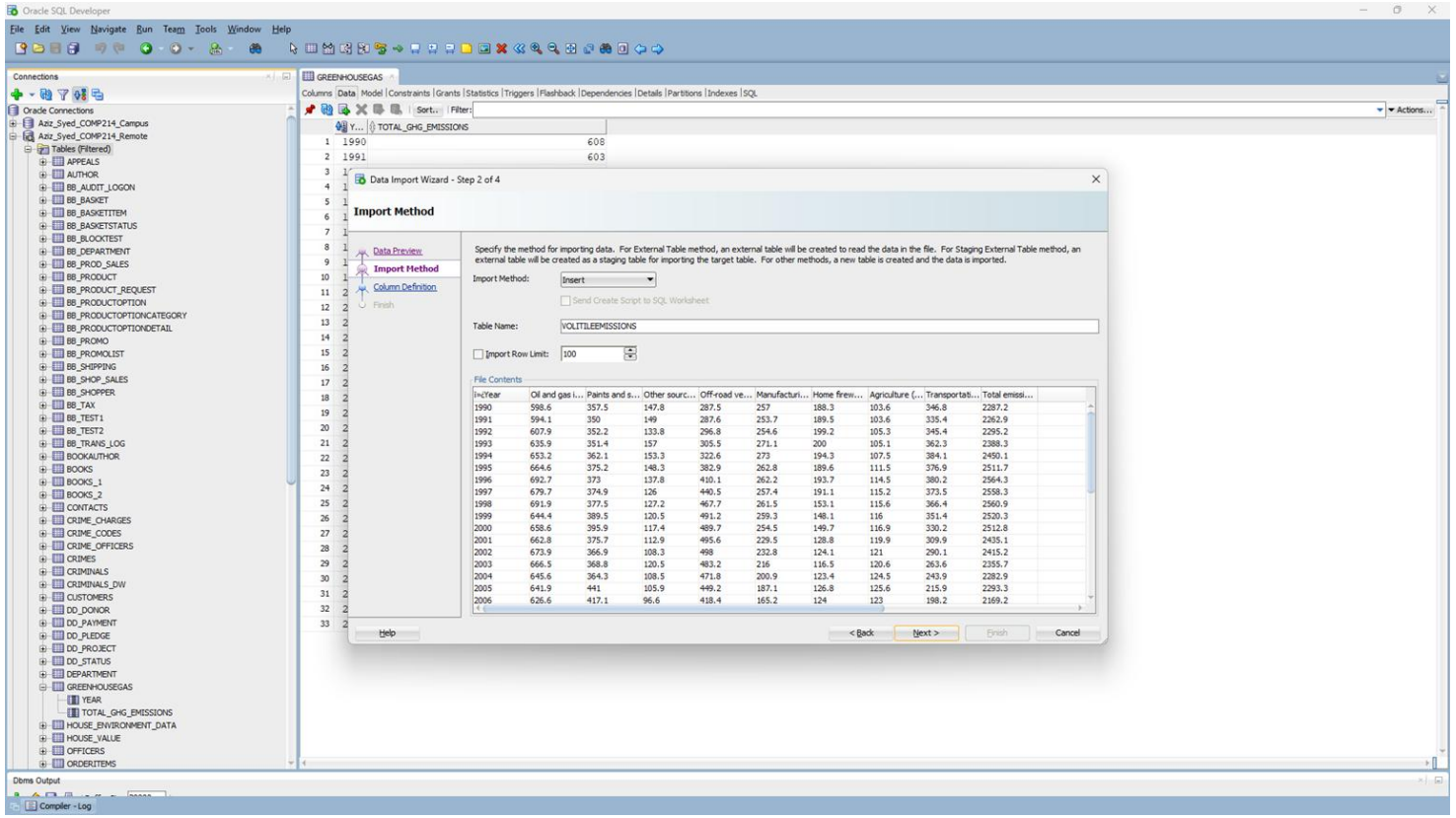
Import





ETL Process

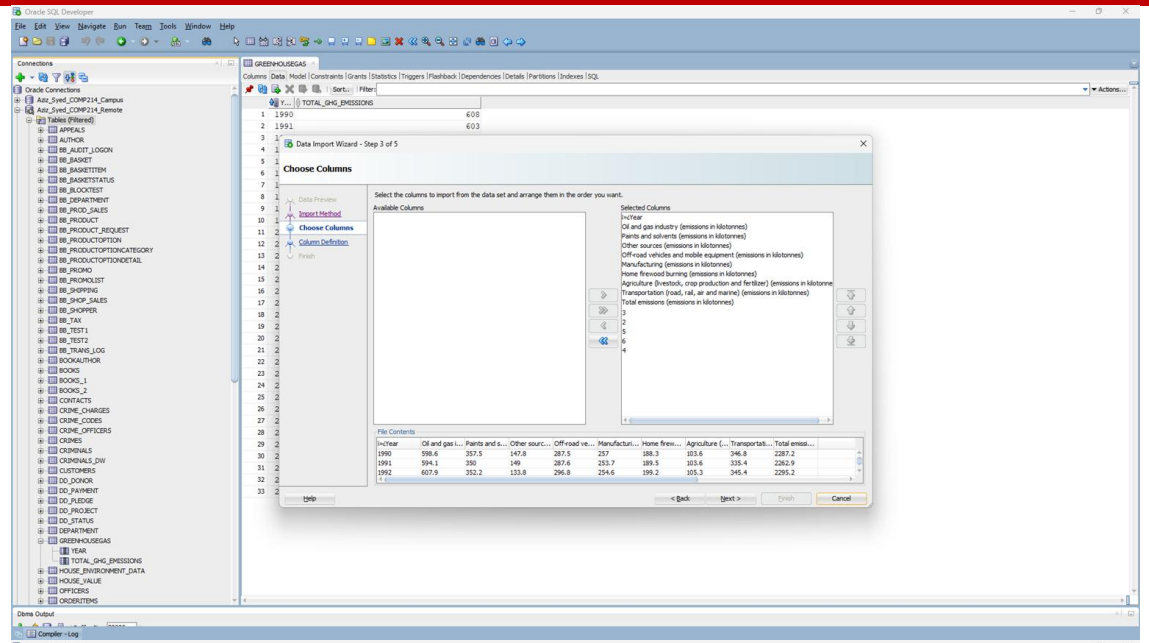
Repeat Process for the volatile emissions data set.

