



#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
```




Query results		 SAVE RESULTS	 EXPLORE DATA ▾
Query complete (0.6 sec elapsed, 1 KB processed)			
Job information		Results	JSON Execution details
Row	fuel_type		
1	gas		
2	diesel		

#The result shows 2 types, This confirms that the fuel\_type column doesn't have any unexpected values so 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:

Untitled ▶ RUN Query completed.

```

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

```

Query results SAVE RESULTS EXPLORE DATA

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";

```

Untitled ▶ RUN Query completed.

```

1 UPDATE `mythic-lead-404523.auto_mobile_data.Automobile_data`
2 SET
3   num_of_doors = "four"
4 WHERE
5   make = "mazda"
6   AND fuel_type = "diesel"
7   AND body_style = "sedan";

```

Query results SAVE RESULTS EXPLORE DATA

1 This statement modified 2 rows in Automobile\_data. GO TO TABLE

Untitled ▶ RUN Query completed.

```

1 UPDATE `mythic-lead-404523.auto_mobile_data.Automobile_data`
2 SET
3   num_of_doors = "four"
4 WHERE
5   make = "dodge"
6   AND fuel_type = "gas"
7   AND body_style = "sedan";

```

Query results SAVE RESULTS EXPLORE DATA

1 This statement modified 3 rows in Automobile\_data. GO TO TABLE

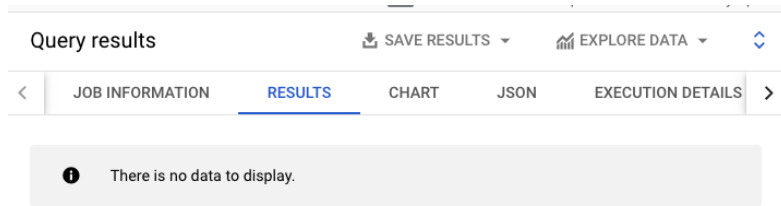
Now, let's run the previous Query again to make sure this was really updated:

```

SELECT
  *
FROM `mythic-lead-404523.auto_mobile_data.Automobile_data`
WHERE

```

```
num_of_doors IS NULL;
```



Query results

SAVE RESULTS EXPLORE DATA

JOB INFORMATION RESULTS CHART JSON EXECUTION DETAILS

There is no data to display.

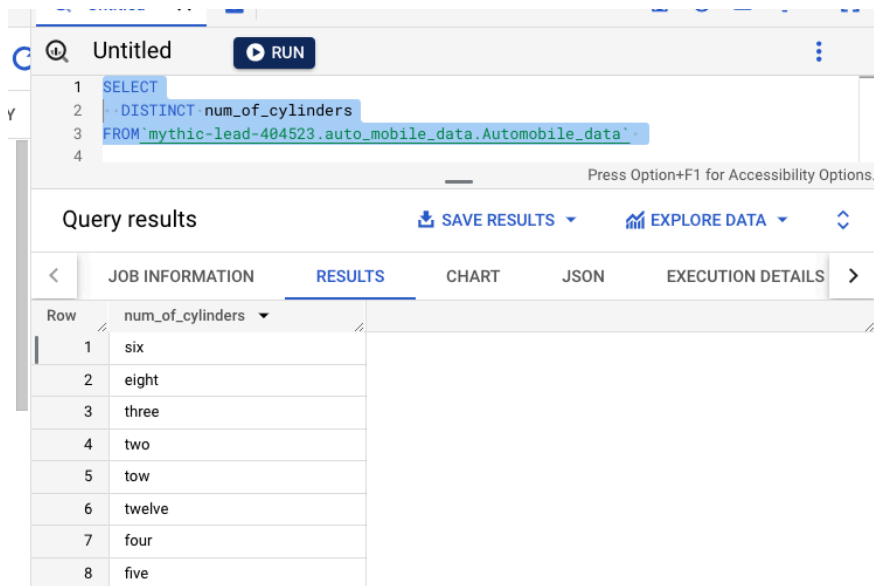
Sweet! Now let's check for potential errors:

```
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. \*\*



Untitled RUN

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

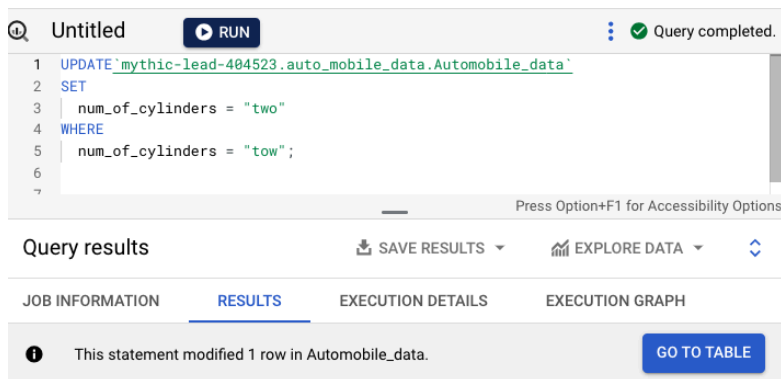
Query results

SAVE RESULTS EXPLORE DATA

JOB INFORMATION RESULTS CHART JSON EXECUTION DETAILS

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:



Untitled RUN Query completed.

