







UCB Project 1

Screenshots of data summaries from Kaggle

About car_prices.csv

The "Vehicle Sales and Market Trends Dataset" provides a comprehensive collection of information pertaining to the sales transactions of various vehicles. This dataset encompasses details such as the year, make, model, trim, body type, transmission type, VIN (Vehicle Identification Number), state of registration, condition rating, odometer reading, exterior and interior colors, seller information, Manheim Market Report (MMR) values, selling prices, and sale dates.

# year 	Δ make 		Δ model 		Δ trim 		Δ body 	
The manufacturing year of the vehicle.	The brand or manufacturer of the vehicle.		The specific model of the vehicle.		Additional designation for the vehicle model.		The body type of the vehicle (e.g., SUV, Sedan).	
	Ford	17%	Altima	3%	Base	10%	Sedan	36%
	Chevrolet	11%	F-150	3%	SE	8%	SUV	21%
	Other (405086)	72%	Other (525009)	94%	Other (459372)	82%	Other (240108)	43%
2015	Kia		Sorento		LX		SUV	
2015	Kia		Sorento		LX		SUV	
2014	BMW		3 Series		328i SULEV		Sedan	
2015	Volvo		S60		T5		Sedan	
2014	BMW		6 Series Gran Coupe		650i		Sedan	

<div> <div>▲</div> <div>transmission</div> <div>≡</div> </div> <div>The type of transmission in the vehicle (e.g., automatic).</div>	<div> <div>▲</div> <div>vin</div> <div>≡</div> </div> <div>Vehicle Identification Number, a unique code for each vehicle.</div>	<div> <div>▲</div> <div>state</div> <div>≡</div> </div> <div>The state where the vehicle is registered.</div>	<div> <div>#</div> <div>condition</div> <div>≡</div> </div> <div>Condition of the vehicle, possibly rated on a scale.</div>	<div> <div>#</div> <div>odometer</div> <div>≡</div> </div> <div>The mileage or distance traveled by the vehicle.</div>
<div> <div>automatic</div> <div>85%</div> </div> <div> <div>[null]</div> <div>12%</div> </div> <div> <div>Other (17570)</div> <div>3%</div> </div>	<div>550298</div> <div>unique values</div>	<div> <div>fl</div> <div>15%</div> </div> <div> <div>ca</div> <div>13%</div> </div> <div> <div>Other (402744)</div> <div>72%</div> </div>		
automatic	5xyktca69fg566472	ca	5	16639
automatic	5xyktca69fg561319	ca	5	9393
automatic	wba3c1c51ek116351	ca	45	1331
automatic	yv1612tb4f1310987	ca	41	14282
automatic	wba6b2c57ed129731	ca	43	2641
automatic	1n4a13ap1fn326013	ca	1	5554

Kaggle Data Source

There is a very good summary of the data in the link below. The analysis is web-based and provides definitions and summary statistics. This is available on a PPT separate from the presentation PPT.

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

Dataset

car_prices.csv

Provides good code for dataset, but not needed

<https://www.kaggle.com/code/yusupibrahim/selling-price-predic-with-stacked-gen-r-2-0-96>

Dataset Description

- The "Vehicle Sales and Market Trends Dataset" provides a comprehensive collection of information pertaining to the sales transactions of various vehicles. This dataset encompasses details such as the year, make, model, trim, body type, transmission type, VIN (Vehicle Identification Number), state of registration, condition rating, odometer reading, exterior and interior colors, seller information, Manheim Market Report (MMR) values, selling prices, and sale dates.
- Note the original data was called "car_prices.csv" but this was cleaned and we used "Data/clean_car_prices.csv"

Key Features:

- **Vehicle Details:** Includes specific information about each vehicle, such as its make, model, trim, and manufacturing year.
- **Transaction Information:** Provides insights into the sales transactions, including selling prices and sale dates.
- **Market Trends:** MMR values offer an estimate of the market value of each vehicle, allowing for analysis of market trends and fluctuations.
- **Condition and Mileage:** Contains data on the condition of the vehicles as well as their odometer readings, enabling analysis of how these factors influence selling prices.

