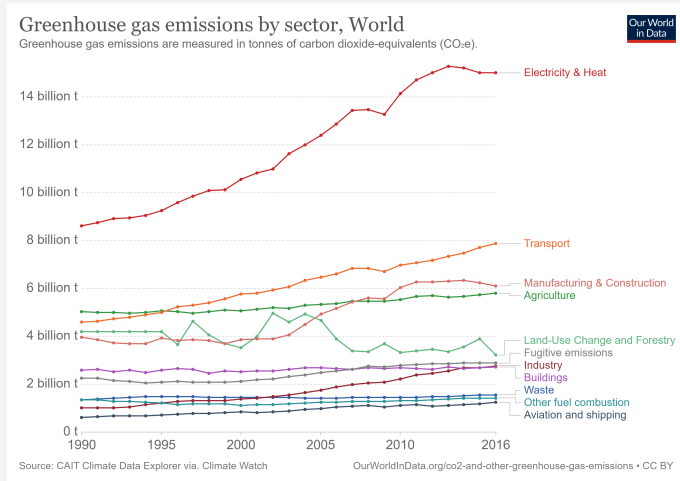


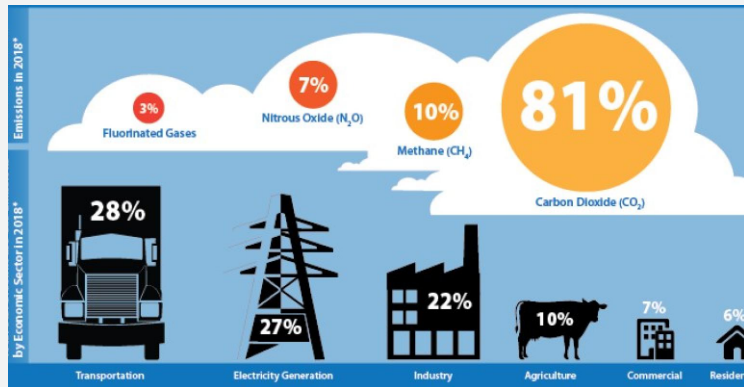
GREENHOUSE GAS EMISSIONS REGRESSION ANALYSIS

By Samuel Robbins

Project Premise



- Greenhouse Gases (GHG) contribute significantly to ongoing climate change and are key factors when considering the magnitude of future warning
- The premise behind this regression analysis is to model GHG emissions as a function of company size, industry, location, and other key features
- Societal Goal: predict the largest emitters based on the given criteria to target legislative, activist, and social lobbying campaigns



Why Model Greenhouse Gas Emissions?

VB VentureBeat

AI can estimate corporate greenhouse gas emissions

Researchers claim to have developed a model that can estimate the greenhouse gas emissions of companies in a range of industries.

7 hours ago



Datasets

	Facility Id	Facility Name_x	State	Industry Type (sectors)	Unit Name	Unit Type	Unit Maximum Rated Heat Input Capacity (mmBTU/hr)	Unit CO2 emissions (non-biogenic)	Unit Methane (CH4) emissions	Unit Nitrous Oxide (N2O) emissions	Unit Biogenic CO2 emissions (metric tons)	Facility Name_y	General Fuel Type	Specific Fuel Type
0	1000112	23rd and 3rd	NY	Power Plants	2301	Electricity Generator	NaN	24768.9	11.50	14.90	0.0	23rd and 3rd	Natural Gas	Natural Gas (Weighted U.S. Average)
1	1000112	23rd and 3rd	NY	Power Plants	2302	Electricity Generator	NaN	19780.8	9.25	11.92	0.0	23rd and 3rd	Natural Gas	Natural Gas (Weighted U.S. Average)
2	1000112	23rd and 3rd	NY	Power Plants	Heatec1	PRH (Process Heater)	7.40000	30.2	0.00	0.00	0.0	23rd and 3rd	Natural Gas	Natural Gas (Weighted U.S. Average)
3	1000112	23rd and 3rd	NY	Power Plants	Heatec2	PRH (Process Heater)	7.40000	33.0	0.00	0.00	0.0	23rd and 3rd	Natural Gas	Natural Gas (Weighted U.S. Average)
4	1003742	31st Street Landfill	IL	Waste	CP-1	OCS (Other combustion source)	0.20125	246.9	0.00	0.00	0.0	31st Street Landfill	Natural Gas	Natural Gas (Weighted U.S. Average)

