While booming economies often face the challenge of rising air pollution due to factors like industrial activity, proactive environmental policies and regulations can mitigate this impact and even improve air quality. Analysing this intricate relationship between manufacturing output, environmental performance, and air pollution levels is key to achieving sustainable economic growth.

Database
- Name: world_air_quality
- Commands to use the database:
    1. SHOW DATABASES;
    2. USE world_air_quality;

Tables
This analysis relies on four tables, three sourced from existing data and one providing AQI category definitions.

```
mysql> SHOW TABLES;
+--------------------------+
| Tables_in_world_air_quality |
+--------------------------+
| AQI                      |
| AirPollution             |
| EnvironmentalPerformance |
| ManufacturingOutput      |
+--------------------------+
4 rows in set (0.00 sec)
```

According to IQAir, the AQI categories are listed from least harmful to human health ('good') to the most harmful ('hazardous'). The 'AQI' table is as follows:

Category	MinPM	MaxPM
Good	0.0	12.0
Moderate	12.1	35.4
UnhealthyForSensitiveGroups	35.5	55.5
Unhealthy	55.6	150.4
VeryUnhealthy	150.5	250.4
Hazardous	250.5	500

IQAir mentions that PM (Particulate Matter) refers to particles found in the air including dust, soot, dirt, smoke, and liquid droplets. It is also added that PM2.5 particles measure 2.5 microns or less in diameter. PM2.5 particles are so small they can only be seen with an electron microscope.

The command "SHOW COLUMNS FROM AQI" lists up the columns with detailed information.

```
mysql> SHOW COLUMNS FROM AQI;
+----------+-------------+------+-----+---------+-------+
| Field    | Type        | Null | Key | Default | Extra |
+----------+-------------+------+-----+---------+-------+
| Category | varchar(30) | NO   | PRI | NULL    |       |
| MinPM    | float       | NO   |     | NULL    |       |
| MaxPM    | float       | NO   |     | NULL    |       |
+----------+-------------+------+-----+---------+-------+
3 rows in set (0.00 sec)
```

The command "SELECT * FROM AQI" brings all the data contained in the table.

```
mysql> SELECT * FROM AQI;
+---------------------------+-------+-------+
| Category                  | MinPM | MaxPM |
+---------------------------+-------+-------+
| Good                      |     0 |    12 |
| Moderate                  |  12.1 |  35.4 |
| UnhealthyForSensitiveGroups | 35.5 |  55.5 |
| Unhealthy                 |  55.6 | 150.4 |
| VeryUnhealthy             | 150.5 | 250.4 |
| Hazardous                 | 250.5 |   500 |
+---------------------------+-------+-------+
```

The remaining tables are three, and they only hold data collected in 2022:
- AirPollution
  - Columns in the original file
    - City
    - Country
    - Average PM 2.5 for each year from 2017 to 2023
  - Columns in the table

Column Name	Datatype	Key
City	VARCHAR(100)	PRIMARY
Country	VARCHAR(100)	PRIMARY
AveragePM	FLOAT	

Here is the list of the columns in the AirPollution table:

```
mysql> SHOW COLUMNS FROM AirPollution;
+-----------+--------------+------+-----+---------+-------+
| Field     | Type         | Null | Key | Default | Extra |
+-----------+--------------+------+-----+---------+-------+
| City      | varchar(100) | NO   | PRI | NULL    |       |
| Country   | varchar(100) | NO   | PRI | NULL    |       |
| AveragePM | float        | NO   |     | NULL    |       |
+-----------+--------------+------+-----+---------+-------+
3 rows in set (0.00 sec)
```

To extract the data from the AirPollution table, "SELECT * FROM AirPollution" is used. In the below screenshot, "LIMIT 10" is added so that only 10 rows can be printed.

```
mysql> SELECT * FROM AirPollution LIMIT 10;
+--------------+-------------+-----------+
| City         | Country     | AveragePM |
+--------------+-------------+-----------+
| Kabul        | Afghanistan |      18.1 |
| Tirana       | Albania     |      14.4 |
| Algiers      | Algeria     |      17.4 |
| Ordino       | Andorra     |       5.3 |
| Luanda       | Angola      |       8.7 |
| Buenos Aires | Argentina   |      14.2 |
| Cordoba      | Argentina   |       8.7 |
| General Pico | Argentina   |       7.2 |
| Mendoza      | Argentina   |       8.2 |
| Rafaela      | Argentina   |      10.7 |
+--------------+-------------+-----------+
```

