

## **Automobile Data Analysis**

### **Scenario**

You are a data analyst working with a used car dealership startup venture. The investors sent a survey to prospective customers and found that the most important detail customers are looking for in a car is high gas mileage. Your task is to identify which car types have the highest gas mileage.

### **Ask/Business Task**

Identify which types of cars have the highest gas mileage.

### **Prepare/Understanding the Dataset**

This is data from an external source that contains historical sales data on car prices and their features.

Contains the following columns:

- make
- fuel\_type
- number\_of\_doors
- body\_style
- drive\_wheels
- engine\_location
- wheel\_base
- length
- width
- height
- curb\_weight
- engine\_type
- num\_of\_cylinders
- engine\_size
- fuel\_system
- compression\_ratio
- horse\_power
- city\_mpg
- highway\_mpg
- price

[Data's description](#)

## Process

### Clean the Data

1. Fill in Missing Data

To begin, I checked **each** column for NULL values. I ran the following query multiple times, replacing the WHERE clause with the specific column.

```
SELECT *
```

```
FROM shrivastava.automobile_data.automobile
```

```
WHERE num_of_doors IS NULL
```

### Results

| Row | make  | fuel_type | num_of_doors | body_style | drive_wheels | engine_location |
|-----|-------|-----------|--------------|------------|--------------|-----------------|
| 1   | dodge | gas       | <i>null</i>  | sedan      | fwd          | front           |
| 2   | mazda | diesel    | <i>null</i>  | sedan      | fwd          | front           |

When checking the num\_of\_doors column, it returned the following NULL results. In this scenario, it instructed me to fill in these missing values with the value “four”. To do this I ran the following queries:

### Update Table

```
UPDATE
```

```
shrivastava.automobile_data.automobile
```

```
SET
```

```
num_of_doors = "four"
```

```
WHERE
```

```
make = "dodge"
```

```
AND fuel_type = "gas"
```

```
AND body_style = "sedan";
```

```
UPDATE
```

```
shrivastava.automobile_data.automobile
```

```
SET
```

```
num_of_doors = "four"
```

```
WHERE
```

```
make = "mazda"
```

```
AND fuel_type = "diesel"
```

```
AND body_style = "sedan";
```

To double check that the table was updated, I then ran this query again.

```
SELECT *  
FROM shrivastava.automobile_data.automobile  
WHERE num_of_doors IS NULL
```

## Results

 There is no data to display.

## 2. Check for Potential Errors

It is important when cleaning data to check for potential errors. I used the SELECT DISTINCT clause to see what values exist in each column.

EX:

```
SELECT  
  DISTINCT num_of_cylinders  
FROM shrivastava.automobile_data.automobile
```

## RESULT

| Row | num_of_cylinders |
|-----|------------------|
| 1   | four             |
| 2   | six              |
| 3   | five             |
| 4   | three            |
| 5   | twelve           |
| 6   | two              |
| 7   | tow              |
| 8   | eight            |

I found that in Row 7, “two” is misspelled. To fix this error I ran the following query:

```
UPDATE  
  shrivastava.automobile_data.automobile  
SET  
  num_of_cylinders = "two"  
WHERE  
  num_of_cylinders = "tow";
```

To double check that the above query was successful, I ran the above SELECT DISTINCT query and found that the update was successful.

When using the SELECT DISTINCT query for drive\_wheels I also noticed the following error:

```
SELECT
  DISTINCT drive_wheels
FROM
  Shrivastava.automobile_data.automobile
```

| Row | drive_wheels |
|-----|--------------|
| 1   | rwd          |
| 2   | fwd          |
| 3   | 4wd          |
| 4   | 4wd          |

4wd is listed as two distinct rows meaning there may be extra whitespace in one of the iterations. To see if this is the case I used the LENGTH statement to determine the length of each string:

```
SELECT
  DISTINCT drive_wheels,
  LENGTH(drive_wheels) AS string_length
FROM
  shrivastava.automobile_data.automobile
```

I saw that some 4wd were showing 4 characters instead of 3, so I used the TRIM function to remove the extra space:

```
UPDATE
  shrivastava.automobile_data.automobile
SET
  drive_wheels = TRIM(drive_wheels)
WHERE TRUE;
```

I then double checked the results using the SELECT DISTINCT query from above.

