cars.sales

November 10, 2024

[58]:
```
from bs4 import BeautifulSoup as bs
from urllib.request import urlopen
```

0.0.1 connecting to the webpage we are trying to scrap

[59]:
```
url = "https://www.voiturenet.ma/en/
 ↪buy-cars-in-Rabat-Sale-Zammour-Zaer-ar-Ribat?page=4"
client = urlopen(url)
html = client.read()
client.close()
```

0.0.2 html parsing with bs

[60]:
```
soup = bs(html,"html.parser")
```

[61]:
```
container = soup.find_all("div",{'class':'ad-specification'})
len(container)
```

[61]: 20

[62]:
```
bs.prettify(container[0])
```

[62]:
```
'<div class="ad-specification">\n <div class="specification-section">\n  <h4
title="Mercedes-Benz Eqe">\n   Mercedes-Benz Eqe\n  </h4>\n  <div
class="location">\n   <i class="fa fa-map-marker">\n   </i>\n   Casablanca\n
</div>\n </div>\n <div class="specification-section xl-wrapper">\n  <div
class="wrapper">\n   <div class="ad-vehicle-price">\n    <div class="title">\n
Price\n    </div>\n    <div class="value">\n     <span class="price-wrap">\n
MAD\n      <span class="price">\n       780 000\n      </span>\n     </span>\n
</div>\n   </div>\n   <div class="ad-vehicle-mileage">\n    <div class="title"
title="Mileage">\n     Mileage\n     <div class="value">\n      79,203 km\n
</div>\n    </div>\n   </div>\n  </div>\n  <div class="ad-vehicle-engine">\n
<div class="engine-capacity">\n    <img alt="Engine xs"
src="https://voiturenet.ma/assets/design/icons/engine-
xs-40c8ae5e5247b08fa64071a29989cfcda52f437c48a4042ce7363e907fcceb40.svg"/>\n
<span>\n     N/A\n    </span>\n    <span>\n     (Ev)\n    </span>\n   </div>\n
<div class="transmission">\n    <img alt="Transmission xs"
src="https://voiturenet.ma/assets/design/icons/transmission-
```

xs-7c13fb8d43ee234f637f71d3910fc74ae8e558713ccb9fbdde19645bfb376907.svg"/>\n \n Automatic\n \n \n \n \n \n \n La Mercedes EQE est une berline électrique premium qui incarne le savoir-faire de Mercedes-Benz en matière de luxe et de technologie. Esthétique moderne : L\'EQE arbore des lignes fluides et un design aérodynamique, offrant une silhouette élégante et futuriste. Éclairage LED : Les phares et feux arrière LED sont sophistiqués et contribuent à l\'identité visuelle de la marque. Motorisation électrique : Proposée avec plusieurs options de puissance, l\'EQE offre une accélération rapide et une conduite silencieuse. Autonomie : Grâce à une batterie de grande capacité, elle propose une autonomie compétitive pour les trajets quotidiens et les longs voyages. Confort et technologie : L\'habitacle est spacieux, luxueux et équipé de matériaux haut de gamme. Le système MBUX (Mercedes-Benz User Experience) intègre des commandes vocales et un affichage numérique avancé. Écran tactile : Un grand écran central permet d\'accéder facilement aux fonctionnalités de navigation, multimédia et de gestion du véhicule. Assistance à la conduite : L\'EQE est dotée de nombreuses aides à la conduite, allant du régulateur de vitesse adaptatif à la gestion semi-autonome. Connectivité : Intégration fluide avec les smartphones et les services en ligne pour un accès à distance aux fonctionnalités du véhicule. Écologique : Conçue avec des matériaux durables et une empreinte carbone réduite, l\'EQE représente l\'engagement de Mercedes envers une mobilité plus verte. L\'EQE est une option attrayante pour ceux qui recherchent une berline électrique alliant luxe, performance et technologie de pointe.\n \n \n \n \n \n Posted about 1 month ago\n \n \n \n \n'

0.0.3 accessing html elements

[63]:
```
#car name
container[0].h4.text
```

[63]: 'Mercedes-Benz Eqe'

[64]:
```
#location
a = container[0].findAll("div",{"class":"location"})
b = a[0].text
b
```

[64]: 'Casablanca'

[65]:
```
#price
a = container[0].findAll("span",{"class":"price"})
b = a[0].text
b
```

[65]: '780 000'

