Predicting CO_2 Emissions

Jack Hlavka and Gabby Walrath

STATS/DATASCI 112

Question

How well can we predict a country's CO_2 emissions?

What about if we don’t use any factors relating to what type of energy it uses?

Data Collection

Our World in Data

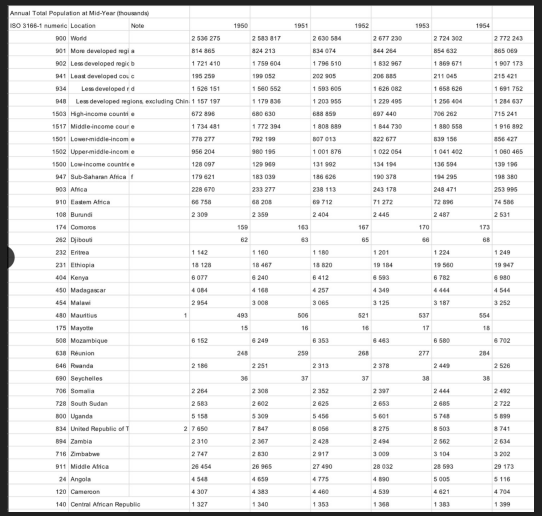
Population Division World Urbanization Prospects 2018

THE WORLD BANK
IBRD • IDA

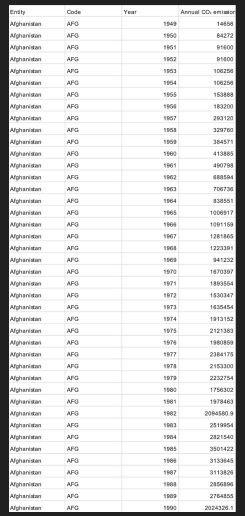
FORTUNE

kaggle

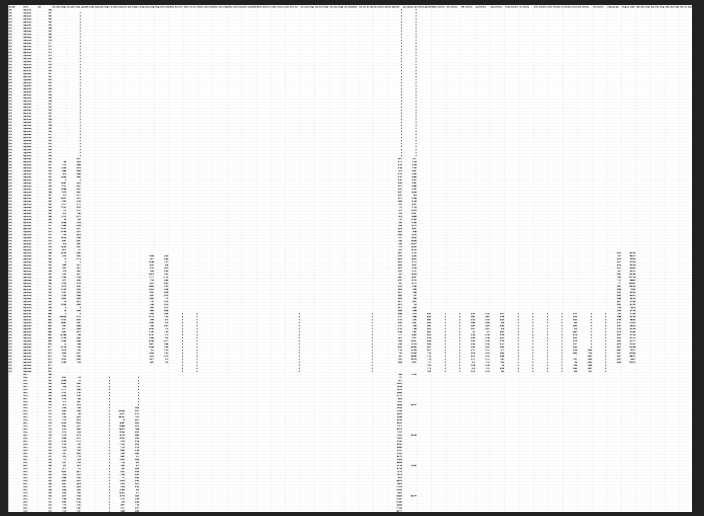
1	Walmart
2	Amazon
3	State Grid
4	China National Petroleum
5	Sinopec Group
6	Saudi Aramco
7	Apple
8	Volkswagen
9	China State Construction Engineering
10	CVS Health



Annual Total Population at Mid-Year (thousands)									
ISO 3166-1 numeric	Location	Note	1950	1951	1952	1953	1954		
900	World		2 536 275	2 583 817	2 630 584	2 677 230	2 724 302	2 772 243	
901	More developed regi	a	814 865	824 213	834 074	844 264	854 632	865 069	
902	Less developed regic	b	1 721 410	1 759 604	1 796 510	1 832 967	1 869 671	1 907 173	
941	Least developed cou	c	195 259	199 052	202 905	206 885	211 045	215 421	
934	Less developed r	d	1 526 151	1 560 552	1 593 605	1 626 082	1 658 626	1 691 752	
948	Less developed regions, excluding Chin		1 157 197	1 179 836	1 203 955	1 229 495	1 256 404	1 284 637	
1503	High-income countri	e	672 896	680 630	688 859	697 440	706 262	715 241	
1517	Middle-income cour	e	1 734 481	1 772 394	1 808 889	1 844 730	1 880 558	1 916 892	
1501	Lower-middle-incom	e	778 277	792 199	807 013	822 677	839 156	856 427	
1502	Upper-middle-incom	e	956 204	980 196	1 001 876	1 022 054	1 041 402	1 060 465	
1500	Low-income countrie	e	128 097	129 969	131 992	134 194	136 594	139 196	
947	Sub-Saharan Africa	f	179 621	183 039	186 626	190 378	194 295	198 380	
903	Africa		228 670	233 277	238 113	243 178	248 471	253 995	
910	Eastern Africa		66 758	68 208	69 712	71 272	72 896	74 586	
108	Burundi		2 309	2 359	2 404	2 445	2 487	2 531	
174	Comoros		159	163	167	170	173		
262	Djibouti		62	63	65	66	68		
232	Eritrea		1 142	1 160	1 180	1 201	1 224	1 249	
231	Ethiopia		18 128	18 467	18 820	19 184	19 560	19 947	
404	Kenya		6 077	6 240	6 412	6 593	6 782	6 980	
450	Madagascar		4 084	4 168	4 257	4 349	4 444	4 544	
454	Malawi		2 954	3 008	3 065	3 125	3 187	3 252	
480	Mauritius	1	493	506	521	537	554		
175	Mayotte		15	16	16	17	18		
508	Mozambique		6 152	6 249	6 353	6 463	6 580	6 702	
638	Réunion		248	259	268	277	284		
646	Rwanda		2 186	2 251	2 313	2 378	2 449	2 526	
690	Seychelles		36	37	37	38	38		
706	Somalia		2 264	2 308	2 352	2 397	2 444	2 492	
728	South Sudan		2 583	2 602	2 625	2 653	2 685	2 722	
800	Uganda		5 158	5 309	5 456	5 601	5 748	5 899	
834	United Republic of T	2	7 650	7 847	8 056	8 275	8 503	8 741	
894	Zambia		2 310	2 367	2 428	2 494	2 562	2 634	
716	Zimbabwe		2 747	2 830	2 917	3 009	3 104	3 202	
911	Middle Africa		26 454	26 965	27 490	28 032	28 593	29 173	
24	Angola		4 548	4 659	4 775	4 890	5 005	5 116	
120	Cameroon		4 307	4 383	4 460	4 539	4 621	4 704	
140	Central African Republic		1 327	1 340	1 353	1 368	1 383	1 399	



Entity	Code	Year	Annual CO_2 emissior
Afghanistan	AFG	1949	14656
Afghanistan	AFG	1950	84272
Afghanistan	AFG	1951	91600
Afghanistan	AFG	1952	91600
Afghanistan	AFG	1953	106256
Afghanistan	AFG	1954	106256
Afghanistan	AFG	1955	163868
Afghanistan	AFG	1956	183200
Afghanistan	AFG	1957	293120
Afghanistan	AFG	1958	329760
Afghanistan	AFG	1959	384571
Afghanistan	AFG	1960	413885
Afghanistan	AFG	1961	490784
Afghanistan	AFG	1962	688584
Afghanistan	AFG	1963	706736
Afghanistan	AFG	1964	838551
Afghanistan	AFG	1965	1006917
Afghanistan	AFG	1966	1091159
Afghanistan	AFG	1967	1281865
Afghanistan	AFG	1968	1223381
Afghanistan	AFG	1969	941232
Afghanistan	AFG	1970	1670397
Afghanistan	AFG	1971	1893554
Afghanistan	AFG	1972	1530347
Afghanistan	AFG	1973	1635454
Afghanistan	AFG	1974	1913152
Afghanistan	AFG	1975	2121383
Afghanistan	AFG	1976	1980859
Afghanistan	AFG	1977	2384175
Afghanistan	AFG	1978	2153300
Afghanistan	AFG	1979	2232754
Afghanistan	AFG	1980	1756302
Afghanistan	AFG	1981	1978463
Afghanistan	AFG	1982	2094580.9
Afghanistan	AFG	1983	2519854
Afghanistan	AFG	1984	2821540
Afghanistan	AFG	1985	3501422
Afghanistan	AFG	1986	3133645
Afghanistan	AFG	1987	3113826
Afghanistan	AFG	1988	2856896
Afghanistan	AFG	1989	2764855
Afghanistan	AFG	1990	2024326.1

