DATA MINING AND MACHINE LEARNING

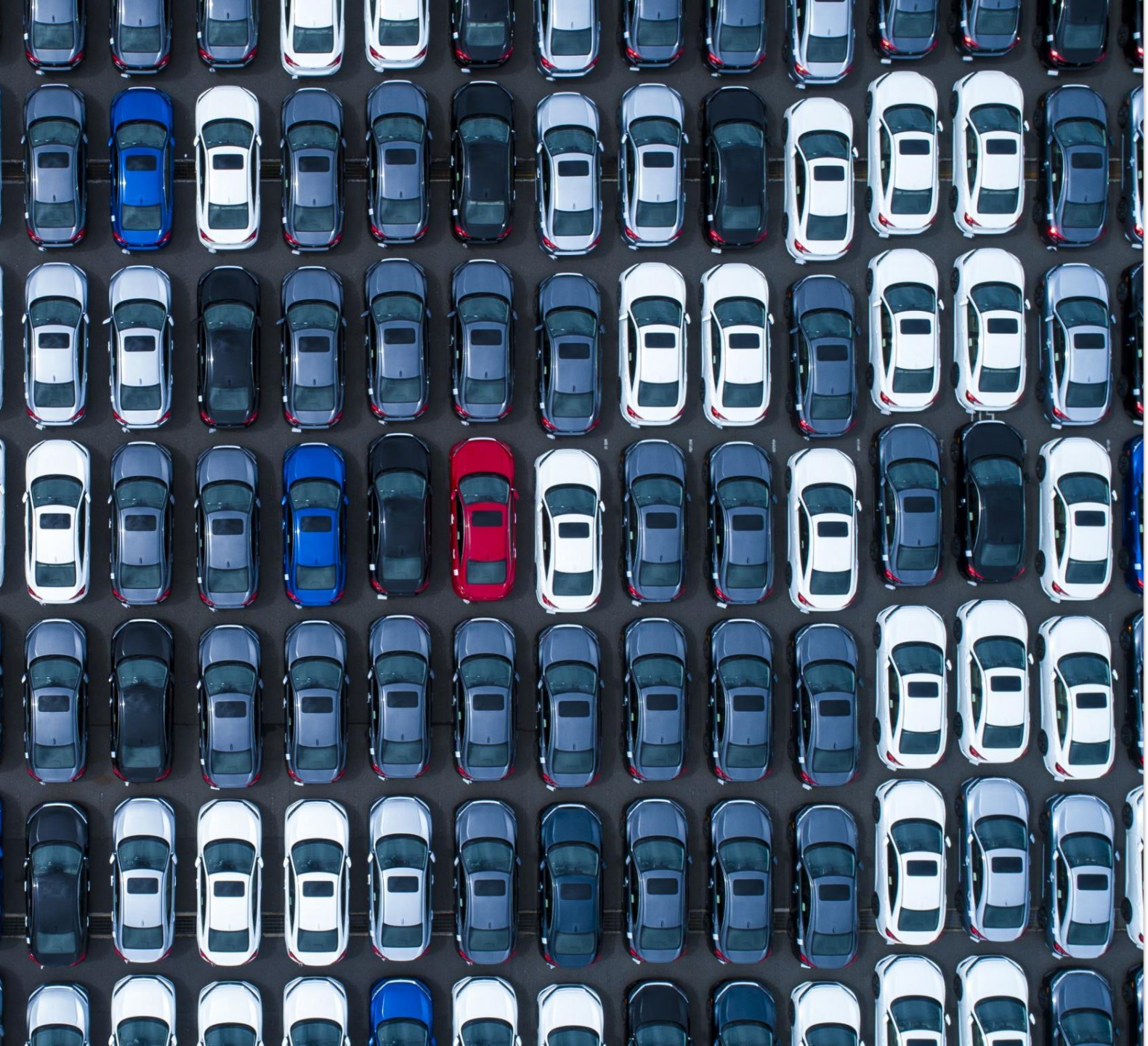
USED CAR VALUE PREDICTOR

2021/2022

Filippo Puccini
Matteo Mugnai

INTRODUCTION AND GOALS

- The aim of our project is to build a model able to predict a fair price for selling a used car, through a regression algorithm based on supervised learning.
- Why customers prefer to buy used car?
  - Long delivery and shipping time for new cars (Semiconductor crisis, pandemic)
  - Save money
  - Uncertain about fuel type price evolution



WHY IS IMPORTANT TO BUILD SUCH A MODEL?

- Customers have a reliable tool to evaluate the price of a car
- Sellers can be sure to put a car on the market with a fair and competitive price

DATASET

- **Source**: the dataset was obtained through scraping on https://www.autoscout24.com/
- 25786 instances and 33 columns
- **Manufacturers**: *Audi, BMW, Ford, Mercedes-Benz, Opel, Renault, Volkswagen, Alfa Romeo, Chevrolet, Chrysler, Citroen, Cupra, Dacia, Daihatsu, Dodge, Fiat, Honda, Hyundai, Infiniti, Jeep, Kia, Lancia, Lexus, Mazda, MINI, Mitsubishi, Nissan, Peugeot, SEAT, Skoda, smart, Suzuki, Toyota, Volvo*
- **Key features extracted:**
  - General (Manufacturer, model, type)
  - Historical (km, year, number of previous owner)
  - Technical (Engine power, transmission, traction, weight, drive type)
  - Consumption (fuel, consumption, emissions)
  - Equipment (comfort, optional, other features)
  - Aesthetic (colors, interiors)
  - Price

SCRAPING PARAMETERS

Selected cars have the following features:

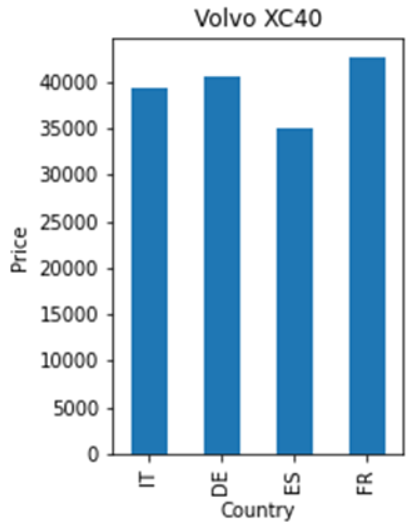
- First registration from 1992 to 2021
- Used
- Comes from all European countries
- Minimum mileage of 2500 km

