

Data Cleaning Project Using SQL

This dataset contained historical sales data, including details such as car features and prices. I inspected the data to make sure it was clean, free from null values, errors, and inconsistencies. I also ran a few queries to make sure the dataset matched the description in the schema.

1.) The fuel_type column should only have 2 unique string values so I ran this query..

```
SELECT
    DISTINCT fuel_type
FROM
    cars.car_info;
```

2.) Then I inspected the max and min column length to make sure it matched the dataset description..

```
SELECT
    MIN(length) AS min_length,
    MAX(length) AS max_length
FROM
    cars.car_info;
```

3.) I checked for missing or null values in the num_of_doors column and updated them.

```
SELECT
    *
FROM
    cars.car_info
WHERE
    num_of_doors IS NULL;

UPDATE
    cars.car_info
SET
    num_of_doors = "four"
WHERE
    make = "dodge"
    AND fuel_type = "gas"
    AND body_style = "sedan";

SELECT
    *
FROM
    cars.car_info
```

WHERE

num_of_doors IS NULL;

4.) I then checked for potential errors and misspelled rows in the dataset and corrected them..

SELECT

DISTINCT num_of_cylinders

FROM

cars.car_info;

UPDATE

cars.car_info

SET

num_of_cylinders = "two"

WHERE

num_of_cylinders = "tow";

SELECT

DISTINCT num_of_cylinders

FROM

cars.car_info;

SELECT

MIN(compression_ratio) AS min_compression_ratio,

MAX(compression_ratio) AS max_compression_ratio

FROM

cars.car_info;

SELECT

MIN(compression_ratio) AS min_compression_ratio,

MAX(compression_ratio) AS max_compression_ratio

FROM

cars.car_info

WHERE

compression_ratio <> 70;

SELECT

COUNT(*) AS num_of_rows_to_delete

FROM

cars.car_info

WHERE

compression_ratio = 70;

```
DELETE cars.car_info
WHERE compression_ratio = 70;
```

5.) I also checked the drive_wheels column for inconsistent data and fixed those as well.

```
SELECT
  DISTINCT drive_wheels
FROM
  cars.car_info;
```

```
SELECT
  DISTINCT drive_wheels,
  LENGTH(drive_wheels) AS string_length
FROM
  cars.car_info;
```

```
UPDATE
  cars.car_info
SET
  drive_wheels = TRIM(drive_wheels)
WHERE TRUE;
```

```
SELECT
  DISTINCT drive_wheels
FROM
  cars.car_info;
```