Defining the best second hand car market price
&
Identifying the best second hand car

(https://github.com/corrado-github/as24_project)

Requirements

A customer want to sell his old car and buy a second hand electric car.

He want to sell a Fiat Punto Evo car

He want to buy a second hand BMW i3 electric car

Questions to answer:

1) **What's a fair price of a specific car on the second hand car market?**
2) He want to buy a second hand BMW i3 car, which will be sold in one year. **Can we identify which cars lose their value on the market less with increasing mileage?**

Fiat Punto Evo

The car he wants to sell is a Fiat Punto Evo

- Purchase year: 2010
- Mileage (km): 162 000
- Power: 57 kW
- Fuel type: Petrol/LPG

We need to collect data for many similar autos and build a machine learning model to predict the price

Data gathering

We collected data from the web site autoscout24.it and set a Python *data scraper* by using the packeges Selenium and BeautifulSoap.

We collected data for:

- 400 second hand cars Fiat Punto Evo
- 246 second hand cars BMW i3

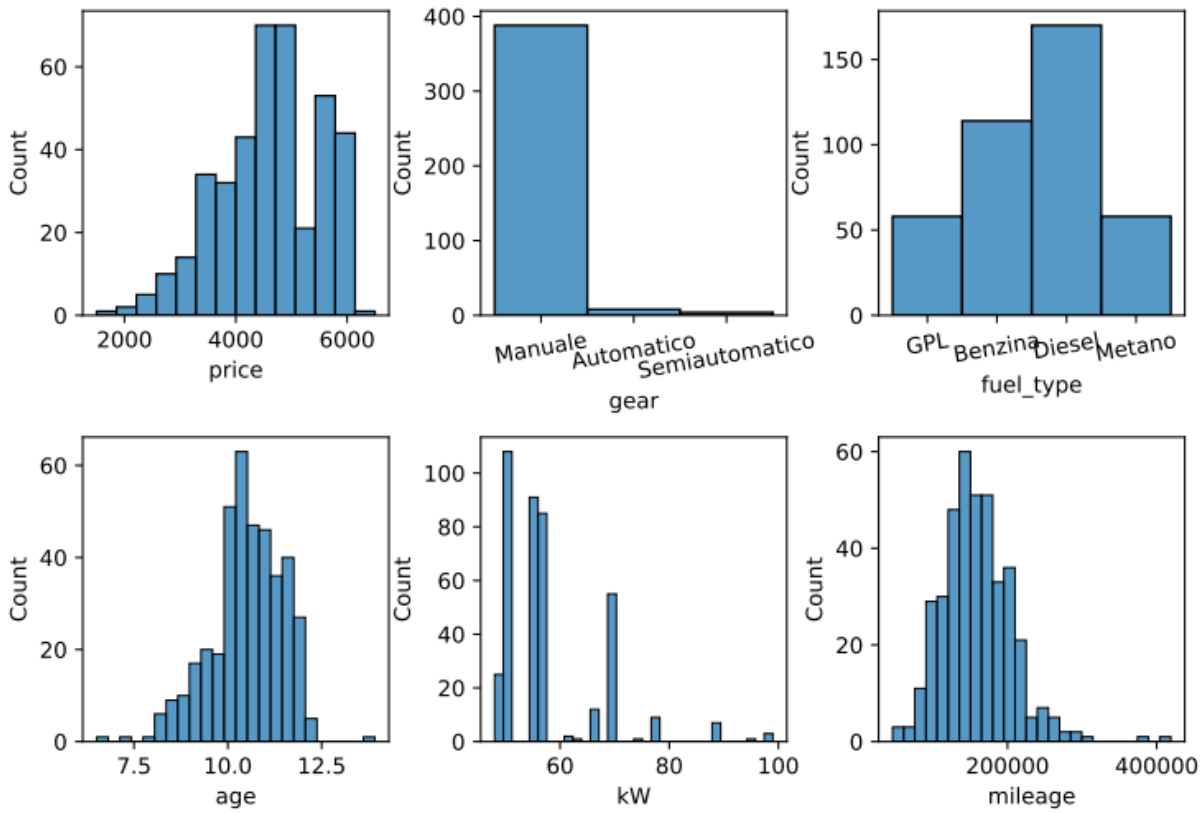
	model	version	equipments	mileage	mmyy	power	use_type	n_owners	gear	fuel_type	price
0	Fiat Punto Evo	1.4 5 porte Dynamic EasyPower	['Sedile posteriore sdoppiato, Fendinebbia']	218000	09/2011	57 kW (77 CV)	Usato	1.0	Manuale	GPL	3390
1	Fiat Punto Evo	1.2 3 porte Active	['Specchietti laterali elettrici, Antifurto']	176539	11/2011	48 kW (65 CV)	Usato	NaN	Manuale	Benzina	3900
2	Fiat Punto Evo	Punto Evo 1.2 Dynamic s	['Volante in pelle, Computer di bordo, Chiusur...	159000	01/2012	51 kW (69 CV)	Usato	NaN	Manuale	Benzina	4500
3	Fiat Punto Evo	1.6 Mjt DPF 3 porte Sport	['Fendinebbia, Pacchetto sportivo, Chiusura ce...	197000	04/2010	88 kW (120 CV)	Usato	NaN	Manuale	Diesel	4900
4	Fiat Punto Evo	1.2 3 porte S&S Dynamic	['Cerchi in lega, Climatizzatore, Autoradio, C...	59900	06/2012	51 kW (69 CV)	Usato	NaN	Manuale	Benzina	5900