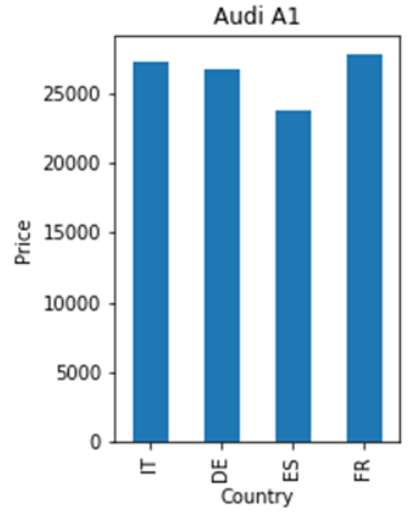
RAW DATASET

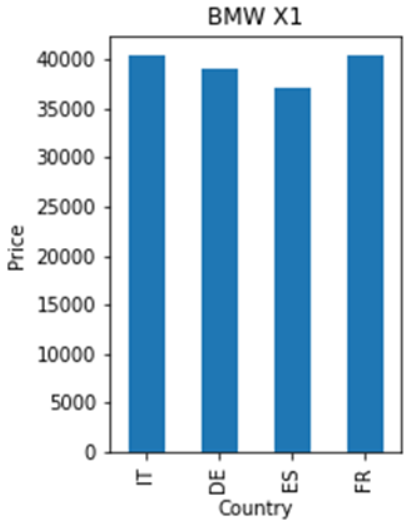
BodyType	CO2Emissi...	Colour	ComfortCo...	Country	Cylinders	Doors	Drivetrain	EmissionCl...	EmptyWei...	EngineSize	Entertainm...	Extras	FirstRegist...	FuelConsu...	FuelType	FullService...	Gearbox	Gears
Categorical	Number	Categorical	Text	Categorical	Number	Number	Number	Number	Number	Number	Text	Text	Categorical	Text	Categorical	Categorical	Categorical	Number
Body type	CO2 Emissi...	Colour	Comfort & ...	Country	Cylinders	Doors	Drivetrain	Emission cl...	Empty weig...	Engine size	Entertainm...	Extras	First registr...	Fuel consu...	Fuel type	Full service ...	Gearbox	Gears
Coupe	306 g/km (c...		Air conditio...	DE	10		Rear	Euro 6d-TE...		5204 cc	Bluetooth; ...	Alloy wheel...	10/2020	12.9 l/100 k...	Super Plus ...	Yes	Automatic	
Coupe		Black	Seat heatin...	ES		3				1984 cc			04/2021	7.3 l/100 k...	Gasoline		Automatic	
Coupe		Black	Air conditio...	ES	8		4WD		1760 kg	4163 cc	Bluetooth; ...	Alloy wheel...	02/2015	12.6 l/100 k...	Gasoline		Automatic	7
Coupe		Grey	Air conditio...	BE	6	2	4WD	Euro 6		2995 cc	Apple CarPl...	Ambient lig...	04/2017		Gasoline	Yes	Automatic	8
Coupe	306 g/km (c...	Red	Air conditio...	BE	10		4WD	Euro 6	1655 kg	5204 cc	Bluetooth; ...	Alloy wheel...	05/2018	13.4 l/100 k...	Gasoline	Yes	Automatic	7
Coupe	287 g/km (c...	Black	Air conditio...	DE	10		4WD	Euro 6	1630 kg	5204 cc	Bluetooth; ...	Alloy wheel...	04/2017	12.3 l/100 k...	Super Plus 98	Yes	Automatic	7
Coupe	222 g/km (c...	Black	360Â° cam...	DE	6	2	4WD	Euro 6d	1770 kg	2894 cc	Android Au...	Alloy wheel...	06/2021	8.8 l/100 k...	Super 95	Yes	Automatic	8
Off-Road/P...		White	Air conditio...	ES	4		Front		1460 kg	1495 cc	Bluetooth;		01/2020	6.1 l/100 k...	Gasoline		Manual	6
Coupe	145 g/km (c...	Silver	Air conditio...	DE		2	Front			1984 cc	Bluetooth; ...	Alloy wheel...	04/2021	6.3 l/100 k...	Super 95		Automatic	
Sedan		Grey	Air conditio...	DE			4WD	Euro 6d		3996 cc	Bluetooth; ...	Alloy wheel...	03/2021		Super 95		Automatic	
Coupe	144 g/km (c...	White	Air conditio...	DE			4WD	Euro 6d-TE...		1984 cc	Android Au...	Alloy wheel...	02/2020	6.3 l/100 k...	Gasoline	Yes	Automatic	
Convertible	163 g/km (c...	Yellow	Air suspensi...	DE	4		4WD	Euro 6	1525 kg	1984 cc	Apple CarPl...	Alloy wheel...	08/2017	7.1 l/100 k...	Gasoline	Yes	Automatic	6
Station wag...	288.6 g/km ...	Black	Air conditio...	DE			4WD	Euro 6d-TE...	2208 kg	3996 cc	Bluetooth; ...	Alloy wheel...	04/2020		Super 95		Automatic	
Sedan		Grey	Air conditio...	ES		5							06/2021		Gasoline		Manual	
Station wag...	290.1 g/km ...	Grey	Air conditio...	DE			4WD	Euro 6d-TE...	2233 kg	3996 cc	Bluetooth; ...	Alloy wheel...	03/2021		Super 95	Yes	Automatic	
Sedan	225.1 g/km ...	Blue	Air conditio...	DE			4WD	Euro 6d	1836 kg	2894 cc	Bluetooth; ...	Alloy wheel...	06/2021		Super 95		Automatic	
Coupe	149 g/km (c...	Black	Air conditio...	BE		2		Euro 6		1984 cc	Bluetooth; ...	Alloy wheel...	01/2015	6.4 l/100 k...	Gasoline (P...	Yes	Automatic	
Sedan		Blue	Air conditio...	DE		4	Front	Euro 6d		1498 cc	Bluetooth; ...	Alloy wheel...	02/2021		Super 95		Automatic	
Off-Road/P...	120 g/km (c...	Green	Air conditio...	DE	4				1365 kg	1498 cc	Bluetooth; ...	Alloy wheel...	04/2021	5.3 l/100 k...	Super 95	Yes	Automatic	7
Off-Road/P...	186 g/km (c...	Black	Air conditio...	DE	4	4	4WD	Euro 6	1845 kg	1968 cc	Bluetooth; ...	Alloy wheel...	02/2020	5.6 l/100 k...	Diesel (Part...	Yes	Automatic	7
Coupe	144 g/km (c...	Violet	Air conditio...	DE	4	3	Front	Euro 6d	1370 kg	1984 cc	Bluetooth; ...	Alloy wheel...	02/2021	6.3 l/100 k...	Super 95	Yes	Automatic	7
Compact	137 g/km (c...	Grey	Air conditio...	DE	4	4	Front	Euro 6		1984 cc	Bluetooth; ...	Alloy wheel...	06/2021	6 l/100 km (...	Super 95	Yes	Automatic	
Station wag...	163 g/km (c...	Black	Air conditio...	DE		4	4WD	Euro 6d-TE...		3996 cc	Android Au...	Alloy wheel...	07/2021	11.5 l/100 k...	Gasoline	Yes	Automatic	
Coupe	146 g/km (c...	Yellow	Air conditio...	BE		2		Euro 6		1984 cc	CD player; ...	Alloy wheel...	10/2017	6.3 l/100 k...	Gasoline (P...	Yes	Automatic	
Off-Road/P...	172 g/km (c...	Grey	Air conditio...	DE	4	4	4WD	Euro 6		1984 cc	Bluetooth; ...	Alloy wheel...	01/2020	7.5 l/100 k...	Super 95	Yes	Automatic	

