



**SCHOOL OF ARTS AND SCIENCES**  
**DEPARTMENT OF COMPUTER, INFORMATION SCIENCES AND MATHEMATICS**

**Analyzing Vehicle Data for Efficiency and Environmental Impact**

A Project Proposal

Presented to the Faculty of the  
Department of Computer, Information Sciences and Mathematics  
University of San Carlos

In Partial Fulfillment  
of the Requirements for the Course  
**IT 3107 - FUNDAMENTALS OF DATA WAREHOUSING AND DATA MINING**

By  
CABOGOY, ZYGUEL PHILIP  
CENIZA, JADEN  
MENDOZA, JEAN RHAYE CHRISTOPHER  
MONSALES, DAN LIUS

**GLENN BULLECER PEPITO**

Instructor

## Table of Contents

List of Figures.....	iii
List of Tables.....	iii
1. Executive Summary.....	1
2. Business Questions.....	2
3. Business Terms and Definitions.....	3-4
4. Data Sources.....	5-7
5. Design.....	8
5.1. Detailed Data Modal.....	8-9
5.2. Fact Table Strategy.....	9-10
5.3. Dimension Load Strategy.....	11-12
5.4. Technical Metadata Mapping.....	12-15
5.5. Physical Capacity Plan.....	16
5.6. Technology and Tools.....	16-17
6. Data Analysis.....	18
6.1. Overview.....	18
6.2. Key Reports and Findings.....	18-25
6.2.1: Average Fuel Cost and CO <sub>2</sub> Emissions.....	18-19
6.2.2: Highest Average Combined MPG per Manufacturers.....	19-21
6.2.3: Average Fuel Price by Fuel Type.....	21-22
6.2.4: Average Emission Scores of Manufacturers.....	22-24
6.2.5: Average MPG by Vehicle Class.....	24-26
6.2.6: Vehicle Class Ranking by Percentage Reduction in CO <sub>2</sub> Emissions...	25-27
7. Conclusion.....	28

## **List of Figures**

Figure 5.1.1: Detailed Data Model.....	9
Figure 6.2.1: Average Fuel Cost and CO <sub>2</sub> Emissions.....	18
Figure 6.2.2: Highest Average Combined MPG per Manufacturers.....	20
Figure 6.2.3: Average Fuel Price by Fuel Type.....	21
Figure 6.2.4: Average Emission Scores of Manufacturers.....	23
Figure 6.2.5: Average MPG by Vehicle Class.....	24
Figure 6.2.6: Vehicle Class Ranking by Percentage Reduction in CO <sub>2</sub> Emissions....	26

## **List of Tables**

Table 4.1.1: Vehicle Dataset Columns and Descriptions.....	5-6
Table 4.1.2: Emissions Dataset Columns and Descriptions.....	7
Table 4.1.3: Diesel Dataset Columns and Descriptions.....	7
Table 4.1.4: Gasoline Dataset Columns and Descriptions.....	7
Table 5.5.1: Physical Capacity Plan.....	16

# **Chapter 1**

## **Introduction**

### **Executive Summary**

This study examines the factors influencing the evolution of vehicle fuel economy, emissions, and fuel prices over the past decade. By analyzing a range of data from reputable sources, the research aims to understand how various technological, economic, and regulatory factors have impacted fuel efficiency and environmental compliance in the automotive sector.

The study integrates datasets covering key aspects such as vehicle specifications, fuel efficiency ratings, emissions standards, and historical fuel prices for both diesel and gasoline. Through this analysis, the study seeks to identify trends in fuel economy, technological advancements, and regulatory impacts, while also exploring the broader implications for vehicle design, consumer behavior, and sustainability.

Using a structured data model, the research examines how changes in fuel costs, fuel types, emissions standards, and vehicle characteristics have shaped the automotive landscape over time. By connecting these data points, the study provides insights into the complex relationships between vehicle performance, environmental standards, and economic factors.

The findings of this study are intended to inform decision-making by stakeholders across the automotive and policy sectors. By highlighting the key drivers of fuel economy improvements and emissions reductions, this research contributes to the ongoing conversation around sustainability and innovation in transportation.

## **Chapter 2**

### **Business Questions**

This section presents key business questions that leverage the data warehouse to analyze vehicle fuel efficiency, emissions, and trends. These questions guide stakeholders in making informed decisions on sustainability and innovation.

#### **2.1 Standard Report:**

- What is the average fuel cost and CO<sub>2</sub> emissions across all vehicles in the datasets by year?

#### **2.2 Report with Aggregation (Windowing):**

- Which manufacturers have the highest average fuel economy (Combined MPG) across their vehicle lineup, and how does this compare with their competitors?

#### **2.3 Report with Analytical (Windowing):**

- How does the average fuel cost of each fuel type (e.g., Regular, Diesel) rank over time, and what trends can we observe?

#### **2.4 Report Showing Hierarchical Results:**

- What is the distribution of vehicle types (e.g., Two-Seaters, SUVs) across manufacturers and their average emissions?

#### **2.5 Additional Business Questions**

- How has the average fuel economy (Combined MPG) for each vehicle class changed over time across all manufacturers?
- Which vehicle classes have shown the greatest improvement in reducing CO<sub>2</sub> emissions per mile over the last 5 years?

#### **2.6 Additional Data Source**

- How does the total cost of maintenance compare across different vehicle makes and models with similar fuel efficiency and emissions?

## Chapter 3

### Business Terms and Definition

Below are the key business terms and definitions relevant to this project:

- **Fuel Economy:** A measure of how efficiently a vehicle uses fuel, typically expressed in miles per gallon (MPG) for gasoline and diesel vehicles or as energy consumption per mile for electric vehicles.
- **Vehicle Class:** Categories of vehicles based on size and design, such as compact cars, midsize cars, SUVs, trucks, and vans. Each class often correlates with different fuel economy and emissions profiles.
- **Miles Per Gallon (MPG):** The distance a vehicle can travel using one gallon of fuel, often broken down into city MPG, highway MPG, and combined MPG to reflect different driving conditions.
- **CO<sub>2</sub> Emissions:** The amount of carbon dioxide released by a vehicle per mile, measured in grams. This metric reflects the environmental impact of fuel consumption.
- **Emission Standards:** Regulatory benchmarks that define allowable limits for pollutants released by vehicles, including CO<sub>2</sub> and smog-forming compounds. Examples include EPA Tier 3 and Euro 6 standards.
- **Fuel Type:** The type of fuel a vehicle uses, such as gasoline, diesel, or electricity. This influences both the vehicle's efficiency and its environmental impact.
- **Smog Rating:** A standardized scale (e.g., 1-10) that measures a vehicle's contribution to smog formation, with higher scores indicating lower emissions.
- **Hybrid Vehicles:** Vehicles that use a combination of an internal combustion engine and an electric motor to improve fuel efficiency and reduce emissions.
- **Electric Vehicles (EVs):** Vehicles powered entirely by electricity, offering zero tailpipe emissions and higher energy efficiency compared to conventional vehicles.
- **Fuel Cost:** The monetary cost associated with operating a vehicle, calculated based on the price of fuel and the vehicle's fuel economy.
- **Vehicle Manufacturer:** The company that produces the vehicle, such as Tesla, Toyota, or Ford. Manufacturers are often analyzed for their average fuel economy and emissions across all models.

