



Big Data Project

Co2 Emission Change between 2004 and 2014

Group Members:

Jorge Alberto Gavilan Benavides (219203275)

Jennifer Gerbl (219201330)

Mushahid Husain Khan (219203201)

Mayur Rameshbhai Waghela (219203180)

TOPIC & TASKS

Dataset used: <https://data.worldbank.org/indicator/EN.ATM.CO2E.PC>

The aim is to analyze the change in CO2 emissions from 2004 to 2014. Perform the following tasks:

1. Import the data to Spark.
2. Preprocess the data: handle null values.
3. Compute the change.
4. Apply k-means on the change. Choose a reasonable amount of clusters.
5. Visualize the data, for instance as in Fig. 1. You can use GeoPandas for this.

APPROACH & EXPLANATION

See the video for explanation of our approach and our code:

<https://cloud.uni-koblenz-landau.de/s/nCnd5qzRdqLaaLB>

RESULTS

Visualizations produced by our code: (for better visibility check figures-folder)

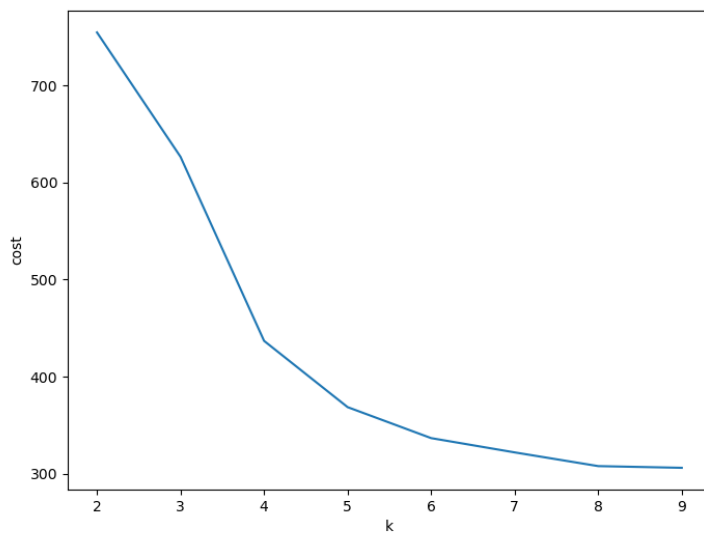


Figure 1: Choosing best k by comparing costs

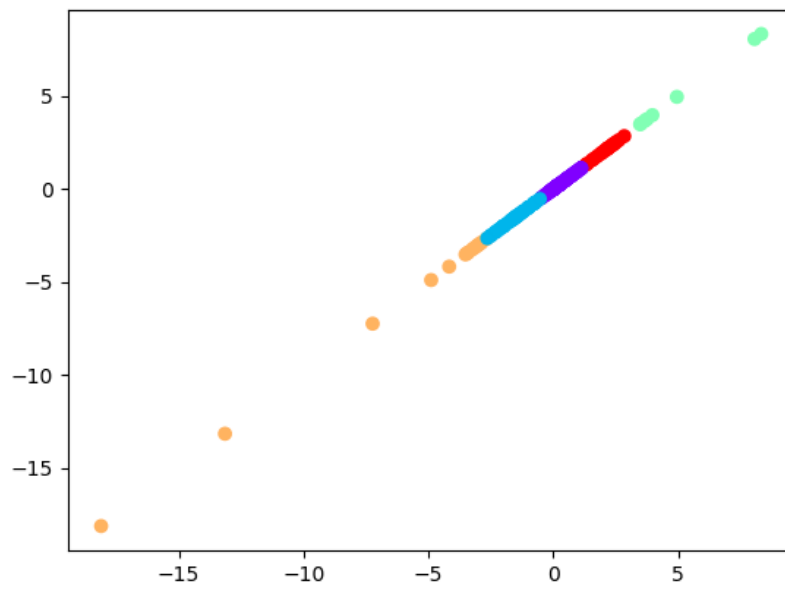


Figure 2: Cluster results

