CAPSTONE PROJECT

CARSALES.COM.AU



1. PROBLEM STATEMENT

	VW Golf	Opel Astra	Ford Focus	Toyota Corolla
Price (1000 \$)	16.2	14.9	14.0	15.2
Fuel Efficiency (per 100 km)	7.2	7.0	7.5	8.2
Power (kW)	66.0	62.0	55	71

Which one to buy if I want powerful, fuel efficient and cheap car?



Method - Multi-Objective Optimization

2. DATA SCRAPING



```
section="Showcase">
▼ <div class="listing-header n_pad-left-15 n_pad-right-20">
  ▶ <div class="considered n_columns">...</div>
  ▼ <div class="n_columns">
    ▼ <div class="n_width-max title ">
      ▼<a href="/dealer/details/Subaru-Outback-2012/OAG-AD-
      13666093/?gts=OAG-AD-13666093&gtssaleid=OAG-AD-
     13666093&rankingType=TopSpot" data-csn-item-history=
      "OAG-AD-13666093" data-webm-clickvalue="sv-view-
      title">
        ▼ <h2>
                                    2012 Subaru Outback 3.6R
           Premium 4GEN Auto AWD
           <span class="has-tooltip" csn-tooltip="+</pre>
           span.tooltip-content" data-tooltip-placement=
           "right" original-title>
                                        MY12
```

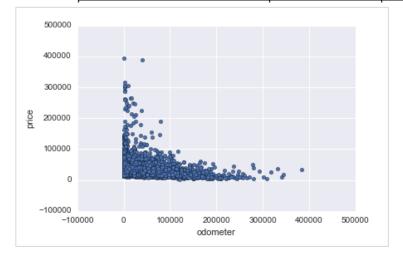
- After cleaning data and removing duplicates: 2905 cars
- Binning by fuel efficiency
- so we can assume that the data set is balanced.

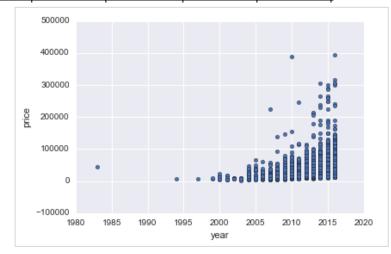
- 4 L/100km or less
- 5 L/100km or less
- 7 L/100km or less
- 9 L/100km or less
- 11L/100km or less
- 11L/100km or more

3. DATA WRANGLING

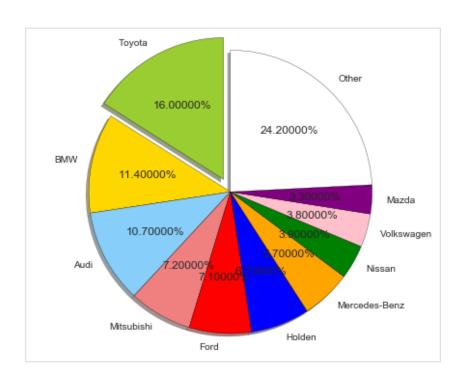
Linear correlation coefficients:

	fuel_efficiency	price	year	odometer	Number of cylinders	capacity
fuel_efficiency	1.900000	-0.005104	-0.382246	0.359213	0.619098	0.712716
price	-0.005104	1.000089	0.349844	-0.356435	0.263397	0.205220
year	-0.382246	0.349844	1 000006	-0.781987	-0.282250	-0.278190
odometer	0.359213	-0.356435	-0.781987	1.000000	9 226298	0.261757
Number of cylinders	0.619098	0.263397	-0.282250	0.226298	1.890000	6.901489
capacity	0.712716	0.205220	-0.278190	0.261757	0.901489	1.089000



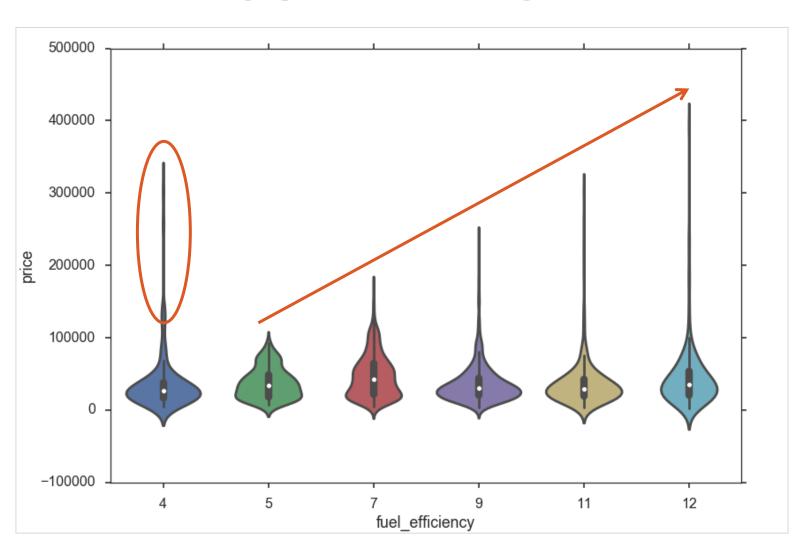


4.DATA VISUALIZATION

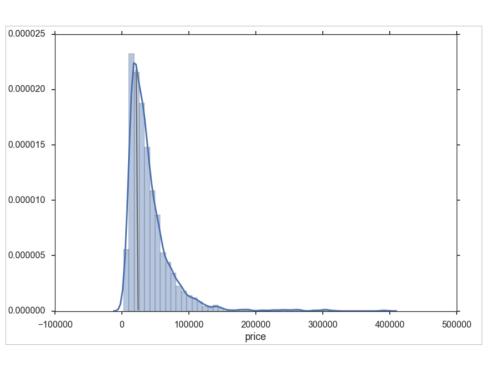


Distribution by car make

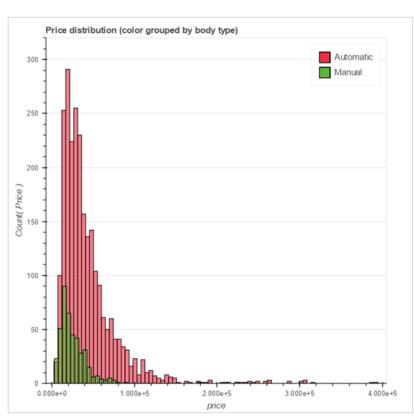
4.DATA VISUALIZATION



4.DATA VISUALIZATION



Price distribution for whole data set



Price distribution by transmission type