



OGTIP
Data Science
Internship

VEHICLE MARKETING ANALYSIS

Joseph Robinson - Data Analytics Professional
09 Sep 24



Visit My Website:

<https://github.com/joerobo2>



CONTENT

- 1 OVERVIEW
 - 2 EXECUTIVE SUMMARY
 - 3 DATA
 - 4 ANALYTICAL APPROACH
 - 5 ASSUMPTIONS & DISCLAIMERS
 - 6 RESEARCH ROADMAP
 - 7 IDENTIFY TRENDS
 - 8 ANALYZE IMPACT
 - 9 UNDERSTAND RELATIONSHIP
 - 10 DETERMINE POPULARITY
- ★ CONCLUSION & RECOMMENDATIONS



OVERVIEW

Business task: Analyze the dataset to uncover market trends, understand pricing dynamics, and explore factors influencing vehicle sales.

Vehicle Marketing Analysis Stages:



Deliverables

- Python Jupyter Notebook
- Presentation

Data

- One vehicle sales dataset
- 472325 rows by 16 columns.
- Selling Price target variable

Analytical Approach

- Descriptive Statistics
- Exploratory Data Analysis
- Data Visualization
- Consolidated Reporting



EXECUTIVE SUMMARY

KEY INSIGHTS:



Vehicle Condition: Primary driver of selling price.



Mileage: Negatively impacts price, but effect varies.



Body Type: Influences condition-price relationship.



Popular Makes and Models: Ford, Toyota, Japanese brands, BMW 3 Series.



MMR: Inaccurate for certain vehicles.



SOURCE

Title: Vehicle Sales Data

Subtitle: Vehicle/Car Sales Trends and Pricing Insights

Author: Syed Anwar

Link: <https://www.kaggle.com/datasets/syedanwarafridi/vehicle-sales-data>

DETAILS

- **Comprehensive:** Includes vehicle details, transaction information, market trends, condition, and mileage.
- **Key Features:** Vehicle details, transaction information, MMR values, condition, mileage.
- **Format:** Tabular format (e.g., CSV).
- **Data Integrity:** Efforts made to ensure accuracy and reliability.

KEY METRICS

- Selling Price

DATA

```
<class 'pandas.core.frame.DataFrame'>
Index: 472325 entries, 0 to 558835
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   year            472325 non-null   int64  
 1   make             472325 non-null   object  
 2   model             472325 non-null   object  
 3   trim              472325 non-null   object  
 4   body              472325 non-null   object  
 5   transmission     472325 non-null   object  
 6   vin               472325 non-null   object  
 7   state             472325 non-null   object  
 8   condition         472325 non-null   float64 
 9   odometer          472325 non-null   float64 
 10  color              472325 non-null   object  
 11  interior          472325 non-null   object  
 12  seller             472325 non-null   object  
 13  mmr                472325 non-null   float64 
 14  sellingprice      472325 non-null   float64 
 15  saledate           472325 non-null   object  
 16  saletime           472325 non-null   object  
dtypes: float64(4), int64(1), object(12)
memory usage: 64.9+ MB
```

A photograph of two people's hands working on a laptop. One hand holds a white stylus pen, pointing at a bar chart on the screen. The other hand holds a black pen, pointing at a pie chart on a printed document. The laptop screen shows a dashboard with various graphs and data. The background is blurred.

ANALYTICAL APPROACH

METHODOLOGY

EDA: Identify patterns, relationships, and outliers.

Statistical Analysis: Quantify variable impact.

Visualization: Present findings clearly.

TOOLS

Analysis tool: Python Programming

Libraries: Pandas, NumPy, Matplotlib, Seaborn, Statsmodels

PROCESS

1. Data cleaning
2. EDA
3. Statistical analysis
4. Visualization
5. Conclusions

ASSUMPTIONS & DISCLAIMERS

Data:

- **Quality:** Intensive data cleaning process using Excel, data seems to have been input from various sources, which could impact accuracy.
- **Representativeness:** Dataset has **52 makes** and **761 models**, the dataset is fairly balanced in that regard.
- **Time Period:** Limited by dataset's time period (**2014-2015**).
- **Regional Variation:** The dataset only has data from **33 US states**, which limits its scope.

Analysis:

- **Causality:** Correlation doesn't always imply causation.
- **Generalizability:** Findings may not apply to all markets or time periods.

Disclaimer:

- Findings based on available data and assumptions. Results may not be applicable to all regions, time periods, or vehicle types. Consider data limitations and potential bias.



