

Data Wrangling - Cleansing

September 19, 2022

```
[1]: import numpy as np
import pandas as pd
```

```
[2]: auto_price = pd.read_csv('https://raw.githubusercontent.com/ammishra08/
↳MachineLearning/master/Datasets/Automobile_price_data__Raw_.csv', sep = ',')
display(auto_price)
```

	symboling	normalized-losses	make	fuel-type	aspiration	\
0	3	?	alfa-romero	gas	std	
1	3	?	alfa-romero	gas	std	
2	1	?	alfa-romero	gas	std	
3	2	164	audi	gas	std	
4	2	164	audi	gas	std	
..	
200	-1	95	volvo	gas	std	
201	-1	95	volvo	gas	turbo	
202	-1	95	volvo	gas	std	
203	-1	95	volvo	diesel	turbo	
204	-1	95	volvo	gas	turbo	

	num-of-doors	body-style	drive-wheels	engine-location	wheel-base	...	\
0	two	convertible	rwd	front	88.6	...	
1	two	convertible	rwd	front	88.6	...	
2	two	hatchback	rwd	front	94.5	...	
3	four	sedan	fwd	front	99.8	...	
4	four	sedan	4wd	front	99.4	...	
..	
200	four	sedan	rwd	front	109.1	...	
201	four	sedan	rwd	front	109.1	...	
202	four	sedan	rwd	front	109.1	...	
203	four	sedan	rwd	front	109.1	...	
204	four	sedan	rwd	front	109.1	...	

	engine-size	fuel-system	bore	stroke	compression-ratio	horsepower	\
0	130	mpfi	3.47	2.68	9.0	111	
1	130	mpfi	3.47	2.68	9.0	111	
2	152	mpfi	2.68	3.47	9.0	154	
3	109	mpfi	3.19	3.40	10.0	102	

4	136	mpfi	3.19	3.40	8.0	115
..
200	141	mpfi	3.78	3.15	9.5	114
201	141	mpfi	3.78	3.15	8.7	160
202	173	mpfi	3.58	2.87	8.8	134
203	145	idi	3.01	3.40	23.0	106
204	141	mpfi	3.78	3.15	9.5	114

	peak-rpm	city-mpg	highway-mpg	price
0	5000	21	27	13495
1	5000	21	27	16500
2	5000	19	26	16500
3	5500	24	30	13950
4	5500	18	22	17450
..
200	5400	23	28	16845
201	5300	19	25	19045
202	5500	18	23	21485
203	4800	26	27	22470
204	5400	19	25	22625

[205 rows x 26 columns]

```
[3]: auto_price.isnull().sum()
```

```
[3]: symboling          0
normalized-losses      0
make                   0
fuel-type              0
aspiration             0
num-of-doors          0
body-style             0
drive-wheels           0
engine-location        0
wheel-base            0
length                0
width                 0
height                0
curb-weight            0
engine-type            0
num-of-cylinders       0
engine-size            0
fuel-system            0
bore                   0
stroke                 0
compression-ratio      0
horsepower             0
```

```

peak-rpm          0
city-mpg          0
highway-mpg       0
price             0
dtype: int64

```

```
[4]: # Replace '?' by np.nan
auto_price.replace('?', np.nan, inplace = True)
```

```
[5]: auto_price.isnull().sum()
```

```

[5]: symboling          0
normalized-losses    41
make                 0
fuel-type            0
aspiration            0
num-of-doors         2
body-style           0
drive-wheels         0
engine-location      0
wheel-base           0
length               0
width                0
height               0
curb-weight           0
engine-type           0
num-of-cylinders      0
engine-size           0
fuel-system           0
bore                  4
stroke                4
compression-ratio     0
horsepower            2
peak-rpm              2
city-mpg              0
highway-mpg           0
price                 4
dtype: int64

```

```
[6]: auto_price.select_dtypes(include = ['int64'])
```

```

[6]:      symboling  curb-weight  engine-size  city-mpg  highway-mpg
0          3         2548         130         21          27
1          3         2548         130         21          27
2          1         2823         152         19          26
3          2         2337         109         24          30
4          2         2824         136         18          22

```

..
200	-1	2952	141	23	28	
201	-1	3049	141	19	25	
202	-1	3012	173	18	23	
203	-1	3217	145	26	27	
204	-1	3062	141	19	25	