- ManufacturingOutput
  - Columns in the original file
    - CountryName
    - CountryCode
    - IndicatorName
    - IndicatorCode
    - Output for each year from 1960 to 2022
  - Columns in the table

Column Name	Datatype	Key
Country	VARCHAR(100)	PRIMARY
ISOCode	CHAR(3)	PRIMARY
Output	FLOAT	

The details of each column are as follows:

```
mysql> SHOW COLUMNS FROM ManufacturingOutput;
+---------+--------------+------+-----+---------+-------+
| Field   | Type         | Null | Key | Default | Extra |
+---------+--------------+------+-----+---------+-------+
| Country | varchar(100) | NO   | PRI | NULL    |       |
| ISOCode | varchar(100) | NO   | PRI | NULL    |       |
| Output  | varchar(255) | YES  |     | NULL    |       |
+---------+--------------+------+-----+---------+-------+
3 rows in set (0.00 sec)
```

Only 10 rows of the ManufacturingOutput table are printed below:

```
mysql> SELECT * FROM ManufacturingOutput LIMIT 10;
+------------------------------+---------+-------------------+
| Country                      | ISOCode | Output            |
+------------------------------+---------+-------------------+
| Aruba                        | ABW     |                   |
| Africa Eastern and Southern  | AFE     | 118372151635.228  |
| Afghanistan                  | AFG     |                   |
| Africa Western and Central   | AFW     | 111386417453.795  |
| Angola                       | AGO     | 7933572389.55225  |
| Albania                      | ALB     | 1236498026.9812   |
| Andorra                      | AND     |                   |
| Arab World                   | ARB     | 465521862185.264  |
| United Arab Emirates         | ARE     |                   |
| Argentina                    | ARG     | 97371319994.1355  |
+------------------------------+---------+-------------------+
```

The World Bank has not released manufacturing output data for certain countries. In examining the connection between industrial activity and air pollution using the air quality dataset, only those countries with reported manufacturing output data and high levels of air pollution are considered.

- EnvironmentalPerformance
    - Columns in the original file
        - code
        - iso
        - country
        - region
        - EPI
    - Columns in the table to use

Column Name	Datatype	Key
ISOCode	CHAR(3)	PRIMARY
Country	VARCHAR(100)	PRIMARY
EPI	FLOAT	

The list of the column details is as below:

```
mysql> SHOW COLUMNS FROM EnvironmentalPerformance;
+---------+--------------+------+-----+---------+-------+
| Field   | Type         | Null | Key | Default | Extra |
+---------+--------------+------+-----+---------+-------+
| ISOCode | char(3)      | NO   | PRI | NULL    |       |
| Country | varchar(100) | NO   | PRI | NULL    |       |
| EPI     | float        | YES  |     | NULL    |       |
+---------+--------------+------+-----+---------+-------+
3 rows in set (0.00 sec)
```

The following screenshot contains only 10 rows of the EnvironmentalPerformance table.

```
mysql> SELECT * FROM EnvironmentalPerformance LIMIT 10;
+---------+----------------------+------+
| ISOCode | Country              | EPI  |
+---------+----------------------+------+
| AFG     | Afghanistan          | 43.6 |
| AGO     | Angola               | 30.5 |
| ALB     | Albania              | 47.1 |
| ARE     | United Arab Emirates | 52.4 |
| ARG     | Argentina            | 41.1 |
| ARM     | Armenia              | 48.3 |
| ATG     | Antigua and Barbuda  | 52.4 |
| AUS     | Australia            | 60.1 |
| AUT     | Austria              | 66.5 |
| AZE     | Azerbaijan           | 38.6 |
+---------+----------------------+------+
```

The additional table for ISO codes does not exist because a single ISO code can correspond to multiple country names, even when those different names represent the same country (e.g., South Korea may be represented as 'Korea, Rep'). The AirPollution table requires finding values for both EPI from the EnvironmentalPerformance table and Output from the ManufacturingOutput table, respectively. However, it's important to note that the country names may differ between them.