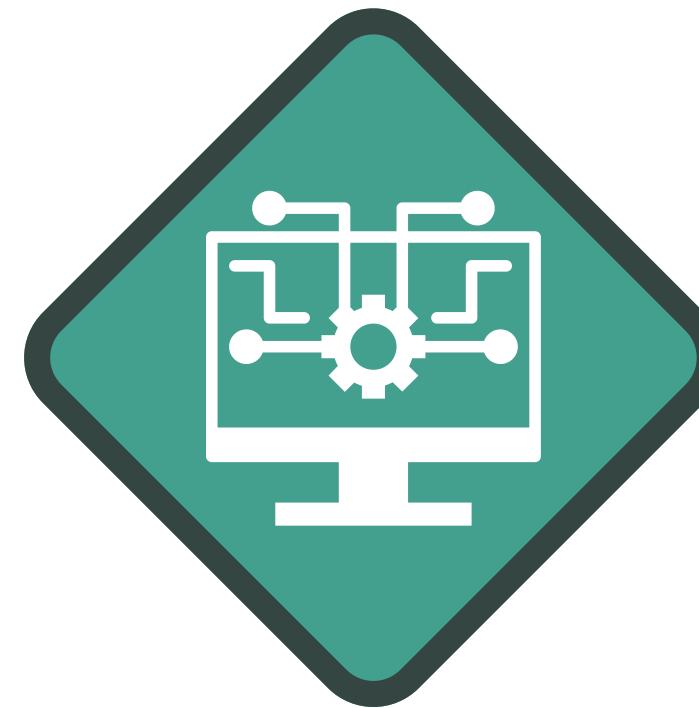
Research Roadmap



Identify Trends in Vehicle Sales Over Time



Analyze the Impact of Vehicle Condition and Mileage on Selling Prices



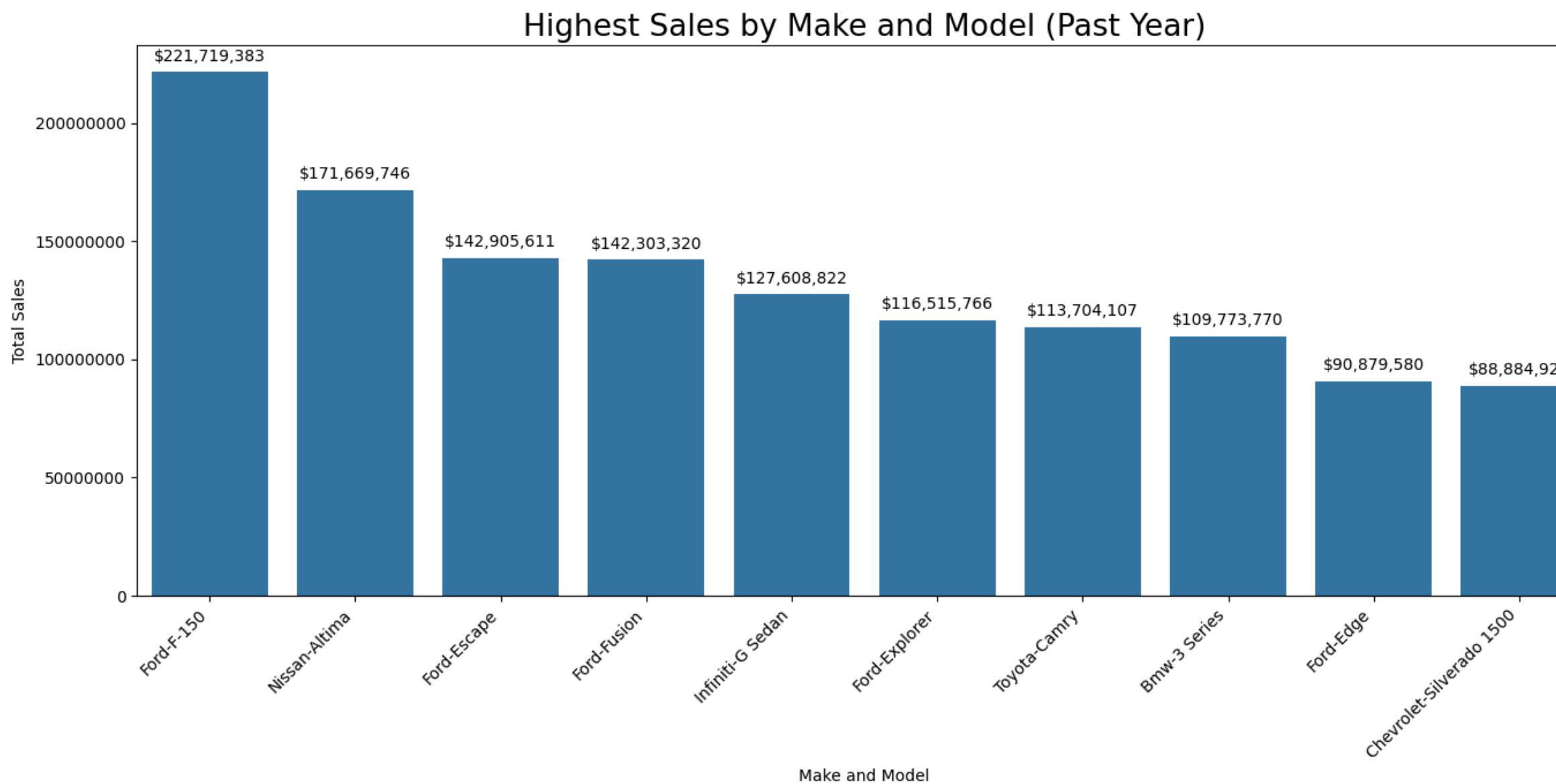
Understand the Relationship Between MMR Values and Actual Selling Prices



Determine the Most Popular Vehicle Makes and Models

IDENTIFYING TRENDS OVER TIME

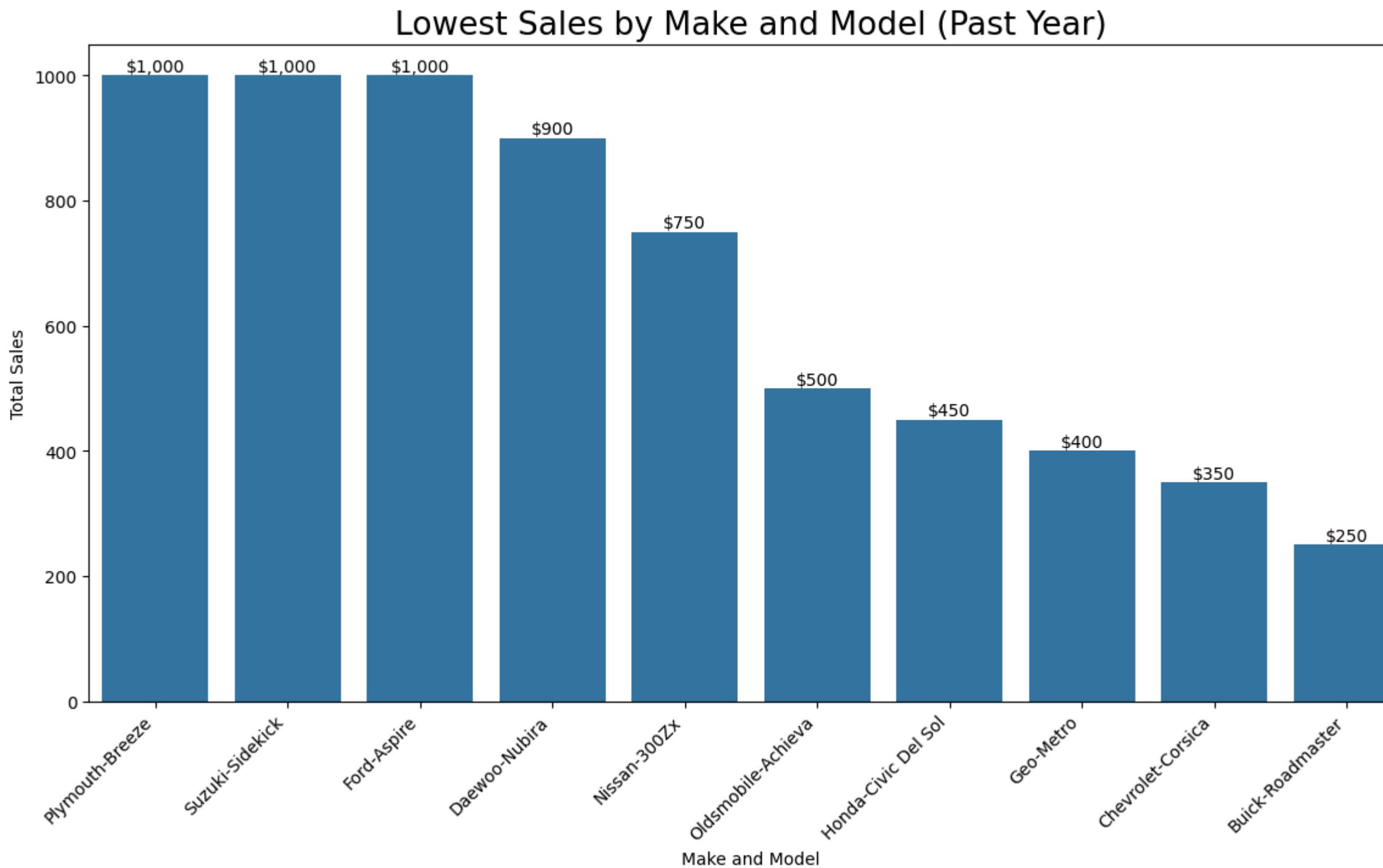
Which makes and models produced the highest sales over the past year?



```
Make and model with the highest sales: ('Ford', 'F-150')
make      model
Ford      F-150      221719383.00
Nissan    Altima     171669746.00
Ford      Escape     142905611.00
          Fusion     142303320.00
Infiniti  G Sedan   127608822.00
Name: sellingprice, dtype: float64
```