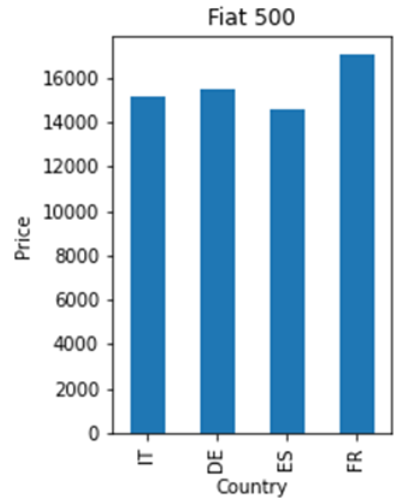
FACTORS THAT AFFECT CAR PRICE: COUNTRY

- List of countries considered: Italy, Germany, Spain, France
- List of cars considered: *Audi A1/A3 , BMW X1/X3, Ford Focus/Fiesta, Volkswagen Golf/Polo, Citroen C1/C3, Renault Captur, Fiat 500, Opel Corsa, Toyota Yaris, Volvo XC40*

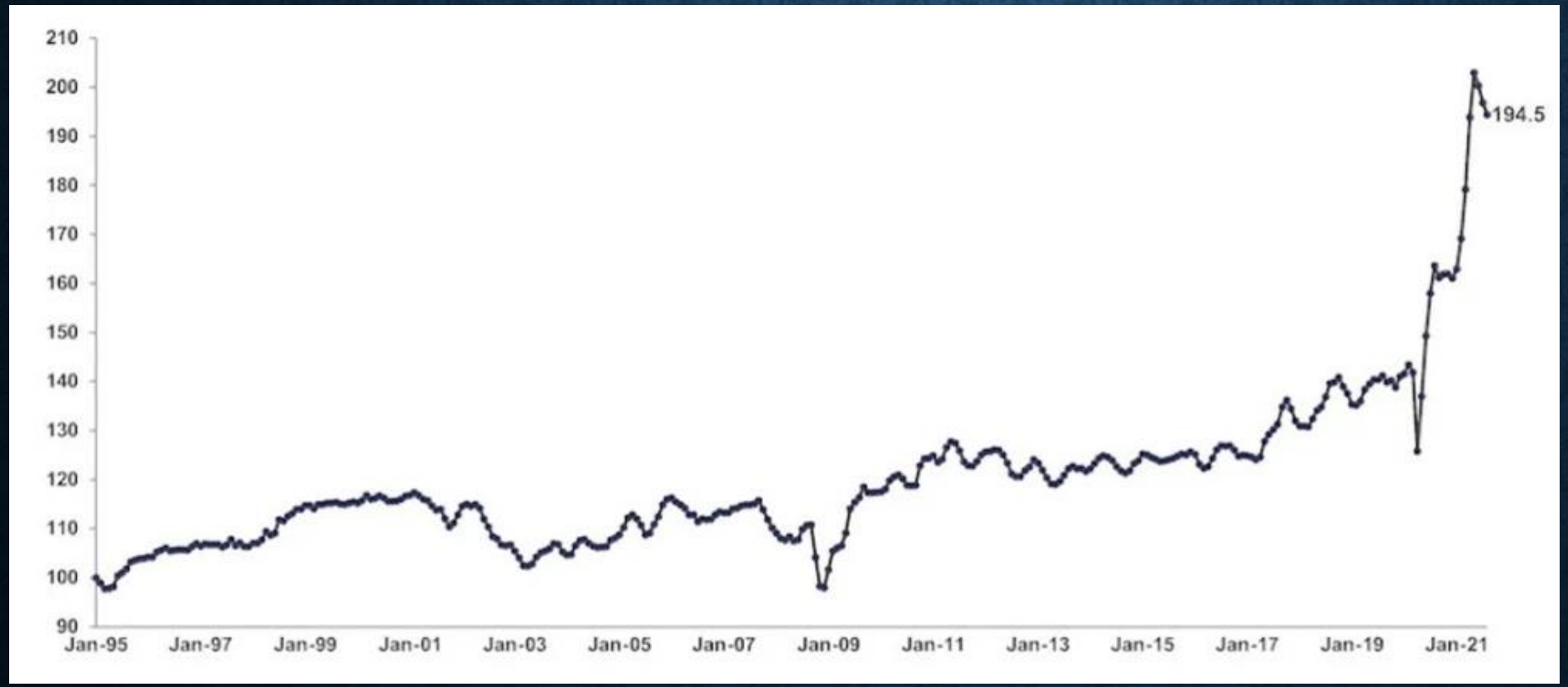








FACTORS THAT AFFECT CAR PRICE: TIME



Car market price fluctuations

FACTORS THAT AFFECT CAR PRICE: SPECIAL CARS

Supercars and vintage cars could strongly condition in some way the final price

→ We perform an additional scraping to test these cars in our classifier

DATA CLEANING: MISSING VALUES

Index	Column	Non-null count	Dtype
0	Body type	25780	object
1	CO2 Emissions	18333	object
2	Colour	24240	object
3	Comfort & Convenience	23776	object
4	Country	25781	object
5	Cylinders	15554	float64
6	Doors	11451	float64
7	Drivetrain	16517	object
8	Emission class	17808	object
9	Empty weight	14612	object
10	Engine size	24708	object
11	Entertainment & Media	22401	object
12	Extras	21928	object
13	First registration	25786	object
14	Fuel consumption	20634	object
15	Fuel type	24489	object
16	Full-service history	11947	object
17	Gearbox	25576	object
18	Gears	16268	float64
19	Manufacturer	25786	object
20	Mileage	25786	object
21	Model	25529	object
22	Other fuel types	2003	object
23	Power	25257	object
24	Previous owner	13309	float64
25	Price	25786	int64
26	Safety & Security	23842	object
27	Seats	23205	float64
28	Seller	0	float64
29	Type	25786	object
30	Upholstery	18498	object
31	Upholstery colour	10748	object
32	Warranty	14420	object