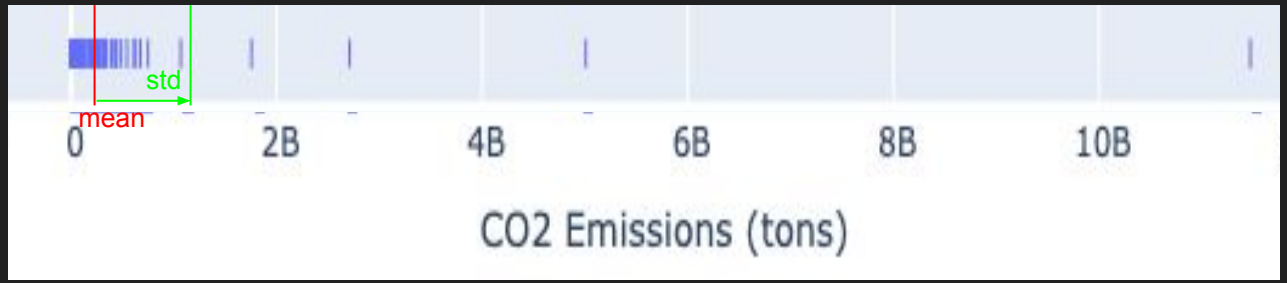


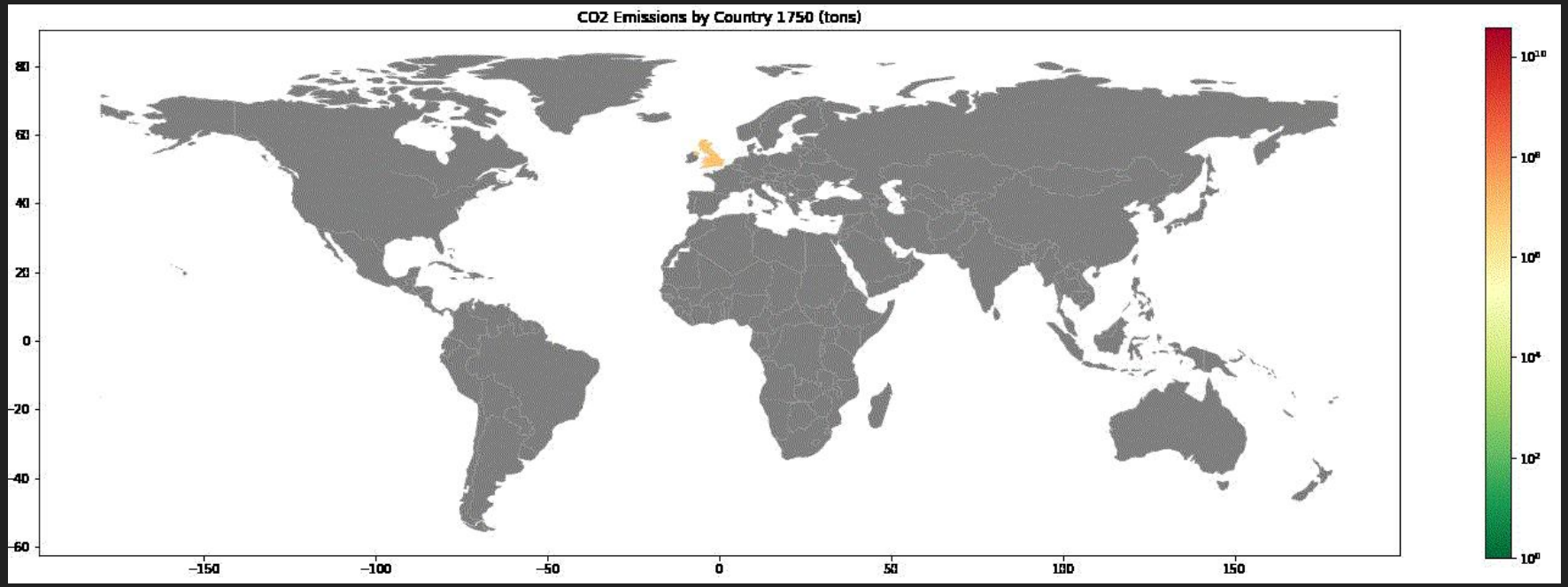
Data Analysis

	Country	Code	Year	Population	Urban Percentage	GDP Per Capita (USI	Electrification Perce	FDI (USD)	Biofuel Share (%)	Coal Share (%)	Fossil Fuel Share (%	Gas Share (%)	Low Carbon Share (%	Oil Share (%)
0	Australia	AUS	1990	17041000	85.4	18249	100	8457776859	0	42.032	95.922	15.869	4.078	38.021
1	Australia	AUS	1991	17272000	85.4	18860	100	2612066526	0	43.151	95.669	15.31	4.331	37.207
2	Australia	AUS	1992	17486000	85.3	18624	100	4941906671	0	42.972	95.648	15.68	4.352	36.996
3	Australia	AUS	1993	17687000	85.2	17700	100	5312435141	0	42.05	95.654	15.806	4.346	37.798
4	Australia	AUS	1994	17883000	85	18129	100	4458484243	0	41.263	95.871	16.881	4.129	37.727
5	Australia	AUS	1995	18077000	84.9	20446	100	13268875155	0	41.562	96.069	16.603	3.931	37.904
6	Australia	AUS	1996	18272000	84.8	22020	100	4563952446	0	42.196	96.058	16.313	3.942	37.549
7	Australia	AUS	1997	18468000	84.6	23645	100	8088068982	0	43.239	96.146	15.623	3.854	37.284
8	Australia	AUS	1998	18665000	84.5	21478	100	7597610928	0	44.139	96.246	15.718	3.754	36.389
9	Australia	AUS	1999	18864000	84.4	20698	100	2210917991	0	44.911	96.283	15.425	3.717	35.948
10	Australia	AUS	2000	19066000	84.2	21853	100	14892978180	0	45.039	96.291	15.649	3.709	35.604
11	Australia	AUS	2001	19269000	84.1	19681	100	10717133150	0	44.177	96.406	16.724	3.594	35.505
12	Australia	AUS	2002	19475000	84.2	20291	100	14656321800	0	44.511	96.341	16.959	3.659	34.87
13	Australia	AUS	2003	19697000	84.3	23705	100	8985246029	0	45.106	96.275	16.839	3.725	34.33
14	Australia	AUS	2004	19948000	84.5	30819	100	42907672820	0	45.491	96.284	16.623	3.716	34.169
15	Australia	AUS	2006	20574000	84.7	36570	100	30551100656	0.063	43.976	96.146	17.407	3.854	34.763
16	Australia	AUS	2007	20947000	84.8	41023	100	44440090037	0.239	43.046	96.19	19.226	3.81	33.918
17	Australia	AUS	2008	21342000	84.9	49679	100	45160024270	0.564	44.14	96.148	18.598	3.852	33.41
18	Australia	AUS	2009	21739000	85.1	42810	100	28683266147	0.835	43.041	95.689	19.152	4.311	33.496
19	Australia	AUS	2010	22120000	85.2	52134	100	35210733743	1.255	39.77	94.963	22.123	5.037	33.071
20	Australia	AUS	2011	22480000	85.3	62596	100	65554890649	1.278	37.415	93.707	22.299	6.293	33.992
21	Australia	AUS	2012	22822000	85.4	68044	100	57550426822	1.152	35.471	93.859	22.643	6.141	35.745
22	Australia	AUS	2013	23151000	85.5	68158	100	54465480054	1.084	33.414	93.14	23.606	6.86	36.119
23	Australia	AUS	2014	23475000	85.6	62513	100	63202728079	1.025	32.667	93.756	25.103	6.244	35.986
24	Australia	AUS	2015	23800000	85.7	56710	100	46892808567	0.917	33.326	93.496	25.97	6.504	34.2
25	Australia	AUS	2016	24126000	85.8	49875	100	42969505135	0.776	33.045	92.793	25.535	7.207	34.213
26	Australia	AUS	2017	24451000	85.9	53936	100	48198614234	0.878	31.977	93.115	25.289	6.885	35.849