IDENTIFYING TRENDS OVER TIME

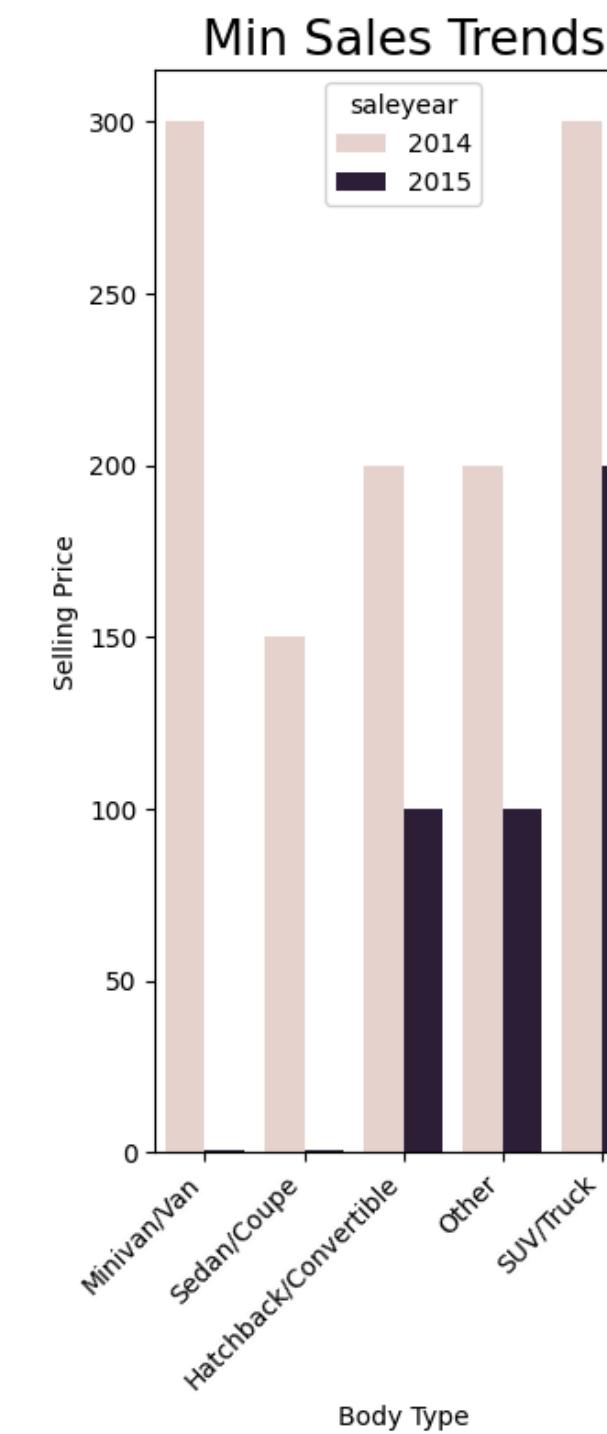
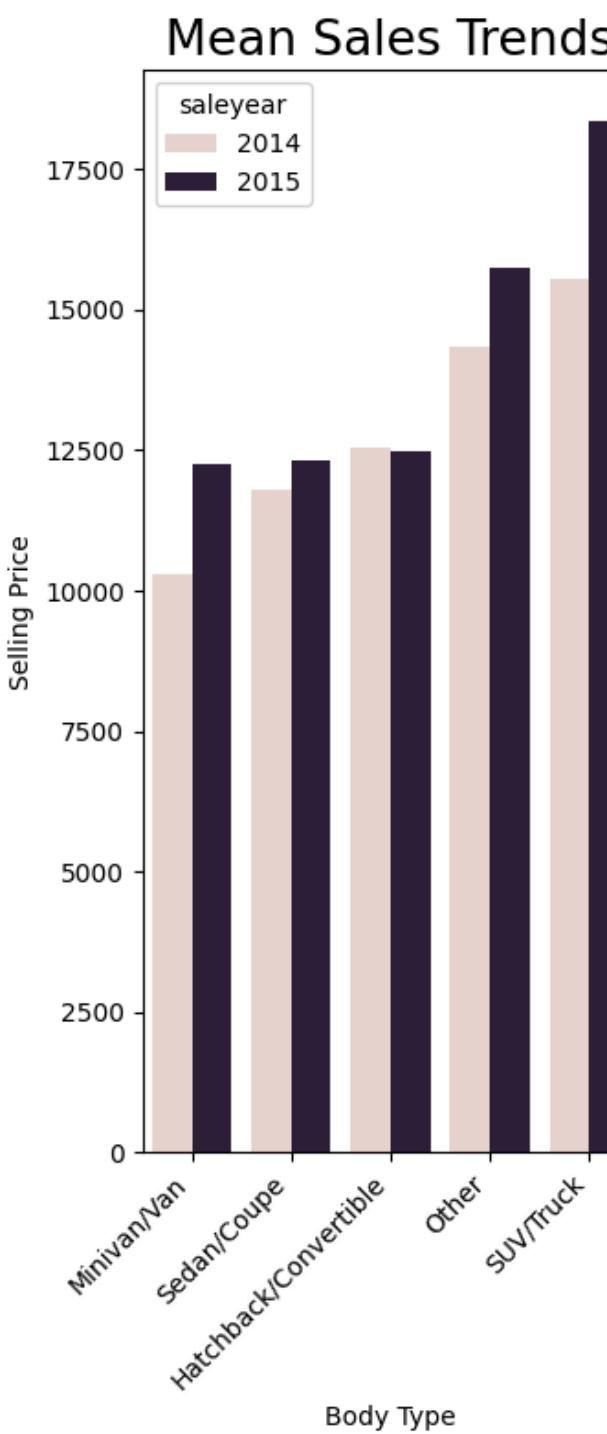
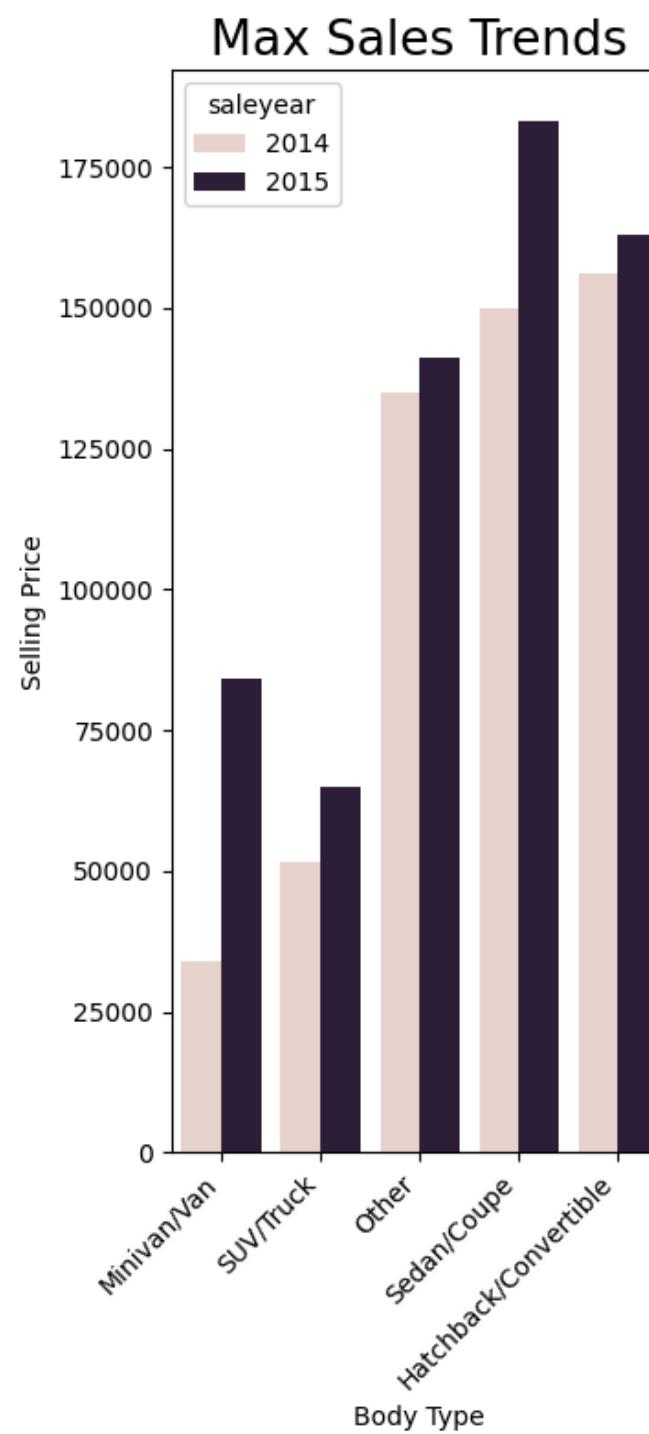
Which makes and models produced the least?



```
Make and model with the lowest sales: ('Buick', 'Roadmaster')
make      model
Buick     Roadmaster    250.00
Chevrolet Corsica     350.00
Geo        Metro        400.00
Honda     Civic Del Sol 450.00
Oldsmobile Achieva    500.00
Name: sellingprice, dtype: float64
```