[66]:
```
#Mileage
a = container[0].findAll("div",{"class":"ad-vehicle-mileage"})
b = a[0].text
b
```

[66]: 'Mileage79,203 km'

[67]:
```
#transmission
a = container[0].findAll("div",{"class":"transmission"})
b = a[0].text
b
```

[67]: 'Automatic'

[68]:
```
a = container[0].findAll("div",{"class":"engine-capacity"})
b = a[0].text
b
```

[68]: 'N/A (Ev)'

[69]:
```
a = container[0].findAll("div",{"class":"md-hidden sort-ad-description"})
b = a[0].text
b
```

[69]: "La Mercedes EQE est une berline électrique premium qui incarne le savoir-faire de Mercedes-Benz en matière de luxe et de technologie. Esthétique moderne : L'EQE arbore des lignes fluides et un design aérodynamique, offrant une silhouette élégante et futuriste. Éclairage LED : Les phares et feux arrière LED sont sophistiqués et contribuent à l'identité visuelle de la marque. Motorisation électrique : Proposée avec plusieurs options de puissance, l'EQE offre une accélération rapide et une conduite silencieuse. Autonomie : Grâce à une batterie de grande capacité, elle propose une autonomie compétitive pour les trajets quotidiens et les longs voyages. Confort et technologie : L'habitacle est spacieux, luxueux et équipé de matériaux haut de gamme. Le système MBUX (Mercedes-Benz User Experience) intègre des commandes vocales et un affichage numérique avancé. Écran tactile : Un grand écran central permet d'accéder facilement aux fonctionnalités de navigation, multimédia et de gestion du véhicule. Assistance à la conduite : L'EQE est dotée de nombreuses aides à la conduite, allant du régulateur de vitesse adaptatif à la gestion semi-autonome. Connectivité : Intégration fluide avec les smartphones et les services en ligne pour un accès à distance aux fonctionnalités du véhicule. Écologique : Conçue avec des matériaux durables et une empreinte carbone réduite, l'EQE représente l'engagement de Mercedes envers une mobilité plus verte. L'EQE est une option attrayante pour ceux qui recherchent une berline électrique alliant luxe, performance et technologie de pointe."

1 SCRAPING PROGRAM

[70]:
```
import pandas as pd
```

[71]:
```
number_pages = 78
brand = []
location = []
price = []
km = []
transmission = []
engine = []
desc = []
for i in range(1,number_pages+1) :
    url = f"https://www.voiturenet.ma/en/
 ↪buy-cars-in-Rabat-Sale-Zammour-Zaer-ar-Ribat?page={i}"
    client = urlopen(url)
    html = client.read()
    client.close()
    soup = bs(html,'html.parser')
    container = soup.find_all("div",{'class':'ad-specification'})
    for j in range(len(container)) :
        brand.append(container[j].h4.text)

        l = container[j].find("div",{"class":"location"})
        location.append(l.text if l else "N/A")

        p = container[j].find("span",{"class":"price"})
        price.append(p.text if p else "N/A")

        k = container[j].find("div",{"class":"ad-vehicle-mileage"})
        km.append(k.text if k else "N/A")

        t = container[j].find("div",{"class":"transmission"})
        transmission.append(t.text if t else "N/A")

        e = container[j].find("div",{"class":"engine-capacity"})
        engine.append(e.text if e else "N/A")

        d = container[j].find("div",{"class":"md-hidden sort-ad-description"})
        desc.append(d.text if d else "N/A")

data = pd.DataFrame({
"Brand" : brand ,
"Location" : location,
"Price" : price,
"Metrage" : km ,
"Transmission" : transmission ,
```

```
"Engine" : engine ,
"Descreption" : desc

})

data.head()
```

[71]:

	Brand	Location	Price	Metrage	Transmission	Engine	Descreption
0	2022 Renault Captur	ad-Dar-al-Bayda	N/A	Mileage41,713 km	Manual	1,5 L (Diesel)	Renault Captur Manuelle 6 rapports D…
1	2014 Dacia Dokker	Casablanca	N/A	Mileage192,000 km	Manual	N/A (Diesel)	Dacia docker jama…
2	Leapmotor C10 530	Ain Sebaa	399 000	N/A	Automatic	N/A (Ev)	La Leap Motor C10 530 Smart Edition est un véh…
3	Mercedes-Benz AMG GLE	al-Ayun Sidi Malluk	N/A	N/A	Manual	N/A (Phev)	Mercedes Benz classe 200c tout option jamais a…
4	Renault Clio	al-Ayun Sidi Malluk	29 000	N/A	Automatic	N/A (Petrol)	…