- **Vehicle Specifications:** Technical details of a vehicle, including engine displacement, number of cylinders, drivetrain type, transmission type, and fuel type. These attributes influence both fuel efficiency and environmental performance.
- **Time Period:** The range of years analyzed in the study, spanning from 2014 to 2024. This temporal dimension is used to track trends in fuel economy, emissions, and fuel prices.
- **Fuel Price Trends:** Historical and current data on fuel prices, including variations by type (e.g., regular, diesel, premium gasoline) and over time.
- **Average MPG:** A calculated measure that combines city and highway MPG to provide a comprehensive view of a vehicle's overall fuel efficiency.
- **Effective Date:** The start date for a record in the data model, used in Type II Slowly Changing Dimensions to track changes over time.
- **End Date:** The end date for a record in the data model, indicating when a particular version of the data became obsolete due to updates or modifications.
- **Emission Score:** A comparative rating that indicates a vehicle's emissions performance relative to other vehicles.

## Chapter 4

### Data Sources

The data employed in this study is sourced from two reputable U.S. Government web services. The **Vehicle and Emissions Datasets** are derived from the **Fuel Economy Web Service**, managed by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy. These datasets provide detailed records of vehicles across the United States, including emissions data, spanning various years. The datasets were meticulously cleaned and curated to meet the objectives of this study, focusing on fuel efficiency, emissions standards, and related vehicle metrics.

- <https://www.fueleconomy.gov/feg/ws/index.shtml>

Additionally, the **Diesel and Gasoline Datasets** are sourced from the **EIA Web Service**, maintained by the U.S. Energy Information Administration. These datasets include historical retail price records for diesel and various grades of gasoline, supporting analyses of fuel price trends over time. The integration of these datasets ensures a comprehensive approach to understanding fuel economy and emissions in the context of fluctuating energy markets.

- <https://www.eia.gov>

### Dataset Columns and Descriptions

The column description of all the datasets used are detailed as follows:

Column Name	Description
barrels08	Annual petroleum consumption in barrels for fuelType1
city08	City MPG for fuelType1
co2TailpipeGpm	Tailpipe CO2 emissions in grams/mile for fuelType1
comb08	Combined MPG for fuelType1
cylinders	Number of engine cylinders
displ	Engine displacement in liters
drive	Drive axle type

engId	EPA model type index
eng_dscr	Engine descriptor
fuelCost08	Annual fuel cost for fuelType1
fuelType	Fuel type with fuelType1 and fuelType2 (if applicable)
fuelType1	Primary fuel type for single or dual-fuel vehicles
highway08	Highway MPG for fuelType1
id	Vehicle record ID
make	Manufacturer (division)
model	Model name (carline)
mpgData	Indicates if "My MPG" data is available
phevBlended	Indicates if the vehicle operates on a gasoline-electricity blend in charge-depleting mode
trany	Transmission type
UCity	Unadjusted city MPG for fuelType1
UHighway	Unadjusted highway MPG for fuelType1
VClass	EPA vehicle size class
year	Model year
youSaveSpend	Cost savings (+) or additional spending (-) over 5 years compared to an average car
baseModel	Base model name
createdOn	Date the vehicle record was created
modifiedOn	Date the vehicle record was last modified
startStop	Indicates if the vehicle has stop-start technology
record_created	Date the record was created

*(Table 4.1.1: Vehicle Dataset Columns and Descriptions)*

Column Name	Description
efid	Engine family ID
id	Vehicle record ID (links to vehicle dataset)
salesArea	EPA sales area code
score	EPA 1-10 smog rating for fuelType1
standard	Vehicle Emission Standard Code
stdText	Vehicle Emission Standard

*(Table 4.1.2: Emissions Dataset Columns and Descriptions)*

Column Name	Description
Day	Date of diesel price record
Price	Diesel price (\$/gal)

*(Table 4.1.3: Diesel Dataset Columns and Descriptions)*

Column Name	Description
Day	Date of gasoline price record
U.S. Midgrade All Formulations Retail Gasoline Prices \$/gal	Retail gasoline price for midgrade formulations
U.S. Premium All Formulations Retail Gasoline Prices \$/gal	Retail gasoline price for premium formulations
U.S. Regular All Formulations Retail Gasoline Prices \$/gal	Retail gasoline price for regular formulations

*(Table 4.1.4: Gasoline Dataset Columns and Descriptions)*

## Chapter 5

### Design

#### 5.1. Detailed Data Model

Below is the detailed data model for the Vehicle Fuel Economy analysis project. It consists of one fact table and four dimension tables focused on tracking and analyzing vehicle fuel efficiency, emissions, and environmental impact between 2014-2024.

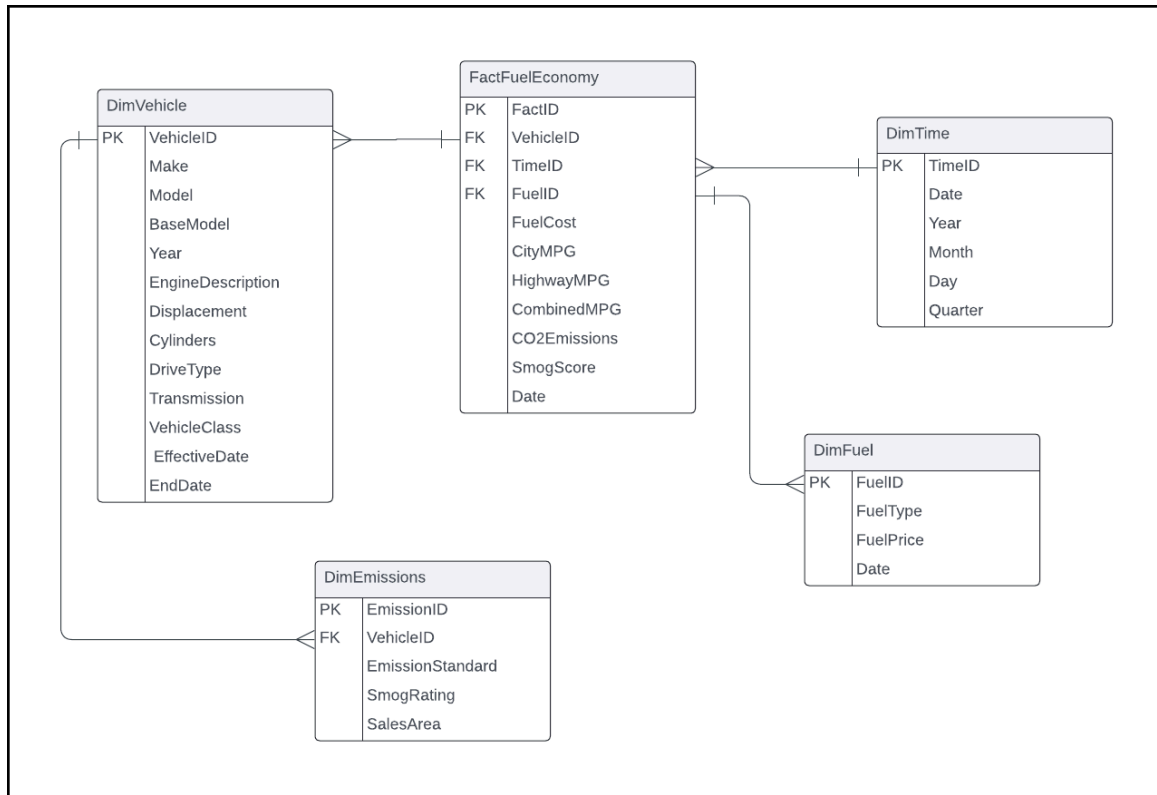
- **FactFuelEconomy**

- The central fact table tracking vehicle fuel economy metrics, containing measures for fuel costs, efficiency ratings, and environmental impact scores. It connects to all dimension tables through foreign keys (VehicleID, TimeID, and FuelID) to enable comprehensive analysis.

