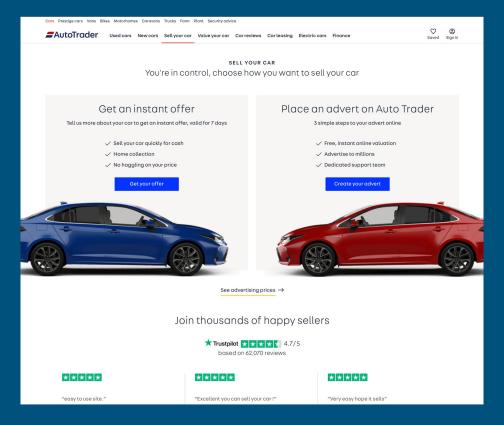# Big Data
for more accurate
prediction of car prices
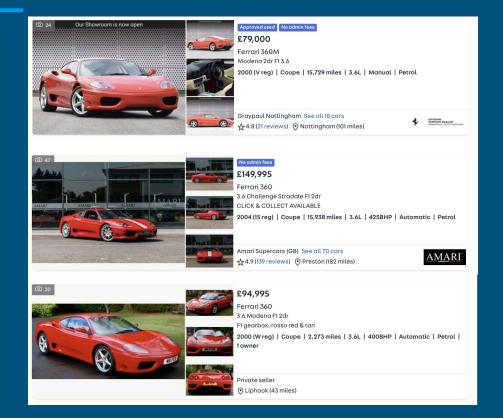
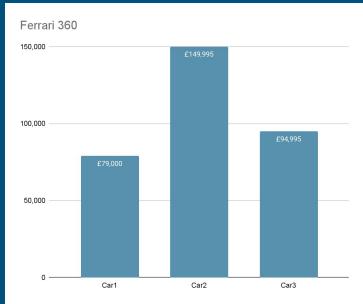# Selling Cars

Selling Cars

Buying Cars - accurate price
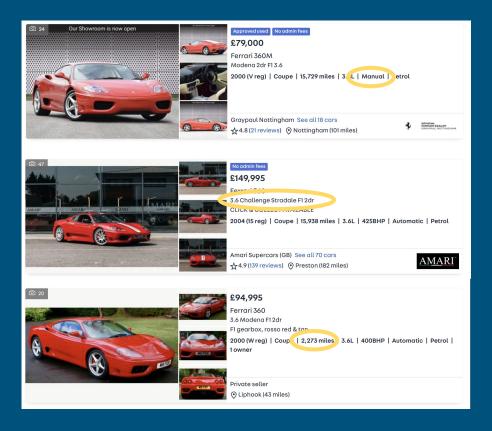
Machine Learning

Big Data

25 Worst results

	Make	MSRP	Predicted price	Error in price	Percentage error
1	Make	MSRP	Predicted price	Error in price	Percentage error
2	Bugatti	2,065,902.00	29,042.00	2,036,860.00	99%
3	Bugatti	1,705,769.00	29,042.00	1676727.001	98%
4	Chrysler	31,465.00	1,697,678.86	1666213.865	5295%
5	Chrysler	35,845.00	1,697,678.86	1661833.865	4636%
6	Chrysler	40,295.00	1,697,678.86	1657383.865	4113%
7	Lamborghini	1,500,000.00	2,056.53	1497943.467	100%
8	Bugatti	1,500,000.00	29,042.00	1470958.001	98%
9	Maserati	109,900.00	1,527,815.83	1417915.831	1290%
10	Maybach	1,382,750.00	49,010.43	1333739.568	96%
11	Maybach	1,380,000.00	49,010.43	1330989.568	96%
12	Scion	15,665.00	1,299,771.84	1284106.838	8197%
13	Lexus	37,325.00	1,298,622.59	1261297.588	3379%
14	Ferrari	643,330.00	35,163.08	608166.9175	95%
15	Lamborghini	548,800.00	2,056.53	546743.4669	100%
16	Lamborghini	548,800.00	36,167.26	512632.7426	93%
17	Lamborghini	535,500.00	23,758.22	511741.7844	96%
18	Rolls-Royce	492,000.00	2,056.53	489943.4669	100%
19	Ford	3,206.00	484,346.09	481140.0875	15007%
20	Chevrolet	43,245.00	524,185.68	480940.6833	1112%
21	Rolls-Royce	479,775.00	2,056.53	477718.4669	100%
22	Chevrolet	50,305.00	526,378.93	476073.9333	946%
23	Ford	3,550.00	478,092.89	474542.8873	13367%
24	Rolls-Royce	474,600.00	2,056.53	472543.4669	100%
25	Toyota	36,470.00	507,748.82	471278.8167	1292%
26	Rolls-Royce	492,425.00	21,683.14	470741.8566	96%