:[72]
```
data.head()
```

[72]:

	Brand	Location	Price	Metrage	Transmission	Engine
0	2022 Renault Captur	ad-Dar-al-Bayda	N/A	Mileage41,713 km	Manual	1,5 L (Diesel)
1	2014 Dacia Dokker	Casablanca	N/A	Mileage192,000 km	Manual	N/A (Diesel)
2	Leapmotor C10 530	Ain Sebaa	399 000	N/A	Automatic	N/A (Ev)
3	Mercedes-Benz AMG GLE	al-Ayun Sidi Malluk	N/A	N/A		
4	Renault Clio	al-Ayun Sidi Malluk	29 000	N/A		

```
3       Manual      N/A (Phev)
4    Automatic    N/A (Petrol)

                                        Descreption
0    Renault Captur Manuelle 6 rapports     D…
1      Dacia docker                 jama…
2  La Leap Motor C10 530 Smart Edition est un véh…
3  Mercedes Benz classe 200c tout option jamais a…
4                              …
```

:[73]
```
data['Metrage'] = data['Metrage'].apply(lambda x: int(''.join(filter(str.
↪isdigit, str(x)))) if any(char.isdigit() for char in str(x)) else None)
```

[74]:
```
data.head()
```

[74]:
```
                  Brand             Location    Price   Metrage Transmission  \
0    2022 Renault Captur      ad-Dar-al-Bayda      N/A   41713.0       Manual
1      2014 Dacia Dokker           Casablanca      N/A  192000.0       Manual
2      Leapmotor C10 530            Ain Sebaa  399 000       NaN    Automatic
3  Mercedes-Benz AMG GLE  al-Ayun Sidi Malluk      N/A       NaN       Manual
4           Renault Clio  al-Ayun Sidi Malluk   29 000       NaN    Automatic

           Engine                                        Descreption
0  1,5 L (Diesel)    Renault Captur Manuelle 6 rapports     D…
1    N/A (Diesel)      Dacia docker                 jama…
2        N/A (Ev)  La Leap Motor C10 530 Smart Edition est un véh…
3      N/A (Phev)  Mercedes Benz classe 200c tout option jamais a…
4    N/A (Petrol)                                …
```

:[75]
```
import numpy as np
```

[76]:
```
data['Price'] = data['Price'].apply(lambda x : np.nan if x is None else x)
data
```

[76]:
```
                      Brand             Location    Price   Metrage  \
0       2022 Renault Captur      ad-Dar-al-Bayda      N/A   41713.0
1         2014 Dacia Dokker           Casablanca      N/A  192000.0
2         Leapmotor C10 530            Ain Sebaa  399 000       NaN
3     Mercedes-Benz AMG GLE  al-Ayun Sidi Malluk      N/A       NaN
4              Renault Clio  al-Ayun Sidi Malluk   29 000       NaN
…                      …                 …       …       …
1554    2009 Toyota Corolla      ad-Dar-al-Bayda   70 000  162000.0
1555        Volkswagen Golf      ad-Dar-al-Bayda   95 000  215000.0
1556      2002 BMW 3-Series      ad-Dar-al-Bayda   53 000  260000.0
1557            Ford Fiesta      ad-Dar-al-Bayda   69 000  140000.0
1558      2007 Daewoo Matiz      ad-Dar-al-Bayda   29 000       NaN
```

```
     Transmission          Engine  \
0          Manual  1,5 L (Diesel)
1          Manual    N/A (Diesel)
2       Automatic        N/A (Ev)
3          Manual      N/A (Phev)
4       Automatic    N/A (Petrol)
…            …             …
1554       Manual    N/A (Petrol)
1555       Manual    N/A (Petrol)
1556       Manual    N/A (Petrol)
1557       Manual    N/A (Petrol)
1558       Manual    N/A (Petrol)

                                              Descreption
0       Renault Captur Manuelle 6 rapports     D…
1         Dacia docker                jama…
2     La Leap Motor C10 530 Smart Edition est un véh…
3     Mercedes Benz classe 200c tout option jamais a…
4                                 …
…                                                   …
1554  Toyota Corolla Essence Modèle : 2009 Cheveux :…
1555  Volkswagen Golf 5 Modèle : 2007 Essence Cheveu…
1556  BMW_e46_318i MODÈLE : 2002 ESSENCE Cheveux : 1…
1557  Fiesta Modèl 2012 Essence Impôt : 350dh Cheveu…
1558  Modèl : 2007 Prix : 29000dh            …

[1559 rows x 7 columns]
```