Datasets

	GHGRP FACILITY ID	FRS ID (FACILITY)	FACILITY NAME	FACILITY STATE	PARENT COMPANY NAME
0	1000355	1.100030e+11	#540 BONANZA CREEK ENERGY - DENVER BASIN	CO	BONANZA CREEK ENERGY INC
1	1004377	1.100438e+11	121 REGIONAL DISPOSAL FACILITY	TX	NORTH TEXAS MUNICIPAL WATER DISTRICT
2	1009238	1.100144e+11	220 Gulf Coast Basin DEC	OK	DEVON ENERGY CORP
3	1000112	1.100198e+11	23rd and 3rd	NY	NEW YORK POWER AUTHORITY
4	1012156	1.100701e+11	260 - East Texas Basin Gathering/Boosting	TX	BP AMERICA INC

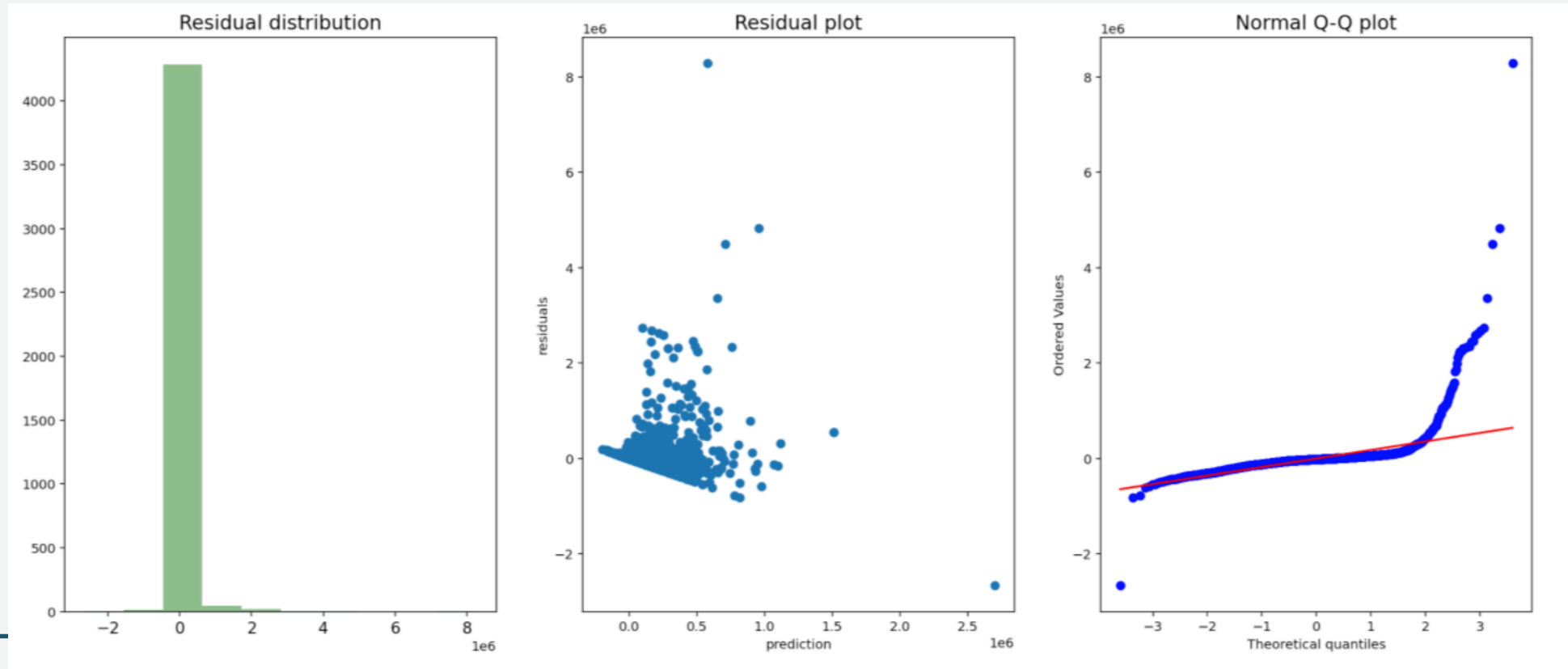
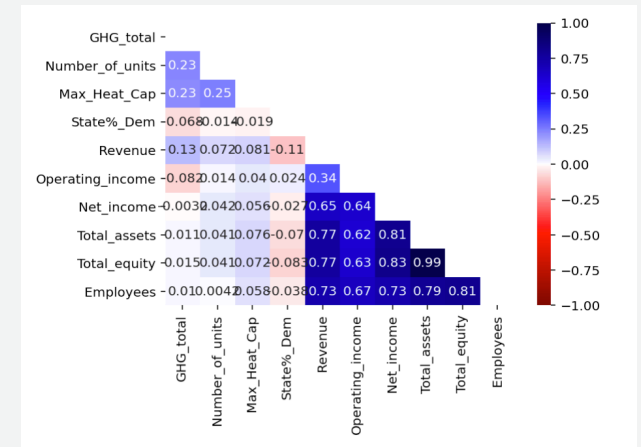
Datasets

