Data Collection Plan

Project number:	7	Project title:	Unveiling the Future of Carbon Emissions Trading: A Machine Learning and Neural Network Perspective on Regional Markets.	Project leader:	Syed Faraaz

Date: 04/15/2024

Description of the data collection
Table 1. State energy-related carbon dioxide emissions by year (1970–2021)
Table 4. Per capita energy-related carbon dioxide emissions by state (1970–2021)
Table 6. Carbon intensity of the energy supply by State (1970–2021)
Table 7. Carbon Intensity of the economy by state (1997–2021)
What will be done with the data once it has been collected?
After obtaining the data, we are selecting and filtering the necessery tables from like 8 different tables and after obtaining the final dataset then we have preprocess the data and carried out necessary data augmentation. We then use to train the models and then prdicting the final values.

Key Variables - A summary of the chosen input variables (Y's) and/or output variables (X's)

		1	2	3	4	5	6	7
What?	**Variable title**	State - Code	State	Table 1: Years (1970 - 2021)	Table 4: Years (1970-2021)	Table 6: Years (1970-2021)	Table 7: Years (1997-2021)	Futue data Prediction
	Input (X) or output (Y) variable?	X	X	X	X	X	X	Y
	Unit of measurement	N/A	N/A	million metric tons	kilograms/million Btu	million metric tons	million metric tons	million metric tons
	Data type	String	String	Numerical	Numerical	Numerical	Numerical	Numerical
	Collection method	U.S. Energy Information Administration						
	If manual	No	No	No	No	No	No	No
Historical data	Historical data exist?	N/A						
	Source of historical data	N/A						
	Historical data representative/reliable?	N/A						
Who?	**Data collector**	Prayag Nikul Purani						
	Operational definition exist?	No	No	No	No	No	No	No
	Data collector trained?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Resources available for data collector?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
When?	**Start date**	4/7/2024	4/7/2024	4/7/2024	4/7/2024	4/7/2024	4/7/2024	4/7/2024
	Due date	4/14/2024	4/14/2024	4/14/2024	4/14/2024	4/14/2024	4/14/2024	4/14/2024
	Duration (in days)	8	8	8	8	8	8	8

Additional Comments - e.g. resolution needed, sampling method, R&R results, storing of data, handling outliers, using filters, etc.
1. The sorting the time-frame which concides with the avalibility of the data and from where the awarness regarding the carbon dioxide and credits where in the market
2. The combinimg all the files into one dataset by using star-schema
3. Sorting the units of different files so can be compiled into single file
4. Selection of the time-frame