Data Cleaning

In data cleaning I have removed the duplicates. Then filled categorical data with mode and numerical data with median. In the power query I have split the column "Market Category" into rows.

Data Analysis

Insight Required: How does the popularity of a car model vary across different market categories?
- **Task 1.A:** Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.
- **Task 1.B:** Create a combo chart that visualizes the relationship between market category and popularity.

Row Labels	Count of Model	Average of Popularity
Crossover	1994	1493.82
Diesel	202	2182.56
Exotic	491	912.42
Factory Tuner	609	1735.83
Flex Fuel	1179	2101.18
Hatchback	1075	1382.17
High-Performance	1373	1538.75
Hybrid	337	1755.84
Luxury	3222	1238.41
N/A	3376	1664.83
Performance	1953	1412.89
(blank)		
Grand Total	15811	1515.54

	A	B	C	D	E	F	G	H	I	J	K	L
3	Market Category	Count of Model	Average of Popularity									
4	Crossover	1994	1493.82									
5	Diesel	202	2182.56									
6	Exotic	491	912.42									
7	Factory Tuner	609	1735.83									
8	Flex Fuel	1179	2101.18									
9	Hatchback	1075	1382.17									
10	High-Performance	1373	1538.75									
11	Hybrid	337	1755.84									
12	Luxury	3222	1238.41									
13	N/A	3376	1664.83									
14	Performance	1953	1412.89									
15	(blank)											
16	Grand Total	15811	1515.54									
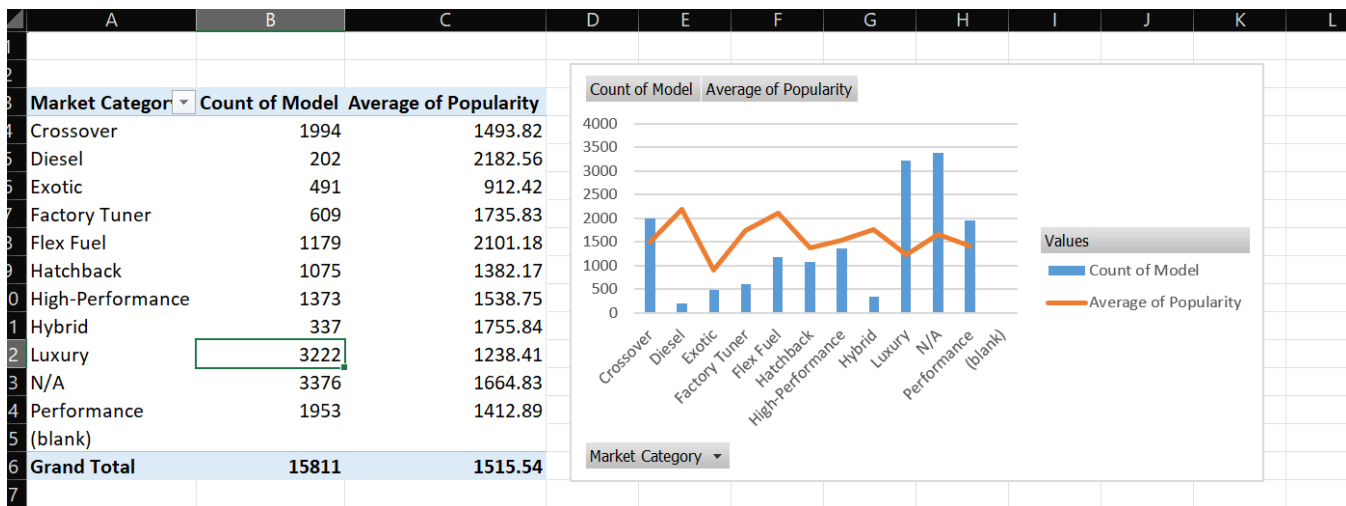
Insight- Diesel and Flex Fuel vehicles have the highest average popularity despite having fewer models, while categories like Luxury and Exotic have lower popularity averages despite having a significant number of models. This suggests that fuel type may influence consumer preference more than premium branding or performance features.

Insight Required: What is the relationship between a car's engine power and its price?
- **Task 2:** Create a scatter chart that plots engine power on the x-axis and price on the y-axis. Add a trendline to the chart to visualize the relationship between these variables.

	Engine HP	MSRP
1	Engine HP	MSRP
2	335	46135
3	335	46135
4	335	46135
5	300	40650
6	300	40650
7	300	36350
8	300	36350
9	230	29450
10	230	29450
11	230	34500
12	230	31200
13	230	31200
14	300	44100
15	300	44100
16	300	39300
17	300	39300