```
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

SAVE RESULTS EXPLORE DATA

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:

The screenshot shows a SQL query editor with the following query:

```
1 SELECT
2   DISTINCT drive_wheels
3 FROM `mythic-lead-404523.auto_mobile_data.Automobile_data`
4
5
6
```

The query is executed, and the results are displayed in a table with the following data:

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

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:

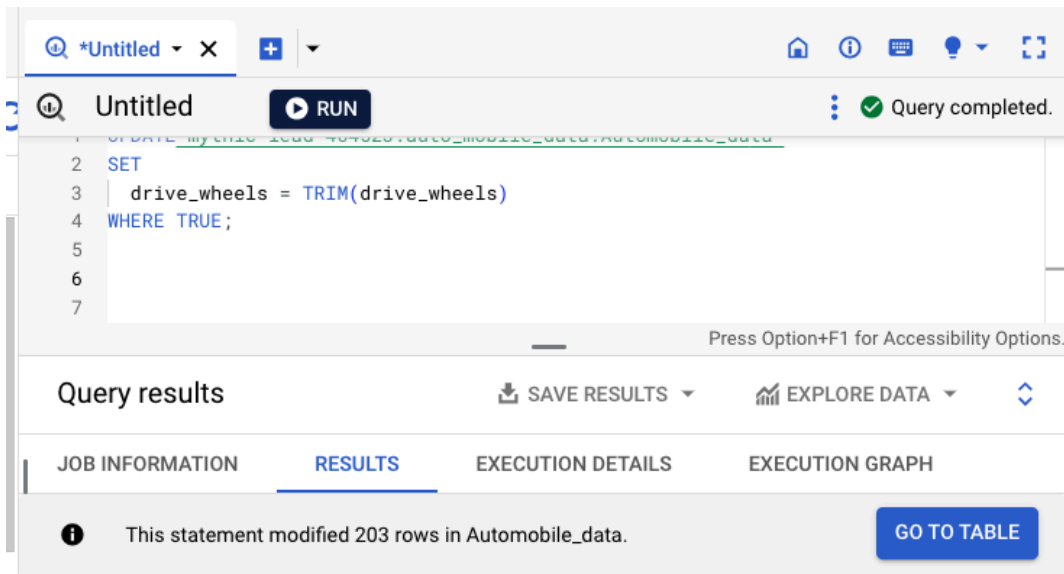
The screenshot shows a SQL query editor with the following query:

```
1 SELECT
2   DISTINCT drive_wheels,
3   LENGTH(drive_wheels) AS string_length
4 FROM `mythic-lead-404523.auto_mobile_data.Automobile_data`
5
6
7
```

The query is executed, and the results are displayed in a table with the following data:

Row	drive_wheels	string_length
1	rwd	3
2	fwd	3
3	4wd	3
4	4wd	4

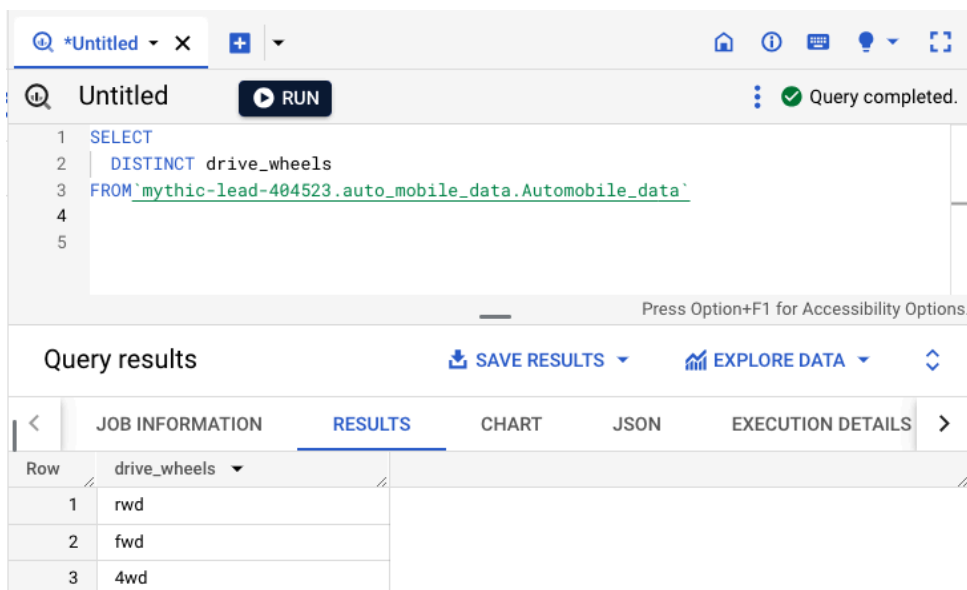
According to these results, some instances of the 4wd string have four characters instead of the expected three (4wd has 3 characters). Let's fix it:



The screenshot shows a SQL editor window titled "Untitled" with a "RUN" button. The SQL query is:

```
1 UPDATE mythic-lead-404523.auto_mobile_data.Automobile_data
2 SET
3   drive_wheels = TRIM(drive_wheels)
4 WHERE TRUE;
```

The query has completed successfully. Below the editor, the "Query results" section shows a message: "This statement modified 203 rows in Automobile\_data." and a "GO TO TABLE" button.



The screenshot shows a SQL editor window titled "Untitled" with a "RUN" button. The SQL query is:

```
1 SELECT
2   DISTINCT drive_wheels
3 FROM `mythic-lead-404523.auto_mobile_data.Automobile_data`
4
5
```

The query has completed successfully. Below the editor, the "Query results" section shows a table with 3 rows and 1 column:

Row	drive_wheels
1	rwd
2	fwd
3	4wd

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