Global Emissions & Electric Vehicles

Group 15 - Green Wave

TEAM GREEN WAVE

Kobe Pho

Ruchita Agarwal

Avneesh Sawhney

Tarini Pal

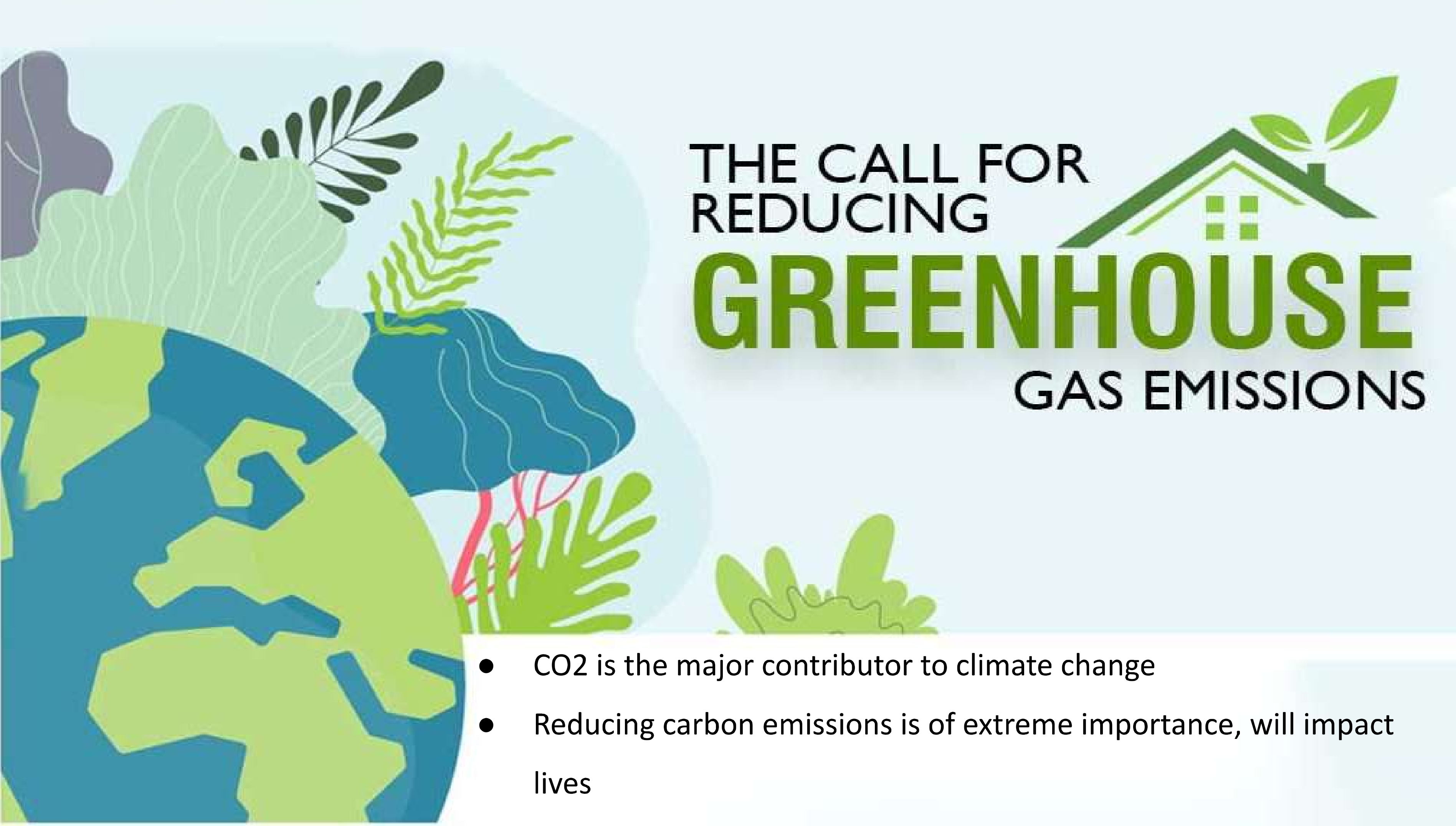
Mikhail Tikhanchikov

Agenda

- Project Objectives
- Understanding Carbon/GHG emissions and Leading Factors
- Data Collection and Preprocessing
- Data Analysis and Outcomes
- Machine Learning Design & Dashboard

