Analyzing Data using PL/SQL

Mansi Thumbar

Analyzing Big Data I

Table of Contents

Consultancy Firm

We at Akatsuki Consultancy believe in creating practical solutions for our clients by utilizing our specialists' understanding of the automobile industry. 'Akatsuki' in Japanese, stands for daybreak or dawn, which reflects our belief in providing a bright start to our clients.

Akatsuki Consultancy firm aims to understand the mindset of consumers when purchasing a car. Investigating historical data and preferences of the consumers allows us to analyze the trends and patterns of car features that are important for different consumer demographics. Ultimately permitting us to aid the current manufacturers focus on car design in accordance with the preferences.

With growing trends towards EVs and cleaner commutation, it is evident that the automobile industry is going through a dramatic change. It is estimated that the decrease in new vehicle sales worldwide is around 11% from 2021 to 2022. Automobile producers all over the world are introducing a new range of EV that can compete with the disruption caused by Tesla. Due to these changes in the industry, we would like to analyze the automobile industry and get a better understanding of the new trends.

Audience

For the purpose of this project, our audience will be limited to car manufacturing companies. The data which we have collected focuses on the sales of 3 major companies in the industry (Ford, BMW, and Toyota). We will be using this data to garner a better understanding of past data and whether specific factors and features of vehicles sold can be attributed to the decline in automobile sales. The company executives of these 3 companies will be our target audience for the project. After seeing our research, these executives will be able to understand what specs and requirements a customer is attracted to when purchasing a new car. Thus, our clients will be able to tailor their manufacturing and future car production to the consumer's desired features, designs and more. This will allow automakers to reduce the amble variety they currently offer to customers, and cut production costs which will ultimately increase their profits.

Data Exploration

The source dataset our consultancy is utilizing consists of multiple columns representing the type of cars, manufacturer, engine type, fuel type etc. We have cherry picked some columns from the source dataset which we perceive will help us extract meaningful trends for our analysis. Below we have provided the list of columns we have selected and what they represent.

Id: A unique identification number for all the cars

Year: Year the model was manufactured

front head room: Distance between the roof of the car and the front seat.

front hip room: Width of the car's front seats

front leg room: Amount of leg space available for the driver

front shoulder room: Shoulder room is the distance from the driver's upper door panel to the front passengers upper door panel.

Price: The price of each individual car

Exterior-color: The color on the exterior of the car

Interior-color: The color on the interior of the car

Engine type: Type of engine installed in the car

Fuel type: Type of fuel used by the car

Fuel capacity: Total capacity of the fuel tank

Range City/Highway: The mileage car gives on a full tank in city v/s highway

Domestic/Imported: Whether the car is locally made or imported.

Country of origin: Country where the car was manufactured

Car classification: The type of car for eg compact car, truck, etc.

Platform code/generation code: The generation of car produced.

Body Type: Chassis of the car

Doors: Total number of doors in the car

Seatings: Total number of seats in the car

Length: Total length of the car

Width: Total width of the car

Height: Total height of the car

Wheel Base: Horizontal distance between the centers of the front and rear wheels

Cylinders: Total number of cylinders in the engine

Engine Size: Size of the engine

HP: Horsepower of the engine

RPM: Revs per minute in the engine

Torque: Torque provided by the engine

Drive Type: The car is either rear wheel drive or all wheel drive.

Transmission: Total number of gears drives in the car

The source dataset from where we are taking our samples has collected data from various sources such as JSON(JavaScript Object Notation), review pages, HTML(HyperText Markup Language), Teoalida. Columns like Price & colors(Price, Interior-color, Exterior-color), Dimensions and weight(Body Type, Doors, Length, Width, Height, Seatings, wheel base), Engine and transmission(Cylinders, Engine Size, HP, RPM, Torque, Drive Type, Transmission), Fuel and mileage(fuel type, fuel capacity, Engine type), Interior Dimensions(front head room,

front hip room, front legroom, front shoulder room) have all been collected via JSON, Bonus Data(Domestic/Imported, Country of Origin, Car Classification, Platform code/generation code) via Teoalida, and Naming(Manufacturer, Model, Year) via HTML.

The front head room, front hip room, front legroom, front shoulder room, fuel type, fuel capacity, Cylinders, Engine Size, HP, RPM, Torque, Doors, Range/Miles, Length, Width, Year, Height, Seatings, wheel base all these columns are numeric-based.