[77]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1559 entries, 0 to 1558
Data columns (total 7 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Brand         1559 non-null   object
 1   Location      1559 non-null   object
 2   Price         1559 non-null   object
 3   Metrage       830 non-null    float64
 4   Transmission  1559 non-null   object
 5   Engine        1559 non-null   object
 6   Descreption   1559 non-null   object
dtypes: float64(1), object(6)
memory usage: 85.4+ KB
```

[78]:
```
data
```

[78]:
```
                        Brand             Location    Price   Metrage  \
0       2022 Renault Captur      ad-Dar-al-Bayda      N/A   41713.0
1         2014 Dacia Dokker           Casablanca      N/A  192000.0
2         Leapmotor C10 530            Ain Sebaa  399 000       NaN
3     Mercedes-Benz AMG GLE  al-Ayun Sidi Malluk      N/A       NaN
4              Renault Clio  al-Ayun Sidi Malluk   29 000       NaN
…                      …                 …      …       …
1554    2009 Toyota Corolla      ad-Dar-al-Bayda   70 000  162000.0
1555        Volkswagen Golf      ad-Dar-al-Bayda   95 000  215000.0
1556      2002 BMW 3-Series      ad-Dar-al-Bayda   53 000  260000.0
1557            Ford Fiesta      ad-Dar-al-Bayda   69 000  140000.0
1558      2007 Daewoo Matiz      ad-Dar-al-Bayda   29 000       NaN

     Transmission          Engine  \
0          Manual  1,5 L (Diesel)
1          Manual    N/A (Diesel)
2       Automatic        N/A (Ev)
3          Manual      N/A (Phev)
4       Automatic    N/A (Petrol)
…            …              …
1554       Manual    N/A (Petrol)
1555       Manual    N/A (Petrol)
1556       Manual    N/A (Petrol)
1557       Manual    N/A (Petrol)
1558       Manual    N/A (Petrol)

                                            Descreption
0       Renault Captur Manuelle 6 rapports     D…
1         Dacia docker                jama…
2     La Leap Motor C10 530 Smart Edition est un véh…
3     Mercedes Benz classe 200c tout option jamais a…
4                                 …
…                                                   …
1554  Toyota Corolla Essence Modèle : 2009 Cheveux :…
1555  Volkswagen Golf 5 Modèle : 2007 Essence Cheveu…
1556  BMW_e46_318i MODÈLE : 2002 ESSENCE Cheveux : 1…
1557  Fiesta Modèl 2012 Essence Impôt : 350dh Cheveu…
1558  Modèl : 2007 Prix : 29000dh            …

[1559 rows x 7 columns]
```

[79]:
```
data[['capacity', 'engine_type']] = data['Engine'].str.split(' ', expand=True,␣
↪n=1)
data.head()
```

[79]:
```
                  Brand             Location    Price   Metrage Transmission  \
0    2022 Renault Captur      ad-Dar-al-Bayda      N/A   41713.0       Manual
```

```
1      2014 Dacia Dokker           Casablanca      N/A  192000.0       Manual
2      Leapmotor C10 530            Ain Sebaa  399 000       NaN    Automatic
3  Mercedes-Benz AMG GLE  al-Ayun Sidi Malluk      N/A       NaN       Manual
4           Renault Clio  al-Ayun Sidi Malluk   29 000       NaN    Automatic

           Engine                                        Descreption capacity  \
0  1,5 L (Diesel)    Renault Captur Manuelle 6 rapports     D…      1,5
1    N/A (Diesel)      Dacia docker                jama…      N/A
2        N/A (Ev)  La Leap Motor C10 530 Smart Edition est un véh…      N/A
3      N/A (Phev)  Mercedes Benz classe 200c tout option jamais a…      N/A
4    N/A (Petrol)                               …       N/A

  engine_type
0  L (Diesel)
1    (Diesel)
2        (Ev)
3      (Phev)
4    (Petrol)
```

[80]:
```
data.drop('Engine',axis=1,inplace=True)
data['capacity'] = data['capacity'].replace('N/A',np.nan)
```

[ ]:
```

```

[81]:
```
import re

data['engine_type'] = data['engine_type'].str.extract(r'\((.*?)\)',␣
 ↪expand=False)
```

[82]:
```
data.head()
```

[82]:
```
                  Brand             Location    Price   Metrage Transmission  \
0    2022 Renault Captur      ad-Dar-al-Bayda      N/A   41713.0       Manual
1      2014 Dacia Dokker           Casablanca      N/A  192000.0       Manual
2      Leapmotor C10 530            Ain Sebaa  399 000       NaN    Automatic
3  Mercedes-Benz AMG GLE  al-Ayun Sidi Malluk      N/A       NaN       Manual
4           Renault Clio  al-Ayun Sidi Malluk   29 000       NaN    Automatic

                                  Descreption capacity engine_type
0    Renault Captur Manuelle 6 rapports     D…      1,5      Diesel
1      Dacia docker                jama…      NaN      Diesel
2  La Leap Motor C10 530 Smart Edition est un véh…      NaN          Ev
3  Mercedes Benz classe 200c tout option jamais a…      NaN        Phev
4                            …       NaN      Petrol
```

:[83]
```
data['engine_type'].unique()
```

[83]:
```
array(['Diesel', 'Ev', 'Phev', 'Petrol', 'Hybrid'], dtype=object)
```

[84]:
```
data['brand'] = data['Brand'].str.split().str[0]
```

[85]:
```
data
```

[85]:

	Brand	Location	Price	Metrage	Transmission	Descreption	capacity	engine_type	brand
0	2022 Renault Captur	ad-Dar-al-Bayda	N/A	41713.0	Manual	Renault Captur Manuelle 6 rapports D…	1,5	Diesel	2022
1	2014 Dacia Dokker	Casablanca	N/A	192000.0	Manual	Dacia docker jama…	NaN	Diesel	2014
2	Leapmotor C10 530	Ain Sebaa	399 000	NaN	Automatic	La Leap Motor C10 530 Smart Edition est un véh…	NaN	Ev	Leapmotor
3	Mercedes-Benz AMG GLE	al-Ayun Sidi Malluk	N/A	NaN	Manual	Mercedes Benz classe 200c tout option jamais a…	NaN	Phev	Mercedes-Benz
4	Renault Clio	al-Ayun Sidi Malluk	29 000	NaN	Automatic	…	NaN	Petrol	Renault
…	…	…	…	…	…	…	…	…	…
1554	2009 Toyota Corolla	ad-Dar-al-Bayda	70 000	162000.0	Manual	Toyota Corolla Essence Modèle : 2009 Cheveux :…	NaN	Petrol	2009
1555	Volkswagen Golf	ad-Dar-al-Bayda	95 000	215000.0	Manual	Volkswagen Golf 5 Modèle : 2007 Essence Cheveu…	NaN	Petrol	Volkswagen
1556	2002 BMW 3-Series	ad-Dar-al-Bayda	53 000	260000.0	Manual	BMW_e46_318i MODÈLE : 2002 ESSENCE Cheveux : 1…	NaN	Petrol	2002
1557	Ford Fiesta	ad-Dar-al-Bayda	69 000	140000.0	Manual	Fiesta Modèl 2012 Essence Impôt : 350dh Cheveu…	NaN	Petrol	Ford
1558	2007 Daewoo Matiz	ad-Dar-al-Bayda	29 000	NaN	Manual	Modèl : 2007 Prix : 29000dh …	NaN	Petrol	2007

[1559 rows x 9 columns]

[86]:
```
data.describe(include='all')
```

[86]:

	Brand	Location	Price	Metrage	Transmission	\
count	1559	1559	1559	830.000000	1559	
unique	1079	32	316	NaN	2	
top	2020 Dacia Logan	ad-Dar-al-Bayda	N/A	NaN	Automatic	
freq	9	511	235	NaN	871	
mean	NaN	NaN	NaN	91630.746988	NaN	
std	NaN	NaN	NaN	46977.415485	NaN	
min	NaN	NaN	NaN	5000.000000	NaN	
25%	NaN	NaN	NaN	80000.000000	NaN	
50%	NaN	NaN	NaN	80000.000000	NaN	
75%	NaN	NaN	NaN	96000.000000	NaN	
max	NaN	NaN	NaN	600000.000000	NaN	

	Descreption	capacity	engine_type	brand
count	1559	28	1559	1559
unique	1420	13	5	62
top	0606064455	1,6	Petrol	2020
freq	21	5	1284	226
mean	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN

[87]:
```
import re
def extract_brand_model(desc):