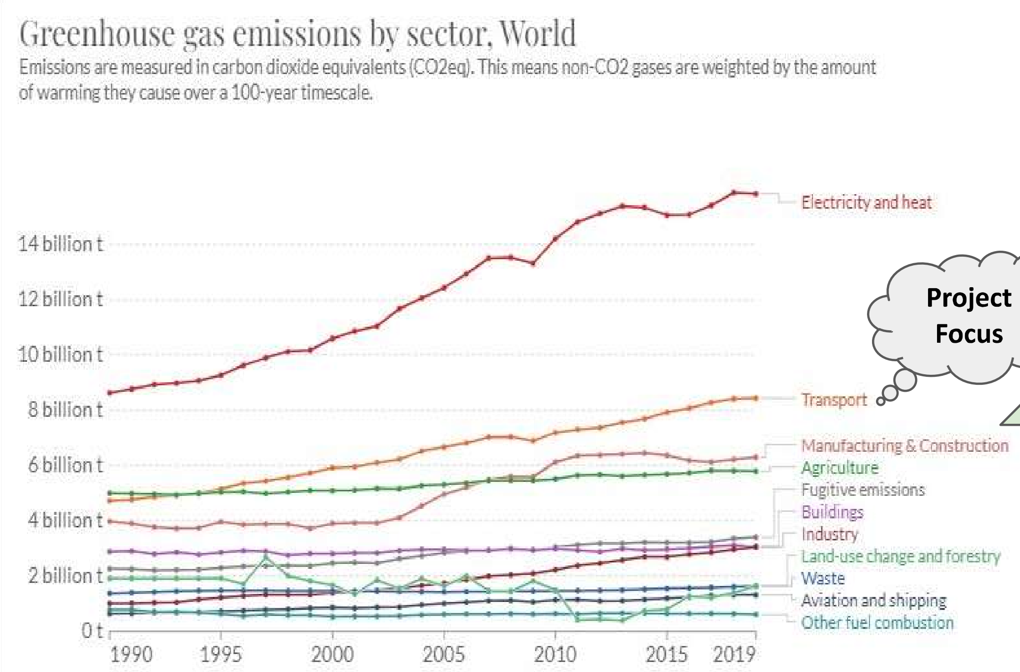
Main Objectives

- Analyzing what factors impact carbon emissions globally
- Finding correlations between emission factors
- Determining if EV introduction results in significant positive impacts on reducing carbon transportation emissions
- Corroborating our hypothesis with a peer research review on use of EV and its impact in Norway



- CO2 is the major contributor to climate change
- Reducing carbon emissions is of extreme importance, will impact lives

Overall Picture for GHG emissions



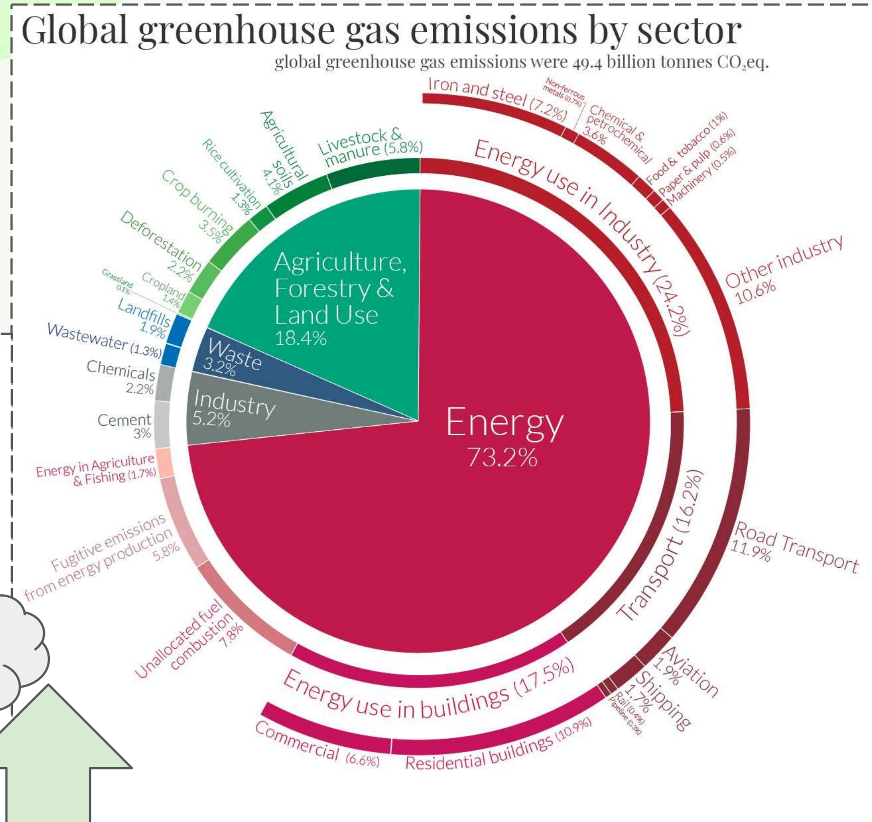
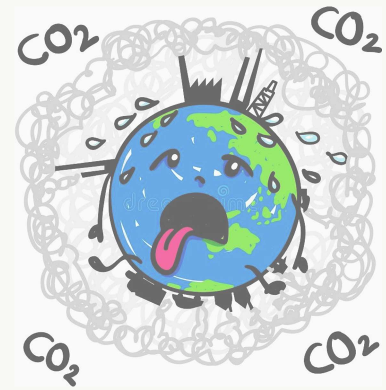
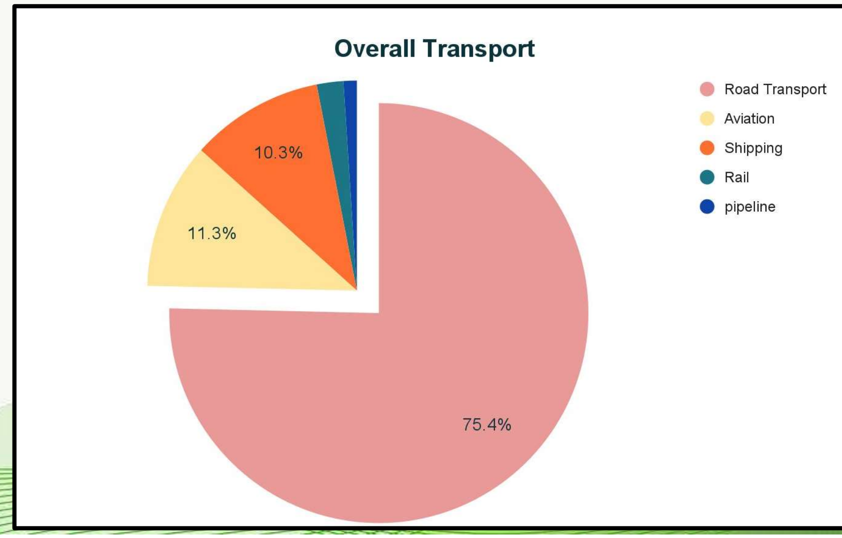