[205 rows x 5 columns]

```
[7]: # convert object columns to numerical columns
cols = ['bore', 'stroke', 'horsepower', 'peak-rpm', 'price']
auto_price[cols] = auto_price[cols].apply(pd.to_numeric)
```

```
[8]: auto_price.dtypes
```

```
[8]: symboling          int64
normalized-losses     object
make                  object
fuel-type             object
aspiration            object
num-of-doors          object
body-style            object
drive-wheels          object
engine-location       object
wheel-base           float64
length               float64
width                float64
height               float64
curb-weight           int64
engine-type           object
num-of-cylinders      object
engine-size           int64
fuel-system           object
bore                  float64
stroke                float64
compression-ratio     float64
horsepower            float64
peak-rpm              float64
city-mpg              int64
highway-mpg           int64
price                 float64
dtype: object
```

0.0.1 Central Tendencies

```
[9]: auto_price['body-style'].value_counts()
```

```
[9]: sedan          96
     hatchback      70
     wagon          25
     hardtop         8
     convertible     6
     Name: body-style, dtype: int64
```

```
[10]: auto_price['body-style'].mode()
```

```
[10]: 0    sedan
     dtype: object
```

```
[11]: # Median Value
     auto_price['price'].median()
```

```
[11]: 10295.0
```

```
[12]: # Mean Value
     auto_price['price'].mean()
```

```
[12]: 13207.129353233831
```

0.0.2 Handling Missing Values

```
[13]: auto_price['normalized-losses'].unique()
```

```
[13]: array([nan, '164', '158', '192', '188', '121', '98', '81', '118', '148',
        '110', '145', '137', '101', '78', '106', '85', '107', '104', '113',
        '150', '129', '115', '93', '142', '161', '153', '125', '128',
        '122', '103', '168', '108', '194', '231', '119', '154', '74',
        '186', '83', '102', '89', '87', '77', '91', '134', '65', '197',
        '90', '94', '256', '95'], dtype=object)
```

```
[14]: # Typecast column from object data types to float
     auto_price['normalized-losses'] = auto_price['normalized-losses'].
     ↪astype('float64')
```

```
[15]: auto_price['normalized-losses'].replace(np.nan, auto_price['normalized-losses'].
     ↪mean(), inplace = True)
```

```
[16]: auto_price.isnull().sum()
```

```
[16]: symboling          0
      normalized-losses  0
      make              0
      fuel-type         0
      aspiration        0
      num-of-doors      2
      body-style        0
      drive-wheels      0
      engine-location   0
      wheel-base        0
      length            0
      width             0
      height            0
      curb-weight       0
      engine-type       0
      num-of-cylinders  0
      engine-size       0
      fuel-system       0
      bore              4
      stroke            4
      compression-ratio 0
      horsepower       2
      peak-rpm          2
      city-mpg          0
      highway-mpg       0
      price             4
      dtype: int64
```

```
[17]: auto_price.dropna(inplace = True)
```

```
[18]: auto_price.shape
```

```
[18]: (193, 26)
```

0.0.3 Basic Imputation Techniques

- imputation by using statistics of each column with missing values
- imputation with constant value

```
[19]: diabetes = pd.read_csv('https://raw.githubusercontent.com/ammishra08/
      ↪MachineLearning/master/Datasets/Diabetes_Preprocessing.csv')
      diabetes.head()
```

```
[19]:   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI   \
0           6    148.0           72.0           35.0      NaN  33.6
1           1     85.0           66.0           29.0      NaN  26.6
```

2	8	183.0	64.0	NaN	NaN	23.3
3	1	89.0	66.0	23.0	94.0	28.1
4	0	137.0	4.0	35.0	168.0	43.1

