Project Description: In this scenario, I'm working with a used car dealership startup venture. I will guide you through the steps I took to clean up this dataset:

Dataset: googlesheet/rawdata

Data's description: https://archive.ics.uci.edu/dataset/10/automobile

```
#Based on the data's description there should only be 2 types of fuel, let's run a
query to find out if that's true:
SELECT
 DISTINCT fuel_type
FROM `mythic-lead-404523.auto_mobile_data.Automobile_data` LIMIT 1000
```



```
#The result shows 2 types, This confirms that the fuel_type column doesn't have any
unexpected valuesso we are good to go!


#Next, we will inspect a column with numerical data. The length column should contain numeric measurements
of the cars.
Let's run this query (results should confirm that 141.1 and 208.1 are the minimum and maximum values
respectively in this column):


SELECT
 MIN(length) AS min_length,
 MAX(length) AS max_length
FROM `mythic-lead-404523.auto_mobile_data.Automobile_data` LIMIT 1000
```

Query results

SAVE RESULTS ▼   **EXPLORE DATA** ▼ ↕

< | JOB INFORMATION | **RESULTS** | CHART | JSON | EXECUTION DETAILS | >

Row	min_length ▼	max_length ▼	
1	141.1	208.1	

Now, let's check for missing data. For example, I will check to see if the num_of_doors column contains null values using this query:

```
1  SELECT
2  *
3  FROM `mythic-lead-404523.auto_mobile_data.Automobile_data`
4  WHERE
5    num_of_doors IS NULL;
6
```

Query completed.

Press Option+F1 for Accessibility Options.

Query results

SAVE RESULTS ▼   **EXPLORE DATA** ▼ ↕

< | JOB INFORMATION | **RESULTS** | CHART | JSON | EXECUTION DETAILS | >

Row	make ▼	fuel_type ▼	num_of_doors ▼
1	dodge	gas	null
2	mazda	diesel	null

Now let's fill in those missing values:

```
UPDATE `mythic-lead-404523.auto_mobile_data.Automobile_data`
SET
 num_of_doors = "four"
WHERE
 make = "dodge"
 AND fuel_type = "gas"
```

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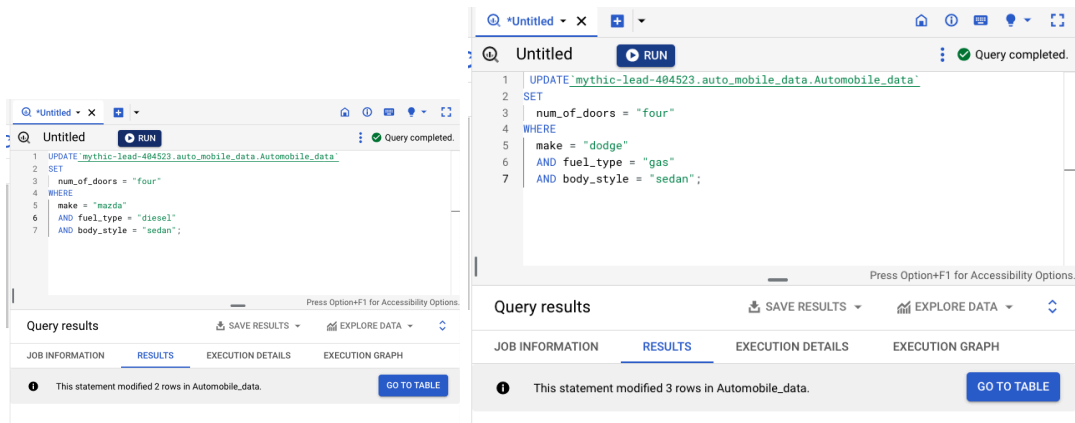

Now, let's run the previous Query again to make sure this was really updated:

```
SELECT
 *
FROM `mythic-lead-404523.auto_mobile_data.Automobile_data`
WHERE
 num_of_doors IS NULL;
```
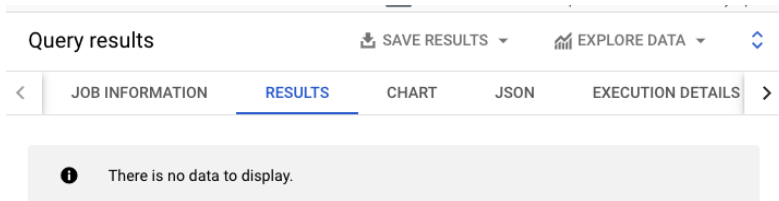

Sweet! Now let's check for potential errors:

```
SELECT
 DISTINCT num_of_cylinders
FROM `mythic-lead-404523.auto_mobile_data.Automobile_data`
```

**There are two entries for two cylinders: rows 6 and 7. But the *two* in row 7 is misspelled. **

```
1   SELECT
2    DISTINCT num_of_cylinders
3   FROM `mythic-lead-404523.auto_mobile_data.Automobile_data`
4
```

Query results

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

Let's fix that error:


```
1   UPDATE `mythic-lead-404523.auto_mobile_data.Automobile_data`
2   SET
3     num_of_cylinders = "two"
4   WHERE
5     num_of_cylinders = "tow";
6
7
```

Query results

JOB INFORMATION | RESULTS | EXECUTION DETAILS | EXECUTION GRAPH

This statement modified 1 row in Automobile_data.