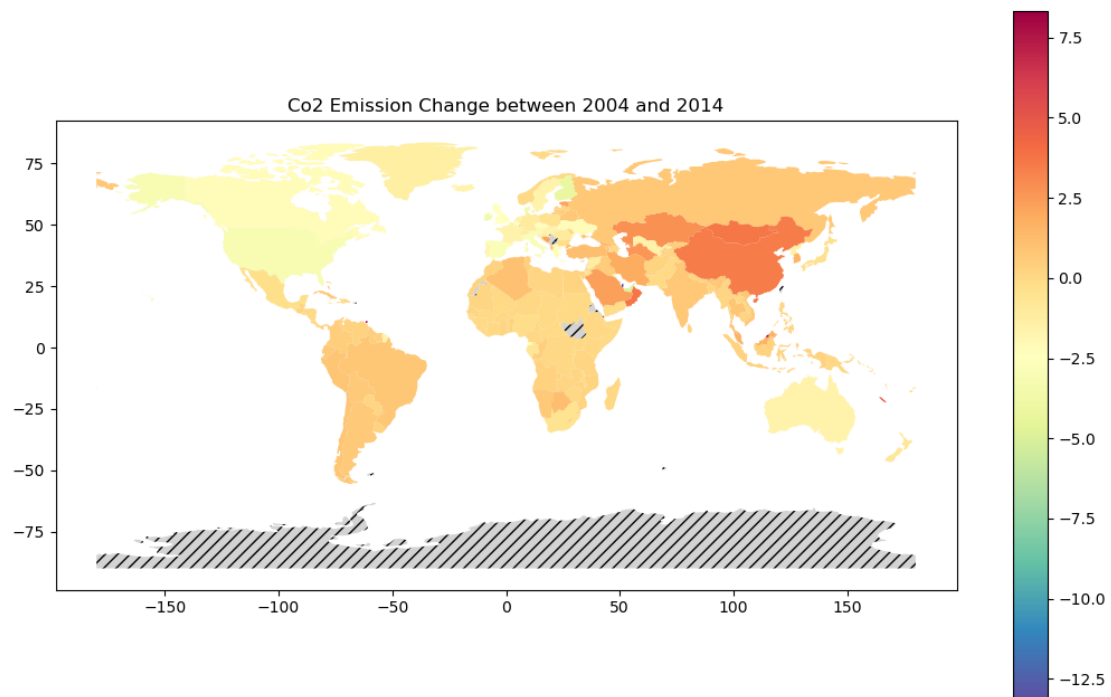


Figure 3: Visualization of Change Value

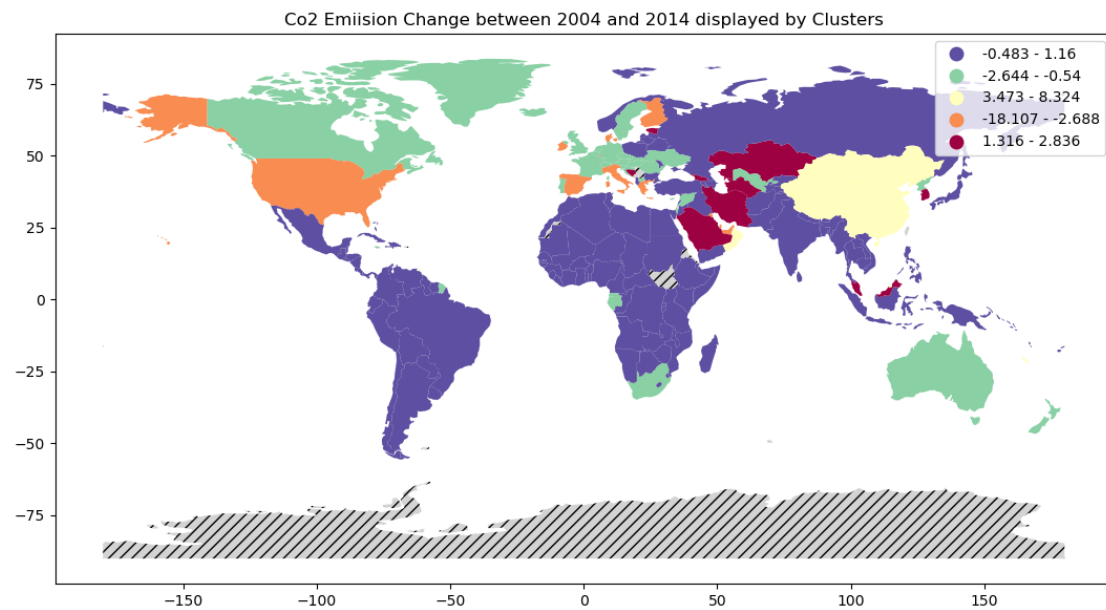


Figure 4: Visualization of our clusters

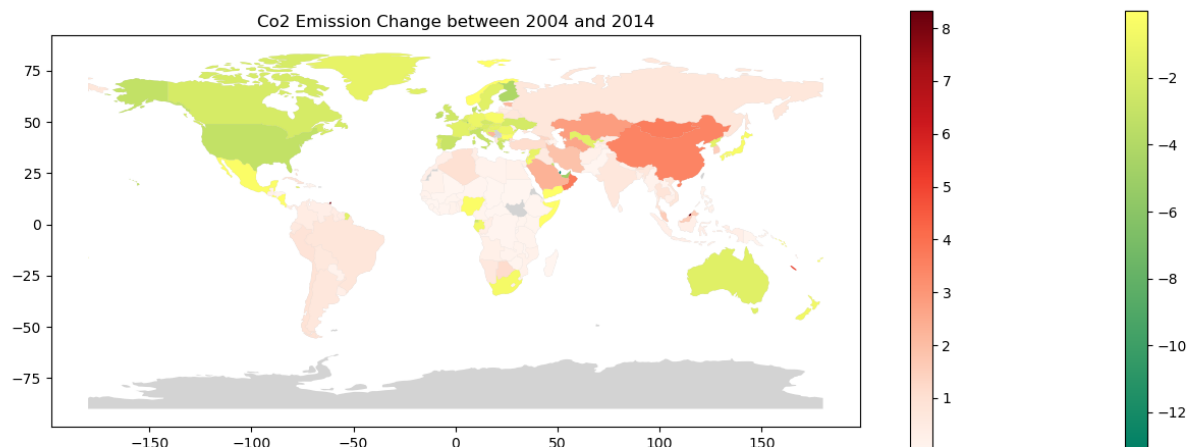


Figure 5: Visualization of Change (Improvement)

Questions asked:

Which top 3 countries have the biggest co2 emission in year 2004 and year 2014?

Which country (top 3) reduced co2 emissions the most?

Which country (top 3) increased his co2 emissions the most?

How many countries increased /reduces their emissions? Also which?

What's the sum of change?

For personal interest data of: Data of Germany, USA, France, China, Sweden?

Answers:

Which top 3 countries have the biggest co2 emission in year 2004 and year 2014?

In 2004:

Country Name	Country Code	2004	2014	change	isReduced	features	cluster
Qatar	QAT	57.0055738888727	43.8573079516167	-13.148265937256006	true	[-13.148265937256...]	3
Kuwait	KWT	28.8728348023987	25.8492292753449	-3.0236055270537996	true	[-3.0236055270537...]	3
United Arab Emirates	ARE	27.8330290495186	22.9396063130991	-4.893422736419499	true	[-4.8934227364194...]	3
