

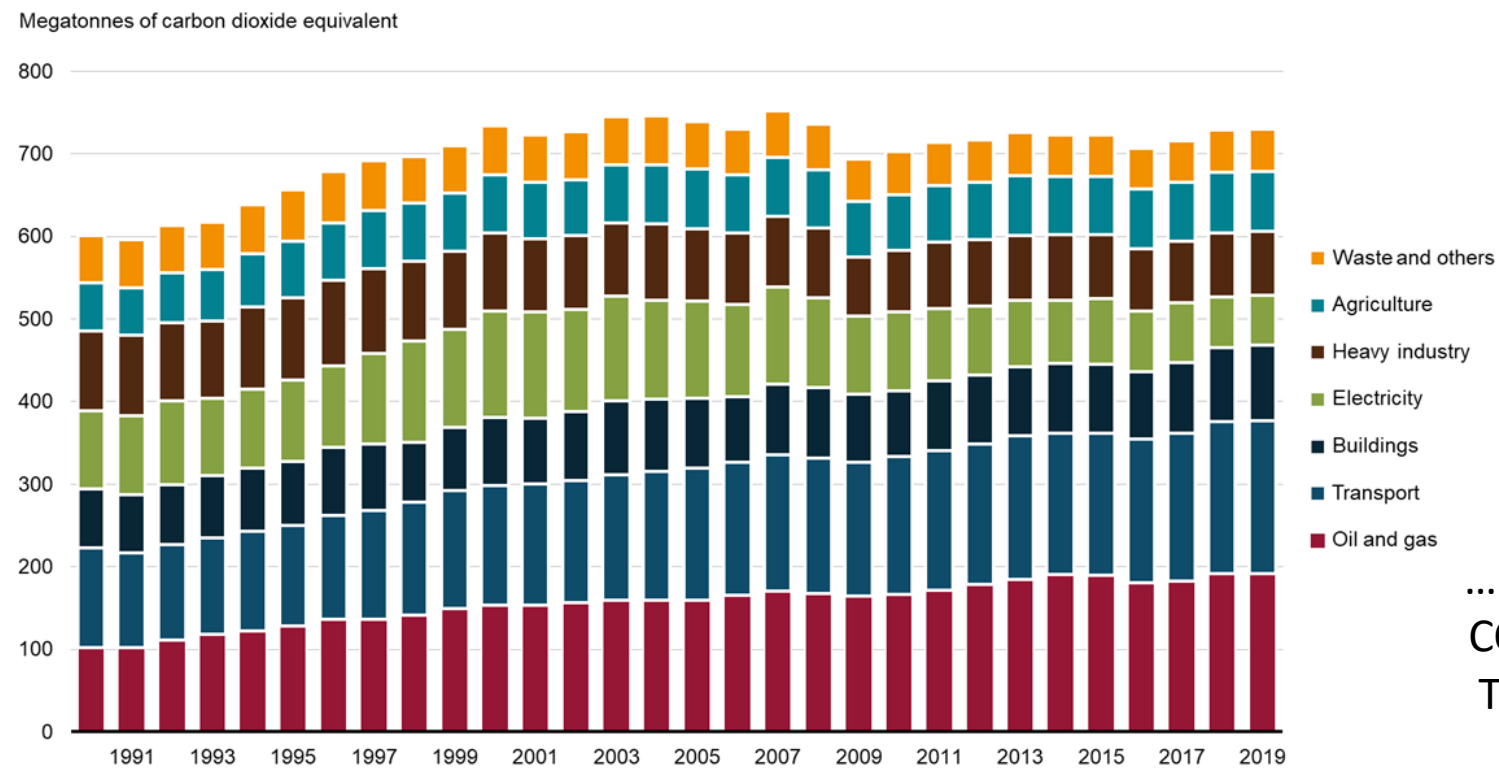
# CARBON GAS EMISSIONS PREDICTOR

# AGENDA

1. Business Problem / Data Problem
2. Stakeholder Identification
3. Process Workflow
4. Dataset Characteristics
5. Exploratory Data Analysis
6. Model Building & Selection
7. Model Deployment
8. Limitations