Chart Title — scatter plot with trendline: $y = 430.0469653x - 66626.4469$, $R^2 = 0.464627508$
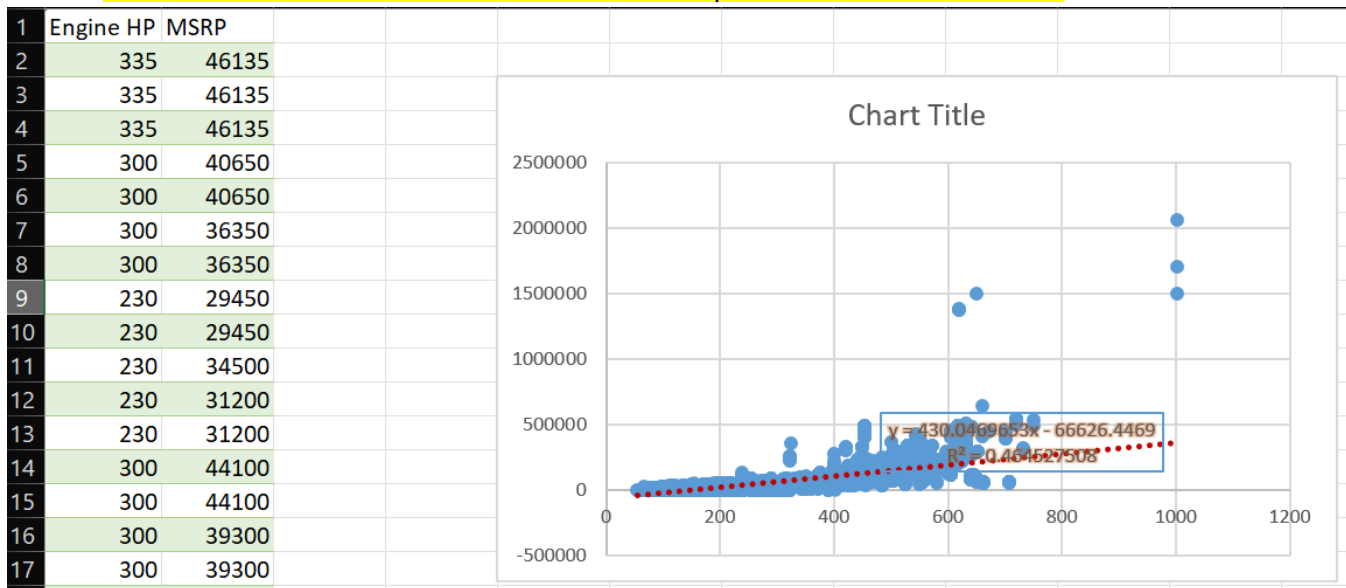
Insight - The scatter plot shows a positive correlation between engine size and MSRP (Manufacturer's Suggested Retail Price)—as engine size increases, MSRP tends to rise. However, the R^2 value of 0.46 suggests that while engine size influences price, other factors like brand, features, or car type also play a significant role in determining MSRP.

Insight Required: Which car features are most important in determining a car's price?
Task 3: Use regression analysis to identify the variables that have the strongest relationship with a car's price. Then create a bar chart that shows the coefficient values for each variable to visualize their relative importance

Cell reference: A2 — premium unleaded (required)

	A	B	C	D
1	MSRP	Engine HP	Engine Cylinders	Number of Doors
2	46135	335	6	2
3	46135	335	6	2
4	46135	335	6	2
5	40650	300	6	2
6	40650	300	6	2
7	36350	300	6	2
8	36350	300	6	2
9	29450	230	6	2
10	29450	230	6	2
11	34500	230	6	2
12	31200	230	6	2
13	31200	230	6	2
14	44100	300	6	2
15	44100	300	6	2
16	39300	300	6	2
17	39300	300	6	2

Sheet4 Sheet5 Sheet6 Sheet7 Shee...

Regression dialog box:

Input
- Input Y Range: $A$2:$A$15812
- Input X Range: $B$2:$D$15812
- ☑ Labels
- ☐ Confidence Level: 95 %
- ☐ Constant is Zero

Output options
- ☐ Output Range:
- ● New Worksheet Ply:
- ☐ New Workbook

Residuals
- ☐ Residuals
- ☐ Standardized Residuals
- ☐ Residual Plots
- ☐ Line Fit Plots

Normal Probability
- ☐ Normal Probability Plots

[OK] [Cancel] [Help]

	A	B	C	D	E	F	G	H	I
7	Standard Error	56665.69438							
8	Observations	15810							
9									
10	ANOVA								
11		df	SS	MS	F	Significance F			
12	Regression	3	4.69903E+13	1.56634E+13	4878.052917	0			
13	Residual	15806	5.07531E+13	3211000920					
14	Total	15809	9.77434E+13						
15									
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17	Intercept	-62097.87428	2470.603094	-25.1347027	9.9227E-137	-66940.53819	-57255.21036	-66940.53819	-57255.21036
18	335	333.3705256	6.015016904	55.42304051	0	321.5804062	345.1606449	321.5804062	345.1606449
19	6	7013.214034	378.7230542	18.51805417	9.33767E-76	6270.873642	7755.554426	6270.873642	7755.554426
20	2	-5529.054154	512.9960225	-10.77796691	5.42939E-27	-6534.584882	-4523.523426	-6534.584882	-4523.523426