27	Australia	AUS	2018	24772000	86	57207	100	60685908478	0.952	30.764	91.647	24.873	8.353	36.01
28	Australia	AUS	2019	25089000	86.1	54541	100	38955497515	0.87	27.819	91.433	30.184	8.567	33.43
42	Iceland	ISL	1992	260000	91.1	27124	100	8648079	0	2.904	40.448	0	59.552	37.545
43	Iceland	ISL	1994	265000	91.5	24018	100	2394198	0	2.821	40.389	0	59.611	37.568
44	Iceland	ISL	1996	270000	91.8	27614	100	83381484	0	3.006	41.624	0	58.376	38.619
45	Iceland	ISL	1997	273000	92	27919	100	148012830	0	2.976	39.155	0	60.845	36.179
46	Iceland	ISL	1998	275000	92.1	31030	100	154082591	0	2.47	36.741	0	63.259	34.271
47	Iceland	ISL	1999	278000	92.3	32381	100	66639312	0	2.637	33.872	0	66.128	31.234
48	Iceland	ISL	2000	280000	92.4	32096	100	155160734	0	3.435	33.419	0	66.581	29.984
49	Iceland	ISL	2001	283000	92.5	28897	100	163445772	0	3.688	31.539	0	68.461	27.851
50	Iceland	ISL	2002	285000	92.7	32409	100	91905493	0	3.537	31.43	0	68.57	27.893
51	Iceland	ISL	2003	288000	92.8	39476	100	335355205	0	3.456	31.51	0	68.49	28.054
52	Iceland	ISL	2004	291000	92.9	47334	100	755317000	0	3.548	32.307	0	67.693	28.759
53	Iceland	ISL	2005	295000	93	56794	100	3093680067	0	3.373	32.543	0	67.457	29.17
54	Iceland	ISL	2006	300000	93.2	57492	100	3876554983	0	3.146	30.161	0	69.839	27.015
55	Iceland	ISL	2007	305000	93.3	69495	100	6872169567	0	2.946	26.672	0	73.328	23.727
56	Iceland	ISL	2008	311000	93.4	56943	100	1203686964	0	2.054	19.666	0	80.334	17.612
57	Iceland	ISL	2009	316000	93.5	41301	100	63588140	0	1.903	18.254	0	81.746	16.351
58	Iceland	ISL	2010	320000	93.6	43237	100	256982201	0	2.006	17.666	0	82.334	15.66
59	Iceland	ISL	2011	323000	93.6	47714	100	1106958185	0	2.005	17.452	0	82.545	15.447
60	Iceland	ISL	2012	326000	93.6	45995	100	1024164328	0	2.031	17.162	0	82.835	15.13
61	Iceland	ISL	2013	327000	93.6	49804	100	472607545	0	2.265	17.588	0	82.367	15.324
62	Iceland	ISL	2014	328000	93.7	54576	100	766775539	0	1.858	17.905	0	82.011	16.047
63	Iceland	ISL	2015	330000	93.7	52951	100	1145767444	0	1.969	18.744	0	81.019	16.775
127	Singapore	SGP	1990	3013000	100	11861	100	5574738855	0	0.091	99.914	0	0.086	99.822
128	Singapore	SGP	1991	3097000	100	14502	100	4887094440	0	0.053	99.755	0	0.245	99.702
129	Singapore	SGP	1992	3189000	100	16135	100	2204342221	0	0.072	99.772	3.501	0.228	96.199
130	Singapore	SGP	1993	3288000	100	18290	100	4686312017	0	0.073	99.791	4.37	0.209	95.349
131	Singapore	SGP	1994	3385000	100	21553	100	8550165226	0	0.076	99.818	3.821	0.182	95.92
132	Singapore	SGP	1995	3479000	100	24914	100	11942852103	0	0.034	99.825	3.665	0.175	96.126
133	Singapore	SGP	1996	3566000	100	26233	100	11432363956	0	0.001	99.827	3.632	0.173	96.194

CO_2 Emissions

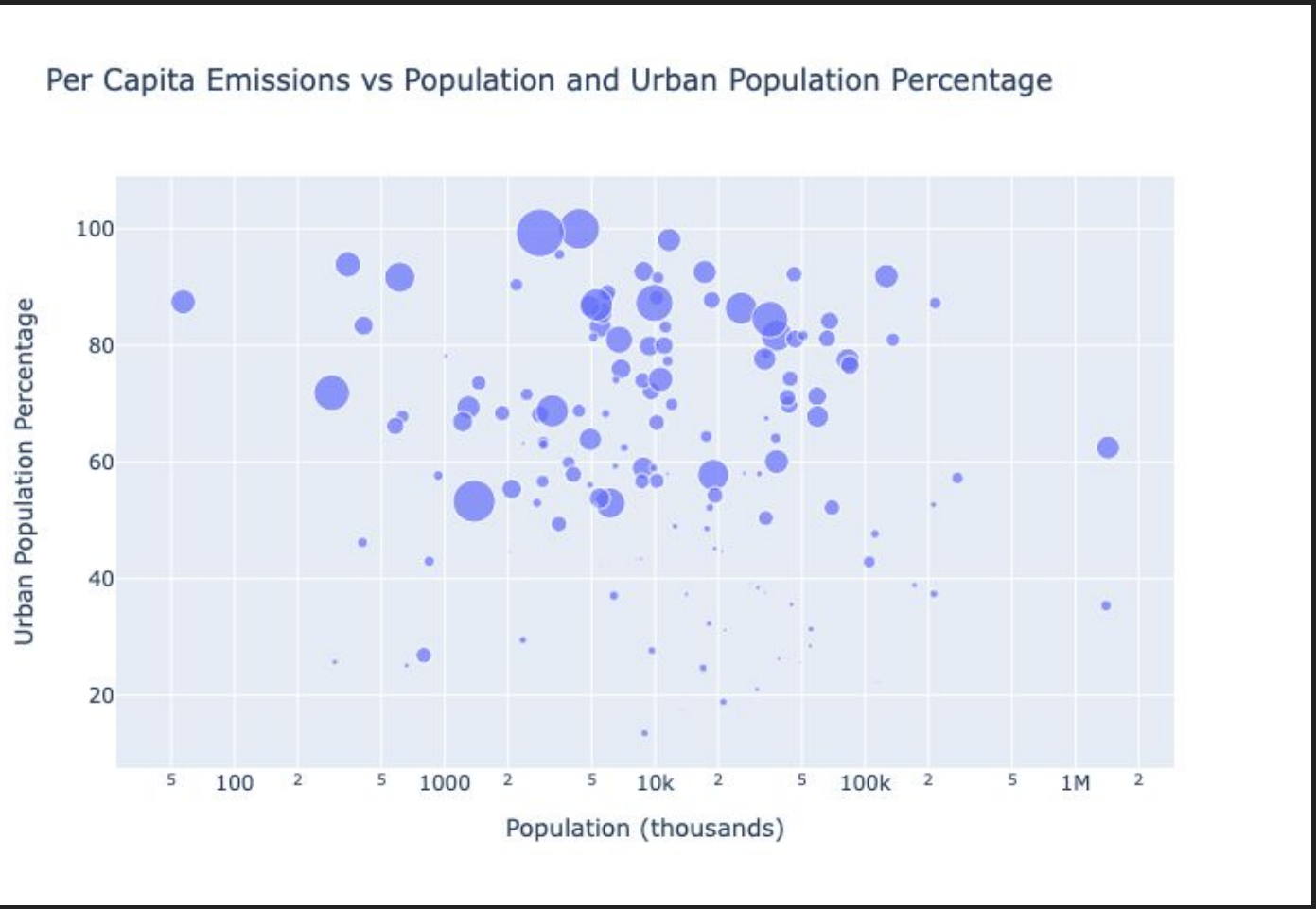
count	1.710000e+02
mean	2.102999e+08
std	9.869335e+08
min	0.000000e+00
25%	5.548892e+06
50%	2.032710e+07
75%	8.857515e+07
max	1.147237e+10