# BUSINESS PROBLEM / DATA PROBLEM

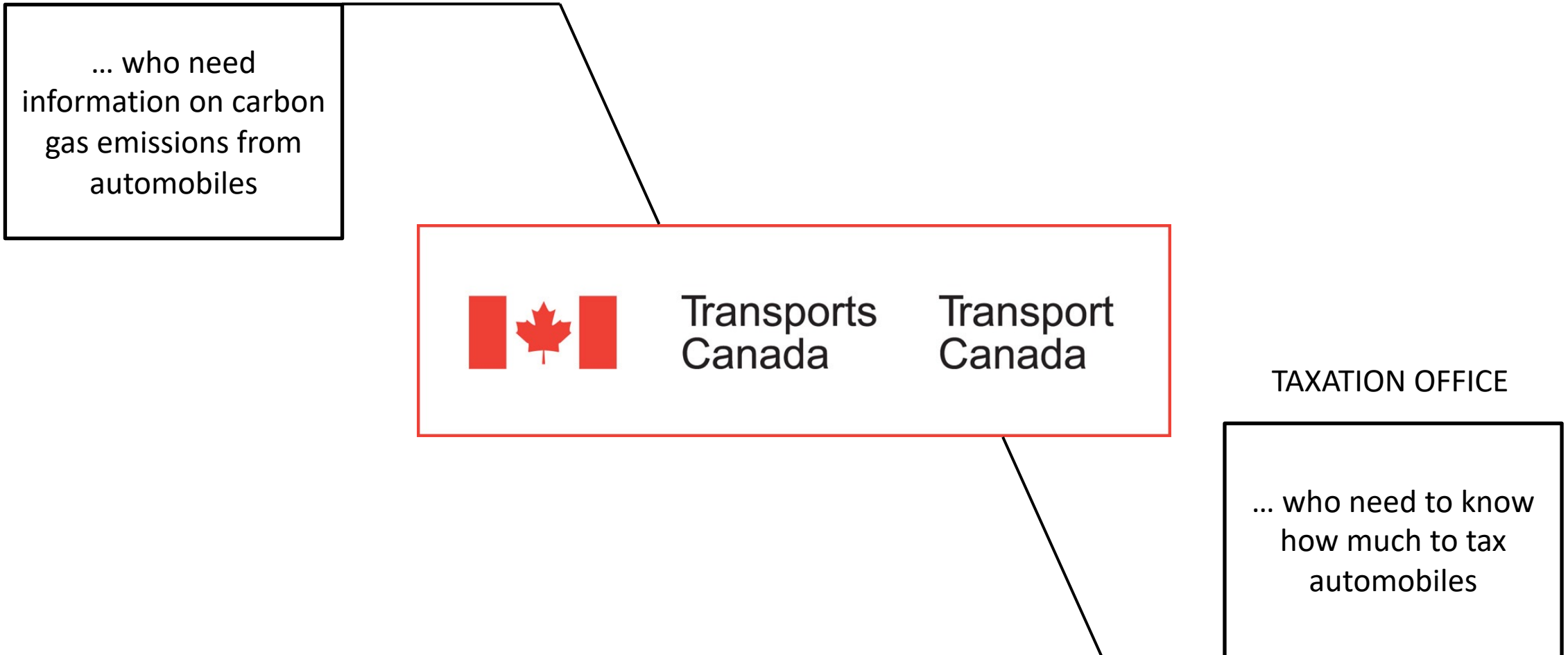
- As carbon emissions from cars make a significant proportion of carbon emissions, policy makers need a quick method to assess and forecast carbon gas emissions emitted by automobiles in order to properly tax and disincentivize consumers from owning certain vehicles
- Collecting primary historical data from a specific geographic location, as many characteristics can impact carbon emissions.