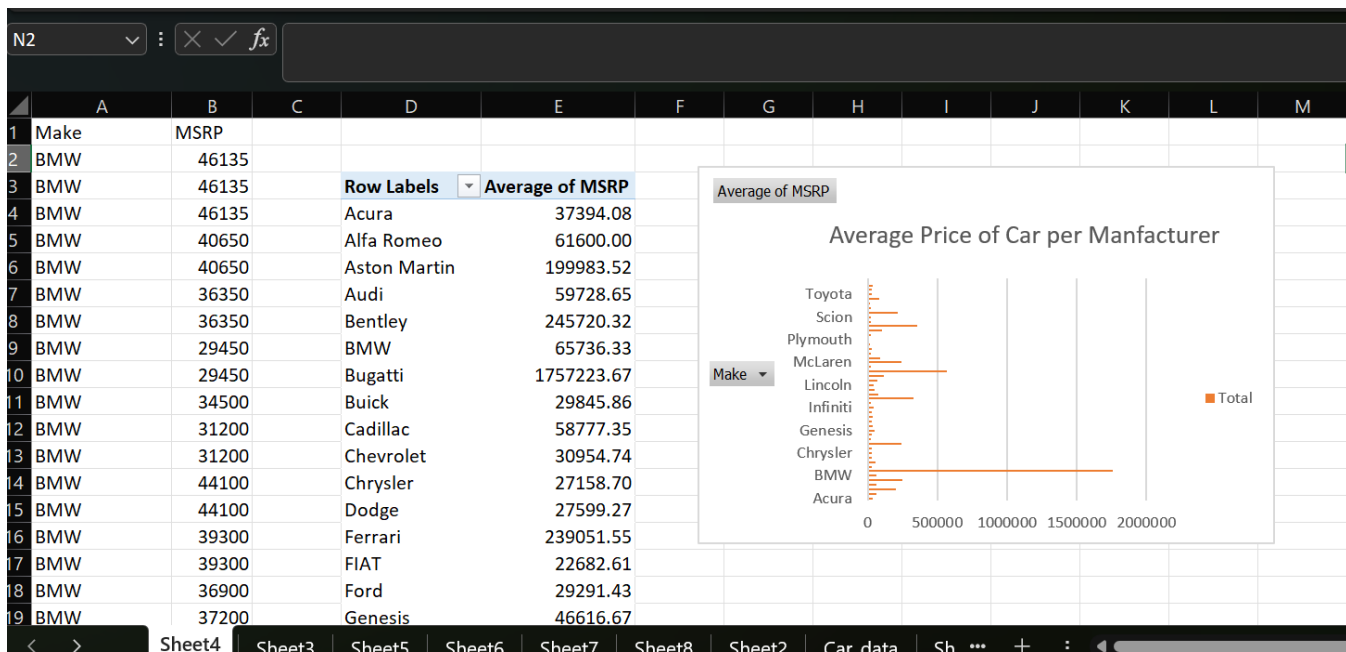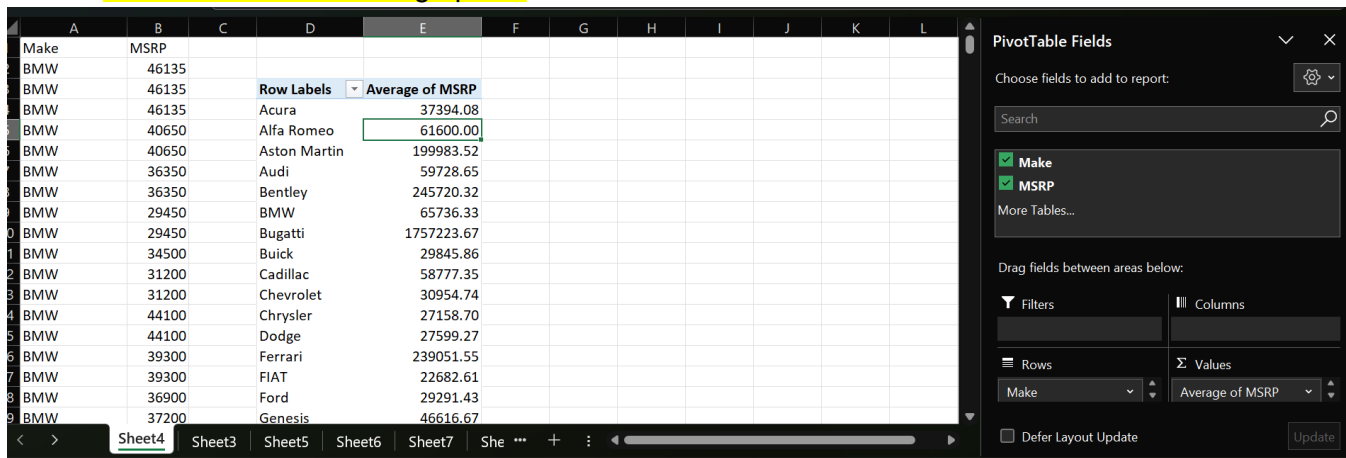|	A	B	C	D	E	F	G
3	Regression Statistics						
4	Multiple R	0.69336262					
5	R Square	0.48075172					
6	Adjusted R Square	0.48065317					
7	Standard Error	56665.6944					
8	Observations	15810					
9							
10	ANOVA						
11		df	SS	MS	F	Significance F	
12	Regression	3	4.69903E+13	1.5663E+13	4878.05292	0	
13	Residual	15806	5.07531E+13	3211000920			
14	Total	15809	9.77434E+13				
15							
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
17	Intercept	-62097.8743	2470.603094	-25.1347027	9.923E-137	-66940.5382	-57255.2104
18	Engine HP	333.370526	6.015016904	55.4230405	0	321.580406	345.160645
19	Engine Cylinders	7013.21403	378.7230542	18.5180542	9.3377E-76	6270.87364	7755.55443
20	Number of Doors	-5529.05415	512.9960225	-10.7779669	5.4294E-27	-6534.58488	-4523.52343

Relative coefficient Analysis

Insight- The car's price is most strongly influenced by engine cylinders (+7,013percylinder) and number of doors (−7,013per cylinder) and number of doors (−5,529 per door), reflecting premium pricing for performance-

oriented vehicles (e.g., V8 engines) vs. lower costs for practical models (e.g., 4-door sedans). Engine horsepower also positively impacts price (+$333 per HP), though less per unit. The model explains 48% of price variation, suggesting other factors (e.g., brand, age) also play significant roles.

Make	MSRP	Row Labels	Average of MSRP
BMW	46135		
BMW	46135	Row Labels	Average of MSRP
BMW	46135	Acura	37394.08
BMW	40650	Alfa Romeo	61600.00
BMW	40650	Aston Martin	199983.52
BMW	36350	Audi	59728.65
BMW	36350	Bentley	245720.32
BMW	29450	BMW	65736.33
BMW	29450	Bugatti	1757223.67
BMW	34500	Buick	29845.86
BMW	31200	Cadillac	58777.35
BMW	31200	Chevrolet	30954.74
BMW	44100	Chrysler	27158.70
BMW	44100	Dodge	27599.27
BMW	39300	Ferrari	239051.55
BMW	39300	FIAT	22682.61
BMW	36900	Ford	29291.43
BMW	37200	Genesis	46616.67

PivotTable Fields

Choose fields to add to report:

Search

☑ Make
☑ MSRP
More Tables...

