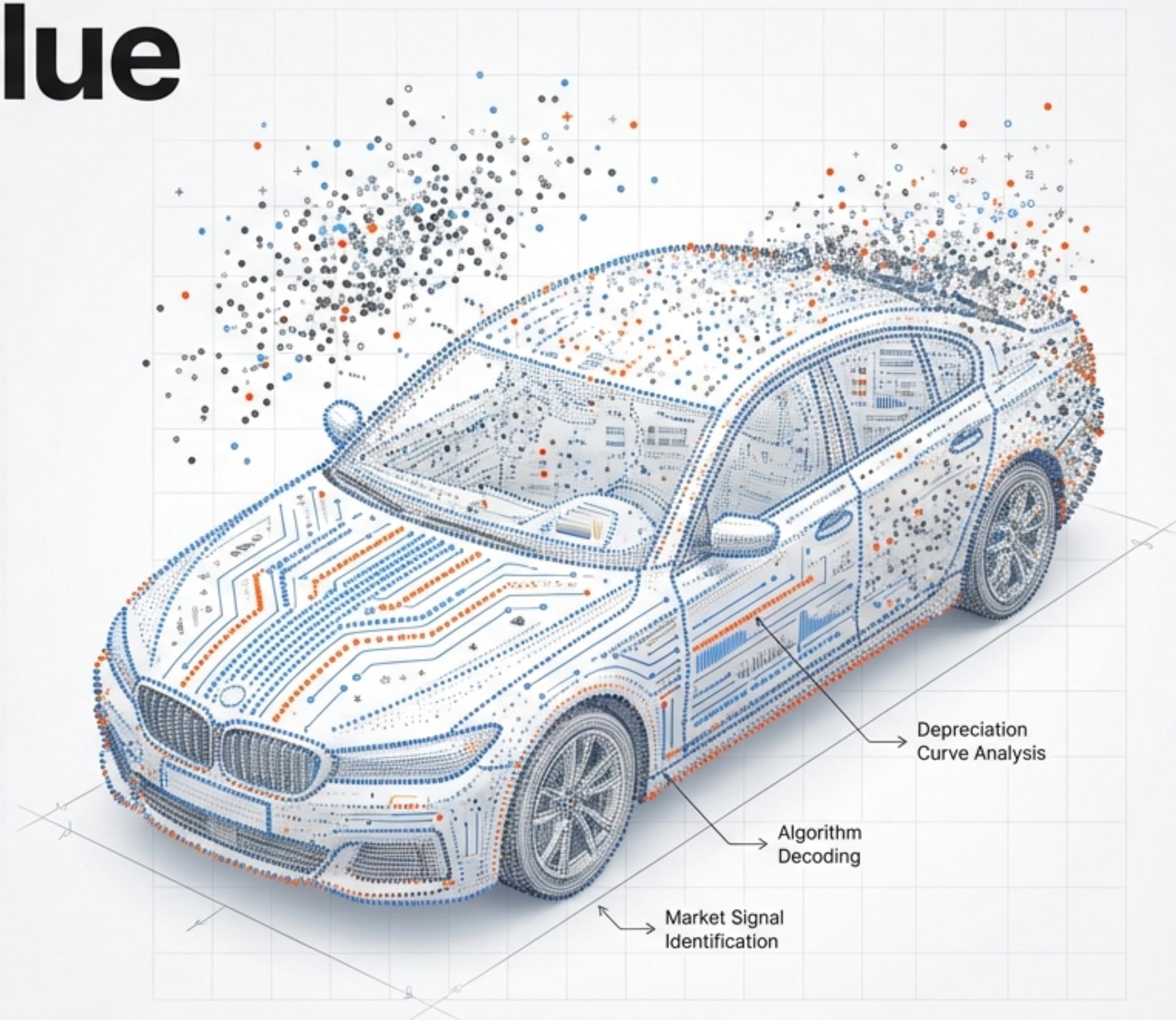


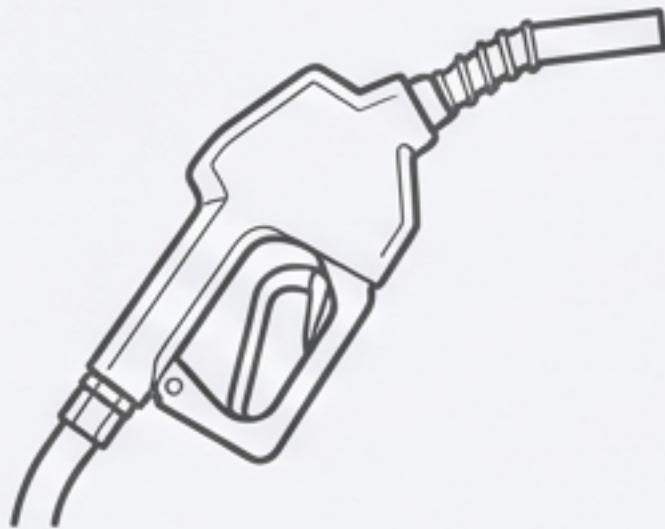
Unlocking Value in the Used Car Market

A data-driven journey through 426,000+ listings to decode pricing algorithms, depreciation curves, and hidden market signals.

- **DATASET:** 426,880 ENTRIES
- **SOURCE:** CRAIGSLIST METADATA
- **TOOLS:** PYTHON / PANDAS / SEABORN

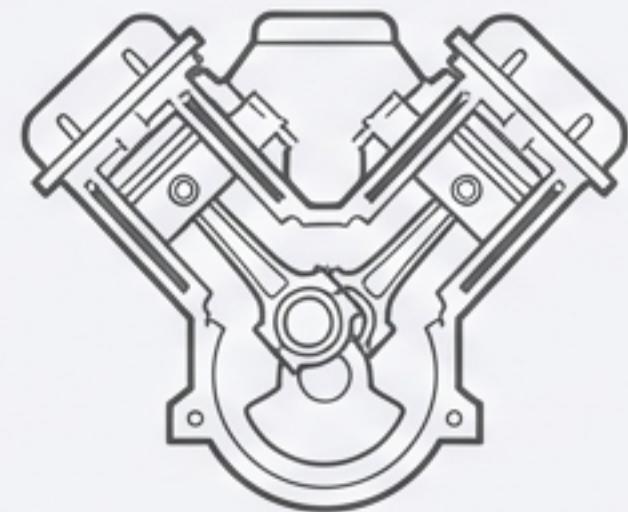


The Signal in the Noise: Key Findings



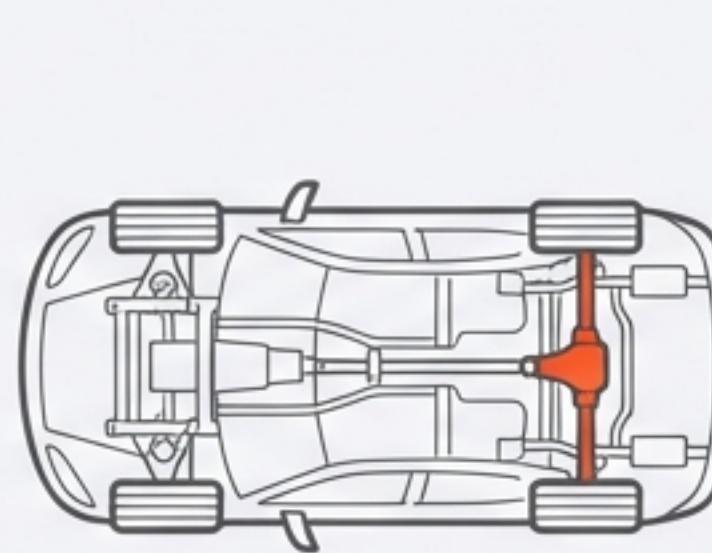
The Diesel Premium **\$33,000 Avg**

Diesel vehicles command the highest market value, outperforming Electric (\$25k) and Gas (\$17k) models.