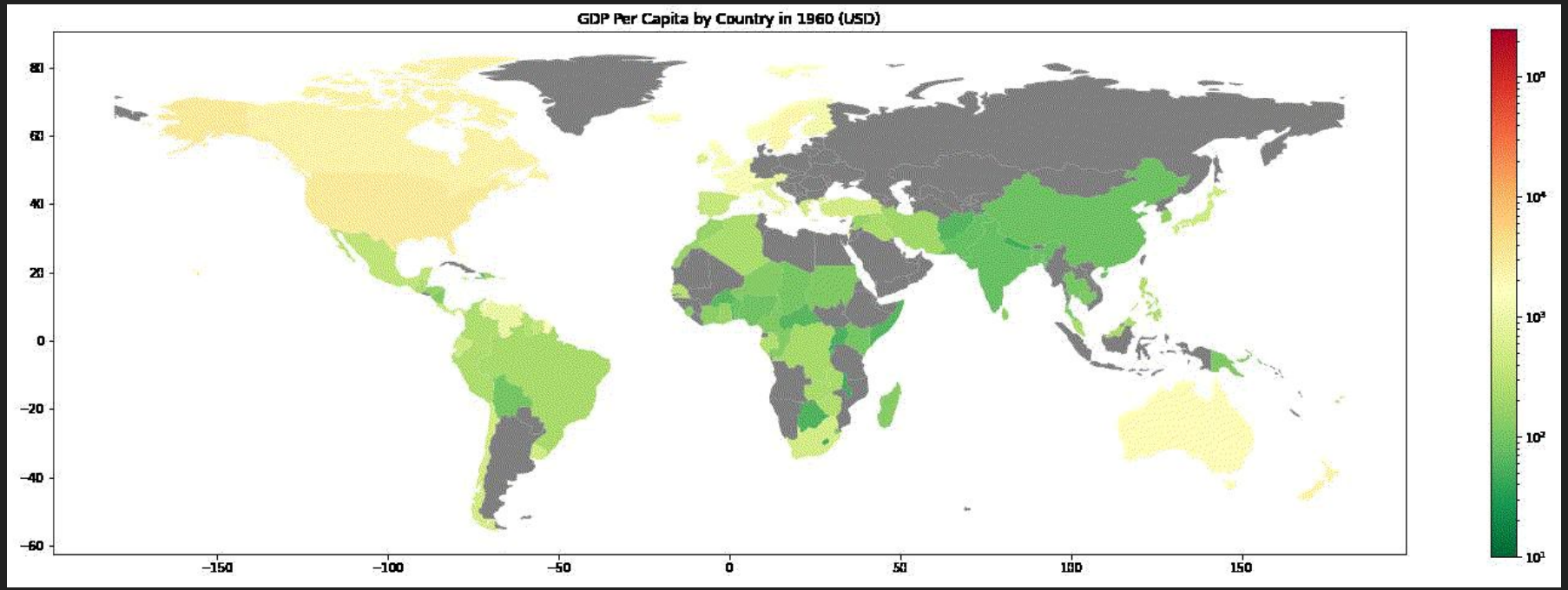

CO2 Emissions by Country 1750 (tons)
80
60
40
20
0
-20
-40
-60
-150
-100
-50
0
50
100
150
10^{10}
10^{8}
10^{6}
10^{4}
10^{2}
10^{0}

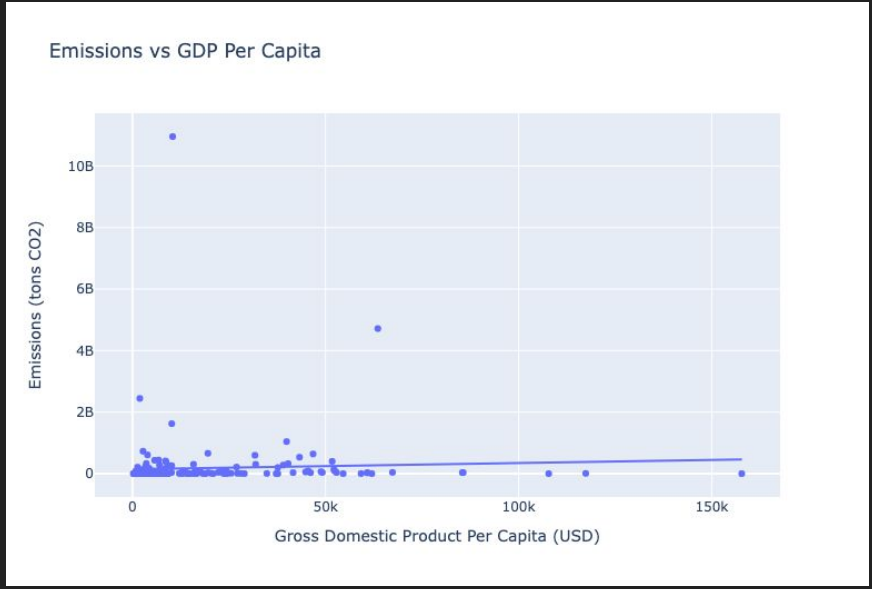
Predicting Variables

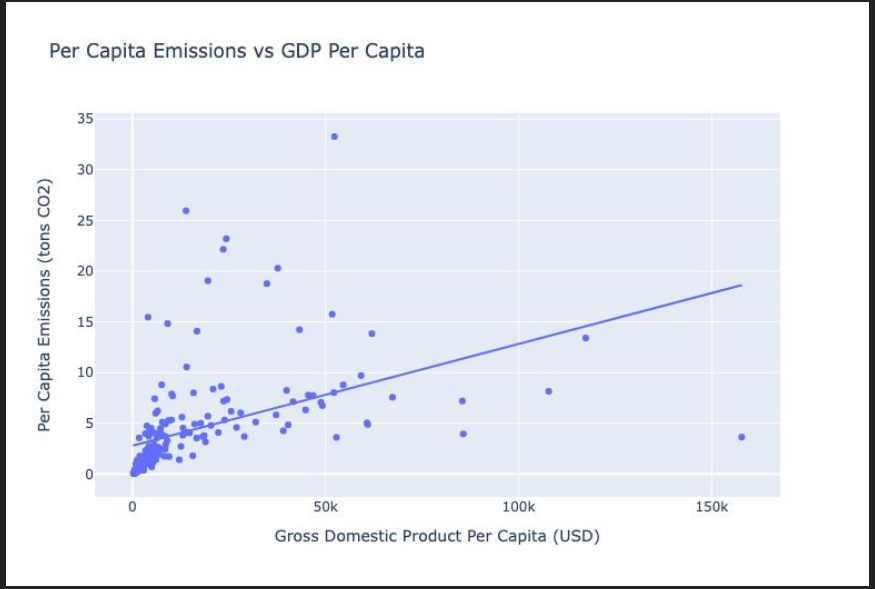
1. Population


Per Capita Emissions vs Population and Urban Population Percentage
Urban Population Percentage
100
80
60
40
20
5
100
2
5
1000
2
5
10k
2
5
100k
2
5
1M
2
Population (thousands)