Congo, Dem. Rep.	COD	0.0225953327072288	0.0633308890302249	0.040735556322996094	false	[0.04073555632299...]	0
Burundi	BDI	0.0277659175738496	0.0446999922899522	0.0169340747161026	false	[0.01693407471610...]	0
Afghanistan	AFG	0.0384100431744103	0.293946407148718	0.2555363639743077	false	[0.25553636397430...]	0

In 2014:

Country Name	Country Code	2004	2014	change	isReduced	features	cluster
Qatar	QAT	57.0055738888727	43.8573079516167	-13.148265937256006	true	[-13.148265937256...]	3
Trinidad and Tobago	TTO	25.9138384479619	33.966414453933	8.052576005971098	false	[8.05257600597109...]	2
Kuwait	KWT	28.8728348023987	25.8492292753449	-3.0236055270537996	true	[-3.0236055270537...]	3
Burundi	BDI	0.0277659175738496	0.0446999922899522	0.0169340747161026	false	[0.01693407471610...]	0
Somalia	SOM	0.0586416337692985	0.045347230872012	-0.01329440289728...	true	[-0.0132944028972...]	0
Chad	TCD	0.0387991823533116	0.0534072418467253	0.014608059493413705	false	[0.01460805949341...]	0

Which country (top 3) reduced co2 emissions the most?

