Abstract

OBJECTIVES: This project is an analysis of air pollution in Nova Scotia, as air pollution is detrimental to health, the environment, and the economy. In this era of green and smart cities initiatives, it is important to have accurate information about air pollution. This project involves acquiring the datasets, cleaning, extracting, transforming, loading, and visualizing the results of the analysis.
METHODS: This project involved obtaining data from an open data portal on the internet, loading the data into our analytical environment, exploring the data using analytical software, cleaning the data, building a data warehouse, determining feature and outcome variables, and visualizing the data. Cloud Dataprep was utilized to perform exploratory analysis and cleaning. Oracle 11g Express Edition was used to build the data warehouse. Tableau was used for visualization. The datasets are air pollution data in Nova Scotia.
RESULTS: After preprocessing in Cloud Dataprep and ETL (Extract Transform Load) in Oracle DB, Tableau was utilized to give insights into regions in Nova Scotia which have the most pollution now and in the past, and gave information on pollution levels in the future.
CONCLUSION: In this era of green and smart cities initiatives, it is important for stakeholders to have access to important air quality indicators, such as air pollution from fine particulate matter, smaller than 2.5 micrometers per cubic metre. As such, stakeholders need information regarding problem pollution areas in Nova Scotia, so that interventions can be initiated to decrease the air pollution in those locales. Also, accurate information is needed to determine the effects that interventions will have on decreasing air pollution. Finally, accurate forecasts are needed to anticipate problem pollution areas, so that stakeholders can address pollution problems before they occur.

Objectives

This project is an analysis of air pollution in Nova Scotia, as air pollution is detrimental to health, the environment, and the economy. In this era of green and smart cities initiatives, it is important to have accurate information about air pollution. This project involves acquiring the datasets, cleaning, extracting, transforming, loading, and visualizing the results of the analysis.

Problem

Air Pollution is problematic in Nova Scotia, as it is detrimental to health (i.e. asthma) (Weerasinghe 2017), detrimental to the environment, and detrimental to the economy of Nova Scotia. As such, it is important that we fully understand the current and future air quality conditions, as many areas are impacted by this variable.

Business Case

There has been a recent push for green and smart cities, focusing on clean industries and technologies. This green initiative is especially important for Nova Scotia, as its economy is based on tourism and its world-renown nature. Federal and Provincial Ministers have been focused on these green initiatives, and want to bring it to Nova Scotia. To carry out these green initiatives, policy makers need:

- Accurate and convenient access to current and future air pollution conditions,
- to highlight pollution problems areas,
- to measure the effects of interventions to reduce pollution, and
- to predict future pollution problems.

This project will allow the Government of Nova Scotia to track pollution (fine particulate matter) in various locations around the province. This project will analyze which regions have the most pollution, and which ones have the least. In addition, the project will determine trends over time, to see if pollution in certain areas are decreasing, increasing, or staying the same.

Methods

This study involved obtaining an air pollution dataset from an open data portal on the internet, loading the data into our analytical environment, cleaning the data, extracting the data, transforming the data, loading the data, and visualizing the data.

The team divided the work into the following tasks:

- Project topic selection (C.C., F.A., N.K.)
- Dataset selection (C.C.)
- Data cleaning (C.C.)
- ETL (F.A.)

- Visualization (N.K.)
- Research, report writing, and editing (C.C., N.K., F.A.)

Analytic Environment: Software

- Oracle Database 11g Express Edition Release 11.2.0.2.0 - 64bit Production
- Oracle SQL Developer
- SQL*Loader
- MS Excel
- Cloud Dataprep
- Tableau
- Draw.io

Analytic Environment: Hardware

Apple MacBook Air Processor: 1.6 GHz Intel Core i5; Number of Processors: 1; Total Number of Cores: 2; L2 Cache (per Core): 256 KB; L3 Cache: 3 MB; Memory: 8 GB 1600 MHz DDR3; Operating System: Apple Mac OS X v10.0

Acer Aspire ES1-512 Processor: 2.16GHz Intel Celeron; Number of Processors: 1; Total Number of Cores: 2; Memory: 4 GB DDR3; Operating System: Microsoft Windows 10 Home

Acquire the Air Pollution Data

The data set was obtained from the Nova Scotia Open Data Portal: https://data.novascotia.ca.

To obtain the data, the following procedure was followed:

1. Go to https://data.novascotia.ca
2. Click on button 'Data Catalogue'
3. In the search bar, search for 'pollution air particulate'
4. You will see 30 results: sort by 'Most Relevant'
5. You want the first 12 results (ignore the 11th result, which is a lookup table)

This will leave you with 11 datasets for analysis.

Motivation:

We chose these datasets as it contained enough data for us to extract insights into a business problem. When we were searching for datasets to analyze at the start, we noticed that many of the open datasets available had much sparcity or was not populated fully, and also the datasets were small (small number of rows and columns). We ended up extracting the air particulate pollution datasets from the Nova Scotia Open Data Portal, as there were several dataset on air particulate matter from all around the Province. We chose these datasets as it gave us a chance to extract multiple datasets and perform ETL and analytics on them.

Data Description

The 11 csv files have a total of 44 MB and 643,453 rows:

Nova_Scotia_Provincial_Ambient_Fine_Particulate_Matter__Kentville_BAM.csv, 498 KB, 8,785 rows

Nova_Scotia_Provincial_Ambient_Fine_Particulate_Matter__Halifax_BAM.csv, 4.7 MB, 69,958 rows

Nova_Scotia_Provincial_Ambient_Fine_Particulate_Matter__Port Hawkesbury_BAM.csv, 4.1 MB, 55,559 rows

Nova_Scotia_Provincial_Ambient_Fine_Particulate_Matter__Sable_Island_BAM.csv, 4.6 MB, 64,496 rows

Nova_Scotia_Provincial_Ambient_Fine_Particulate_Matter__Pictou_BAM.csv, 6.8 MB, 104,033 rows

Nova_Scotia_Provincial_Ambient_Fine_Particulate_Matter__Sydney_BAM.csv, 3.4 MB, 50,752 rows

Nova_Scotia_Provincial_Ambient_Fine_Particulate_Matter__Lake_Major_TEOM.csv, 4 MB, 60,691 rows

Nova_Scotia_Provincial_Ambient_Fine_Particulate_Matter__Halifax_TEOM.csv, 790 KB, 12,494 rows

Nova_Scotia_Provincial_Ambient_Fine_Particulate_Matter__Aylesford_BAM.csv, 4.7 MB, 68,025 rows

Nova_Scotia_Provincial_Ambient_Fine_Particulate_Matter__Lake_Major_BAM.csv, 6.7 MB, 96,433 rows

Nova_Scotia_Provincial_Ambient_Fine_Particulate_Matter__Sydney_TEOM.csv, 3.8 MB, 60,514 rows

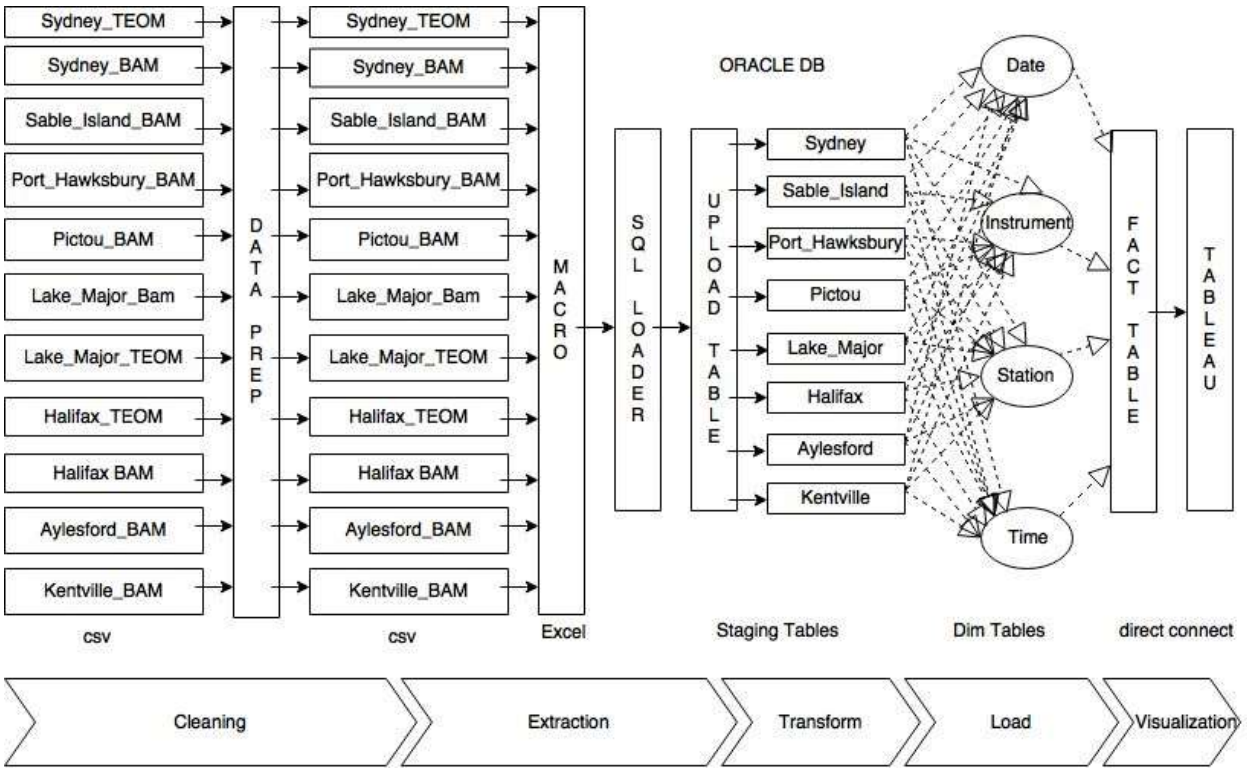
The following shows a sample of the data, and description of each column:

Column	Sample Data	Description

Date_Time	07/24/1998 03:00:00 PM	date-time stamp
Pollutant	PM2.5	fine particulate matter
Unit	¬µg/m3	micrograms per cubic metre
Station	Sydney	location of instrument
Instrument	TEOM	instrument type
Average	24	average concentration

Process Flow

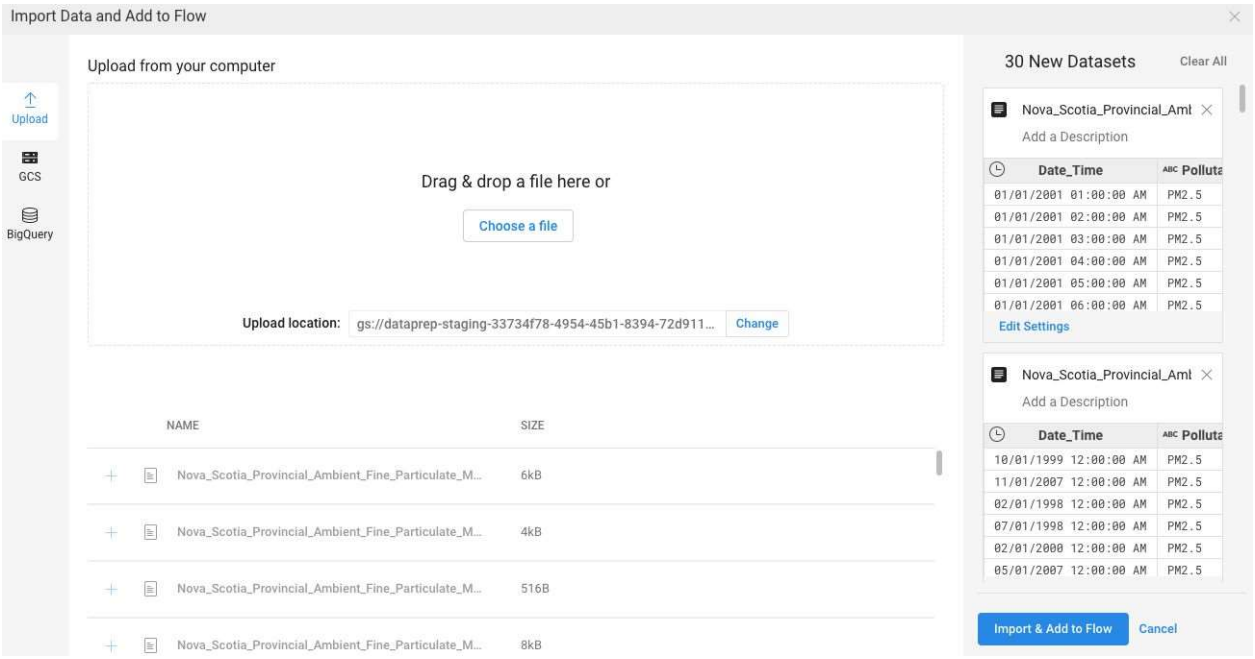
The following flowchart maps the process flow in the project:



Cleaning

We chose to perform preprocessing of the data in Cloud Dataprep. We uploaded the 11 csv files that were extracted from the Nova Scotia Open Data Portal. Below are our preprocessing steps:

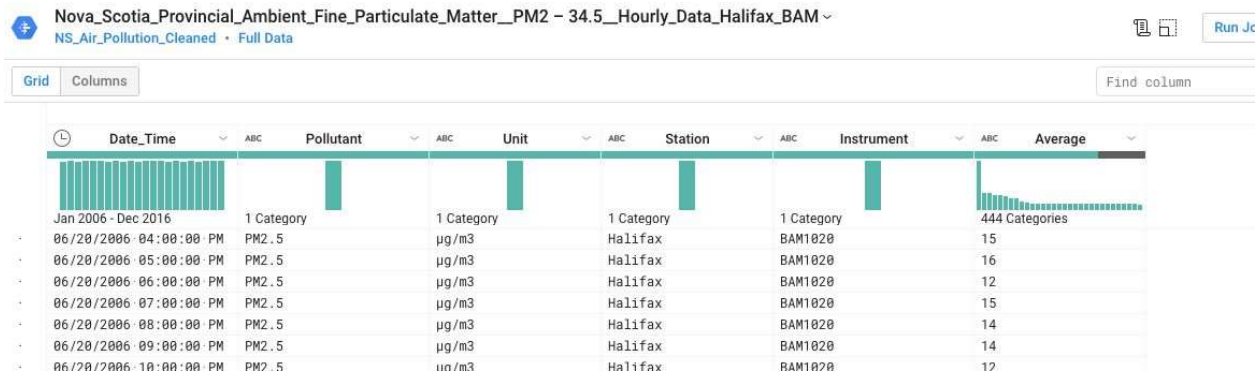
- Import the Data and Add to Flow. We upload each dataset into Cloud Dataprep:



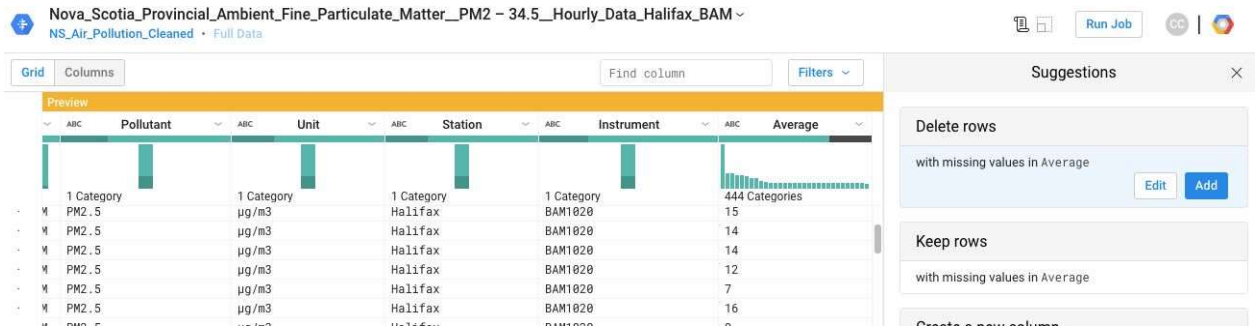
- Add new recipe. Here, we add a recipe to remove rows with null values:

- Edit new recipe. Here, we edit a recipe to remove rows with null values:

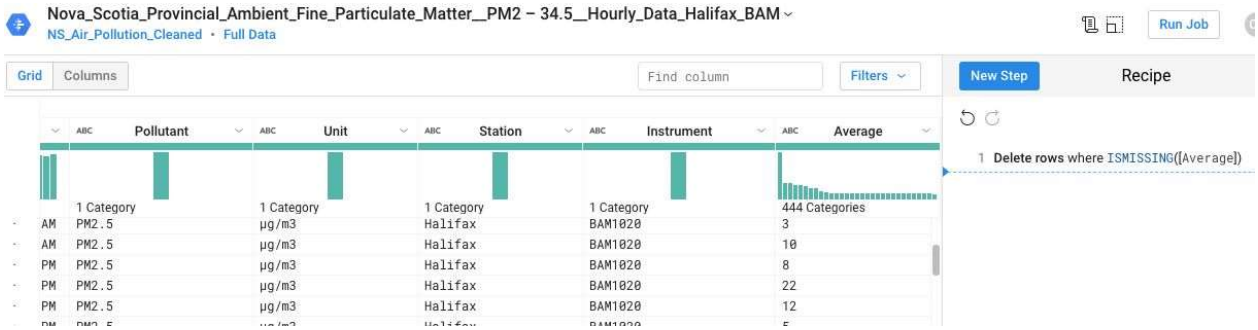
- View columns:



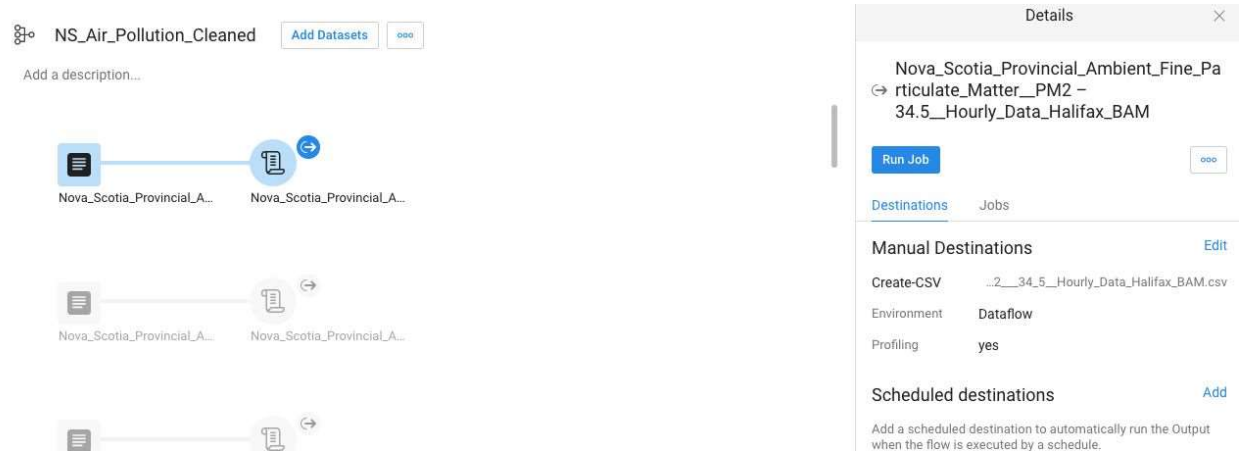
- Delete rows with missing values:



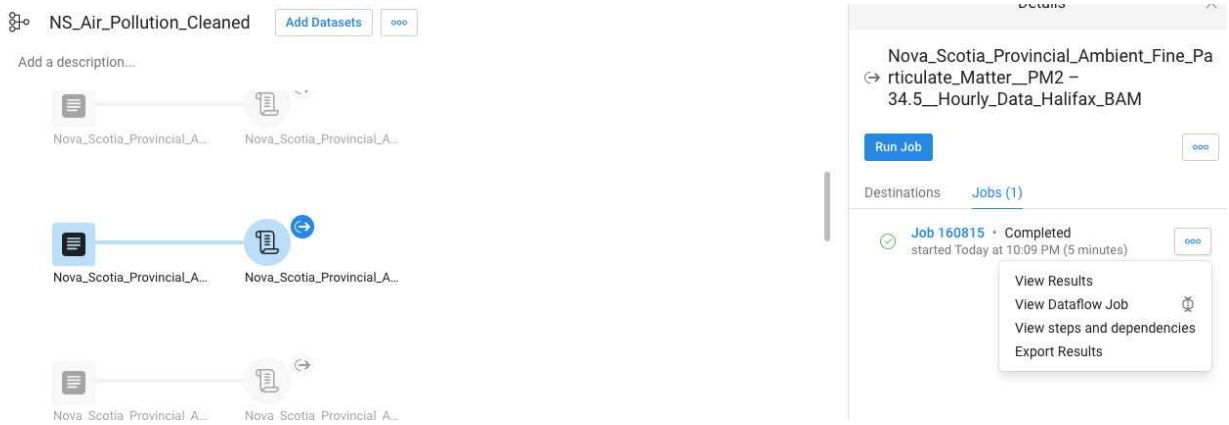
- Rows with missing values removed:



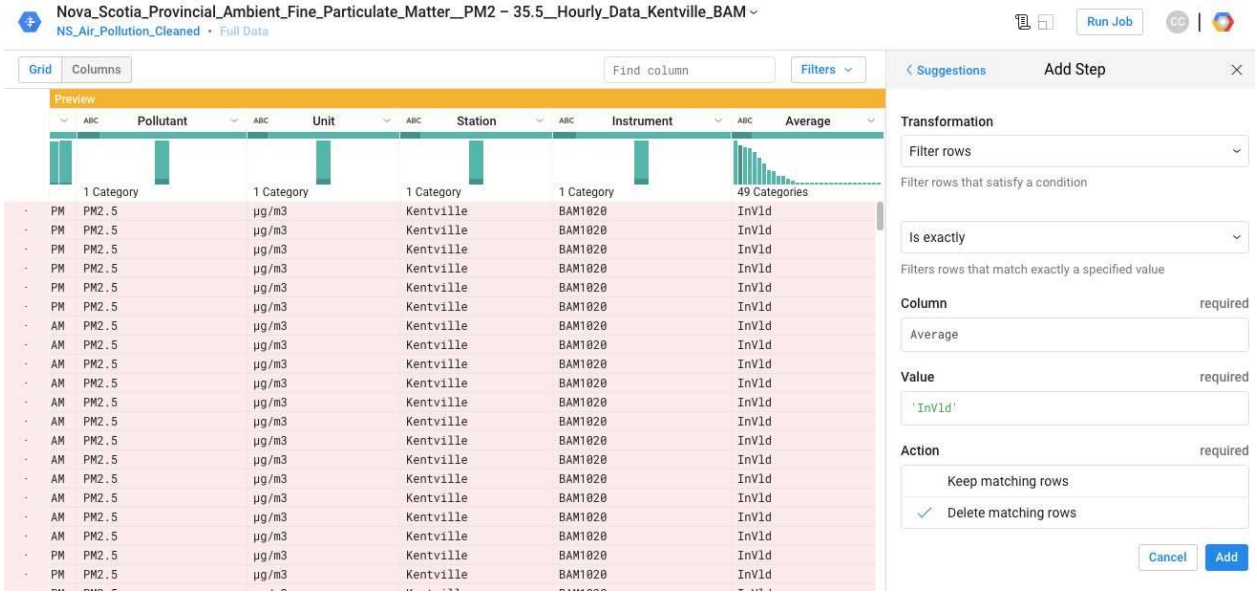
- Set up the data to be extracted to a csv file:



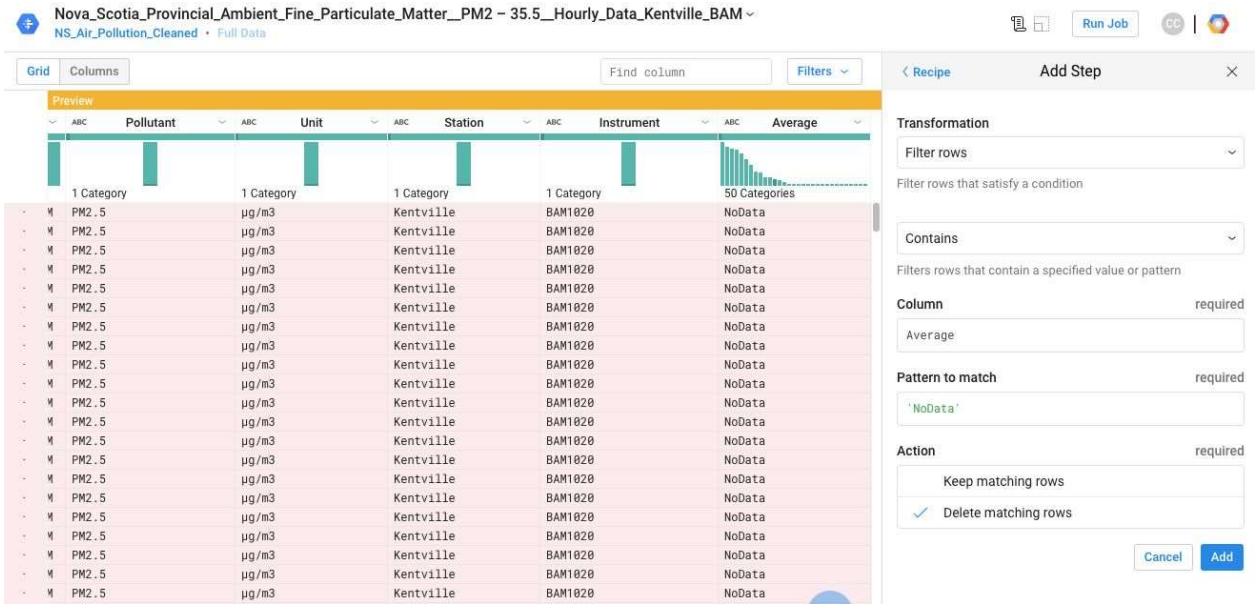
- Download csv file to localhost:



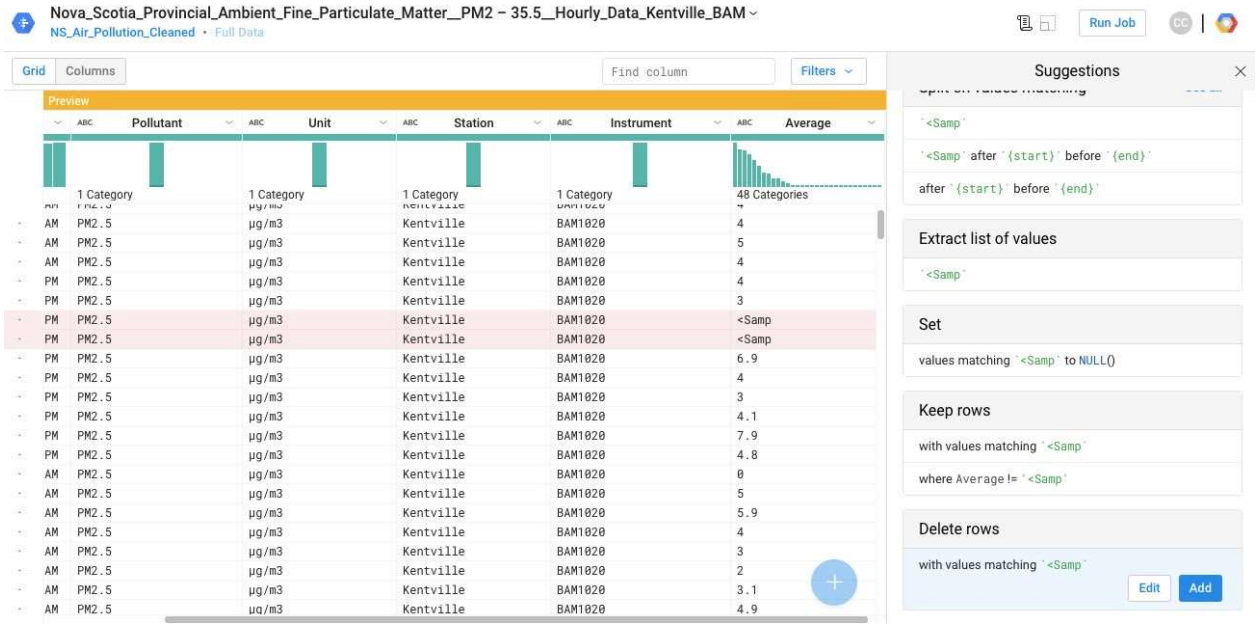
- Further preprocessing of dataset- remove rows with ‘InVld’:



- Further preprocessing of dataset- remove rows with ‘NoData’:



- Further preprocessing of dataset- remove rows with ‘<Samp’:

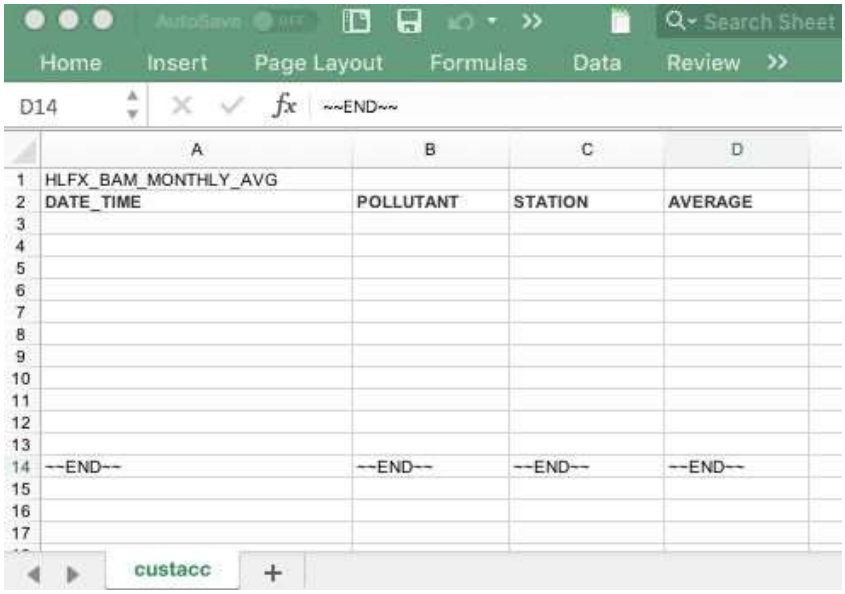


Extraction

After the data are preprocessed in Cloud Dataprep, the data is exported as 11 cleaned csv files. Utilizing localhost Oracle Database 11g Express Edition, these 11 csv files are then extracted to staging tables via Excel Macro using SQL*Loader. Before the data is extracted, the upload table is created (please see the Appendix for the create upload table script):

DWUSER.UPLOAD_TABLE	
DATE_TIME	VARCHAR2 (250 BYTE)
POLLUTANT	VARCHAR2 (250 BYTE)
UNIT	VARCHAR2 (250 BYTE)
STATION	VARCHAR2 (250 BYTE)
INSTRUMENT	VARCHAR2 (250 BYTE)
AVERAGE	FLOAT (126)

Next, an Excel Macro is utilized to load the csv files via SQL Loader:




```
-- Run SQL*Loader via control files, in conjunction with Excel Macros:

LOAD DATA

INFILE *
```

```
APPEND

INTO TABLE HLFX_BAM_MONTHLY_AVG

Fields terminated by '~'

Trailing Nullcols

(DATE_TIME,POLLUTANT,STATION,AVERAGE)

BEGINDATA

9-Jan-2008~PM2.5~Halifax~4.814259
```

All 11 csv files can be extracted via this method, utilizing the Excel Macro via SQL*Loader. The biggest advantage of using the Macro is that the extraction process is automated, rather than manually having to extract each csv, each with its own control file for SQL*Loader.

Testing

Before proceeding with the rest of the workflow, testing was performed to ensure the csv files were extracted properly. Here are the contents of the log file of the initial extraction of csv files into the Oracle DB table:

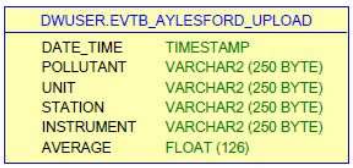
```
SQL*Loader: Release 11.2.0.2.0 - Production on Fri Mar 9 14:05:54 2018

Copyright (c) 1982, 2009, Oracle and/or its affiliates.  All rights reserved.

Control File:   C:\JUNK\INTREVERSE.CTL

Data File:      C:\JUNK\INTREVERSE.CTL

  Bad File:       C:\JUNK\INTREVERSE.bad

  Discard File:  none specified

 (Allow all discards)
Number to load: ALL
Number to skip: 0
Errors allowed: 50
```

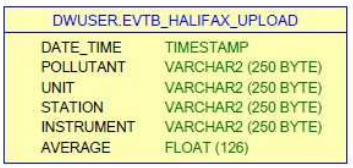
```
Bind array:       64 rows, maximum of 256000 bytes
Continuation:     none specified
Path used:        Conventional

Table HLFX_BAM_MONTHLY_AVG, loaded from every logical record.
Insert option in effect for this table: APPEND
TRAILING NULLCOLS option in effect

   Column Name                  Position   Len  Term Encl Datatype
------------------------------ ---------- ----- ---- ----
---------------------
DATE_TIME                           FIRST      *   ~       CHARACTER
POLLUTANT                           NEXT   *   ~       CHARACTER
STATION                              NEXT      *   ~       CHARACTER
AVERAGE                              NEXT      *   ~       CHARACTER


Table HLFX_BAM_MONTHLY_AVG:
  132 Rows successfully loaded.
  0 Rows not loaded due to data errors.
  0 Rows not loaded because all WHEN clauses were failed.
  0 Rows not loaded because all fields were null.
```