- **Dimension Tables:**

- **DimVehicle:** Contains detailed vehicle specifications including Make, Model, BaseModel, and technical details such as EngineDescription, Displacement, Cylinders, DriveType, and Transmission. Tracks changes through EffectiveDate and EndDate fields.
- **DimTime:** Manages temporal aspects of the analysis with TimeID as the primary key, containing Date, Year, Month, Day, and Quarter fields for time-based analysis within the 2014-2024 range.
- **DimFuel:** Tracks fuel-related information including FuelID (primary key), FuelType, FuelPrice, and Date, enabling analysis of fuel costs and types over time.
- **DimEmissions:** Records environmental compliance and impact metrics with EmissionID as primary key, containing EmissionStandard, SmogRating, and SalesArea fields, linked to vehicles through VehicleID.

This dimensional model enables comprehensive analysis of vehicle fuel efficiency, costs, and environmental impact across various dimensions while maintaining historical accuracy through appropriate tracking of changes over the specified time period.



(Figure 5.1.1: Detailed Data Modal)

## 5.2 Fact Table Strategy

The data warehouse for this project has one primary fact table:

- **FactFuelEconomy**
  - **Granularity:** This fact table's granularity is at the vehicle-fuel-time level, capturing measurements of fuel economy, emissions, and costs for each vehicle with specific fuel type at a given point in time between 2014 and 2024.
- **Dimensions and Relationships:**
  - **DimVehicle:** Creates a relationship between vehicle characteristics and performance data for vehicles manufactured between 2014-2024. This dimension contains the VehicleID as primary key, along with Make,

Model, BaseModel attributes. It includes engine specifications through EngineDescription, Displacement, and Cylinders fields. Additional vehicle attributes include Transmission, DriveType, and VehicleClass. The dimension tracks changes through EffectiveDate and EndDate fields, all constrained within the 2014-2024 timeframe.

- **DimTime:** Creates a relationship for temporal analysis within the 2014-2024 range. This dimension contains the TimeID as primary key, and includes Date, Year (filtered to 2014-2024), Month, Day, and Quarter fields for various levels of time-based analysis.
- **DimFuel:** Creates a relationship for fuel-specific analysis during the 2014-2024 period. This dimension contains the FuelID as primary key, and includes FuelType, FuelPrice, and Date fields, with all dates falling within the 2014-2024 range.
- **DimEmissions:** Creates a relationship for emissions and environmental metrics for vehicles from 2014-2024. This dimension contains the EmissionID as primary key, and includes EmissionStandard, SmogRating, and SalesArea fields, with a foreign key relationship to VehicleID.
- **Facts and Measures:**
  - **FuelCost:** The cost of fuel per vehicle, providing insights into operational expenses and efficiency from a financial perspective.
  - **CityMPG and HighwayMPG:** Individual measurements for city and highway fuel efficiency, allowing for specific analysis of vehicle performance in different driving conditions.
  - **CombinedMPG:** A weighted average of city and highway fuel efficiency, providing a comprehensive measure of the vehicle's overall fuel economy.
  - **CO2Emissions:** Measurement of carbon dioxide emissions, quantifying the environmental impact of each vehicle.
  - **SmogScore:** A standardized rating indicating the vehicle's contribution to smog formation, helping assess environmental compliance and impact.
- **Fact Table Type:**
  - The FactFuelEconomy is a **Periodic Snapshot Fact Table**, capturing fuel economy metrics, costs, and environmental measurements at regular time intervals between 2014-2024. This design enables trend analysis and performance tracking over time, with each record representing a specific measurement point for a vehicle-fuel combination.

### 5.3 Dimensional Load Strategy

Deciding on a dimension load strategy is crucial for answering the business questions for this project. Before starting the implementation of the dimension tables, it is essential to have a clear idea about:

- Columns that should never change
- Columns needed for merging different fact tables
- Columns needed to keep track of the date and time of certain entries

The loading strategy of our dimension tables is based on the above three points. For this project, we have implemented the following strategies:

#### SCD 1: Slowly Changing Dimension Type I

- **DimTime:** Implements SCD Type 1 as time attributes are static and standardized. The TimeID, Date, Year, Month, Day, and Quarter fields will never change once loaded.
- **DimFuel:** Uses SCD Type 1 for FuelType and FuelPrice attributes. Any changes in fuel prices or types will overwrite existing values, as historical fuel price tracking is handled in the fact table.

#### SCD 2: Slowly Changing Dimension Type II

- **DimVehicle:** Implements SCD Type 2 to maintain historical records of vehicle specifications. Changes to Make, Model, BaseModel, EngineDescription, and other vehicle attributes are tracked using EffectiveDate and EndDate fields, preserving the history of vehicle modifications and updates between 2014-2024.
- **DimEmissions:** Uses SCD Type 2 to track changes in emission standards and ratings over time. The EmissionStandard, SmogRating, and SalesArea attributes maintain historical records as environmental regulations and ratings may change.

#### Assumptions

- Time dimension values are static and won't require historical tracking.
- Fuel types and current prices don't need historical tracking in the dimension (price history is maintained in facts).

- Vehicle specifications and emission standards require historical tracking due to potential changes in regulations and vehicle modifications between 2014-2024.
- The **EffectiveDate** and **EndDate** in DimVehicle ensure accurate historical analysis of vehicle changes over the specified time period.

## 5.4 Technical Metadata Mapping

We have created a source database before the implementation of the Data warehouse for Vehicle Fuel Economy data for the following reasons:

- **Vehicle Data** was provided through a CSV file and separated by columns. Data was filtered to include only vehicles manufactured between 2014-2024. Make, Model, BaseModel, and technical specifications were standardized. Engine specifications were validated to ensure consistency in Displacement and Cylinder values. Each vehicle has a unique VehicleID with associated EffectiveDate and EndDate for tracking changes.
- **Time Data** was structured to support analysis between 2014-2024, with standardized date formats. The data includes TimeID as a primary key, with derived fields for Year, Month, Day, and Quarter. Each time entry is unique and formatted consistently for accurate temporal analysis.
- **Fuel Data** was sourced from standardized fuel price records, containing FuelID, FuelType, and FuelPrice information. Price data was normalized and validated for consistency. Date fields were standardized to match the 2014-2024 analysis period.
- **Emissions Data** was compiled from environmental compliance records, with standardized EmissionStandard classifications and SmogRating scales. SalesArea data was validated against geographic standards. Each emission record is linked to a specific vehicle through VehicleID.

### Data Cleaning and Processing:

1. **Initial Inspection:** The dataset was inspected to understand its structure and content. Columns with all values as 0, -1, or NaN were replaced and dropped.

2. **Filtering Data by Year:** Data was filtered to focus on vehicles manufactured between 2014 and 2024, aligning with the project's timeframe.
3. **Handling Missing Values:** Columns with less than 60% filled values were removed to ensure data quality.
4. **Dropping Irrelevant Columns:** Specific columns such as co2, co2A, eng\_dscr, evMotor, and mfrCode were dropped due to irrelevance or insufficient data.
5. **Feature Engineering:** An average\_mpg column was created to calculate the mean of city and highway MPG, providing a comprehensive measure of fuel efficiency.
6. **Final Cleaning Steps:** Remaining missing values in essential columns were addressed, and duplicates were removed to ensure data uniqueness and integrity.