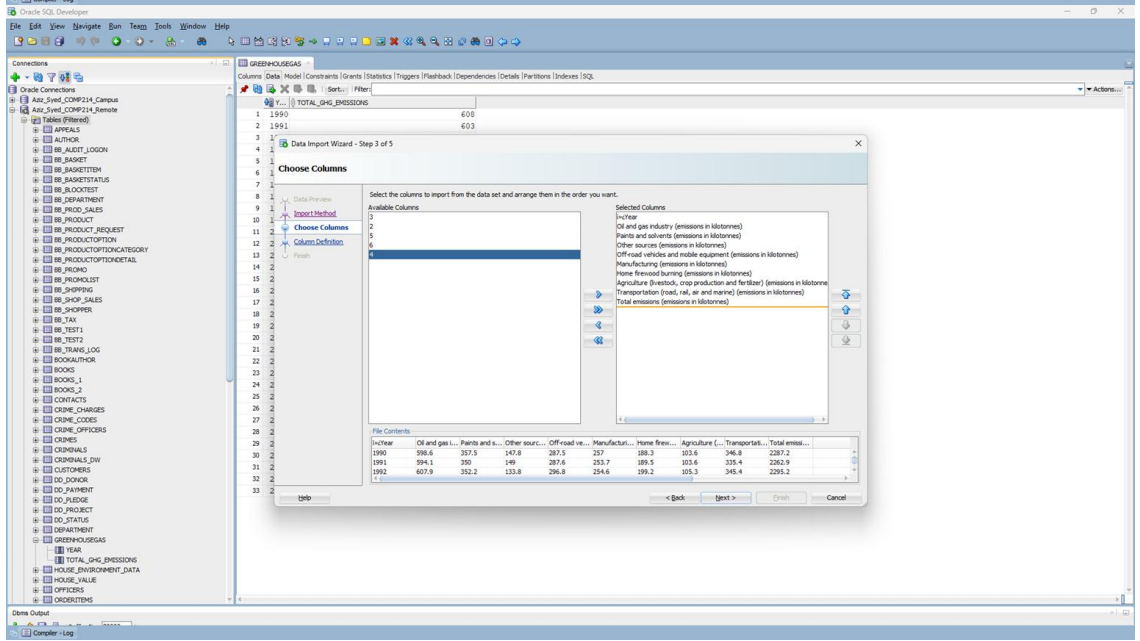


ETL Process

The Volatile Emissions data set had empty columns of data

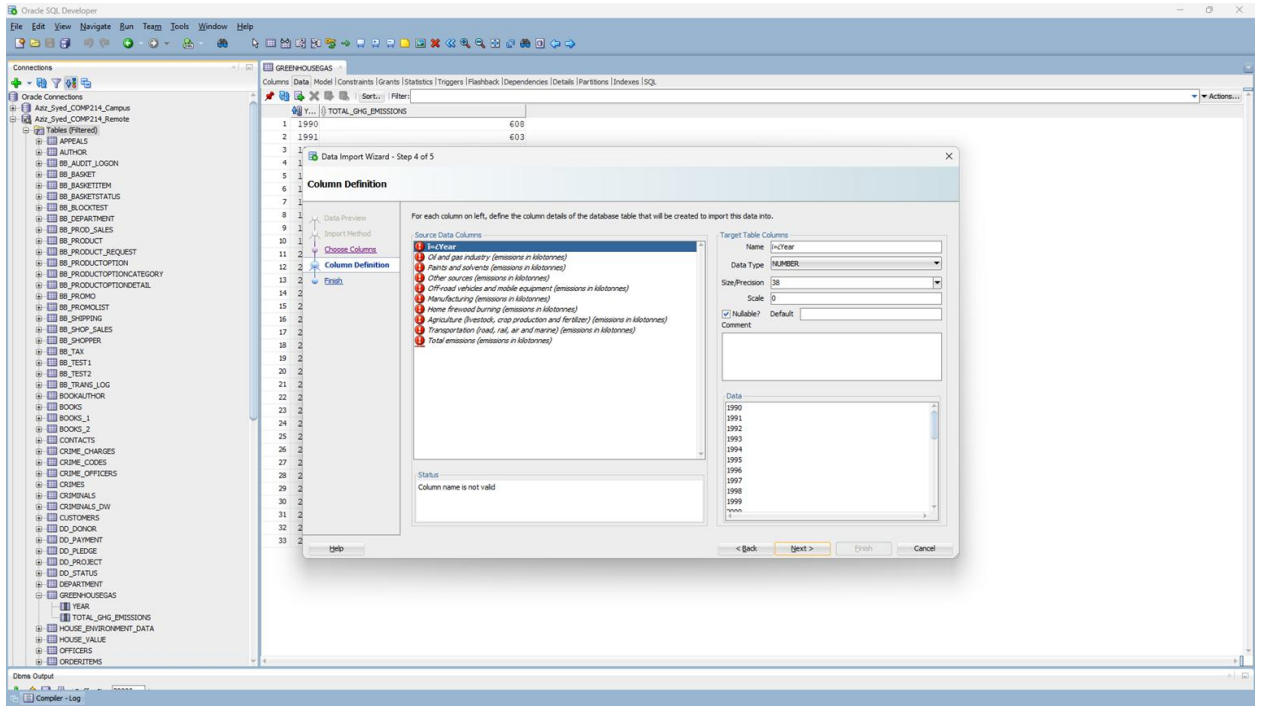
We can easily move columns that are not needed to the “Available Columns” section so they don’t show up when we import.





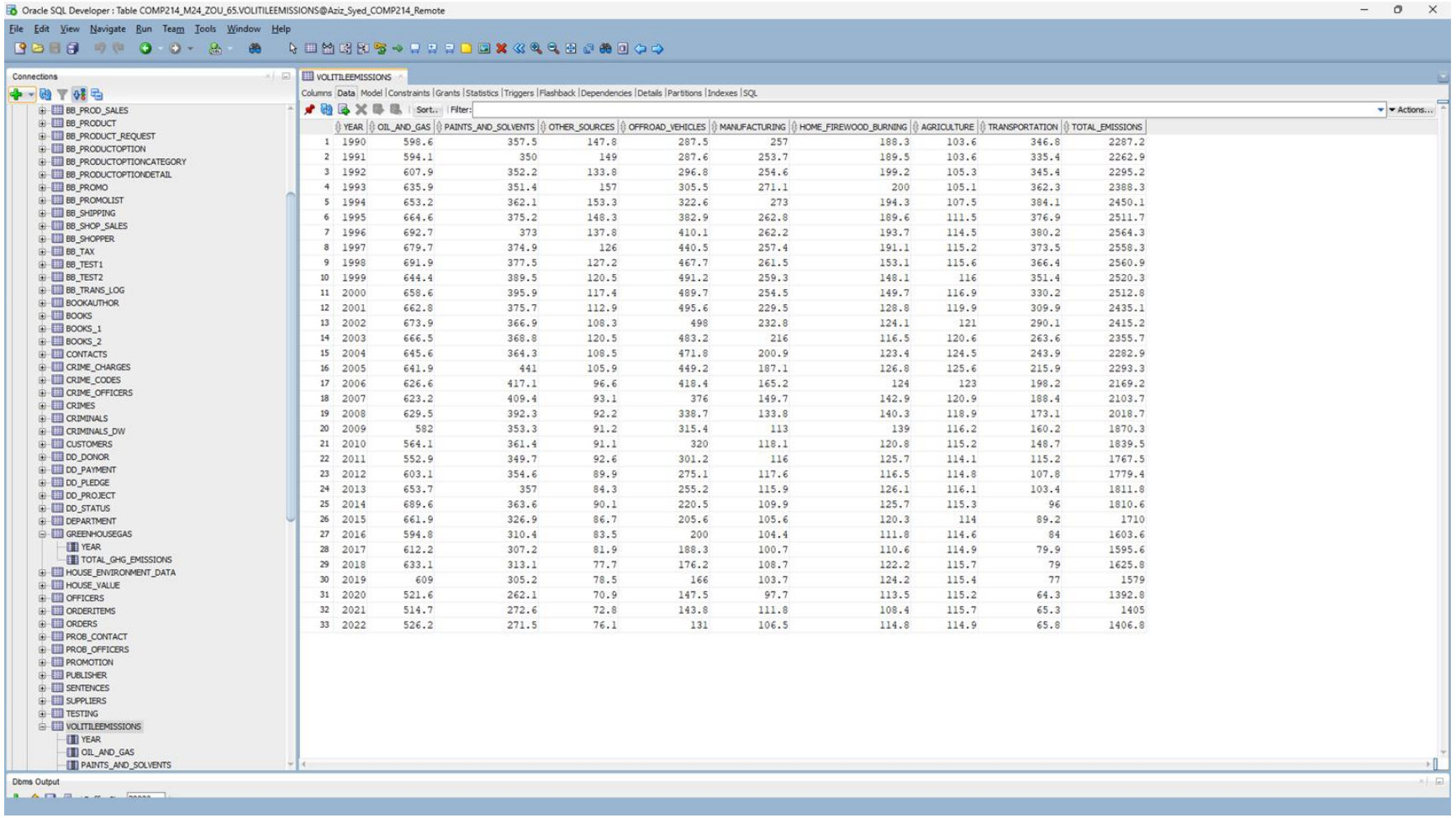
ETL Process

Change names by removing spaces to fit SQL format.