	DiabetesPedigreeFunction	Age	Outcome
0	0.627	5	1
1	0.351	31	0
2	0.672	32	1
3	0.167	21	0
4	2.288	33	1

```
[20]: diabetes.shape
```

```
[20]: (768, 9)
```

```
[21]: diabetes.isnull().sum().sort_values(ascending = False)
```

```
[21]: Insulin          374
      SkinThickness  227
      BloodPressure  35
      BMI           11
      Glucose        5
      Outcome        0
      Age            0
      DiabetesPedigreeFunction  0
      Pregnancies    0
      dtype: int64
```

```
[22]: diabetes['Insulin'].replace(np.nan, diabetes['Insulin'].mean(), inplace = True)
```

```
[23]: diabetes.isnull().sum().sort_values(ascending = False)
```

```
[23]: SkinThickness  227
      BloodPressure  35
      BMI           11
      Glucose        5
      Outcome        0
      Age            0
      DiabetesPedigreeFunction  0
      Insulin        0
      Pregnancies    0
      dtype: int64
```

```
[24]: from sklearn.impute import SimpleImputer
      # strategy = 'mean', 'median', 'most_frequent'
      impute = SimpleImputer(strategy='median')
      # fit = train, transform = apply transformation
```

```
data_array = impute.fit_transform(diabetes)
```

```
[25]: diabetes_df = pd.DataFrame(data_array, columns = diabetes.columns)
diabetes_df
```

```
[25]:
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	\
0	6.0	148.0	72.0	35.0	105.659898	33.6	
1	1.0	85.0	66.0	29.0	105.659898	26.6	
2	8.0	183.0	64.0	27.0	105.659898	23.3	
3	1.0	89.0	66.0	23.0	94.000000	28.1	
4	0.0	137.0	4.0	35.0	168.000000	43.1	
..	
763	10.0	11.0	76.0	48.0	18.000000	32.9	
764	2.0	122.0	7.0	27.0	105.659898	36.8	
765	5.0	121.0	72.0	23.0	112.000000	26.2	
766	1.0	126.0	6.0	27.0	105.659898	3.1	
767	1.0	93.0	7.0	31.0	105.659898	3.4	

	DiabetesPedigreeFunction	Age	Outcome
0	0.627	5.0	1.0
1	0.351	31.0	0.0
2	0.672	32.0	1.0
3	0.167	21.0	0.0
4	2.288	33.0	1.0
..
763	0.171	63.0	0.0
764	0.340	27.0	0.0
765	0.245	3.0	0.0
766	0.349	47.0	1.0
767	0.315	23.0	0.0

```
[768 rows x 9 columns]
```

```
[26]: diabetes_df.isnull().sum()
```

```
[26]: Pregnancies      0
Glucose              0
BloodPressure        0
SkinThickness        0
Insulin              0
BMI                  0
DiabetesPedigreeFunction  0
Age                  0
Outcome              0
dtype: int64
```

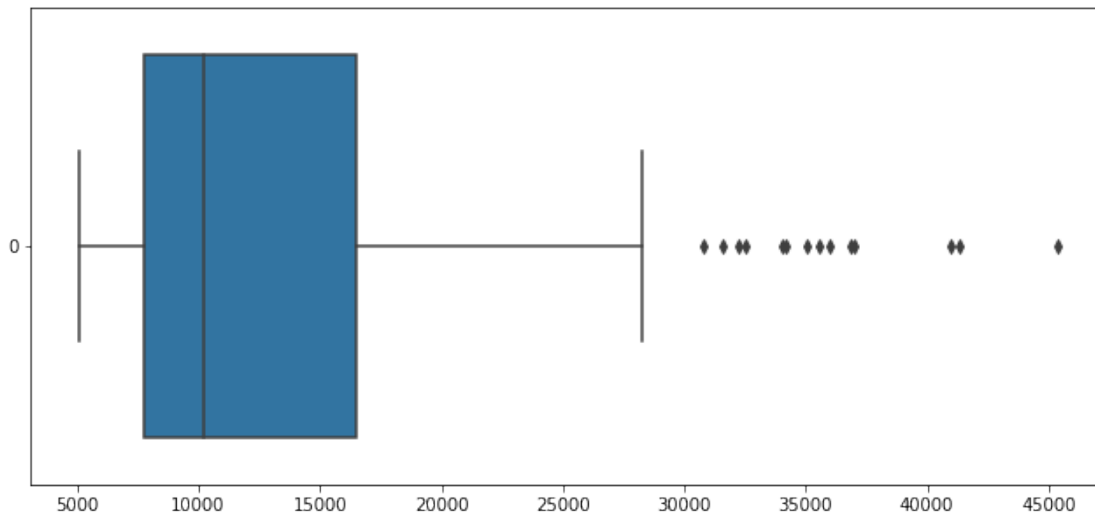


```
[27]: import matplotlib.pyplot as plt
import seaborn as sns
```

Matplotlib is building the font cache; this may take a moment.