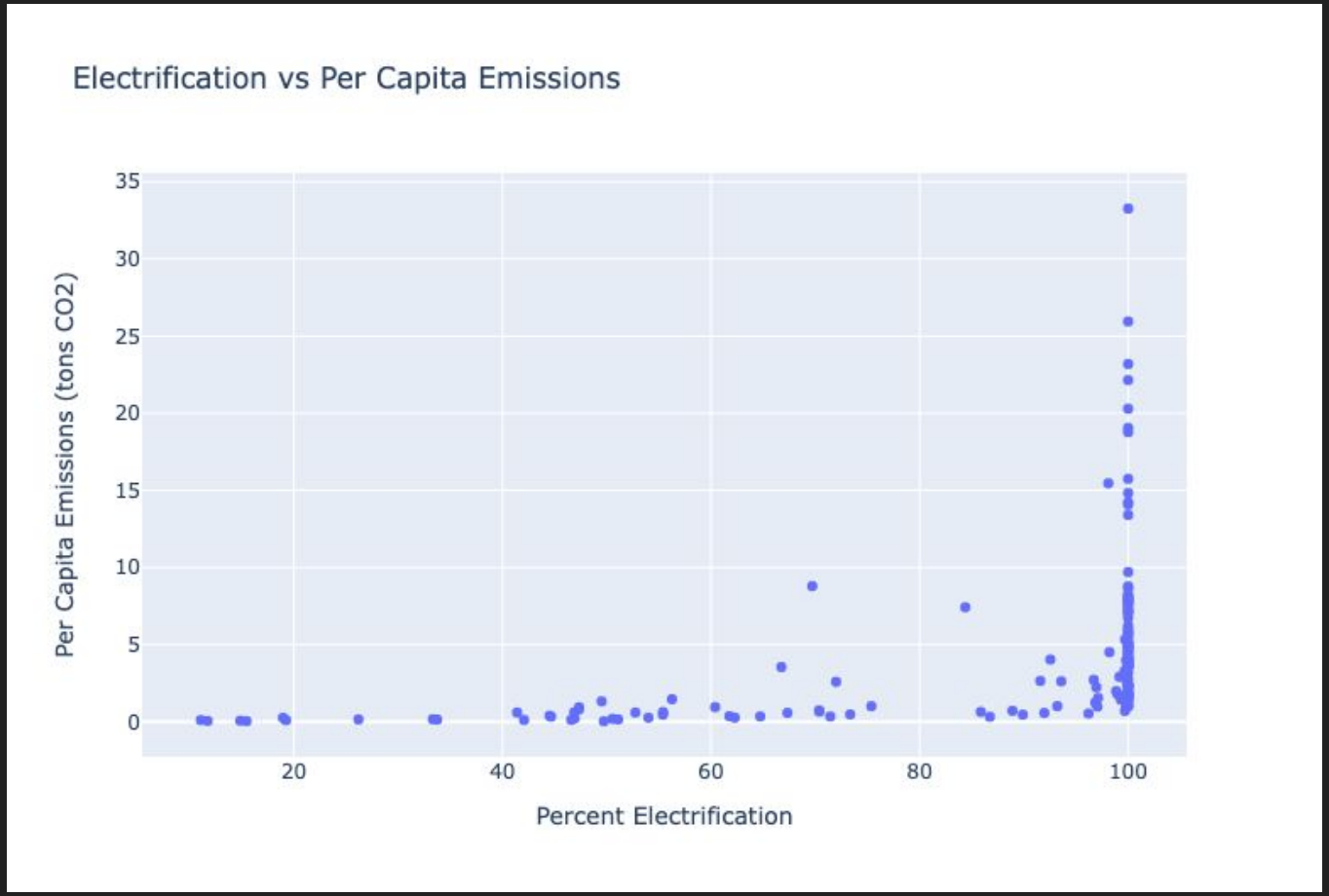
2. GDP Per Capita


GDP Per Capita by Country in 1960 (USD)
80
60
40
20
0
−20
−40
−60
−150
−100
−50
0
50
100
150
10^3
10^4
10^3
10^2
10^1

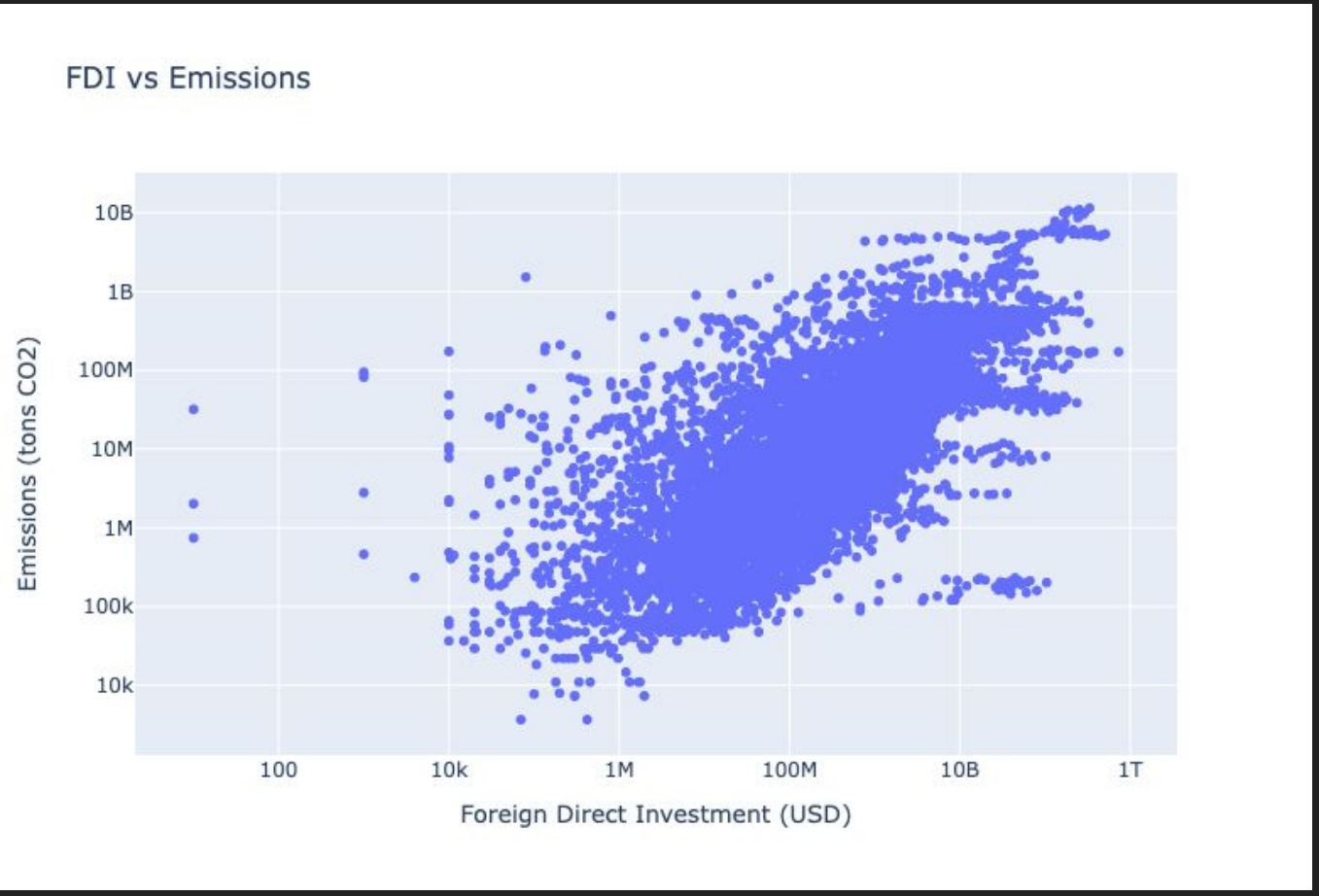

Emissions vs GDP Per Capita
Emissions (tons CO2)
10B
8B
6B
4B
2B
0
0
50k
100k
150k
Gross Domestic Product Per Capita (USD)


Per Capita Emissions vs GDP Per Capita
Per Capita Emissions (tons CO2)
35
30
25
20
15
10
5
0
0
50k
100k
150k
Gross Domestic Product Per Capita (USD)