Country Name	Country Code	2004	2014	change	isReduced	features	cluster
Aruba	ABW	26.5172934158421	8.41006417792511	-18.10722923791699	true	[-18.107229237916...]	3
Qatar	QAT	57.0055738888727	43.8573079516167	-13.148265937256006	true	[-13.148265937256...]	3
Luxembourg	LUX	24.5990264028204	17.3621213728095	-7.236905030010899	true	[-7.2369050300108...]	3

Which country (top 3) increased his co2 emissions the most?

Country Name	Country Code	2004	2014	change	isReduced	features	cluster
Brunei Darussalam	BRN	13.9055707183258	22.229177902672	8.3236071843462	false	[8.3236071843462,...]	2
Trinidad and Tobago	TTO	25.9138384479619	33.966414453933	8.052576005971098	false	[8.05257600597109...]	2
New Caledonia	NCL	11.0611147540984	16.005931729155	4.944816975056602	false	[4.94481697505660...]	2

Reduced:

87

Country Name	Country Code	2004	2014	change	isReduced	features	cluster
Aruba	ABW	26.5172934158421	8.41006417792511	-18.10722923791699	true	[-18.107229237916...]	3
Qatar	QAT	57.0055738888727	43.8573079516167	-13.148265937256006	true	[-13.148265937256...]	3
Luxembourg	LUX	24.5990264028204	17.3621213728095	-7.236905030010899	true	[-7.2369050300108...]	3
United Arab Emirates	ARE	27.8330290495186	22.9396063130991	-4.893422736419499	true	[-4.8934227364194...]	3
Finland	FIN	12.8326749770283	8.66072124349448	-4.171953733533821	true	[-4.1719537335338...]	3
Seychelles	SYC	8.93685359199758	5.41867796276229	-3.518175629235289	true	[-3.5181756292352...]	3
Ireland	IRL	10.7516366268314	7.31393980771791	-3.4376968191134907	true	[-3.4376968191134...]	3
Denmark	DNK	9.36270101912787	5.9357124821143	-3.42698853701357	true	[-3.4269885370135...]	3
Faroe Islands	FRO	15.7157142857143	12.462906588824	-3.252807696890301	true	[-3.2528076968903...]	3
United States	USA	19.6583711815214	16.507265616325	-3.1511055651963993	true	[-3.1511055651963...]	3
North America	NAC	19.4232416179772	16.3708023227419	-3.0524392952353026	true	[-3.0524392952353...]	3
Kuwait	KWT	28.8728348023987	25.8492292753449	-3.0236055270537996	true	[-3.0236055270537...]	3
Italy	ITA	8.21648720132938	5.27086678640296	-2.9456204149264202	true	[-2.9456204149264...]	3
Spain	ESP	7.90805068135971	5.03382448723757	-2.87422619412214	true	[-2.8742261941221...]	3
Greece	GRC	8.86796719458015	6.18033726778447	-2.6876299267956796	true	[-2.6876299267956...]	3
Nauru	NRU	6.6659260755403	4.02185100818358	-2.64407506735672	true	[-2.6440750673567...]	1
Equatorial Guinea	GNQ	7.2782596809536	4.76396715246517	-2.51429252848843	true	[-2.5142925284884...]	1
United Kingdom	GBR	8.98914012749737	6.49853294692396	-2.4906071805734102	true	[-2.4906071805734...]	1
Belgium	BEL	10.6612902219787	8.32815989783976	-2.3331303241389403	true	[-2.3331303241389...]	1
Czech Republic	CZE	11.451844499726	9.16597837581982	-2.28586612390618	true	[-2.2858661239061...]	1

only showing top 20 rows


```

Increased:
157
+-----+-----+-----+-----+-----+-----+-----+
| Country Name|Country Code|      2004|      2014|      change|isReduced|      features|cluster|
+-----+-----+-----+-----+-----+-----+-----+
| Brunei Darussalam|BRN|13.9055707183258|22.229177902672|8.3236071843462|false|[8.3236071843462,...]|2|
| Trinidad and Tobago|TTO|25.9138384479619|33.966414453933|8.052576005971098|false|[8.05257600597109...]|2|
| New Caledonia|NCL|11.0611147540984|16.005931729155|4.944816975056602|false|[4.94481697505660...]|2|
| Gibraltar|GIB|11.6932397959184|15.6574647887324|3.9642249928139996|false|[3.96422499281399...]|2|
| Oman|OMN|11.444019403668|15.1887951113164|3.744775707648401|false|[3.74477570764840...]|2|
| Mongolia|MNG|3.42552268454741|7.08802567796829|3.6625029934208806|false|[3.66250299342088...]|2|
| China|CHN|4.03799065100399|7.54390764144927|3.5059169904452796|false|[3.50591699044527...]|2|
| Singapore|SGP|6.83382557364837|10.306331909983|3.4725063363346296|false|[3.47250633633462...]|2|
| Kazakhstan|KAZ|11.5268924867373|14.3631697418223|2.836277255085001|false|[2.83627725508500...]|4|
| Turkmenistan|TKM|9.89530001800824|12.5170961201011|2.6217961020928584|false|[2.62179610209285...]|4|
| East Asia & Pacif...|TEA|3.32603214720709|5.82837164447171|2.5023394972646202|false|[2.50233949726462...]|4|
| East Asia & Pacif...|EAP|3.32240831031778|5.77612844555788|2.4537201352401|false|[2.4537201352401,...]|4|
| Late-demographic ...|LTE|4.35184185774159|6.75494397274433|2.4031021150027403|false|[2.40310211500274...]|4|
| Bahrain|BHR|21.1002610129133|23.455406320753|2.355145307839699|false|[2.35514530783969...]|4|
| Saudi Arabia|SAU|17.1114752279913|19.4406664826471|2.3291912546558002|false|[2.32919125465580...]|4|
| Bosnia and Herzeg...|BIH|4.14708142035187|6.38493881859933|2.23785739824746|false|[2.23785739824746...]|4|
| East Asia & Pacific|EAS|4.12015502106544|6.29425627174206|2.17410125067662|false|[2.17410125067662...]|4|
| Estonia|EST|12.6839903122821|14.848819173174|2.1648288608919|false|[2.1648288608919,...]|4|
| Other small states|OSS|5.46479898725205|7.62789783009735|2.1630988428453|false|[2.1630988428453,...]|4|
| Upper middle income|UMC|4.39617497084748|6.52757182101923|2.1313968501717504|false|[2.13139685017175...]|4|
+-----+-----+-----+-----+-----+-----+-----+

only showing top 20 rows

Sum of change:
Reduced:
+-----+
| sum(change)|
+-----+
|-148.53868230815294|
+-----+

Increased:
+-----+
| sum(change)|
+-----+
|126.47130645427599|
+-----+

Total Change:
+-----+
| sum(change)|
+-----+
|-22.067375853877|
+-----+

```

```

Out of interest:
+-----+-----+-----+-----+-----+-----+-----+
| Country Name|Country Code|      2004|      2014|      change|isReduced|      features|cluster|
+-----+-----+-----+-----+-----+-----+-----+
| Sweden|SWE|6.06224140440501|4.47818217821374|-1.58405922619127|true|[-1.5840592261912...]|1|
| France|FRA|6.12007856802886|4.57345972943356|-1.5466188385952995|true|[-1.5466188385952...]|1|
| China|CHN|4.03799065100399|7.54390764144927|3.5059169904452796|false|[3.50591699044527...]|2|
| Germany|DEU|9.89868236878404|8.88937039483839|-1.0093119739456498|true|[-1.0093119739456...]|1|
| United States|USA|19.6583711815214|16.507265616325|-3.1511055651963993|true|[-3.1511055651963...]|3|
+-----+-----+-----+-----+-----+-----+-----+

```

HELPFUL RESOURCES

KMeans: <https://spark.apache.org/docs/2.0.0-preview/ml-clustering.html>

KMeans: <https://rsandstroem.github.io/sparkkmeans.html>

Installing GeoPandas on Windows: <https://geoffboeing.com/2014/09/using-geopandas-windows/>

ColorMaps: https://matplotlib.org/examples/color/colormaps_reference.html