The metadata mapping for each column and its corresponding table is explained in detail in the mapping document. The following sections outline the column mapping and transformation applied:

#### **Mapping Document Overview:**

- **DimVehicle**
  - **VehicleID:** Integer, Primary Key. Unique identifier for each vehicle.
  - **Make:** Name of the vehicle manufacturer (e.g., Toyota, Honda).
  - **Model:** Name of the vehicle model (e.g., Corolla, Civic).
  - **BaseModel:** Specific variant of the model (e.g., LE, SE).
  - **Year:** Integer, year of manufacture.
  - **EngineDescription:** Text description of the engine specifications.
  - **Displacement:** Float, engine displacement in liters.
  - **Cylinders:** Integer, number of engine cylinders.
  - **DriveType:** Type of drivetrain (e.g., front-wheel drive, all-wheel drive).

- **Transmission:** Type of transmission system (e.g., automatic, manual).
  - **VehicleClass:** Type of vehicle (e.g., sedan, SUV, truck).
  - **EffectiveDate:** Date, start date for the record (Type II dimension).
  - **EndDate:** Date, end date for the record (Type II dimension).
- **DimFuel**
    - **FuelID:** Integer, Primary Key. Unique identifier for each fuel type.
    - **FuelType:** Text, type of fuel (e.g., gasoline, diesel, electric).
    - **FuelPrice:** Float, price of the fuel in dollars per gallon.
    - **Date:** Date, the date when the fuel price was recorded.
- **DimTime**
    - **TimeID:** Integer, Primary Key. Unique identifier for each time period.
    - **Date:** Date, specific calendar date.
    - **Year:** Integer, year of the record.
    - **Month:** Integer, month of the record.
    - **Day:** Integer, day of the record.
    - **Quarter:** Integer, quarter of the year (1 to 4).
- **DimEmissions**
    - **EmissionID:** Integer, Primary Key. Unique identifier for each emission record.
    - **VehicleID:** Integer, Foreign Key to DimVehicle.
    - **EmissionStandard:** Text, emission standard compliance level (e.g., Euro 6, EPA Tier 3).

- **SmogRating:** Integer, smog score indicating environmental impact.
  - **SalesArea:** Text, geographic area where the emissions data applies.
- **FactFuelEconomy**
    - **FactID:** Integer, Primary Key. Unique identifier for each fact record.
    - **VehicleID:** Integer, Foreign Key to DimVehicle.
    - **TimeID:** Integer, Foreign Key to DimTime.
    - **FuelID:** Integer, Foreign Key to DimFuel.
    - **FuelCost:** Float, total cost of fuel consumed.
    - **CityMPG:** Float, miles per gallon in city driving conditions.
    - **HighwayMPG:** Float, miles per gallon on highways.
    - **CombinedMPG:** Float, combined miles per gallon (city and highway).
    - **CO2Emissions:** Float, tailpipe CO2 emissions in grams per mile.
    - **SmogScore:** Integer, smog score indicating environmental impact.
    - **Date:** Date, the date associated with the record.

The steps mentioned above are also referred to as the Data Cleaning and Data Processing steps as we started loading the data into the proposed data warehouse. This proposed document is submitted, along with the attached appendices, as a deliverable.

## 5.5 Physical Capacity Plan

Sr. No.	Table Name	Table Type	Row Count (Approx)	Physical Size (Approx)
1	DimVehicle	Dimension	48231	10.00 MB
2	DimFuel	Dimension	263	0.05 MB
3	DimTime	Dimension	48231	3.0 MB
4	DimEmissions	Dimension	7860	3.00 MB
5	FactFuelEconomy	Fact	48231	3.00 MB
	<b>Total</b>			19.05 MB

*(Table 5.5.1: Physical Capacity Plan)*

This approximation was made using the file sizes generated by the python script. The physical sizes and row counts are based on the filtered data from the cleaned\_vehicle.csv, cleaned\_emissions.csv, diesel\_unified.csv, gasoline\_unified.csv files.

## 5.6 Technology and Tools

These are the following technologies and tools that were used to achieve the desired results in our data warehouse project:

- **PyCharm Professional Edition:** An integrated development environment (IDE) used extensively for Python development. It offers advanced coding features, database tools, and integration capabilities which facilitated efficient data manipulation and analysis throughout the project.
- **Anaconda:** A distribution of Python and R for scientific computing. It simplifies package management and deployment, ensuring that all necessary libraries and dependencies are installed and compatible. Anaconda was crucial for setting up the development environment.
- **Jupyter Notebook:** An open-source web application that allows creating and sharing documents containing live code, equations, visualizations, and narrative text. It was used for exploratory data analysis, data cleaning, and visualizations.

- **SQL Server 2019 (Dev Edition):** A relational database management system used for storing, querying, and managing the data warehouse. It provided robust data integration, advanced analytics, and built-in security features necessary for handling large datasets.
- **SQL Server Management Studio (SSMS):** A software application for monitoring, querying, and managing databases. It was used for tasks such as database configuration, querying, performance management, security management, backup and restore, and integration with other SQL Server functions.
- **Lucidchart:** A cloud-based application used for creating and visualizing charts and diagrams. It was instrumental in creating the detailed data model and visualizing the relationships between tables in the star schema.
- **Python Libraries:**
  - **Pandas:** An open-source data analysis and manipulation library, used primarily for cleaning up the collected datasets, handling duplicate and missing data, and transforming data for analysis.
  - **NumPy:** A fundamental package for scientific computing with Python, used for numerical operations and handling arrays.
  - **Matplotlib:** A plotting library used for creating static, interactive, and animated visualizations.
  - **Seaborn:** A Python visualization library based on Matplotlib, used for making statistical graphics.
  - **OS:** A standard Python library used for interacting with the operating system, such as handling file paths and sizes.

This combination of tools and technologies ensured that the data warehouse project was efficient, reliable, and capable of meeting the objectives of analyzing vehicle fuel efficiency trends from 2014 to 2024.

## Chapter 6

### Data Analysis

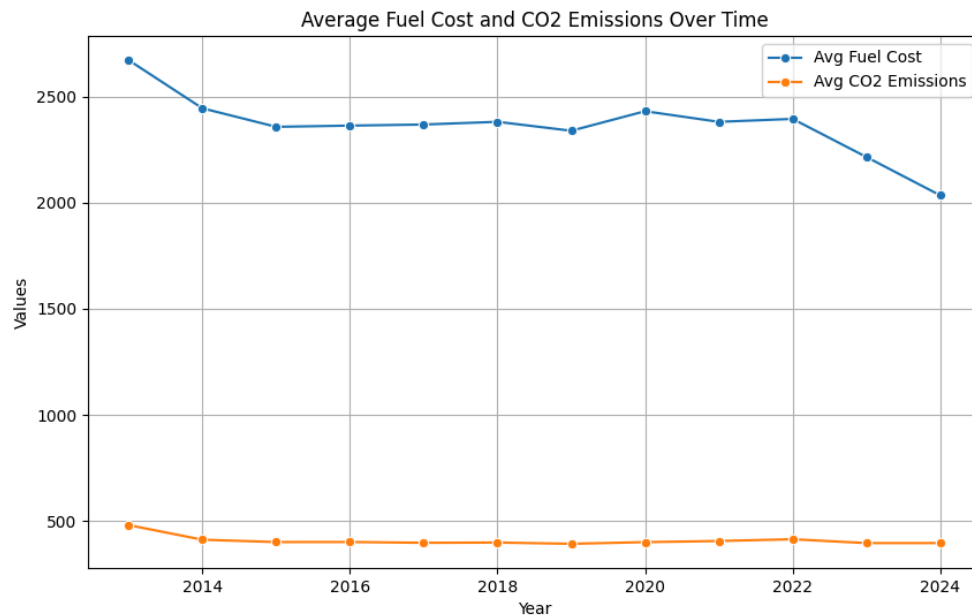
#### 6.1 Overview

This chapter focuses on the analysis of key metrics such as fuel costs, CO<sub>2</sub> emissions, and fuel efficiency trends derived from the data warehouse. The analysis evaluates the performance of vehicles across various classes, manufacturers, and fuel types over time, spanning the years 2014 to 2024. Insights are generated through reports and trends to understand how technological advancements, regulatory standards, and consumer preferences have influenced fuel economy and emissions. The findings provide valuable information to stakeholders, enabling informed decisions to promote sustainability and innovation in the automotive industry.