ETL Process

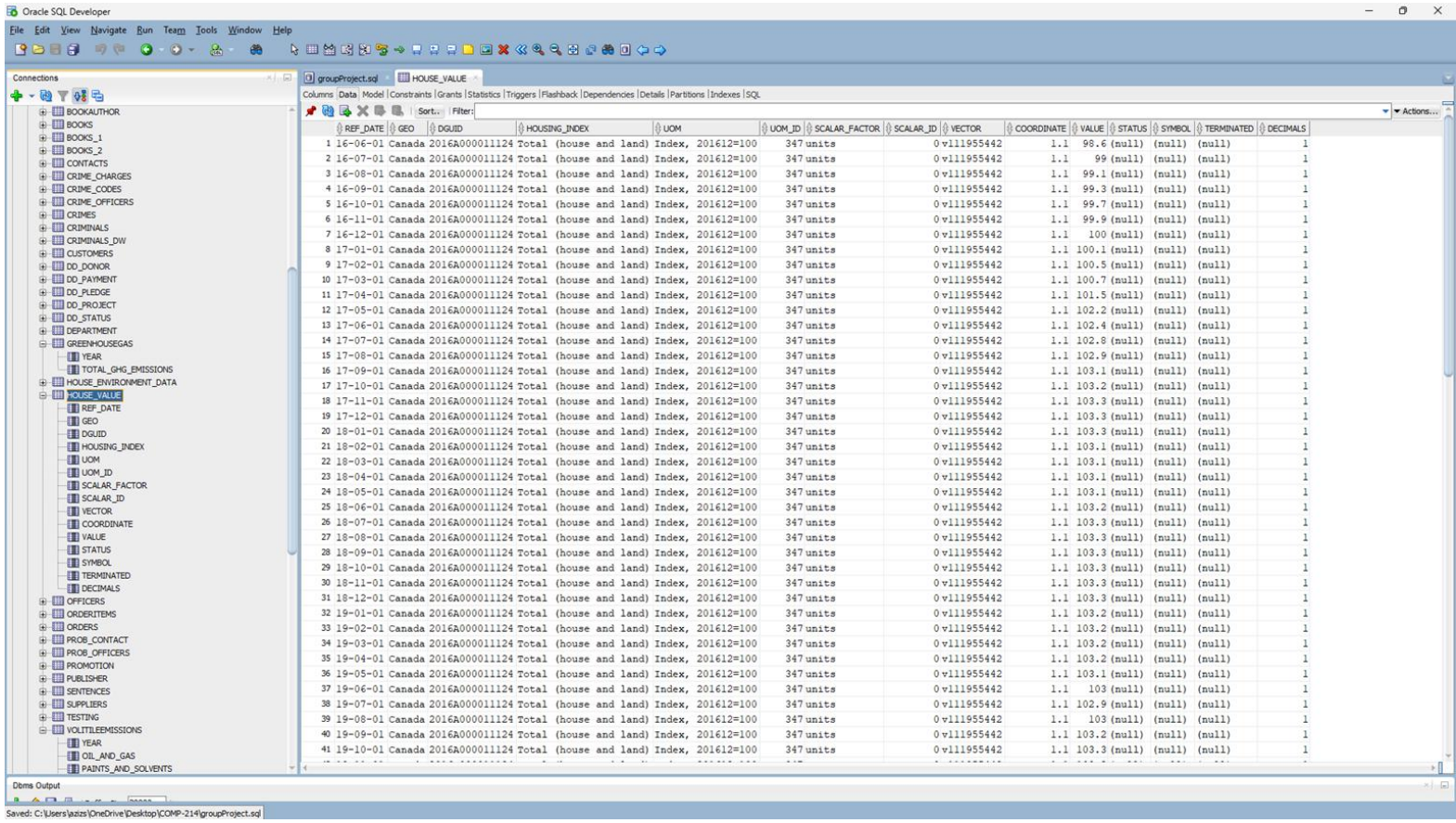
Import



	YEAR	OIL_AND_GAS	PAINTS_AND_SOLVENTS	OTHER_SOURCES	OFFROAD_VEHICLES	MANUFACTURING	HOME_FIREWOOD_BURNING	AGRICULTURE	TRANSPORTATION	TOTAL_EMISSIONS
1	1990	598.6	357.5	147.8	287.5	257	188.3	103.6	346.8	2287.2
2	1991	594.1	350	149	287.6	253.7	189.5	103.6	335.4	2262.9
3	1992	607.9	352.2	133.8	296.8	254.6	199.2	105.3	345.4	2295.2
4	1993	635.9	351.4	157	305.5	271.1	200	105.1	362.3	2388.3
5	1994	653.2	362.1	153.3	322.6	273	194.3	107.5	384.1	2450.1
6	1995	664.6	375.2	148.3	382.9	262.8	189.6	111.5	376.9	2511.7
7	1996	692.7	373	137.8	410.1	262.2	193.7	114.5	380.2	2564.3
8	1997	679.7	374.9	126	440.5	257.4	191.1	115.2	373.5	2558.3
9	1998	691.9	377.5	127.2	467.7	261.5	153.1	115.6	366.4	2560.9
10	1999	644.4	389.5	120.5	491.2	259.3	148.1	116	351.4	2520.3
11	2000	658.6	395.9	117.4	489.7	254.5	149.7	116.9	330.2	2512.8
12	2001	662.8	375.7	112.9	495.6	229.5	128.8	119.9	309.9	2435.1
13	2002	673.9	366.9	108.3	498	232.8	124.1	121	290.1	2415.2
14	2003	666.5	368.8	120.5	483.2	216	116.5	120.6	263.6	2355.7
15	2004	645.6	364.3	108.5	471.8	200.9	123.4	124.5	243.9	2282.9
16	2005	641.9	441	105.9	449.2	187.1	126.8	125.6	215.9	2293.3
17	2006	626.6	417.1	96.6	418.4	165.2	124	123	198.2	2169.2
18	2007	623.2	409.4	93.1	376	149.7	142.9	120.9	188.4	2103.7
19	2008	629.5	392.3	92.2	338.7	133.8	140.3	118.9	173.1	2018.7
20	2009	582	353.3	91.2	315.4	113	139	116.2	160.2	1870.3
21	2010	564.1	361.4	91.1	320	118.1	120.8	115.2	148.7	1839.5
22	2011	552.9	349.7	92.6	301.2	116	125.7	114.1	115.2	1767.5
23	2012	603.1	354.6	89.9	275.1	117.6	116.5	114.8	107.8	1779.4
24	2013	653.7	357	84.3	255.2	115.9	126.1	116.1	103.4	1811.8
25	2014	689.6	363.6	90.1	220.5	109.9	125.7	115.3	96	1810.6
26	2015	661.9	326.9	86.7	205.6	105.6	120.3	114	89.2	1710
27	2016	594.8	310.4	83.5	200	104.4	111.8	114.6	84	1603.6
28	2017	612.2	307.2	81.9	188.3	100.7	110.6	114.9	79.9	1595.6
29	2018	633.1	313.1	77.7	176.2	108.7	122.2	115.7	79	1625.8
30	2019	609	305.2	78.5	166	103.7	124.2	115.4	77	1579
31	2020	521.6	262.1	70.9	147.5	97.7	113.5	115.2	64.3	1392.8
32	2021	514.7	272.6	72.8	143.8	111.8	108.4	115.7	65.3	1405
33	2022	526.2	271.5	76.1	131	106.5	114.8	114.9	65.8	1406.8

ETL Process

Repeat same processing for the Housing Data set.