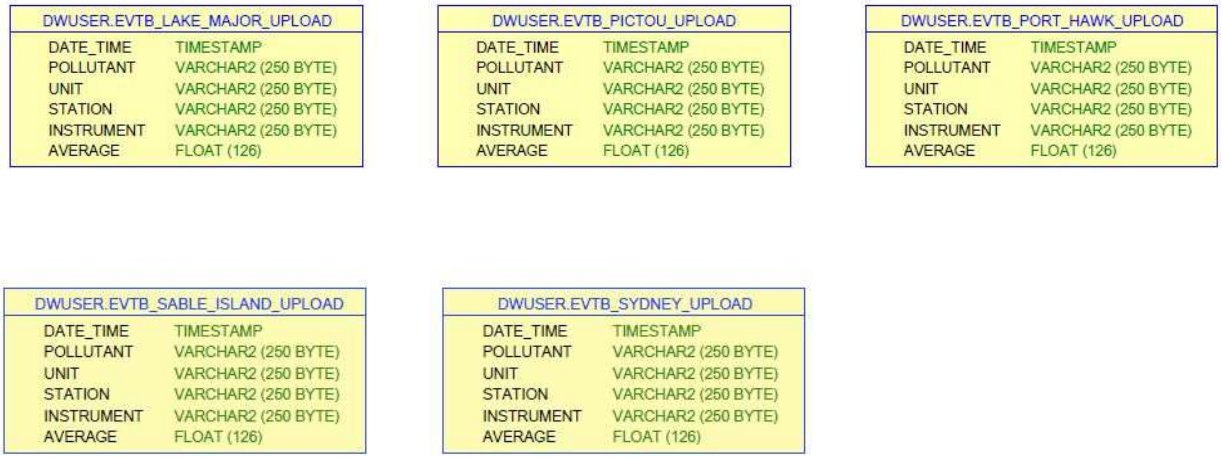
As the testing revealed no errors with the initial extraction, the workflow commenced.

Transformation

After extraction of the data into the one upload table, transformation commenced with the partitioning of the upload table into tables partitioned by STATION (location). Below are the created upload tables partitioned by STATION:





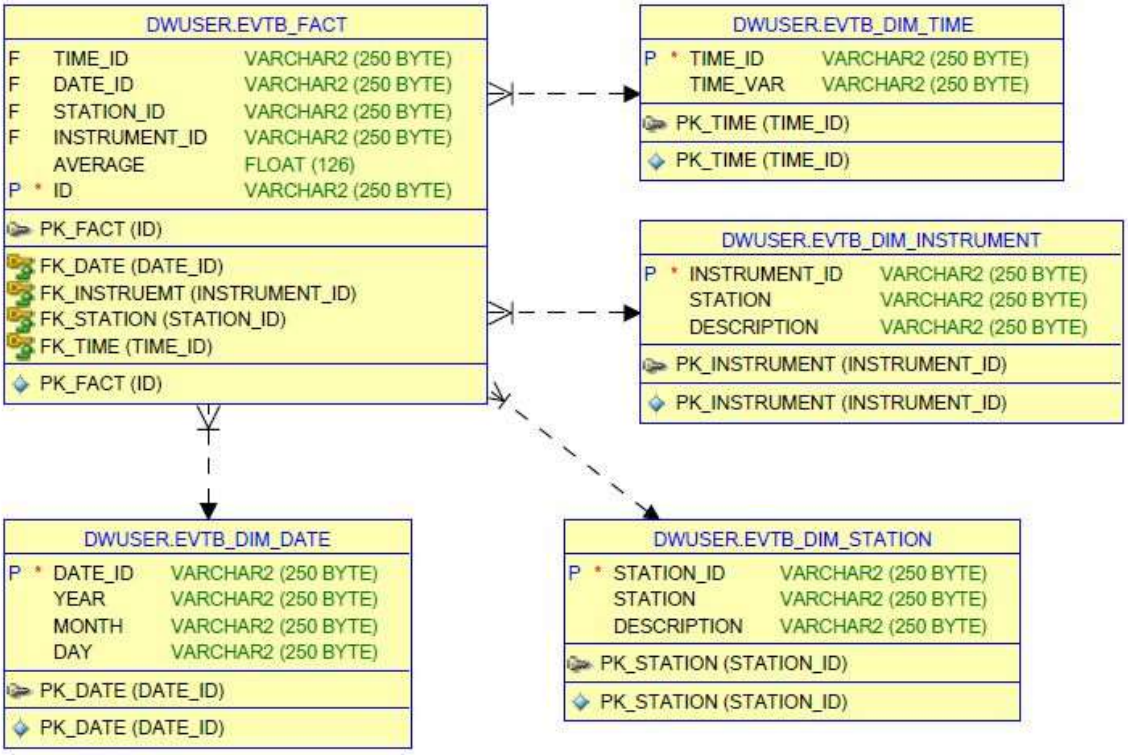


Please see the Appendix for the create table scripts for all the upload tables partitioned by STATION.

Load

After the data are uploaded into the partitioned STATION upload tables, the next step is to load the DIM (dimension) tables, which will feed the FACT table. The DIM tables are comprised of TIME, INSTRUMENT, STATION, and DATE. The FACT table has all of the inputs from the DIM tables, each contributing to the FACT table measurement AVERAGE (pollution average measurement), and the child FACT table containing all the FK (foreign keys) referencing back to the parent DIM tables (see Star Schema below):

Star Schema



Please see the appendix for the complete create table scripts for FACT and DIM tables.

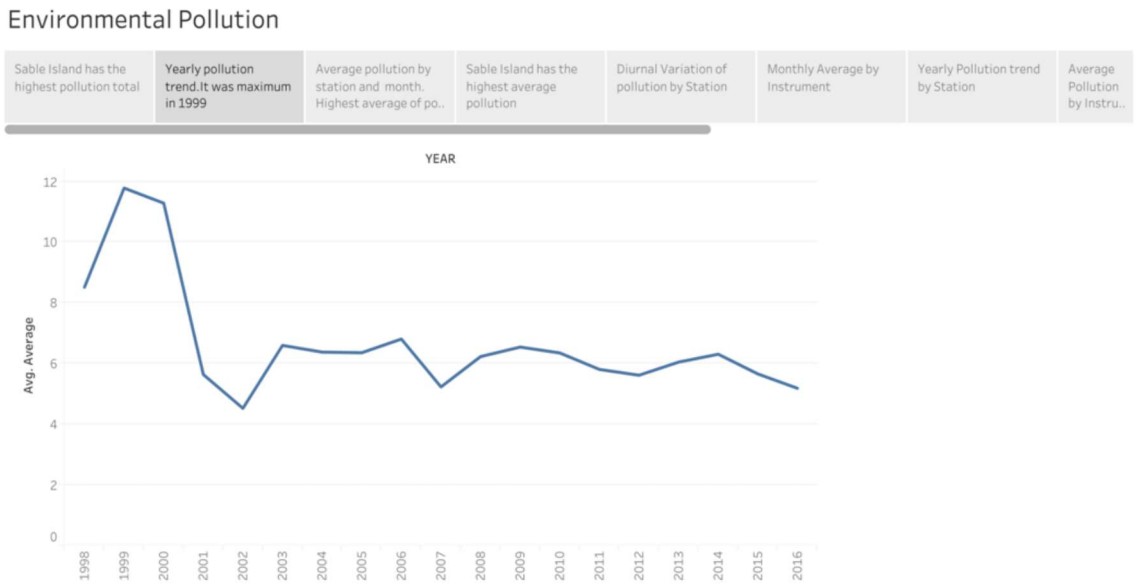
Results

The FACT table and Dim tables from the Oracle database are subsequently connected directly to Tableau for visualization.

Visualize the Data

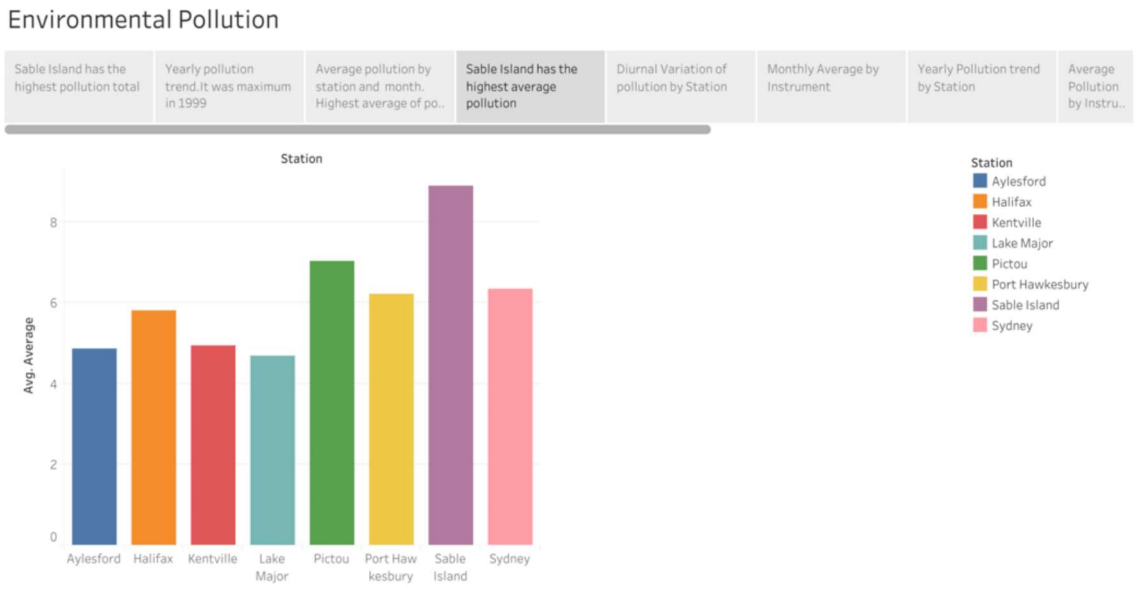
The following graphs are obtained from Tableau, where the FACT table and DIM tables are directly loaded from the Oracle DB into Tableau. When the FACT table and DIM tables are loaded into Tableau, the DIM tables are joined to the FACT tables, according to their Primary Key / Foreign Key relationships.