```
[29]: # figsize = (width,height)
plt.figure(figsize = (11,5))
# orient = orientation of plot h = horizontal
sns.boxplot(data = auto_price['price'], orient = 'h')
```

[29]: <AxesSubplot:>



```
[30]: def find_outliers(df, cols):
    q1 = df[cols].quantile(.25)
    q3 = df[cols].quantile(.75)
    iqr = q3 - q1
    min_r = q1 - 1.5 * iqr
    max_r = q3 + 1.5 * iqr
    outliers_indices = df.index[(df[cols] < min_r) | (df[cols] > max_r)]
    return outliers_indices
```

```
[31]: find_outliers(auto_price, 'price')
```

```
[31]: Int64Index([15, 16, 17, 47, 48, 49, 70, 71, 72, 73, 74, 126, 127, 128],
dtype='int64')
```

0.0.4 Treating Outliers

- Remove/Trim the outliers

- Trim the data (5% to 95%)
- Mean/median apply imputation

```
[32]: import numpy as np
      from scipy import stats
```

```
[33]: auto_price_numeric = auto_price.select_dtypes(include = ['int64', 'float64'])
```

```
[34]: z = np.abs(stats.zscore(auto_price_numeric))
```

```
[35]: # z-score > +3 or z-score < -3
      np.where((z > 3) | (z < -3))
```

```
[35]: (array([ 15, 17, 17, 28, 28, 43, 44, 45, 45, 64, 64, 64, 65,
              65, 81, 97, 102, 104, 122, 146, 147, 153, 154, 162, 170, 172,
              175, 178, 180, 191]),
      array([15, 13, 14, 13, 14, 7, 7, 7, 11, 2, 7, 15, 7, 15, 13, 1, 9,
              9, 9, 10, 10, 12, 12, 10, 10, 10, 10, 1, 10, 10]))
```