CO2 Emissions --> replace with avg values obtained by taking into consideration the cars of the same model, manufacturer and with the same fuel type

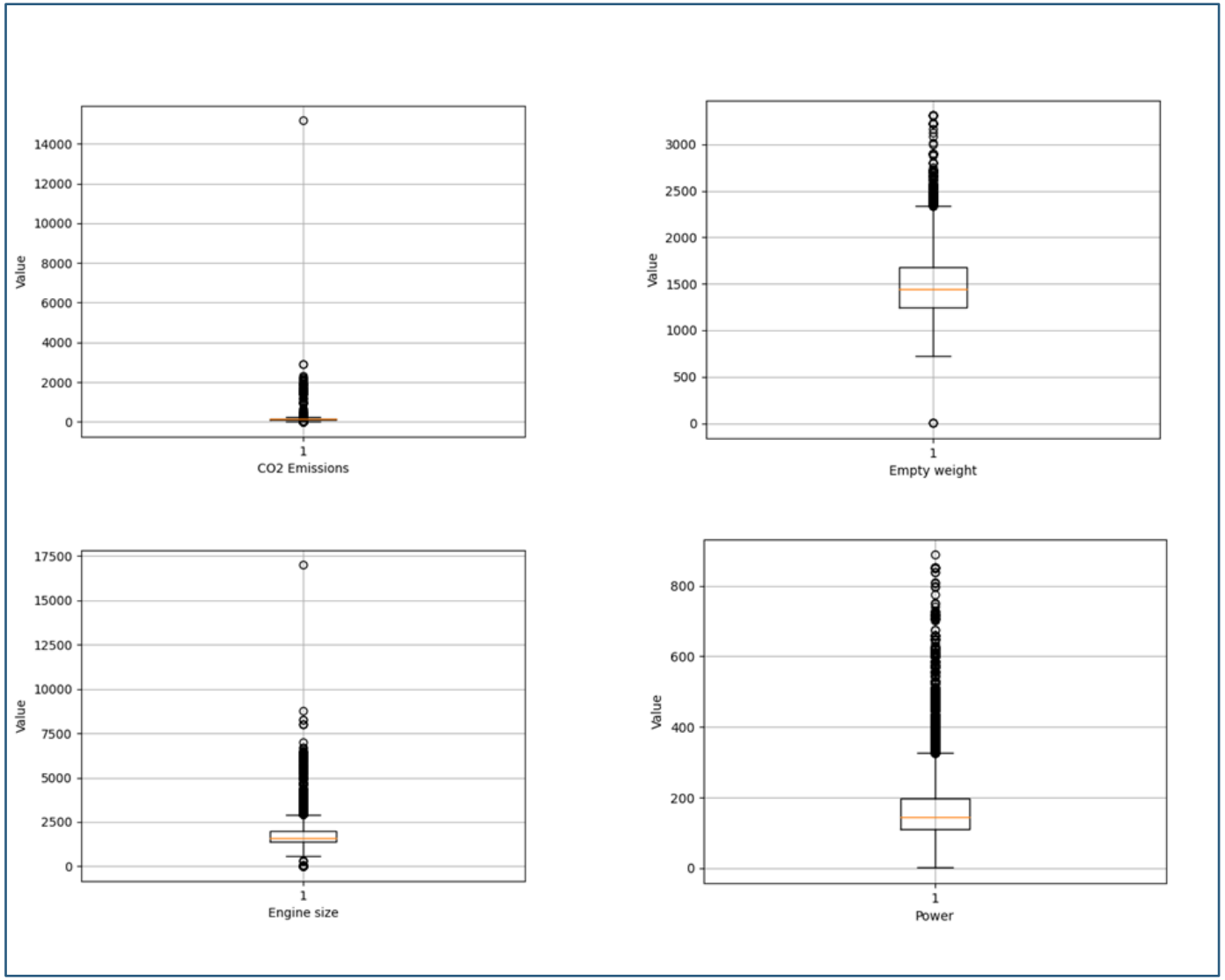
Cylinders, Empty weight, Engine size, Power, Upholstery --> replace with the more common value among cars of the same model and manufacturer

Fuel type, other fuel types --> merge the values of these columns avoiding null values

Emission class --> we consider the year of the car
Doors, Drivetrain, Gears, Gearbox, Seats --> replace with a common value

Body type, Country, Model --> records dropped

DATA CLEANING: OUTLIER DETECTION



Quantile filter approach to remove outliers from:

- CO2 Emissions
- Empty weight
- Engine size
- Power

Python rule-based selection to remove outliers from:

- Mileage
- Gears
- Seats

Index	Column	Non-null count	Dtype
0	Body type	23080	object
1	CO2 Emissions	23080	float64
2	Colour	23080	object
3	Comfort & Convenience	23080	int64
4	Country	23080	object
5	Cylinders	23080	int64
6	Doors	23080	int64
7	4WD	23080	int64
8	Emission class	23080	int64
9	Empty weight	23080	int64
10	Engine size	23080	int64
11	Entertainment & Media	23080	int64
12	Extras	23080	int64
13	First registration	23080	int64
14	Fuel consumption	23080	float64
15	Fuel type	23080	int64
16	Gearbox	23080	int64
17	Gears	23080	int64
18	Manufacturer	23080	object
19	Mileage	23080	int64
20	Model	23080	object
21	Power	23080	int64
22	Previous owner	23080	int64
23	Price	23080	int64
24	Safety & Security	23080	int64
25	Seats	23080	int64
26	Upholstery	23080	int64
27	Upholstery colour	23080	object
28	Warranty	23080	int64