Image source: ourworlddata.org

Impacts of Road Transportation

- Majority of global GHG emissions within transport industry come from road vehicles
- For each gallon of gas that a car burns, it releases about 19 pounds of carbon dioxide through tailpipe emissions





Data Collection / Sources

Data Description / Preparation

- Overall GHG emissions by sector
- Market demand of all vehicles types (Diesel, Electric, Hybrid, etc)
- Merge multiple datasets based on country and year as join keys

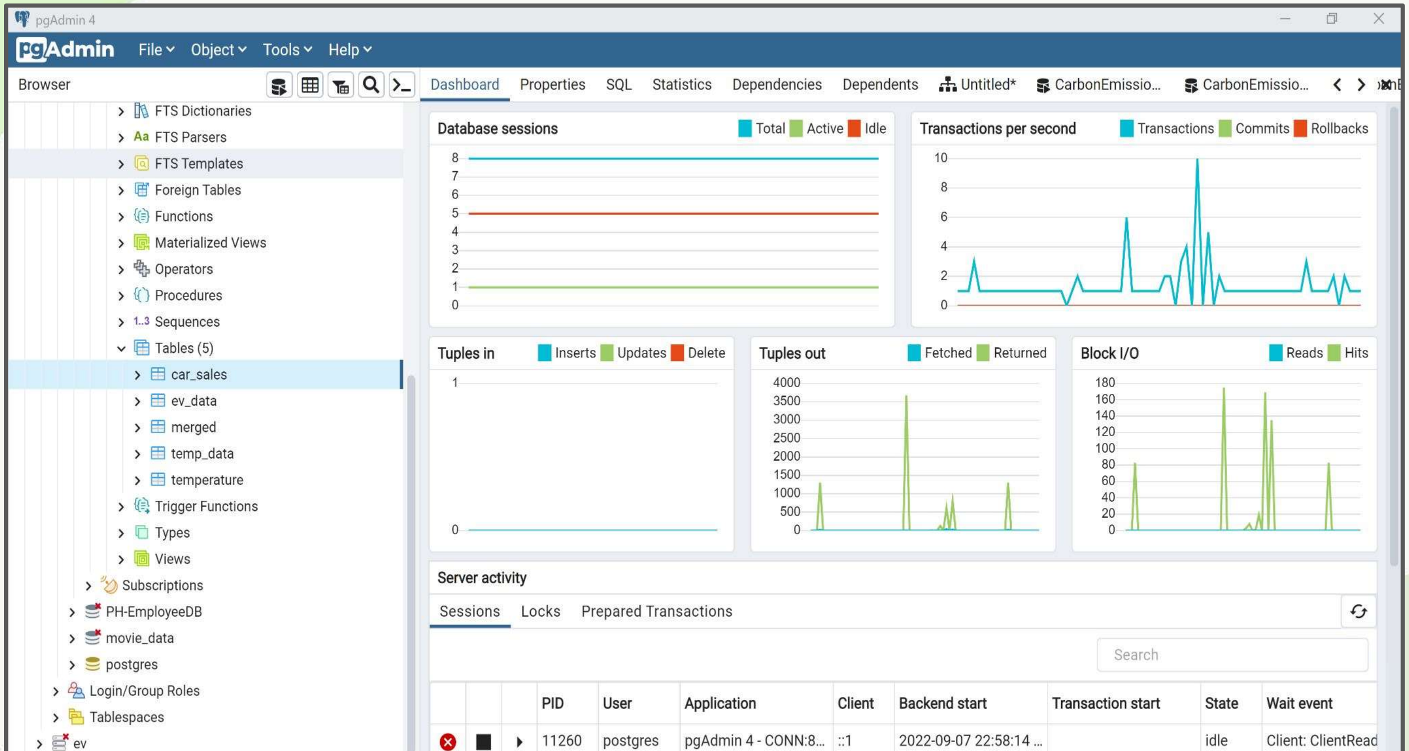
Data Source:

https://ourworldindata.org/

https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

https://www.iea.org/articles/global-ev-data-explorer

Database: pgAdmin



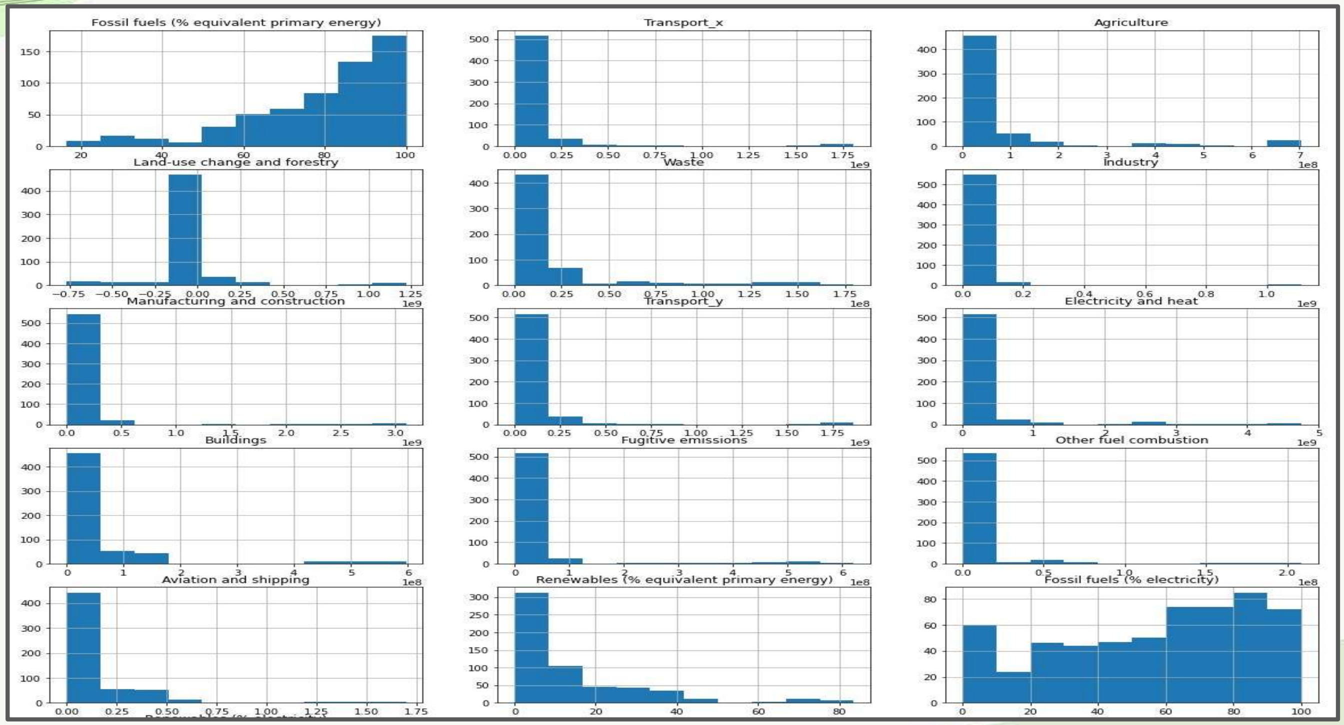
Tools:

- PostgreSQL
- Scikit-learn
- Jupyter Notebook
- Python
  - *Pandas*
  - *Matplotlib*
- AWS
- pgAdmin

Data Cleaning/Wrangling

- Dropped empty values(200 Nan values)
- Data type corrections(object to int)
- Matched mapping of country names and year range (Entity, 2015-2019)
- As a result, our final dataset consists of 59 different countries and emissions data for each individual year (200+ rows)

Univariate Analysis



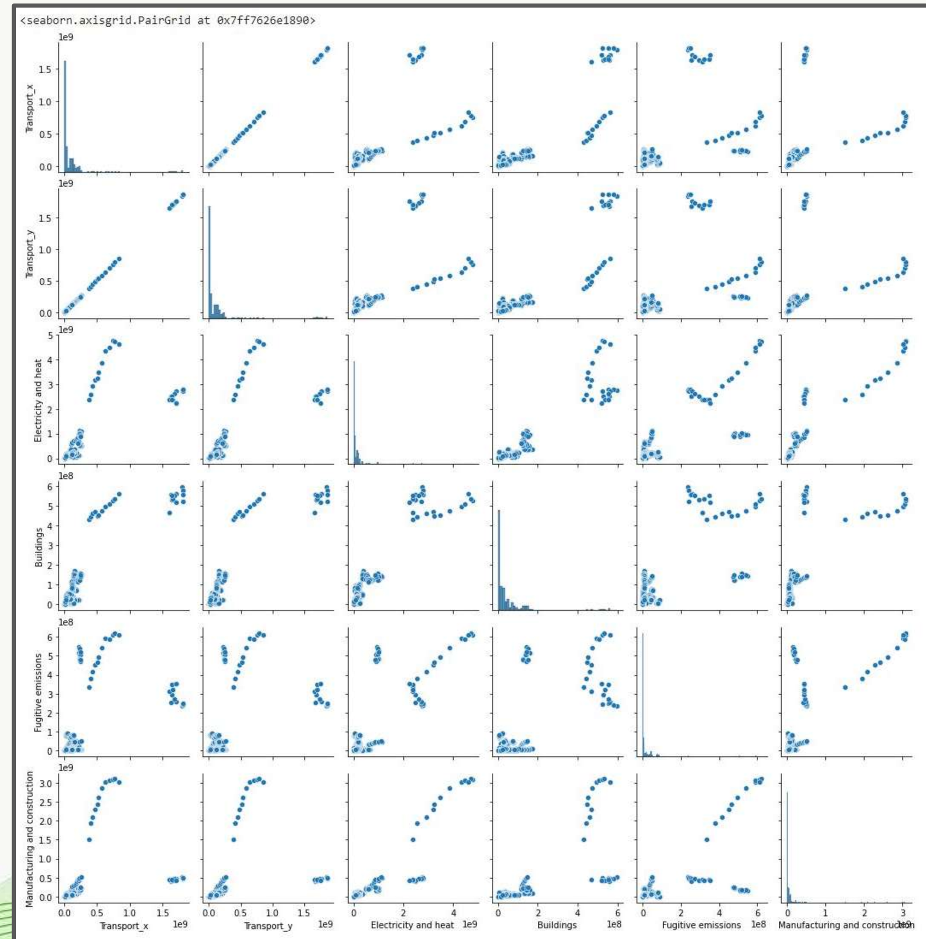
Used univariate analysis to detect and remove outliers and applied log transformation to skewed features