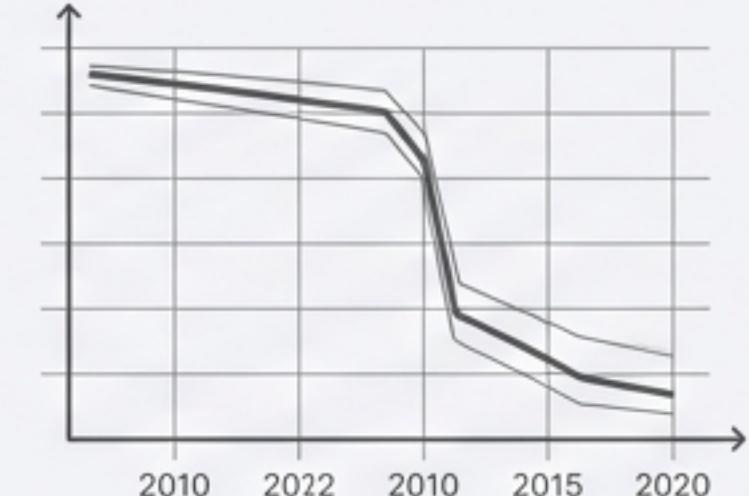
The Cylinder Gap **2x Value**

8-cylinder engines average ~\$23.9k, more than double the value of 4-cylinder economy engines (~\$11k).



RWD Advantage **+\$8,719**

Rear-Wheel Drive configurations carry a massive premium over Front-Wheel Drive counterparts.



Depreciation Cliff **2010**

The market bifurcates at the 2010 model year. Post-2010 cars avg \$23k; Pre-2010 drop to \$9k.

The Raw Data: A Diamond in the Rough

The initial dataset of 426,880 rows was riddled with noise, requiring significant cleaning before insights could be trusted.

Real-world data is messy; we found missing VINs, null manufacturer names, and placeholder prices.

```
df.isna().sum()
df['price'] = df['price'].replace(0, np.nan)
df.dropna(subset=['vin', 'manufacturer'])
df.info()
```

INDEX	VIN	MANUFACTURER	MODEL	YEAR	CYLINDERS	CONDITION	ODOMETER	PRICE	POSTED_DATE
0	NaNAV3SBVXXXAN	None	Generic	2019	NaN	NaN	50000	\$12000	2023-01-15
1	NaNAZ3XXXXXXAN	None	Generic	2019	NaN	NaN	5000	\$0	2023-01-15
2	NaNAZ3XXXXXXAN	None	Generic	2019	NoN	NaN	12000	\$0	2023-01-15
3	NaNAZ3XXXXXXAN	None	Generic	2019	NoN	NaN	30000	\$0	2023-01-15
4	NaNAZ3XXXXXXAN	None	Generic	2019	NoN	NaN	50000	\$12000	2023-01-15
5	NaNAZ3XXXXXXAN	None	Generic	2019	NoN	NaN	20000	\$5000	2023-01-15
6	NaNAZ3XXXXXXAN	None	Generic	2019	NaN	NaN	30000	\$12000	2023-01-15
7	NaNAK230VXXXAN	None	Generic	2019	NoN	NaN	50000	\$0	2023-01-15
8	Nan	None	Generic	2019	NoN	NaN	50000	\$0	2023-01-15
9	NaNAV238VXXXAN	None	Generic	2019	NoN	NaN	50000	\$0	2023-01-15
10	NaNAV339VXXXAN	None	Generic	2019	NaN	NaN	50000	\$0	2023-01-15
11	NaNAV339VXXXAN	None	Generic	2019	NaN	NaN	50000	\$0	2023-01-15
112	NaNAV339VXXXAN	None	Generic	2019	NaN	NaN	50000	\$0	2023-01-15
113	NaNAV339VXXXAN	None	Generic	2019	NaN	NaN	50000	\$0	2023-01-15
114	NaNAV339VXXXAN	None	Generic	2019	NaN	NaN	50000	\$0	2023-01-15
...
115	NaNAV339VXXXAN	None	Generic	2019	NaN	NaN	25000	\$0	2023-01-15
116	NaNAV330VXXXAN	None	Generic	2019	NoN	NaN	20000	\$0	2023-01-15
117	NaNAV2309EXXAN	NaN	Generic	NaN	NaN	NaN	50000	\$12000	2023-01-15
118	NaNAV2309EXXAN	None	Generic	NaN	NaN	NaN	50000	\$0	2023-01-15
119	NaNAV2309EXXAN	None	Generic	NaN	NaN	NaN	60000	\$0	2023-01-15

```
df.isna().sum()
df['price'] = df['price'].replace(0, np.nan)
df.dropna(subset=['vin', 'manufacturer'])
df.info()
```

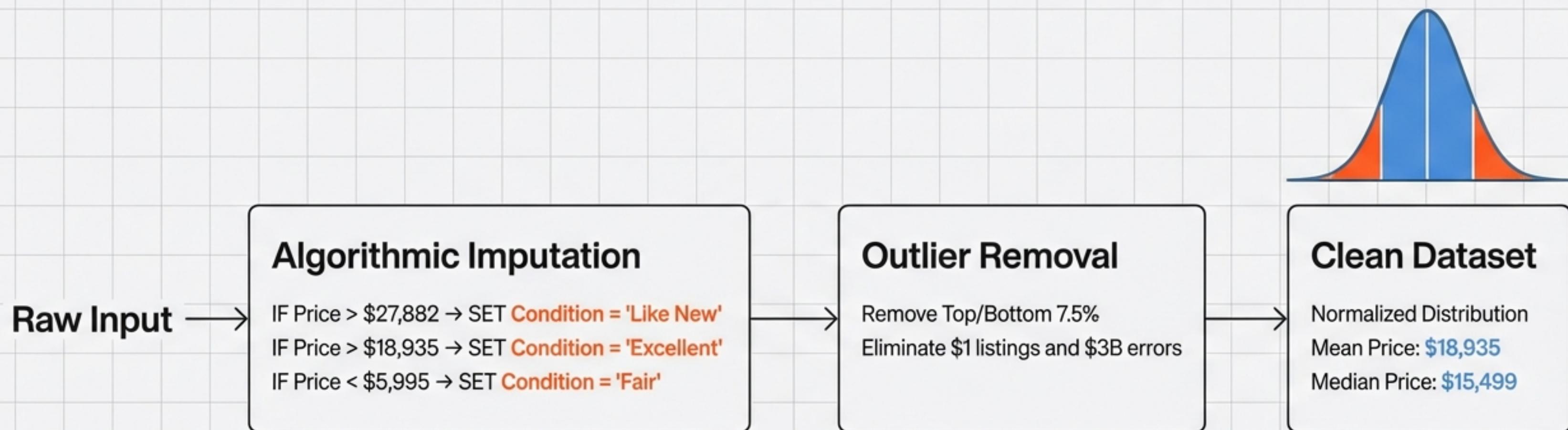
MISSING ESSENTIALS
177,678 rows missing 'cylinders',
174,104 missing 'condition'.

THE ZERO-DOLLAR PROBLEM
Thousands of listings with
\$0 asking price.

STRUCTURAL GAPS
1,205 listings missing
the year.

From Chaos to Clarity: The Cleanup Logic

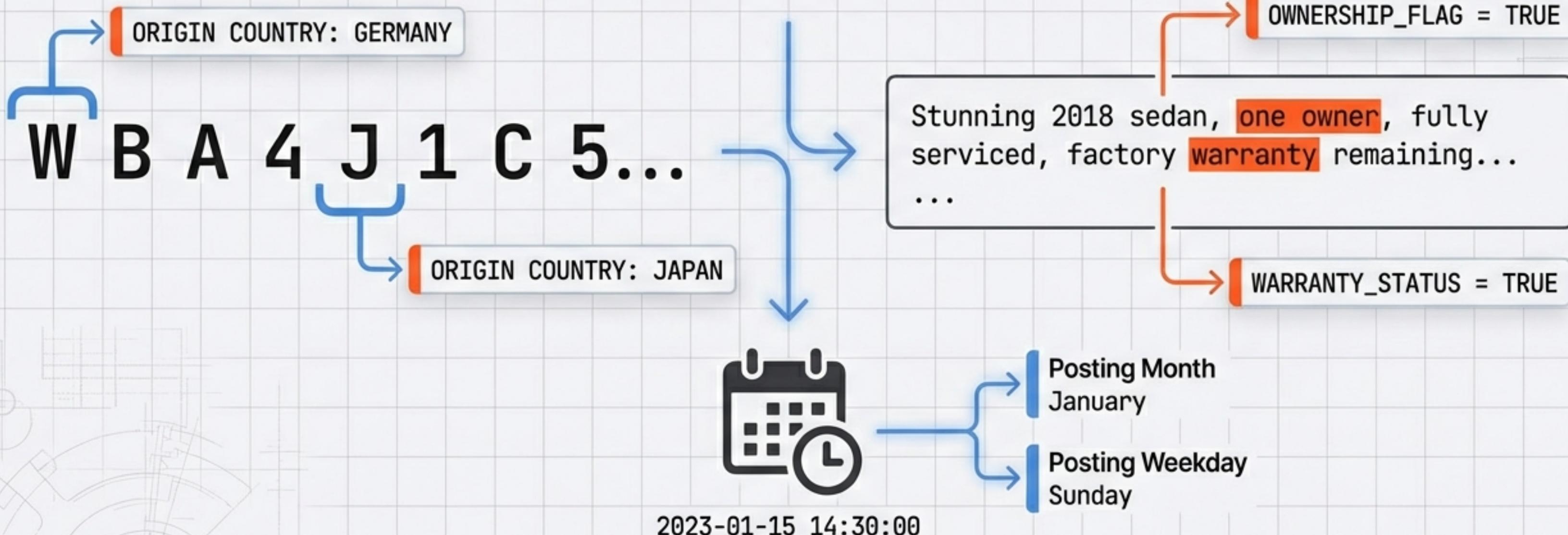
We didn't just delete data—we used **price deductively** to impute missing variables.



