Cleaning Dataset

#searching for duplicates
```
SELECT DISTINCT *
FROM `capstone-346915.US_Cars.cars`
```
No duplicates found

```
# looking at min and max price
SELECT min(price) as min_price, max(price) as max_price
FROM `capstone-346915.US_Cars.cars`
```
Min price was zero. Will look to see if valid. Max = 84000


```
# looking at all cars with zero price to determine if any outliers
SELECT *
FROM `capstone-346915.US_Cars.cars`
WHERE price = 0
```
43 cars listed with zero price

```
# Looking for character length in vin number to determine if any extra spaces
SELECT length(vin)
FROM `capstone-346915.US_Cars.cars`
```
#length is 19

```
#finding any cars with vin greater or less than 19
SELECT *
FROM `capstone-346915.US_Cars.cars`
WHERE length(vin) > 19
or length(vin) < 19
```
One car returned with a vin length of 15

Examining data

```
#Examining brands average price
SELECT brand, avg(price) as avg_price
FROM `capstone-346915.US_Cars.cars`
group by brand
```

#Examining number of cars in each state.

```sql
SELECT state, COUNT(*) as cars_state
FROM `capstone-346915.US_Cars.cars`
group by state
order by cars_state desc

#Examining number of car models
SELECT model, COUNT(*) as carmodel
FROM `capstone-346915.US_Cars.cars`
group by model
order by carmodel desc

#Examining average price of cars by year
SELECT year, avg(price) as avg_yearly
FROM `capstone-346915.US_Cars.cars`
group by year
order by year desc

#Examining number of cars from each brand
SELECT brand, count(*) as brand_count
FROM `capstone-346915.US_Cars.cars`
group by brand
order by brand_count desc
```