Multivariate Analysis

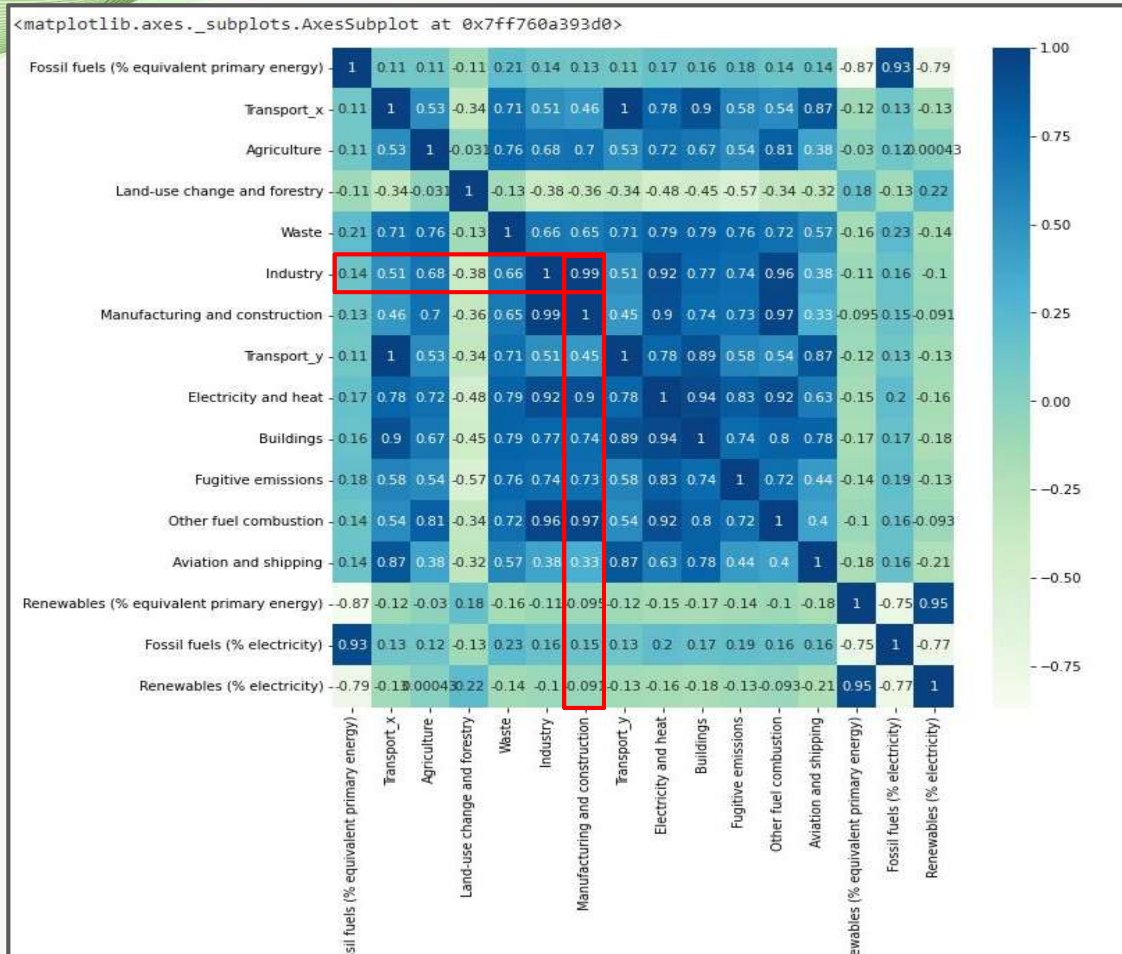
- Used multivariate analysis to detect feature significance and correlation across features

- Transport_x is only CO2 emissions while Transport_y is overall GHG emissions including CO2
    - This is evident from the scatter plots

- We can clearly see correlations between various sectors that emit greenhouse gases



Correlation Matrix



Machine Learning Model Choice

Linear Regression Model

Limitations:

- Accuracy was initially low when including electric bike data
- For some of the features, standard error was really high, which required dropping columns

Benefits:

- Accuracy increased by 20%
- R-squared variance also increased by 20%

Machine Learning Design

Problem Statement:
Create a model that predicts how much CO2-eq by volume is offset by introducing EVs in a market (country+year).

Hypothesis: EV vehicles lead to reduction in overall Greenhouse gases.

