

Spark Individual Assignment

Global Emissions by Country



February 18, 2020

Mohamed Amer

Table of Contents

[1 Background/ Dataset Description 2](#_Toc32785657)

[1.1 My Data 2](#_Toc32785658)

[1.2 Data Profiling 3](#_Toc32785659)

[2 Analysis Goal 3](#_Toc32785660)

[3 Analysis Deep Dive 4](#_Toc32785661)

[3.1 Which is the most emissions polluting the earth? 4](#_Toc32785662)

[3.2 Which countries generates most of the world’s CO2 emissions? 5](#_Toc32785663)

[3.3 Is there a change in emissions since 2000? 5](#_Toc32785664)

[4 Conclusion 6](#_Toc32785665)

[5 Resources 7](#_Toc32785666)

# Background/ Dataset Description

In this assignment I’m going to analyse the global emission dataset for the period between 2000 and 2012 and I’ll be trying to answer some questions about the Top countries in emission, CO2 emission over the years, and if there is a specific decrease in emission or not?

to begin with that I need to highlight here what emissions I’m considering and their categories to just clarify more what are these abbreviations for, as follows:

Emissions are calculated for the following substances: 1) Direct greenhouse gases: Carbon Dioxide (CO2), Methane (CH4), Nitrous Oxide (N2O), Hydrofluorocarbons (HFC-23, 32, 125, 134a, 143a, 152a, 227ea, 236fa, 245fa, 365mfc, 43-10-mee), Perfluorocarbons (PFCs: CF4, C2F6, C3F8, c-C4F8, C4F10, C5F12, C6F14, C7F16), Sulfur Hexafluoride (SF6), Nitrogen Trifluoride (NF3) and Sulfuryl Fluoride (SO2F2); 2) Ozone precursor gases: Carbon Monoxide (CO), Nitrogen Oxides (NOx), Non-Methane Volatile Organic Compounds (NMVOC) and Methane (CH4). 3) Acidifying gases: Ammonia (NH3), Nitrogen oxides (NOx) and Sulfur Dioxide (SO2). 4) Primary particulates: Fine Particulate Matter (PM10) - Carbonaceous speciation (BC , OC) is under progress. 5) Stratospheric Ozone Depleting Substances: Chlorofluorocarbons (CFC-11, 12, 113, 114, 115), Halons (1211, 1301, 2402), Hydrochlorofluorocarbons (HCFC-22, 124, 141b, 142b), Carbon Tetrachloride (CCl4), Methyl Bromide (CH3Br) and Methyl Chloroform (CH3CCl2). Emissions (EM) for a country C are calculated for each compound x on an annual basis (y) and sector wise (for i sectors, multiplying on the one hand the country-specific activity data (AD), quantifying the human activity for each of the i sectors, with the mix of j technologies (TECH) for each sector i, and with their abatement percentage by one of the k end-of-pipe (EOP) measures for each technology j, and on the other hand the country-specific emission factor (EF) for each sector i and technology j with relative reduction (RED) of the uncontrolled emission by installed abatement measure k. Emissions in are calculated by individual countries using country-specific information. The countries are organized in different world regions for illustration purposes. Emissions of some small countries are presented together with other countries depending on country definition and availability of activity statistics.

## My Data

My dataset almost speaks about country and regional-wise emission from the period 2000 until 2012 and they are as follows:

**COUNTRY\_ID**: ID of the country

**COUNTRY\_NAME**: Name of the country

**COUNTRY\_NOTES**: Geographical Area for the country

**COUNTRY\_REGIONID**: Region Id

**DATE**: Date the report taken

**FREQUENCY**: Report frequency

**INDICATOR\_ID**:

**INDICATOR\_NAME**: Emission (CO2,HC4,..etc)

**IPCC\_ID**: id for the sector the emission been measured in

**IPCC\_NAME**: name sector the emission been measured in

**VALUE**: emission value measured

## Data Profiling

Since I’m having almost 1.9 M records, I preferred to reduce it to around 500 K by choosing only all data from the year 2000 and in order to do that I have just made a filter to subset the data between only 2000 and 2012 and my discovery is as follows:

**Metrics:**

My only metric was the value columns and it can be described as follows:

emmission value summary:

+-------+------------------+

|summary| VALUE|

+-------+------------------+

| count| 518789|

| mean|3251.4186273294995|

| stddev| 73873.94031694348|

| min| -496000.0|

| 25%| 0.0162442|

| 50%| 0.47762|

| 75%| 11.815299832|

| max| 9918456.19197339|

+-------+------------------+

Also there were no null data found.

**Dimensions:**

My dimensions are COUNTRY\_ID, COUNTRY\_NAME, COUNTRY\_NOTES, COUNTRY\_REGIONID, DATE, FREQUENCY, INDICATOR\_ID, INDICATOR\_NAME, IPCC\_ID, IPCC\_NAME and the taken insights is as follows:

Checking amount of distinct values in columns COUNTRY\_ID,COUNTRY\_NAME,COUNTRY\_NOTES,COUNTRY\_REGIONID,DATE,FREQUENCY,INDICATOR\_ID,INDICATOR\_NAME,IPCC\_ID,IPCC\_NAME:

+----------+------------+-------------+----------------+----+---------+------------+--------------+-------+---------+

|COUNTRY\_ID|COUNTRY\_NAME|COUNTRY\_NOTES|COUNTRY\_REGIONID|DATE|FREQUENCY|INDICATOR\_ID|INDICATOR\_NAME|IPCC\_ID|IPCC\_NAME|

+----------+------------+-------------+----------------+----+---------+------------+--------------+-------+---------+

| 225| 225| 25| 221| 13| 1| 40| 40| 64| 64|

+----------+------------+-------------+----------------+----+---------+------------+--------------+-------+---------+

Checking also if there is nulls on other columns in the dataset :

+----------+------------+-------------+----------------+----+---------+------------+--------------+-------+---------+

|COUNTRY\_ID|COUNTRY\_NAME|COUNTRY\_NOTES|COUNTRY\_REGIONID|DATE|FREQUENCY|INDICATOR\_ID|INDICATOR\_NAME|IPCC\_ID|IPCC\_NAME|

+----------+------------+-------------+----------------+----+---------+------------+--------------+-------+---------+

| 0| 0| 7462| 4752| 0| 0| 0| 0| 0| 0|

+----------+------------+-------------+----------------+----+---------+------------+--------------+-------+---------+

I have almost **225** countries been used for this emission measurements during these years and I have almost **7,462** and **4,752** **null** values for COUNTRY\_NOTES and COUNTRY\_REGIONID

In terms of occurrence it is worth to mention that **Australia** is the most country that occurred in this dataset which might be an indication of the safety measure they take for this measurement they took for all sectors.

# Analysis Goal

My objective from this analysis to analyse the global emission dataset for the period between 2000 and 2012 to answer some questions about the Top countries in emission, CO2 emission over the years, and if there is a specific decrease in emission or not?

Also, in my analysis I’m trying to see which countries are causing the [global warming](https://en.wikipedia.org/wiki/Global_warming) the most as there was lots of changes in weather during the last three decades and reports are highlighting that if no one takes an actions then there will be lots of cities removed from the global map during the upcoming fifty years.

And to see exactly what is going to be happening by global warming the below image illustrates my hometown Alexandria, Egypt, founded by Alexander the Great around 330 B.C. could be lost to rising waters.