IDENTIFYING TRENDS OVER TIME

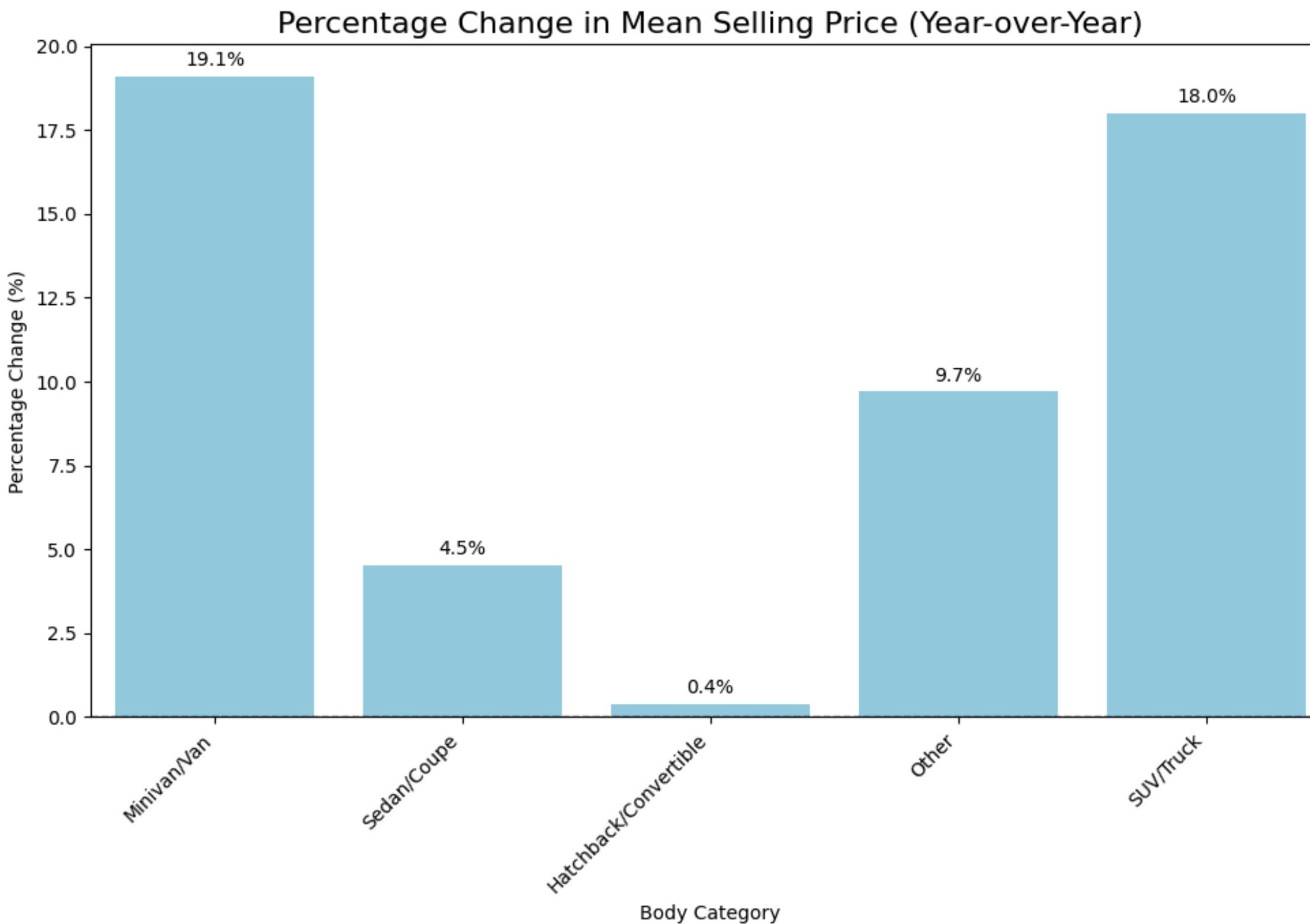
How have sales volumes for different vehicle segments changed over the past year?



- There is an **increase** in sale volumes in **2015** for both mean and max sales across body types.
- **Sedans/Coupes** have the highest overall sales volumes
- **SUV/Trucks** with the highest average sales volumes

IDENTIFYING TRENDS CONT.