In this location graph, Sable Island has the highest cumulative pollution total.

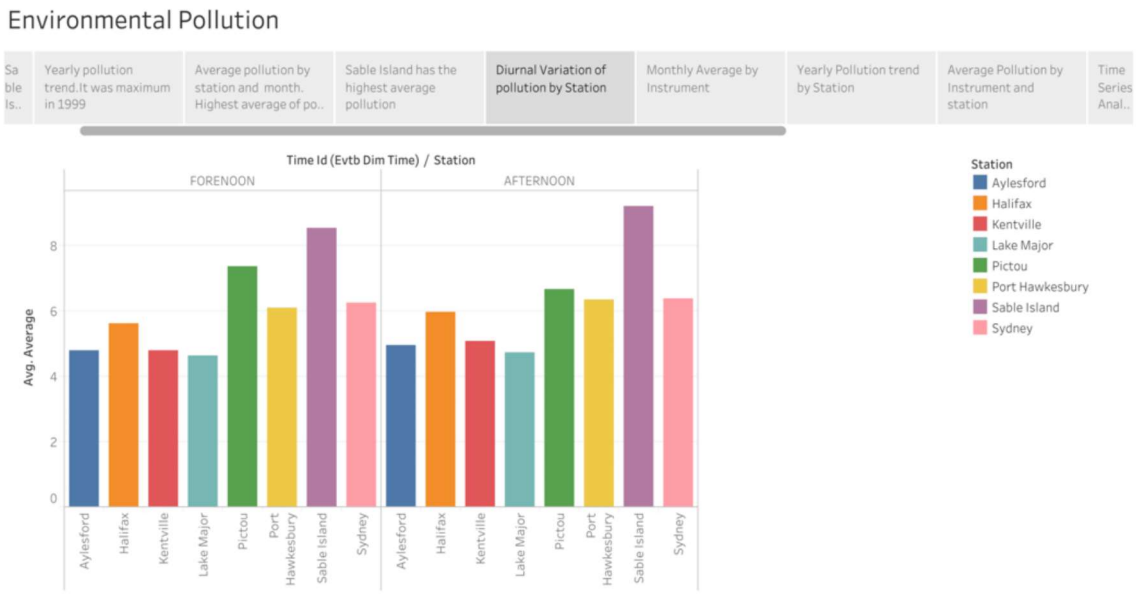


In this yearly pollution trend graph, it was maximum in 1999, then dropped drastically between 2000 and 2002.

This graph shows average pollution by station and month. The highest average is in Sable Island for the month of July.

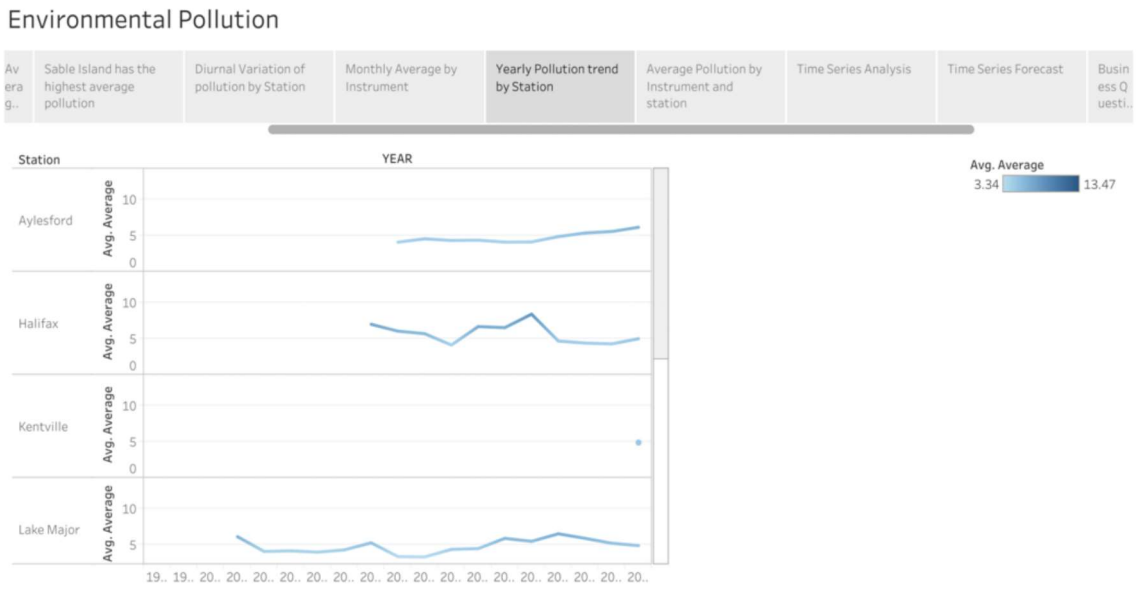


This graph shows Sable Island has the highest average pollution.



This graph shows the diurnal variation of pollution by station.

This graph shows the average pollution grouped by instrument and month. July is the month with the highest average.

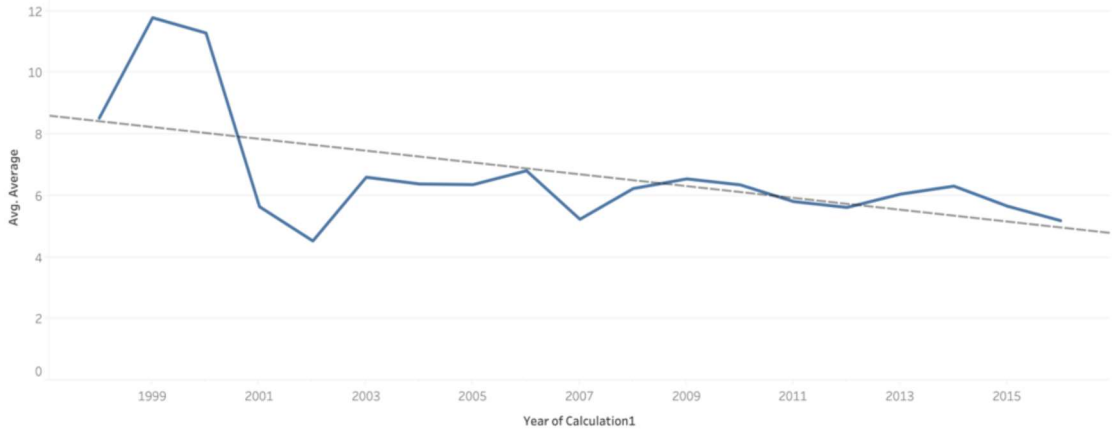


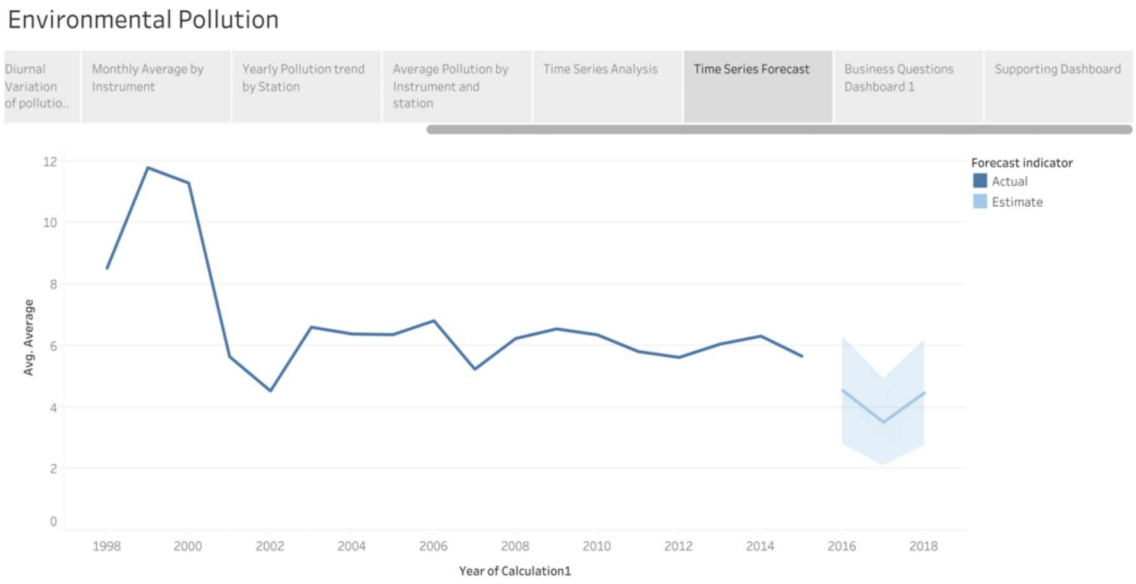
Here is the yearly pollution trend by station.