Feature Engineering: Creating New Intelligence

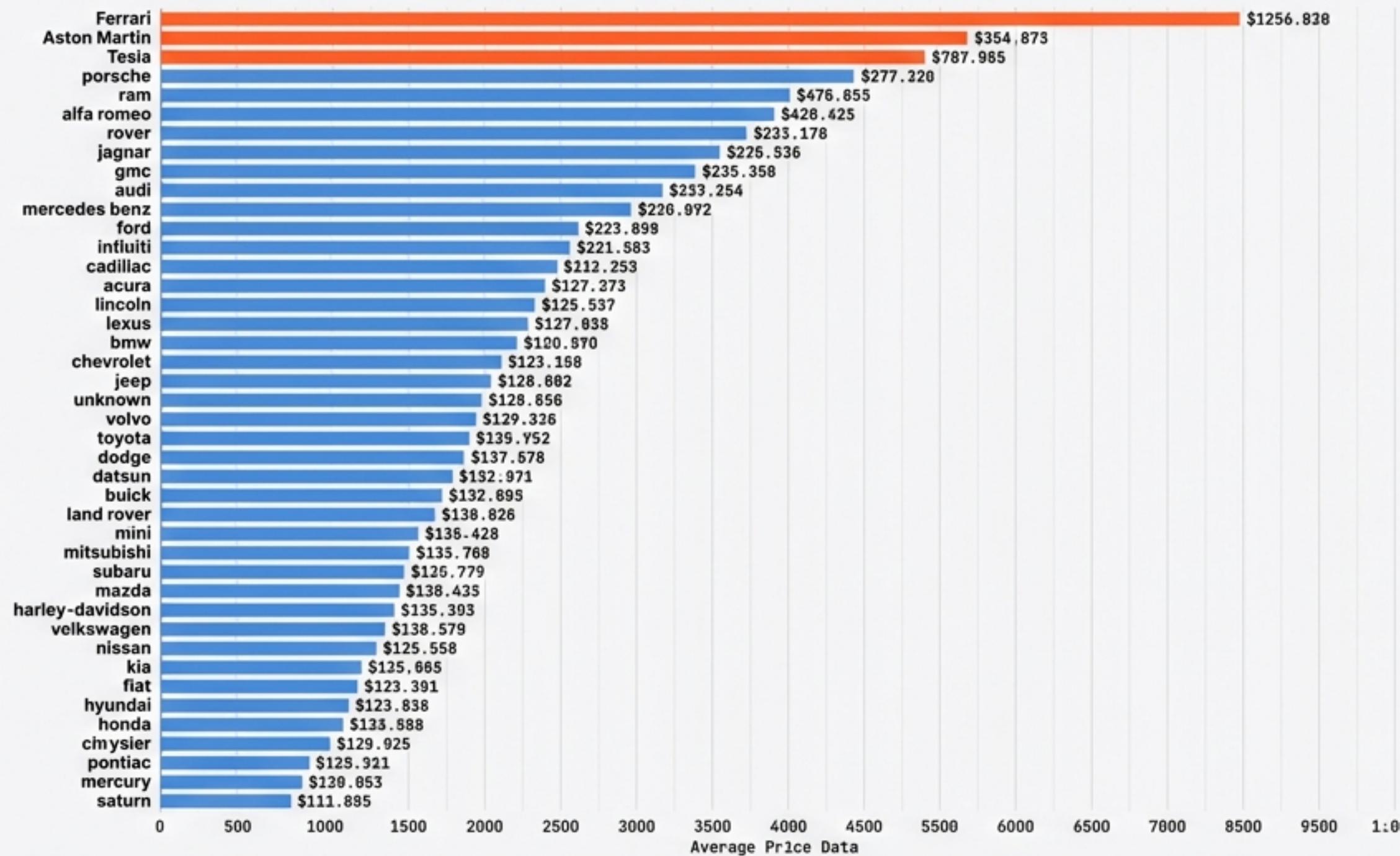
We derived new columns to reveal hidden attributes by extracting intelligence from text fields.

Data Extraction



The Price Hierarchy: Brand Dominance

Average Price per Car Brand



Top 5 Most Expensive Car Brands (Max Price)



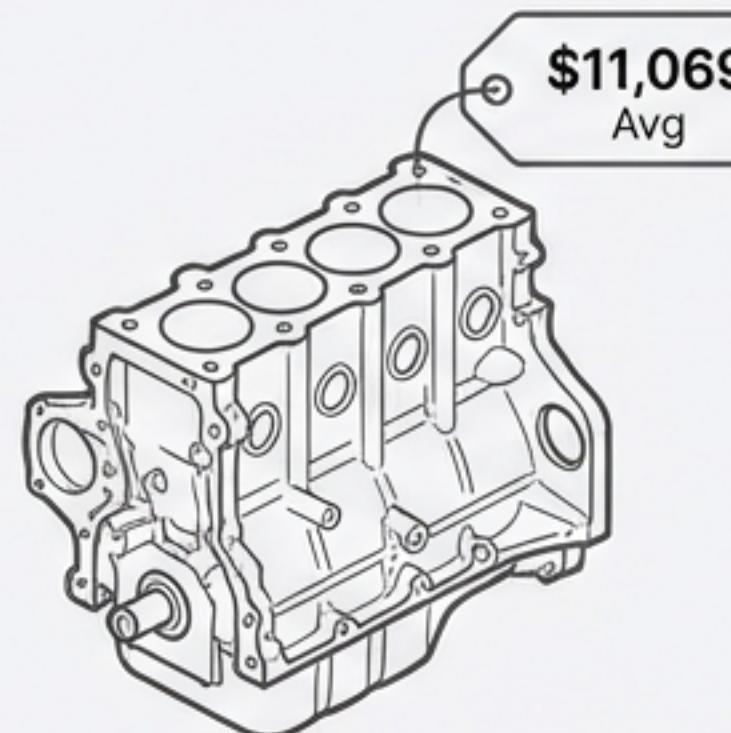
Insights

THE EXOTICS: Ferrari sits alone at the top, followed by Aston Martin.

THE TECH PREMIUM: Tesla ranks third, outpricing traditional luxury stalwarts like Porsche and Rover.

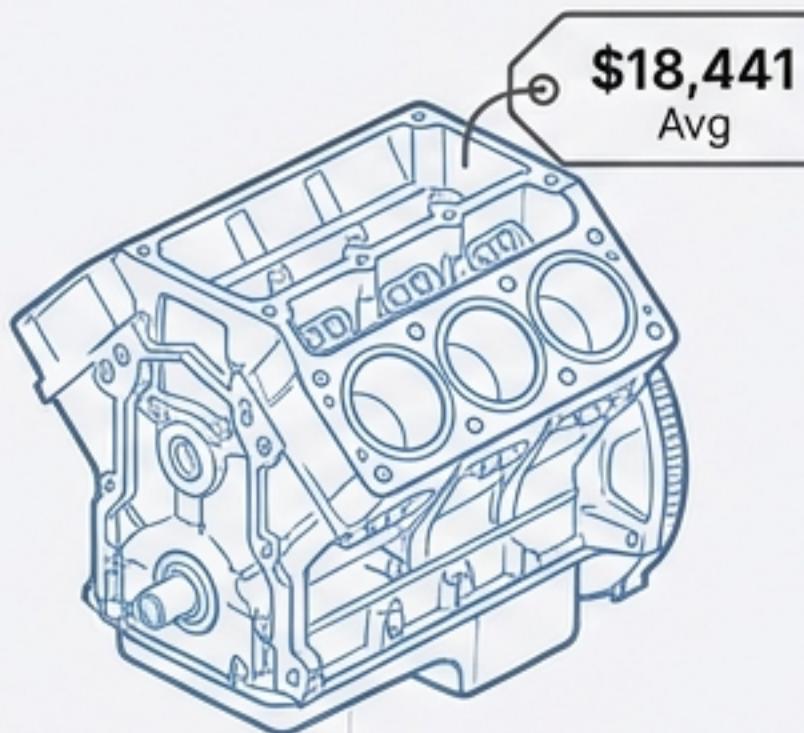
VOLUME VS VALUE: Mainstream brands (Ford, Toyota) show lower averages due to high volume, despite their truck segments.