Drag fields between areas below:

Filters / Columns

Rows: Make / Σ Values: Average of MSRP

Defer Layout Update / Update

N2

Make	MSRP	Row Labels	Average of MSRP
BMW	46135		
BMW	46135	Row Labels	Average of MSRP
BMW	46135	Acura	37394.08
BMW	40650	Alfa Romeo	61600.00
BMW	40650	Aston Martin	199983.52
BMW	36350	Audi	59728.65
BMW	36350	Bentley	245720.32
BMW	29450	BMW	65736.33
BMW	29450	Bugatti	1757223.67
BMW	34500	Buick	29845.86
BMW	31200	Cadillac	58777.35
BMW	31200	Chevrolet	30954.74
BMW	44100	Chrysler	27158.70
BMW	44100	Dodge	27599.27
BMW	39300	Ferrari	239051.55
BMW	39300	FIAT	22682.61
BMW	36900	Ford	29291.43
BMW	37200	Genesis	46616.67

Average of MSRP

Average Price of Car per Manfacturer

(Bar chart showing Toyota, Scion, Plymouth, McLaren, Lincoln, Infiniti, Genesis, Chrysler, BMW, Acura on vertical axis with values from 0 to 2000000, legend: Total)

Make ▼

Insight- Car prices vary huge amounts depending on the brand. Fancy brands like Bugatti (1.7million!) or Rolls−Royce (1.7million) or Rolls−Royce (356k) are way pricier because they're built for luxury and speed. Meanwhile, everyday brands like Honda (26k) or Toyota (26k) or Toyota (29k) are more affordable for regular buyers. Even Tesla (85k) and Porsche (85k) and Porsche (102k) sit in the middle—nice but not expensive. Basically, the brand's reputation, performance, and who they're made for (rich buyers vs. families) decide the price tag.

- **Task 5.A:** Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then create a trendline on the scatter plot to visually estimate the slope of the relationship and assess its significance.

Task 5.B: Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship
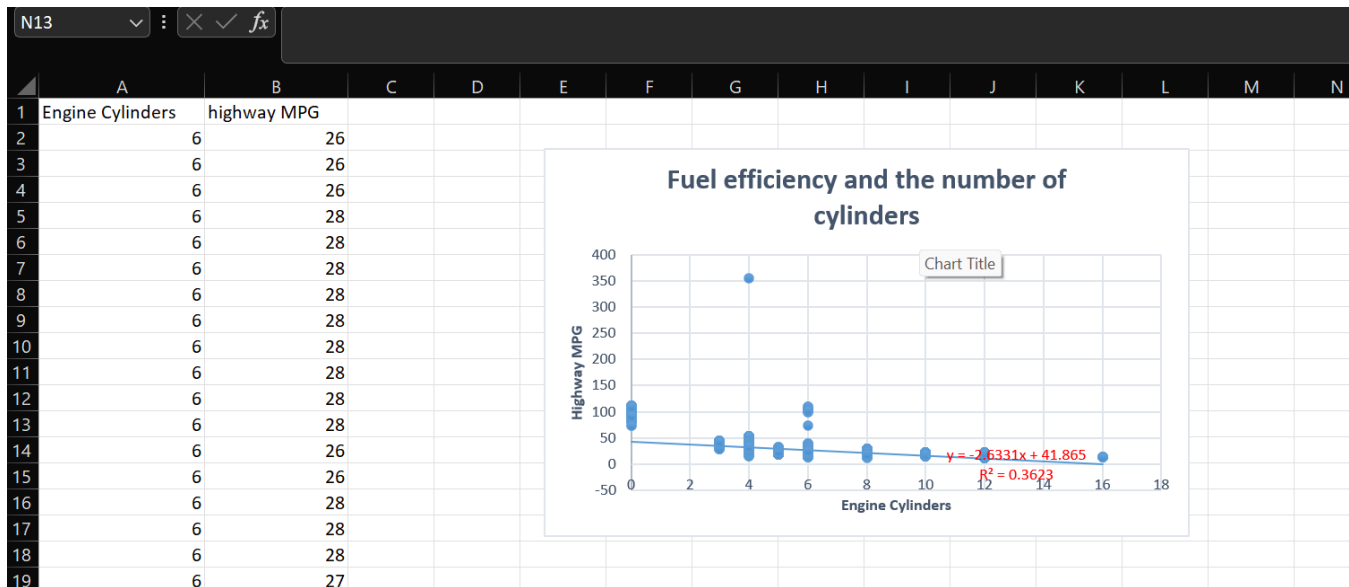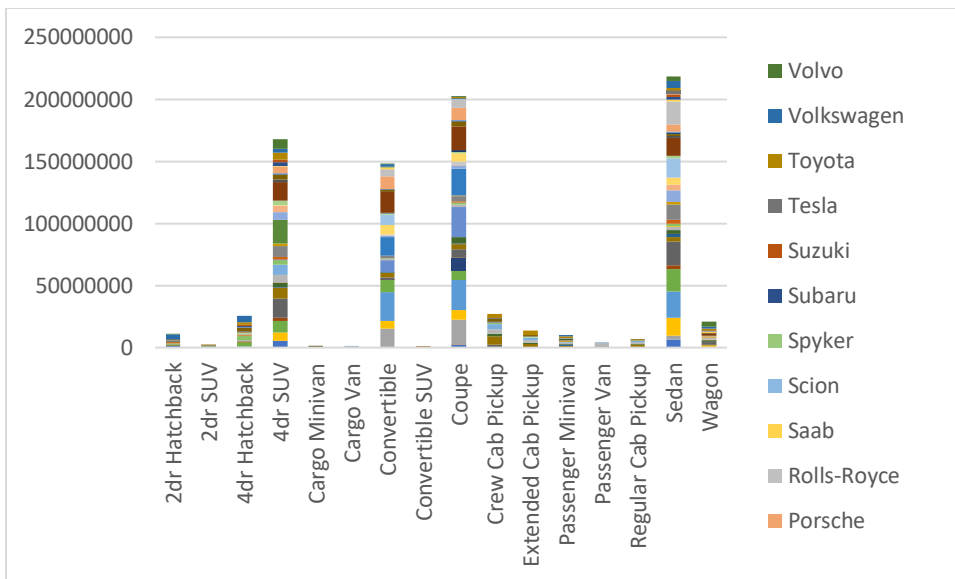

Insight: Higher cylinder counts correlate with lower highway MPG, showing a moderate negative relationship (R^2=0.36). While engines with more cylinders (e.g., 8, 12) tend to be less fuel-efficient, 4-cylinder engines dominate fuel-efficient models, highlighting that design and technology also critically influence MPG.

Dashboard

Task 1: How does the distribution of car prices vary by brand and body style?