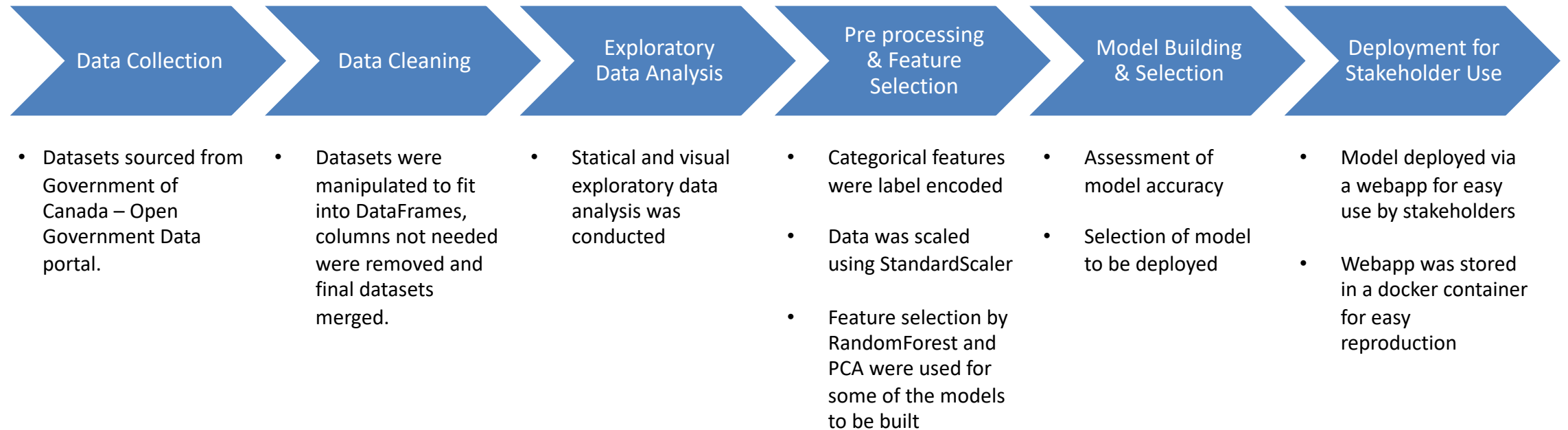
... Average of 25% of  
CO2 emissions from  
Transport Sector in  
recent years

# STAKEHOLDER IDENTIFICATION

## STRATEGY PLANNERS



# PROCESS WORKFLOW



# DATASET CHARACTERISTICS

## Dataset statistics

Number of variables	16
Number of observations	2891
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	361.5 KiB
Average record size in memory	128.0 B