```
[36]: auto_price.groupby(['make'])
```

```
[36]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x7f62ec4c1390>
```

```
[37]: auto_price.groupby(['make']).first()
```

```
[37]:      symboling  normalized-losses  fuel-type  aspiration  num-of-doors  \
make
alfa-romero      3             122.0      gas      std         two
audi             2             164.0      gas      std         four
bmw              2             192.0      gas      std         two
chevrolet        2             121.0      gas      std         two
dodge            1             118.0      gas      std         two
honda            2             137.0      gas      std         two
isuzu            0             122.0      gas      std         four
jaguar           0             145.0      gas      std         four
mazda            1             104.0      gas      std         two
mercedes-benz    -1              93.0    diesel    turbo         four
mercury          1             122.0      gas    turbo         two
mitsubishi       2             161.0      gas      std         two
nissan            1             128.0      gas      std         two
peugot           0             161.0      gas      std         four
plymouth         1             119.0      gas      std         two
porsche          3             186.0      gas      std         two
saab             3             150.0      gas      std         two
subaru           2              83.0      gas      std         two
toyota           1              87.0      gas      std         two
volkswagen       2             122.0    diesel    std         two
```

volvo	-2	103.0	gas	std	four
	body-style	drive-wheels	engine-location	wheel-base	length \
make					
alfa-romero	convertible	rwd	front	88.6	168.8
audi	sedan	fwd	front	99.8	176.6
bmw	sedan	rwd	front	101.2	176.8
chevrolet	hatchback	fwd	front	88.4	141.1
dodge	hatchback	fwd	front	93.7	157.3
honda	hatchback	fwd	front	86.6	144.6
isuzu	sedan	rwd	front	94.3	170.7
jaguar	sedan	rwd	front	113.0	199.6
mazda	hatchback	fwd	front	93.1	159.1
mercedes-benz	sedan	rwd	front	110.0	190.9
mercury	hatchback	rwd	front	102.7	178.4
mitsubishi	hatchback	fwd	front	93.7	157.3
nissan	sedan	fwd	front	94.5	165.3
peugot	sedan	rwd	front	107.9	186.7
plymouth	hatchback	fwd	front	93.7	157.3
porsche	hatchback	rwd	front	94.5	168.9
saab	hatchback	fwd	front	99.1	186.6
subaru	hatchback	fwd	front	93.7	156.9
toyota	hatchback	fwd	front	95.7	158.7
volkswagen	sedan	fwd	front	97.3	171.7
volvo	sedan	rwd	front	104.3	188.8

	...	engine-size	fuel-system	bore	stroke	compression-ratio \
make	...					
alfa-romero	...	130	mpfi	3.47	2.68	9.00
audi	...	109	mpfi	3.19	3.40	10.00
bmw	...	108	mpfi	3.50	2.80	8.80
chevrolet	...	61	2bbl	2.91	3.03	9.50
dodge	...	90	2bbl	2.97	3.23	9.41
honda	...	92	1bbl	2.91	3.41	9.60
isuzu	...	111	2bbl	3.31	3.23	8.50
jaguar	...	258	mpfi	3.63	4.17	8.10
mazda	...	91	2bbl	3.03	3.15	9.00
mercedes-benz	...	183	idi	3.58	3.64	21.50
mercury	...	140	mpfi	3.78	3.12	8.00
mitsubishi	...	92	2bbl	2.97	3.23	9.40
nissan	...	97	2bbl	3.15	3.29	9.40
peugot	...	120	mpfi	3.46	3.19	8.40
plymouth	...	90	2bbl	2.97	3.23	9.40
porsche	...	151	mpfi	3.94	3.11	9.50
saab	...	121	mpfi	3.54	3.07	9.31
subaru	...	97	2bbl	3.62	2.36	9.00
toyota	...	92	2bbl	3.05	3.03	9.00

volkswagen	...	97	idi	3.01	3.40	23.00
volvo	...	141	mpfi	3.78	3.15	9.50

	horsepower	peak-rpm	city-mpg	highway-mpg	price
make					
alfa-romero	111.0	5000.0	21	27	13495.0
audi	102.0	5500.0	24	30	13950.0
bmw	101.0	5800.0	23	29	16430.0
chevrolet	48.0	5100.0	47	53	5151.0
dodge	68.0	5500.0	37	41	5572.0
honda	58.0	4800.0	49	54	6479.0
isuzu	78.0	4800.0	24	29	6785.0
jaguar	176.0	4750.0	15	19	32250.0
mazda	68.0	5000.0	30	31	5195.0
mercedes-benz	123.0	4350.0	22	25	25552.0
mercury	175.0	5000.0	19	24	16503.0
mitsubishi	68.0	5500.0	37	41	5389.0
nissan	69.0	5200.0	31	37	5499.0
peugot	97.0	5000.0	19	24	11900.0
plymouth	68.0	5500.0	37	41	5572.0
porsche	143.0	5500.0	19	27	22018.0
saab	110.0	5250.0	21	28	11850.0
subaru	69.0	4900.0	31	36	5118.0
toyota	62.0	4800.0	35	39	5348.0
volkswagen	52.0	4800.0	37	46	7775.0
volvo	114.0	5400.0	23	28	12940.0

[21 rows x 25 columns]