GO TO TABLE

Finally, let's check our data for inconsistencies:
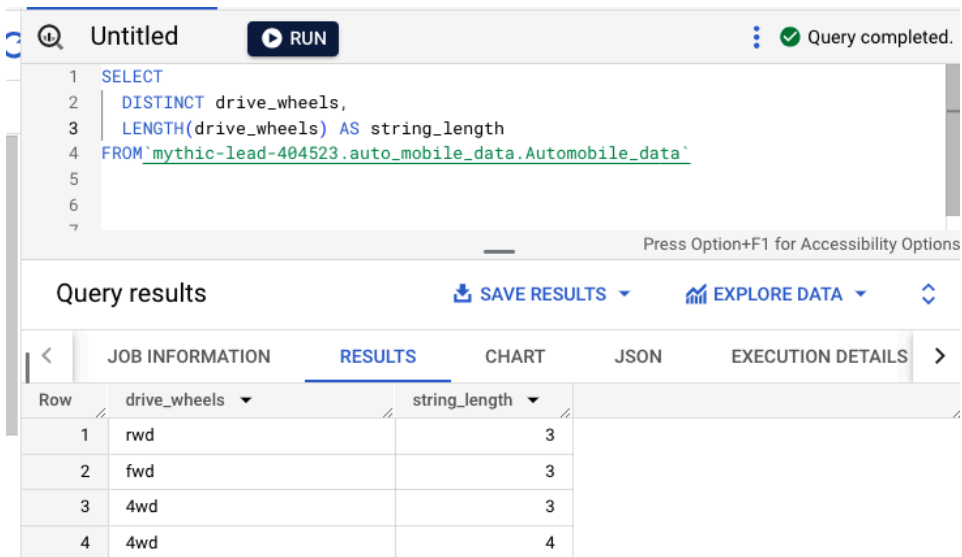
It appears that 4wd appears twice in results. However, because I used a SELECT DISTINCT statement to return unique values, this probably means there's an extra space in one of the 4wd entries that makes it different from the other 4wd. Let's check:
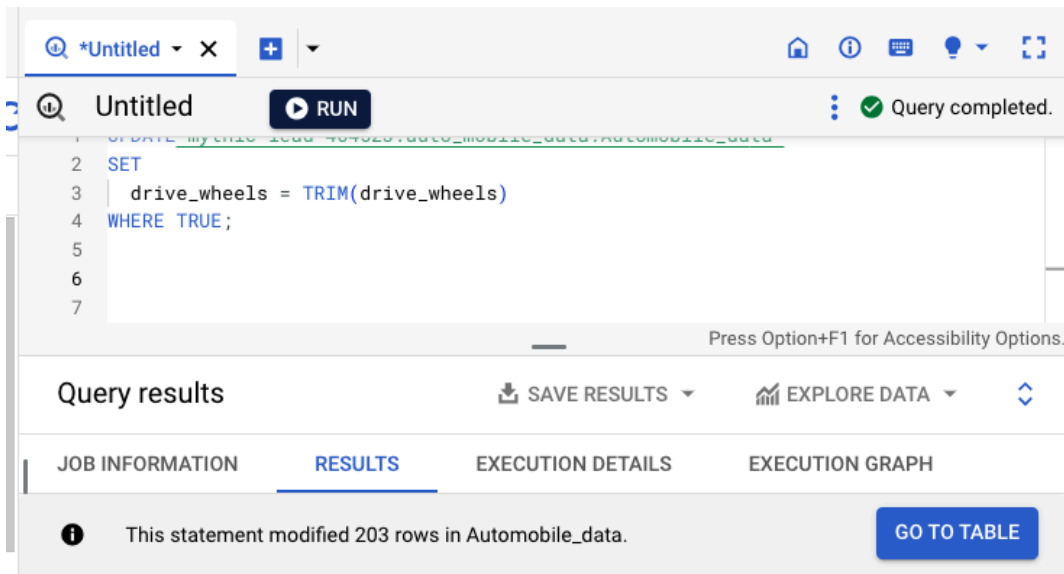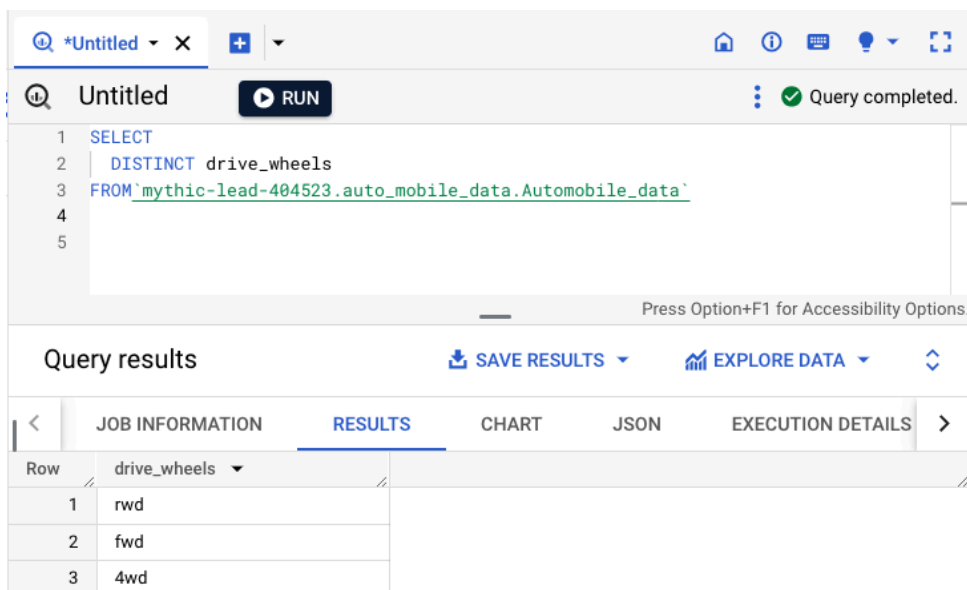
According to these results, some instances of the 4wd string have four characters instead of the expected three (4wd has 3 characters). Let's fix it:

And now there are only be three unique values in this column! This means the data is clean, consistent, and ready for analysis! :)