This graph on average pollution grouped by station and instrument shows Sable Island to have the highest average pollution.

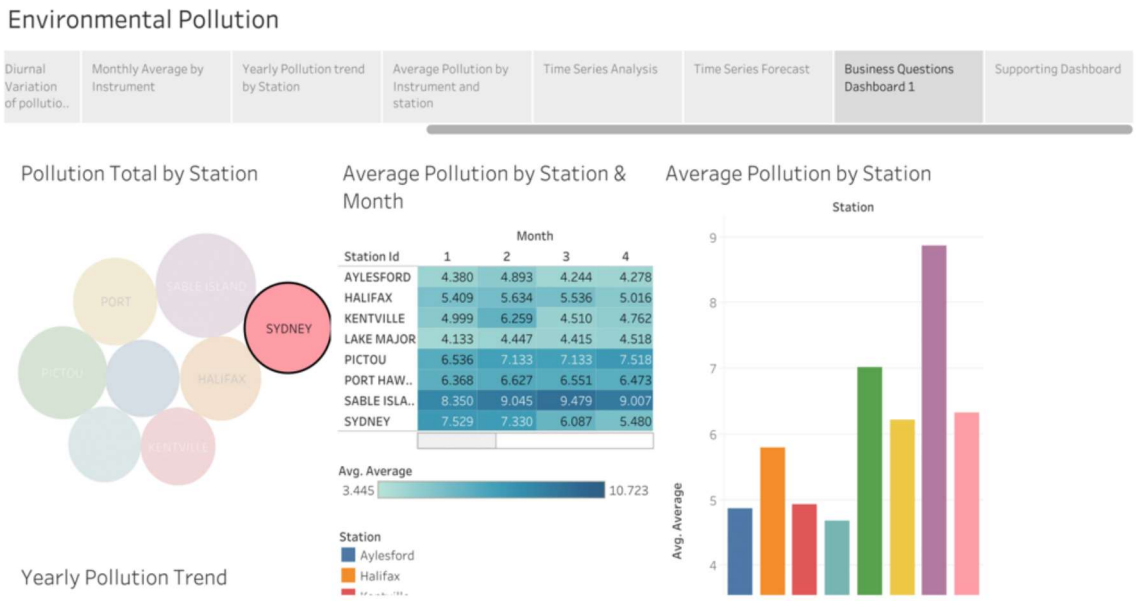
Environmental Pollution

Diurnal Variation of pollutio.. | Monthly Average by Instrument | Yearly Pollution trend by Station | Average Pollution by Instrument and station | **Time Series Analysis** | Time Series Forecast | Business Questions Dashboard 1 | Supporting Dashboard

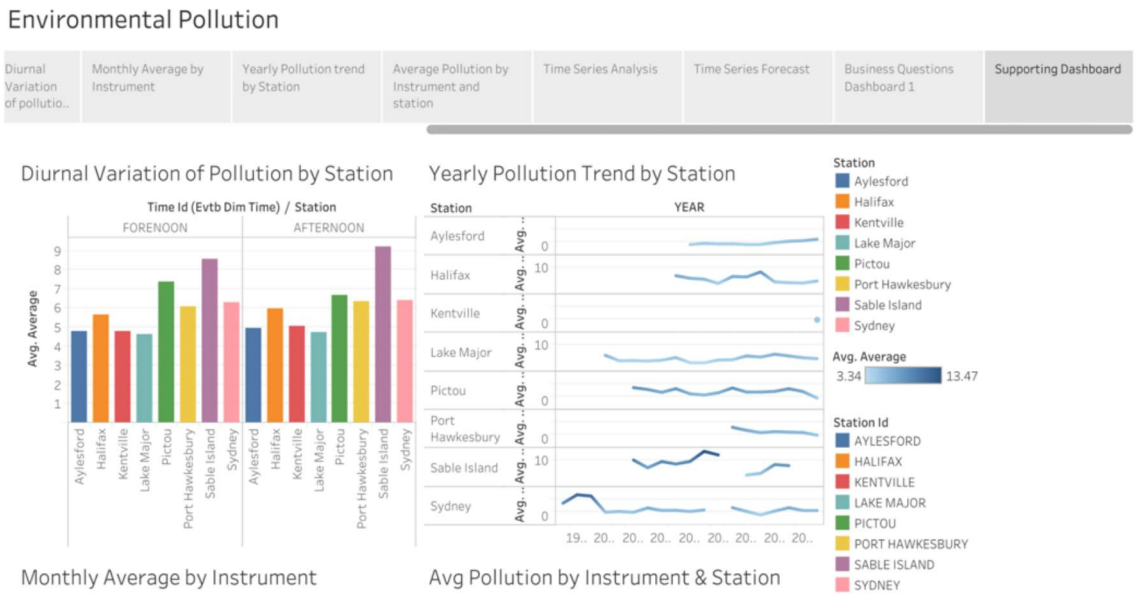




This time series forecast shows that there is a trend for lower pollution in the next couple of years.



Here is a dashboard of the visualizations.



Conclusion

In this era of green and smart cities initiatives, it is important for stakeholders to have access to important air quality indicators, such as air pollution from fine particulate matter, smaller than 2.5 micrometers per cubic metre. As such, stakeholders need information regarding problem pollution areas in Nova Scotia, so that interventions can be initiated to decrease the air pollution in that locale. Also, accurate information is needed to determine the effects that interventions will have on decreasing air pollution. Finally, accurate forecasts are needed to anticipate problem pollution areas, so that stakeholders can address pollution problems before they occur.

This project is the proof of concept that the Nova Scotia government should look into developing further, given its importance to health, the environment, and the economy.

References

Nova Scotia Open Data Portal: https://data.novascotia.ca, accessed March 25, 2018.

Weerasinghe S (2017). Statistical modeling of complex health outcomes and air pollution data: Application of air quality health indexing for asthma risk assessment. *Epidemiology Biostatistics and Public Health*, Volume 14, Number 1.

Appendix

Powerpoint Presentation of Analysis

https://www.slideshare.net/carlocarandang/analysis-of-air-pollution-in-nova-scotia-presentation