#### 6.2 Key Reports and Findings

##### 6.2.1 Average Fuel Cost and CO<sub>2</sub> Emissions by Year:

- What is the average fuel cost and CO<sub>2</sub> emissions across all vehicles in the datasets by year?



(Figure 6.2.1: Average Fuel Cost and CO<sub>2</sub> Emissions)

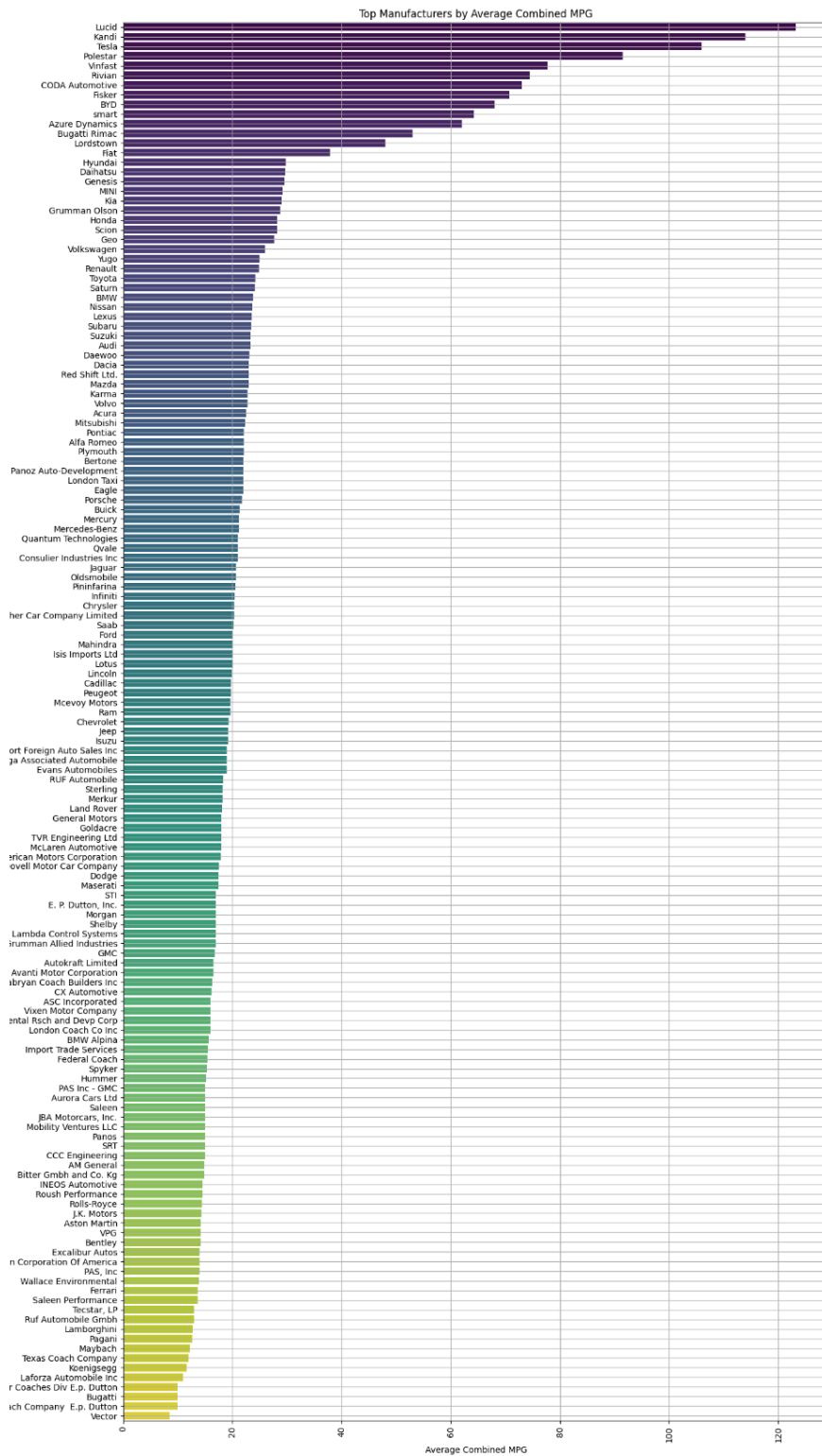
The table shows the average fuel cost (AvgFuelCost) and CO<sub>2</sub> emissions (AvgCO2Emissions) for vehicles across the years 2013 to 2024.

**Trends:**

- Fuel cost has generally decreased over the years, with a noticeable dip in 2023 and 2024. This could be attributed to factors like changes in fuel prices or improvements in vehicle fuel efficiency.
- CO<sub>2</sub> emissions have shown a fluctuating trend, with a decline between 2013 and 2019, followed by a rise in 2020, possibly due to the impact of the pandemic or other external factors like fuel mix.
- The highest emissions are recorded in 2013, with a downward trend until 2019, followed by a slight increase.

**6.2.2 Manufacturer Comparison by Combined MPG:**

- Which manufacturers have the highest average fuel economy (Combined MPG) across their vehicle lineup, and how does this compare with their competitors?



(Figure 6.2.2: Highest Average Combined MPG per Manufacturers)

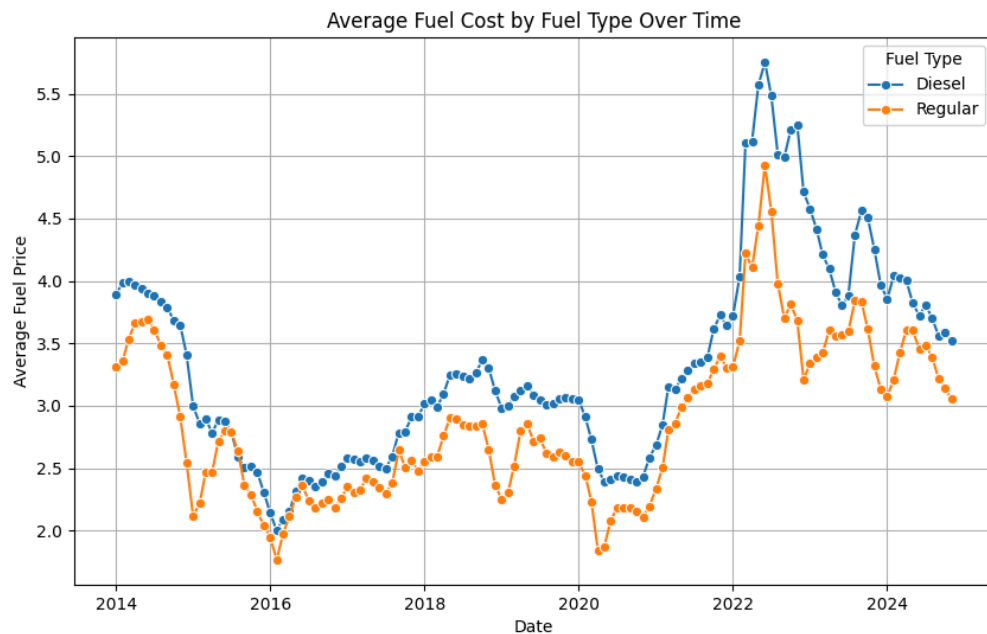
This result lists manufacturers with the highest average Combined MPG (miles per gallon), focusing on Lucid, Kandi, Tesla, Polestar, and Vinfast at the top.