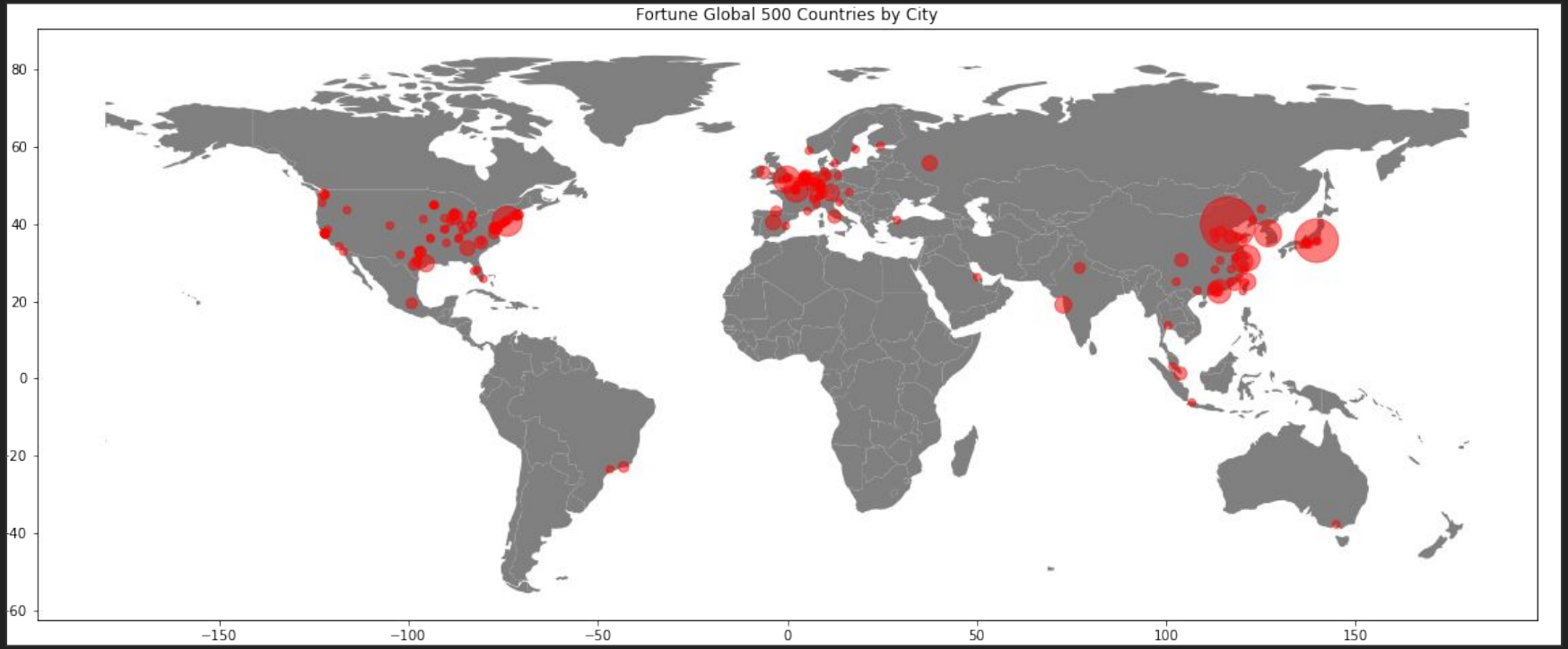
3. Electrification

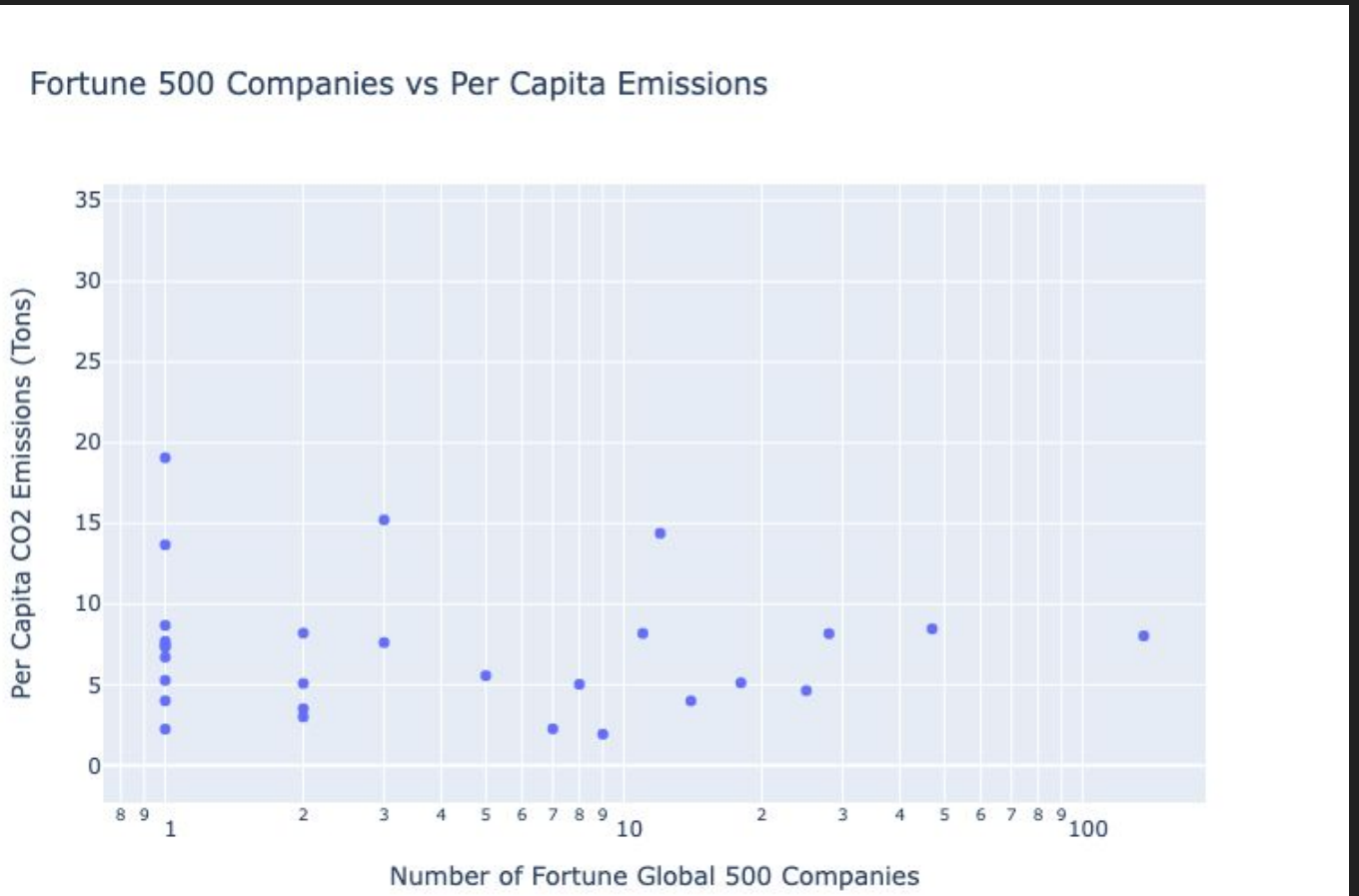

Electrification vs Per Capita Emissions
Per Capita Emissions (tons CO2)
35
30
25
20
15
10
5
0
20
40
60
80
100
Percent Electrification

4. Foreign Direct Investment


FDI vs Emissions
Emissions (tons CO2)
10B
1B
100M
10M
1M
100k
10k
100
10k
1M
100M
10B
1T
Foreign Direct Investment (USD)

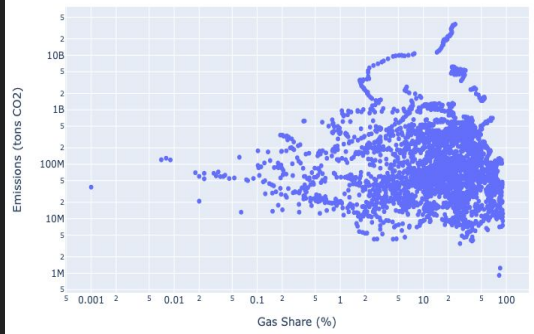
5. Fortune 500 Companies


Fortune Global 500 Countries by City
80
60
40
20
0
-20
-40
-60
−150
−100
−50
0
50
100
150