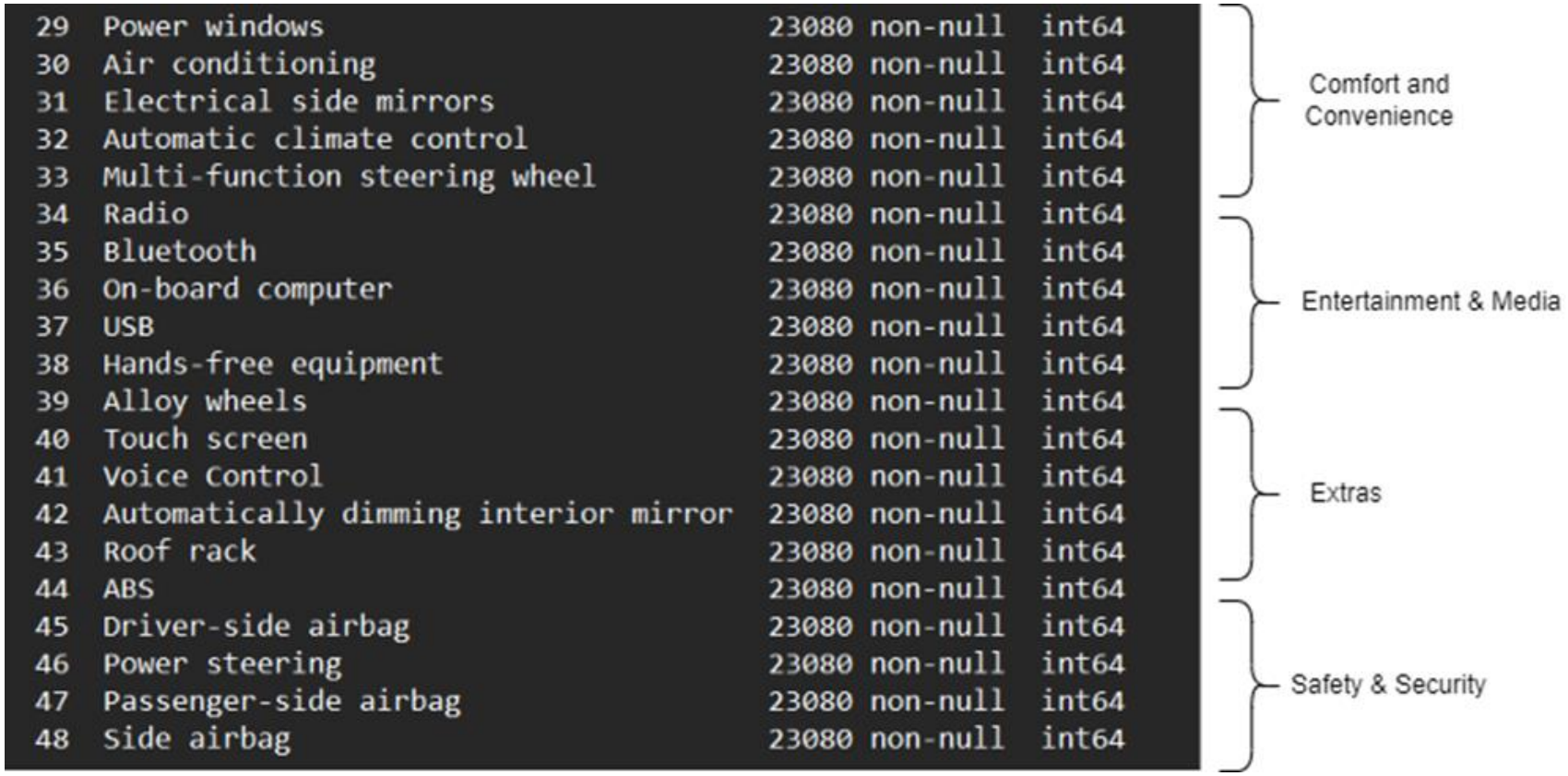
DATA REDUCTION

Dimensionality reduction:

- Deleted some records because they did not have values for attributes that we considered fundamental
- Deleted some records because they had too many missing values
- Dropped the following columns: Seller, Other fuel types, Type and Full Service History

Number of records at the beginning: 25786

Number of records after pre-processing: 23080



DATA TRANSFORMATION: BINARIZATION

Comfort and Convenience, Entertainment and media, Extras, Safety and Security: these attributes are a list of strings divided by ";" delimiter

- Assign to those attributes the length of the relative list
- Through a query we retrieve the 5 most common optionals for each category
- Binarization of them by adding new columns (One Hot Encoding)