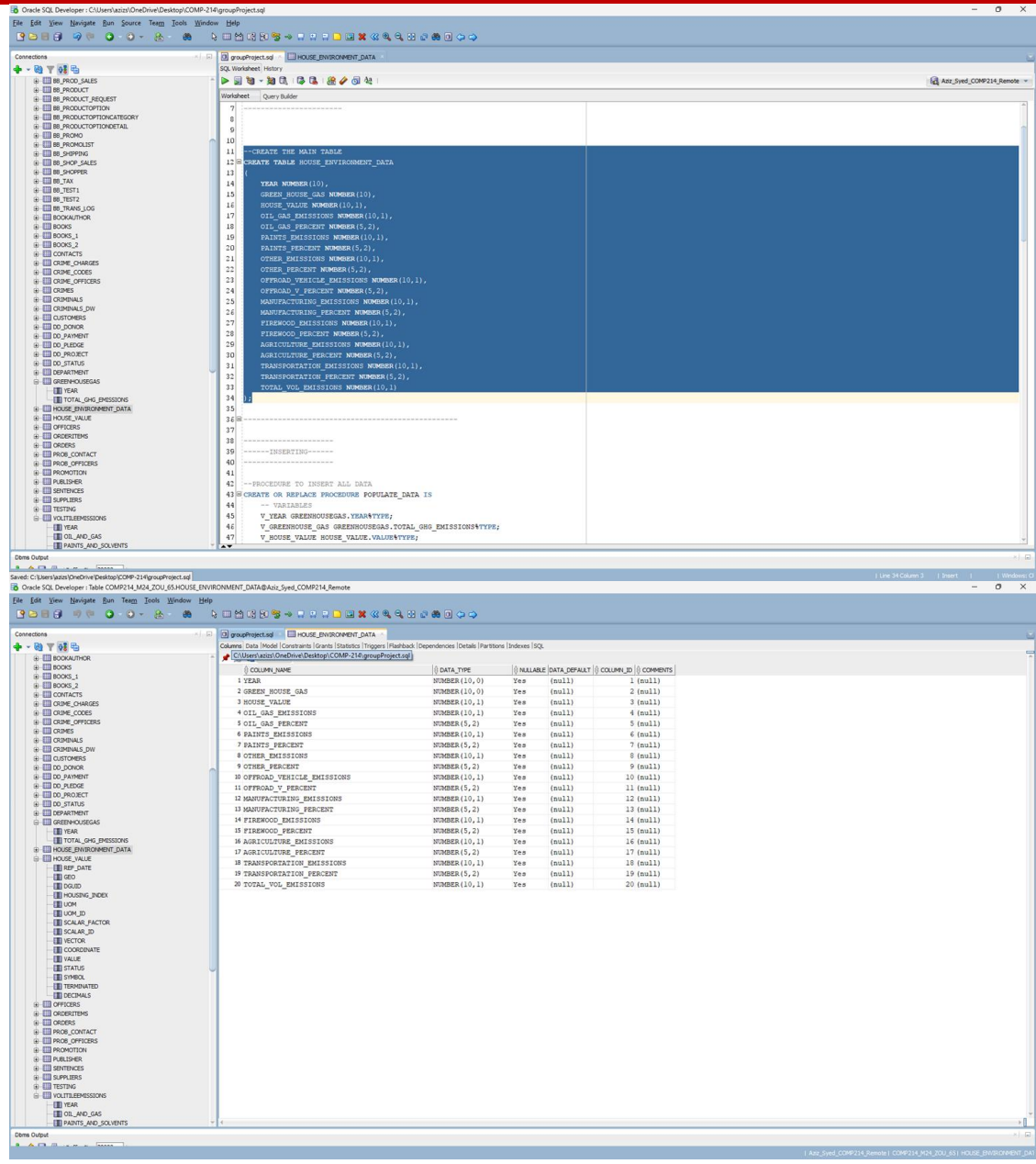


ETL Process

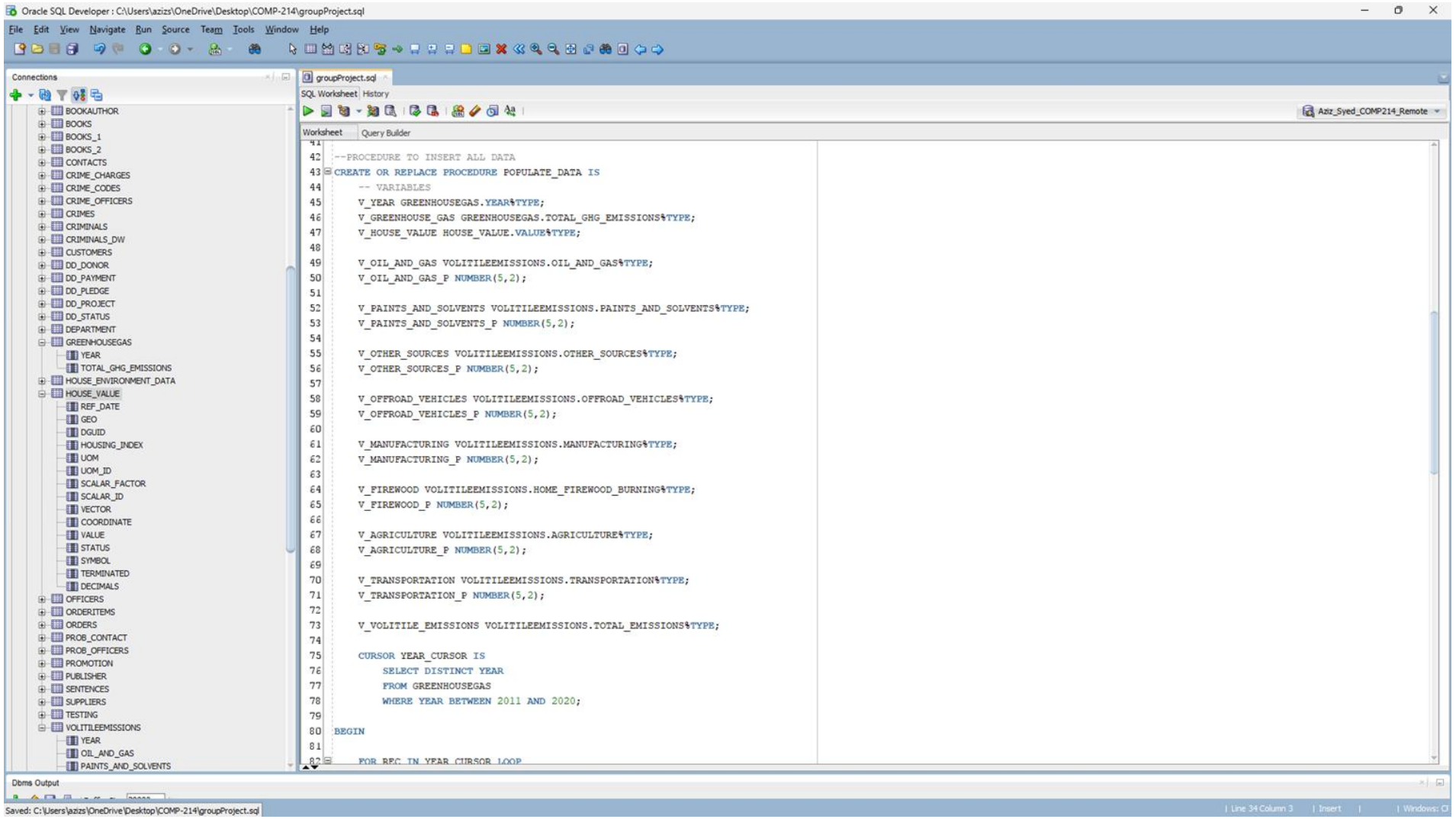
Finally we have to manually create the table in which we want to integrate our data sets into. We chose to use 10 years worth of data, from 2011-2020.

This table also includes the percentage of how much each volatile emission contributes to the total emission.

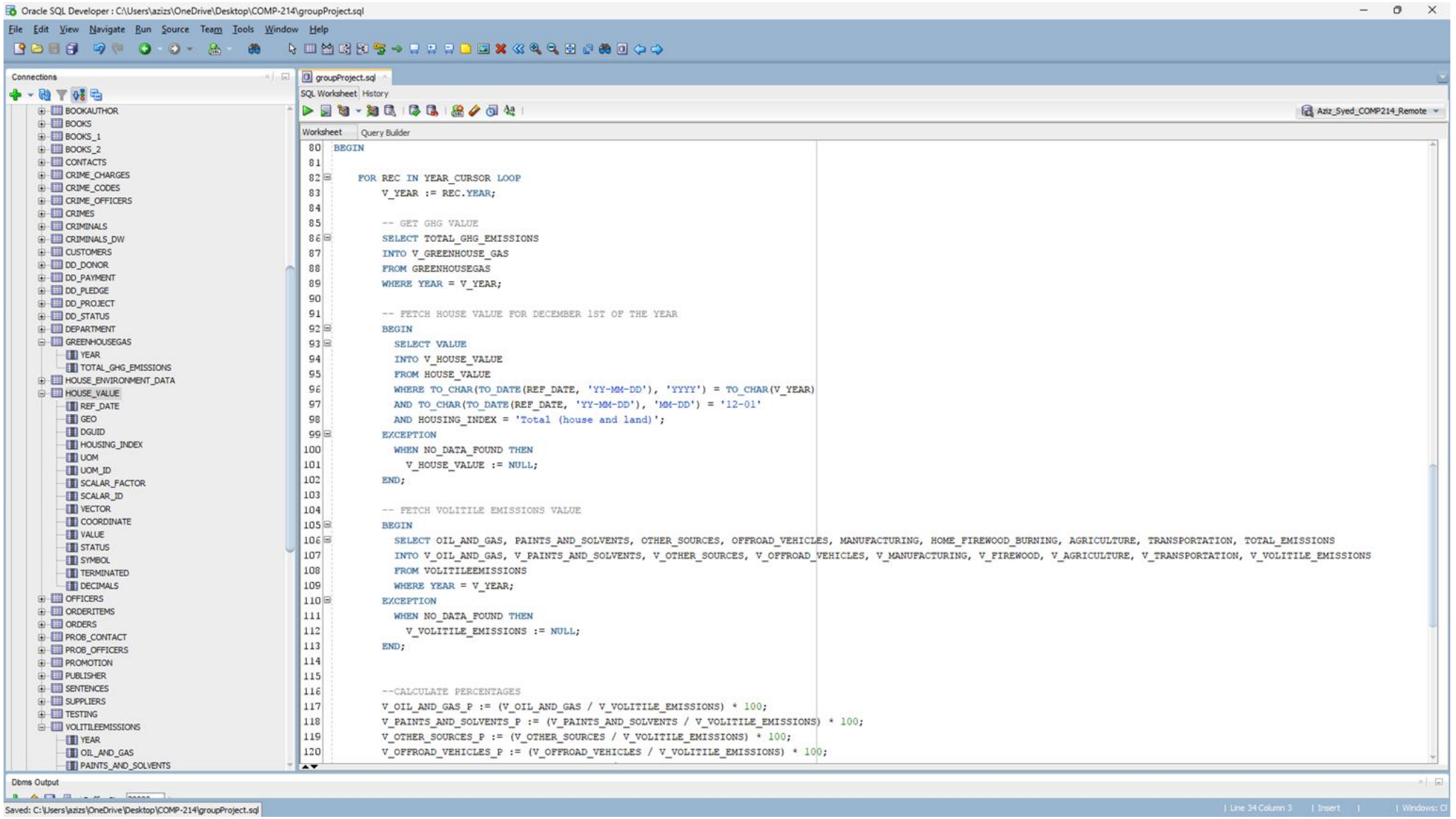
To integrate all of the data sets we are using a procedure that filters data that we need and inserts in into the table.



