

## U.S. Car Sales Data – SQL and Excel Project

Project Description: Create SQL queries using car sales data in the U.S. to determine any potential insights. Create visualizations in Excel highlighting those same queries created in SQL.

- Link to dataset: <https://www.kaggle.com/datasets/juanmerinobermejo/us-sales-cars-dataset/data>

--

The contents of the dataset include one table and contains the following columns:

- Brand: The name of the manufacturer of the vehicle
- Model: The name of the vehicle
- Year: The year the vehicle was manufactured
- Status: Indicates whether the vehicle is new or used
- Mileage: The mileage of the vehicle at the time it was sold (new vehicles have a null value for mileage)
- Dealer: Name of the dealership that sold the vehicle
- Price: How much the vehicle was sold for

Brand	Model	Year	Status	Mileage	Dealer	Price
Mazda	MX-5 Miata	2023	New		Greenway Mazda	32950
Mazda	Mazda3	2023	New		Greenway Mazda	23703
Volkswagen	Golf	2023	New			40185
Mazda	Mazda3	2024	New		Greenway Mazda	32110
Mazda	CX-5	2023	New		Greenway Mazda	36703
Mazda	CX-5	2023	New		Greenway Mazda	41643
BMW	xD	2014	Used	94110	Sterling McCall Lexus	13295
Toyota	Highlander	2019	Used	39063	Sterling McCall Lexus	27495
Hyundai	Tucson	2017	Used	84375	Sterling McCall Lexus	16995
Lexus	RX	2015	Used	95005	Sterling McCall Lexus	19295

### Average Price of Cars Sold (Separate queries for “New” and “Used” cars)

```
SELECT Brand, COUNT(Brand) AS Number_of_Cars_Sold, ROUND(AVG(Price),2) AS  
Average_Sale_Price
```

```
FROM cars
```

```
WHERE Status = "Used"
```

```
GROUP BY Brand
```

```
ORDER BY Number_of_Cars_Sold DESC
```

	Brand	Number_of_Cars_Sold	Average_Sale_Price
►	Toyota	5457	31246.67
	Ford	5348	34518.40
	Chevrolet	3932	37620.77
	Lexus	3049	37357.07
	BMW	2961	42449.99
	Honda	2126	21772.04
	Mercedes	2092	50156.64
	Audi	2051	40459.21
	Jeep	1795	33308.67

```
SELECT Brand, COUNT(Brand) AS Number_of_Cars_Sold, ROUND(AVG(Price),2) AS
Average_Sale_Price
```

```
FROM cars
```

```
WHERE Status = "New"
```

```
GROUP BY Brand
```

```
ORDER BY Number_of_Cars_Sold DESC
```

	Brand	Number_of_Cars_Sold	Average_Sale_Price
►	Ford	10939	62190.71
	Chevrolet	9883	57544.20
	Toyota	7534	49200.72
	Jeep	6249	59783.17
	GMC	5338	72808.29
	RAM	4309	67780.15
	BMW	3714	88557.79
	Kia	3378	39602.06
	Honda	3301	39279.02

One of the immediate ways to analyze this dataset is by differentiating used car sales with that of new car sales. Just from these two queries, you can see more new cars were sold and are sold at a higher average price. Ordering by Number\_of\_Cars\_Sold rather than Average\_Sale\_Price was important in order to prevent possible errors in analysis.

	Brand	Number_of_Cars_Sold	Average_Sale_Price
►	Bugatti	1	3399000.00
	Lamborghini	85	287725.31
	Ferrari	64	256057.28
	Rolls-Royce	45	248156.07
	McLaren	31	208451.10
	Aston Martin	52	137969.06
	Bentley	96	111951.01
	Lucid	13	93054.85
	Rivian	32	87675.06
	Porsche	1222	77492.23

The more luxury/expensive car brands sell for a higher average price, but do not sell as many vehicles.