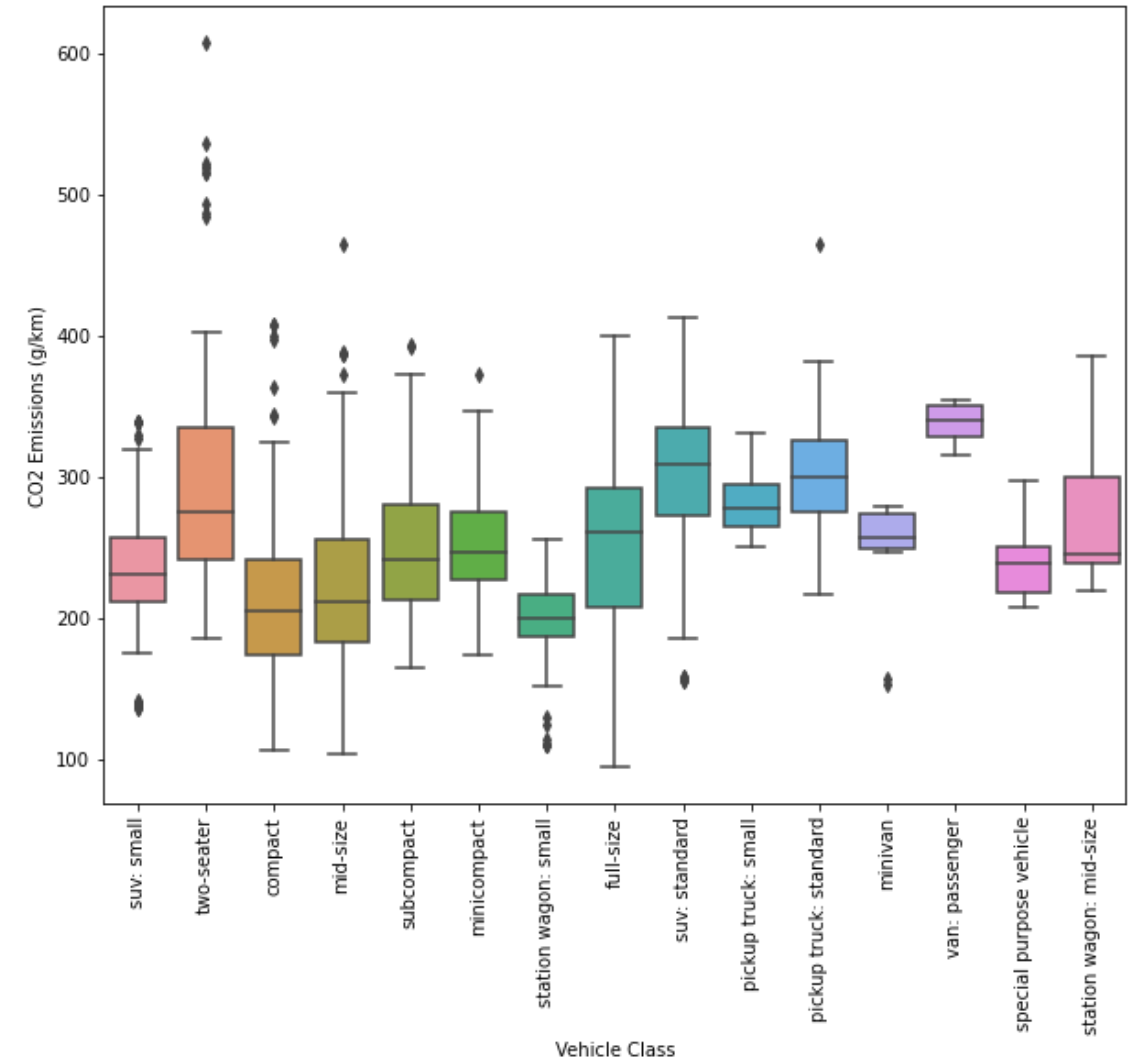
## Variable types

Numeric	9
Categorical	7

- Data downloaded from: <https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64#wb-auto-6>