25 Best Results

	Make	MSRP	Predicted price	Error in price	Percentage error
1	Make	MSRP	Predicted price	Error in price	Percentage error
2	Toyota	2,064.00	2,056.53	7.466896734	0%
3	Mercedes-Benz	2,105.00	2,114.35	9.349841301	0%
4	Chrysler	23,950.00	23,940.53	9.468908899	0%
5	Mercedes-Benz	2,124.00	2,114.35	9.650158699	0%
6	Chevrolet	26,695.00	26,684.50	10.50177564	0%
7	Dodge	2,046.00	2,056.53	10.53310327	1%
8	Mitsubishi	29,495.00	29,508.60	13.60329002	0%
9	Nissan	28,910.00	28,923.61	13.61078289	0%
10	Dodge	2,073.00	2,056.53	16.46689673	1%
11	Toyota	29,065.00	29,042.00	23.00114387	0%
12	Infiniti	2,081.00	2,056.53	24.46689673	1%
13	Mazda	40,170.00	40,196.21	26.20640179	0%
14	Honda	25,030.00	25,001.80	28.1985325	0%
15	Acura	29,900.00	29,869.43	30.57419396	0%
16	Dodge	22,695.00	22,662.86	32.14326696	0%
17	Pontiac	22,220.00	22,254.66	34.66404973	0%
18	Honda	22,065.00	22,029.80	35.19541745	0%
19	Dodge	2,093.00	2,056.53	36.46689673	2%
20	Suzuki	29,079.00	29,042.00	37.00114387	0%
21	Volvo	29,000.00	29,042.00	41.99885613	0%
22	Nissan	16,780.00	16,735.82	44.17810097	0%
23	Mazda	21,430.00	21,384.57	45.42972491	0%
24	Subaru	32,495.00	32,448.54	46.4636507	0%
25	Subaru	28,995.00	29,042.00	46.99885613	0%
26	Lincoln	28,995.00	29,042.00	46.99885613	0%

Results

Random Forest

- R² score of 0.9943 on the validation set
- Mean Absolute Error = £37,448.56
- Quite bad
- DataSet was too varied

Our original Ferrari

	Make	Model	Year	Vehicle Style	MSRP	y_hat	Error in price £	Percentage error
4465	Ferrari	360	2004	Coupe	187,124.00	34087.2364	153036.7636	81.78361067
4512	Ferrari	360	2004	Convertible	180,408.00	34890.0115	145517.9885	80.66049647
4574	Ferrari	360	2003	Convertible	165,986.00	34890.0115	131095.9885	78.98014801
4603	Ferrari	360	2004	Coupe	157,767.00	34087.2364	123679.7636	78.39393767
4652	Ferrari	360	2002	Coupe	150,694.00	34087.2364	116606.7636	77.37983174
4655	Ferrari	360	2003	Coupe	154,090.00	34890.0115	119199.9885	77.35738106
4697	Ferrari	360	2003	Coupe	143,860.00	34087.2364	109772.7636	76.30527154
4772	Ferrari	360	2003	Convertible	176,287.00	44943.2966	131343.7034	74.50560927
4804	Ferrari	360	2002	Convertible	170,829.00	44943.2966	125885.7034	73.69106147
4810	Ferrari	360	2004	Convertible	169,900.00	44948.1698	124951.8302	73.54433795
4854	Ferrari	360	2002	Convertible	160,829.00	44340.1744	116488.8256	72.43023684
5020	Ferrari	360	2004	Coupe	147,332.00	44943.2966	102388.7034	69.49522399
5096	Ferrari	360	2002	Coupe	140,615.00	44948.1698	95666.83017	68.03458391

Future extensions

- More data & separate datasets
- Frontend that inputs make + model
- Database of cars

Data Science Aspects

Data Cleaning

- Mean Encoding: Make + Model

- One Hot Encoding:
    - Engine Fuel Type
    - Transmission Type
    - Driven_Wheels
    - Vehicle Size
    - Vehicle Style
    - Market Category

Validation

- Split into train, test and validation
- Train - training the model
- Test - check the score of the model
- Validation - after fitting training data and testing it -- validate the final output

Different models

Linear regression (baseline) - 0.9839

Random-Forest (best)

- Without tuning: 0.9940
- With Tuning: 0.9943

Model	Validation R² Score	Mean Absolute Error
K-Nearest	0.969	£37,126
Linear Regressor	0.984	£37,649
Decision Tree	0.992	£37,467
RandomForest	0.994	£37,448

Hyper tuning

- Adjust parameters on each model

- DecisionTree + RandomForest
    - Max-Leaf nodes
    - Max-Depth

- K-Nearest
    - Algorithm
    - N - neighbors
    - Leaf Size

Results

Random Forest

- R² score of 0.9943 on the validation set
- Mean Absolute Error = £37,448.56
- RMSE = £83,330.84

Extension Work

- Larger data-set

- Feature engineering

- Missing data

- Stacked ensemble
    - Filter by make+model and then use regression

Conclusion