The Engine Room: Cylinders = Currency



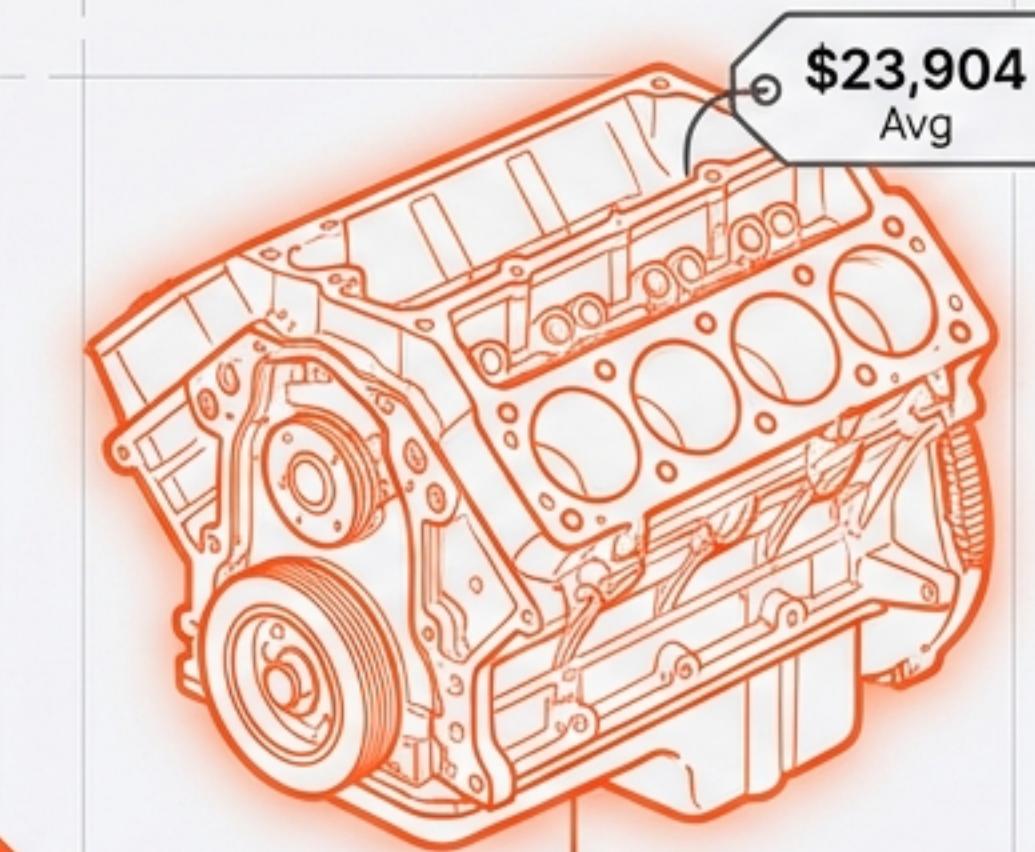
4 Cylinders

\$11,069
Avg



6 Cylinders

\$18,441
Avg



8 Cylinders

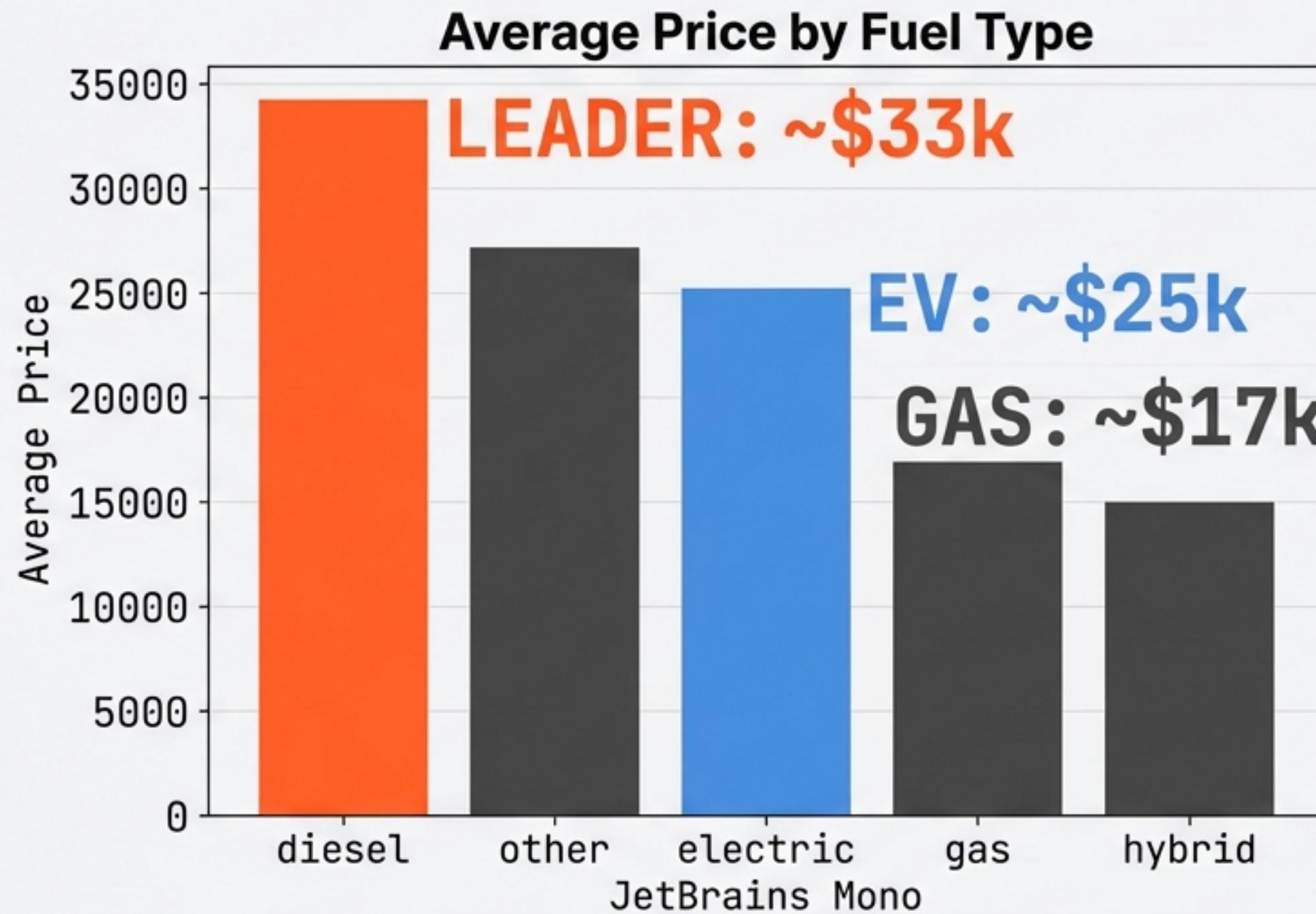
\$23,904
Avg

+\$12,835 PREMIUM



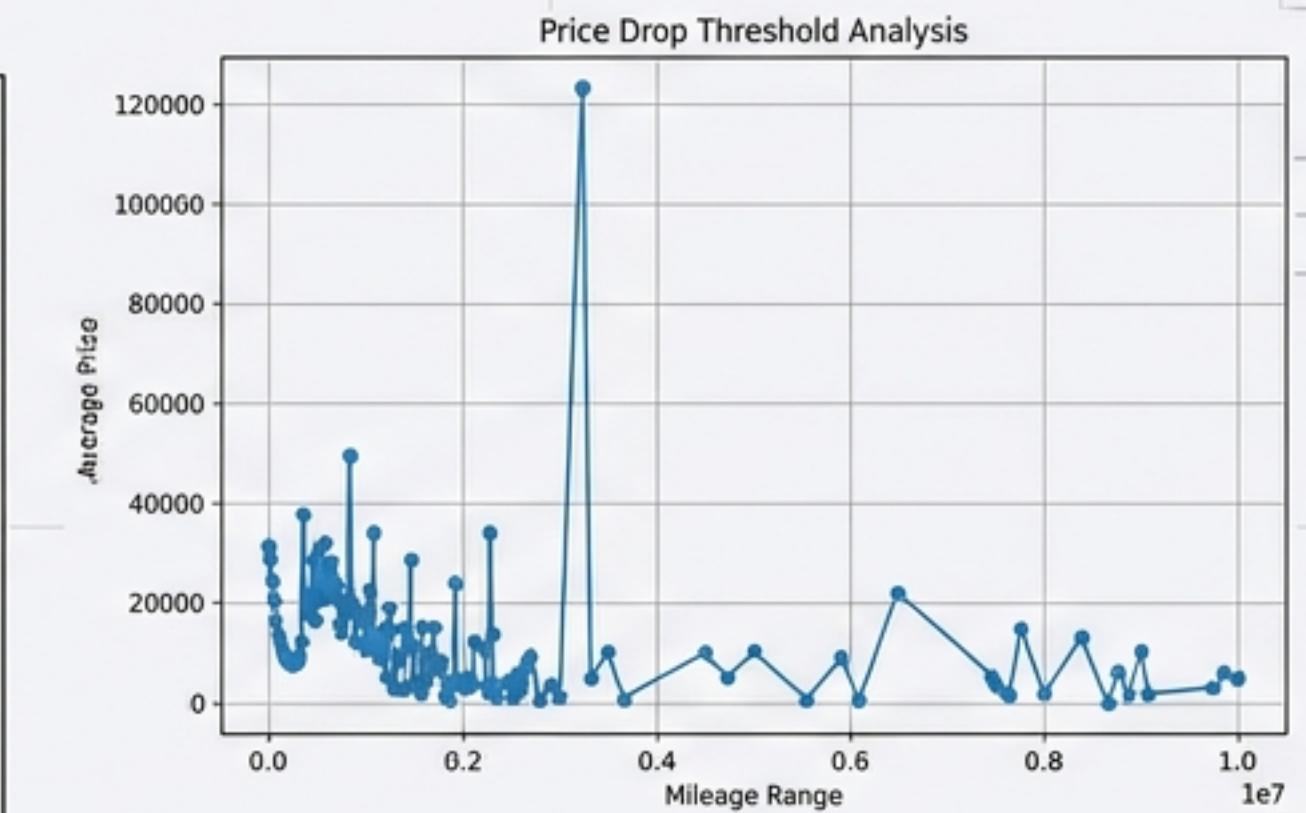
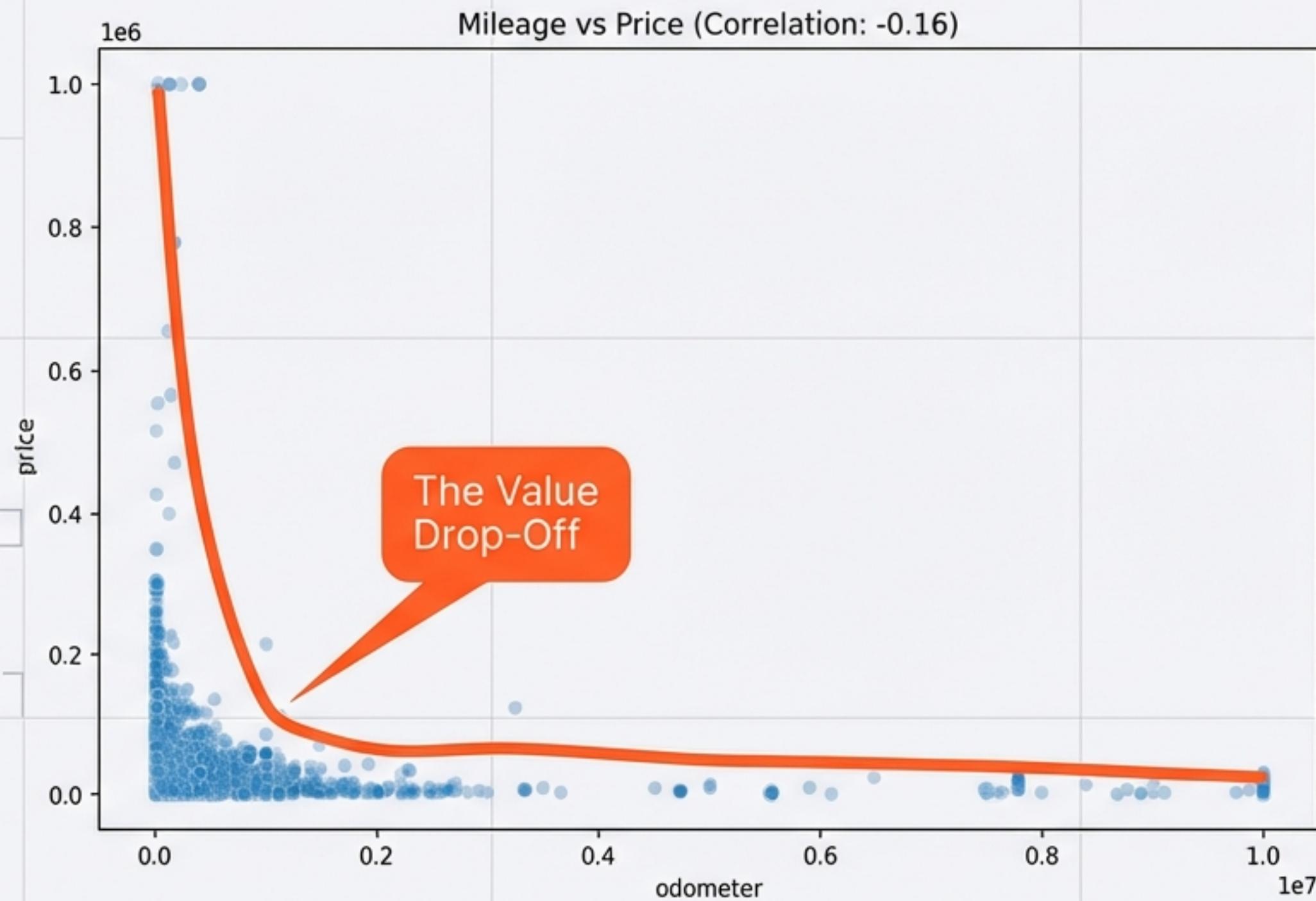
MARKET VOLUME INSIGHT: Despite the premium, 'Full Size' (8-Cyl) and 'Mid Size' engines dominate market share, far exceeding 4-cylinder ('Compact') options.

Fuel Type Economics: The Diesel Premium



Counter-Intuitive Finding:
While EVs get the headlines,
Diesel utility vehicles hold the
strongest residual value in the
used market, likely driven by
heavy-duty trucks and engine
longevity.