I then used the [data description](#) to check that each column had a value in the correct range using the MIN and MAX function. For example the compression\_ratio column values should be between 7 and 23, but when I ran the query below I found that the maximum was 70.

```
SELECT
  MIN(compression_ratio) AS min_compression_ratio,
  MAX(compression_ratio) AS max_compression_ratio
FROM
  shrivastava.automobile_data.automobile
```

I then checked how many rows contained a maximum of compression\_ratio of 70 using the following query and found only one row contained this error.

```
SELECT
  COUNT(*) AS num_of_rows_to_delete
FROM
  shrivastava.automobile_data.automobile
WHERE
  compression_ratio = 70
```

In a real-life scenario I would check what the correct compression\_ratio was for this car and update the table. For this example though, since I am unable to check the correct number, I updated the compression\_ratio to 7.0 as that was likely the correct compression\_ratio. To do this I wrote the following query:

```
UPDATE
  shrivastava.automobile_data.automobile
SET
  compression_ratio = 7.0
WHERE
  compression_ratio = 70.0
```

I then double checked my data using the MAX query above and found that the rest of the data fell within the correct range.

## Analysis

Identify the cars with the Highest Gas Mileage for each body style

From my preparation I know that the five body\_styles are: hardtop, wagon, sedan, hatchback, convertible. In order to identify which cars of the same body type and cylinders have the highest gas mileage I ran the following query for each body\_style:

```
SELECT make, body_style, num_of_cylinders, city_mpg, highway_mpg
FROM
    shrivastava.automobile_data.automobile
WHERE body_style="hardtop"
ORDER BY city_mpg DESC, highway_mpg DESC
```

## Results

| Row | make          | body_style | num_of_cylinders | city_mpg | highway_mpg |
|-----|---------------|------------|------------------|----------|-------------|
| 1   | nissan        | hardtop    | four             | 31       | 37          |
| 2   | toyota        | hardtop    | four             | 24       | 30          |
| 3   | toyota        | hardtop    | four             | 24       | 30          |
| 4   | toyota        | hardtop    | four             | 24       | 30          |
| 5   | mercedes-benz | hardtop    | five             | 22       | 25          |
| 6   | porsche       | hardtop    | six              | 17       | 25          |
| 7   | porsche       | hardtop    | six              | 17       | 25          |
| 8   | mercedes-benz | hardtop    | eight            | 14       | 16          |

```
WHERE body_style="wagon"
```

## Results

| 1   | make        | body_style | num_of_cyli | city_mpg | highway_mpg |
|-----|-------------|------------|-------------|----------|-------------|
| 2   | nissan      | wagon      | four        | 31       | 37          |
| 3   | nissan      | wagon      | four        | 31       | 37          |
| 4   | toyota      | wagon      | four        | 31       | 37          |
| 5   | honda       | wagon      | four        | 30       | 34          |
| 6   | subaru      | wagon      | four        | 28       | 32          |
| 7   | toyota      | wagon      | four        | 27       | 32          |
| 8   | toyota      | wagon      | four        | 27       | 32          |
| 9   | subaru      | wagon      | four        | 25       | 31          |
| 10  | volkswagen  | wagon      | four        | 25       | 31          |
| 11  | peugot      | wagon      | four        | 25       | 25          |
| 12  | peugot      | wagon      | four        | 25       | 25          |
| 13  | dodge       | wagon      | four        | 24       | 30          |
| 14  | plymouth    | wagon      | four        | 24       | 30          |
| 15  | volvo       | wagon      | four        | 24       | 28          |
| 16  | subaru      | wagon      | four        | 23       | 29          |
| 17  | volvo       | wagon      | four        | 23       | 28          |
| 18  | subaru      | wagon      | four        | 23       | 23          |
| 19  | mercedes-be | wagon      | five        | 22       | 25          |
| 20  | audi        | wagon      | five        | 19       | 25          |
| ... |             |            |             |          |             |

WHERE body\_style="sedan"