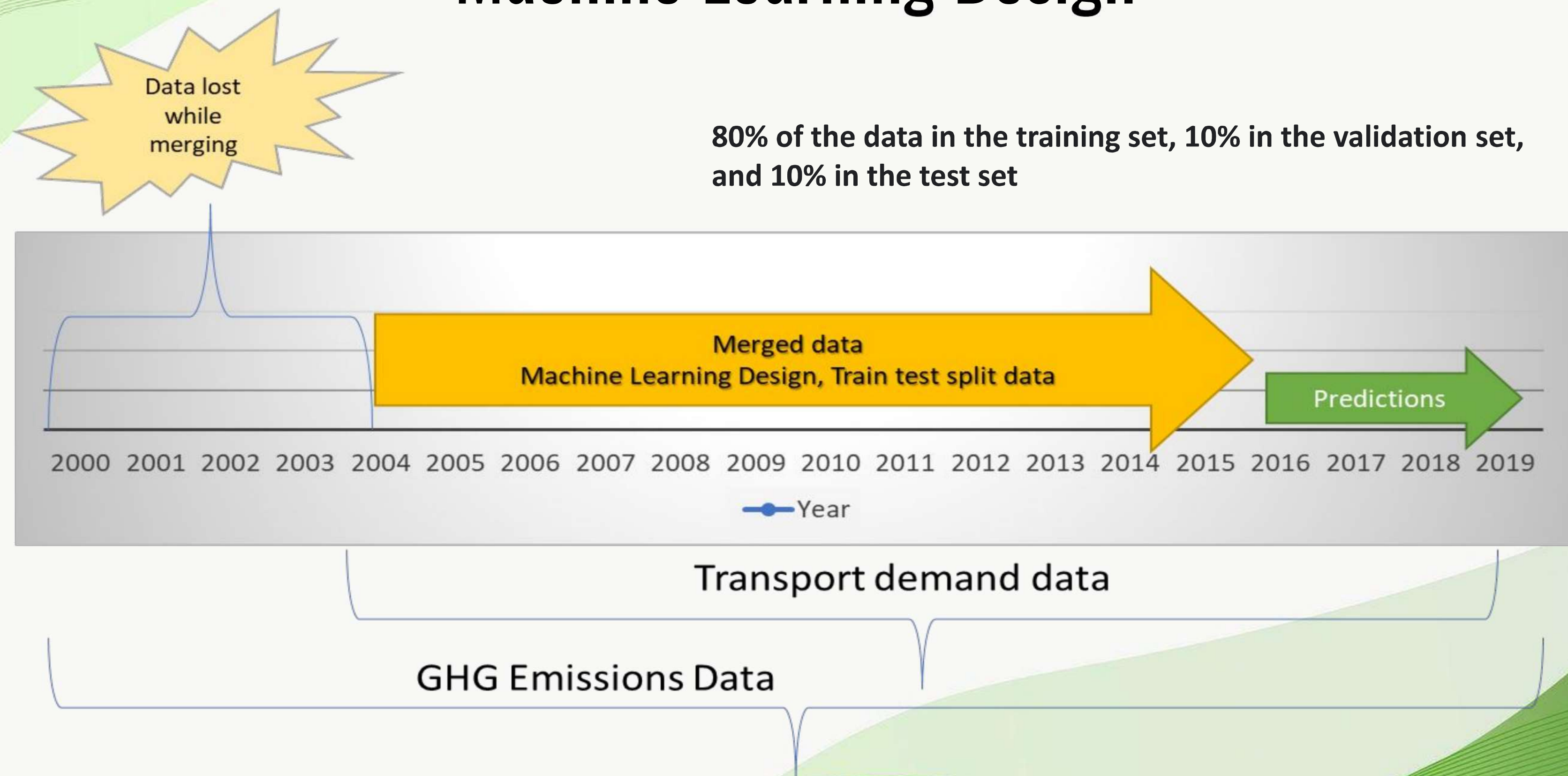
X: *GHG emissions* by **sector** and *energy production* by **fuel types**
Y: CO2 offset due to EV

Target Variable(Y):

$$ev\_offset = transport\_co2eq\_emissions * \frac{ev\_demand\_percent}{gas\_vehicle\_demand\_percent}$$

Machine Learning Design

80% of the data in the training set, 10% in the validation set, and 10% in the test set



Linear Regression Model Performance

```
lm.score(X,y)
0.4954363469172296
```

Dropping Electric bike data improved performance by 10%

```
lm.score(X,y)
0.6014439142014962
```

Train-Test-Validation Splits

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=1)