### Average Price of Cars Sold by Status, Brand, Model and Year

```
SELECT Brand, Model, Year, ROUND(AVG(Price),2) AS Average_Sale_Price
```

```
FROM cars
```

```
WHERE Status = "Used"
```

```
GROUP BY Brand, Model, Year
```

```
ORDER BY Brand ASC, Model ASC, Year ASC
```

	Brand	Model	Year	Average_Sale_Price
►	Acura	ILX	2013	13900.00
	Acura	ILX	2014	10996.00
	Acura	ILX	2015	13650.00
	Acura	ILX	2016	15298.00
	Acura	ILX	2017	17183.83
	Acura	ILX	2018	18793.25
	Acura	ILX	2019	22391.00
	Acura	ILX	2020	22487.20
	Acura	ILX	2021	24883.00

```
SELECT Brand, Model, Year, ROUND(AVG(Price),2) AS Average_Sale_Price
```

```
FROM cars
```

```
WHERE Status = "New"
```

```
GROUP BY Brand, Model, Year
```

```
ORDER BY Brand ASC, Model ASC, Year ASC
```

	Brand	Model	Year	Average_Sale_Price
►	Acura	ILX	2022	32845.00
	Acura	Integra	2023	35815.09
	Acura	Integra	2024	41464.03
	Acura	MDX	2023	64627.96
	Acura	MDX	2024	62579.36
	Acura	NSX	2022	188494.00
	Acura	RDX	2021	47825.00
	Acura	RDX	2023	49764.12
	Acura	RDX	2024	50574.88

The previous queries provided a snapshot of average price by brand. However, you can drill down further to gather more information by adding in the model and years for both used and new cars. This further highlights the differences in average price, within the same brand just different models and years when these cars were manufactured.

### **Total Revenue by Status, Brand, Model and Year**

```
SELECT Brand, Model, Year, ROUND(SUM(Price),2) AS Total_Revenue_by_Brand
FROM cars
WHERE Status = "Used"
GROUP BY Brand, Model, Year
ORDER BY Brand ASC, Model ASC, Year ASC
```

	Brand	Model	Year	Total_Revenue_by_Brand
►	Acura	ILX	2013	13900
	Acura	ILX	2014	10996
	Acura	ILX	2015	13650
	Acura	ILX	2016	45894
	Acura	ILX	2017	103103
	Acura	ILX	2018	75173
	Acura	ILX	2019	268692

```
SELECT Brand, Model, Year, ROUND(SUM(Price),2) AS Total_Revenue_by_Brand
FROM cars
WHERE Status = "New"
GROUP BY Brand, Model, Year
ORDER BY Brand ASC, Model ASC, Year ASC
```

	Brand	Model	Year	Total_Revenue_by_Brand
►	Acura	ILX	2022	32845
	Acura	Integra	2023	823747
	Acura	Integra	2024	6053748
	Acura	MDX	2023	9177171
	Acura	MDX	2024	11264285
	Acura	NSX	2022	188494
	Acura	RDX	2021	47825

Another important metric to analyze would be total revenue. This can also be further drilled down by status, brand, model and year.

### Total Revenue by Brand, Percentage of Cars Sold and Percentage of Total Revenue

```
SELECT Brand, COUNT(Brand) AS Number_of_Cars_Sold, SUM(Price) AS
Total_Revenue_by_Brand,
    ROUND(COUNT(Brand) * 100 / (SELECT COUNT(Brand) FROM cars WHERE Status
= "Used"),4) AS Percentage_of_Cars_Sold,
    ROUND(SUM(Price) * 100 / (SELECT SUM(Price) FROM cars Where Status =
"Used"),4) AS Total_Revenue_Percentage
FROM cars
WHERE Status = "Used"
GROUP BY Brand
ORDER BY Total_Revenue_by_Brand DESC
```

	Brand	Number_of_Cars_Sold	Percentage_of_Cars_Sold	Total_Revenue_by_Brand	Total_Revenue_Percentage
►	Ford	5348	11.0732	184604414	10.2521
	Toyota	5457	11.2988	170513078	9.4695
	Chevrolet	3932	8.1413	147924859	8.2151
	BMW	2961	6.1308	125694411	6.9805
	Lexus	3049	6.3130	113901720	6.3256
	Mercedes	2092	4.3315	104927682	5.8272

```
SELECT Brand, COUNT(Brand) AS Number_of_Cars_Sold, SUM(Price) AS
Total_Revenue_by_Brand,
    ROUND(COUNT(Brand) * 100 / (SELECT COUNT(Brand) FROM cars WHERE Status
= "New"),4) AS Percentage_of_Cars_Sold,
    ROUND(SUM(Price) * 100 / (SELECT SUM(Price) FROM cars Where Status =
"New"),4) AS Total_Revenue_Percentage
FROM cars
WHERE Status = "New"
```