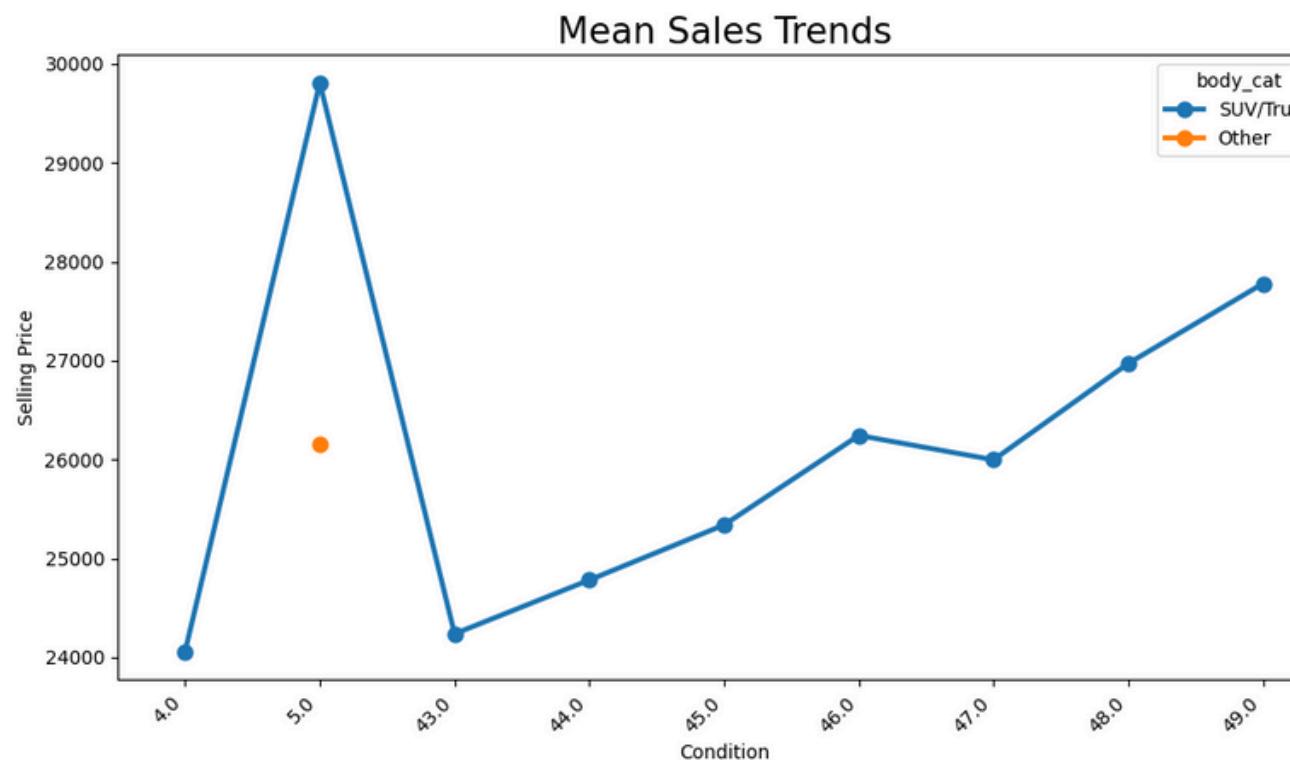
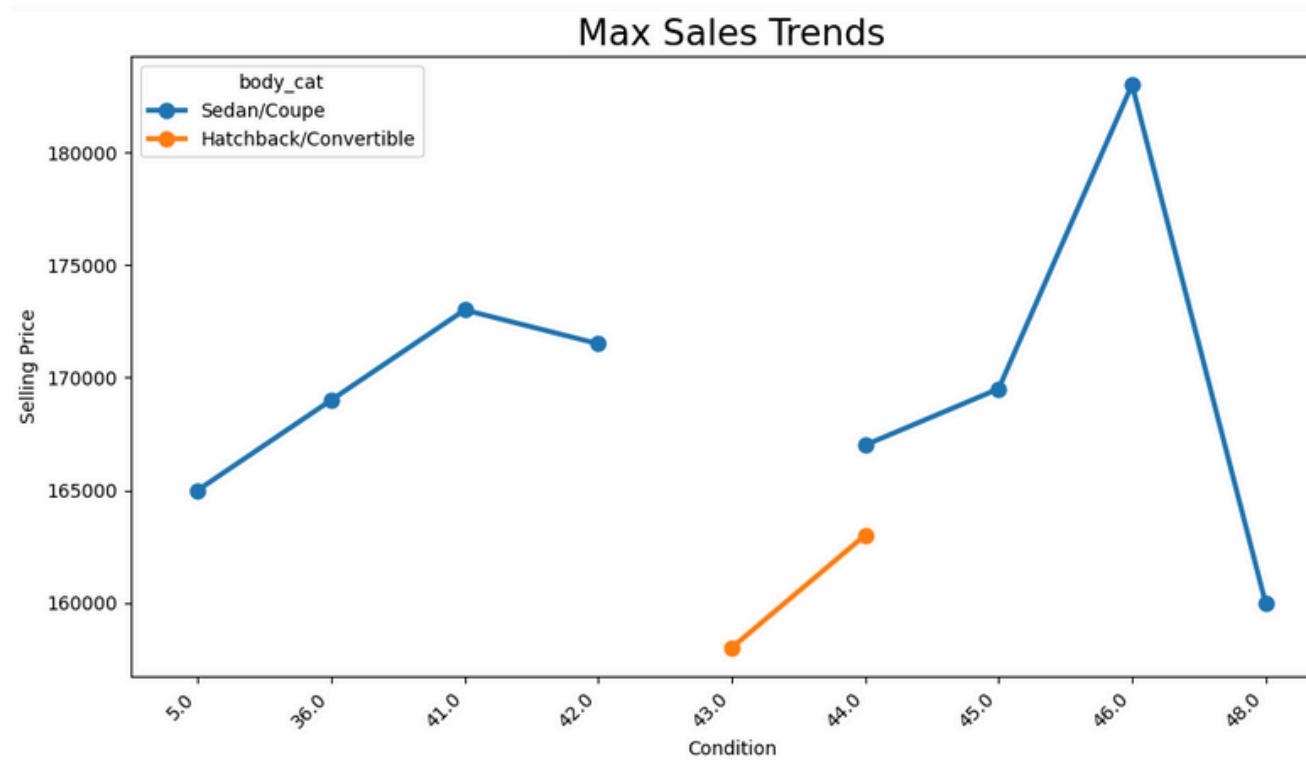
How have sales volumes for different vehicle segments changed over the past year?



- Year-over-year **increase 19.1%** and **18%** for Vans and SUV/Trucks, respectively.
- There is an **increase** in sales for each segment.

ANALYZING THE IMPACT

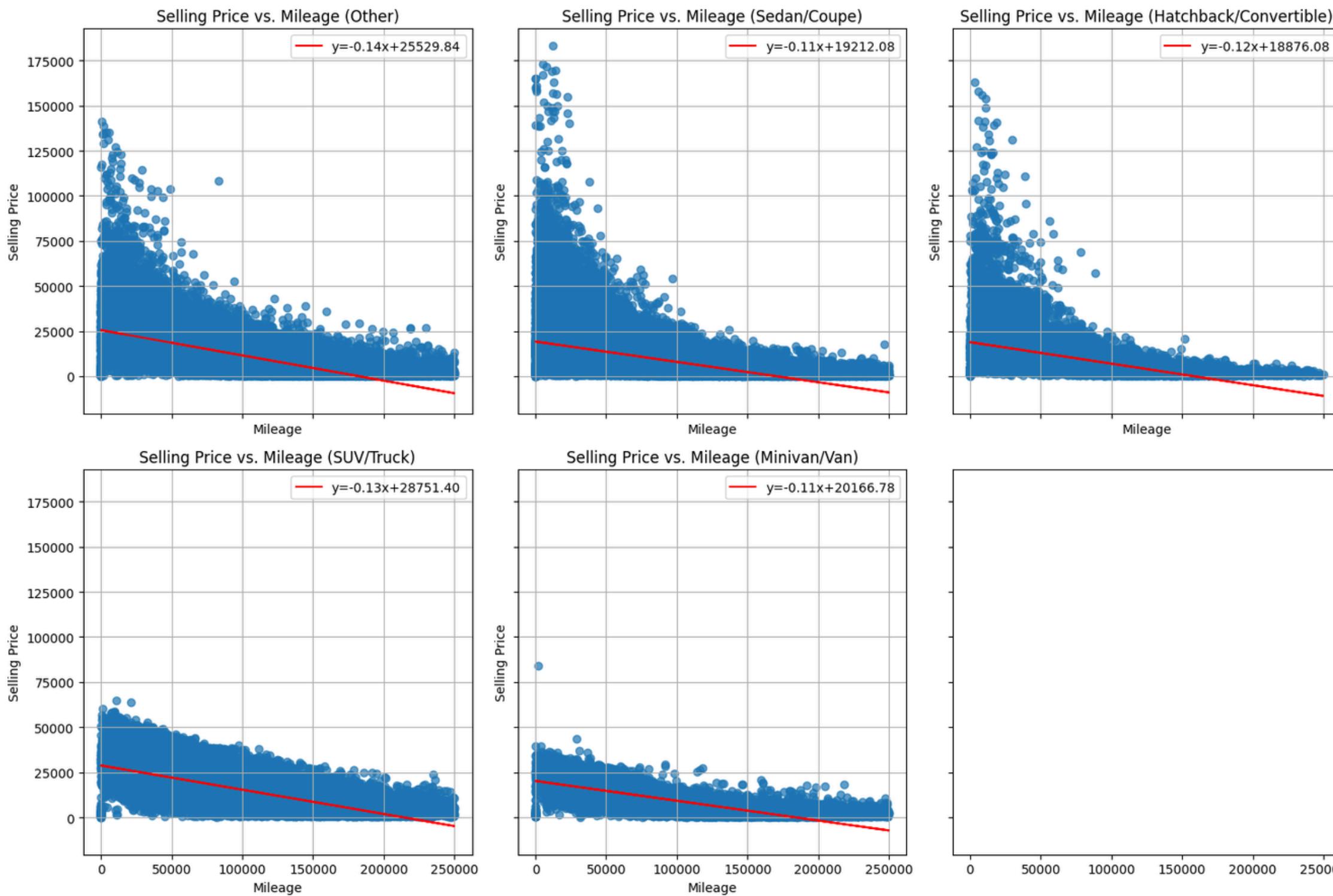
How does vehicle condition affect selling price for different car body types?