Entity-Relationship Diagram
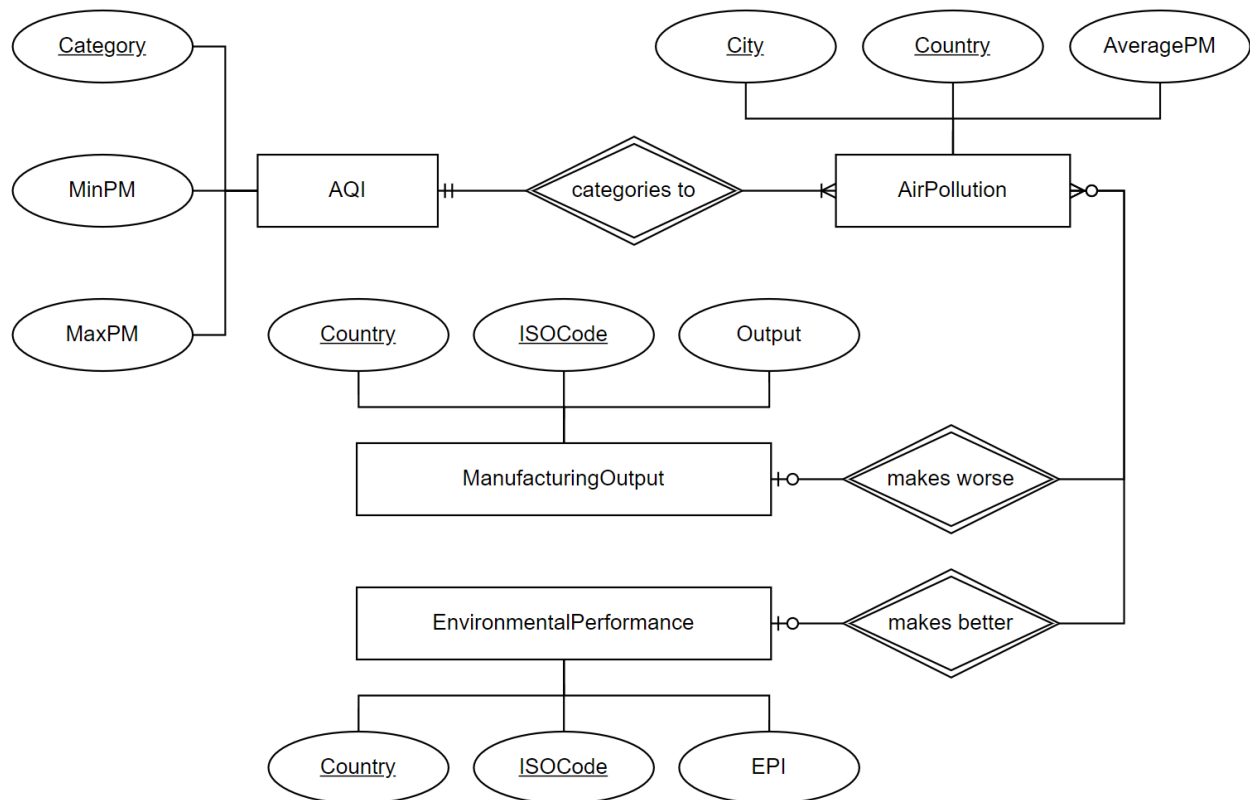
Visualisation in the web

Steps:
1. Populate Country and AveragePM columns
   : Select a single city with the highest average PM 2.5 level per country using the 'group by' command.
2. Populate the ISOCode column
   : Find the ISO codes for each country by joining ManufacturingOutput and EnvironmentalPerformance tables on the country names.
3. Populate the ManufacturingOutput (millions of $US) and EnvironmentalPerformanceIndex columns
4. Populate the category of the average PM 2.5 level

Attribute	How To Use
Country	ManufacturingOutput uses different country names compared to AirPollution and EnvironmentalPerformance. By matching these names, the ISO codes from the latter tables can be extracted.
ISOCode	
PM Category	Clarify the air quality status row by row
AveragePM	Rank countries based on the average PM 2.5, starting with the most polluted
Output	Manufacturing output worsens air quality, while environmental performance improves it. Understanding this relationship is crucial for analysing the interconnectedness between air pollution and national output.
EPI	