GROUP BY Brand

ORDER BY Total\_Revenue\_by\_Brand DESC

	Brand	Number_of_Cars_Sold	Percentage_of_Cars_Sold	Total_Revenue_by_Brand	Total_Revenue_Percentage
►	Ford	10939	12.8612	680304202	13.0011
	Chevrolet	9883	11.6197	568709305	10.8684
	GMC	5338	6.2760	388650632	7.4274
	Jeep	6249	7.3471	373585006	7.1395
	Toyota	7534	8.8579	370678233	7.0839
	BMW	3714	4.3666	328903637	6.2856

This query is useful for creating a snapshot of total revenue by brand. When we create a subquery, we can also find the percentages of the Number\_of\_Cars\_Sold and Total\_Revenue\_by\_Brand. This adds more information detailing market share of cars sold based on the brand that sold them.

### Average Price of Cars Sold Based on Mileage

SELECT

CASE

WHEN Mileage BETWEEN 0 AND 80000 THEN "Low Miles"

WHEN Mileage BETWEEN 80001 AND 150000 THEN "Medium Miles"

WHEN Mileage > 150000 THEN "High Miles"

ELSE "N/A"

END AS Miles\_Status,

ROUND(AVG(Price),2) AS Average\_Price, COUNT(Brand) AS Number\_of\_Cars\_Sold,

ROUND(COUNT(Brand) \* 100 / (SELECT COUNT(Brand) FROM cars WHERE Status = "Used"),2) AS Percentage\_of\_Cars\_Sold

FROM cars

WHERE Status = "Used"

GROUP BY Miles\_Status

	Miles_Status	Average_Price	Number_of_Cars_Sold	Percentage_of_Cars_Sold
►	Low Miles	43618.09	36361	75.29
	Medium Miles	19322.06	9824	20.34
	High Miles	11760.25	2112	4.37

## Total Revenue of Cars Sold Based on Mileage

```
SELECT
CASE
    WHEN Mileage BETWEEN 0 AND 80000 THEN "Low Miles"
    WHEN Mileage BETWEEN 80001 AND 150000 THEN "Medium Miles"
    WHEN Mileage > 150000 THEN "High Miles"
    ELSE "N/A"
END AS Miles_Status,
    ROUND(SUM(Price),2) AS Total_Revenue, COUNT(*) AS Number_of_Cars_Sold,
    ROUND(COUNT(*) * 100 / (SELECT COUNT(*) FROM cars WHERE Status =
    "Used"),2) AS Percentage_of_Cars_Sold
FROM cars
WHERE Status = "Used"
GROUP BY Miles_Status
```

	Miles_Status	Total_Revenue	Number_of_Cars_Sold	Percentage_of_Cars_Sold
►	Low Miles	1585997372	36361	75.29
	Medium Miles	189819881	9824	20.34
	High Miles	24837655	2112	4.37

When we create a case statement based off of mileage when used cars were sold, we can also analyze the average price as well as the total revenue of these used cars. Lower mileage will sell for more on average, but it's also important to see what cars generally sell for based on their mileage. Also, it's important to see how much market share a used car has based on its mileage.

--

Once we've created the SQL queries, we can also create visualizations in Excel of these same queries. Once we grabbed the imported the data to Excel we first created some calculated columns and measures in Power Pivot in order to create some Pivot Tables and Pivot Charts.