	Company	Revenue	Operating_income	Net_income	Total_assets	Total_equity	Employees
0	Air_Products_&_Chemicals	8.856000e+09	2.144000e+09	1.931000e+09	2.516920e+10	1.244300e+10	19275.0
1	Alliant_Energy_Corp	3.416000e+09	7.400000e+08	6.240000e+08	1.771031e+10	5.888000e+09	3375.0
2	Ameren_Corp	5.794000e+09	1.300000e+09	8.710000e+08	3.203031e+10	9.080000e+09	8615.0
3	American_Electric_Power	8.879000e+09	1.507000e+09	1.064000e+09	8.075731e+10	2.077400e+10	17666.0
4	Anheuser-Busch	1.558800e+10	NaN	9.811202e+09	NaN	NaN	NaN
...

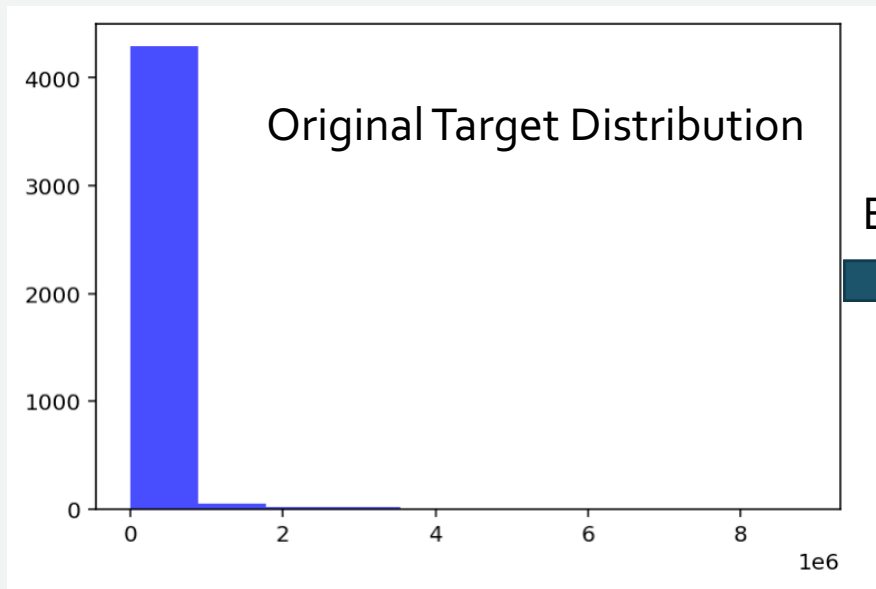
Datasets

	State	Dem_margin%_2000	Dem_margin%_2004	Dem_margin%_2008	Dem_margin%_2012	Dem_margin%_2016	Dem_margin%_2020	Dem_margin_avg%
0	AL	-14.92	-25.62	-21.58	-22.19	-27.73	-25.46	-14.92
1	AK	-30.95	-25.55	-21.54	-13.99	-14.73	-10.06	-30.95
2	AZ	-6.29	-10.47	-8.52	-9.06	-3.50	0.31	-6.29
3	AR	-5.44	-9.76	-19.85	-23.69	-26.92	-27.62	-5.44
4	CA	11.80	9.95	24.06	23.12	30.11	29.16	11.80