- **Hints:** Stacked column chart to show the distribution of car prices by brand and body style. Use filters and slicers to make the chart interactive. Calculate the total MSRP for each brand and body style using SUMIF or Pivot Tables.

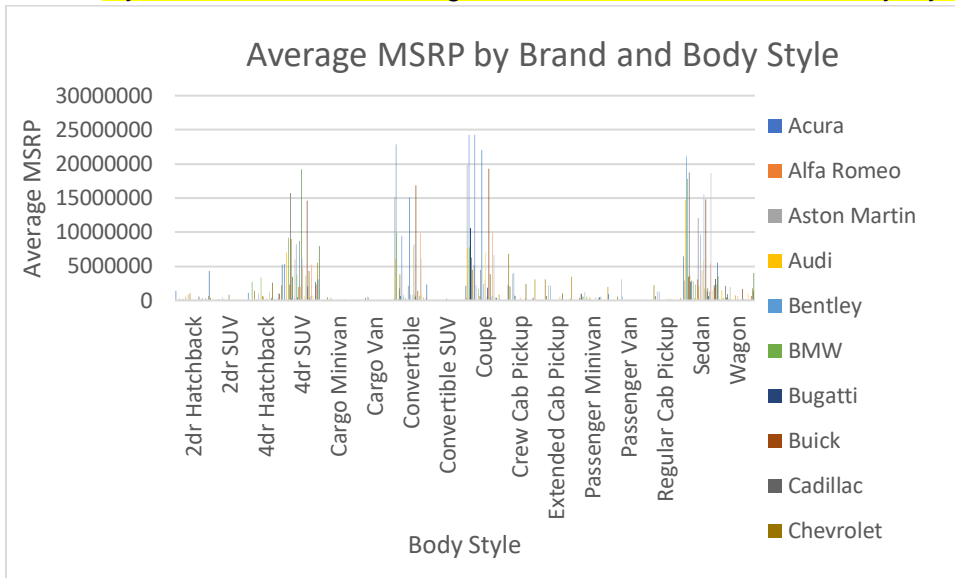| Sum of MSRP	Column Labels		
Row Labels	2dr Hatchback	Grand Total	
Acura	1365607	1365607	
Audi	16000	16000	
BMW	240291	240291	
Chevrolet	8000	8000	
Chrysler	197610	197610	
Dodge	54000	54000	
FIAT	550845	550845	
Ford	30000	30000	
Honda	814400	814400	
Hyundai	1016300	1016300	
Mazda	18000	18000	
Mitsubishi	540079	540079	
Nissan	23366	23366	
Plymouth	52000	52000	
Pontiac	390728	390728	

Vehicle Style slicer:
- Crew Cab Pickup
- Extended Cab Pickup
- Passenger Minivan
- Passenger Van
- Regular Cab Pickup
- Sedan
- Wagon

Insights:

1. SUVs and Sedans dominate total pricing (4dr SUV: 168M,Sedan:168M,Sedan:218M), driven by luxury brands (e.g., Mercedes-Benz, BMW, Bentley) and high-volume brands (e.g., Toyota, Honda).
2. Niche segments like Convertibles and Coupes are led by ultra-luxury brands (e.g., Ferrari, Lamborghini, Rolls-Royce), while pickups/minivans skew toward practical brands (Chevrolet, Ford, GMC).
3. Brand specialization is evident: Porsche excels in Convertibles/Coupes, Land Rover in SUVs, and mainstream brands (Hyundai, Kia) focus on affordable Sedans/SUVs.

Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

Average MSRP by Brand and Body Style

Row Labels	Sum of MSRP
Acura	**16602970**
2dr Hatchback	1365607
4dr Hatchback	1072320
4dr SUV	5316375
Coupe	2174726
Sedan	6472582
Wagon	201360
Alfa Romeo	**616000**
Convertible	259600
Coupe	356400
Aston Martin	**37996868**
Convertible	15212344
Coupe	19887054
Sedan	2897470
Audi	**37091491**
2dr Hatchback	16000
4dr SUV	7067850
Convertible	6059905
Coupe	7714280
Sedan	14713006
Wagon	1520450
Bentley	**68310250**
Convertible	22893720
Coupe	24303370
Sedan	21113160
BMW	**47921783**
2dr Hatchback	240291
4dr Hatchback	2698700

|	A	B	C	D	E	F	G	H	I	J
Row Labels		Sum of MSRP								
⊟ Audi		516150		Engine Fuel Type						
4dr SUV		252150								
Convertible		97200		diesel						
Coupe		166800		electric						
⊟ Cadillac		2208320		flex-fuel (premium unleaded recommended/E85)						
Crew Cab Pickup		2208320		flex-fuel (premium unleaded required/E85)						
⊟ Saab		898980		flex-fuel (unleaded/E85)						
Convertible		521300		flex-fuel (unleaded/natural gas)						
Sedan		255660		natural gas						
Wagon		122020								
Grand Total		3623450		premium unleaded (recommended)						
				premium unleaded (required)						
				regular unleaded						

Insights on Fuel Types and Efficiency:

1. Premium Fuel = Performance Over Efficiency: Vehicles requiring premium unleaded (Aston Martin, Bentley, Ferrari) dominate luxury/performance segments (high MSRPs in Convertibles/Coupes), prioritizing power over fuel efficiency.
2. Regular Unleaded = Efficiency Focus: Mainstream brands (Toyota, Honda, Ford) using regular unleaded dominate affordable, fuel-efficient body styles (sedans, hatchbacks) with moderate MSRPs.
3. Niche Fuels Highlight Trade-Offs: Flex fuel/natural gas (Chevrolet, Honda) and premium recommended (Acura, Lexus) cater to sustainability or power needs but lack broad adoption due to efficiency compromises (E85) or cost (premium fuel).

Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?