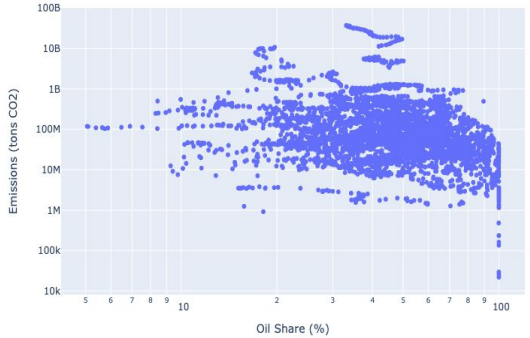

Fortune 500 Companies vs Per Capita Emissions
Per Capita CO2 Emissions (Tons)
35
30
25
20
15
10
5
0
8 9 1 2 3 4 5 6 7 8 9 10 2 3 4 5 6 7 8 9 100
Number of Fortune Global 500 Companies

6. Energy Sectors

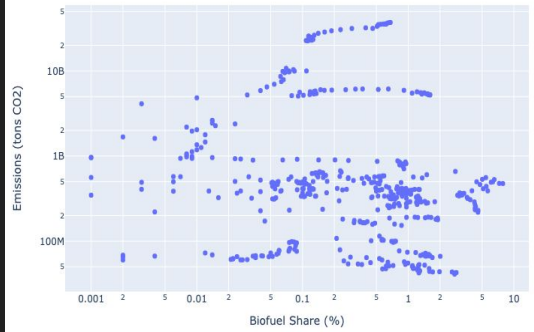
Gas Share (%) vs Emissions



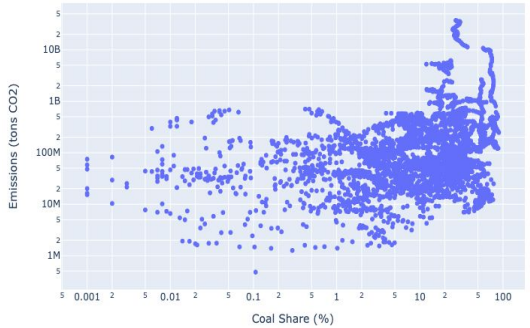
Oil Share (%) vs Emissions



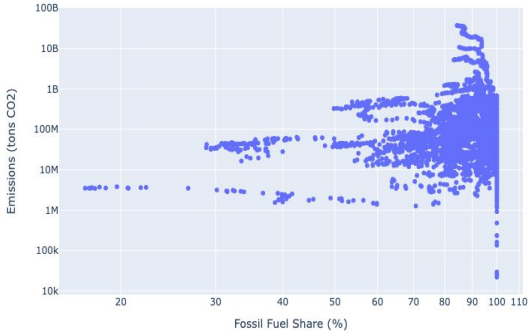
Biofuel Share (%) vs Emissions



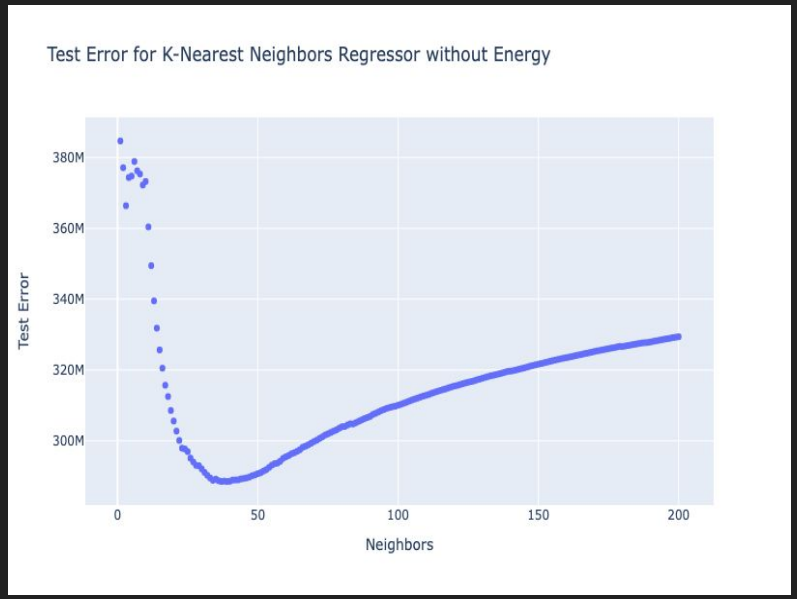
Coal Share (%) vs Emissions

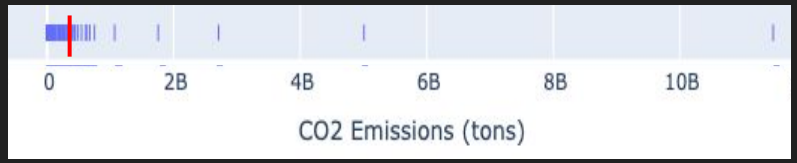


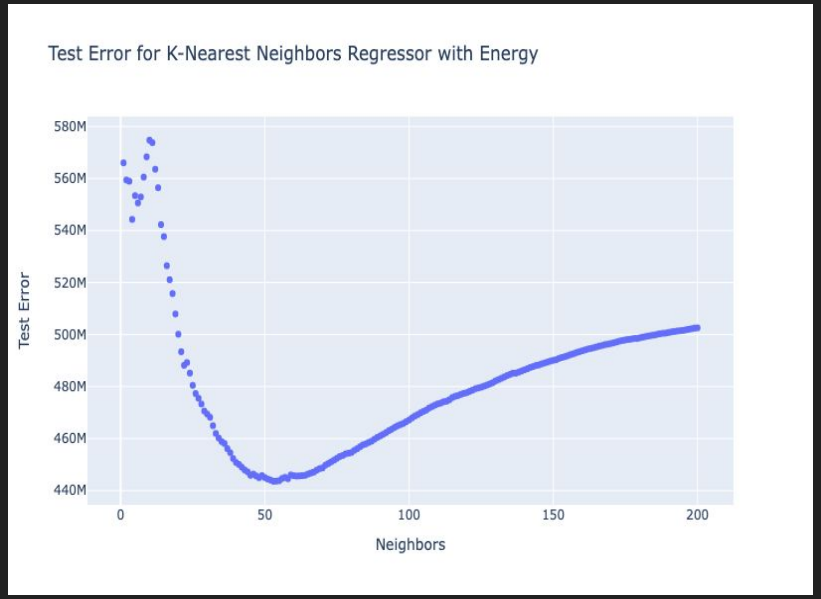
Fossil Fuel Share (%) vs Emissions

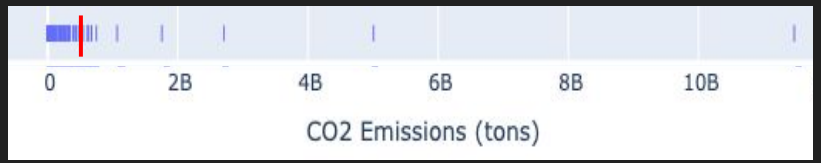


Machine Learning


Test Error for K-Nearest Neighbors Regressor without Energy
Test Error
380M
360M
340M
320M
300M
0
50
100
150
200
Neighbors