### Trends:

- **Lucid** stands out with the highest average Combined MPG of 123.18, followed by **Kandi** and **Tesla** with significant fuel economy values.
- Many manufacturers, particularly in luxury or performance segments (e.g., Bugatti, Laforza), are at the bottom of the list with very low Combined MPG.

### 6.2.3 Fuel Cost Trends by Fuel Type:

- How does the average fuel cost of each fuel type (e.g., Regular, Diesel) rank over time, and what trends can we observe?



(Figure 6.2.3: Average Fuel Price by Fuel Type)

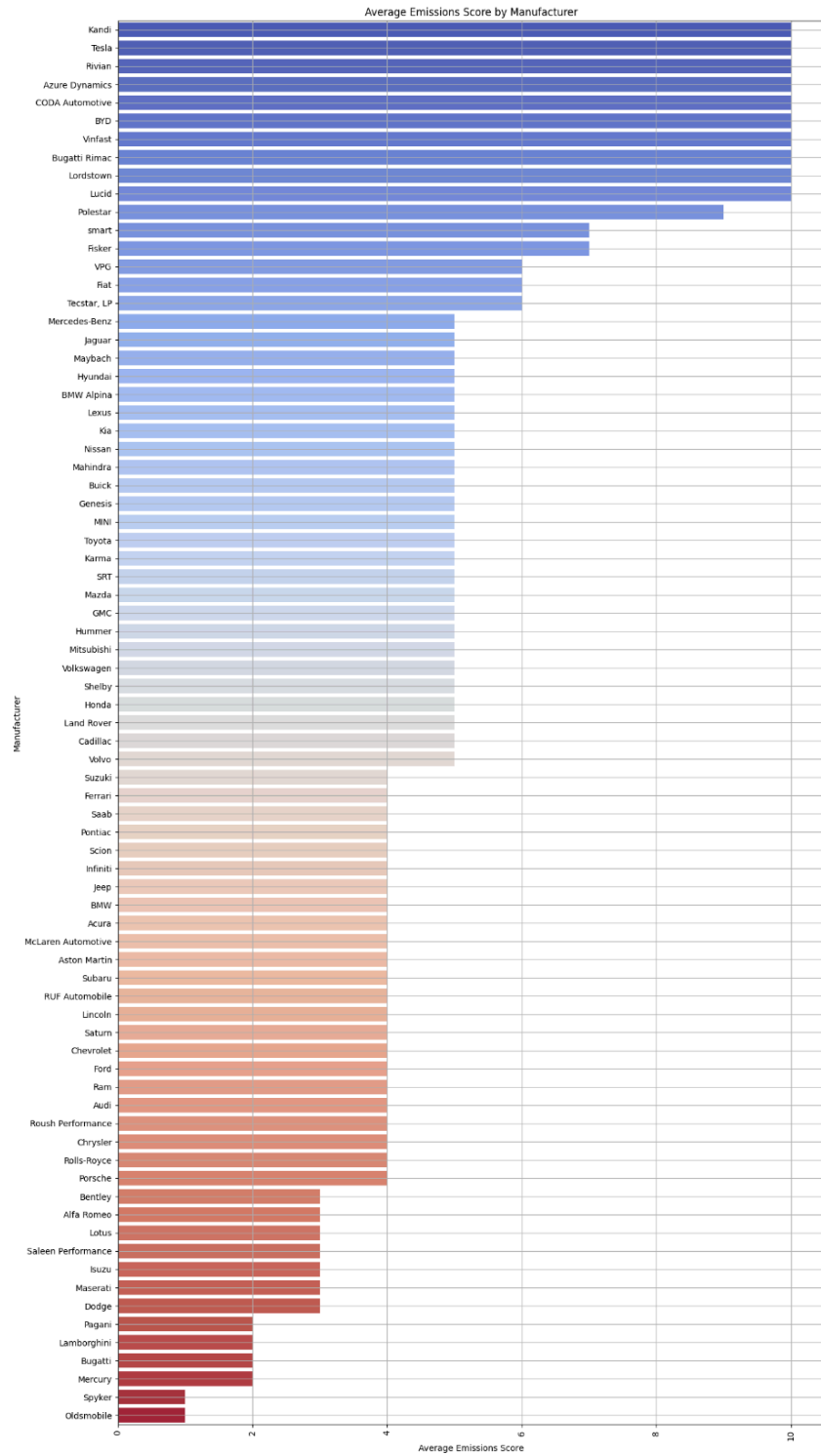
The data presents the average fuel price (AvgFuelPrice) by fuel type (Diesel, Regular) over time, with a ranking (FuelRank) based on time.

**Trends:**

- Diesel fuel prices tend to be higher than Regular fuel, with fluctuations in both over the years.
- Fuel prices show some seasonal patterns, with the highest ranks appearing in the 2014-2015 period and fluctuating until 2024. Diesel has shown slightly more volatility, but both fuel types follow similar price trends over time.

**6.2.4 Vehicle Type Distribution and Emissions:**

- What is the distribution of vehicle types (e.g., Two-Seaters, SUVs) across manufacturers and their average emissions?



(Figure 6.2.4: Average Emission Scores of Manufacturers)

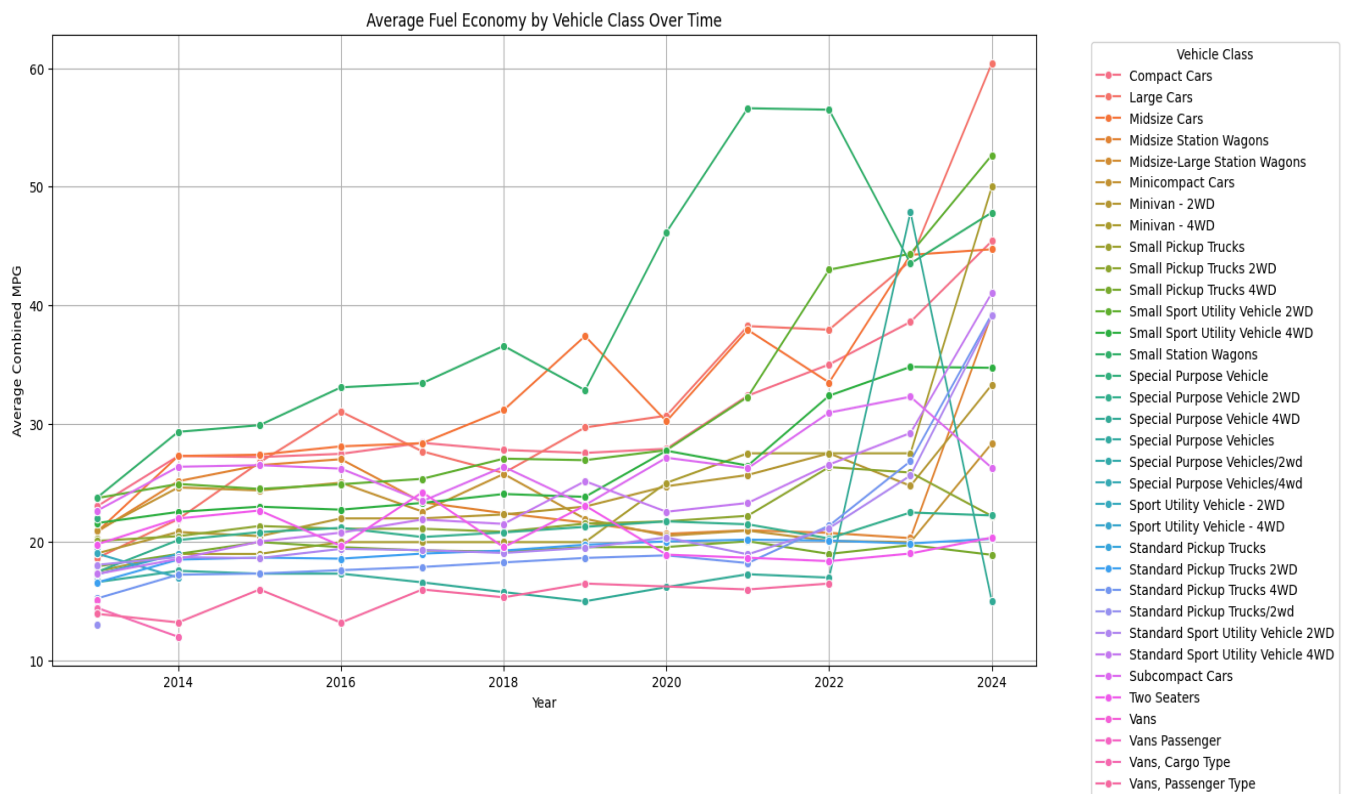
This query provides insights into manufacturers and their average emissions scores, with a focus on Kandi, Tesla, and Rivian having low emissions.

