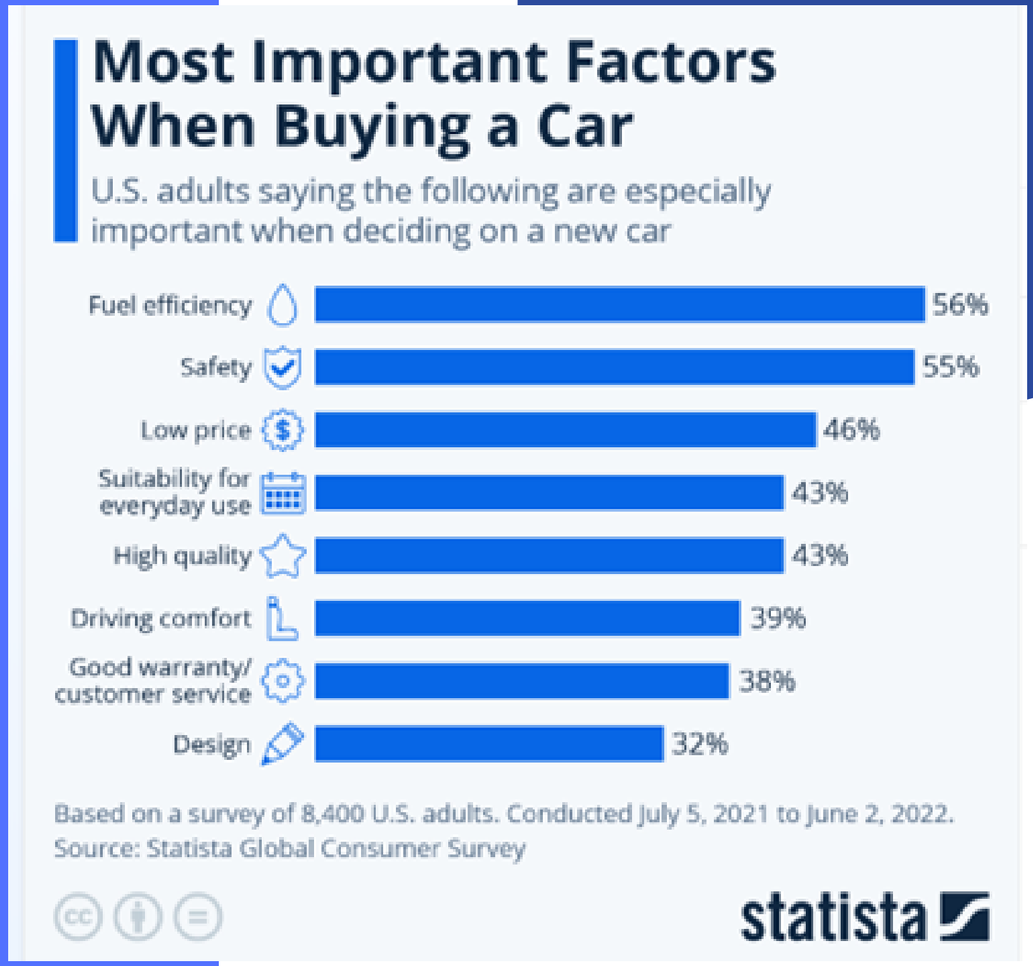
EXPLORING THE LINK BETWEEN SAFETY RATINGS AND CAR PRICE

MARTA BRASOLA

INTRODUCTION

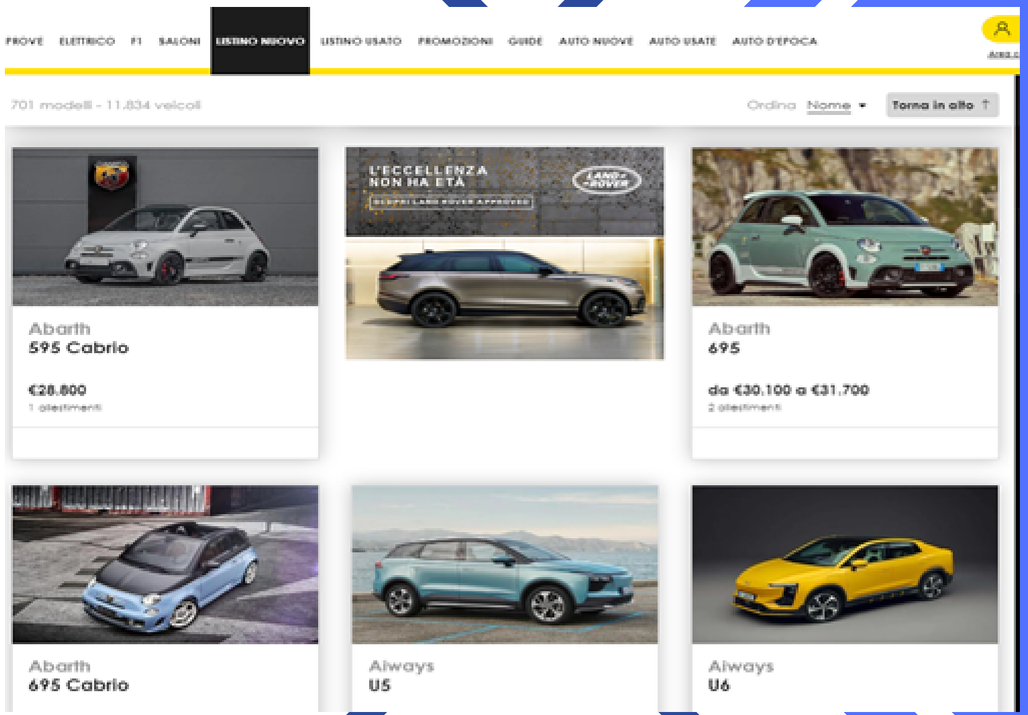


CONTENT INDEX

DATA ACQUISITION