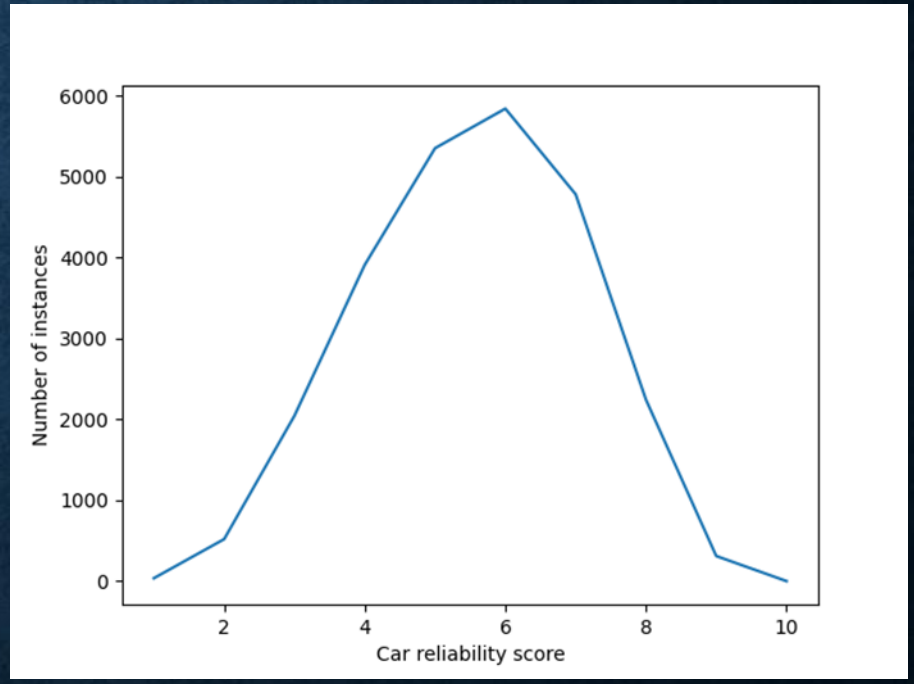
DATA TRANSFORMATION: NOMINAL TO NUMERIC

- **Power:** this feature was composed by two values, horsepower and power: we decided to keep only the first one.
- **Fuel type:** we had a lot of different values, so we parse these and traced them back to one of these 6 values: “electric”, “hybrid”, “lpg”, “methane”, “diesel”, “gasoline”; after that we transformed these in numeric values.
- **Emission class, Gearbox, Upholstery:** we have few different values, so we transform these in numeric values.
- **Drivetrain:** we keep only the value “4WD”, binarizing it, because other values not affecting the classification; we rename this attribute in 4WD.
- **Fuel consumption:** initially for this feature we had three values corresponding to the type of use of the car: city, country and combined. We have replaced the values (eliminating the units of measurement) with the average of the previous three and set it to zero in the case of electric cars.

DATA INTEGRATION

$$Reliability\ Score = \frac{BrandScore * SafeSecurity * Warranty}{Mileage * PreviousOwner * CarAge}$$

- **Brand score** is based on data taken from the extended car warranty data held by Warranty Wise (UK warranty provider) and the repair claims they have received, considering also car's age, repair cost, the time taken for it to be repaired, and the frequency of repairs.
- We applied to this feature a logarithmic transformation, because the distribution of values was skewed, and after this, we normalize it in a range from 1 to 10 with min-max normalization.



FINAL DATASET

BodyType	CO2Emissi...	Colour	ComfortCo...	Country	Cylinders	Doors	WD	EmissionCl...	EmptyWei...	EngineSize	Entertainm...	Extras	FirstRegist...	FuelConsu...	FuelType	Gearbox	Gears	Manufactu...	Mileage
Categorical	Number	Categorical	Number	Categorical	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Categorical	Number
Body type	CO2 Emissi...	Colour	Comfort & ...	Country	Cylinders	Doors	4WD	Emission cl...	Empty weig...	Engine size	Entertainm...	Extras	First registr...	Fuel consu...	Fuel type	Gearbox	Gears	Manufacturer	Mileage
Coupe	160.6	Black	1	ES	4	3	0	6	1530	1984	0	0	2021	7.0	1	3	7	Audi	9987
Coupe	301.3	Black	4	ES	8	5	1	6	1760	4163	2	2	2015	13.6	1	3	7	Audi	44730
Coupe	188.5	Grey	11	BE	6	3	1	6	1735	2995	5	5	2017	8.5	1	3	8	Audi	49500
Coupe	222.0	Black	25	DE	6	3	1	6	1770	2894	13	7	2021	9.2	1	3	8	Audi	17500
Off-Road/P...	246.8	White	3	ES	4	5	0	6	1460	1495	1	0	2020	5.9	1	1	6	Audi	49791
Coupe	145.0	Silver	13	DE	4	3	0	6	1530	1984	5	5	2021	6.6	1	3	7	Audi	13692
Coupe	144.0	White	18	DE	4	5	1	6	1520	1984	8	8	2020	6.6	1	3	7	Audi	33000
Convertible	163.0	Yellow	19	DE	4	5	1	6	1525	1984	12	13	2017	7.3	1	3	6	Audi	5610
Sedan	132.6	Grey	1	ES	4	5	0	6	1730	999	0	0	2021	6.3	1	1	5	Audi	25875
Coupe	149.0	Black	18	BE	4	3	0	6	1530	1984	8	5	2015	6.7	1	3	7	Audi	32966
Sedan	132.6	Blue	15	DE	4	5	0	6	1730	1498	7	3	2021	6.3	1	3	7	Audi	11755
Off-Road/P...	120.0	Green	15	DE	4	5	0	6	1365	1498	10	4	2021	5.4	1	3	7	Audi	9627
Off-Road/P...	186.0	Black	17	DE	4	5	1	6	1845	1968	5	8	2020	5.7	2	3	7	Audi	38600
Coupe	144.0	Violet	20	DE	4	3	0	6	1370	1984	7	6	2021	6.6	1	3	7	Audi	21000
Compact	137.0	Grey	19	DE	4	5	0	6	1730	1984	8	7	2021	6.3	1	3	7	Audi	8900
Coupe	146.0	Yellow	16	BE	4	3	0	6	1530	1984	7	5	2017	6.6	1	3	7	Audi	48625
Off-Road/P...	172.0	Grey	24	DE	4	5	1	6	1845	1984	8	9	2020	7.8	1	3	7	Audi	27600
Coupe	194.0	White	18	DE	5	3	1	6	1515	2480	7	5	2018	8.6	1	3	7	Audi	12500
Sedan	190.0	Green	20	BE	5	5	0	6	1605	2480	3	4	2021	2.8	1	3	7	Audi	11734
Off-Road/P...	122.0	Grey	15	DE	4	5	0	6	1420	1498	6	5	2020	5.5	1	3	7	Audi	24500
Sedan	119.0	Black	19	BE	4	5	0	6	1520	1968	7	3	2021	4.6	2	3	7	Audi	11736
Off-Road/P...	123.0	Black	21	DE	4	5	0	6	1622	1968	14	10	2021	5.6	2	3	7	Audi	24840
Coupe	150.0	White	17	DE	4	3	0	6	1325	1984	11	6	2019	6.9	1	1	6	Audi	22266
Off-Road/P...	205.0	Black	28	DE	8	5	1	6	2345	3956	11	14	2020	7.9	2	3	7	Audi	43900
Sedan	117.0	Black	14	DE	4	5	0	6	1335	1498	7	7	2020	5.3	1	3	7	Audi	22600