[383] r2_score(y_test, y_predictions)
0.7284700452410147
```

- *Train-Test-Validation split improved performance by further 12%*

Hypothesis Testing

With a P-value threshold of 5 percent, we can see that the expected sectors are significant while predicting CO2 offset from EVs(P-value *less than* 5%)

- Transport
- Industry
- Manufacturing
- Electricity and Heat
- Buildings
- Fugitive Emissions

We can also see that Primary and Secondary Energy production from Renewables have some significance (P-value *close to* 5%)

OLS Regression Results

Dep. Variable:	ev_savings	**R-squared (uncentered):**	0.606
Model:	OLS	**Adj. R-squared (uncentered):**	0.594
Method:	Least Squares	**F-statistic:**	53.74
Date:	Sat, 10 Sep 2022	**Prob (F-statistic):**	6.63e-102
Time:	07:01:15	**Log-Likelihood:**	-7066.8
No. Observations:	576	**AIC:**	1.417e+04
Df Residuals:	560	**BIC:**	1.424e+04
Df Model:	16		
Covariance Type:	nonrobust		

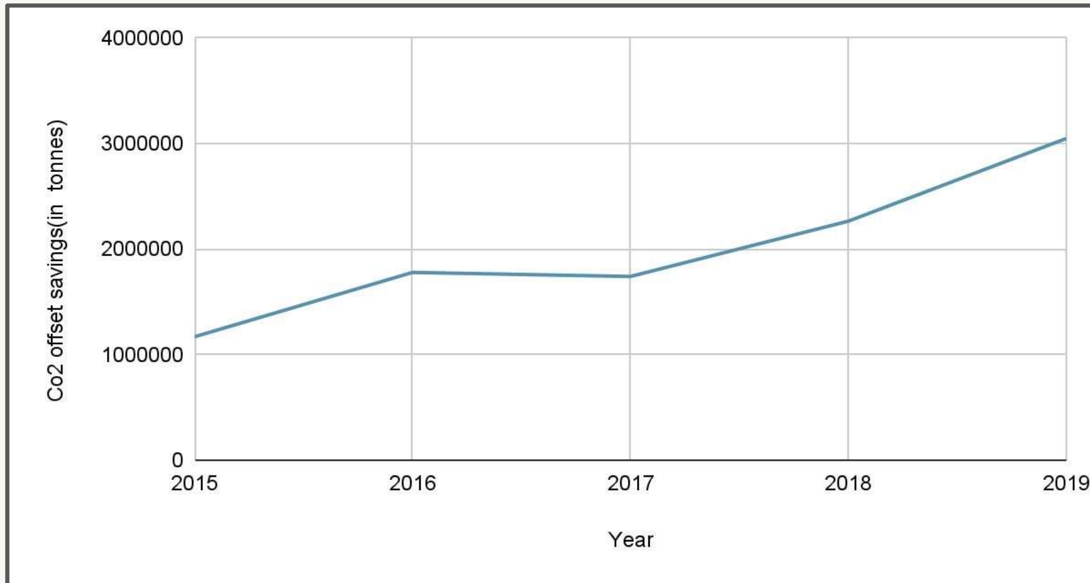
	coef	std err	t	P>|t|	[0.025	0.975]
Fossil fuels (% equivalent primary energy)	171.3201	188.151	0.911	0.363	-198.248	540.888
Transport_x	-0.0944	0.006	-16.298	0.000	-0.106	-0.083
Agriculture	3.493e-05	4.07e-05	0.858	0.391	-4.51e-05	0.000
Land-use change and forestry	-1.824e-05	1.96e-05	-0.932	0.352	-5.67e-05	2.02e-05
Waste	-0.0002	0.000	-0.993	0.321	-0.001	0.000
Industry	0.0009	0.000	4.898	0.000	0.001	0.001
Manufacturing and construction	-0.0003	7.58e-05	-3.361	0.001	-0.000	-0.000
Transport_y	0.0915	0.006	16.380	0.000	0.081	0.103
Electricity and heat	-0.0001	4.06e-05	-3.303	0.001	-0.000	-5.44e-05
Buildings	0.0006	0.000	5.249	0.000	0.000	0.001
Fugitive emissions	0.0004	5.71e-05	7.740	0.000	0.000	0.001
Other fuel combustion	0.0004	0.001	0.765	0.444	-0.001	0.002
Aviation and shipping	-3.848e-05	0.000	-0.179	0.858	-0.000	0.000
Renewables (% equivalent primary energy)	921.3666	473.462	1.946	0.052	-8.612	1851.346
Fossil fuels (% electricity)	12.0675	205.053	0.059	0.953	-390.699	414.834
Renewables (% electricity)	-675.0721	332.980	-2.027	0.043	-1329.114	-21.030

Omnibus:	417.162	**Durbin-Watson:**	0.531
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	37756.473
Skew:	2.397	**Prob(JB):**	0.00
Kurtosis:	42.373	**Cond. No.**	2.41e+08

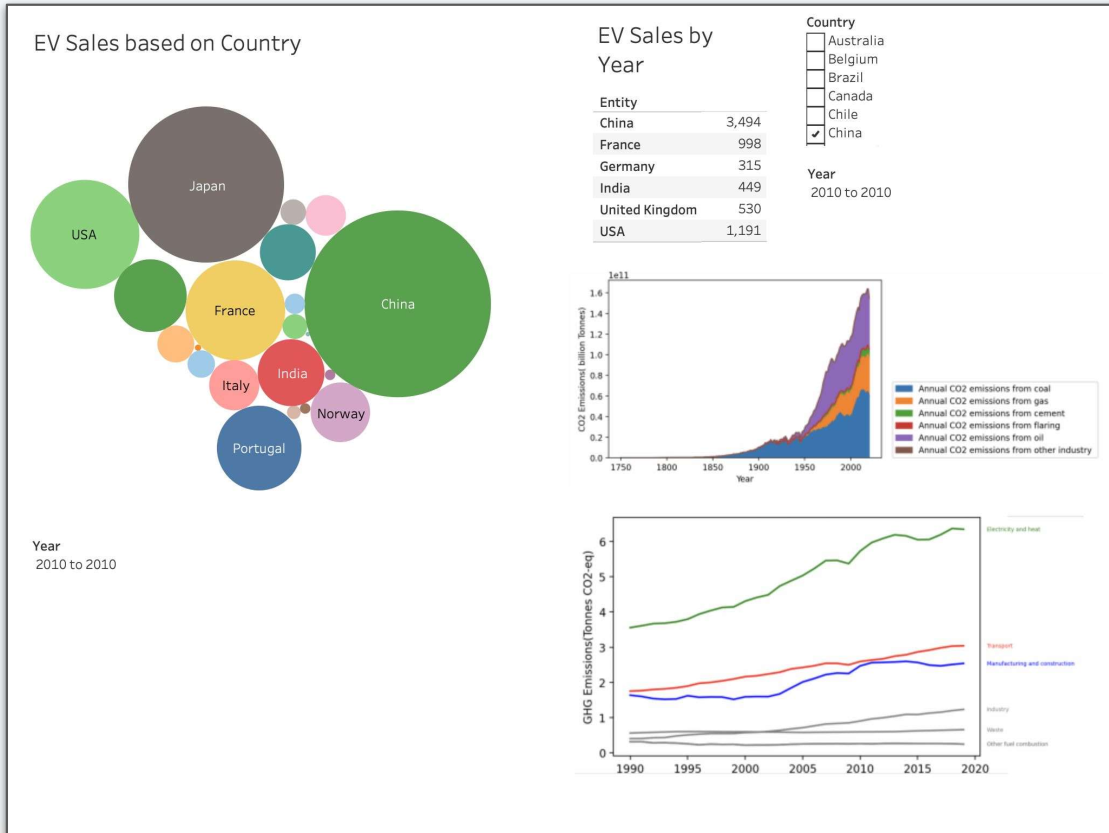
Forecasting/Predictions

- Although we do not have the y-values of the target variable(co2_offset/ev_savings) from 2015-19, we do have the input features(X values)

- We used the model to generate predictions and see that the CO2 offset due to EVs is predicted to be increasing year over year, as expected

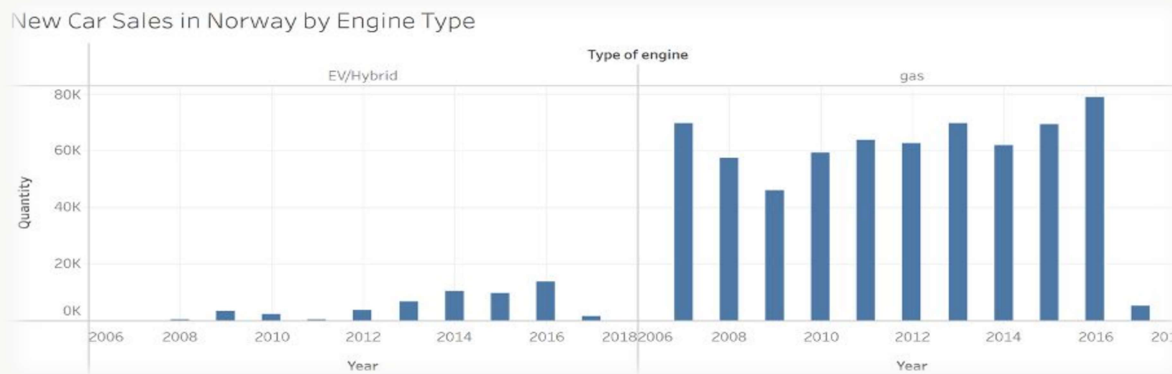


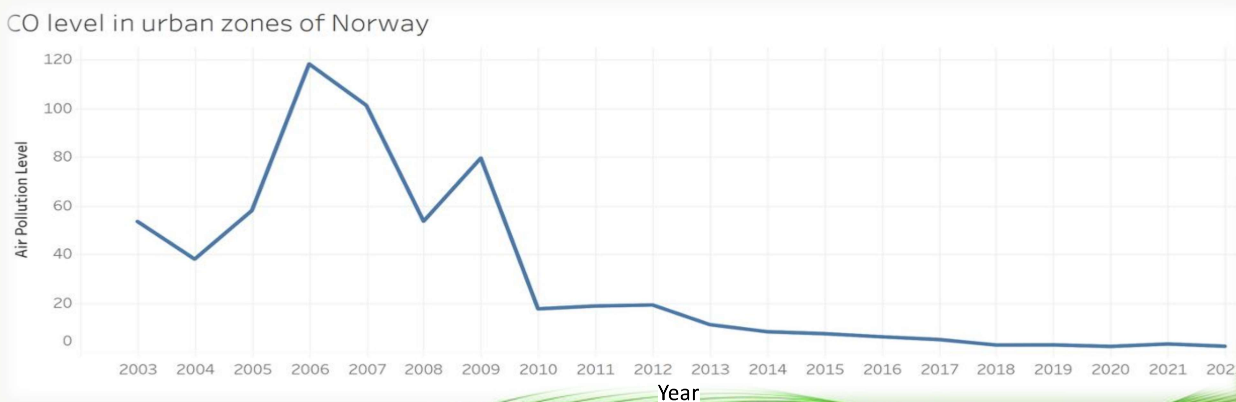
Dashboard: Tableau



Norway EV Sales: Real World Example

Norway has been pursuing the EV popularization policy such a reducing taxes and banning gas cars for some urban areas





Conclusion

- Contradicting theories around negating impact due to energy in recharging
- EV infrastructure is still at its nascent stage and evolving
- Charging your EV with renewable sources results in zero emissions for both when the car is operating and during power generation

Appendix

Database: Tables, Join, Connection String