### Trends:

- Manufacturers like **Kandi**, **Tesla**, and **Rivian** dominate the top of the list with the lowest emissions scores, suggesting they prioritize low-emission vehicles.
- At the other end, **Lamborghini** and **Bugatti** show higher emissions, indicating that their vehicles are less efficient and produce more emissions.

### 6.2.5 Average Fuel Economy by Vehicle Class:

- How has the average fuel economy (Combined MPG) for each vehicle class changed over time across all manufacturers?



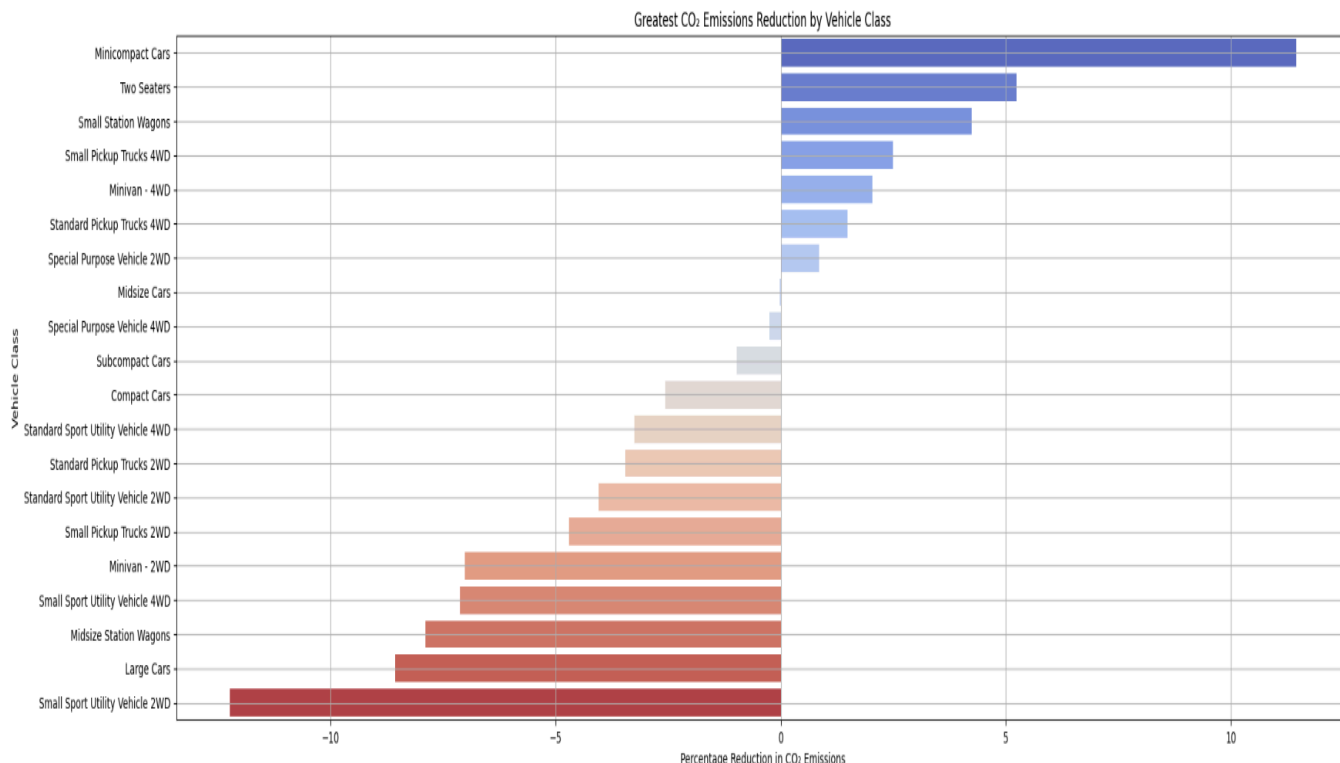
(Figure 6.2.5: Average MPG by Vehicle Class)

This visual shows the trend of Average MPG by Vehicle Class Over Time:

- **Compact Cars:** The fuel economy for compact cars shows a steady increase over the years, rising from about 23.01 MPG in 2013 to 45.42 MPG in 2024. This significant improvement indicates advances in fuel efficiency for smaller vehicles, likely driven by improved technology, lightweight materials, and better engine performance.
- **Large Cars:** Large cars have also seen a noticeable improvement in fuel economy, from 18.65 MPG in 2013 to 60.44 MPG in 2024. This is a very large increase, suggesting that manufacturers have been making strides in making traditionally fuel-inefficient vehicles more eco-friendly, possibly through hybrid or electric powertrains.
- **Midsize Cars:** Midsize cars show a similar trend, with an increase in fuel efficiency, from 20.85 MPG in 2013 to 44.73 MPG in 2024. This growth could reflect the widespread adoption of hybrid and electric versions of midsize vehicles.
- **SUVs, Pickup Trucks, and Vans:** Some vehicle types, such as pickup trucks (both standard and small), SUVs, and vans, show smaller improvements in MPG over time compared to smaller passenger cars. For example:
  - **Small Sport Utility Vehicles (SUVs):** These have also seen improvements, going from 23.72 MPG in 2013 to 52.68 MPG in 2024.
  - **Standard Pickup Trucks:** These vehicles had a more modest improvement in MPG, especially 2WD models, which started at about 15 MPG in 2013 and reached around 20 MPG in 2024. 4WD models saw an increase from about 15.23 MPG to 39.25 MPG over the same period.
  - **Vans:** Fuel efficiency for vans remains relatively lower, particularly for the "Passenger Type" van, which remained around 15-16 MPG throughout the years, except for a small peak in 2024.

#### 6.2.6 Vehicle Classes with Best and Worst Performance:

- Which vehicle classes have shown the greatest improvement in reducing CO<sub>2</sub> emissions per mile over the last 5 years?



(Figure 6.2.6: Vehicle Class Ranking by Percentage Reduction in CO<sub>2</sub> Emissions)

This query ranks vehicle classes by their percentage reduction in CO<sub>2</sub> emissions over the last 5 years.

#### Trends:

- **Minicompact Cars** show the greatest improvement with an 11.45% reduction in emissions, followed by **Two Seaters** and **Small Station Wagons**, both with notable reductions.
- Classes like **Standard Pickup Trucks 2WD**, **Standard Sport Utility Vehicle 2WD**, and **Minivan - 2WD** show reductions in emissions, while others such as **Midsize Cars** and **Compact Cars** have seen slight increases, suggesting limited progress in those categories.
- The **smallest vehicles** (such as **Minicompact Cars** and **Subcompact Cars**) have shown consistent improvements in fuel economy, which is expected due to the push for fuel-efficient urban transportation.