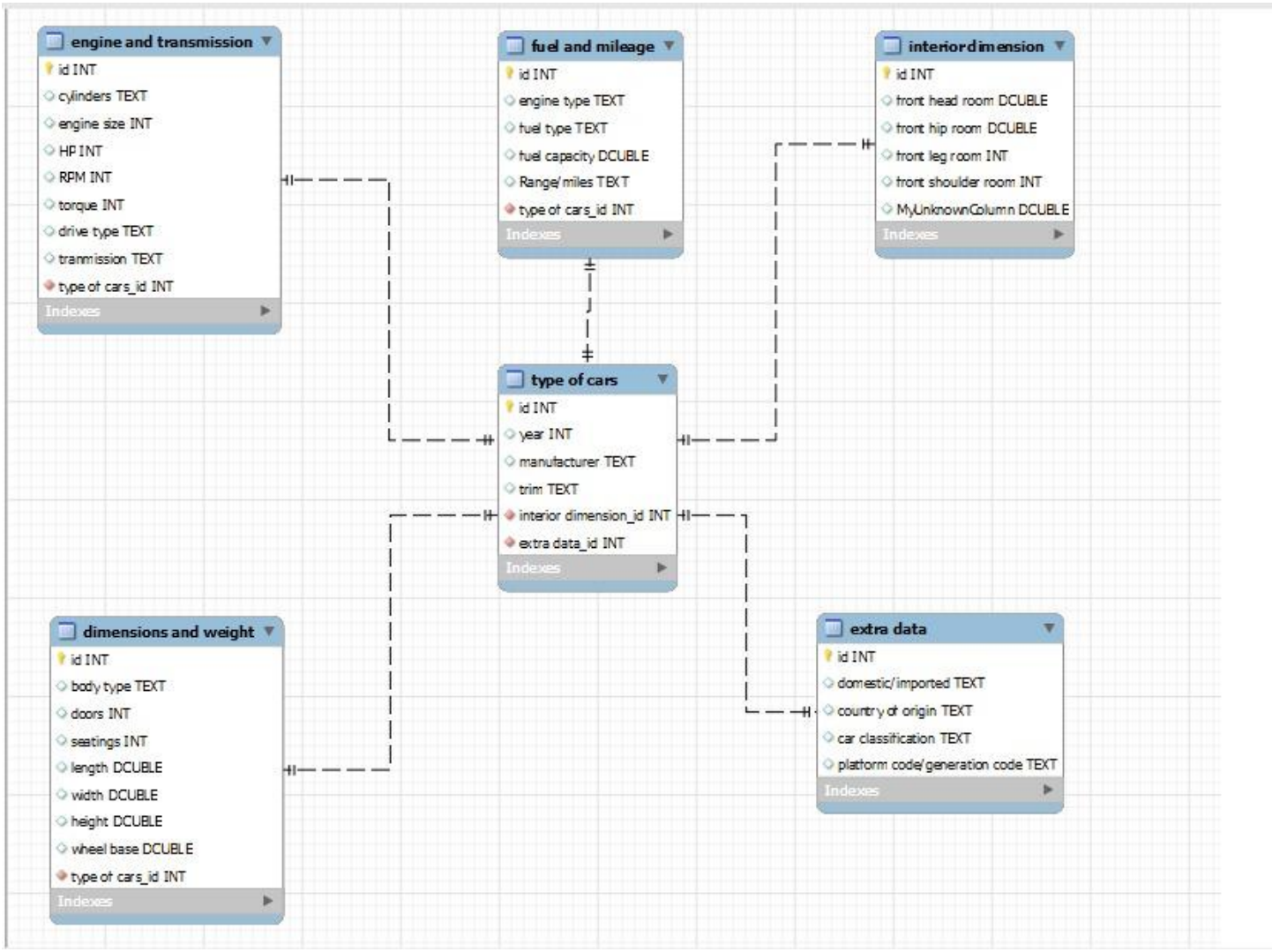
Engine type, fuel type, Transmission, manufacturer, Interior-color, Exterior-color, Domestic/Imported, Country of Origin, Car Classification, Platform code/generation code, Manufacturer, Model, Body Type, Drive Type, Transmission all these columns are string-based.

Since the data has been collected from various sources(JSON, review pages, HTML, Teolida) there have been multiple instances of null rows in the dataset. The ‘null’ value rows in the numerical-based columns have been replaced by the average of all the values in that column. The ‘null’ value rows in string-based columns have been replaced by random string values.

The original dataset was in the csv format with multiple columns representing various features of the car.

The dataset has details of all cars manufactured by Ford, BMW and Toyota so the first table contains a list of cars under these three manufacturers with a unique id named as type of cars. The next step towards normalizing data in the 1NF form was to divide the relevant features into one table with a single value in each column and each row associated with the type of cars table’s row by common id.
There is no composite key used to uniquely identify the row in the table so the dataset is in the 2NF and 3NF form.

EER



This diagram shows the one to one relationship between the main schema in the database that is titled ‘type of cars’ and various schemas that are the features of these cars connected by one-to-one relation to the main schema.

Queries

```
SELECT
   exterior_color, AVG(price) AS avg_price
FROM
   price_colors
GROUP BY exterior_color
ORDER BY avg_price DESC;
```

- ***Output***:

exterior_color	avg_price
Smoked Topaz Metallic	66500.0000
Black Sapphire Metallic	63500.0000
Le Mans Blue Metallic	61940.0000
Space Gray Metallic	60725.0000
Alaska Blue Metallic	59110.0000
Iconic Silver Metallic(	57842.5000
Magma Red Metallic	57382.5000
Star White Metallic Tri-Coat	56910.8333
Steel Blue Metallic	56700.0000
Carbon Black Metallic	56600.0000

- ***What the Query Does***: The query above shows the average price of a car based on the car's exterior color.
- ***Why the Query is Important***: The query is beneficial to show any relevant relationship between the color of a car and the color of it. This would demonstrate if a color is more popular than another color. If there is a relationship between the color of a car and the price, this will help us determine which color a producer should focus on making and which colors a producer should limit producing.

```
SELECT DISTINCT
   exterior_color
FROM
   price_colors;
SELECT DISTINCT
   model
FROM
   type_of_cars;
SELECT DISTINCT
   body_type
FROM
   dimensions_weight;
```

- ***Output***:

exterior_color	model	body_type
Mineral White Metallic	328i SULEV - Sedan	Sedan
Glacier Silver Metallic	328i xDrive - Sedan	Wagon
(Ind) Dravit Grey Metallic	328i xDrive SULEV - Sedan	Coupe
Alpine White	335i - Sedan	Convertible
Smoked Topaz Metallic	335i xDrive - Sedan	Hatchback
Jet Black	328d - Sedan	Truck
Mediterranean Blue Metallic	330i - Sedan	
Black Sapphire Metallic	330i xDrive - Sedan	
Imperial Blue Metallic	M340i - Sedan	
Liquid Blue Metallic	M340i xDrive - Sedan	

- ***What the Query Does***: The queries return the different types of exterior colors, models, and body types available to customers.
- ***Why the Query is Important***: This query will help demonstrate the vast amount of basic options and features available to customers. We can use this information to show that automakers have created a wide breadth of options for customers but in doing so have made it more difficult on themselves as they must provide and produce a plethora of options, which is expensive.

```
SELECT
   id,
   manufacturer,
   model,
   CASE
      WHEN HP BETWEEN 350 AND 500 THEN 'Powerful'
      WHEN hp BETWEEN 200 AND 349 THEN 'Average'
      ELSE 'Weak'
   END AS power
FROM
   engine_transmission
      INNER JOIN
   type_of_cars USING (id);
```

- ***Output***:

id	manufactur...	model	power
1	BMW	328i SULEV - Sedan	Average
2	BMW	328i xDrive - Sedan	Powerful
3	BMW	328i xDrive SULEV - Sedan	Average
4	BMW	335i - Sedan	Powerful
5	BMW	335i xDrive - Sedan	Average
6	BMW	328d - Sedan	Average
7	BMW	330i - Sedan	Average
8	BMW	330i xDrive - Sedan	Average
9	BMW	M340i - Sedan	Powerful
10	BMW	M340i xDrive - Sedan	Powerful