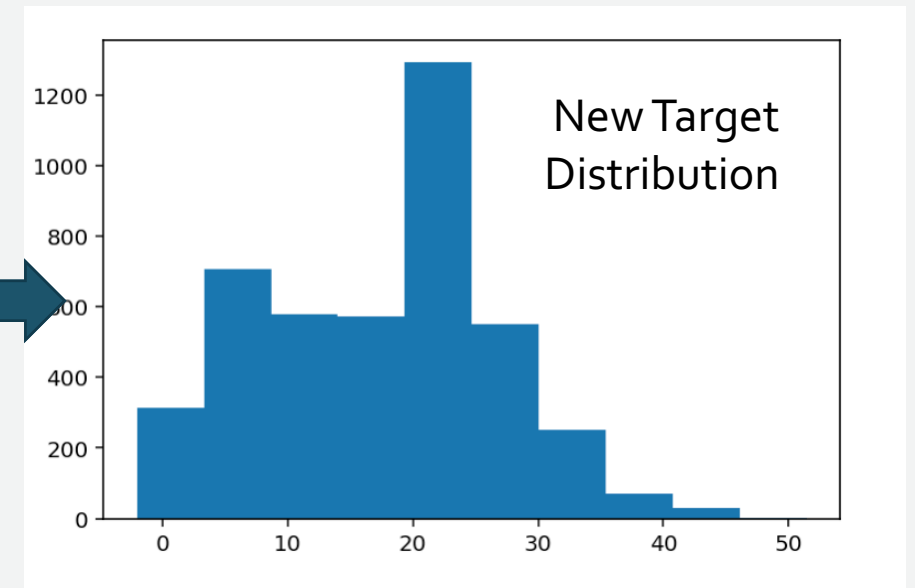
Modeling and Feature Transformations



Modeling and Feature Transformations

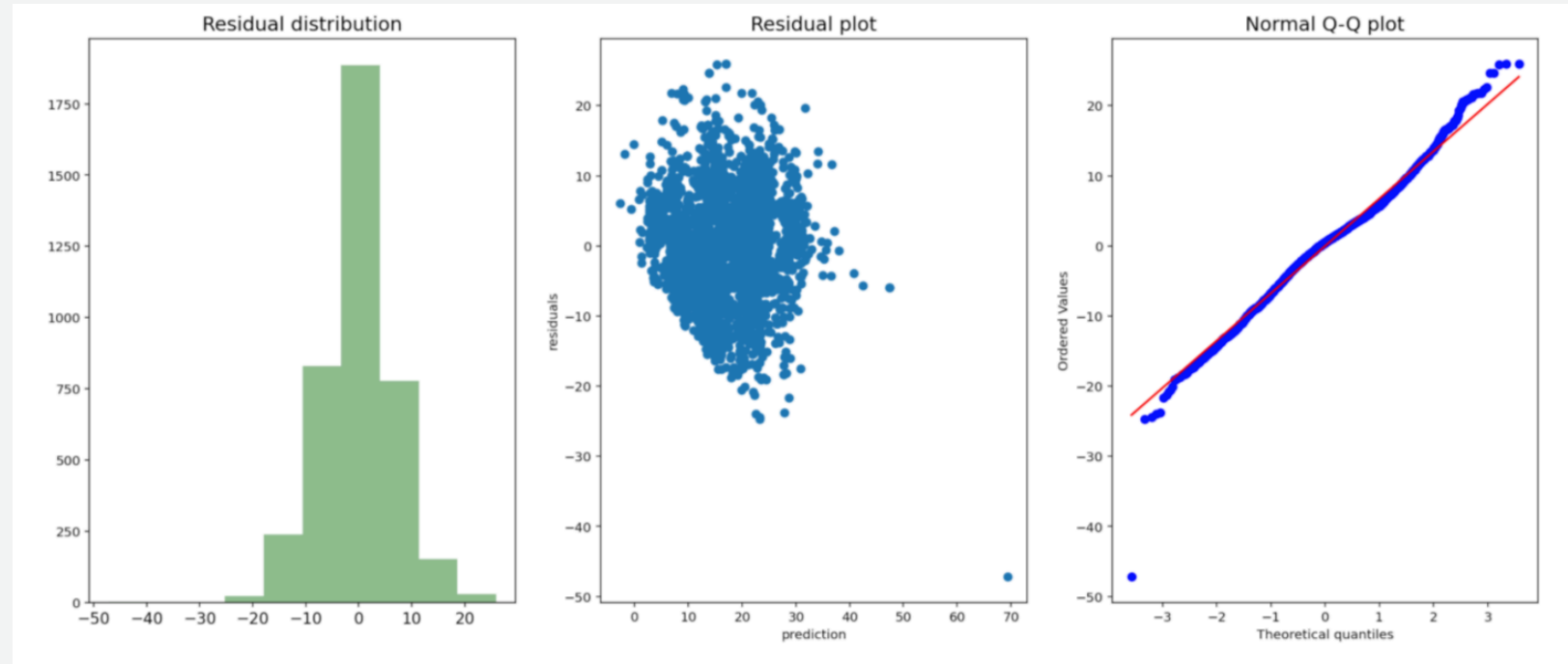