A picture containing person

Description automatically generated

# Analysis Deep Dive

In order to get into the data I will try to answer the most important questions needed to understand fully the case we are examining here and these questions are as follows:

## Which is the most emissions polluting the earth?

TOP20 **most emmissions types poluted the earth** (in Megatons):

A screenshot of text

Description automatically generated

Now, it can be seen that CO2 (KN.A2, KN.A3) are both the most type that is affecting the whole earth, so let's see the most two countries affecting the earth as well but only for CO2, both together are formulating 97.4% of the overall emissions on earth for the period 2000 till 2012

## Which countries generates most of the world’s CO2 emissions?

TOP10 **'COUNTRY-Wise' Most to lowest emissions** (in Megatons):

A screenshot of text

Description automatically generated

From the above table, It Can be seen that China, US, India, Russia and Brazil are the most countries that produces CO2 on earth with almost 44% of the over all emission in the period from 2000 till 2012

## Is there a change in emissions since 2000?

TOP20 **Country-wise** over year from 2000 till 2012(in Megatons):

A close up of text on a white background

Description automatically generated

We can see from the above that the united states of America were at 1st place in emission till 2004 as China took the lead till now in polluting the whole world,

there is also something happened in 2009 as there were a big drop by almost 50% in all countries and I think this is related to the economic crisis happened in 2008.

TOP5 **Indicators** over year from 2000 till 2012(in Megatons):

A screenshot of a cell phone

Description automatically generated

It can be seen from the above table that CO2 in both types are decreased dramatically to half of the percentage recorded in 2000, however, there is a slight increase in CH4 which is more dangerous even with small doses to our earth, and to really imagine how CH4 is problematic to earth just imagine that its impact is 34 times greater than CO2 over a 100-year period, according to the latest IPCC Assessment Report.

TOP20 **emission sectors** over year from 2000 till 2012(in Megatons):

A close up of text on a white background

Description automatically generated

Here, I think some countries are not segregating the sectors and that’s why we have the Total indicator row indicating the total emission in this year in this country which would have been better if they tried to have it segregated for better analysis sector wise.

# Conclusion

In conclusion, it can be seen that CO2 (KN.A2, KN.A3) are both the most type that is affecting the whole earth, both together are formulating 97.4% of the overall emissions on earth for the period 2000 till 2012 but it can be seen from the above table that CO2 in both types are decreased dramatically to half of the percentage recorded in 2000, however, there is a slight increase in CH4 which is more dangerous even with small doses to our earth, and to really imagine how CH4 is problematic to earth just imagine that its impact is 34 times greater than CO2 over a 100-year period, according to the latest IPCC Assessment Report.

Country wise, China, US, India, Russia and Brazil are the most countries that produces CO2 on earth with almost 44% of the overall emission in the period from 2000 till 2012 which will need to work on reducing more there emissions by searching for another source of energy as they are affecting not only their area but also all earth.

Lastly, united states of America were at 1st place in emission till 2004 as China took the lead till now in polluting the whole world, there is also something happened in 2009 as there were a big drop by almost 50% in all countries and I think this is related to the economic crisis happened in 2008.

Note: Methane is a greenhouse gas as is carbon dioxide. Human activity has increased the amount of methane in the atmosphere, contributing to climate change. Methane is particularly problematic as its impact is 34 times greater than CO2 over a 100-year period, according to the latest IPCC Assessment Report. A significant source of human-made methane emissions is fossil fuel production. For example, methane is a key by-product of the rapidly rising global extraction and processing of natural gas. Other top sources of methane come from the digestive process of livestock and from landfills, which emit it as waste decomposes.

# Resources

My Dataset In this [link](https://datasource.kapsarc.org/explore/dataset/global-emissions-by-country-2013/download/?format=csv&timezone=Europe/Berlin&lang=en&use_labels_for_header=true&csv_separator=%3B)

Why Methane Matters:

<https://unfccc.int/news/new-methane-signs-underline-urgency-to-reverse-emissions>

The Secret History of Lead

<https://www.thenation.com/article/archive/secret-history-lead/>

Global CO2 emissions in 2019

<https://www.iea.org/articles/global-co2-emissions-in-2019>

Rising Seas Will Erase More Cities by 2050, New Research Shows

<https://www.nytimes.com/interactive/2019/10/29/climate/coastal-cities-underwater.html>