- Condition metric is measured **1 to 49**, with 1 being the best.
- The **correlation** between Condition and Selling Price: **0.31**, which is **low**. Meaning Condition alone is not a strong indicator of Selling Price.
- **Max Sales:** We found body type, and more importantly make is a stronger indicator for higher sales
- **Mean Sales:** We found that condition is a stronger indicator for consistent sales.

ANALYZING THE IMPACT

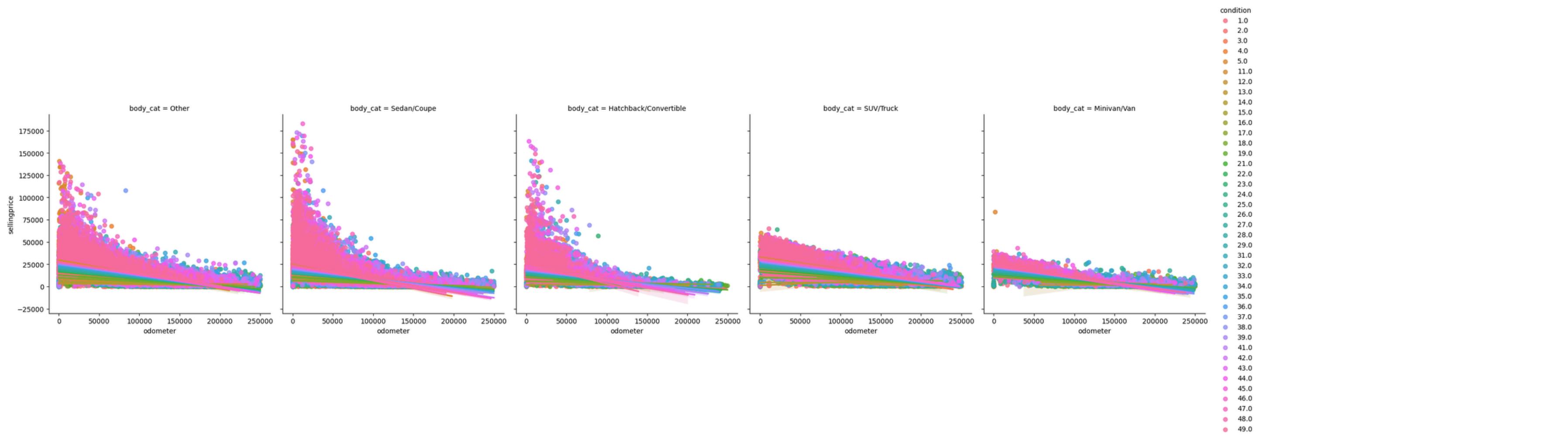
What is the relationship between mileage and selling price for vehicles of the same year?



- Odometer metric is measured **0 to 250,000**.
- The **correlation** between Odometer and Selling Price: **-0.59**, which is a **moderate negative trend**.
- Overall, lower Mileage trends with higher selling prices, though varies between body types.
- **Sedan/Coupe** has the **highest** correlation with this trend.

ANALYZING THE IMPACT

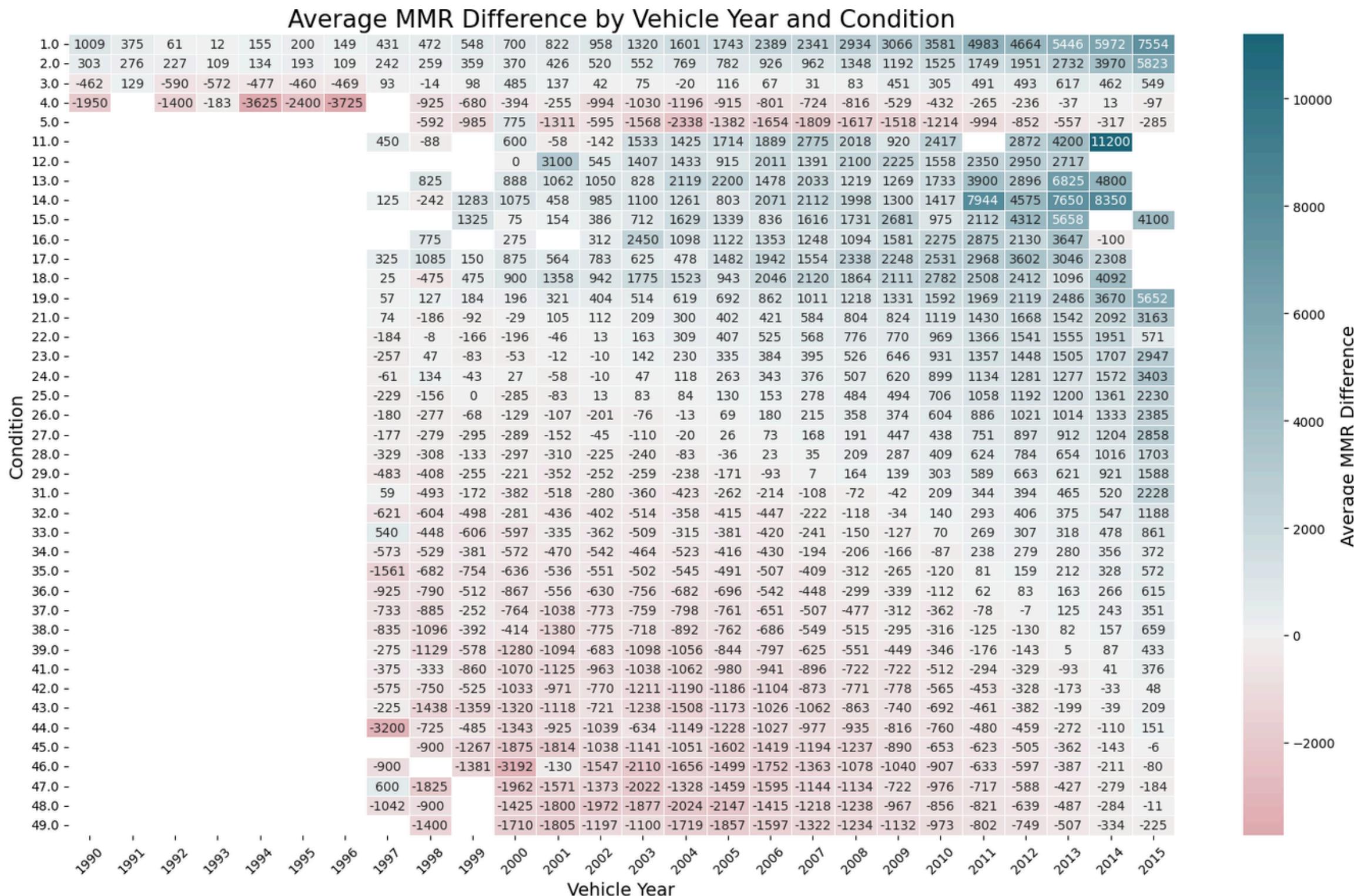
Is there an interaction effect between vehicle condition and mileage on selling price?