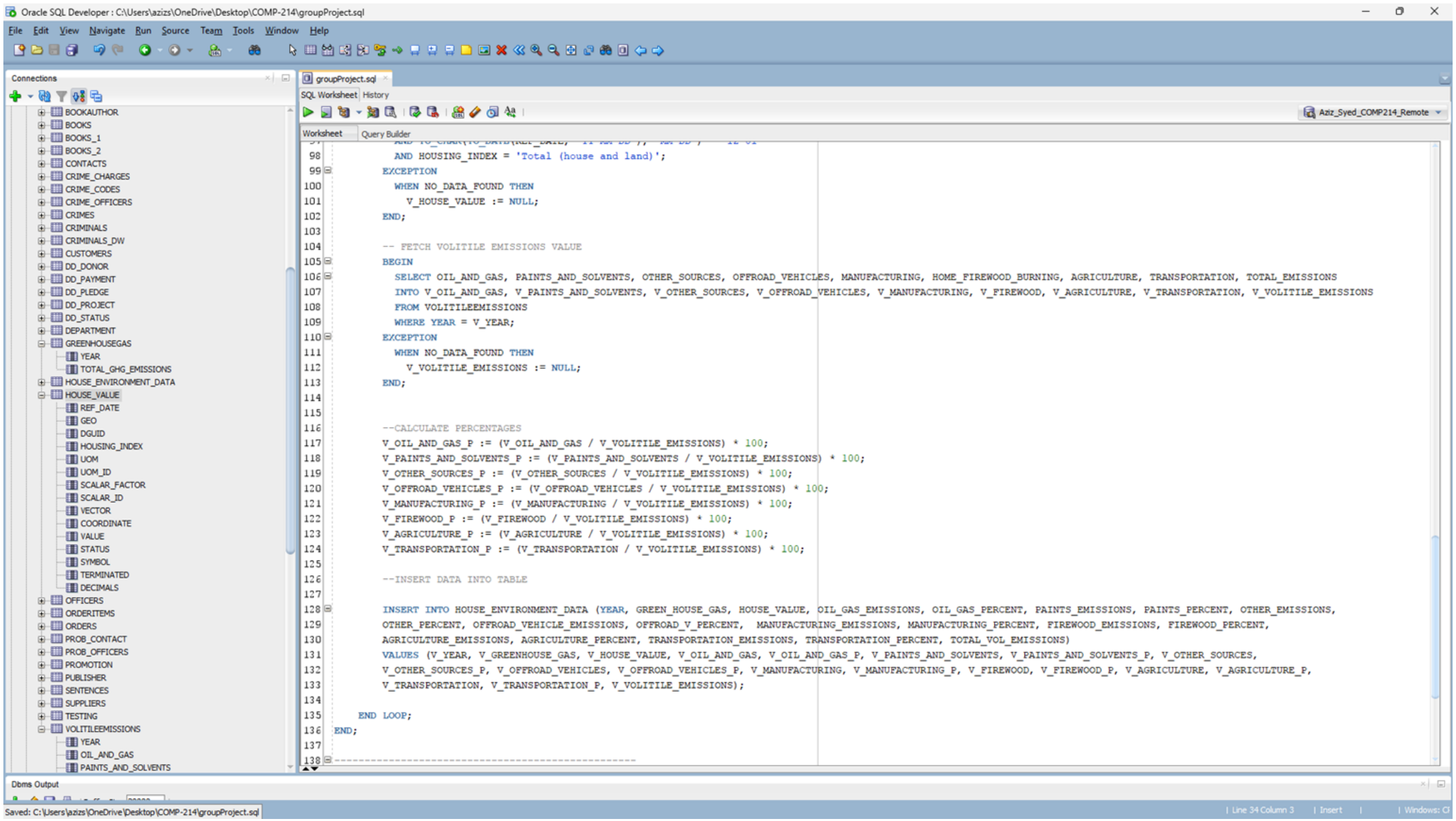
ETL Process


Oracle SQL Developer : C:\Users\azizs\OneDrive\Desktop\COMP-214\groupProject.sql
File Edit View Navigate Run Source Team Tools Window Help
Connections
BOOKAUTHOR
BOOKS
BOOKS_1
BOOKS_2
CONTACTS
CRIME_CHARGES
CRIME_CODES
CRIME_OFFICERS
CRIMES
CRIMINALS
CRIMINALS_DW
CUSTOMERS
DD_DONOR
DD_PAYMENT
DD_PLEDGE
DD_PROJECT
DD_STATUS
DEPARTMENT
GREENHOUSEGAS
YEAR
TOTAL_GHG_EMISSIONS
HOUSE_ENVIRONMENT_DATA
HOUSE_VALUE
REF_DATE
GEO
DGUID
HOUSING_INDEX
UOM
UOM_ID
SCALAR_FACTOR
SCALAR_ID
VECTOR
COORDINATE
VALUE
STATUS
SYMBOL
TERMINATED
DECIMALS
OFFICERS
ORDERITEMS
ORDERS
PROB_CONTACT
PROB_OFFICERS
PROMOTION
PUBLISHER
SENTENCES
SUPPLIERS
TESTING
VOLITILEEMISSIONS
YEAR
OIL_AND_GAS
PAINTS_AND_SOLVENTS
groupProject.sql
SQL Worksheet History
Aziz_Syed_COMP214_Remote
Worksheet Query Builder
--PROCEDURE TO INSERT ALL DATA
CREATE OR REPLACE PROCEDURE POPULATE_DATA IS
-- VARIABLES
V_YEAR GREENHOUSEGAS.YEAR%TYPE;
V_GREENHOUSE_GAS GREENHOUSEGAS.TOTAL_GHG_EMISSIONS%TYPE;
V_HOUSE_VALUE HOUSE_VALUE.VALUE%TYPE;
V_OIL_AND_GAS VOLITILEEMISSIONS.OIL_AND_GAS%TYPE;
V_OIL_AND_GAS_P NUMBER(5,2);
V_PAINTS_AND_SOLVENTS VOLITILEEMISSIONS.PAINTS_AND_SOLVENTS%TYPE;
V_PAINTS_AND_SOLVENTS_P NUMBER(5,2);
V_OTHER_SOURCES VOLITILEEMISSIONS.OTHER_SOURCES%TYPE;
V_OTHER_SOURCES_P NUMBER(5,2);
V_OFFROAD_VEHICLES VOLITILEEMISSIONS.OFFROAD_VEHICLES%TYPE;
V_OFFROAD_VEHICLES_P NUMBER(5,2);
V_MANUFACTURING VOLITILEEMISSIONS.MANUFACTURING%TYPE;
V_MANUFACTURING_P NUMBER(5,2);
V_FIREWOOD VOLITILEEMISSIONS.HOME_FIREWOOD_BURNING%TYPE;
V_FIREWOOD_P NUMBER(5,2);
V_AGRICULTURE VOLITILEEMISSIONS.AGRICULTURE%TYPE;
V_AGRICULTURE_P NUMBER(5,2);
V_TRANSPORTATION VOLITILEEMISSIONS.TRANSPORTATION%TYPE;
V_TRANSPORTATION_P NUMBER(5,2);
V_VOLITILE_EMISSIONS VOLITILEEMISSIONS.TOTAL_EMISSIONS%TYPE;
CURSOR YEAR_CURSOR IS
SELECT DISTINCT YEAR
FROM GREENHOUSEGAS
WHERE YEAR BETWEEN 2011 AND 2020;
BEGIN
FOR REC IN YEAR_CURSOR LOOP
Dbms Output
Saved: C:\Users\azizs\OneDrive\Desktop\COMP-214\groupProject.sql
Line 34 Column 3
Insert