- ***What the Query Does***: This query creates a new column that ranks the power of each car based on its horsepower by categorizing each one as either weak, average, or powerful.
- ***Why the Query is Important***: This query can be used to show the power consumers prefer their cars to have and the minimum necessary requirement that an automaker should consider when producing a car.

```
SELECT
    manufacturer,
    body_type,
    AVG(price) AS avg_price,
    IF(AVG(price) > 40000,
        'higher income group',
        'lower income group') AS targeted_segment
FROM
    price_colors
        INNER JOIN
    dimensions_weight USING (id)
        INNER JOIN
    type_of_cars USING (id)
GROUP BY manufacturer , body_type;
```

- ***Output***:

manufacturer	body_type	avg_price	targeted_segment
BMW	Sedan	41702.1605	higher income group
BMW	Wagon	38252.7778	lower income group
BMW	Coupe	43241.4634	higher income group
BMW	Convertible	46482.6667	higher income group
BMW	Hatchback	55315.0000	higher income group
Ford	Truck	35148.6976	lower income group
Toyota	Hatchback	26117.0732	lower income group
Toyota	Sedan	19995.0000	lower income group

- ***What the Query Does***: The query creates a new column to categorize based on the average price of the targeted income group for each manufacturer and their body types they produce. The information is taken from multiple tables so an inner join was necessary since all information is necessary for the analysis.
- ***Why the Query is Important***: The query will help highlight the body types being produced in each targeted segment and the average price of the cars being sold in each respected targeted segment. This will give automakers a better reference and understanding of where to price their cars depending on the body type they are producing and the income group of the targeted customer.

```
SELECT
   drive_type,
   manufacturer,
   cylinders,
   tranmission,
   AVG(price),
   AVG(fuel_capacity)
FROM
   engine_transmission AS et
      INNER JOIN
   price_colors AS pc USING (id)
      INNER JOIN
   fuel_mileage AS fm USING (id)
      INNER JOIN
   type_of_cars AS cars USING (id)
GROUP BY drive_type
ORDER BY AVG(price) DESC;
```

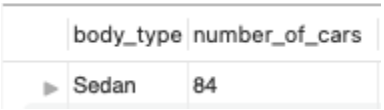
- ***Output***:

drive_type	manufacturer	cylinders	tranmission	AVG(price)	AVG(fuel_capacity)
all wheel drive	BMW	I6	8-speed shiftable automatic	39260.0926	15.214814814814826
rear wheel drive	BMW	I4	8-speed shiftable automatic	37855.3276	21.681034482758598
four wheel drive	Ford	V6	10-speed shiftable automatic	36632.3214	28.972619047619013
front wheel drive	Toyota	I4	continuously variable-speed automatic	25738.7975	11.648101265822774

- ***What the Query Does***: The query returns information about the transmission used, the average price and the average fuel capacity (the number gallons a car can hold) based on the drive type of the car. It is ordered from highest average price to lowest.
- ***Why the Query is Important***: This information will allow the automaker, assuming all else being equal, to have a reference of how much a customer values a certain kind of drive type. For example, assuming all else being equal, a customer is willing to pay more for a car with all wheel drive than front wheel drive by about $14,000. On the other hand, a customer is only willing to pay about $1,200 more for rear wheel drive than four wheel drive, depicting the marginal increase in revenue an automaker would receive if it was considering between the two options.

```
SELECT
   body_type, COUNT(body_type) AS number_of_cars
FROM
   dimensions_weight
WHERE
   body_type LIKE 's%'
GROUP BY body_type;
```

- ***Output***:



	body_type	number_of_cars
▶	Sedan	84

- ***What the Query Does***: The query returns information regarding the number of sedans. This query focuses on body types starting with the letter “s” in order to isolate sedans.
- ***Why the Query is Important***: This information allows an automaker to quickly reference the volume of a specific vehicle body type, in this case a sedan. This is valuable to an automaker because they can see which segments of the market are congested versus segments which could present an opportunity to target.

```
SELECT
   manufacturer,
   ROUND(AVG((`Range` / fuel_capacity)), 2) AS mileage_city,
   ROUND(AVG((Miles / fuel_capacity)), 2) AS mileage_highway
FROM
   type_of_cars AS t
      INNER JOIN
   fuel_mileage AS f USING (id)
GROUP BY manufacturer;
```

- ***Output***:

manufacturer	mileage_city	mileage_highway
BMW	19.35	28.45
Ford	15.34	19.7
Toyota	51.79	48.33

- ***What the Query Does***: The query returns the average miles per gallon a car would get in the city and on the highway, and it is grouped based on the car manufacturer.
- ***Why the Query is Important***: The query will allow an automaker to compare its cars fuel efficiency relative to its competitors. The query shows the dramatic difference of the fuel efficiency between each manufacturer. This would thus let a car producer with a higher fuel efficiency understand its most similar rival would be Toyota while an automaker with lower fuel efficiency may be more likely to be competing with Ford.

```
SELECT
   AVG(HP) AS avg_hp, MAX(HP) AS max_hp, MIN(HP) AS min_hp
FROM
   engine_transmission;
```

- ***Output***:

avg_hp	max_hp	min_hp
236.7327	453	110

- ***What the Query Does***: The query returns the average horsepower of cars produced and also the lowest and highest horsepower of the cars in the entire dataset.
- ***Why the Query is Important***: The query is relevant as customers care about the horsepower of the cars they purchase because horsepower directly correlates to the performance of a car. The higher the horsepower, the more torque and greater acceleration the car can handle. Automakers therefore must know the average amount of horsepower that consumers are expecting from a car they purchase.

```
SELECT
   tof.manufacturer,
   ROUND(AVG(price), 2) AS avg_price,
   body_type,
   ROUND(AVG(fuel_capacity), 2) AS avg_fuel_cap
FROM
   type_of_cars AS tof
      INNER JOIN
   price_colors AS p USING (id)
      INNER JOIN
   dimensions_weight AS dw USING (id)
      INNER JOIN
   fuel_mileage AS fm USING (id)
WHERE
   manufacturer = 'BMW'
      OR manufacturer = 'Toyota'
      OR manufacturer = 'Ford'
GROUP BY dw.body_type , tof.manufacturer;
```