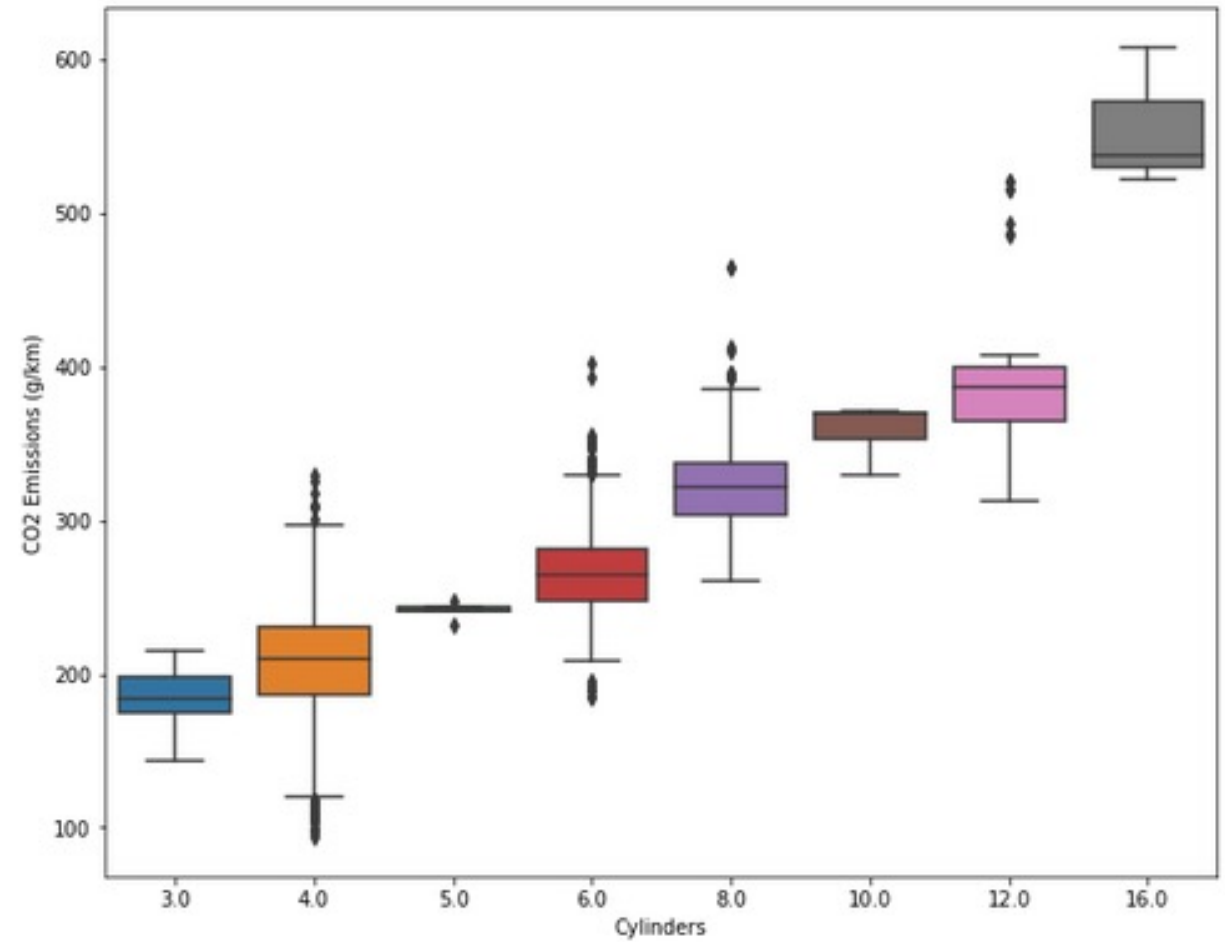
# EXPLORATORY DATA ANALYSIS

- Two-seater vehicles released highest amount of carbon gas emissions.
- Passenger vans had the highest mean carbon emissions released.



# EXPLORATORY DATA ANALYSIS

- Cars with higher number of cylinders released the most carbon emissions.