AUTOMOTO.IT

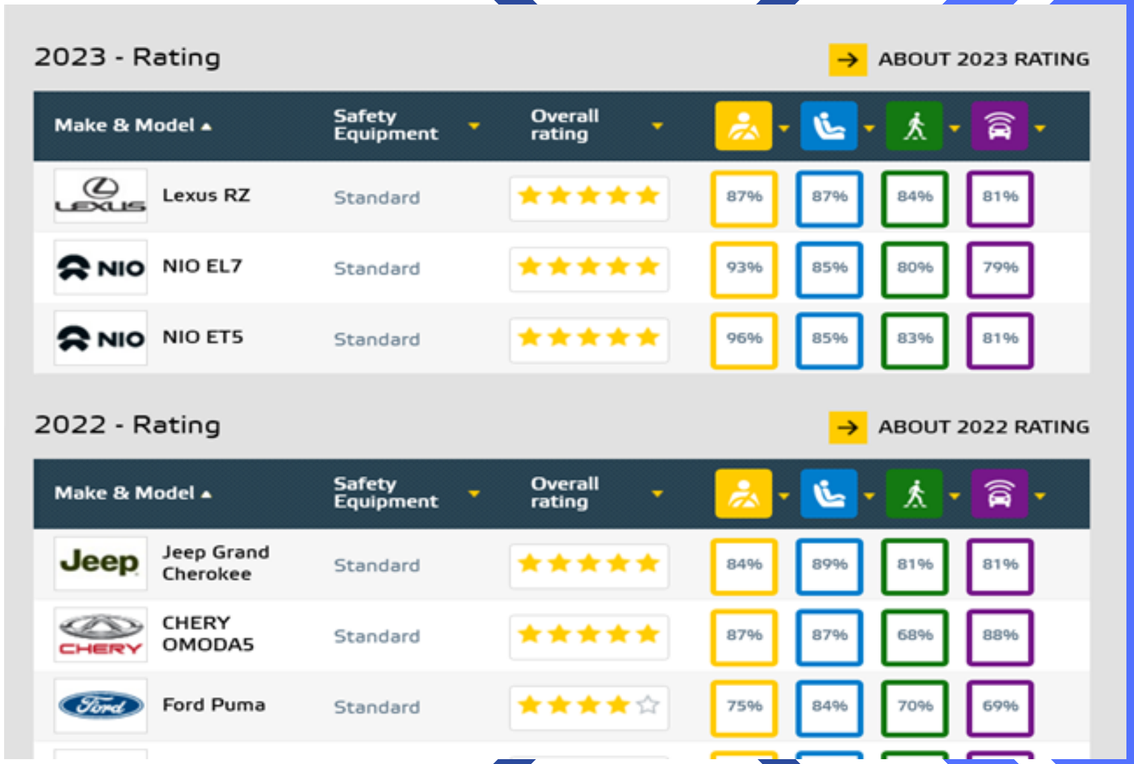
- scraping the price of new cars
- picking the price of the base model for every car setup
- automoto.it had more models than other websites