Fiat Punto Evo

Workflow:

1) Data checking, wrangling, cleaning
2) Data visualizazion, correlations
3) Machine Learning model setting
4) Training and perfomances test of the model
5) Market price prediction

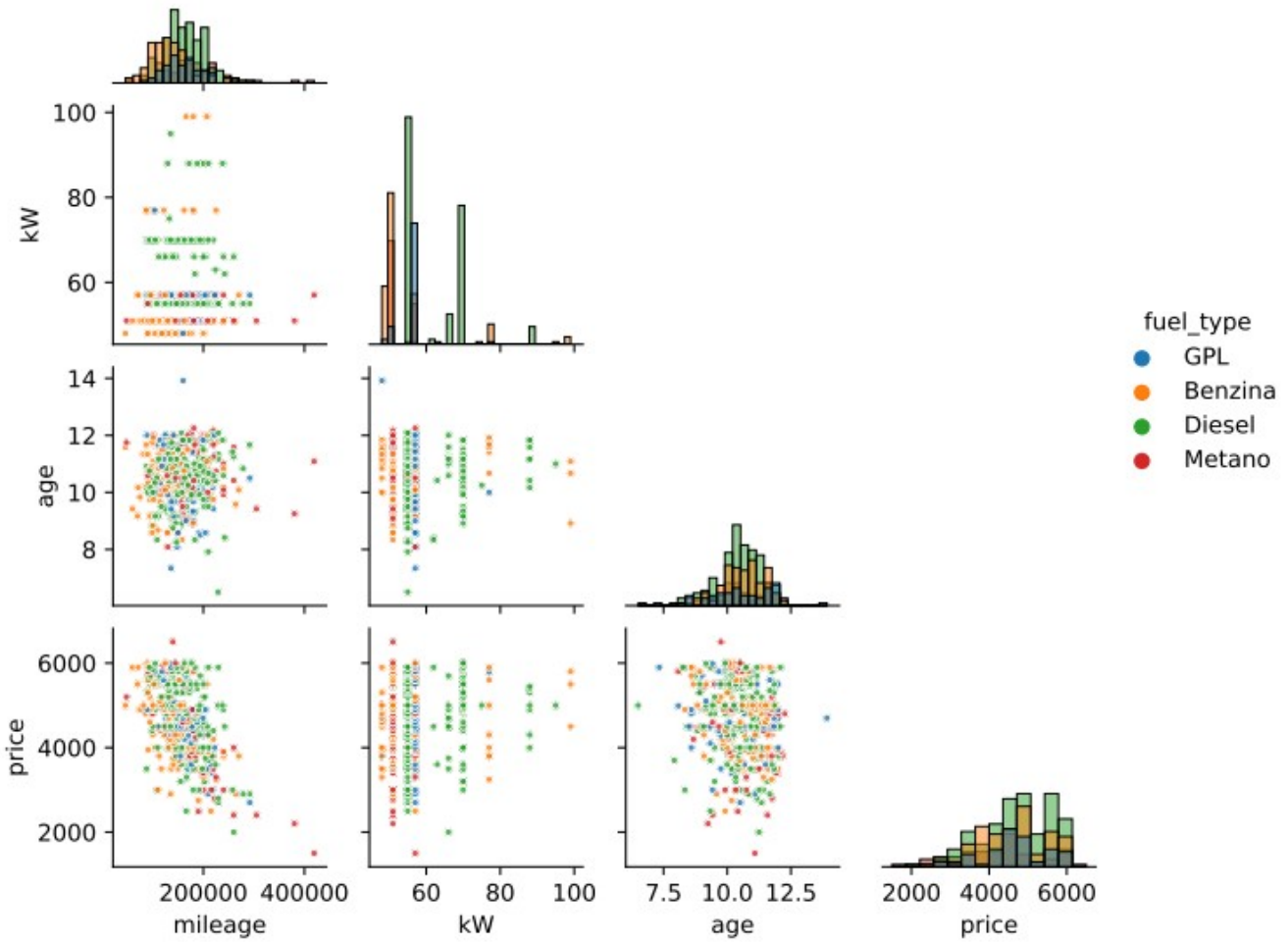
Data Visualization



Distributions of the main vaiables

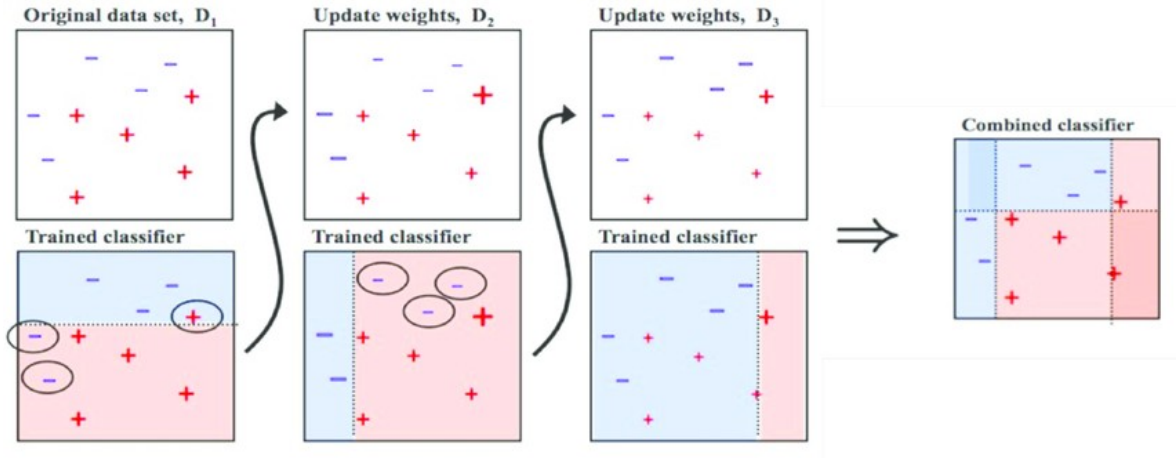
Data Visualization

Scatter plot of the main variables



Machine Learning Model

Machine Learning Model: *XGBoost*



XGBoost regressor: This algorithm builts a series of *decision tree* which following tree corrects the errors of the previous tree. Then, all the trees are summed up to get the final result.

Machine Learning Model

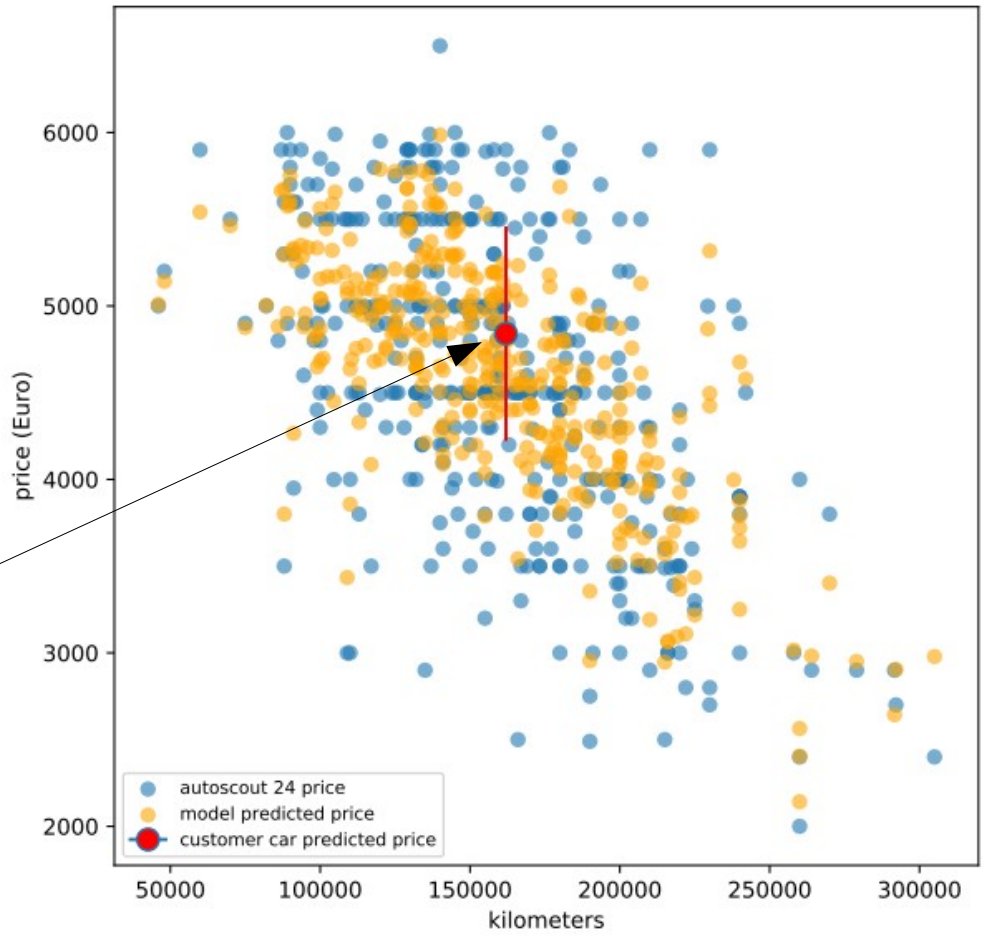
After the training, we use the XGBoost model to predict the price of the customer's car

Result:

Esteemed market price

Fiat Punto Evo:

4840 ± 618 Euro



Identification of the best second hand BMW i3 car

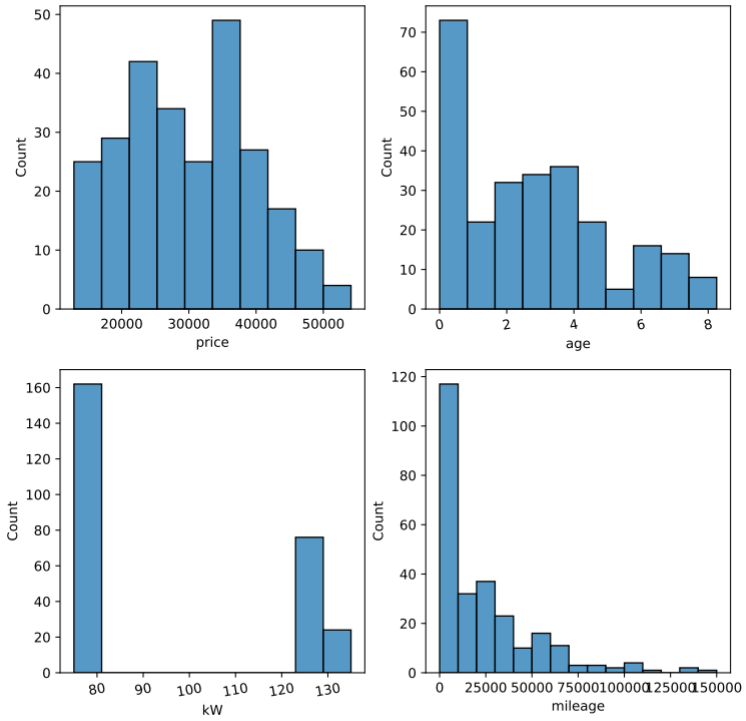
BMW i3

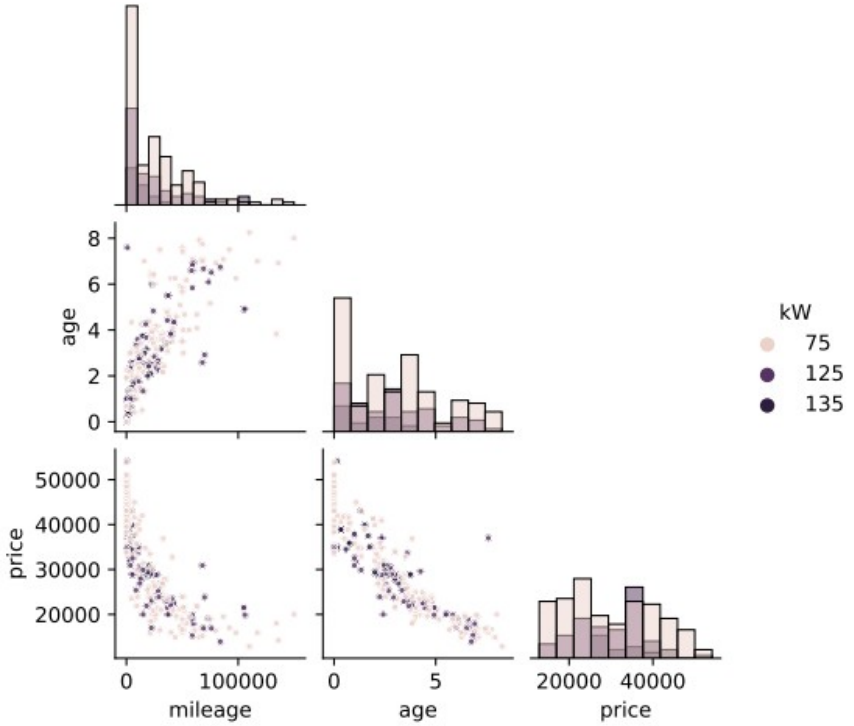
We want to find out which BMW i3 in the second hand market satisfy the following conditions best:

1) Smallest mileage
2) Smallest price loss with mileage

BMW: Data visualization

After the data wrangling and cleaning, we visualize the main variables distributions



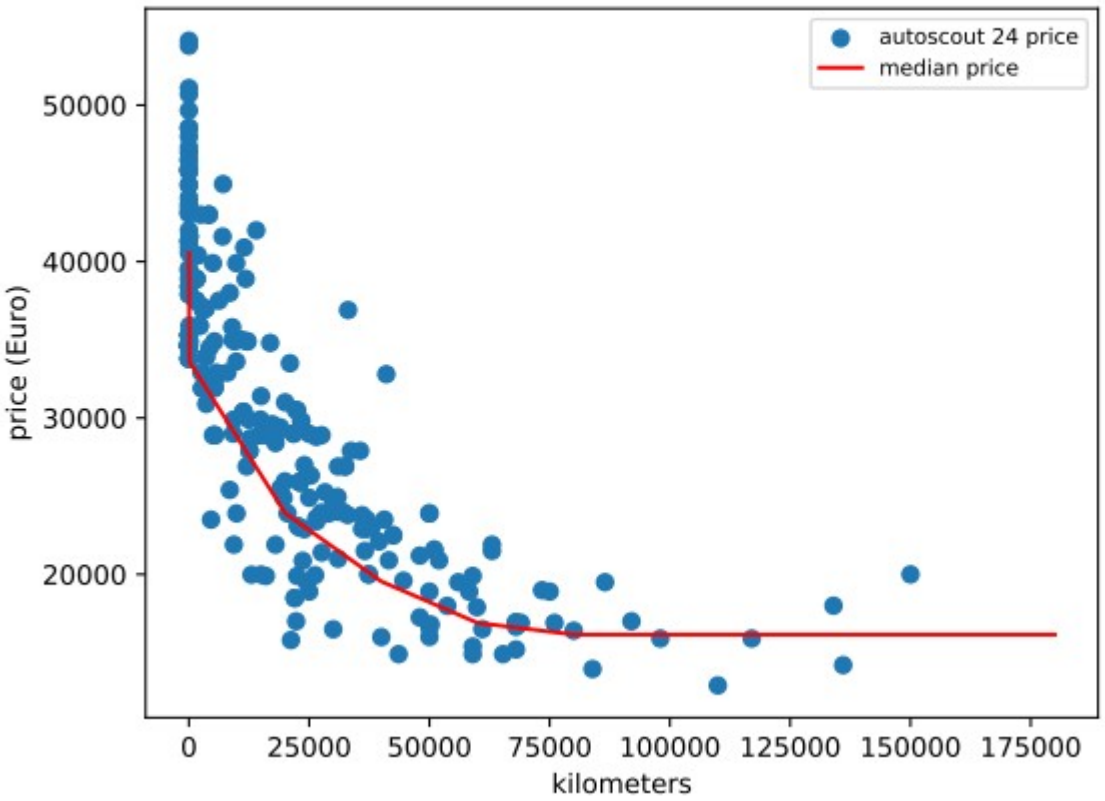


BMW: Data visualization

- We notice that, with increasing mileage:

1) The average price diminishes
2) The price variation diminishes (the first derivative approaces zero)

- The price quickly drop in the first 20’000km and become nearly constant after 60’000km.
- The given conditions are satisfied for those cars under the red “knee” in the figure.