- ***Output***:

manufacturer	avg_price	body_type	avg_fuel_cap
BMW	41702.16	Sedan	15.74
BMW	38252.78	Wagon	16.01
BMW	43241.46	Coupe	16.4
BMW	46482.67	Convertible	16.41
BMW	55315.00	Hatchback	14.1
Ford	35148.70	Truck	27.42
Toyota	26117.07	Hatchback	11.56
Toyota	19995.00	Sedan	11.87

- ***What the Query Does***: The query returns columns from multiple data sets in order to show the average price of cars with certain body types and also specific to each manufacturer. The average fuel capacity was also calculated.
- ***Why the Query is Important***: The query will illustrate the average fuel capacity of a certain type of car. This will allow automakers to have a benchmark of where their car ranks among its competitors. Since fuel capacity is important to customers, as they don't want to constantly be filling up their tank, the automaker could charge a premium if the car they manufacture can store more fuel relative to its rivals.

```
SELECT
   manufacturer,year, AVG(HP) AS avg_HP
FROM
   type_of_cars AS t
      INNER JOIN
   engine_transmission AS e USING (id)
GROUP BY year
HAVING avg_HP > 110 ;
```

- ***Output***:

manufacturer	year	avg_HP
BMW	2021	308.8545
BMW	2020	318.8889
BMW	2018	212.8500
BMW	2016	205.2105
BMW	2014	206.0000
BMW	2012	248.8421
BMW	2010	276.4861
BMW	2008	260.1111
BMW	2006	229.0833
BMW	2004	213.8462

- ***What the Query Does***: The query displays the years that the average horsepower for a manufacturer is greater than 110, and it also displays what the average horsepower was for each year that meets that criteria.
- ***Why the Query is Important***: The query is helpful to show historical trends about the amount of horsepower that customers are demanding from automakers. The data shows uniquely that the demand for horsepower comes in waves as it increased from 2000-2010 but then declined from 2010-2018 until it began to rise to an all time high. More analysis needs to be done to uncover what is the cause for this trend, but understanding this cyclicality could lead to very useful information for automakers to predict consumer demands.

```
SELECT
   manufacturer,
   AVG(fuel_capacity) AS avg_fuel_cap,
   AVG(Miles) AS avg_miles,
   (SELECT
         AVG(fuel_capacity)
      FROM
         fuel_mileage) AS overall_fuel_cap
FROM
   type_of_cars
      INNER JOIN
   fuel_mileage AS fm USING (id)
GROUP BY manufacturer;
```

- ***Output***:

manufacturer	avg_fuel_cap	avg_miles	overall_fuel_cap
BMW	16.023837209302293	451.4418604651158	21.821827411167522
Ford	27.415568862275524	538.0688622754492	21.821827411167522
Toyota	11.574117647058813	558.9235294117645	21.821827411167522

- ***What the Query Does***: The query returns the average fuel capacity, average miles total distance (in miles) a car can travel on a full tank of gas, and the overall fuel capacity.
- ***Why the Query is Important***: As customers have become more environmentally conscious, they have put a heightened focus on how fuel efficient their car has become and less on solely the fuel capacity of the car. This is evident as the average total distance a Toyota can travel is 558 miles while the average total distance a BMW can travel is 451 miles. However, a Toyota only has a fuel capacity of about 11.5 gallons on average while a BMW has the fuel capacity of roughly 16 gallons average. This illustrates that automakers can focus on how efficient their cars are instead of how much fuel it can store.

```
SELECT
   manufacturer,
   COUNT(model) AS models,
   CASE
      WHEN id IN (SELECT id FROM dimensions_weight WHERE
            length > (SELECT
                  AVG(length)
               FROM
                  dimensions_weight))
      THEN
         'Longer'
      WHEN
         id IN (SELECT
               id
            FROM
               dimensions_weight
            WHERE
               length < (SELECT
                     AVG(length)
                  FROM
                     dimensions_weight))
      THEN
         'Shorter'
   END AS car_length
FROM
   type_of_cars
GROUP BY manufacturer;
```

- ***Output***:

manufacturer	models	car_length
BMW	172	Shorter
Ford	334	Longer
Toyota	85	Shorter

- ***What the Query Does***: The query returns information displaying the number of models per manufacturer then provides a breakdown of car length relative to the industry's average car length.
- ***Why the Query is Important***: This information allows automakers to better understand the market as a whole. We learn which manufacturers are selling longer or shorter vehicles than the industry average. This is important information when deciding which manufacturers to produce for and also gives insight into which manufacturers are likely competing for the same consumers.

WITH year_comp (manufacturer, year, number_of_cars) AS (SELECT manufacturer, year, COUNT(id) AS number_of_cars FROM type_of_cars GROUP BY manufacturer, year) SELECT manufacturer, year, number_of_cars from year_comp GROUP BY manufacturer, year;

- ***Output***:

manufacturer	year	number_of_cars
BMW	2021	6
BMW	2020	4
BMW	2018	13
BMW	2016	13
BMW	2014	13
BMW	2012	14
BMW	2010	16
BMW	2008	15
BMW	2006	11
BMW	2004	12
BMW	2002	12

- ***What the Query Does***: This query shows the number of models being produced each year by each manufacturer.
- ***Why the Query is Important***: This shows the historical trends of each manufacturer and how many each one produces in a given year. The query illustrates that BMW tends to come out with new models biennially and we can clearly see the impact of COVID-19 on their production.

```
SELECT manufacturer, price, AVG(price)
OVER (PARTITION BY manufacturer) AS avg_price, CASE
        WHEN AVG(price) OVER (partition by manufacturer) > price THEN 'below avg price'
        WHEN AVG(price) OVER (PARTITION BY manufacturer) < price THEN 'above avg
price'
   END AS price_comp
FROM type_of_cars
INNER JOIN price_colors using(id);
```