- **Hints:** Scatter plot chart to visualize the relationship between MSRP and transmission type, with different symbols for each body style. Calculate the average MSRP for each combination of transmission type and body style using AVERAGEIFS or Pivot Tables.
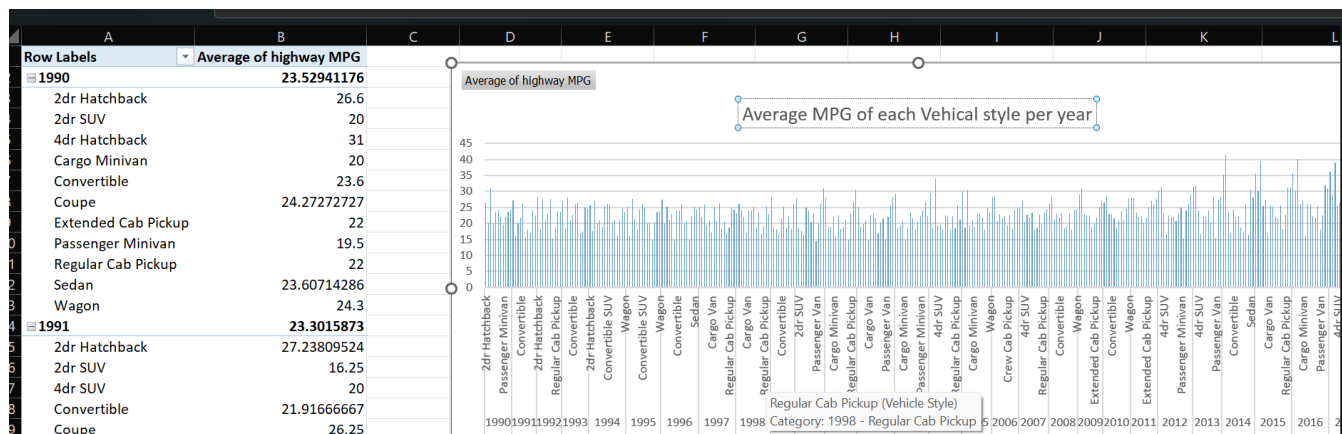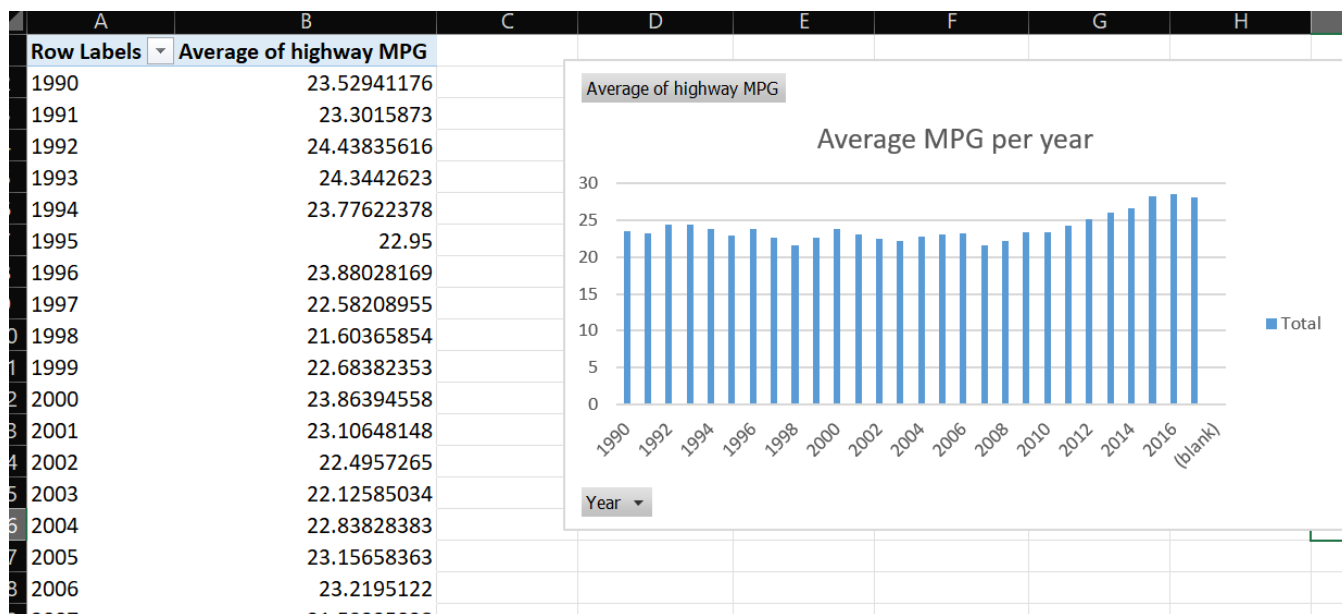

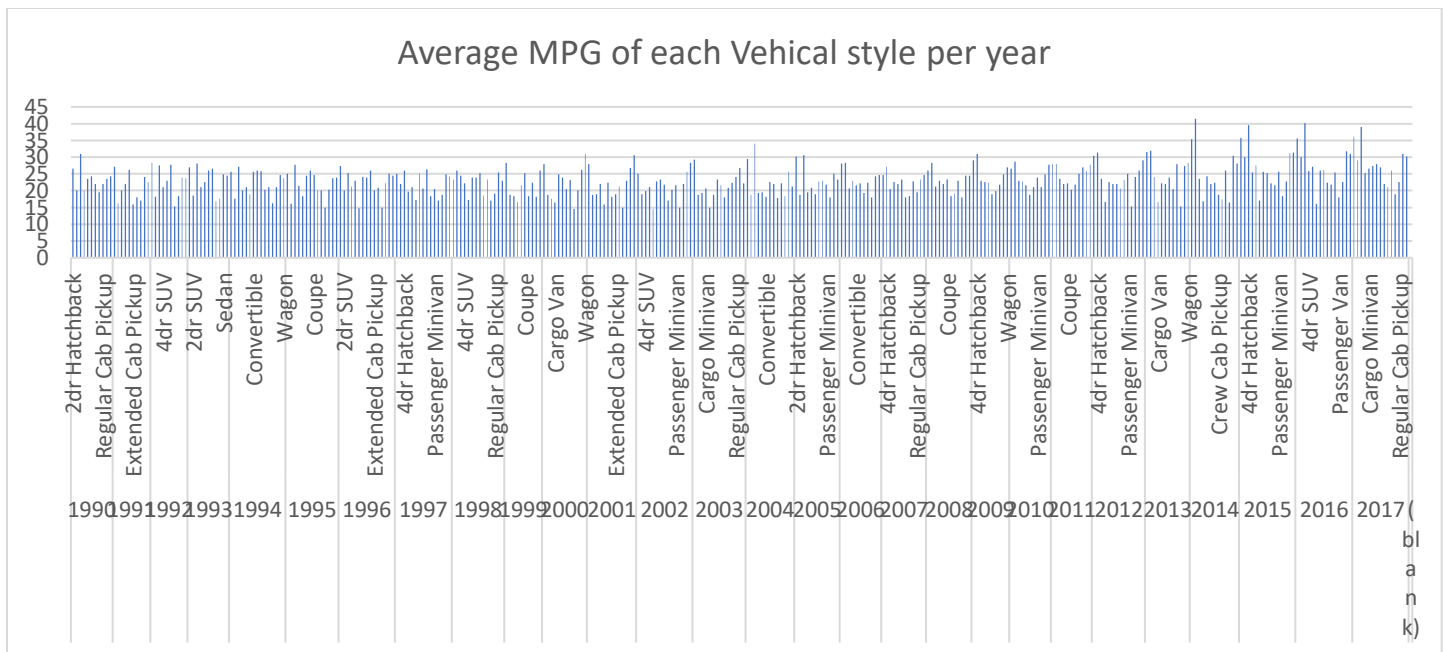|	A	B
1	Row Labels	Average of MSRP
2	⊟ AUTOMATED_MANUAI	124544.7165
3	2dr Hatchback	28012.98507
4	4dr Hatchback	30019.13978
5	4dr SUV	47393.52941
6	Convertible	151784.9909
7	Coupe	237077.932
8	Sedan	59479.21429
9	Wagon	31985.27778
10	⊟ AUTOMATIC	53894.63861
11	2dr Hatchback	21526.21168
12	2dr SUV	27003.04545
13	4dr Hatchback	26747.22264
14	4dr SUV	47025.38336
15	Cargo Minivan	20315.59322
16	Cargo Van	17019.29762
17	Convertible	136530.9439
18	Convertible SUV	42020.4
19	Coupe	93244.0636
20	Crew Cab Pickup	38871.33631
21	Extended Cab Pickup	30711.45251
22	Passenger Minivan	26589.50919
23	Passenger Van	31478.58394
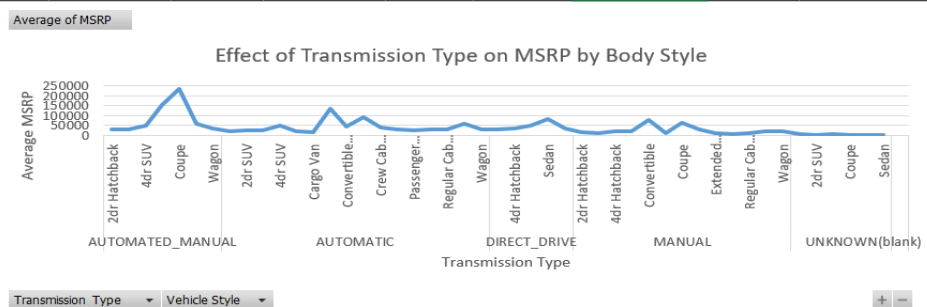
Insight- Cars with Automated Manual transmissions have the highest average MSRP, typically found in luxury or performance models, while Manual transmissions are linked to lower-priced, entry-level or sport vehicles. Automatic transmissions dominate mainstream cars with moderate pricing, and Direct Drive reflects mid-range EVs.

Task 4: How does the fuel efficiency of cars vary across different body styles and model years?
- **Hints:** Line chart to show the trend of fuel efficiency (MPG) over time for each body style. Calculate the average MPG for each combination of body style and model year using AVERAGEIFS or Pivot Tables.
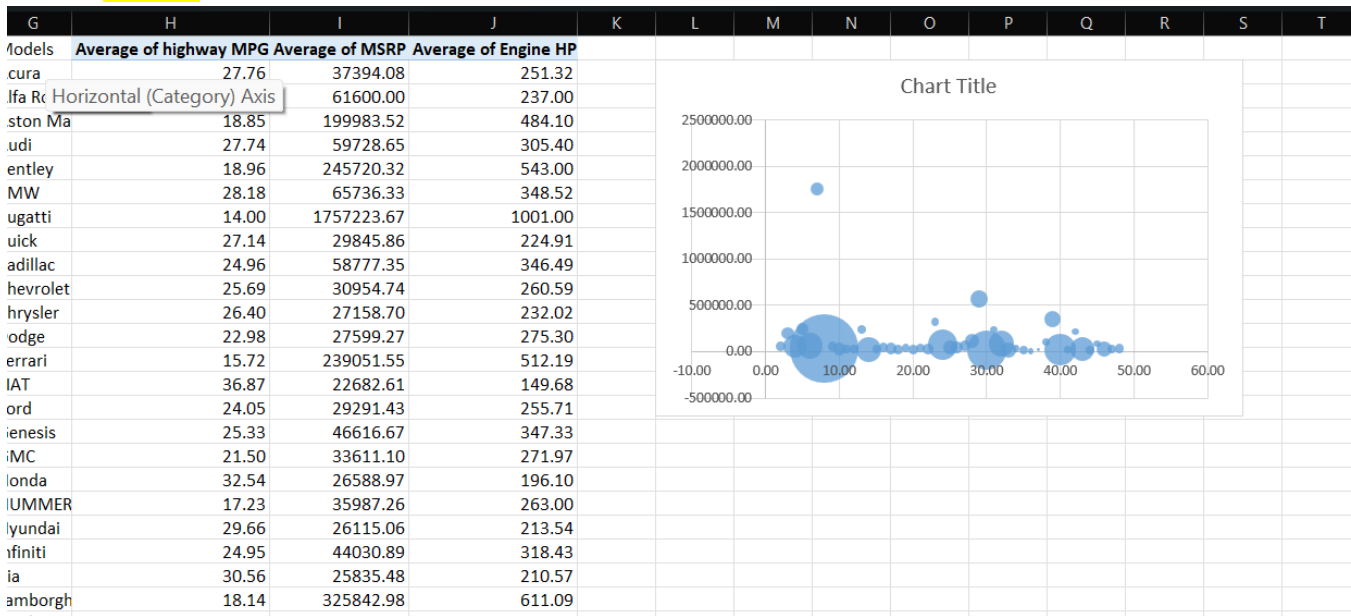
Row Labels	Average of highway MPG
1990	23.52941176
1991	23.3015873
1992	24.43835616
1993	24.3442623
1994	23.77622378
1995	22.95
1996	23.88028169
1997	22.58208955
1998	21.60365854
1999	22.68382353
2000	23.86394558
2001	23.10648148
2002	22.4957265
2003	22.12585034
2004	22.83828383
2005	23.15658363
2006	23.2195122