Box-Cox Transformation



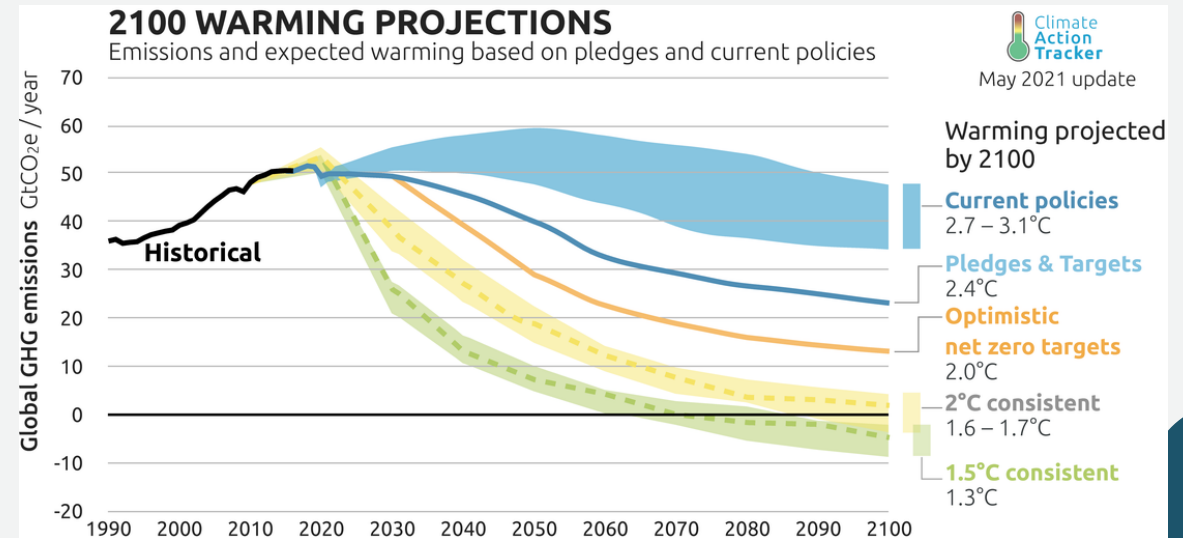
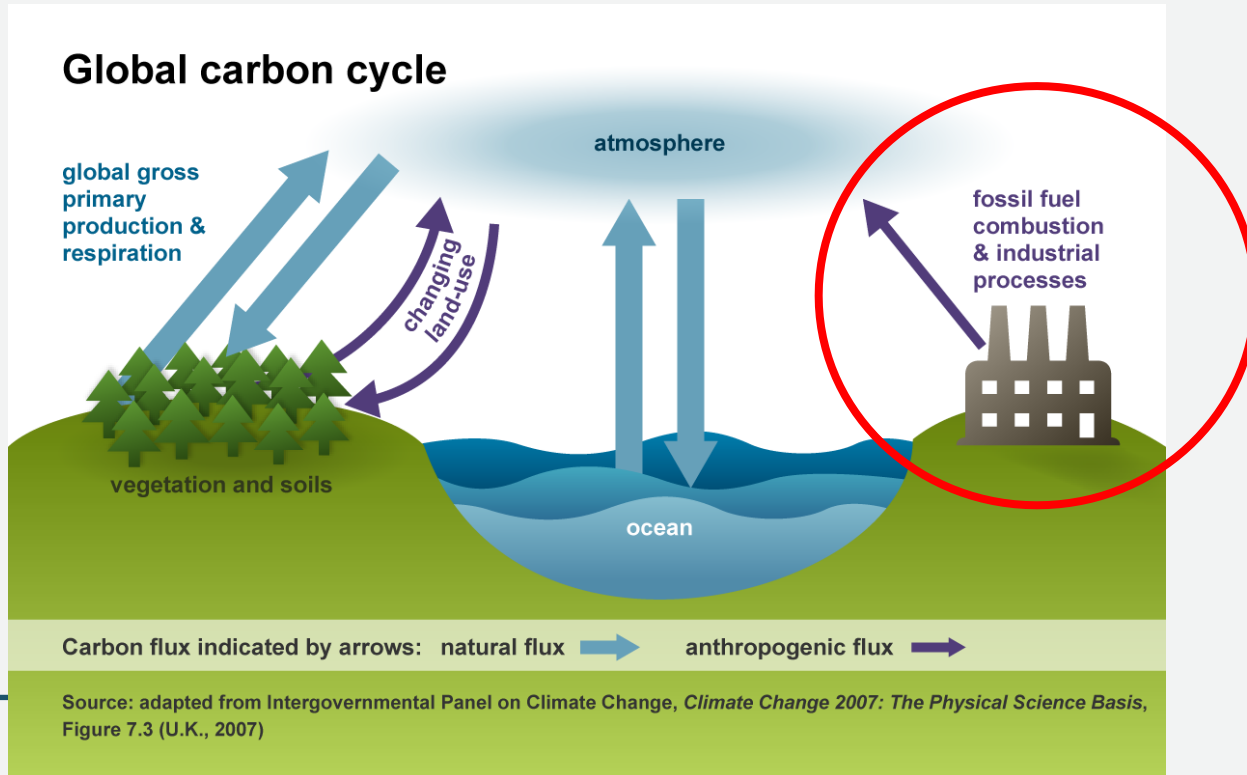
Final Model