CORRELATION MATRIX

	Price
Power	0.679775
Empty weight	0.524184
Gears	0.520177
Car reliability score	0.478688
Gearbox	0.455307
First registration	0.449459
Comfort & Convenience	0.423377
Engine size	0.402226
Safety & Security	0.351583
Emission class	0.346662
Upholstery	0.346175
Entertainment & Media	0.339812
4WD	0.338036
Cylinders	0.331068
Extras	0.321377
Voice Control	0.293532
Automatic climate control	0.280594
Hands-free equipment	0.262498
Automatically dimming interior mirror	0.236797
Multi-function steering wheel	0.236592
Bluetooth	0.233093
USB	0.230613
Touch screen	0.195156
Fuel type	0.157799
Roof rack	0.141700
On-board computer	0.107170
Alloy wheels	0.100292
CO2 Emissions	0.089333
Warranty	0.075698
Electrical side mirrors	0.074771
Fuel consumption	0.063368
Radio	0.061440
Side airbag	0.057916
Seats	0.045735
Passenger-side airbag	0.032301
Previous owner	0.014895
ABS	0.003918
Power steering	0.000850
Doors	-0.006415
Power windows	-0.024004
Driver-side airbag	-0.026870
Air conditioning	-0.042171
Mileage	-0.438882

CLASSIFICATION STEPS

Train and Test splitting:

10-FOLD CROSS VALIDATION

Attribute selection:

CFS SUBSET EVAL + BEST FIRST

CFS SUBSET EVAL + GREEDY STEPWISE

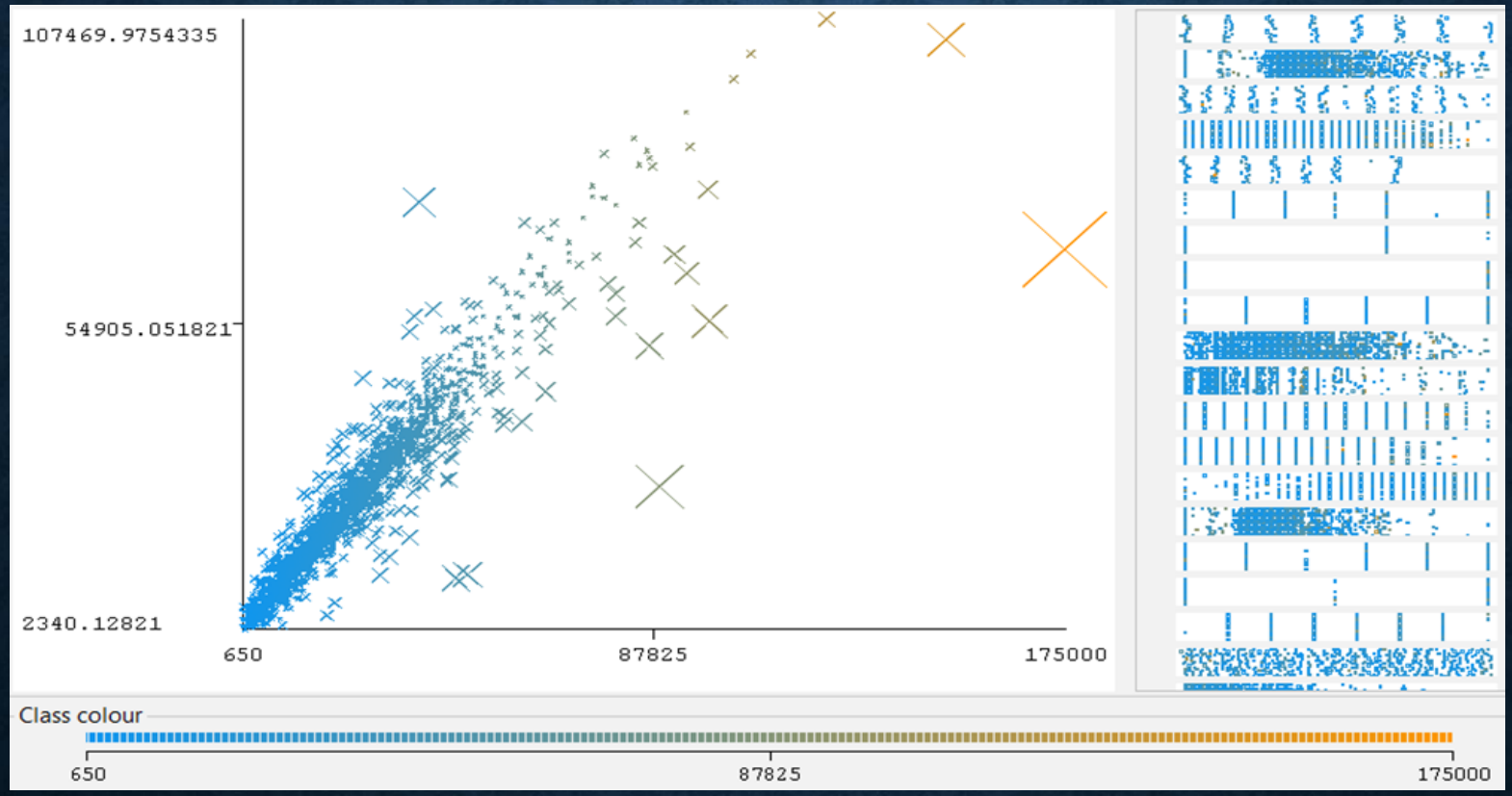
Regression classification algorithms:

- LINEAR REGRESSION
- 5-NN
- M5 RULES
- RANDOM FOREST
- REP TREE

CLASSIFICATION RESULTS

	NO FEATURE SELECTION	CFS SUBSET EVAL + GREEDY STEPWISE	CFS SUBSET EVAL + BEST FIRST
LINEAR REGRESSION	OUT OF MEMORY (EVEN WITH 50 % OF SAMPLES)	CC = 0.8524 MAE = 4819.6601 RMSE = 8129.42899 RAE = 43.641 % RRSE = 52.2844 %	CC = 0.8503 MAE = 4854.37 RMSE = 8183.0349 RAE = 43.9553 % RRSE = 52.6291 %
5-NN	CC = 0.844 MAE = 4731.6987 RMSE = 8367.3542 RAE = 42.8446 % RRSE = 53.8146 %	CC = 0.8404 MAE = 4705.6345 RMSE = 8435.5513 RAE = 42.6086 % RRSE = 54.2532 %	CC = 0.8399 MAE = 4700.002 RMSE = 8449.9122 RAE = 42.5576 % RRSE = 54.3456 %
M5-RULES	OUT OF MEMORY (EVEN WITH 50 % OF SAMPLES)	CC = 0.9068 MAE = 3563.7948 RMSE = 6556.9792 RAE = 32.2694 % RRSE = 42.1712 %	CC = 0.9055 MAE = 3594.5742 RMSE = 6600.8324 RAE = 32.5481 % RRSE = 42.4532 %
REP TREE	CC = 0.8902 MAE = 3820.3797 RMSE = 7105.9627 RAE = 34.5928 % RRSE = 45.702 %	CC = 0.8967 MAE = 3702.4068 RMSE = 6907.7698 RAE = 33.5245 % RRSE = 44.4273 %	CC = 0.8942 MAE = 3741.6653 RMSE = 6982.81 RAE = 33.88 % RRSE = 44.9099 %
RANDOM FOREST	DONE WITH 50 % OF SAMPLES : CC = 0.955 MAE = 2306.9286 RMSE = 4763.2676 RAE = 21.2634 % RRSE = 31.3194 %	**CC = 0.9383** **MAE = 2727.4627** **RMSE = 5456.9642** **RAE = 24.6966 %** **RRSE = 35.0964 %**	CC = 0.9368 MAE = 2759.6865 RMSE = 5518.5499 RAE = 24.9884 % RRSE = 35.4925 %