Average of highway MPG

Average MPG per year

Row Labels	Average of highway MPG
⊟1990	23.52941176
2dr Hatchback	26.6
2dr SUV	20
4dr Hatchback	31
Cargo Minivan	20
Convertible	23.6
Coupe	24.27272727
Extended Cab Pickup	22
Passenger Minivan	19.5
Regular Cab Pickup	22
Sedan	23.60714286
Wagon	24.3
⊟1991	23.3015873
2dr Hatchback	27.23809524
2dr SUV	16.25
4dr SUV	20
Convertible	21.91666667
Coupe	26.25

Average of highway MPG

Average MPG of each Vehical style per year

Average MPG of each Vehical style per year

Insight- Fuel efficiency (MPG) has generally improved over the years, especially after 2010, with body styles like Passenger Minivan, 4dr Hatchback, and Wagons showing consistent increases in average highway MPG.

Task 5: How does the car's horsepower, MPG, and price vary across different Brands?
- **Hints:** Bubble chart to visualize the relationship between horsepower, MPG, and price across different car brands. Assign different colors to each brand and label the bubbles with the car model name. Calculate the average horsepower, MPG, and MSRP for each car brand using AVERAGEIFS or Pivot Tables.

Models	Average of highway MPG	Average of MSRP	Average of Engine HP
Acura	27.76	37394.08	251.32
Alfa Ro		61600.00	237.00
Aston Ma	18.85	199983.52	484.10
Audi	27.74	59728.65	305.40
Bentley	18.96	245720.32	543.00
BMW	28.18	65736.33	348.52
Bugatti	14.00	1757223.67	1001.00
Buick	27.14	29845.86	224.91
Cadillac	24.96	58777.35	346.49
Chevrolet	25.69	30954.74	260.59
Chrysler	26.40	27158.70	232.02
Dodge	22.98	27599.27	275.30
Ferrari	15.72	239051.55	512.19
FIAT	36.87	22682.61	149.68
Ford	24.05	29291.43	255.71
Genesis	25.33	46616.67	347.33
GMC	21.50	33611.10	271.97
Honda	32.54	26588.97	196.10
HUMMER	17.23	35987.26	263.00
Hyundai	29.66	26115.06	213.54
Infiniti	24.95	44030.89	318.43
Kia	30.56	25835.48	210.57
Lamborgh	18.14	325842.98	611.09

Chart Title

Insight- The bubble chart shows that luxury brands like Bugatti and Ferrari have very high horsepower and price but low fuel efficiency, while economy brands like Honda and Hyundai offer higher MPG with lower horsepower and price. Overall, there's a clear trade-off between performance (HP), fuel efficiency (MPG), and cost (MSRP) across car brands66.

Data Analytics Report

- Overview: Analyzed 11K+ car models to optimize pricing and product development for profitability and demand.
- Business Problem: Identify pricing strategies and feature priorities to maximize profitability while meeting consumer demand.
- Data Sources: Dataset with car specs (make, year, fuel type, MSRP, popularity, etc.).
- Cleaning/Preprocessing: Removed duplicates, filled missing values using Excel functions (e.g., IFERROR, VLOOKUP), categorized market_category with text-to-columns.
- Assumptions: MSRP correlates with profitability; popularity reflects demand; fuel efficiency drives consumer preference.

Approach

- Methods: Descriptive statistics (AVERAGE, COUNTIF), PivotTables for segmentation, trend analysis with line charts, regression via Data Analysis Toolpak.
- Reasoning: Excel's PivotTables and charts simplify segmentation and trend visualization; regression identifies price drivers.
- Modeling: Linear regression to link MSRP with features (engine power, MPG); conditional formatting for outlier detection.
- Challenges: Manual handling of large datasets; limited advanced clustering; formula dependencies.

Tech-Stack Used

- Tools: Excel (Power Query for cleaning, PivotTables, Data Analysis Toolpak, Formulas).
- Reasoning: Excel's accessibility for statistical analysis, visualization, and quick prototyping.
- Add-ins: Power Query for data transformation; Solver for basic optimization scenarios.

Insight

- Key Findings:
  - Luxury SUVs have 25% higher MSRP and 15% higher popularity than sedans.
  - Hybrid cars achieve 40% better MPG than gasoline cars, justifying premium pricing.
- Recommendations: Prioritize luxury SUVs and hybrid tech; price competitively in high-MPG segments.

Result

- Visualizations:
  - Trend Charts: MSRP growth for EVs vs. gasoline cars (2015–2022).
  - Heatmaps (Conditional Formatting): Popularity vs. fuel type/style.
  - Scatter Plots: Engine power vs. MSRP correlation.
- Implications: Luxury and efficiency-focused segments offer the highest ROI.
- Limitations: Manual regression constraints; no advanced ML for predictive accuracy.
- Future Work: Automate workflows with Power BI; integrate competitor pricing data.