- Raised R^2 from baseline of 0.2099 to 0.4815
- Model parameters:
 - dropped columns for Operating Income, Total Equity, and Net Income due to their multicollinearity
 - Transformed y via box-cox transformation
- Another good model was the log of financial with a box-cox transformation on y

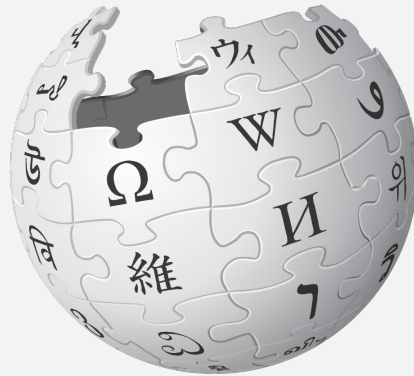


Closing Thoughts

- Very difficult problem to model
- Possible future work: narrow parent companies down to subsidiaries
- EX. Berkshire Hathaway (Warren Buffets company) is a giant multinational corporation so that financial data is being trickled down to even their smallest facility



References



WIKIPEDIA
The Free Encyclopedia

- <https://www.epa.gov/ghgreporting/ghg-reporting-program-data-sets>
 - Reported Parent Companies (xls)
 - Emissions by Unit and Fuel Type (zip)
- <https://en.wikipedia.org/wiki/ExxonMobil>
 - Parent Company Financial Data
- https://en.wikipedia.org/wiki/2000_United_States_presidential_election#Results
 - Election Results since 2000

APPENDIX