BMW: ranking the cars

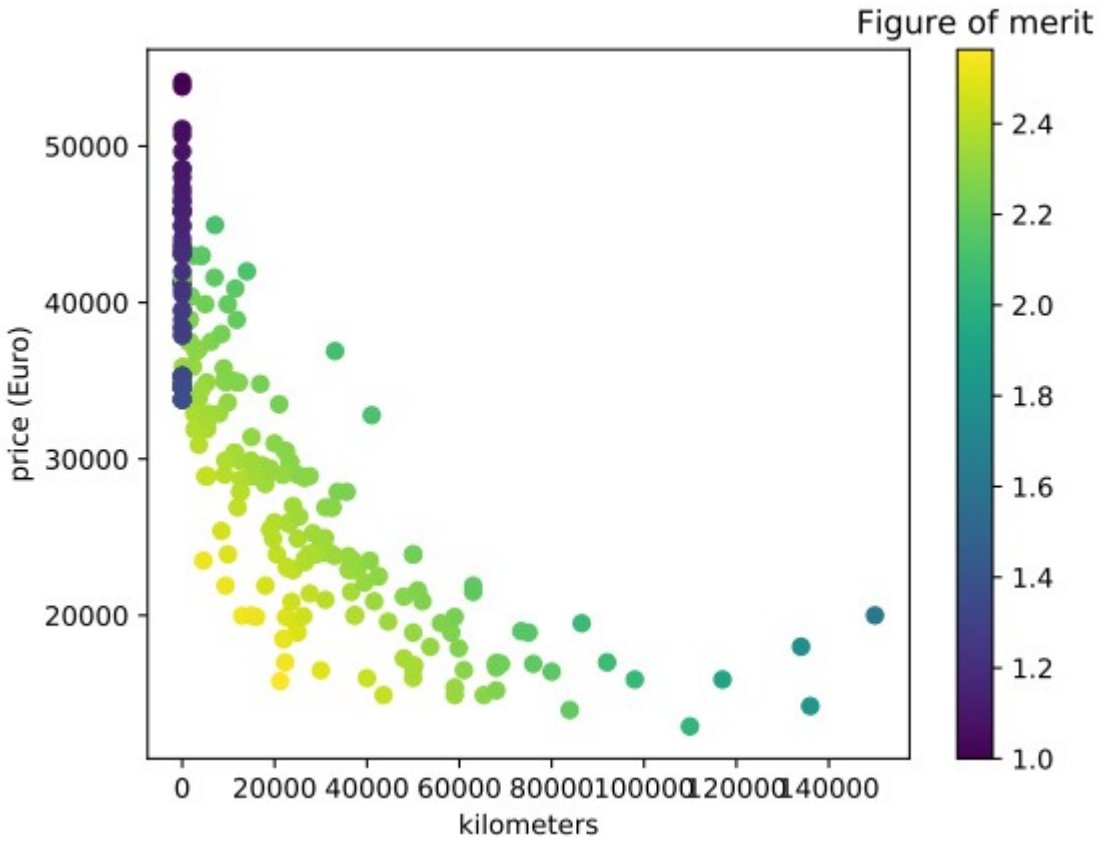
We define a linear function (*figure of merit*) that rank higher the cars having lower price, mileage, and price loss as follow:

$$figure\ of\ merit = \left(1 - \frac{prezzo}{max(prezzo)}\right) + \left(1 - \frac{km}{max(km)}\right) + \left(1 - |(1\,D)| / max(|(1\,D)|)\right)$$

Km: mileage in kilometers
1D: price first derivative
(see Jupyter notebook on GitHub for details)

BMW: ranking the cars

The cars with higher figure of merit (yellow points on the low left corner for the figure) are the cars that satisfied the requirements best



Risultati

Ordinando le auto con *figure of merit* discendente, ecco la lista delle 5 auto che meglio rispondono al secondo quesito

	model	version	mileage	power	mmyy	price	seller_company_name	seller_city	seller_CAP	fig_of_merit
3	BMW i3	i3 *KM 21.000	21215.0	75 kW (102 CV)	2015-06-01	15800	Bc Car srl	Mesagne BR	72023	2.563471
8	BMW i3	94Ah	13010.0	75 kW (102 CV)	2017-08-01	19980	Okay Cars Srl	Quattro Castella - Reggio Emilia - Re	42020	2.537077
39	BMW i3	FULL ELECTRIC 170CV IVA 22% DEDUCIBILE COMPRESA	22300.0	125 kW (170 CV)	2015-11-01	17000	Autoecommerce Srl	San Giovanni in Persiceto - Bologna - Bo	40017	2.534065
86	BMW i3	94 Ah	4600.0	75 kW (102 CV)	2017-07-01	23500	Extramotors - Gruppo Maccianti srl	Follonica - Grosseto - Gr	58022	2.528103
12	BMW i3	i3	9350.0	75 kW (102 CV)	2018-11-01	21900	Tuacar srl	Moncalieri -To	10024	2.526000