#### Results

| 1  | make       | body_style | num_of_cyls | city_mpg | highway_mpg |
|----|------------|------------|-------------|----------|-------------|
| 2  | nissan     | sedan      | four        | 45       | 50          |
| 3  | toyota     | sedan      | four        | 38       | 47          |
| 4  | chevrolet  | sedan      | four        | 38       | 43          |
| 5  | isuzu      | sedan      | four        | 38       | 43          |
| 6  | isuzu      | sedan      | four        | 38       | 43          |
| 7  | volkswagen | sedan      | four        | 37       | 46          |
| 8  | volkswagen | sedan      | four        | 37       | 46          |
| 9  | volkswagen | sedan      | four        | 37       | 42          |
| 10 | mazda      | sedan      | four        | 36       | 42          |
| 11 | toyota     | sedan      | four        | 34       | 36          |
| 12 | volkswagen | sedan      | four        | 33       | 38          |
| 13 | subaru     | sedan      | four        | 32       | 37          |
| 14 | mazda      | sedan      | four        | 31       | 39          |
| 15 | dodge      | sedan      | four        | 31       | 38          |
| 16 | dodge      | sedan      | four        | 31       | 38          |
| 17 | mazda      | sedan      | four        | 31       | 38          |
| 18 | mazda      | sedan      | four        | 31       | 38          |
| 19 | plymouth   | sedan      | four        | 31       | 38          |
| 20 | plymouth   | sedan      | four        | 31       | 38          |

...

WHERE body\_style="hatchback"

#### Results

| 1  | make       | body_style | num_of_cyls | city_mpg | highway_mpg |
|----|------------|------------|-------------|----------|-------------|
| 2  | honda      | hatchback  | four        | 49       | 54          |
| 3  | chevrolet  | hatchback  | three       | 47       | 53          |
| 4  | toyota     | hatchback  | four        | 38       | 47          |
| 5  | chevrolet  | hatchback  | four        | 38       | 43          |
| 6  | honda      | hatchback  | four        | 38       | 42          |
| 7  | dodge      | hatchback  | four        | 37       | 41          |
| 8  | mitsubishi | hatchback  | four        | 37       | 41          |
| 9  | plymouth   | hatchback  | four        | 37       | 41          |
| 10 | toyota     | hatchback  | four        | 35       | 39          |
| 11 | dodge      | hatchback  | four        | 31       | 38          |
| 12 | dodge      | hatchback  | four        | 31       | 38          |
| 13 | honda      | hatchback  | four        | 31       | 38          |
| 14 | mazda      | hatchback  | four        | 31       | 38          |
| 15 | mazda      | hatchback  | four        | 31       | 38          |
| 16 | mitsubishi | hatchback  | four        | 31       | 38          |
| 17 | mitsubishi | hatchback  | four        | 31       | 38          |
| 18 | plymouth   | hatchback  | four        | 31       | 38          |
| 19 | toyota     | hatchback  | four        | 31       | 38          |
| 20 | toyota     | hatchback  | four        | 31       | 38          |

...

WHERE body\_style="convertible"

## Results

| Row | make          | body_style  | num_of_cylinders | city_mpg | highway_mpg |
|-----|---------------|-------------|------------------|----------|-------------|
| 1   | toyota        | convertible | four             | 24       | 30          |
| 2   | volkswagen    | convertible | four             | 24       | 29          |
| 3   | alfa-romero   | convertible | four             | 21       | 27          |
| 4   | alfa-romero   | convertible | four             | 21       | 27          |
| 5   | porsche       | convertible | six              | 17       | 25          |
| 6   | mercedes-benz | convertible | eight            | 16       | 18          |

In order to identify the cars with the highest gas mileage, I created the following query:

SELECT make, body\_style, num\_of\_cylinders, city\_mpg, highway\_mpg

FROM

shrivastava.automobile\_data.automobile

ORDER BY city\_mpg DESC, highway\_mpg DESC

After downloading the results of my queries, I created a new column in Excel titled average mpg.

| city_mpg | highway_mpg | Combined_Average_mpg |
|----------|-------------|----------------------|
| 49       | 54          | =(D2+E2)/2           |