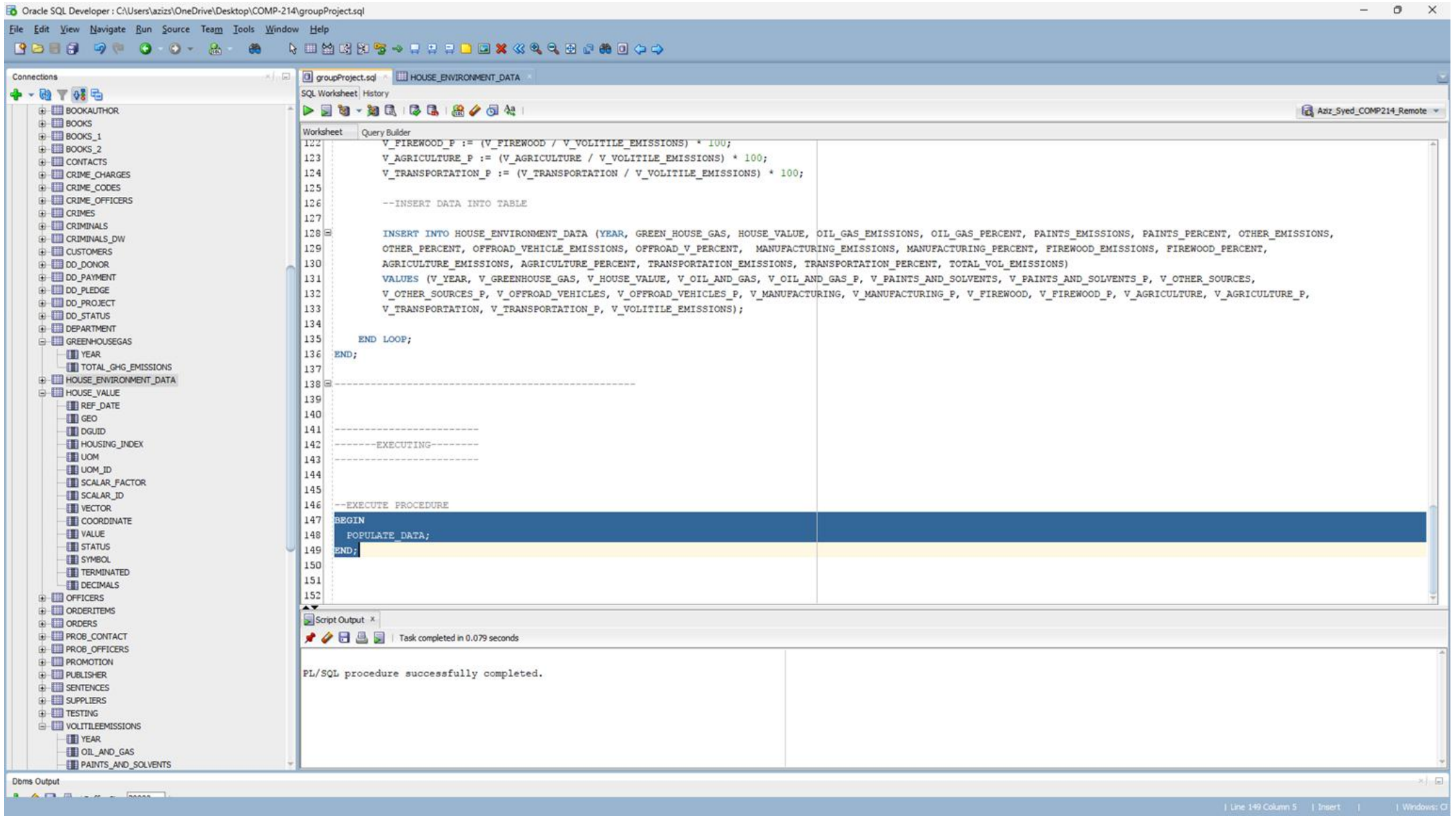
ETL Process


Oracle SQL Developer : C:\Users\azizs\OneDrive\Desktop\COMP-214\groupProject.sql
File Edit View Navigate Run Source Team Tools Window Help
Connections
groupProject.sql
SQL Worksheet History
Worksheet Query Builder
Aziz_Syed_COMP214_Remote
BOOKAUTHOR
BOOKS
BOOKS_1
BOOKS_2
CONTACTS
CRIME_CHARGES
CRIME_CODES
CRIME_OFFICERS
CRIMES
CRIMINALS
CRIMINALS_DW
CUSTOMERS
DD_DONOR
DD_PAYMENT
DD_PLEDGE
DD_PROJECT
DD_STATUS
DEPARTMENT
GREENHOUSEGAS
YEAR
TOTAL_GHG_EMISSIONS
HOUSE_ENVIRONMENT_DATA
HOUSE_VALUE
REF_DATE
GEO
DGUID
HOUSING_INDEX
UOM
UOM_ID
SCALAR_FACTOR
SCALAR_ID
VECTOR
COORDINATE
VALUE
STATUS
SYMBOL
TERMINATED
DECIMALS
OFFICERS
ORDERITEMS
ORDERS
PROB_CONTACT
PROB_OFFICERS
PROMOTION
PUBLISHER
SENTENCES
SUPPLIERS
TESTING
VOLITILEEMISSIONS
YEAR
OIL_AND_GAS
PAINTS_AND_SOLVENTS
BEGIN
FOR REC IN YEAR_CURSOR LOOP
V_YEAR := REC.YEAR;
-- GET GHG VALUE
SELECT TOTAL_GHG_EMISSIONS
INTO V_GREENHOUSE_GAS
FROM GREENHOUSEGAS
WHERE YEAR = V_YEAR;
-- FETCH HOUSE VALUE FOR DECEMBER 1ST OF THE YEAR
BEGIN
SELECT VALUE
INTO V_HOUSE_VALUE
FROM HOUSE_VALUE
WHERE TO_CHAR(TO_DATE(REF_DATE, 'YY-MM-DD'), 'YYYY') = TO_CHAR(V_YEAR)
AND TO_CHAR(TO_DATE(REF_DATE, 'YY-MM-DD'), 'MM-DD') = '12-01'
AND HOUSING_INDEX = 'Total (house and land)';
EXCEPTION
WHEN NO_DATA_FOUND THEN
V_HOUSE_VALUE := NULL;
END;
-- FETCH VOLITILE EMISSIONS VALUE
BEGIN
SELECT OIL_AND_GAS, PAINTS_AND_SOLVENTS, OTHER_SOURCES, OFFROAD_VEHICLES, MANUFACTURING, HOME_FIREWOOD_BURNING, AGRICULTURE, TRANSPORTATION, TOTAL_EMISSIONS
INTO V_OIL_AND_GAS, V_PAINTS_AND_SOLVENTS, V_OTHER_SOURCES, V_OFFROAD_VEHICLES, V_MANUFACTURING, V_FIREWOOD, V_AGRICULTURE, V_TRANSPORTATION, V_VOLITILE_EMISSIONS
FROM VOLITILEEMISSIONS
WHERE YEAR = V_YEAR;
EXCEPTION
WHEN NO_DATA_FOUND THEN
V_VOLITILE_EMISSIONS := NULL;
END;
--CALCULATE PERCENTAGES
V_OIL_AND_GAS_P := (V_OIL_AND_GAS / V_VOLITILE_EMISSIONS) * 100;
V_PAINTS_AND_SOLVENTS_P := (V_PAINTS_AND_SOLVENTS / V_VOLITILE_EMISSIONS) * 100;
V_OTHER_SOURCES_P := (V_OTHER_SOURCES / V_VOLITILE_EMISSIONS) * 100;
V_OFFROAD_VEHICLES_P := (V_OFFROAD_VEHICLES / V_VOLITILE_EMISSIONS) * 100;
Dbms Output
Saved: C:\Users\azizs\OneDrive\Desktop\COMP-214\groupProject.sql
Line 34 Column 3
Insert
Windows:

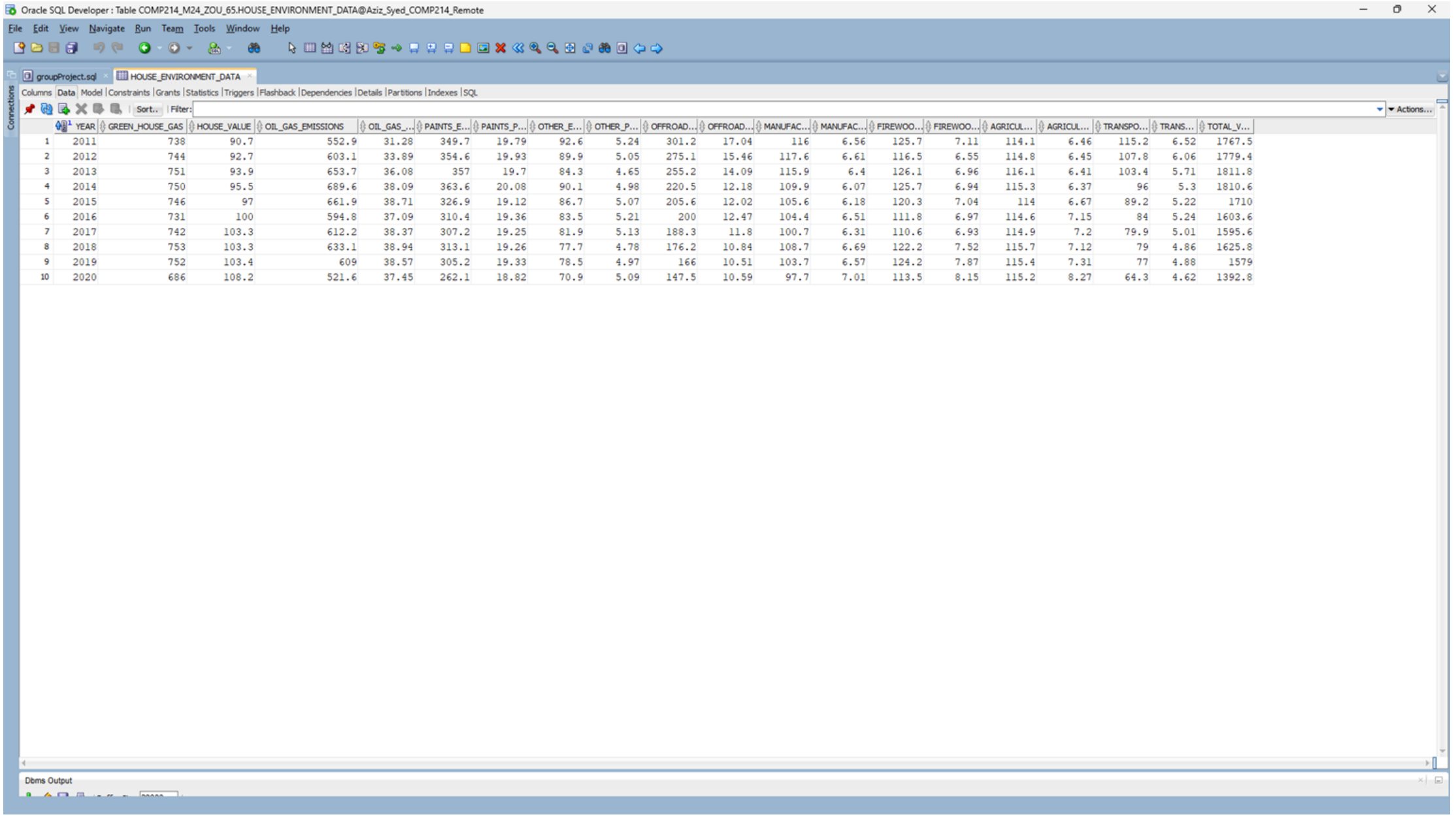
ETL Process


Oracle SQL Developer : C:\Users\azizs\OneDrive\Desktop\COMP-214\groupProject.sql
File Edit View Navigate Run Source Team Tools Window Help
Connections
BOOKAUTHOR
BOOKS
BOOKS_1
BOOKS_2
CONTACTS
CRIME_CHARGES
CRIME_CODES
CRIME_OFFICERS
CRIMES
CRIMINALS
CRIMINALS_DW
CUSTOMERS
DD_DONOR
DD_PAYMENT
DD_PLEDGE
DD_PROJECT
DD_STATUS
DEPARTMENT
GREENHOUSEGAS
YEAR
TOTAL_GHG_EMISSIONS
HOUSE_ENVIRONMENT_DATA
HOUSE_VALUE
REF_DATE
GEO
DGUID
HOUSING_INDEX
UOM
UOM_ID
SCALAR_FACTOR
SCALAR_ID
VECTOR
COORDINATE
VALUE
STATUS
SYMBOL
TERMINATED
DECIMALS
OFFICERS
ORDERITEMS
ORDERS
PROB_CONTACT
PROB_OFFICERS
PROMOTION
PUBLISHER
SENTENCES
SUPPLIERS
TESTING
VOLITILEEMISSIONS
YEAR
OIL_AND_GAS
PAINTS_AND_SOLVENTS
groupProject.sql
SQL Worksheet History
Aziz_Syed_COMP214_Remote
Worksheet Query Builder
AND HOUSING_INDEX = 'Total (house and land)';
EXCEPTION
WHEN NO_DATA_FOUND THEN
V_HOUSE_VALUE := NULL;
END;
-- FETCH VOLITILE EMISSIONS VALUE
BEGIN
SELECT OIL_AND_GAS, PAINTS_AND_SOLVENTS, OTHER_SOURCES, OFFROAD_VEHICLES, MANUFACTURING, HOME_FIREWOOD_BURNING, AGRICULTURE, TRANSPORTATION, TOTAL_EMISSIONS
INTO V_OIL_AND_GAS, V_PAINTS_AND_SOLVENTS, V_OTHER_SOURCES, V_OFFROAD_VEHICLES, V_MANUFACTURING, V_FIREWOOD, V_AGRICULTURE, V_TRANSPORTATION, V_VOLITILE_EMISSIONS
FROM VOLITILEEMISSIONS
WHERE YEAR = V_YEAR;
EXCEPTION
WHEN NO_DATA_FOUND THEN
V_VOLITILE_EMISSIONS := NULL;
END;
--CALCULATE PERCENTAGES
V_OIL_AND_GAS_P := (V_OIL_AND_GAS / V_VOLITILE_EMISSIONS) * 100;
V_PAINTS_AND_SOLVENTS_P := (V_PAINTS_AND_SOLVENTS / V_VOLITILE_EMISSIONS) * 100;
V_OTHER_SOURCES_P := (V_OTHER_SOURCES / V_VOLITILE_EMISSIONS) * 100;
V_OFFROAD_VEHICLES_P := (V_OFFROAD_VEHICLES / V_VOLITILE_EMISSIONS) * 100;
V_MANUFACTURING_P := (V_MANUFACTURING / V_VOLITILE_EMISSIONS) * 100;
V_FIREWOOD_P := (V_FIREWOOD / V_VOLITILE_EMISSIONS) * 100;
V_AGRICULTURE_P := (V_AGRICULTURE / V_VOLITILE_EMISSIONS) * 100;
V_TRANSPORTATION_P := (V_TRANSPORTATION / V_VOLITILE_EMISSIONS) * 100;
--INSERT DATA INTO TABLE
INSERT INTO HOUSE_ENVIRONMENT_DATA (YEAR, GREEN_HOUSE_GAS, HOUSE_VALUE, OIL_GAS_EMISSIONS, OIL_GAS_PERCENT, PAINTS_EMISSIONS, PAINTS_PERCENT, OTHER_EMISSIONS,
OTHER_PERCENT, OFFROAD_VEHICLE_EMISSIONS, OFFROAD_V_PERCENT, MANUFACTURING_EMISSIONS, MANUFACTURING_PERCENT, FIREWOOD_EMISSIONS, FIREWOOD_PERCENT,
AGRICULTURE_EMISSIONS, AGRICULTURE_PERCENT, TRANSPORTATION_EMISSIONS, TRANSPORTATION_PERCENT, TOTAL_VOL_EMISSIONS)
VALUES (V_YEAR, V_GREENHOUSE_GAS, V_HOUSE_VALUE, V_OIL_AND_GAS, V_OIL_AND_GAS_P, V_PAINTS_AND_SOLVENTS, V_PAINTS_AND_SOLVENTS_P, V_OTHER_SOURCES,
V_OTHER_SOURCES_P, V_OFFROAD_VEHICLES, V_OFFROAD_VEHICLES_P, V_MANUFACTURING, V_MANUFACTURING_P, V_FIREWOOD, V_FIREWOOD_P, V_AGRICULTURE, V_AGRICULTURE_P,
V_TRANSPORTATION, V_TRANSPORTATION_P, V_VOLITILE_EMISSIONS);
END LOOP;
END;
Dbms Output
Saved: C:\Users\azizs\OneDrive\Desktop\COMP-214\groupProject.sql
Line 34 Column 3
Insert