- ***Output***:

manufacturer	price	avg_price	price_comp
BMW	41250	42700.2035	below avg price
BMW	56700	42700.2035	above avg price
BMW	43250	42700.2035	above avg price
BMW	54700	42700.2035	above avg price
BMW	44550	42700.2035	above avg price
BMW	46550	42700.2035	above avg price
BMW	40750	42700.2035	below avg price
BMW	42750	42700.2035	above avg price
BMW	54000	42700.2035	above avg price
BMW	56000	42700.2035	above avg price

- ***What the Query Does***: The query displays the prices of each car and adds another column that displays whether the car's price is above or below the average price of the cars produced by each manufacturer.
- ***Why the Query is Important***: The query is useful to descriptively show with ease whether a car being manufactured is above or below the average price. This will be beneficial to have easily identified and be able to access as we can reference whether certain cars are considered pricey.

```
UPDATE engine_transmission
SET
   tranmission = NULL
WHERE
   tranmission LIKE '';
SELECT
   *
FROM
   engine_transmission
WHERE
   tranmission IS NULL;
```

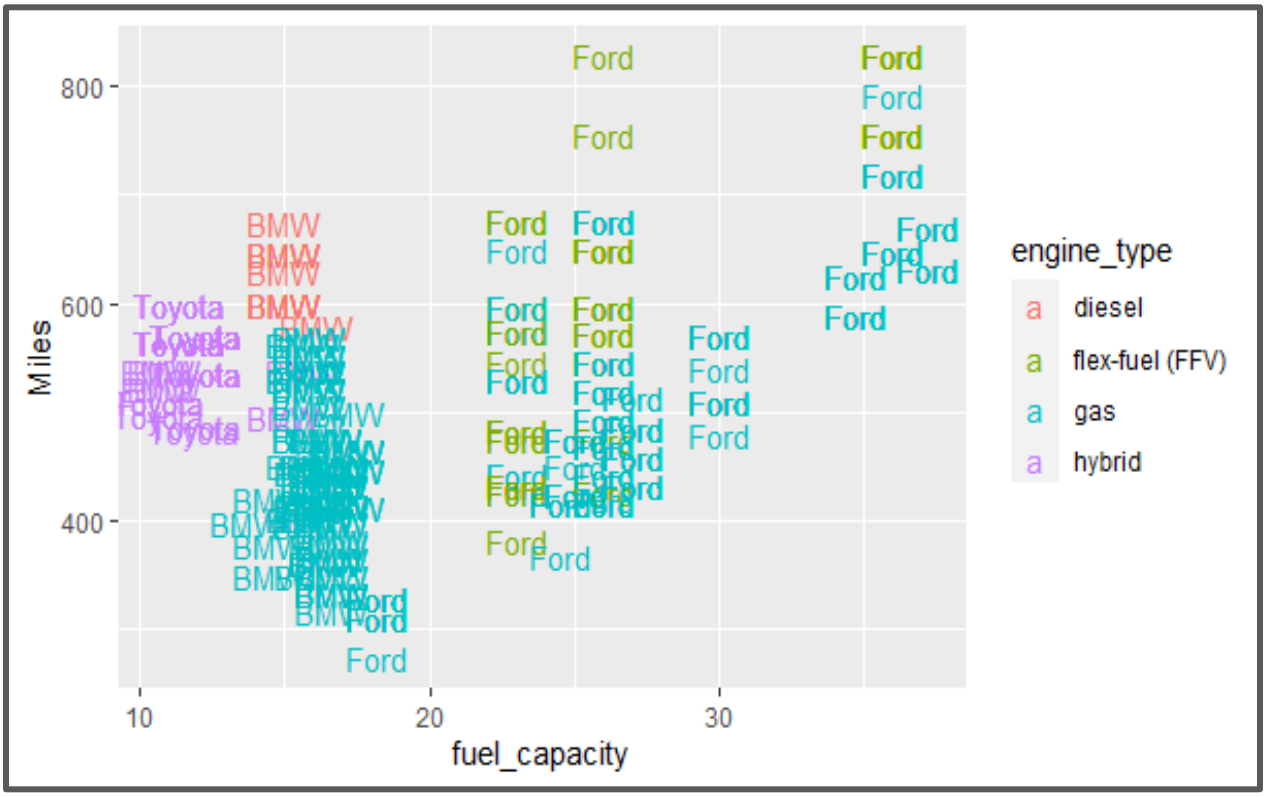
- ***Output***:

id	cylinders	engine_size	HP	RPM	torque	drive_type	tranmission
481	l6	5	145	3400	265	rear wheel drive	NULL
482	l6	5	145	3400	265	rear wheel drive	NULL
483	l6	5	145	3400	265	four wheel drive	NULL
484	l6	5	145	3400	265	four wheel drive	NULL
485	l6	5	145	3400	265	rear wheel drive	NULL
486	l6	5	145	3400	265	rear wheel drive	NULL
487	l6	5	145	3400	265	four wheel drive	NULL
488	l6	5	145	3400	265	four wheel drive	NULL
489	l6	5	145	3400	265	rear wheel drive	NULL
490	l6	5	145	3400	265	rear wheel drive	NULL
491	l6	5	145	3400	265	four wheel drive	NULL