RANDOM FOREST CLASSIFICATION ERRORS



ATTRIBUTE SELECTED CLASSIFIER FOR EXPERIMENTS:

CFS SUBSET EVAL + GREEDY STEPWISE AND RANDOM FOREST

DATASET WITH OPTIONAL BINARIZATION	DATASET WITHOUT OPTIONAL BINARIZATION
Correlation coefficient = 0.9383	Correlation coefficient = 0.9291
Mean absolute error = 2727.4627	Mean absolute error = 3144.1751
Root mean squared error = 5456.9642	Root mean squared error = 5748.0347
Relative absolute error = 24.6966 %	Relative absolute error = 27.3678 %
Root relative squared error = 35.0964 %	Root relative squared error = 39.315 %

EXPERIMENTS: OPTIONAL BINARIZATION

DATASET WITH CAR REL. SCORE	DATASET WITHOUT CAR REL. SCORE
Correlation coefficient = 0.9383	Correlation coefficient = 0.9353
Mean absolute error = 2727.4627	Mean absolute error = 2827.8752
Root mean squared error = 5456.9642	Root mean squared error = 5582.8502
Relative absolute error = 24.6966 %	Relative absolute error = 25.6058 %
Root relative squared error = 35.0964 %	Root relative squared error = 35.9061 %

EXPERIMENTS: RELIABILITY SCORE

EXPERIMENTS: VINTAGE CARS

TRAINING SET: *FINAL DATASET* (50 COL)

TEST SET : *VINTAGE CARS DATASET* (≈3000 ROWS)

RESULTS:

- Correlation coefficient = 0.7266
- Mean absolute error = 9252.6056
- Root mean squared error = 16585.3462
- Relative absolute error = 52.6837 %
- Root relative squared error = 78.7437 %

NOT SUITABLE

EXPERIMENTS: SUPERCARS

TRAINING SET: *FINAL DATASET* (50 COL)

TEST SET : *SUPERCARS DATASET* (≈3000 ROWS)

RESULTS:

- Correlation coefficient = 0.5411
- Mean absolute error = 97902.0542
- Root mean squared error = 141641.0652
- Relative absolute error = 86.0528 %
- Root relative squared error = 89.5059 %

NOT SUITABLE

DATASET WITH SPANISH RECORDS	DATASET WITHOUT SPANISH RECORDS
Correlation coefficient = 0.9383	Correlation coefficient = 0.9368
Mean absolute error = 2727.4627	Mean absolute error = 2885.7304
Root mean squared error = 5456.9642	Root mean squared error = 5710.5958
Relative absolute error = 24.6966 %	Relative absolute error = 25.3488 %
Root relative squared error = 35.0964 %	Root relative squared error = 35.7756 %

EXPERIMENTS: SPANISH CARS

DATASET WITH 50 ATTRIBUTES	DATASET WITH 114 ATTRIBUTES
Correlation coefficient = 0.9383	Correlation coefficient = 0.925
Mean absolute error = 2727.4627	Mean absolute error = 3115.2237
Root mean squared error = 5456.9642	Root mean squared error = 6069.3776
Relative absolute error = 24.6966 %	Relative absolute error = 28.2077 %
Root relative squared error = 35.0964 %	Root relative squared error = 39.0352 %

EXPERIMENTS: ALL NUMERIC COLUMNS

FINAL DATASET	FINAL DATASET WITH 0-1 NORMALIZATION
Correlation coefficient = 0.9383	Correlation coefficient = 0.9079
Mean absolute error = 2727.4627	Mean absolute error = 0.0186
Root mean squared error = 5456.9642	Root mean squared error = 0.0321
Relative absolute error = 24.6966 %	Relative absolute error = 34.2094 %
Root relative squared error = 35.0964 %	Root relative squared error = 41.9479 %

FINAL DATASET 5-NN	**FINAL DATASET 5-NN 0-1 NORMALIZATION**
Correlation coefficient = 0.8404	**Correlation coefficient = 0.8699**
Mean absolute error = 4705.6345	**Mean absolute error = 0.0221**
Root mean squared error = 8435.5513	**Root mean squared error = 0.0377**
Relative absolute error = 42.6086 %	**Relative absolute error = 40.6411 %**
Root relative squared error = 54.2532 %	**Root relative squared error = 49.3924 %**

EXPERIMENTS: 0-1 NORMALIZED COLUMNS

T-TEST

	RANDOM FOREST	LINEAR REGRESSION	5-NN	M5 RULES	REP TREE
CORRELATION COEFF.	0.94	0.87 *	0.93 *	0.94 *	0.94 *
MAE	0.37	0.59 v	0.42 v	0.40 v	0.40 v
RMSE	0.50	0.75 v	0.56 v	0.51 v	0.53 v
ELAPSED TIME TRAINING	9.90	0.53 *	0.69 *	10.73	1.43 *
ELAPSED TIME TESTING	0.23	0.00 *	1.61 v	0.04 *	0.00 *

FINAL CONSIDERATIONS

From this project we have understood that is possible to build an accurate model with the aim of determining the price of a used car. This task could be very useful in a real context because it can help a user to decide at what price to sell his or her car, or whether a car is being sold at a fair price.

- It is not possible to classify every car type using the same classifier
- Keep the data fresh with further scrapings in order to maintain reliable results, that reflects the car market price variation.
- Another possible integration could be the introduction of a car's accident history, so that a buyer could have a complete background of the car they would like to buy.