- Utilized Linear Regression via seaborn library Implot() to explore these relationships.
- The graphs indicate an interactive effect due to:
 - The negative correlation
 - Non-parallel and crossing lines
 - Differential patterns

UNDERSTANDING RELATIONSHIPS

How does the difference between MMR values and selling prices vary across different vehicle years and conditions?



UNDERSTANDING RELATIONSHIPS

Are there specific makes and models where MMR values consistently overestimate or underestimate selling prices?

	make	model	mmr_diff
7	Acura	Rlx	1223.076923
28	Audi	R8	1664.062500
29	Audi	Rs 4	1733.333333
46	Bentley	Continental Gtc	3240.000000
48	Bentley	Continental Supersports	14000.000000
54	Bmw	4 Series Gran Coupe	1163.636364
63	Bmw	Activehybrid X6	2360.000000
64	Bmw	I8	6555.555556
70	Bmw	M6 Gran Coupe	2450.000000
101	Cadillac	Cts-V Coupe	1272.269231
106	Cadillac	Elr	4183.333333
127	Chevrolet	Caprice	1121.428571
146	Chevrolet	Malibu Hybrid	1005.208333
154	Chevrolet	Silverado 1500 Hybrid	1050.000000
160	Chevrolet	Silverado 3500 Classic	1650.000000
164	Chevrolet	Spark Ev	3050.000000
219	Ferrari	California	1846.153846
220	Ferrari	F430	5750.000000
224	Fisker	Karma	2638.888889
232	Ford	E-350	1100.000000
273	Ford	Transit Wagon	1160.000000
289	Gmc	Sierra 1500 Hybrid	1200.000000
293	Gmc	Sierra 2500Hd Classic	1202.941176
294	Gmc	Sierra 3500	1845.000000
358	Infiniti	G37 Coupe	1168.181818
417	Lamborghini	Gallardo	2166.666667
450	Lexus	Ls 600H L	1187.500000

	make	model	mmr_diff
14	Aston Martin	Db9	-2883.333333
15	Aston Martin	Rapide	-8250.000000
36	Audi	S7	-1664.285714
47	Bentley	Continental Gtc Speed	-2000.000000
61	Bmw	Activehybrid 5	-2825.000000
62	Bmw	Activehybrid 7	-2723.076923
68	Bmw	M5	-1031.465116
80	Bmw	Z4 M	-1340.000000
99	Cadillac	Cts Wagon	-1031.818182
102	Cadillac	Cts-V Wagon	-8200.000000
285	Gmc	Savana	-2265.909091
320	Hummer	H1	-2350.000000
322	Hummer	H2 Sut	-1710.483871
324	Hummer	H3T	-2466.666667
370	Infiniti	Q60 Convertible	-1151.250000
375	Infiniti	Qx50	-1012.500000
399	Jeep	Grand Cherokee Srt	-1700.000000
465	Lincoln	Blackwood	-1562.500000
481	Maserati	Coupe	-1150.000000
487	Maserati	Spyder	-2066.666667
517	Mercedes-Benz	400-Class	-1275.000000
542	Mercury	Marauder	-2843.750000

Makes with Overestimated MMR Values:

- Luxury brands:** Acura, Audi, Bentley, Ferrari, Lamborghini, Lexus, Maserati, Mercedes-Benz, Porsche, Rolls-Royce
- Electric vehicles:** Chevrolet Spark EV, Nissan Leaf, Tesla Model S, Tesla Model X

Makes with Underestimated MMR Values:

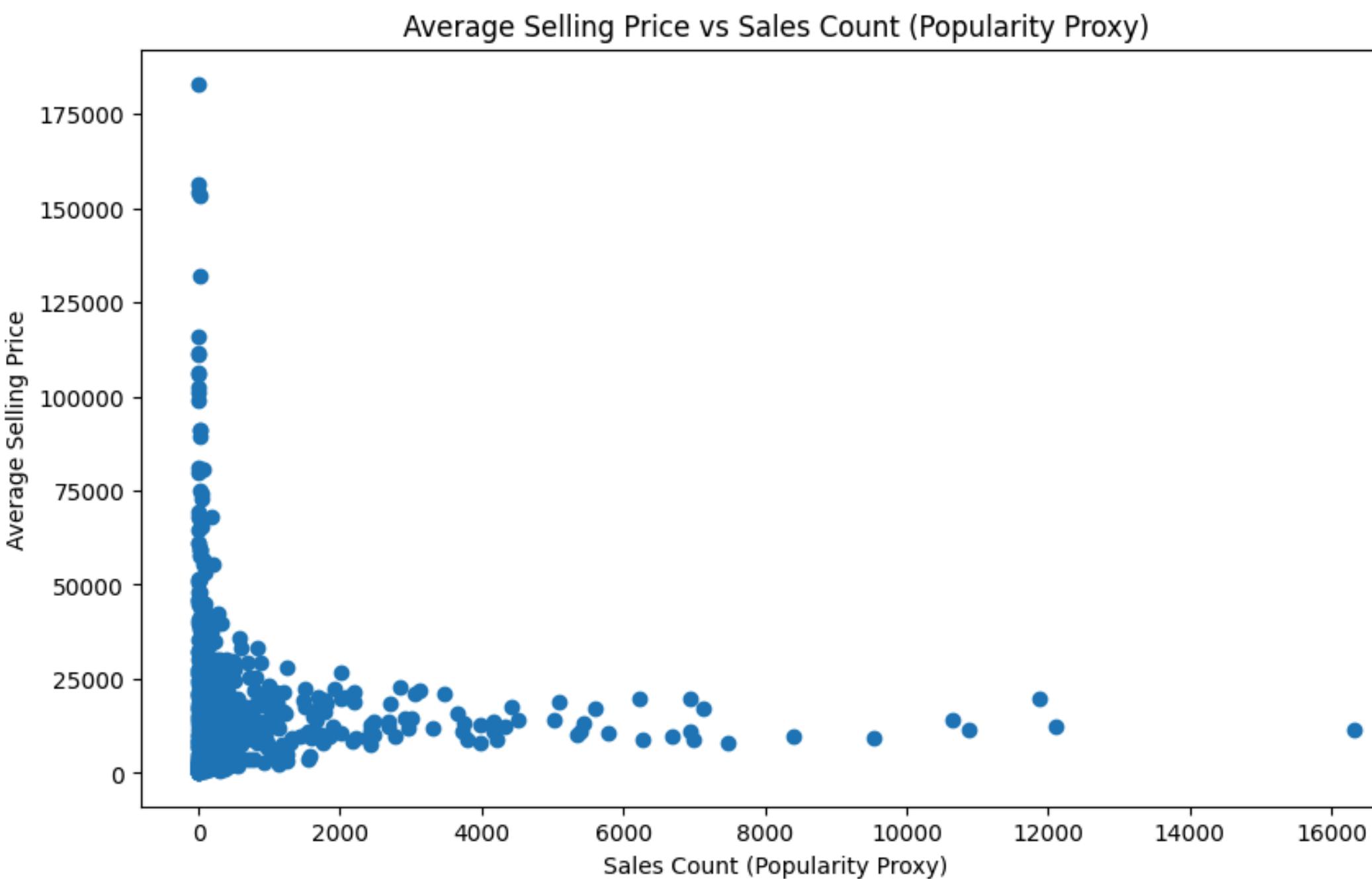
- Luxury performance brands:** Aston Martin, BMW (M series), Cadillac (V series), Hummer, Jeep (SRT), Lincoln, Maserati, Mercedes-Benz (AMG), Porsche

Potential reasons for these discrepancies could include:

- Brand prestige:** Luxury brands might have a higher perceived value, leading to inflated MMR estimates.
- Performance and features:** High-performance models might have unique features that are not fully reflected in MMR calculations.
- Market trends:** Changes in market demand for certain makes or models could affect the accuracy of MMR estimates.