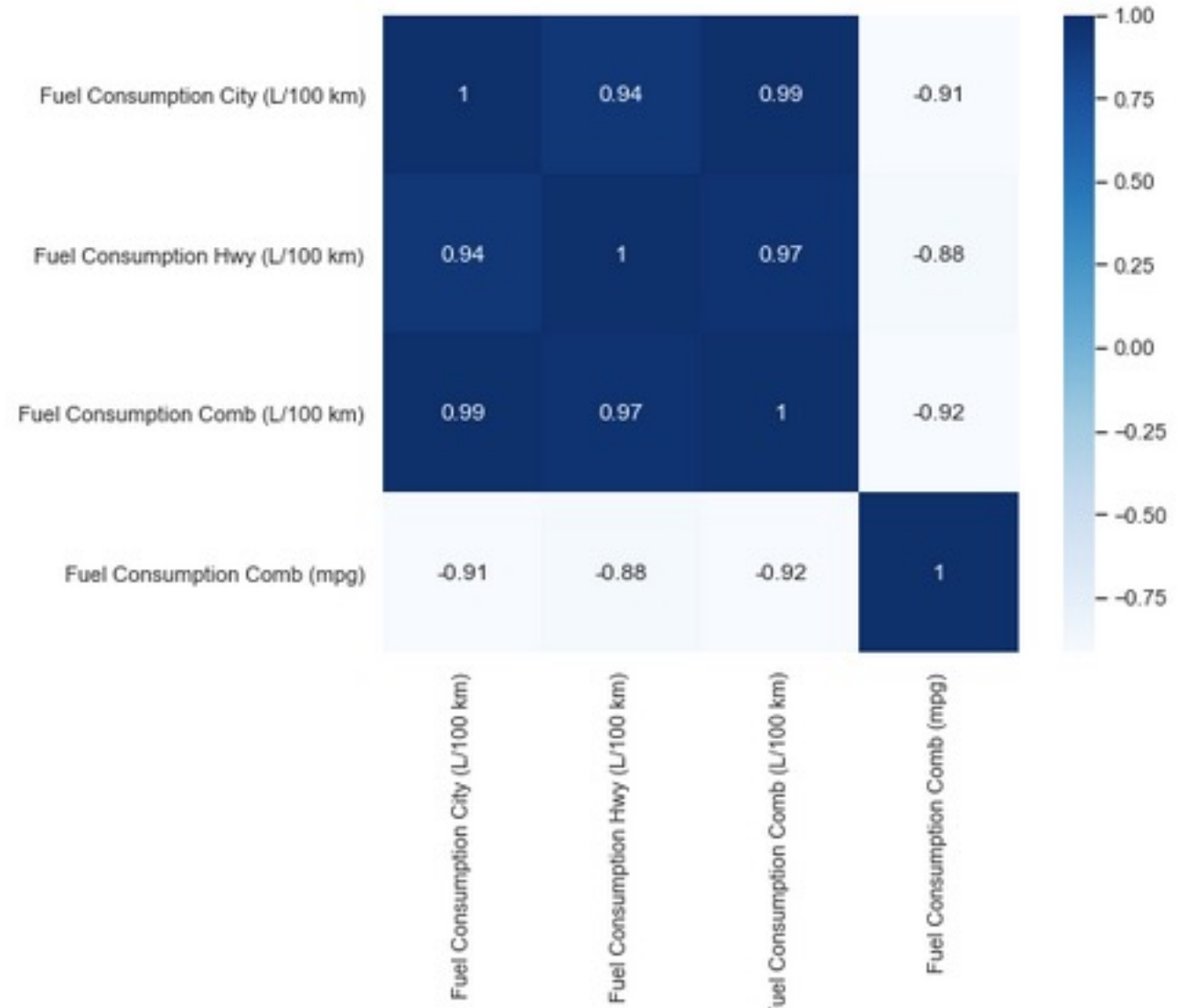
# EXPLORATORY DATA ANALYSIS – FEATURE CORRELATION

- Fuel consumption found to be most highly correlated with carbon emissions.

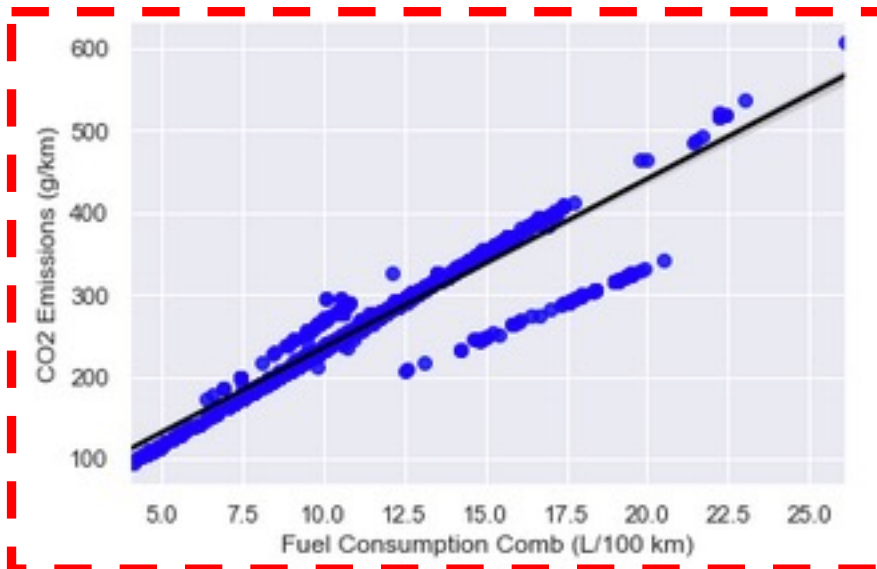
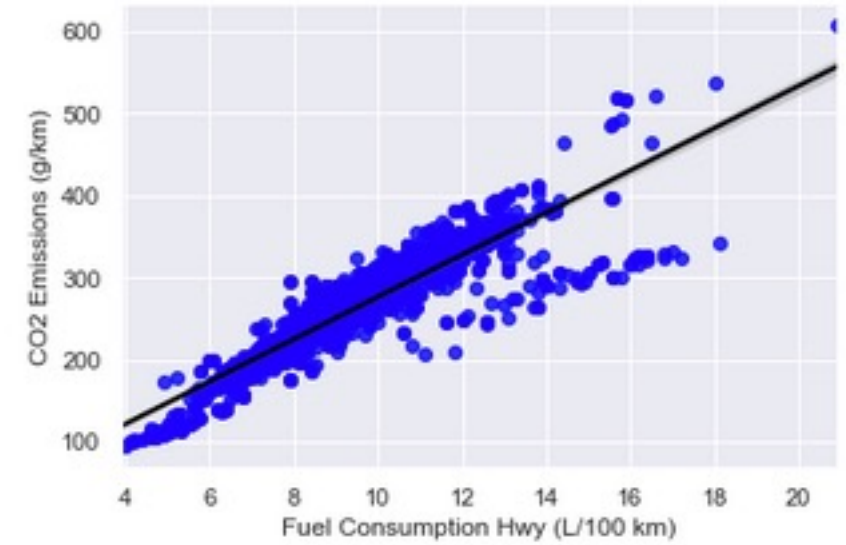
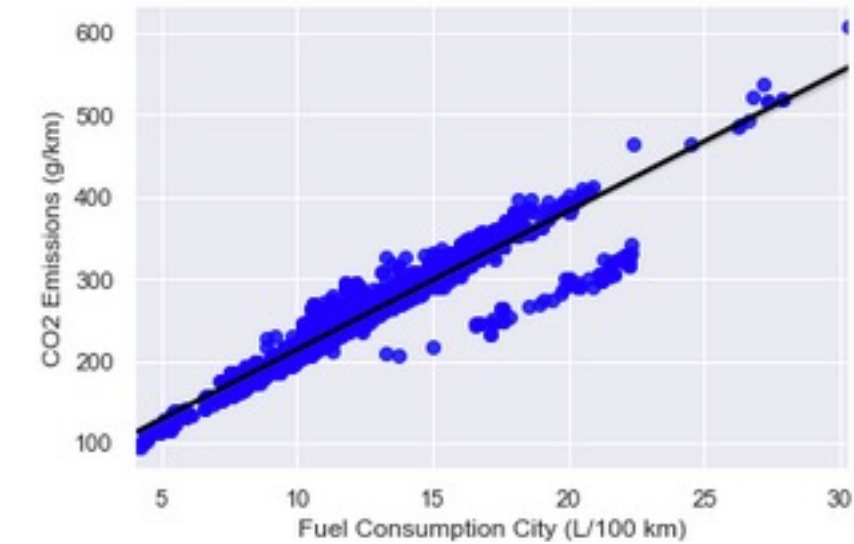
CO2 Emissions (g/km)	
CO2 Emissions (g/km)	1.000000
Fuel Consumption Comb (L/100 km)	0.951656
Fuel Consumption City (L/100 km)	0.948930
Fuel Consumption Hwy (L/100 km)	0.917122
Engine Size_L	0.835887
Cylinders	0.834904
Smog Rating	-0.514353
Fuel Consumption Comb (mpg)	-0.910652
CO2 Rating	-0.954422

# EXPLORATORY DATA ANALYSIS – FEATURE CORRELATION

- All 4 feature columns labelled “Fuel Consumption” are highly correlated.



# EXPLORATORY DATA ANALYSIS – FEATURE CORRELATION



# MODEL SELECTION

- Selected model can predict price with an **upper/lower bound of 16.71 CO2 g/km (7% of mean CO2 g/km of dataset)**

Model Type	Features	R2	RMSE (CO2 Emissions (g/km))
Simple LR (Not Scaled)	1	0.87	21.08
Simple LR (Scaled)	1	0.87	21.08
Multi LR (Features Selected by RandomForest)	3	0.90	17.08
Multi LR (All features – not scaled)	7	0.92	16.71

# RECOMMENDATIONS

---

- We propose a 4-tier carbon tax pricing system based on the quartiles of the dataset.
- An additional carbon tax for luxury cars with high number of cylinders.
- An additional carbon tax for passenger vans.

CO2 Emissions (g/km)	
count	2339.000000
mean	254.582300
std	60.485776
min	94.000000
25%	213.000000
50%	252.000000
75%	294.000000
max	608.000000

# MODEL DEPLOYMENT

---

- Our model was deployed onto a webapp that makes the predictions:  
<https://share.streamlit.io/adireksa/streamlit-linregapp-project2/main/app.py>
- The app was also stored into a docker container for easy reproducibility.

# IMPROVEMENTS FOR FUTURE

---

- Model deployed should be able to calculate carbon tax amount per vehicle.
- Expand dataset to include hybrid vehicles and electric vehicles.
- Compare other more advanced models with higher accuracy.

**Thank you**