The SQL command is located at the 35 line in the index.js file as below:

```js
26    function templateRenderer(template, response){
27        return function(error, results, fields){
28            if(error){
29                throw error;
30            }
31            response.render(template, { data: JSON.stringify(results) });
32        }
33    }
34
35    app.get('/', function (req, res) {
36        const data = {};
37        queryCommand = "SELECT * FROM (SELECT tt1.Country, tt1.AveragePM, tt2.Output, tt2.EPI FROM \
38            (SELECT Country, MAX(AveragePM) AS AveragePM FROM AirPollution GROUP BY Country) AS tt1 \
39            INNER JOIN (SELECT t1.ISOCode, t1.Country AS CountryL, t2.Country AS CountryR, t2.EPI, t1.Output \
40                FROM ManufacturingOutput t1 INNER JOIN EnvironmentalPerformance t2 USING (ISOCode)) AS tt2 \
41                WHERE tt1.AveragePM != \"\" \
42                    AND tt2.Output != \"\" \
43                    AND tt2.EPI != \"\" \
44                    AND (tt1.Country=tt2.CountryL OR tt1.Country=tt2.CountryR)) AS resT \
45                    INNER JOIN AQI \
46                        WHERE resT.AveragePM >= AQI.MinPM AND resT.AveragePM <= AQI.MaxPM ORDER BY AveragePM DESC";
47        db.query(queryCommand, templateRenderer('index', res));
48    });
```

Here is the preview of the web below:

Examining Air Quality: A Pollution Analysis

To what extent do manufacturing output and environmental performance impact air pollution?

Top 5 Countries with the highest average PM 2.5

Country	PM Catetory	Average PM2.5	Manufacturing Output	Environmental Performance Index
Pakistan	Unhealthy	97.4	49804942059.37(1.0%)	24.6(32.2%)
China	Unhealthy	95.0	4975614263593.69(100.0%)	28.4(37.1%)
India	Unhealthy	93.3	450861876466.75(9.1%)	18.9(24.7%)
Iraq	Unhealthy	87.3	4725924137.93(0.1%)	27.8(36.3%)
Kazakhstan	Unhealthy	78.5	29105930481.46(0.6%)	40.9(53.5%)

Bottom 5 with the lowest average PM 2.5

Country	PM Catetory	Average PM2.5	Manufacturing Output	Environmental Performance Index
Estonia	Good	6.4	4862199354.90(0.1%)	61.4(80.3%)
Finland	Good	6.0	44508153739.30(0.9%)	76.5(100.0%)
Belize	Good	5.6	169617579.79(0.0%)	50.0(65.4%)
Iceland	Good	3.9	2908616979.90(0.1%)	62.8(82.1%)
Grenada	Good	3.8	39870370.37(0.0%)	47.9(62.6%)

Manufacturing Output

China was reported as the country with the worst air pollution, attributed to its highest manufacturing output.

China's EPI is 62.9% lower compared to Finland, which has the highest EPI.

EPI

Finland was identified as the country with the most proactive environmental performance, credited to its highest EPI (76.5).

Finland's manufacturing output is 99.1% less compared to China, which has the highest manufacturing output.

This analysis focuses on the impact of manufacturing output and environmental performance on air pollution. As expected, the country with the highest average PM 2.4 level also exhibited the largest manufacturing output but had a relatively low environmental performance index. Conversely, the bottom 5 countries, which have the lowest average PM 2.5 level, were identified as having proactive environmental performance, with their manufacturing output much lower than that of the top countries.

Datasets

1. Air Quality
   https://www.kaggle.com/datasets/andreinovikov/air-pollution
2. Manufacturing Output
   https://data.worldbank.org/indicator/NV.IND.MANF.CD?most_recent_value_desc=false
3. Environmental Performance Index
   https://epi.yale.edu/epi-results/2022/component/epi

References

"PM2.5" (IQAir Staff Writers), https://www.iqair.com/newsroom/pm2-5 (22 Jan, 2022)

"What is the air quality index (AQI)?" (IQAir Staff Writers),
https://www.iqair.com/newsroom/what-is-aqi (4 Aug. 2018)