DETERMINE POPULARITY

Which vehicle makes and models have the highest average selling prices, and how does this relate to their popularity?



	make	model	avg_selling_price	sales_count
218	Ferrari	458 Italia	183000.00	1
538	Mercedes-Benz	Sls Amg Gt	156500.00	1
64	Bmw	I8	154222.22	9
650	Rolls-Royce	Ghost	153456.25	16
219	Ferrari	California	131846.15	13
...
610	Oldsmobile	Cutlass Ciera	358.33	6
133	Chevrolet	Corsica	350.00	1
269	Ford	Tempo	333.33	3
552	Mercury	Tracer	312.50	4
212	Dodge	Spirit	300.00	1

[770 rows x 4 columns]

Expensive Luxury Cars Lead: Ferrari, Mercedes-Benz, Rolls-Royce, and Bmw top the list for average selling price.

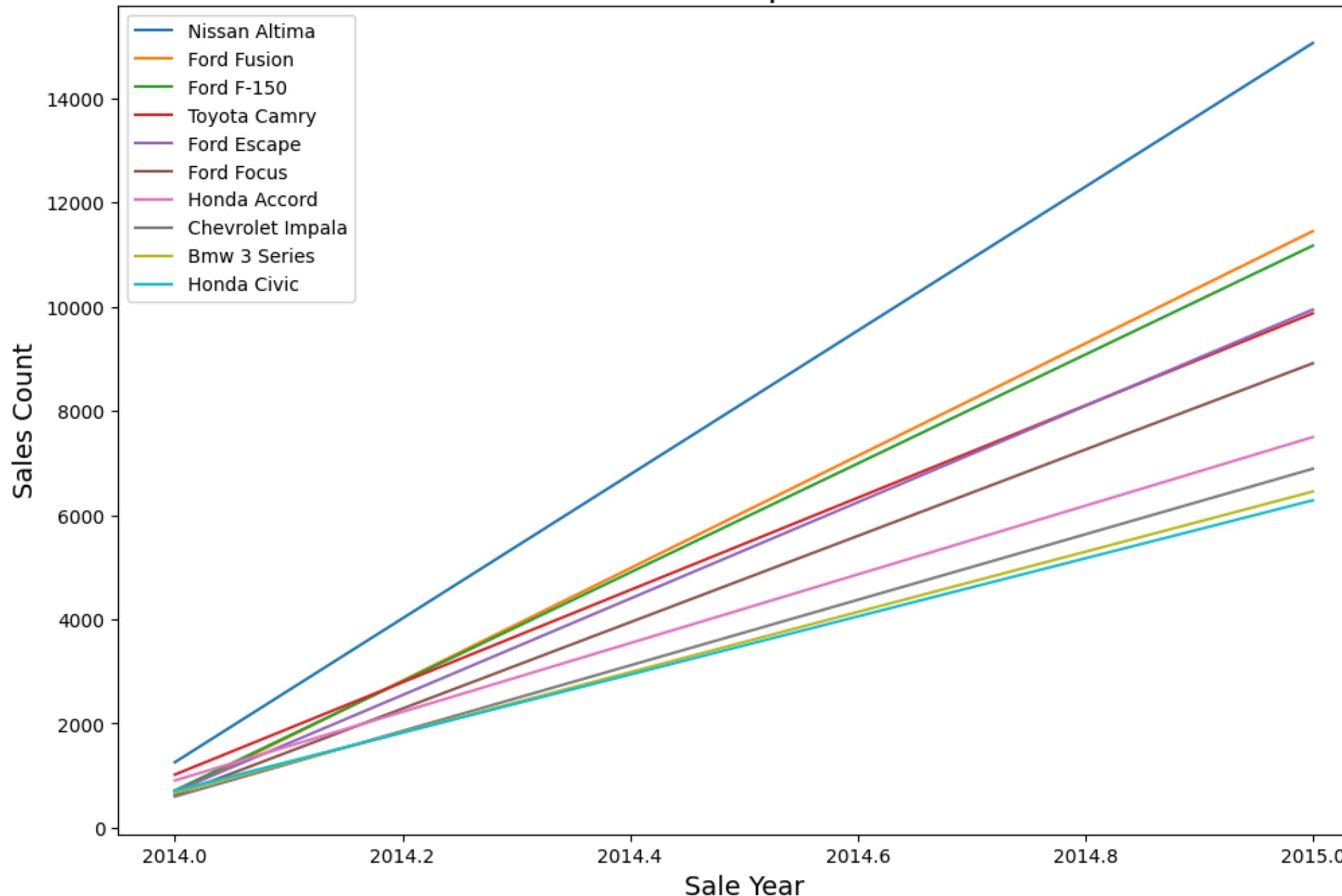
Popularity Varied: Some high-priced models (e.g., Ferrari 458 Italia) sold very few units, while others (e.g., Bmw I8) sold more.

- Scatter plot (not shown here) reveals:
- No clear correlation between average price and sales count.
- Some expensive cars sold well (potentially due to brand or high performance).
- Other factors likely influence price besides popularity (e.g., features, rarity).

DETERMINE POPULARITY

Are there any emerging trends in the popularity of certain makes and models over recent years?

Sales Trends for Top Makes/Models



Key Observations:

- **Steady Growth:** Most models experienced a steady increase in sales during this time.
- **Ford Dominance:** Ford models, including the F-150, Fusion, Escape, and Focus, were consistently popular.
- **Toyota Camry:** The Toyota Camry maintained a strong presence throughout the period.
- **Nissan Altima:** The Nissan Altima also showed consistent sales growth.

Overall, the graph suggests:

- **Strong demand:** There was a high demand for these models during the specified period.
- **Market stability:** The overall market for these vehicles remained relatively stable, with no significant shifts in popularity.

Conclusion & Recommendations



Conclusion

- **Vehicle condition:** the critical factor influencing selling price.
- **Mileage:** played a secondary role.
- The relationship between condition and price varied across body types.
- **Popular makes and models:** Ford vehicles, the Toyota Camry, and Japanese brands.
- **MMR values:** Found to be inaccurate for certain vehicles, highlighting the need for a comprehensive analysis beyond MMR.



Recommendations

- **Buyers:** Prioritize vehicles with high condition and low mileage to maximize value. Consider factors beyond MMR when evaluating prices.
- **Sellers:** Accurately assess vehicle condition and adjust pricing strategies accordingly.
- **Industry:** Conduct further research to improve the accuracy of MMR valuation models.
- **Policymakers:** Implement policies that promote the sale of higher-condition vehicles and address factors that contribute to MMR inaccuracies.



THANK YOU FOR YOUR ATTENTION



Visit Our Website:

<https://github.com/joerobo2>