```
[38]: auto_price.groupby(['make', 'body-style']).first()
```

```
[38]:
```

make	body-style	symboling	normalized-losses	fuel-type	aspiration \
alfa-romero	convertible	3	122.0	gas	std
	hatchback	1	122.0	gas	std
audi	sedan	2	164.0	gas	std
	wagon	1	122.0	gas	std
bmw	sedan	2	192.0	gas	std
chevrolet	hatchback	2	121.0	gas	std
	sedan	0	81.0	gas	std
dodge	hatchback	1	118.0	gas	std
	sedan	1	148.0	gas	std
	wagon	-1	110.0	gas	std
honda	hatchback	2	137.0	gas	std
	sedan	0	110.0	gas	std
	wagon	0	78.0	gas	std
isuzu	hatchback	2	122.0	gas	std

	sedan	0	122.0	gas	std
jaguar	sedan	0	145.0	gas	std
mazda	hatchback	1	104.0	gas	std
	sedan	1	113.0	gas	std
mercedes-benz	convertible	3	142.0	gas	std
	hardtop	0	93.0	diesel	turbo
	sedan	-1	93.0	diesel	turbo
	wagon	-1	93.0	diesel	turbo
mercury	hatchback	1	122.0	gas	turbo
mitsubishi	hatchback	2	161.0	gas	std
	sedan	1	125.0	gas	std
nissan	hardtop	2	168.0	gas	std
	hatchback	1	128.0	gas	std
	sedan	1	128.0	gas	std
	wagon	1	103.0	gas	std
peugot	sedan	0	161.0	gas	std
	wagon	0	122.0	gas	std
plymouth	hatchback	1	119.0	gas	std
	sedan	1	154.0	gas	std
	wagon	-1	74.0	gas	std
porsche	convertible	3	122.0	gas	std
	hardtop	3	122.0	gas	std
	hatchback	3	186.0	gas	std
saab	hatchback	3	150.0	gas	std
	sedan	2	104.0	gas	std
subaru	hatchback	2	83.0	gas	std
	sedan	0	102.0	gas	std
	wagon	0	89.0	gas	std
toyota	convertible	2	134.0	gas	std
	hardtop	2	134.0	gas	std
	hatchback	1	87.0	gas	std
	sedan	0	91.0	gas	std
	wagon	0	77.0	gas	std
volkswagen	convertible	3	122.0	gas	std
	hatchback	3	256.0	gas	std
	sedan	2	122.0	diesel	std
	wagon	0	122.0	gas	std
volvo	sedan	-2	103.0	gas	std
	wagon	-1	74.0	gas	std

make	body-style	num-of-doors	drive-wheels	engine-location	\
alfa-romero	convertible	two	rwd	front	
	hatchback	two	rwd	front	
audi	sedan	four	fwd	front	
	wagon	four	fwd	front	
bmw	sedan	two	rwd	front	

chevrolet	hatchback	two	fwd	front
	sedan	four	fwd	front
dodge	hatchback	two	fwd	front
	sedan	four	fwd	front
	wagon	four	fwd	front
honda	hatchback	two	fwd	front
	sedan	four	fwd	front
	wagon	four	fwd	front
isuzu	hatchback	two	rwd	front
	sedan	four	rwd	front
jaguar	sedan	four	rwd	front
mazda	hatchback	two	fwd	front
	sedan	four	fwd	front
mercedes-benz	convertible	two	rwd	front
	hardtop	two	rwd	front
	sedan	four	rwd	front
	wagon	four	rwd	front
mercury	hatchback	two	rwd	front
mitsubishi	hatchback	two	fwd	front
	sedan	four	fwd	front
nissan	hardtop	two	fwd	front
	hatchback	two	fwd	front
	sedan	two	fwd	front
	wagon	four	fwd	front
peugot	sedan	four	rwd	front
	wagon	four	rwd	front
plymouth	hatchback	two	fwd	front
	sedan	four	fwd	front
	wagon	four	fwd	front
porsche	convertible	two	rwd	rear
	hardtop	two	rwd	rear
	hatchback	two	rwd	front
saab	hatchback	two	fwd	front
	sedan	four	fwd	front
subaru	hatchback	two	fwd	front
	sedan	four	fwd	front
	wagon	four	fwd	front
toyota	convertible	two	rwd	front
	hardtop	two	rwd	front
	hatchback	two	fwd	front
	sedan	four	fwd	front
	wagon	four	fwd	front
volkswagen	convertible	two	fwd	front