ETL Process


Oracle SQL Developer : C:\Users\azizs\OneDrive\Desktop\COMP-214\groupProject.sql
File Edit View Navigate Run Source Team Tools Window Help
Connections
BOOKAUTHOR
BOOKS
BOOKS_1
BOOKS_2
CONTACTS
CRIME_CHARGES
CRIME_CODES
CRIME_OFFICERS
CRIMES
CRIMINALS
CRIMINALS_DW
CUSTOMERS
DD_DONOR
DD_PAYMENT
DD_PLEDGE
DD_PROJECT
DD_STATUS
DEPARTMENT
GREENHOUSEGAS
YEAR
TOTAL_GHG_EMISSIONS
HOUSE_ENVIRONMENT_DATA
HOUSE_VALUE
REF_DATE
GEO
DGUID
HOUSING_INDEX
UOM
UOM_ID
SCALAR_FACTOR
SCALAR_ID
VECTOR
COORDINATE
VALUE
STATUS
SYMBOL
TERMINATED
DECIMALS
OFFICERS
ORDERITEMS
ORDERS
PROB_CONTACT
PROB_OFFICERS
PROMOTION
PUBLISHER
SENTENCES
SUPPLIERS
TESTING
VOLITILEEMISSIONS
YEAR
OIL_AND_GAS
PAINTS_AND_SOLVENTS
groupProject.sql
HOUSE_ENVIRONMENT_DATA
SQL Worksheet History
Aziz_Syed_COMP214_Remote
Worksheet Query Builder
V_FIREWOOD_P := (V_FIREWOOD / V_VOLITILE_EMISSIONS) * 100;
V_AGRICULTURE_P := (V_AGRICULTURE / V_VOLITILE_EMISSIONS) * 100;
V_TRANSPORTATION_P := (V_TRANSPORTATION / V_VOLITILE_EMISSIONS) * 100;
--INSERT DATA INTO TABLE
INSERT INTO HOUSE_ENVIRONMENT_DATA (YEAR, GREEN_HOUSE_GAS, HOUSE_VALUE, OIL_GAS_EMISSIONS, OIL_GAS_PERCENT, PAINTS_EMISSIONS, PAINTS_PERCENT, OTHER_EMISSIONS,
OTHER_PERCENT, OFFROAD_VEHICLE_EMISSIONS, OFFROAD_V_PERCENT, MANUFACTURING_EMISSIONS, MANUFACTURING_PERCENT, FIREWOOD_EMISSIONS, FIREWOOD_PERCENT,
AGRICULTURE_EMISSIONS, AGRICULTURE_PERCENT, TRANSPORTATION_EMISSIONS, TRANSPORTATION_PERCENT, TOTAL_VOL_EMISSIONS)
VALUES (V_YEAR, V_GREENHOUSE_GAS, V_HOUSE_VALUE, V_OIL_AND_GAS, V_OIL_AND_GAS_P, V_PAINTS_AND_SOLVENTS, V_PAINTS_AND_SOLVENTS_P, V_OTHER_SOURCES,
V_OTHER_SOURCES_P, V_OFFROAD_VEHICLES, V_OFFROAD_VEHICLES_P, V_MANUFACTURING, V_MANUFACTURING_P, V_FIREWOOD, V_FIREWOOD_P, V_AGRICULTURE, V_AGRICULTURE_P,
V_TRANSPORTATION, V_TRANSPORTATION_P, V_VOLITILE_EMISSIONS);
END LOOP;
END;
--EXECUTE PROCEDURE
BEGIN
POPULATE_DATA;
END;
Script Output
Task completed in 0.079 seconds
PL/SQL procedure successfully completed.
Dbms Output
Line 149 Column 5 Insert

ETL Process



	YEAR	GREEN_HOUSE_GAS	HOUSE_VALUE	OIL_GAS_EMISSIONS	OIL_GAS_...	PAINTS_E...	PAINTS_P...	OTHER_E...	OTHER_P...	OFFROAD...	OFFROAD...	MANUFAC...	MANUFAC...	FIREWOO...	FIREWOO...	AGRICUL...	AGRICUL...	TRANSPO...	TRANS...	TOTAL_V...
1	2011	738	90.7	552.9	31.28	349.7	19.79	92.6	5.24	301.2	17.04	116	6.56	125.7	7.11	114.1	6.46	115.2	6.52	1767.5
2	2012	744	92.7	603.1	33.89	354.6	19.93	89.9	5.05	275.1	15.46	117.6	6.61	116.5	6.55	114.8	6.45	107.8	6.06	1779.4
3	2013	751	93.9	653.7	36.08	357	19.7	84.3	4.65	255.2	14.09	115.9	6.4	126.1	6.96	116.1	6.41	103.4	5.71	1811.8
4	2014	750	95.5	689.6	38.09	363.6	20.08	90.1	4.98	220.5	12.18	109.9	6.07	125.7	6.94	115.3	6.37	96	5.3	1810.6
5	2015	746	97	661.9	38.71	326.9	19.12	86.7	5.07	205.6	12.02	105.6	6.18	120.3	7.04	114	6.67	89.2	5.22	1710
6	2016	731	100	594.8	37.09	310.4	19.36	83.5	5.21	200	12.47	104.4	6.51	111.8	6.97	114.6	7.15	84	5.24	1603.6
7	2017	742	103.3	612.2	38.37	307.2	19.25	81.9	5.13	188.3	11.8	100.7	6.31	110.6	6.93	114.9	7.2	79.9	5.01	1595.6
8	2018	753	103.3	633.1	38.94	313.1	19.26	77.7	4.78	176.2	10.84	108.7	6.69	122.2	7.52	115.7	7.12	79	4.86	1625.8
9	2019	752	103.4	609	38.57	305.2	19.33	78.5	4.97	166	10.51	103.7	6.57	124.2	7.87	115.4	7.31	77	4.88	1579
10	2020	686	108.2	521.6	37.45	262.1	18.82	70.9	5.09	147.5	10.59	97.7	7.01	113.5	8.15	115.2	8.27	64.3	4.62	1392.8

Application Of This Dataset In Data Science

- ## Predicting Real Estate Market Trends
  - **Can be used to predict real estate market by analyzing any trends from the past 10 years. We can study how fluctuations in greenhouse gas emissions or oil and gas emissions affect property values.**
- ## Market Timing and Decision-Making
  - **You can analyze the trends to see dips in the markets to see the perfect time to buy or sell, or maybe even build a house.**
- ## Environmental Advocacy
  - **Environmental organizations could demonstrate how reducing emissions from oil and gas or other sectors could potentially increase property values in Canada.**

Application Example

By using a simple SELECT statement, we can compare the value of a house, greenhouse gas emissions, and total volatile emissions:

```
SELECT YEAR, HOUSE_VALUE, GREEN_HOUSE_GAS, TOTAL_VOL_EMISSIONS
FROM HOUSE_ENVIRONMENT_DATA
ORDER BY YEAR ASC;
```

Query Result

SQL | All Rows Fetched: 10 in 0.045 seconds

	YEAR	HOUSE_VALUE	GREEN_HOUSE_GAS	TOTAL_VOL_EMISSIONS
1	2011	90.7	738	1767.5
2	2012	92.7	744	1779.4
3	2013	93.9	751	1811.8
4	2014	95.5	750	1810.6
5	2015	97	746	1710
6	2016	100	731	1603.6
7	2017	103.3	742	1595.6
8	2018	103.3	753	1625.8
9	2019	103.4	752	1579
10	2020	108.2	686	1392.8

We can clearly see as emission levels lower, the value of a house rises, meaning that the better the environment is in Canada, the higher the value of a house is.

Or, we can see how a specific emission correlates with the value of a house:

```
SELECT YEAR, HOUSE_VALUE, OFFROAD_VEHICLE_EMISSIONS, OFFROAD_V_PERCENT
FROM HOUSE_ENVIRONMENT_DATA
ORDER BY YEAR ASC;
```

Query Result

SQL | All Rows Fetched: 10 in 0.032 seconds

	YEAR	HOUSE_VALUE	OFFROAD_VEHICLE_EMISSIONS	OFFROAD_V_PERCENT
1	2011	90.7	301.2	17.04
2	2012	92.7	275.1	15.46
3	2013	93.9	255.2	14.09
4	2014	95.5	220.5	12.18
5	2015	97	205.6	12.02
6	2016	100	200	12.47
7	2017	103.3	188.3	11.8
8	2018	103.3	176.2	10.84
9	2019	103.4	166	10.51
10	2020	108.2	147.5	10.59

Again, we can see as off road vehicle emission levels lower, the value of a house rises, and each year off road vehicle emissions contribute less and less to total emissions.

Data Maintenance and Management

- ## Regular Data Updates
  - **We set up a regular annual schedule for importing new data from Environment and Climate Change Canada and Statistics Canada.**
- ## Data Cleaning Procedures
  - **Implement data cleaning procedures to handle missing, duplicate, or incorrect data. This procedure will ensure that any changes applied to the data are consistent with previous versions.**
- ## Error Logging and Resolution
  - **We can keep a log of any errors that occurred and create solutions to resolve said errors so they won't persist in the future.**

Thank you!

References (DataSets Used)

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810020501

https://open.canada.ca/data/en/dataset/ad009bb9-118d-49ce-a57a-f7b79e3e032f/resource/bb4803ec-a4f7-4b5a-b36c-170b6a5c3423

https://open.canada.ca/data/en/dataset/faee4b68-b62a-45c1-88a8-c974f5f19a50/resource/6e0cf73f-e44c-40d8-9673-1dca09860821