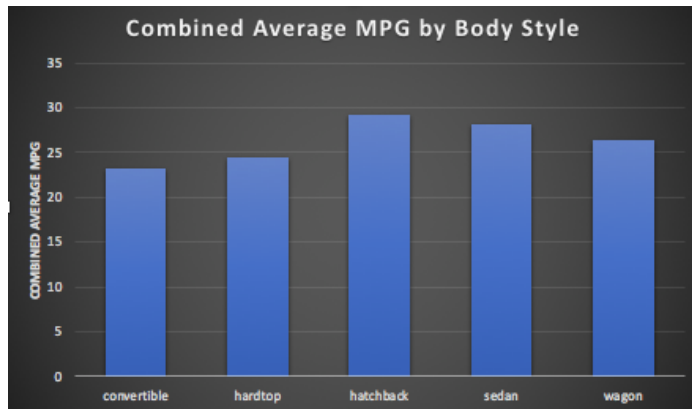
I then sorted my data by highest Combined Average mpg and found that the 10 cars with the highest combined average mpg were:

| 1  | make       | body_style | num_of_cylinders | Combined_Average_mpg |
|----|------------|------------|------------------|----------------------|
| 2  | honda      | hatchback  | 4                | 51.5                 |
| 3  | chevrolet  | hatchback  | 3                | 50                   |
| 4  | nissan     | sedan      | 4                | 47.5                 |
| 5  | toyota     | sedan      | 4                | 42.5                 |
| 6  | toyota     | hatchback  | 4                | 42.5                 |
| 7  | volkswagen | sedan      | 4                | 41.5                 |
| 8  | volkswagen | sedan      | 4                | 41.5                 |
| 9  | chevrolet  | hatchback  | 4                | 40.5                 |
| 10 | chevrolet  | sedan      | 4                | 40.5                 |

I then created the Pivot Table below, using Microsoft Excel, to identify which car body styles have the highest mpg.

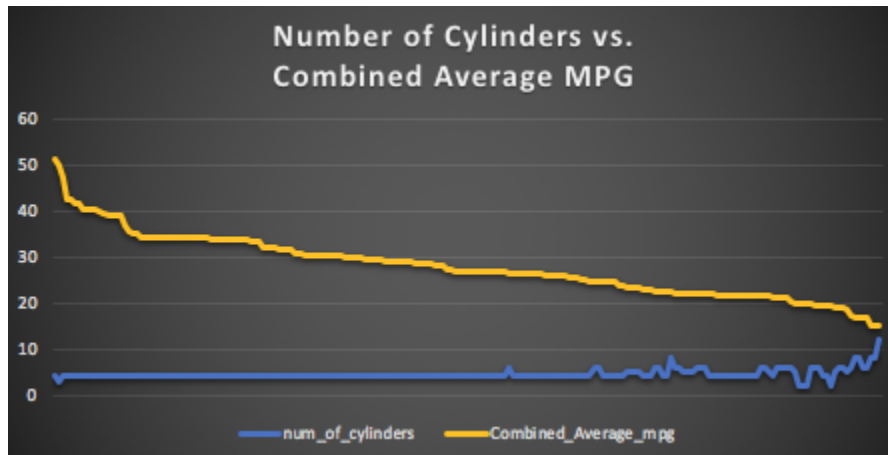
| Row Labels         | Average of Combined_Average_mpg |
|--------------------|---------------------------------|
| convertible        | 23.25                           |
| hardtop            | 24.4375                         |
| hatchback          | 29.27536232                     |
| sedan              | 28.078125                       |
| wagon              | 26.35416667                     |
| <b>Grand Total</b> | <b>27.99507389</b>              |

I then used Microsoft Excel to create a bar graph to visualize the above data.



### Identify the relationship between Number of Cylinders and Combined Average MPG

While looking at my data I noticed that the lower the number of cylinders a car had the higher the mpg usually was, so I used Microsoft Excel, to visualize this data.



### **Business Task Questions**

Identify which types of cars have the highest gas mileage.

### **Conclusion**

- Hatchbacks have the highest gas mileage, followed by Sedan, Wagon, Hardtop and Convertible
- The lower the number of cylinders, typically, the higher the mpg will be.

My recommendation would be to stock the five highest gas mileage cars for each body style. If customers are looking for high mileage cars it will be important to have a higher stock of those vehicles. It is also important to keep a wide range of body styles though, as different people have different uses for their car.