DATA ACQUISITION

EURO NCAP

- Euro NCAP follows European guide lines in terms of safety tests
- the rating system is summed up in stars



2023 - Rating

ABOUT 2023 RATING

Make & Model	Safety Equipment	Overall rating				
Lexus RZ	Standard	★★★★★	87%	87%	84%	81%
NIO EL7	Standard	★★★★★	93%	85%	80%	79%
NIO ET5	Standard	★★★★★	96%	85%	83%	81%

2022 - Rating

ABOUT 2022 RATING

Make & Model	Safety Equipment	Overall rating				
Jeep Grand Cherokee	Standard	★★★★★	84%	89%	81%	81%
CHERY OMODA5	Standard	★★★★★	87%	87%	68%	88%
Ford Puma	Standard	★★★★☆	75%	84%	70%	69%

DATA INTEGRATION

DATA CLEANING

DATA MERGING

DATA CLEANING

AUTOMOTO.IT

- splitting the column “price_range” into the minimum and maximum price
- splitting “brand_model”

	brand_model	price_range	setups
0	Abarth 500e	da €37.950\n a €43.000	3 allestimenti
1	Abarth 500e Cabrio	da €37.950\n a €43.000	3 allestimenti
2	Abarth 595	€26.800	1 allestimenti
3	Abarth 595 Cabrio	€28.800	1 allestimenti
4	Abarth 695	da €30.100\n a €31.700	2 allestimenti

DATA CLEANING

EURO NCAP DUPLICATES MODELS

- same model but with different rating years, in this case I kept the most recent one
- same model but with different safety pack equipement; in this case I kept the one with the standard safety equipement

[ ]:

	brand_model	safety_equipement	rating_year
201	mazda 6	Standard	2018
408	mazda 6	no_info	2013
580	mazda 6	no_info	2009

DATA MERGING

- merging the datasets based on the brand and model columns
- the result of the merge is a dataset with 169 non null values

DATA STORAGE

MONGODB

- mainly for the flexibility advantages in case I would like to enrich the data with more informations

DATA STORAGE

QUERIES

```
[
  {
    $group: {
      _id: "$brand",
      average_min_price: {
        $avg: "$min_price",
      },
    },
  },
  {
    $project: {
      _id: 0,
      brand: "$_id",
      average_min_price: 1,
    },
  },
```


Stage 1 $project
/**
* specifications: The fields to
* include or exclude.
*/
{
"_id": 0,
"brand_model": 1,
"child_occupant_safety": 1
}
Output after $project stage (Sample of 10 documents)
brand_model: "aiways u5"
child_occupant_safety: 70


Stage 2 $sort
/**
* Provide any number of field/order pair
*/
{
"child_occupant_safety": -1
}
Output after $sort stage (Sample of 10 documents)
brand_model: "mazda cx-60"
child_occupant_safety: 91