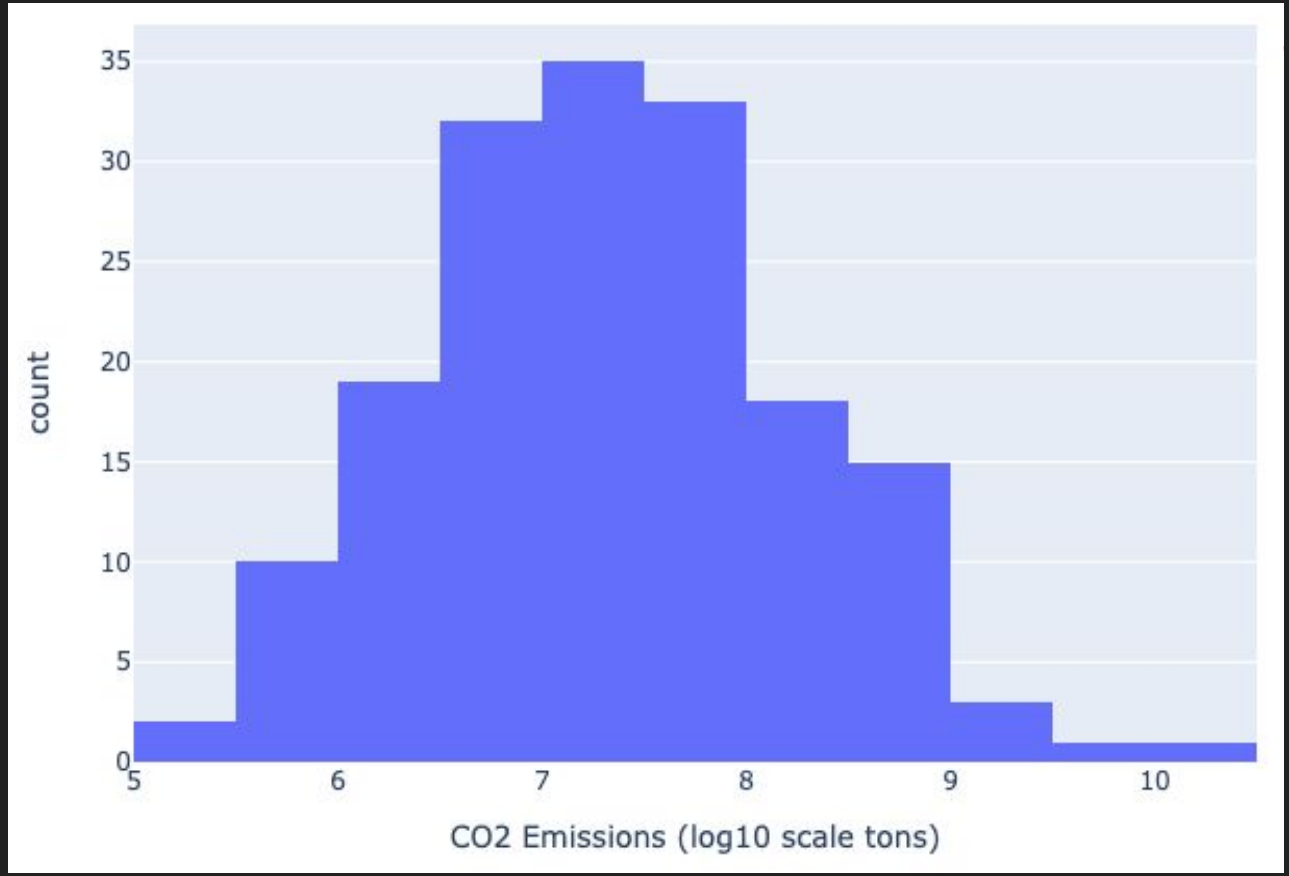

0
2B
4B
6B
8B
10B
CO2 Emissions (tons)


Test Error for K-Nearest Neighbors Regressor with Energy
Test Error
580M
560M
540M
520M
500M
480M
460M
440M
0
50
100
150
200
Neighbors


0
2B
4B
6B
8B
10B
CO2 Emissions (tons)

	Without Energy Data	With Energy Data
K-Nearest Neighbors	2.9 x 108 tons CO2	4.4 x 10^8 tons CO_2
Linear Regression	4.4 x 10^8 tons CO_2	7.9 x 10^8 tons CO_2
Voting Ensemble	3.3 x 10^8 tons CO_2	5.6 x 10^8 tons CO_2
XGBoost	3.2 x 10^8 tons CO_2	6.4 x 10^8 tons CO_2
XGBoost with Hyperparameter Tuning	2.7 x 10^8 tons CO_2	5.1 x 10^8 tons CO_2

Log Normal Distribution


35
30
25
20
15
10
5
0
count
5
6
7
8
9
10
CO2 Emissions (log10 scale tons)

Normal RMSE

$$y\_pred + RMSE = y\_true$$

Off by Sum

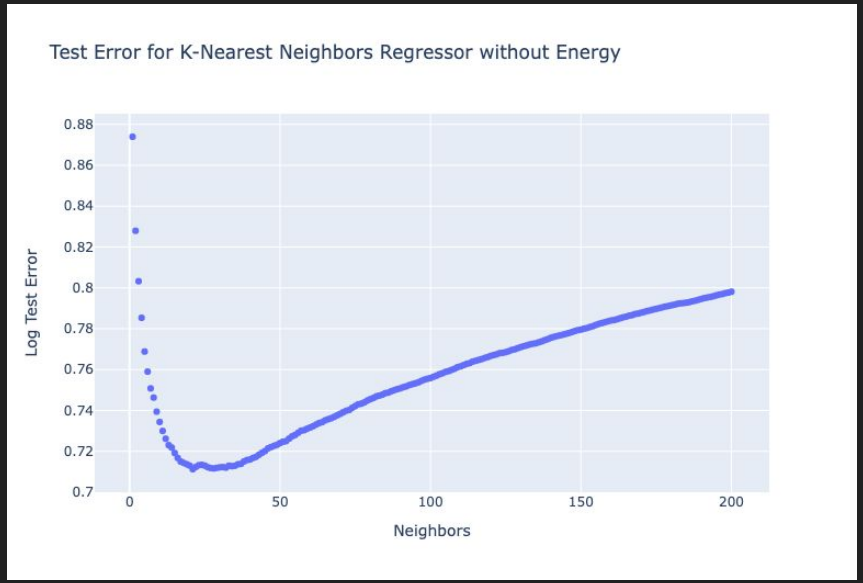
Log Scale RMSE

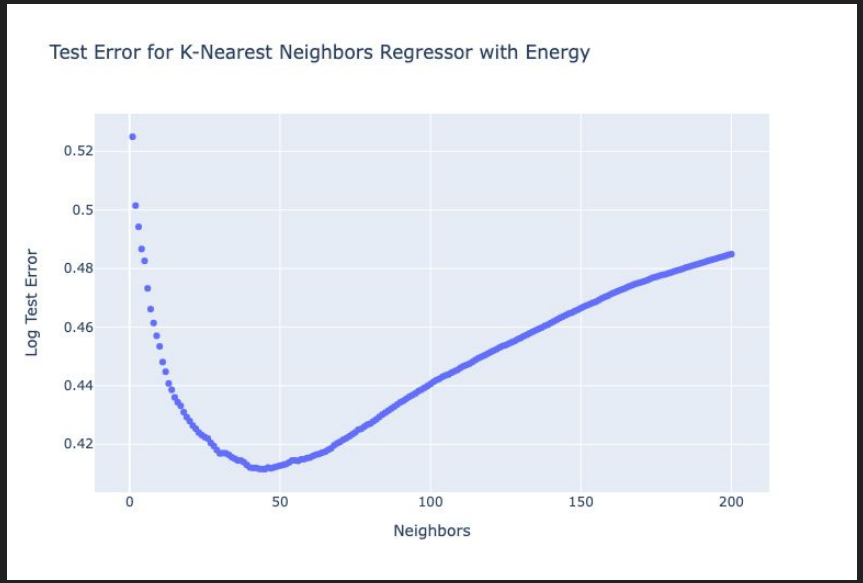
$$\log_{10}(y\_pred) + RMSE = \log_{10}(y\_true)$$

$$10^{\log10(y\_pred) + RMSE} = 10^{\log10(y\_true)}$$

$$y\_pred * 10^{RMSE} = y\_true$$

Off by Factor


Test Error for K-Nearest Neighbors Regressor without Energy
Log Test Error
0.88
0.86
0.84
0.82
0.8
0.78
0.76
0.74
0.72
0.7
0
50
100
150
200
Neighbors


Test Error for K-Nearest Neighbors Regressor with Energy
Log Test Error
0.52
0.5
0.48
0.46
0.44
0.42
0
50
100
150
200
Neighbors

	Without Energy Data	With Energy Data
K-Nearest Neighbors	0.711	0.411
Linear Regression	0.846	0.423
Voting Ensemble	0.739	0.374
Stacking Ensemble	0.707	0.385
XGBoost	0.325	0.268
XGBoost with Hyperparameter Tuning	0.301	0.258

Accuracy

Normal RMSE

RMSE > Mean

RMSE ≈ 30% std

ML is about 3x as accurate as guessing the average value

Log Scale RMSE

RMSE = 0.258

$y\_pred * 10^{0.258} = y\_true$

$y\_pred * 1.8 = y\_true$

ML is within a factor of 2 on average