	hatchback	two	fwd	front
	sedan	two	fwd	front
	wagon	four	fwd	front
volvo	sedan	four	rwd	front

	wagon	four		rwd	front	
		wheel-base	length	width	...	engine-size \
make	body-style				...	
alfa-romero	convertible	88.6	168.8	64.1	...	130
	hatchback	94.5	171.2	65.5	...	152
audi	sedan	99.8	176.6	66.2	...	109
	wagon	105.8	192.7	71.4	...	136
bmw	sedan	101.2	176.8	64.8	...	108
chevrolet	hatchback	88.4	141.1	60.3	...	61
	sedan	94.5	158.8	63.6	...	90
dodge	hatchback	93.7	157.3	63.8	...	90
	sedan	93.7	157.3	63.8	...	90
	wagon	103.3	174.6	64.6	...	122
honda	hatchback	86.6	144.6	63.9	...	92
	sedan	96.5	163.4	64.0	...	92
	wagon	96.5	157.1	63.9	...	92
isuzu	hatchback	96.0	172.6	65.2	...	119
	sedan	94.3	170.7	61.8	...	111
jaguar	sedan	113.0	199.6	69.6	...	258
mazda	hatchback	93.1	159.1	64.2	...	91
	sedan	93.1	166.8	64.2	...	91
mercedes-benz	convertible	96.6	180.3	70.5	...	234
	hardtop	106.7	187.5	70.3	...	183
	sedan	110.0	190.9	70.3	...	183
	wagon	110.0	190.9	70.3	...	183
mercury	hatchback	102.7	178.4	68.0	...	140
mitsubishi	hatchback	93.7	157.3	64.4	...	92
	sedan	96.3	172.4	65.4	...	122
nissan	hardtop	95.1	162.4	63.8	...	97
	hatchback	94.5	165.6	63.8	...	97
	sedan	94.5	165.3	63.8	...	97
	wagon	94.5	170.2	63.8	...	97
peugot	sedan	107.9	186.7	68.4	...	120
	wagon	114.2	198.9	68.4	...	120
plymouth	hatchback	93.7	157.3	63.8	...	90
	sedan	93.7	167.3	63.8	...	90
	wagon	103.3	174.6	64.6	...	122
porsche	convertible	89.5	168.9	65.0	...	194
	hardtop	89.5	168.9	65.0	...	194
	hatchback	94.5	168.9	68.3	...	151
saab	hatchback	99.1	186.6	66.5	...	121
	sedan	99.1	186.6	66.5	...	121
subaru	hatchback	93.7	156.9	63.4	...	97
	sedan	97.2	172.0	65.4	...	108
	wagon	97.0	173.5	65.4	...	108
toyota	convertible	98.4	176.2	65.6	...	146

volkswagen	hardtop	98.4	176.2	65.6	...	146
	hatchback	95.7	158.7	63.6	...	92
	sedan	95.7	166.3	64.4	...	98
	wagon	95.7	169.7	63.6	...	92
	convertible	94.5	159.3	64.2	...	109
	hatchback	94.5	165.7	64.0	...	109
	sedan	97.3	171.7	65.5	...	97
	wagon	100.4	183.1	66.9	...	109
volvo	sedan	104.3	188.8	67.2	...	141
	wagon	104.3	188.8	67.2	...	141

make	body-style	fuel-system	bore	stroke	compression-ratio	\
alfa-romero	convertible	mpfi	3.47	2.68		9.00
	hatchback	mpfi	2.68	3.47		9.00
audi	sedan	mpfi	3.19	3.40		10.00
	wagon	mpfi	3.19	3.40		8.50
bmw	sedan	mpfi	3.50	2.80		8.80
chevrolet	hatchback	2bbl	2.91	3.03		9.50
	sedan	2bbl	3.03	3.11		9.60
dodge	hatchback	2bbl	2.97	3.23		9.41
	sedan	2bbl	2.97	3.23		9.40
	wagon	2bbl	3.34	3.46		8.50
honda	hatchback	1bbl	2.91	3.41		9.60
	sedan	1bbl	2.91	3.41		9.20
	wagon	1bbl	2.92	3.41		9.20
isuzu	hatchback	spfi	3.43	3.23		9.20
	sedan	2bbl	3.31	3.23		8.50
jaguar	sedan	mpfi	3.63	4.17		8.10
mazda	hatchback	2bbl	3.03	3.15		9.00
	sedan	2bbl	3.03	3.15		9.00
mercedes-benz	convertible	mpfi	3.46	3.10		8.30
	hardtop	idi	3.58	3.64		21.50
	sedan	idi	3.58	3.64		21.50
	wagon	idi	3.58	3.64		21.50
mercury	hatchback	mpfi	3.78	3.12		8.00
mitsubishi	hatchback	2bbl	2.97	3.23		9.40
	sedan	2bbl	3.35	3.46		8.50