Mileage Analysis: The Depreciation Curve

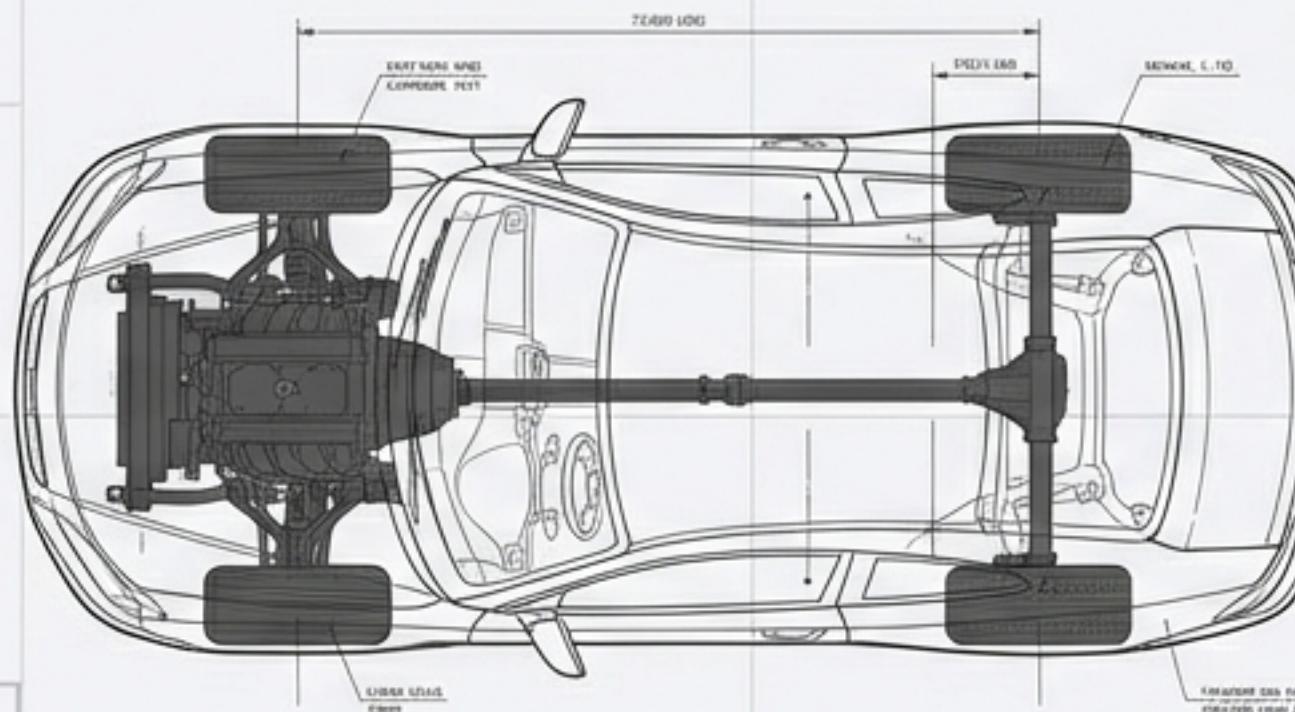


CORRELATION:
-0.16
JetBrains Mono

5-YEAR BENCHMARK:
54,223 Miles
(Avg for 2017
models)



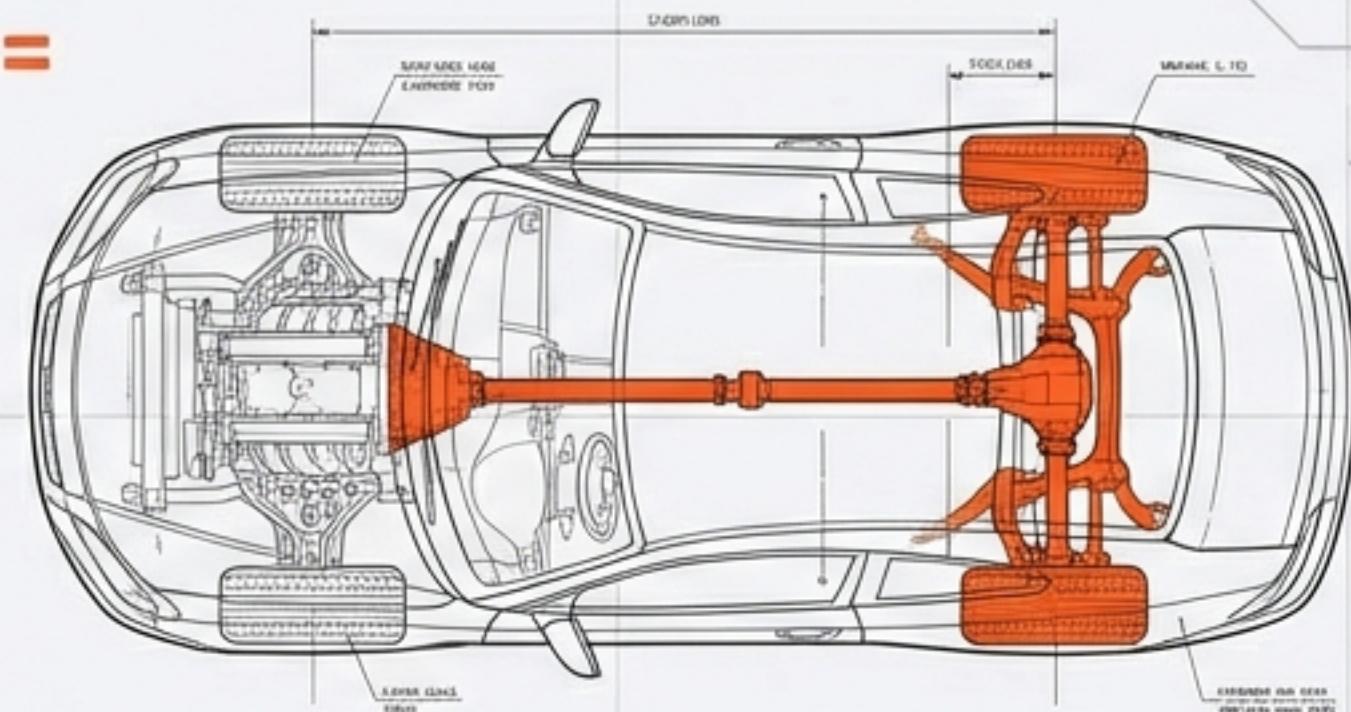
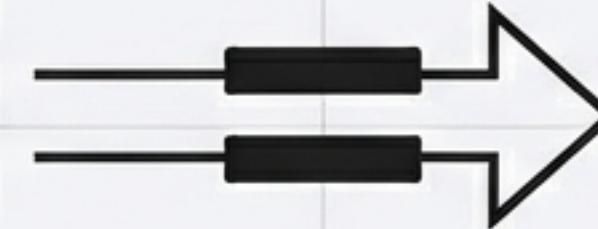
Drivetrain Dynamics: FWD vs. RWD



FWD (Economy Standard)

\$ Low Value ↴

RWD Premium =
+\$8,719



RWD (Performance/Luxury)

\$ High Value ↗

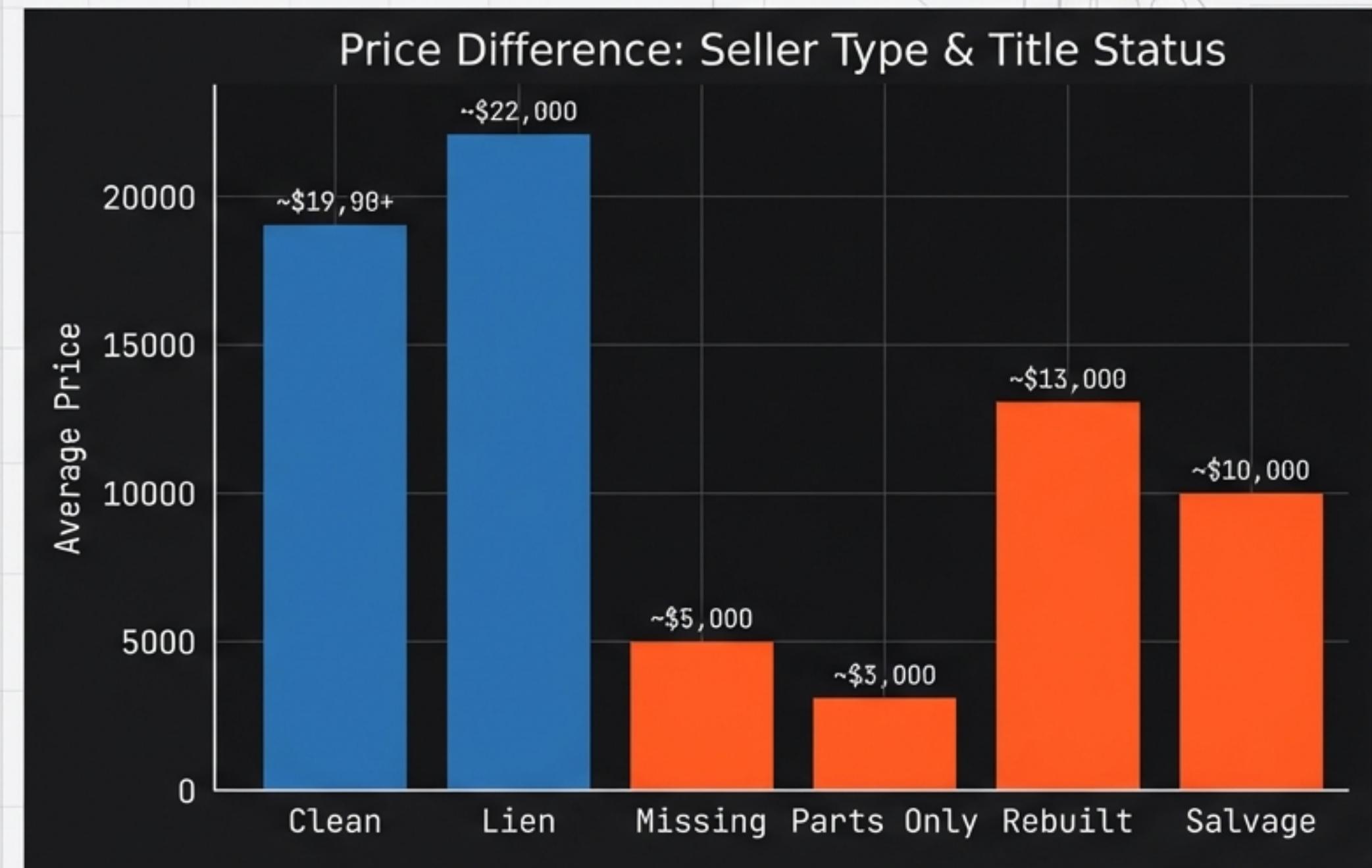
The gap highlights that buyers pay significantly for performance characteristics and towing capability over basic transportation.