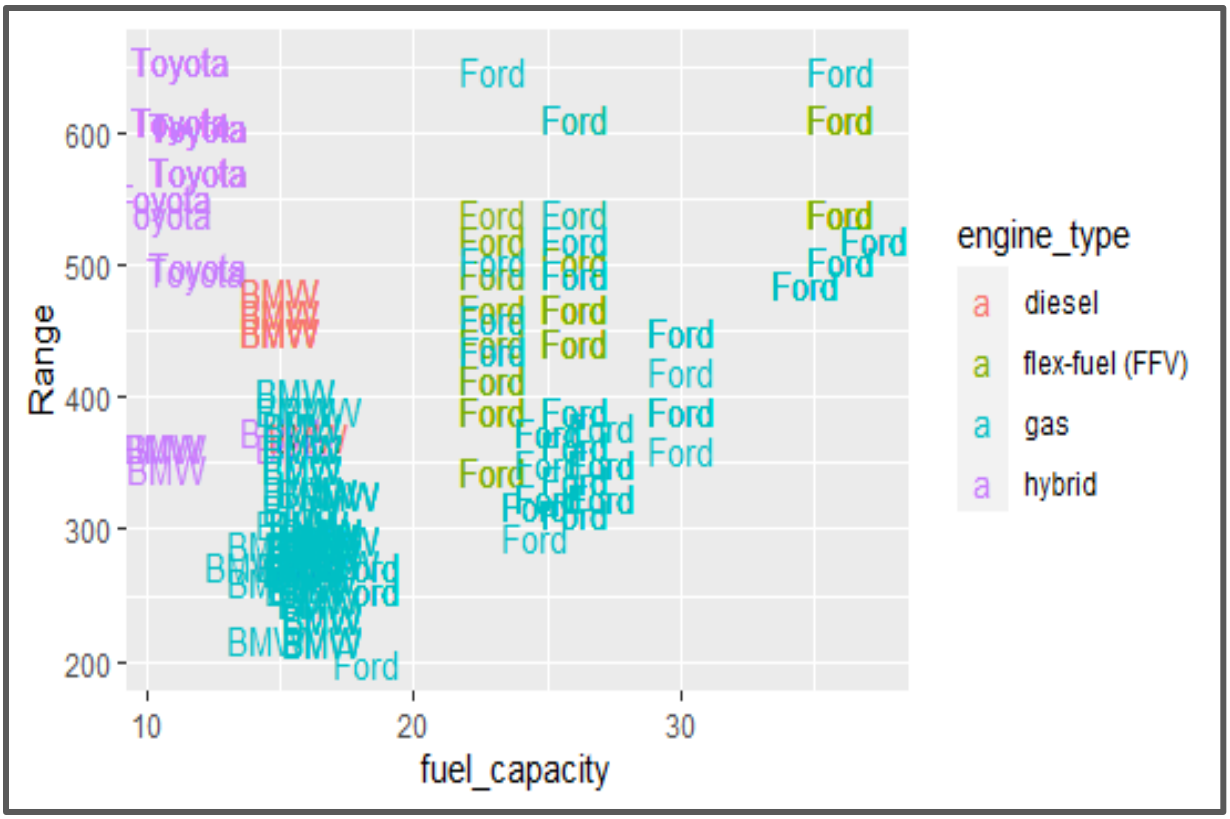
- ***What the Query Does***: The query returns the car models that do not have a value in their transmission column.
- ***Why the Query is Important***: This query is important because it allows us to understand how much data is being missed and get better context onto how encompassing queries are that use transmission. Since there is only a small proportion of models missing the transmission information, we will still be able to use this data to analyze any trends.

Visualization

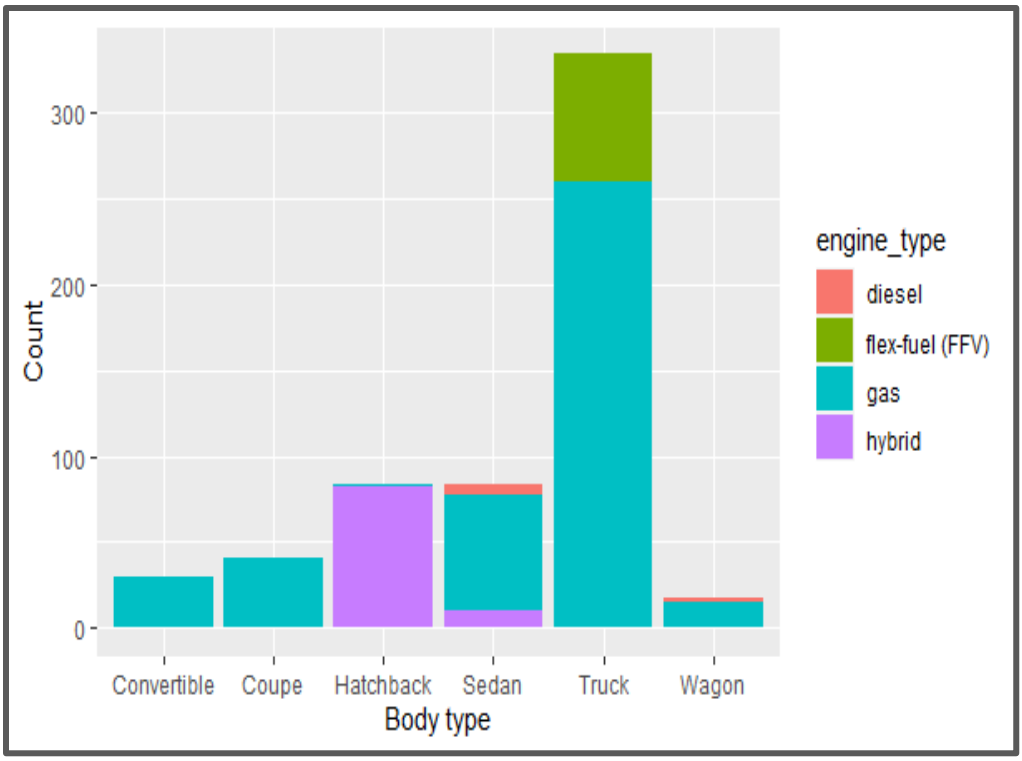
It is valuable to analyze the data beyond its various numerical aspects, so we will provide multiple visualizations. Visualizing the data in different plots can help us identify trends and factors that are most repeated, as well as interpret outliers in the given dataset.