Oracle Database Script

```
-- Create DIM Tables
  CREATE TABLE "DWUSER"."EVTB_DIM_DATE"
   (  "DATE_ID" VARCHAR2(250 BYTE),
      "YEAR" VARCHAR2(250 BYTE),
      "MONTH" VARCHAR2(250 BYTE),
      "DAY" VARCHAR2(250 BYTE),
       CONSTRAINT "PK_DATE" PRIMARY KEY ("DATE_ID")
  USING INDEX PCTFREE 10 INITRANS 2 MAXTRANS 255 COMPUTE STATISTICS
  STORAGE(INITIAL 65536 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "SYSTEM"  ENABLE
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 65536 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "SYSTEM" ;

    CREATE TABLE "DWUSER"."EVTB_DIM_INSTRUMENT"
   (  "INSTRUMENT_ID" VARCHAR2(250 BYTE),
      "STATION" VARCHAR2(250 BYTE),
      "DESCRIPTION" VARCHAR2(250 BYTE),
       CONSTRAINT "PK_INSTRUMENT" PRIMARY KEY ("INSTRUMENT_ID")
  USING INDEX PCTFREE 10 INITRANS 2 MAXTRANS 255 COMPUTE STATISTICS
  STORAGE(INITIAL 65536 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "SYSTEM"  ENABLE
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 65536 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
```

```
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "SYSTEM" ;

    CREATE TABLE "DWUSER"."EVTB_DIM_STATION"
   (  "STATION_ID" VARCHAR2(250 BYTE),
      "STATION" VARCHAR2(250 BYTE),
      "DESCRIPTION" VARCHAR2(250 BYTE),
       CONSTRAINT "PK_STATION" PRIMARY KEY ("STATION_ID")
  USING INDEX PCTFREE 10 INITRANS 2 MAXTRANS 255 COMPUTE STATISTICS
  STORAGE(INITIAL 65536 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "SYSTEM"  ENABLE
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 65536 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "SYSTEM" ;

    CREATE TABLE "DWUSER"."EVTB_DIM_TIME"
   (  "TIME_ID" VARCHAR2(250 BYTE),
      "TIME_VAR" VARCHAR2(250 BYTE),
       CONSTRAINT "PK_TIME" PRIMARY KEY ("TIME_ID")
  USING INDEX PCTFREE 10 INITRANS 2 MAXTRANS 255 COMPUTE STATISTICS
  STORAGE(INITIAL 65536 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "SYSTEM"  ENABLE
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 65536 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "SYSTEM" ;

-- Create FACT Table
    CREATE TABLE "DWUSER"."EVTB_FACT"
   (  "TIME_ID" VARCHAR2(250 BYTE),
      "DATE_ID" VARCHAR2(250 BYTE),
      "STATION_ID" VARCHAR2(250 BYTE),
      "INSTRUMENT_ID" VARCHAR2(250 BYTE),
```

```
      "AVERAGE" FLOAT(126),
      "ID" VARCHAR2(250 BYTE),
       CONSTRAINT "PK_FACT" PRIMARY KEY ("ID")
  USING INDEX PCTFREE 10 INITRANS 2 MAXTRANS 255 COMPUTE STATISTICS
  STORAGE(INITIAL 65536 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "SYSTEM"  ENABLE,
       CONSTRAINT "FK_DATE" FOREIGN KEY ("DATE_ID")
        REFERENCES "DWUSER"."EVTB_DIM_DATE" ("DATE_ID") ENABLE,
       CONSTRAINT "FK_INSTRUEMT" FOREIGN KEY ("INSTRUMENT_ID")
        REFERENCES "DWUSER"."EVTB_DIM_INSTRUMENT" ("INSTRUMENT_ID") ENABLE,
       CONSTRAINT "FK_STATION" FOREIGN KEY ("STATION_ID")
        REFERENCES "DWUSER"."EVTB_DIM_STATION" ("STATION_ID") ENABLE,
       CONSTRAINT "FK_TIME" FOREIGN KEY ("TIME_ID")
        REFERENCES "DWUSER"."EVTB_DIM_TIME" ("TIME_ID") ENABLE
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 65536 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "SYSTEM" ;

-- Create upload tables partitioned by STATION
    CREATE TABLE "DWUSER"."EVTB_HALIFAX_UPLOAD"
   (  "DATE_TIME" TIMESTAMP (6),
      "POLLUTANT" VARCHAR2(250 BYTE),
      "UNIT" VARCHAR2(250 BYTE),
      "STATION" VARCHAR2(250 BYTE),
      "INSTRUMENT" VARCHAR2(250 BYTE),
      "AVERAGE" FLOAT(126)
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 5242880 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "FITWORLDGYM_STG" ;

   CREATE TABLE "DWUSER"."EVTB_KENTVILLE_UPLOAD"
   (  "DATE_TIME" TIMESTAMP (6),
      "POLLUTANT" VARCHAR2(250 BYTE),
      "UNIT" VARCHAR2(250 BYTE),
      "STATION" VARCHAR2(250 BYTE),
```

```
      "INSTRUMENT" VARCHAR2(250 BYTE),
      "AVERAGE" FLOAT(126)
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 524288 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "FITWORLDGYM_STG" ;

    CREATE TABLE "DWUSER"."EVTB_LAKE_MAJOR_UPLOAD"
   (  "DATE_TIME" TIMESTAMP (6),
      "POLLUTANT" VARCHAR2(250 BYTE),
      "UNIT" VARCHAR2(250 BYTE),
      "STATION" VARCHAR2(250 BYTE),
      "INSTRUMENT" VARCHAR2(250 BYTE),
      "AVERAGE" FLOAT(126)
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 8388608 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "FITWORLDGYM_STG" ;

    CREATE TABLE "DWUSER"."EVTB_PICTOU_UPLOAD"
   (  "DATE_TIME" TIMESTAMP (6),
      "POLLUTANT" VARCHAR2(250 BYTE),
      "UNIT" VARCHAR2(250 BYTE),
      "STATION" VARCHAR2(250 BYTE),
      "INSTRUMENT" VARCHAR2(250 BYTE),
      "AVERAGE" FLOAT(126)
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 6291456 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "FITWORLDGYM_STG" ;

    CREATE TABLE "DWUSER"."EVTB_PORT_HAWK_UPLOAD"
   (  "DATE_TIME" TIMESTAMP (6),
      "POLLUTANT" VARCHAR2(250 BYTE),
      "UNIT" VARCHAR2(250 BYTE),
      "STATION" VARCHAR2(250 BYTE),
      "INSTRUMENT" VARCHAR2(250 BYTE),
```

```
      "AVERAGE" FLOAT(126)
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 4194304 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "FITWORLDGYM_STG" ;


  CREATE TABLE "DWUSER"."EVTB_SABLE_ISLAND_UPLOAD"
   (  "DATE_TIME" TIMESTAMP (6),
      "POLLUTANT" VARCHAR2(250 BYTE),
      "UNIT" VARCHAR2(250 BYTE),
      "STATION" VARCHAR2(250 BYTE),
      "INSTRUMENT" VARCHAR2(250 BYTE),
      "AVERAGE" FLOAT(126)
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 4194304 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "FITWORLDGYM_STG" ;

    CREATE TABLE "DWUSER"."EVTB_SYDNEY_UPLOAD"
   (  "DATE_TIME" TIMESTAMP (6),
      "POLLUTANT" VARCHAR2(250 BYTE),
      "UNIT" VARCHAR2(250 BYTE),
      "STATION" VARCHAR2(250 BYTE),
      "INSTRUMENT" VARCHAR2(250 BYTE),
      "AVERAGE" FLOAT(126)
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 6291456 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "FITWORLDGYM_STG" ;

-- Create upload table
    CREATE TABLE "DWUSER"."UPLOAD_TABLE"
   (  "DATE_TIME" VARCHAR2(250 BYTE),
      "POLLUTANT" VARCHAR2(250 BYTE),
      "UNIT" VARCHAR2(250 BYTE),
      "STATION" VARCHAR2(250 BYTE),
```

```
      "INSTRUMENT" VARCHAR2(250 BYTE),
      "AVERAGE" FLOAT(126)
   ) SEGMENT CREATION IMMEDIATE
  PCTFREE 10 PCTUSED 40 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING
  STORAGE(INITIAL 46137344 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
  PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL DEFAULT FLASH_CACHE
DEFAULT CELL_FLASH_CACHE DEFAULT)
  TABLESPACE "FITWORLDGYM_STG" ;

    CREATE OR REPLACE FORCE VIEW "DWUSER"."EVVW_ALL" ("DATE_TIME",
"POLLUTANT", "UNIT", "STATION", "INSTRUMENT", "AVERAGE") AS
  SELECT "DATE_TIME","POLLUTANT","UNIT","STATION","INSTRUMENT","AVERAGE" FROM
EVTB_AYLESFORD_UPLOAD
  UNION ALL
  SELECT "DATE_TIME","POLLUTANT","UNIT","STATION","INSTRUMENT","AVERAGE" FROM
EVTB_HALIFAX_UPLOAD
  UNION ALL
  SELECT "DATE_TIME","POLLUTANT","UNIT","STATION","INSTRUMENT","AVERAGE" FROM
EVTB_LAKE_MAJOR_UPLOAD
  UNION ALL
  SELECT "DATE_TIME","POLLUTANT","UNIT","STATION","INSTRUMENT","AVERAGE" FROM
EVTB_PICTOU_UPLOAD
  UNION ALL
  SELECT "DATE_TIME","POLLUTANT","UNIT","STATION","INSTRUMENT","AVERAGE" FROM
EVTB_PORT_HAWK_UPLOAD
  UNION ALL
  SELECT "DATE_TIME","POLLUTANT","UNIT","STATION","INSTRUMENT","AVERAGE" FROM
EVTB_SABLE_ISLAND_UPLOAD
  UNION ALL
  SELECT "DATE_TIME","POLLUTANT","UNIT","STATION","INSTRUMENT","AVERAGE" FROM
EVTB_SYDNEY_UPLOAD
  UNION ALL
  SELECT "DATE_TIME","POLLUTANT","UNIT","STATION","INSTRUMENT","AVERAGE" FROM
EVTB_KENTVILLE_UPLOAD;
```