JetBrains Mono

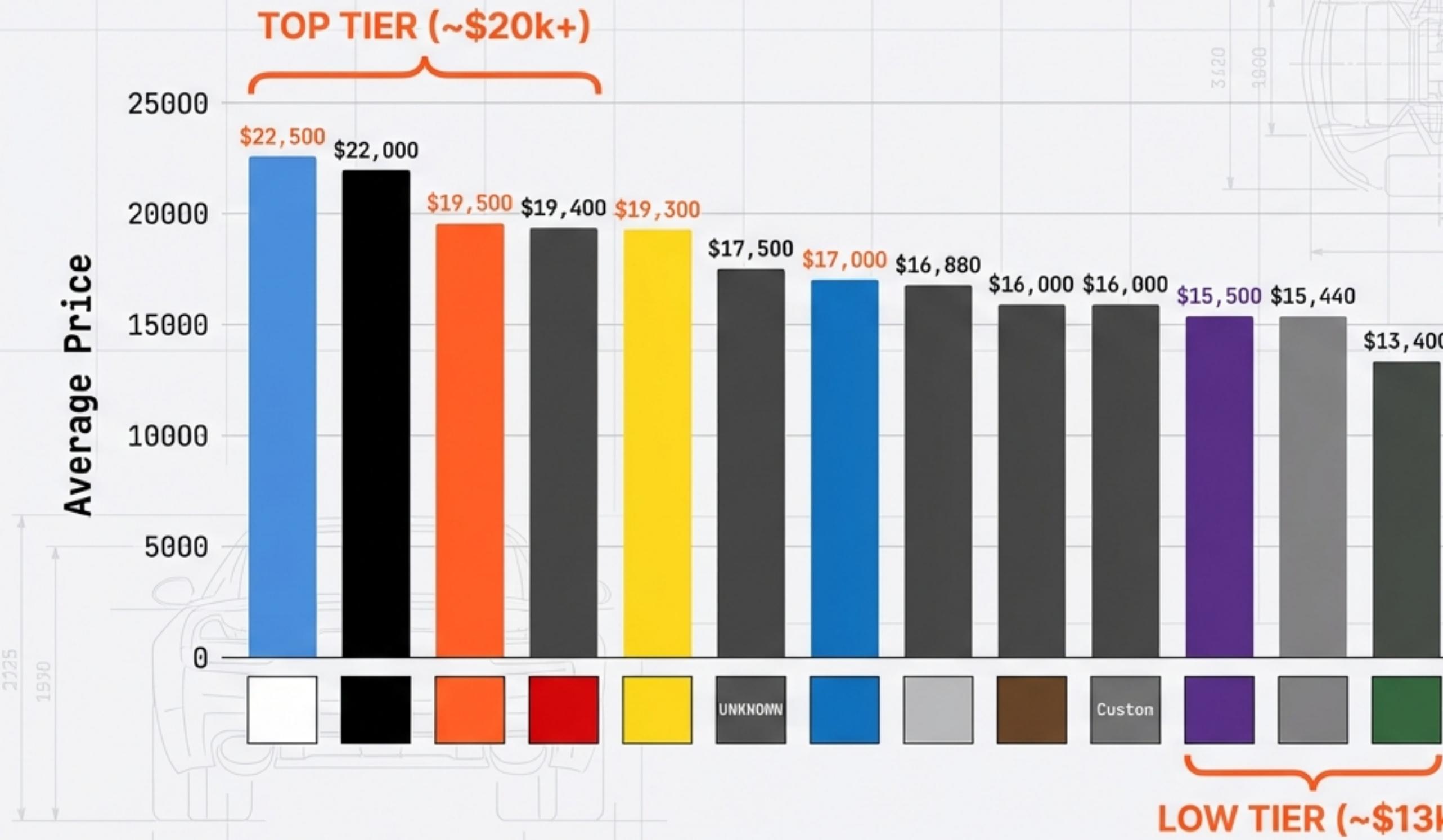
Seller Types & Title Status

- 1. CLEAN TITLE:** ~\$19k+
(Market Standard)
- 2. REBUILT:** ~\$13k
(High Risk/Reward)
- 3. SALVAGE:** ~\$3k-\$10k
(Parts Value)

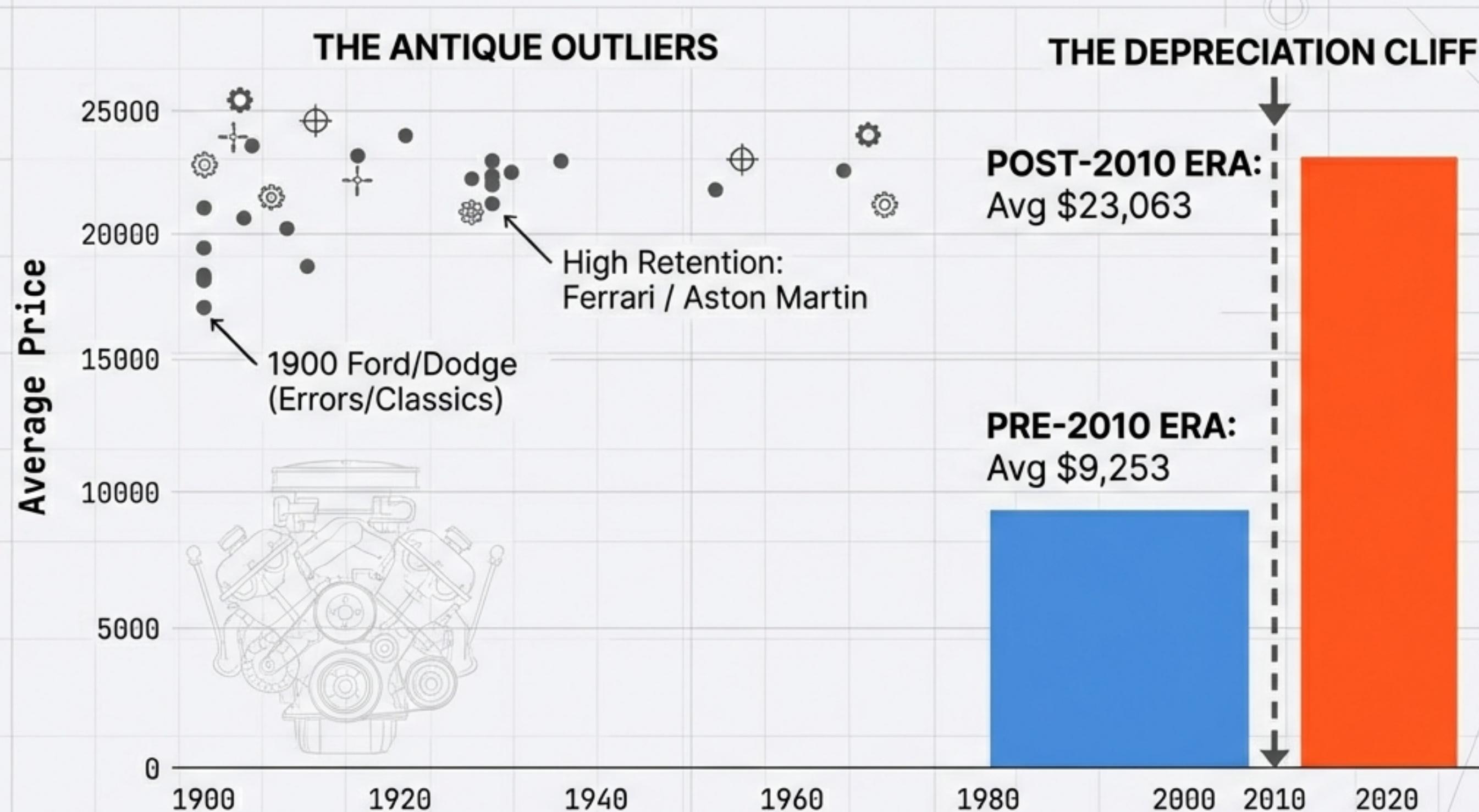
Insight: "Lien" listings average higher (~\$22k) as they represent newer vehicles still under financing.