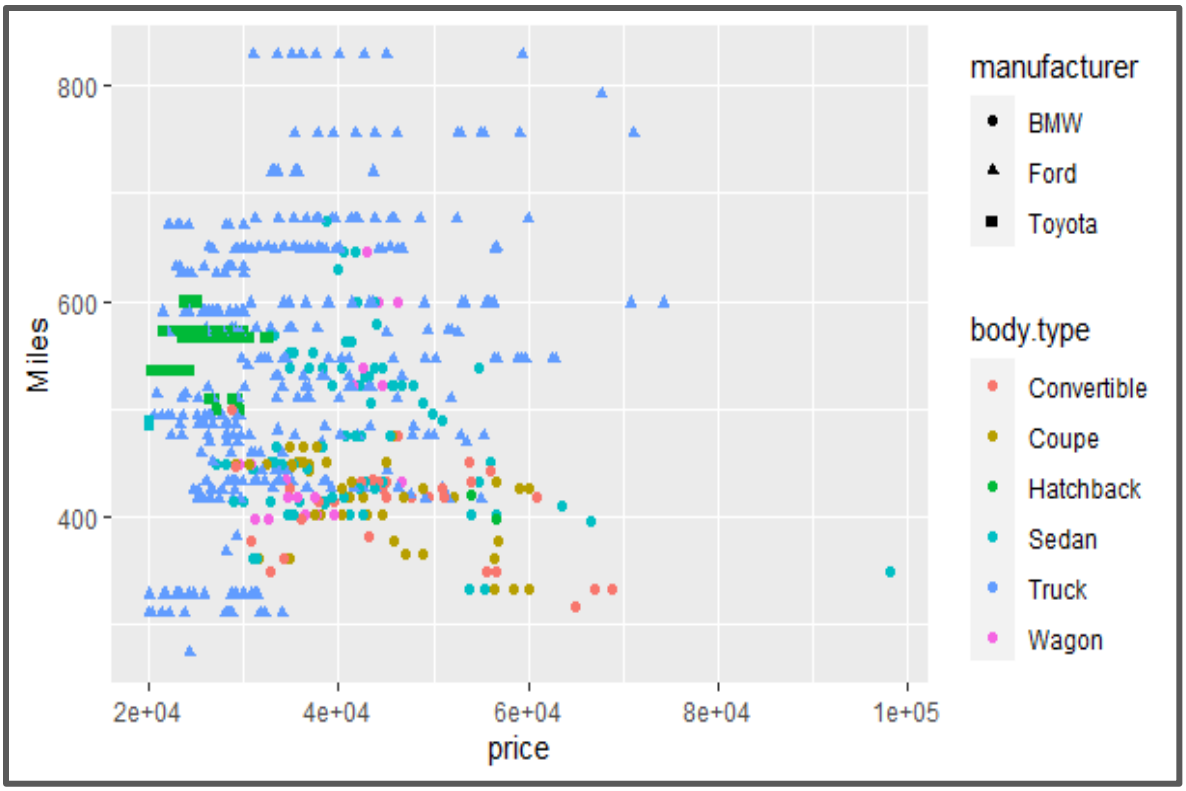
In the graph above, we have plotted the fuel capacity of a car on the x-axis and the total miles it can run on a full tank of gas on the highway. This data can tell us how efficient the car is at long distance travels and which car can give us the longest distance traveled. The graph tells us that Ford has the overall highest miles covered in comparison to any other vehicle. A few of its vehicles, with fuel capacities of either 26 or 36 gallons, can go more than 800 miles with a full tank. However, after looking deeper into the per gallon mileage, we know that Ford vehicles have the least mileage when compared to other vehicles. The best feature of Ford trucks is that it has higher fuel capacity than other vehicles.



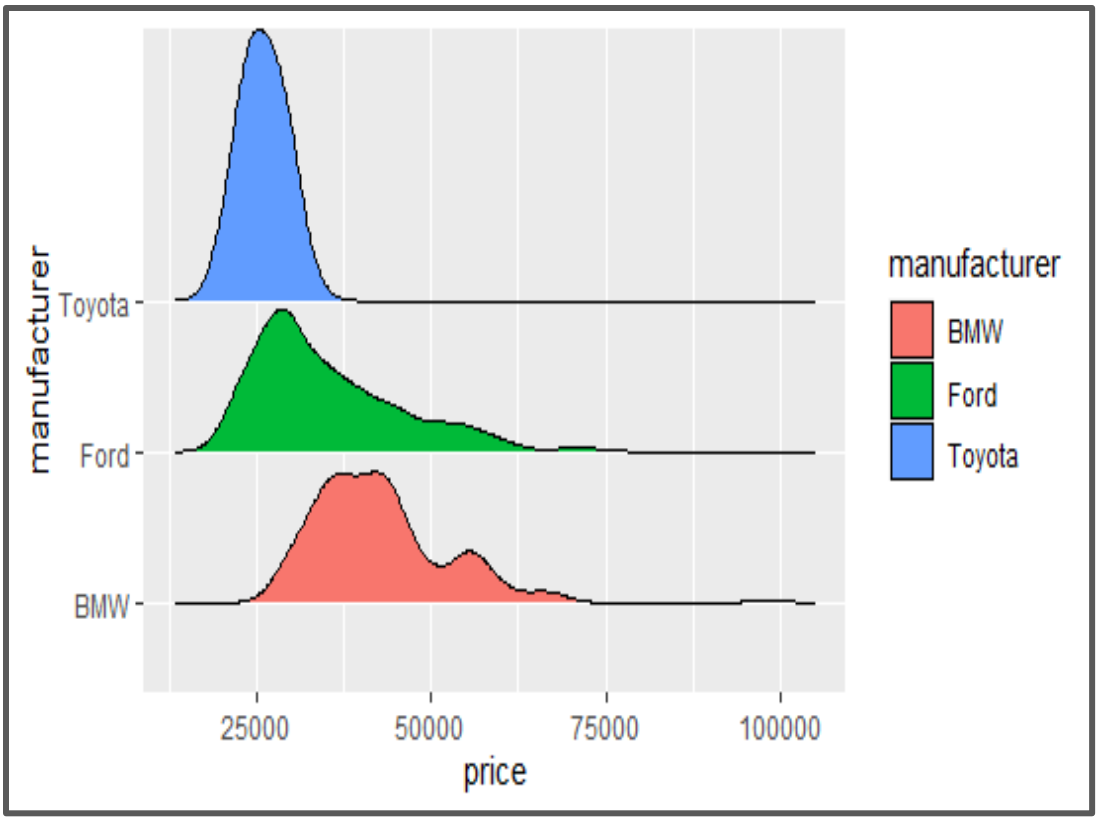
The above graph is similar to the former graph with the exception of range meaning the miles a car can cover in the city with a full tank of gas. As we can see, the mileage given by Toyota on both city and highway is somewhat similar and is on par with Ford's trucks even with extremely lower fuel capacity. This suggests that the Toyota vehicles utilize the hybrid function by running on electricity. This also makes it the most efficient car amongst the batch with the best mileage. Another thing to notice is that the BMW cars, which have better performance on the highway, do not perform as well in the city. Hence, it is safe to say that BMW cars are best suited for highway transit when it comes to mileage.



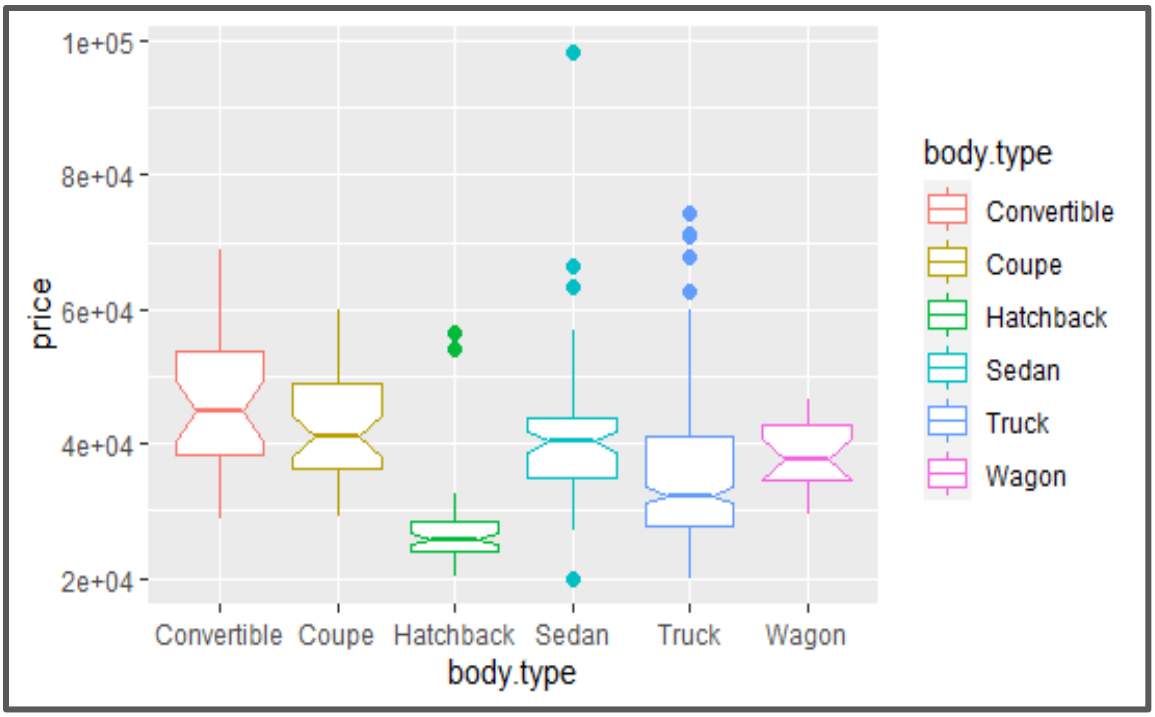
The above graph gives us a segregation of the body type of the cars along with the fuel type of the cars. As we can see, trucks are the most common type of vehicle present in the dataset and they all belong to the Ford family. Apart from a few trucks, most hatchbacks, a few sedans and wagons, the rest of the vehicles in our dataset are all gas or diesel operated. These are one of the worst fuel types when it comes to carbon emissions. This dataset has cars from 1990 to 2020, which suggests that even with the massive popularity of Tesla cars, competition has barely entered the electric vehicle market and most body types are not suitable for hybrid/electric vehicles.



The graph above takes into account the price of the vehicle and mileage provided on the highway. This tells us which manufacturer and vehicle type provide better mileage and at what cost. As we can see, Ford cars have the most miles covered due to high fuel capacity and are mostly available at an average price which makes them popular among the price sensitive market. Toyota cars are the cheapest among the bunch providing the best mileage and hence are the most cost efficient cars. BMW cars are relatively expensive, especially a few of their sedan cars, which are more about luxury and less about the vehicle efficiency.



The above graph compares the price range of the given cars and the curve represents the density of vehicles in the given price range. As we established earlier, Toyota cars are the cheapest of the bunch with all of the cars ranging from $20,000 to $30,000. Ford cars are also starting from the same range, however, they do offer cars finishing at prices just shy of $75,000. And BMW cars have the highest average prices and have cars ranging to $100,000.



This graph gives us the price range of cars based on their body types. As we can see, the most affordable cars are the hatchbacks and the majority of the cars finish just above $30,000. Other cars are in a similar range of $25,000 to $60,000 with only convertible cars having overall pricer models ranging till $70,000. The sedan class has the most dynamic outliers, with a car priced at $20,000 which is substantially lower than its normal range and one car priced at around $100,000 which is the most expensive vehicle in the dataset.

Recommendations

- Price- When it comes to price, we can conclude that the majority of the vehicles in the dataset are under the price of $40,000, which is the threshold limit or the average price of all the vehicles. This suggests that cheaper cars are more in demand, especially for Ford and Toyota. As for BMW, the optimum price range is in between $30k-$60k.
- Dimensions- The length and size of cars for the consumers vary based on the type of car. Apart from trucks, the majority of cars are shorter than the average dimensions of the vehicles present in the dataset. Sedans are also long, but not as much as the trucks. Thus, the average size of the vehicles should be on the lower side, except for trucks, which are bigger.
- Performance- As for the performance of the vehicle, it hugely depends on the fuel type and gas is the most preferred fuel type in the dataset. However, Toyota hybrid vehicles are the best in class when it comes to mileage and fuel consumption. This shows that the vehicles based on electric engines are much better in terms of fuel consumption and reduced emissions. With the advent of the popularity of electric vehicles and the growing environmental consciousness of consumers, we anticipate their preferences to change and a shift toward electric vehicles.
- Design- In terms of the design of vehicles, metallic colored cars specifically stood out to us. Given the fact that metallic colors are more expensive than factory finished matte colors, it shows that customers prefer a shinier finish to their cars. This was an interesting trend that showed us that the preference of exterior paint of vehicles is skewed more towards archaic metallic rather than modern matte.