- **Special Purpose Vehicles** (like 2WD and 4WD variants) show more varied results. For instance, 4WD special purpose vehicles had high fluctuations but still maintained a relatively consistent improvement trend. By 2024, their MPG is still lower compared to other vehicle classes, like compact cars and SUVs.
- **Vehicle Categories with Lower MPG:** Trucks and larger SUVs, particularly those with 4WD, still lag behind in fuel efficiency compared to compact cars. These vehicles require larger engines and more energy for off-road and hauling purposes, limiting their fuel economy.

### Notable Observations:

#### Improvement in Specific Years:

- From **2019 to 2024**, there seems to be a sharp increase in MPG for many vehicle categories. This could be attributed to the increased adoption of hybrid or electric powertrains and stricter fuel efficiency standards.
- **SUVs and Pickup Trucks**, which were historically less fuel-efficient, saw larger jumps after 2020, possibly due to changes in consumer preferences and technological advancements in powertrains.
- In **2020**, for example, the **Compact Cars** had an impressive average of 27.88 MPG, while the **Small Sport Utility Vehicles** improved to 27.72 MPG. This shift towards higher efficiency is indicative of the increasing popularity of hybrid vehicles and eco-conscious design choices in the auto industry.

#### Vehicle Class Performance in 2024:

- In **2024**, **Large Cars** and **Midsize Cars** saw the highest averages, with large cars reaching 60.44 MPG. This is a remarkable shift, reflecting significant progress in technology for traditionally fuel-inefficient vehicle categories.
- **Small Station Wagons** and **SUVs** also performed well, with averages above 40 MPG in many cases.

- **Special Purpose Vehicles, Pickup Trucks, and Vans** still lag behind in fuel efficiency in comparison, but their trends are improving, especially with 4WD vehicles showing increased efficiency.

## Conclusion

Overall, the data reveals a **gradual improvement** in fuel economy across nearly all vehicle classes, driven by advances in **fuel-efficient technologies, hybrid powertrains, and lighter materials**. While **compact cars** and **midsize cars** have seen the most significant increases, larger vehicles such as **large cars, vans, and pickup trucks** are catching up, though their MPG lags behind smaller vehicles. The **biggest gains** occurred from **2019 onward**, particularly in **2020-2024**, showing a shift towards more sustainable, energy-efficient automotive technologies across a wide range of vehicle categories.

## Chapter 7

Q1. What is the average fuel cost and CO<sub>2</sub> emissions across all vehicles in the datasets by year?

Purpose: Summarize trends in fuel cost and environmental impact over time.

Visual: Line chart or bar chart comparing yearly averages.

```
SELECT
    YEAR(Date) AS Year,    -- Replace 'YourDateColumn' with the
                           correct date column
    AVG(FuelCost) AS AvgFuelCost,    -- Replace 'YourFuelCostColumn'
    with the correct column
    AVG(CO2Emissions) AS AvgCO2Emissions    -- Replace
    'YourCO2EmissionsColumn' with the correct column
FROM
    FactFuelEconomy
GROUP BY
    YEAR(Date)
```

```
ORDER BY  
    Year;
```

Q2. Which manufacturers have the highest average fuel economy (Combined MPG) across their vehicle lineup, and how does this compare with their competitors?

Purpose: Identify top-performing manufacturers in terms of fuel efficiency.

Visual: Horizontal bar chart showing top manufacturers by average MPG.

```
SELECT  
    d.make AS Manufacturer,  
    AVG(f.CombinedMPG) AS AvgCombinedMPG  
FROM  
    FactFuelEconomy f  
JOIN  
    DimVehicle d ON f.VehicleID = d.VehicleID -- Replace  
    'vehicle_id' with the actual common column  
GROUP BY  
    d.make  
ORDER BY  
    AvgCombinedMPG DESC;
```

q3. How does the average fuel cost of each fuel type (e.g., Regular, Diesel) rank over time, and what trends can we observe?

Purpose: Rank fuel types by cost to analyze trends over time.

Visual: Line chart with ranking overlay or grouped rank-based summary table.

```

WITH FuelRank AS (
    SELECT
        FuelType,
        Date,
        AVG(FuelPrice) AS AvgFuelPrice,
        RANK() OVER (PARTITION BY FuelType ORDER BY Date) AS FuelRank
    FROM
        DimFuel
    GROUP BY
        FuelType, Date
)
SELECT
    FuelType,
    Date,
    AvgFuelPrice,
    FuelRank
FROM
    FuelRank
ORDER BY
    Date, FuelType;

```

q4. What is the distribution of vehicle types (e.g., Two-Seaters, SUVs) across manufacturers and their average emissions?

Purpose: Analyze emissions performance within a hierarchical structure (vehicle type > manufacturer).

Visual: Treemap or grouped bar chart showing average emissions by manufacturer and vehicle class.

```

SELECT
    d.make AS Manufacturer,
    AVG(e.score) AS AvgEmissionsScore

```

```

FROM
    DimVehicle d
JOIN
    FactFuelEconomy f ON d.VehicleID = f.VehicleID
JOIN
    DimEmissions e ON d.VehicleID = e.VehicleID
GROUP BY
    d.make
ORDER BY
    AvgEmissionsScore DESC;

```

#### Additional Business Questions (from Above Reports)

Question: How has the average fuel economy (Combined MPG) for each vehicle class changed over time across all manufacturers?

Purpose: Understand how vehicle types (e.g., SUVs, sedans) are evolving in terms of fuel economy, identifying trends in improvements or deteriorations in fuel efficiency.

Visual: Line chart or grouped bar chart comparing average MPG over time by vehicle class.

```

SELECT
    YEAR(f.Date) AS Year,
    v.VehicleClass AS VehicleClass,
    AVG(f.CombinedMPG) AS AvgCombinedMPG
FROM
    FactFuelEconomy f
JOIN
    DimVehicle v ON f.VehicleID = v.VehicleID
GROUP BY
    YEAR(f.Date), v.VehicleClass
ORDER BY
    Year, VehicleClass;

```

Question: Which vehicle classes have shown the greatest improvement in reducing CO<sub>2</sub> emissions per mile over the last 5 years?

Purpose: Identify which vehicle classes have been most successful in reducing their environmental impact, helping target areas for further environmental policy focus.

Visual: Line chart showing percentage reduction in emissions for each vehicle class, with a focus on the top performers.

```
WITH EmissionsByYear AS (  
    SELECT  
        YEAR(f.Date) AS Year,  
        v.VehicleClass AS VehicleClass,  
        AVG(f.CO2Emissions) AS AvgCO2Emissions  
    FROM  
        FactFuelEconomy f  
    JOIN  
        DimVehicle v ON f.VehicleID = v.VehicleID  
    WHERE  
        YEAR(f.Date) >= YEAR(GETDATE()) - 5  
    GROUP BY  
        YEAR(f.Date), v.VehicleClass  
)  
SELECT  
    e1.VehicleClass,  
    ((e1.AvgCO2Emissions - e2.AvgCO2Emissions) / e2.AvgCO2Emissions)  
    * 100 AS PercentageReduction  
FROM  
    EmissionsByYear e1  
JOIN  
    EmissionsByYear e2 ON e1.VehicleClass = e2.VehicleClass  
    AND e1.Year = e2.Year + 5  
ORDER BY  
    PercentageReduction DESC;
```