nissan	hardtop	2bbl	3.15	3.29		9.40
	hatchback	2bbl	3.15	3.29		9.40
	sedan	2bbl	3.15	3.29		9.40
	wagon	2bbl	3.15	3.29		9.40
peugot	sedan	mpfi	3.46	3.19		8.40
	wagon	mpfi	3.46	3.19		8.40
plymouth	hatchback	2bbl	2.97	3.23		9.40
	sedan	2bbl	2.97	3.23		9.40
	wagon	2bbl	3.35	3.46		8.50

porsche	convertible	mpfi	3.74	2.90	9.50
	hardtop	mpfi	3.74	2.90	9.50
	hatchback	mpfi	3.94	3.11	9.50
saab	hatchback	mpfi	3.54	3.07	9.31
	sedan	mpfi	3.54	3.07	9.30
subaru	hatchback	2bbl	3.62	2.36	9.00
	sedan	2bbl	3.62	2.64	9.50
	wagon	2bbl	3.62	2.64	9.00
toyota	convertible	mpfi	3.62	3.50	9.30
	hardtop	mpfi	3.62	3.50	9.30
	hatchback	2bbl	3.05	3.03	9.00
	sedan	2bbl	3.19	3.03	9.00
	wagon	2bbl	3.05	3.03	9.00
volkswagen	convertible	mpfi	3.19	3.40	8.50
	hatchback	mpfi	3.19	3.40	8.50
	sedan	idi	3.01	3.40	23.00
	wagon	mpfi	3.19	3.40	9.00
volvo	sedan	mpfi	3.78	3.15	9.50
	wagon	mpfi	3.78	3.15	9.50

make	body-style	horsepower	peak-rpm	city-mpg	highway-mpg	price
alfa-romero	convertible	111.0	5000.0	21	27	13495.0
	hatchback	154.0	5000.0	19	26	16500.0
audi	sedan	102.0	5500.0	24	30	13950.0
	wagon	110.0	5500.0	19	25	18920.0
bmw	sedan	101.0	5800.0	23	29	16430.0
chevrolet	hatchback	48.0	5100.0	47	53	5151.0
	sedan	70.0	5400.0	38	43	6575.0
dodge	hatchback	68.0	5500.0	37	41	5572.0
	sedan	68.0	5500.0	31	38	6692.0
	wagon	88.0	5000.0	24	30	8921.0
honda	hatchback	58.0	4800.0	49	54	6479.0
	sedan	76.0	6000.0	30	34	7295.0
	wagon	76.0	6000.0	30	34	7295.0
isuzu	hatchback	90.0	5000.0	24	29	11048.0
	sedan	78.0	4800.0	24	29	6785.0
jaguar	sedan	176.0	4750.0	15	19	32250.0
mazda	hatchback	68.0	5000.0	30	31	5195.0
	sedan	68.0	5000.0	31	38	6695.0
mercedes-benz	convertible	155.0	4750.0	16	18	35056.0
	hardtop	123.0	4350.0	22	25	28176.0
	sedan	123.0	4350.0	22	25	25552.0
	wagon	123.0	4350.0	22	25	28248.0
mercury	hatchback	175.0	5000.0	19	24	16503.0
mitsubishi	hatchback	68.0	5500.0	37	41	5389.0
	sedan	88.0	5000.0	25	32	6989.0

nissan	hardtop	69.0	5200.0	31	37	8249.0
	hatchback	69.0	5200.0	31	37	7799.0
	sedan	69.0	5200.0	31	37	5499.0
	wagon	69.0	5200.0	31	37	7349.0
peugot	sedan	97.0	5000.0	19	24	11900.0
	wagon	97.0	5000.0	19	24	12440.0
plymouth	hatchback	68.0	5500.0	37	41	5572.0
	sedan	68.0	5500.0	31	38	6692.0
	wagon	88.0	5000.0	24	30	8921.0
porsche	convertible	207.0	5900.0	17	25	37028.0
	hardtop	207.0	5900.0	17	25	32528.0
	hatchback	143.0	5500.0	19	27	22018.0
saab	hatchback	110.0	5250.0	21	28	11850.0
	sedan	110.0	5250.0	21	28	12170.0
subaru	hatchback	69.0	4900.0	31	36	5118.0
	sedan	82.0	4800.0	32	37	7126.0
	wagon	82.0	4800.0	28	32	7463.0
toyota	convertible	116.0	4800.0	24	30	17669.0
	hardtop	116.0	4800.0	24	30	8449.0
	hatchback	62.0	4800.0	35	39	5348.0
	sedan	70.0	4800.0	30	37	6938.0
	wagon	62.0	4800.0	31	37	6918.0
volkswagen	convertible	90.0	5500.0	24	29	11595.0
	hatchback	90.0	5500.0	24	29	9980.0
	sedan	52.0	4800.0	37	46	7775.0
	wagon	88.0	5500.0	25	31	12290.0
volvo	sedan	114.0	5400.0	23	28	12940.0
	wagon	114.0	5400.0	23	28	13415.0

[53 rows x 24 columns]

[]: