

Mini Data Analysis & Reporting Activity

1. Title Page

- Activity Title: Mini Data Analysis & Reporting Activity
- Dataset: `car_prices.csv`
- Students: Justin Louis Amper Dampal and Greg Danielle Cabanda Duetes

2. Introduction

The dataset contains used-car sales transactions with fields for vehicle details, seller, condition, odometer reading, market reference value (MMR), final selling price, and sale date. It represents a dealership/auction sales context where each transaction captures a specific sale event.

The data likely originates from an operational automotive sales platform (e.g., dealer auction management or vehicle marketplace transaction system).

This report converts raw operational records into analytical summaries by category, time, and pricing behavior.

3. Data Dictionary

The original dataset contains 16 columns. Each column is described below:

Column	Description	Type	Example	Notes
<code>year</code>	Model year of the vehicle	int	<code>2015</code>	Ranges from ~1990 to 2015 in this dataset
<code>make</code>	Manufacturer / brand of the vehicle	string	<code>FORD</code>	Uppercased during cleaning; 50+ unique makes
<code>model</code>	Specific model name within the make	string	<code>F - 150</code>	Over 1,000 unique model values
<code>trim</code>	Trim level or package variant	string	<code>SE</code>	Indicates feature/ equipment tier; may be missing

Column	Description	Type	Example	Notes
body	Body style classification of the vehicle	string	SEDAN	Uppercased; common values: SEDAN, SUV, CREW CAB
transmission	Transmission type	string	automatic	Typically 'automatic' or 'manual'
vin	Vehicle Identification Number (unique 17-char ID)	string	1FAHP3F2...	Unique per vehicle; serves as a natural key
state	U.S. state where the sale occurred	string	FL	Uppercased two-letter abbreviation
condition	Numeric condition rating of the vehicle	float	3.3	Scale approximately 1.0 (poor) to 5.0 (excellent)
odometer	Mileage reading at time of sale	float	36368.0	Measured in miles; higher values indicate more use
color	Exterior color of the vehicle	string	black	Free-text; some missing values
interior	Interior color of the vehicle	string	gray	Free-text; some missing values
seller	Name of the selling entity	string	hertz auto sales	Dealer or auction house name
mmr	Manheim Market Report value (independent market estimate)	float	14525.0	Benchmark price used to evaluate deal quality
sellingprice	Actual transaction price the vehicle sold for	float	13500.0	Primary metric for pricing analysis
saledate	Full timestamp of the sale event	string	Tue Dec 16 2014 12:30:00 GMT-0800	Parsed into sale_year, sale_month during cleaning

4. OLTP Perspective

One row represents one completed car sale transaction. It stores transactional attributes (what car was sold, by whom, when, and for how much).

This is OLTP data because records are event-level, high-volume, and designed for day-to-day operational recording rather than summarized reporting.

A typical system storing this data is a vehicle sales transaction management system integrated with dealer inventory and auction workflows.

5. Data Analysis (OLAP Results)

5.1 Transactions per Category (Make)

make	transactions
FORD	93997
CHEVROLET	60587
NISSAN	54017
TOYOTA	39966
DODGE	30955
HONDA	27351
HYUNDAI	21836
BMW	20793
KIA	18084
CHRYSLER	17485

5.2 Transactions per Time Period (Year)

sale_year	transactions
2014	53725
2015	505074

5.3 Transactions per Time Period (Month)

year_month	transactions
2014-01	206
2014-02	1
2014-12	53518
2015-01	140611
2015-02	163052
2015-03	46277
2015-04	1450
2015-05	52444
2015-06	99938
2015-07	1302

5.4 Transactions per Category (Body Type)

body	transactions
SEDAN	241343
SUV	143844
HATCHBACK	26237
MINIVAN	25529
COUPE	17752
CREW CAB	16394
WAGON	16129
CONVERTIBLE	10476
SUPERCREW	9033
G SEDAN	7417

5.5 Transactions per Category (State)

state	transactions
FL	82945
CA	73148
PA	53907
TX	45913
GA	34750
NJ	27784
IL	23486
NC	21845
OH	21575
TN	20895

5.6 Additional Aggregation: Average Selling Price by Make

make	avg_sellingprice
ROLLS-ROYCE	153488.24
FERRARI	127210.53
LAMBORGHINI	112625.00
BENTLEY	74367.67
AIRSTREAM	71000.00
TESLA	67054.35
ASTON MARTIN	54812.00
FIKER	46461.11
MASERATI	44947.06
LOTUS	40800.00

5.7 Additional Aggregation: Average Price vs MMR by Make

(Positive values indicate selling above MMR; negative values indicate below MMR.)

make	avg_price_vs_mmr
AIRSTREAM	41500.00
ASTON MARTIN	1252.00
LAMBORGHINI	1125.00
MAZDA TK	1125.00
LOTUS	500.00
GMC TRUCK	429.55
HUMMER	252.62
HYUNDAI TK	225.00
LANDROVER	197.22
CHEV TRUCK	75.00

6. Findings and Insights

6.1 Volume and Scale

The dataset contains **558,837** individual sale transactions, representing a substantial operational record suitable for identifying market-wide patterns. The sheer volume confirms that the source system handles high-frequency, day-to-day transactional activity typical of an OLTP environment.

6.2 Category Patterns

FORD leads all manufacturers with **93,997** transactions, followed by Chevrolet and Nissan. This dominance reflects Ford's historically large share of the U.S. vehicle market, its wide model lineup (sedans, trucks, SUVs), and high fleet and rental-channel volume that feeds into auction resale pipelines.

The dominant body type is **SEDAN** at **241,343** transactions, accounting for roughly 43% of all sales. SUVs follow at approximately 26%. This split mirrors broader U.S. consumer preferences during the 2014-2015 period, when sedans still held the largest share of the used-car market before the industry-wide shift toward crossovers and trucks.

6.3 Geographic Distribution

FL is the most active state with **82,945** transactions, followed by CA and PA. Florida's leading position is likely driven by its large population, favorable climate (which reduces vehicle corrosion and preserves condition), and concentration of major auction houses such as Manheim. California follows due to its status as the largest U.S. auto market by population.

6.4 Time-Based Trends

The vast majority of transactions — **505,074** — occurred in **2015**, with only **53,725** in the prior year. This suggests the dataset captures a system that either ramped up operations or expanded data collection during this period.

The peak month is **2015-02** with **163,052** transactions. Monthly data reveals a general upward trend from late 2014 through early 2015, which may reflect seasonal patterns in wholesale auction activity where volumes typically increase in Q1 as dealers replenish inventory after year-end.

6.5 Pricing Insights

The average selling price across all valid records is **\$13,611.36**. More notably, vehicles sold at an average of **\$-158.02** relative to MMR (Manheim Market Report), meaning most cars transacted slightly below independent market valuation. This negative gap suggests a buyer-favorable market at scale, which is typical of high-volume wholesale auctions where speed of sale is prioritized over maximizing individual unit profit.

Luxury and specialty brands (Rolls-Royce, Ferrari, Lamborghini) command dramatically higher average prices but represent a very small fraction of total volume. Some niche segments like Airstream and Aston Martin show positive price-vs-MMR gaps, indicating that scarcity and collector demand can push certain vehicles above their reference value.

7. Recommendation

1. **Prioritize high-volume inventory.** Ford, Chevrolet, and Nissan collectively account for a large share of all transactions. Dealers and auction operators should ensure adequate supply and competitive pricing for these makes to maintain turnover and buyer engagement.
2. **Tighten pricing for below-MMR segments.** The overall negative price-vs-MMR gap indicates an opportunity to improve pricing discipline. By identifying which specific makes and models consistently sell below market value, sellers can adjust reserve prices or improve vehicle presentation to narrow the gap.

3. **Leverage seasonal peaks for operational planning.** The strong monthly fluctuations — particularly the surge in early 2015 — suggest that staffing, logistics, and marketing budgets should be aligned with predictable high-volume periods rather than spread evenly across the year.
4. **Monitor niche segments for premium opportunities.** Brands that consistently sell above MMR (such as specialty and luxury makes) represent margin opportunities. Tracking which segments sustain above-market pricing can guide targeted acquisition strategies for higher-margin inventory.

Appendix: Method Notes

- Missing values in `sellingprice`, `mmr`, and `saledate` were handled with coercion and excluded only where required per metric.
- Time-series summaries were derived from parsed `saledate` values (`sale_year`, `sale_month`).
- Full summary tables are available in the `/outputs/tables` folder.