DATA QUALITY

COMPLETENESS

- my final dataset contains 27% of the initial data available

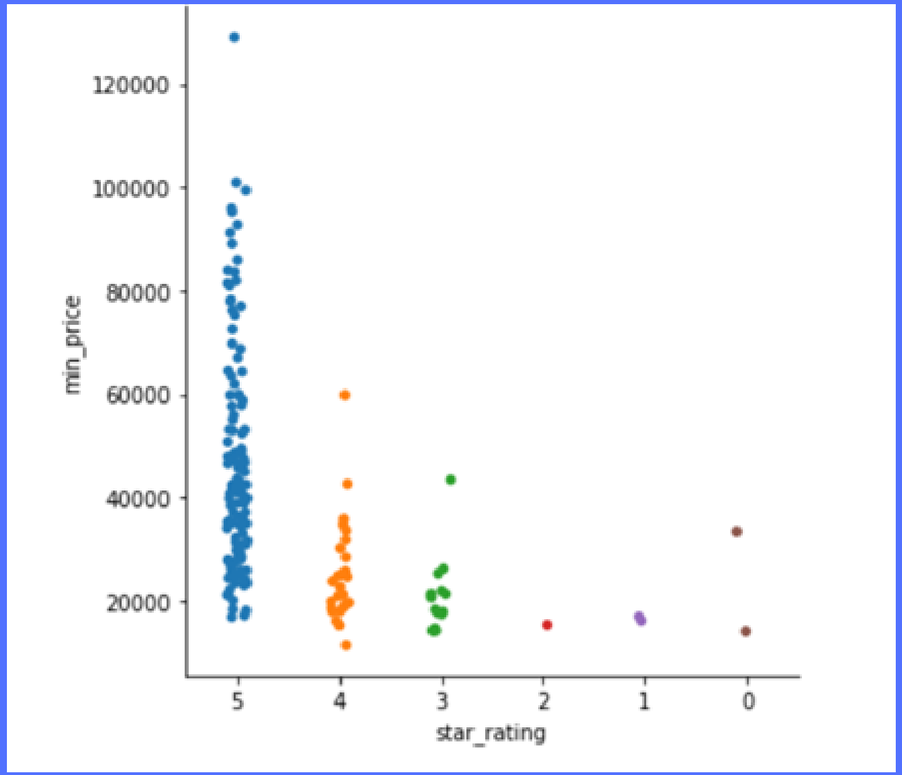
CONSISTENCY

- the purpose was to understand why only 27% of values found a match

reason	
not_present	0.767442
version	0.162791
simbols	0.069767

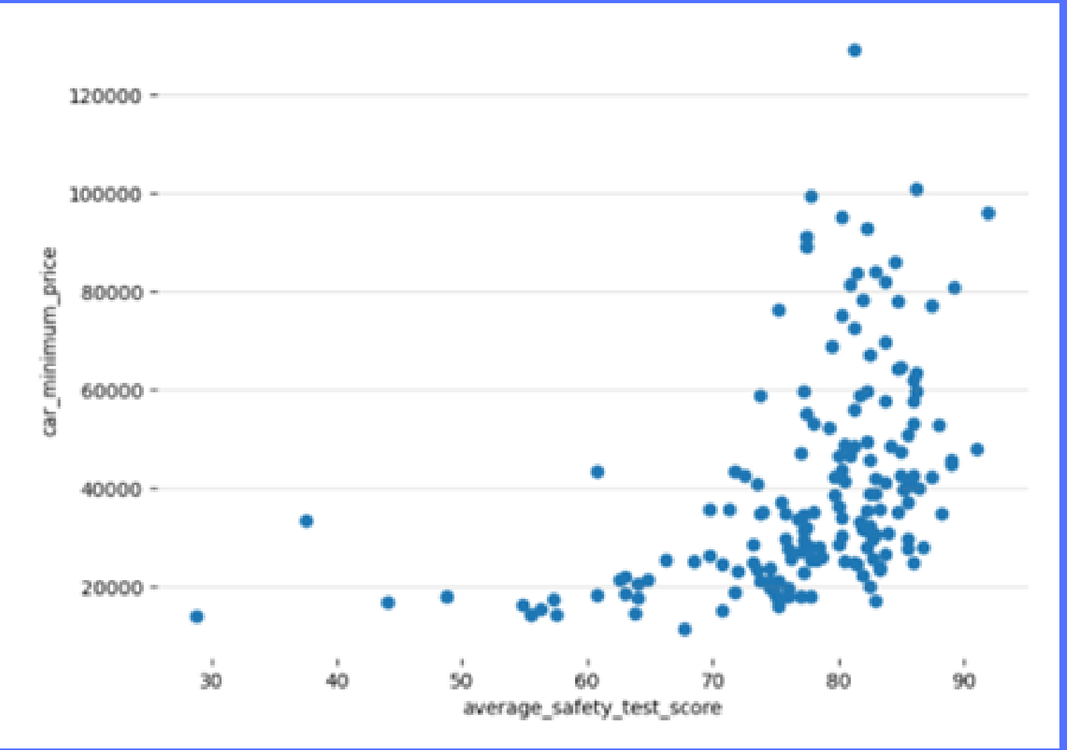
EDA

relationship between price and the star rating



EDA