```sql
CREATE TABLE IF NOT EXISTS public.temp_data
(
    "Country_year" character varying COLLATE pg_catalog."default" NOT NULL,
    "2000" double precision,
    "2001" double precision,
    "2002" double precision,
    "2003" double precision,
    "2004" double precision,
    "2005" double precision,
    "2006" double precision,
    "2007" double precision,
    "2008" double precision,
    "2009" double precision,
    "2010" double precision,
    "2011" double precision,
    "2012" double precision,
    "2013" double precision,
    CONSTRAINT temp_data_pkey PRIMARY KEY ("Country_year")
)

TABLESPACE pg_default;

ALTER TABLE IF EXISTS public.temp_data
    OWNER to postgres;
```

```python
import psycopg2
engine = psycopg2.connect(
    database="postgres",
    user="postgres",
    password="qt3WxeF5SzWtQV3",
    host="database-ev.cirez3rkweci.us-west-1.rds.amazonaws.com",
    port='5432'
)
```

- Tables (7)
  - Norway_co_urban
  - Norway_sales
  - car_sales
  - ev_data
  - merged
  - temp_data
  - temperature

Recommendations for Future Analysis

- Time permitting we would like to integrate the CO2 emission data with the EV sales data and show that on global map
- Studies at length the assumptions around future dominating country in EV demand
- Studied Tesla’s infrastructure to prove net impact of Carbon footprint due to it’s energy consumed in charging EVs

Questions?