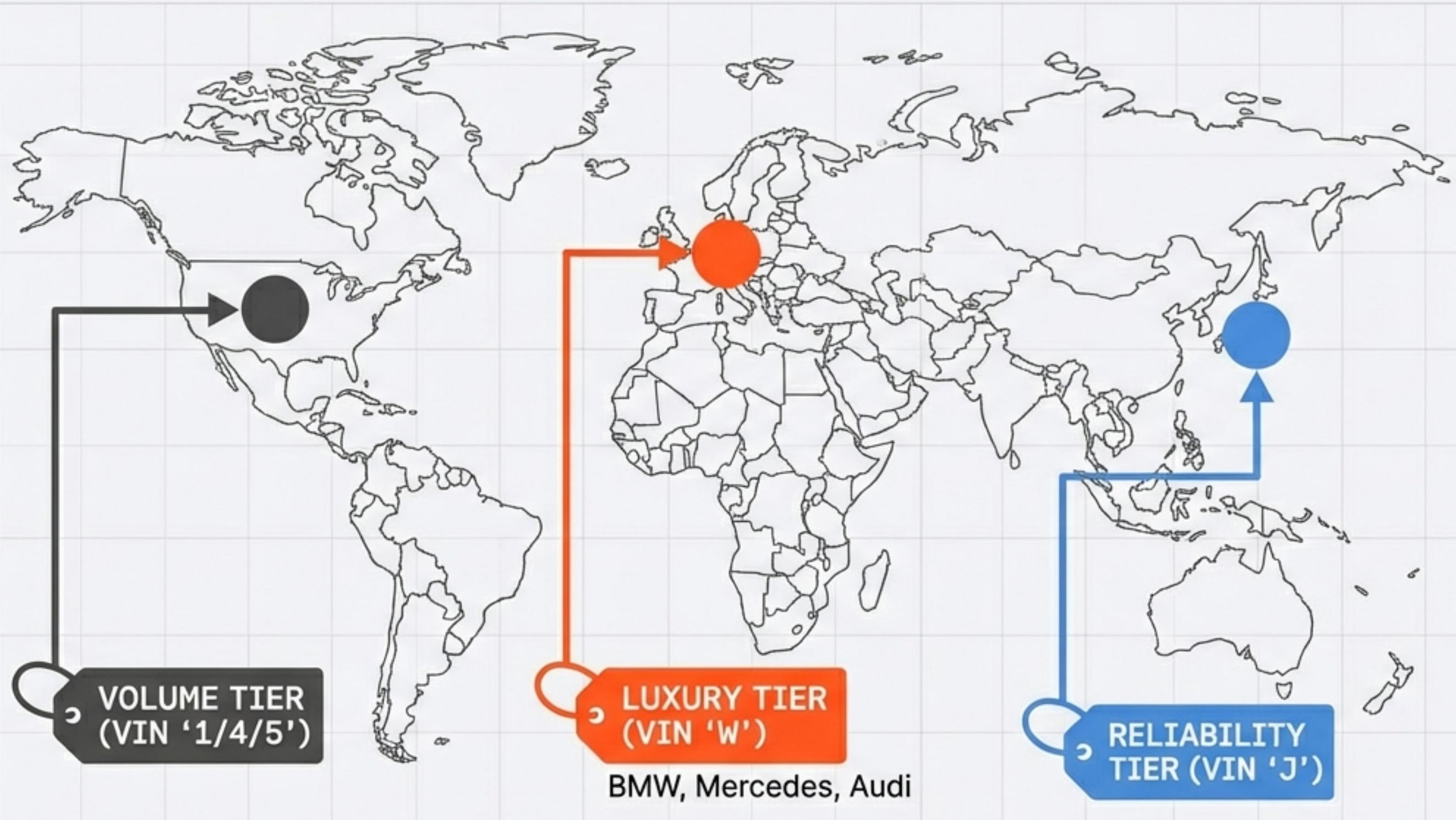
Aesthetic Value: Does Color Matter?



Age & Vintage: The Curve of Time



Global Origins: Value by Geography



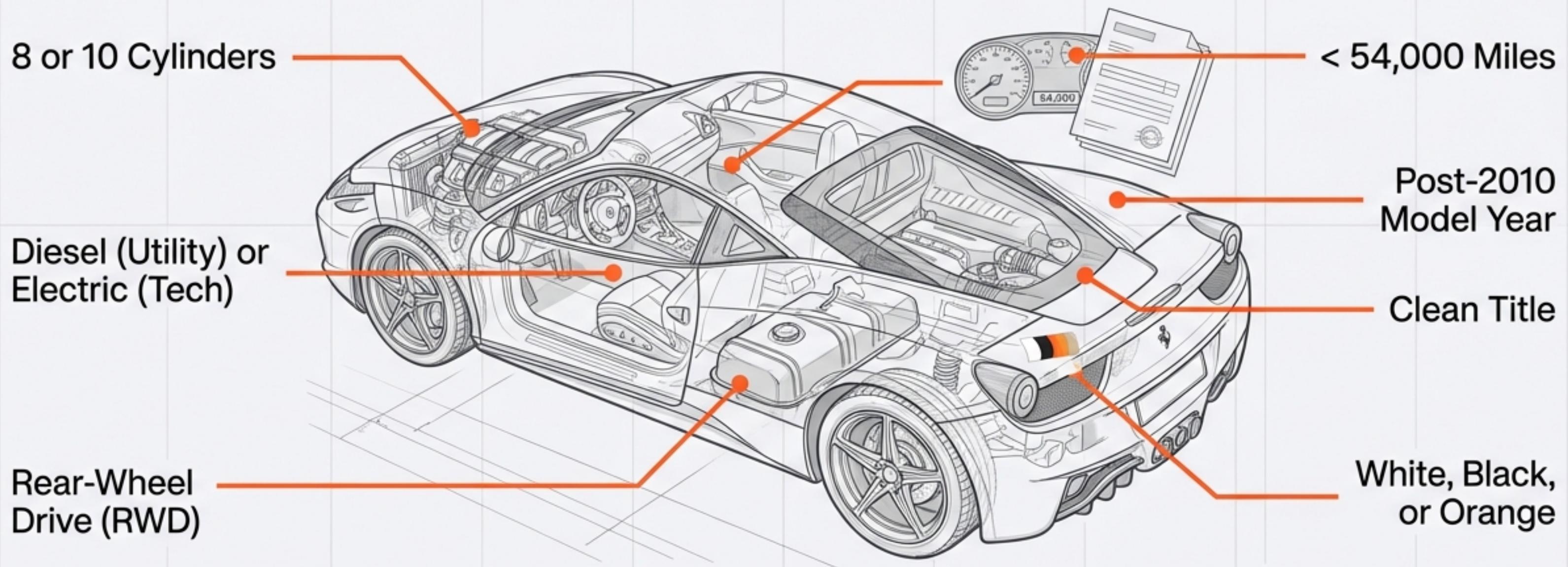
Methodology:

Origin countries decoded from VIN prefixes.

German engineering consistently aligns with higher price brackets in the used market.

The Anatomy of High Value

Constructing the perfect listing based on 426,000 data points.



While brand is king (Ferrari/Tesla), utility (Diesel/Trucks) provides the sturdiest floor for resale value in the general market.