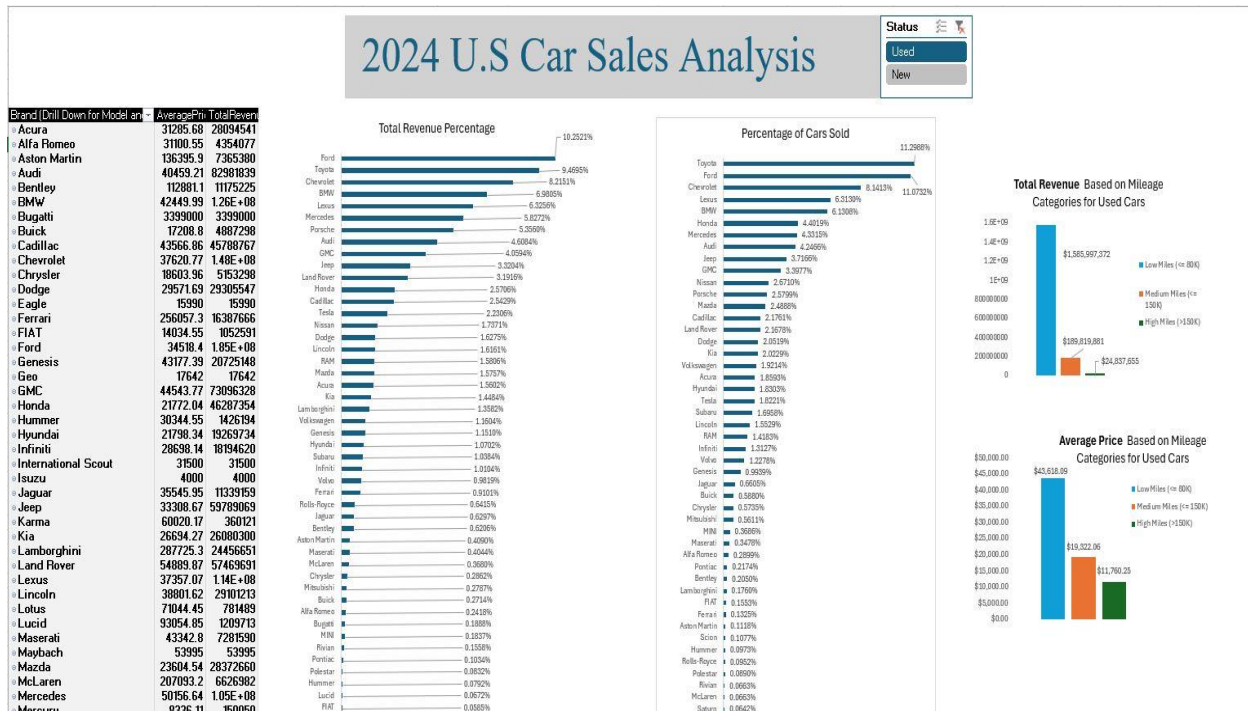
## Power Pivot Formulas for Calculated Columns and Measures

- AveragePrice:=ROUND(AVERAGE(Table1\_3[Price]),2)
- TotalRevenue:=ROUND(SUM(Table1\_3[Price]),4)
- NumberOfCarsSold:=COUNT(Table1\_3[Brand])
- Miles\_Status:=SWITCH(TRUE(),Table1\_3[Mileage] = 0 && Table1\_3[Status] = "New", "N/A",Table1\_3[Mileage]>= 0 && Table1\_3[Mileage] <=80000 && Table1\_3[Status] = "Used", "Low Miles", Table1\_3[Mileage] > 80000 && Table1\_3[Mileage] <=150000 &&



Table1\_3[Status] = "Used", "Medium Miles", Table1\_3[Mileage] > 150000 &&  
Table1\_3[Status] = "Used", "High Miles", "none")

- TotalRevenuePercentagebyBrand:=varPercentageofAllSales  
CALCULATE([TotalRevenue],ALL(Table1\_3[Brand])) RETURN  
DIVIDE([TotalRevenue],PercentageofAllSales)
- PercentageofCarsSoldbyBrand:=var TotalPercentageofCarsSold =  
CALCULATE([NumberofCarsSold], ALL(Table1\_3[Brand])) RETURN  
DIVIDE([NumberofCarsSold],TotalPercentageofCarsSold)



This is the final dashboard. The pivot table on the left side shows the average price and total revenue, but can be drilled down further to show average price and total revenue of the model and manufactured year of each of these brands.

There are two percentage of total bar charts in descending order. One shows the percentage of total revenue each brand is responsible for as well as the percentage of cars sold by each brand.

There are two column bar charts that show total revenue and average price based on the different mileage status designations the used cars were assigned.

- Low miles = <80K
- Medium miles = >80K and <150K
- High miles = >150K

All of these charts and tables can also be filtered by the status of the vehicles sold; new or used.



## **Conclusions**

Looking at both the results of the SQL queries and the visualizations designed in Excel, used cars with less than 80K miles, while more expensive, are sold at drastically higher amount than used cars with more miles. Customers are willing to pay more for a used cars with lower miles.

The car brands that tend to sell more cars also have higher revenue totals than those who sell less cars, even luxury car brands with high average sale prices gather less total revenue overall.