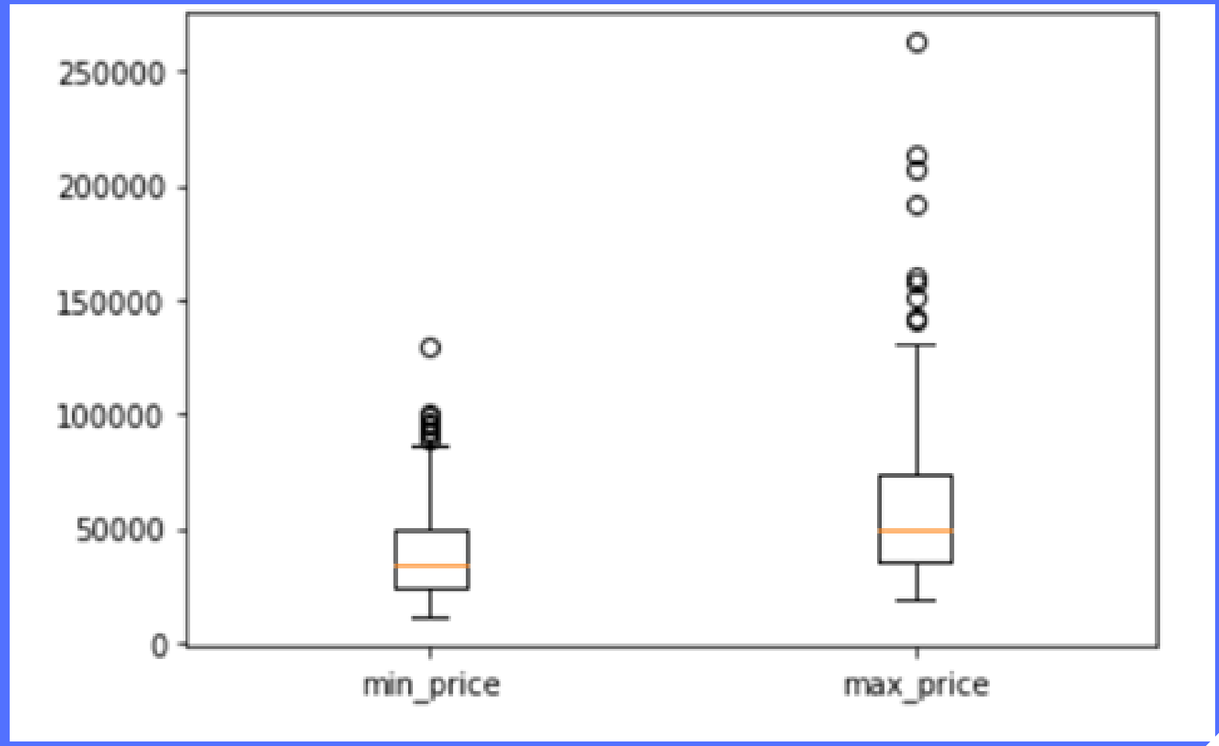
relationship between price and average safety score



	min_price	average_safety
min_price	1.000000	0.445513
average_safety	0.445513	1.000000

EDA

minimum and maximum price distributions



COCLUSIONS AND FUTURE DEVELOPMENTS

- the relationship between the price of a car and its safety tests is weak
- adding more data to explore better this relationship