    desc_cleaned = re.sub(r'\b\d{4}\b', '', desc).strip()


    words = desc_cleaned.split()


    if len(words) > 0:
        brand = words[0]
        model = ' '.join(words[1:])
    else:
        brand, model = None, None

    return brand, model
```

[88]:
```
data[['brand','model']] = data['Brand'].apply(lambda x : pd.
↪Series(extract_brand_model(x)))
```

```
data
```

[88]:
```
                        Brand             Location    Price   Metrage  \
0       2022 Renault Captur      ad-Dar-al-Bayda      N/A   41713.0
1         2014 Dacia Dokker           Casablanca      N/A  192000.0
2         Leapmotor C10 530            Ain Sebaa  399 000       NaN
3     Mercedes-Benz AMG GLE  al-Ayun Sidi Malluk      N/A       NaN
4              Renault Clio  al-Ayun Sidi Malluk   29 000       NaN
…                       …                   …      …       …
1554    2009 Toyota Corolla      ad-Dar-al-Bayda   70 000  162000.0
1555        Volkswagen Golf      ad-Dar-al-Bayda   95 000  215000.0
1556      2002 BMW 3-Series      ad-Dar-al-Bayda   53 000  260000.0
1557            Ford Fiesta      ad-Dar-al-Bayda   69 000  140000.0
1558      2007 Daewoo Matiz      ad-Dar-al-Bayda   29 000       NaN

     Transmission                                        Descreption capacity  \
0          Manual    Renault Captur Manuelle 6 rapports     D…      1,5
1          Manual      Dacia docker                jama…      NaN
2       Automatic  La Leap Motor C10 530 Smart Edition est un véh…      NaN
3          Manual  Mercedes Benz classe 200c tout option jamais a…      NaN
4       Automatic                                …        NaN
…            …                                                  …       …
1554       Manual  Toyota Corolla Essence Modèle : 2009 Cheveux :…      NaN
1555       Manual  Volkswagen Golf 5 Modèle : 2007 Essence Cheveu…      NaN
1556       Manual  BMW_e46_318i MODÈLE : 2002 ESSENCE Cheveux : 1…      NaN
1557       Manual  Fiesta Modèl 2012 Essence Impôt : 350dh Cheveu…      NaN
1558       Manual  Modèl : 2007 Prix : 29000dh            …      NaN

     engine_type          brand     model
0         Diesel        Renault    Captur
1         Diesel          Dacia    Dokker
2             Ev      Leapmotor   C10 530
3           Phev  Mercedes-Benz   AMG GLE
4         Petrol        Renault      Clio
…          …             …       …
1554      Petrol         Toyota   Corolla
1555      Petrol     Volkswagen      Golf
1556      Petrol            BMW  3-Series
1557      Petrol           Ford    Fiesta
1558      Petrol         Daewoo     Matiz

[1559 rows x 10 columns]
```

[89]:
```
data.drop('Descreption',axis=1,inplace=True)
```

[90]:
```
def extract_model_hjh(Brand):
    match = re.search(r'\d{4}',Brand)
```

```
    if match:
        return match.group(0)
    return None
```

[91]:
```
data['production_year'] = data['Brand'].apply(extract_model_hjh)
```

[92]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1559 entries, 0 to 1558
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Brand            1559 non-null   object
 1   Location         1559 non-null   object
 2   Price            1559 non-null   object
 3   Metrage          830 non-null    float64
 4   Transmission     1559 non-null   object
 5   capacity         28 non-null     object
 6   engine_type      1559 non-null   object
 7   brand            1559 non-null   object
 8   model            1559 non-null   object
 9   production_year  1418 non-null   object
dtypes: float64(1), object(9)
memory usage: 121.9+ KB
```

after cleaning the data wa can see that the col capacity has only 27 non null col we can drop the rows instead we well drop the entire col

[93]:
```
data.drop('capacity',axis=1,inplace=True)
```

[94]:
```
df =data.copy()
df.dropna(inplace=True)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 791 entries, 0 to 1556
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Brand            791 non-null    object
 1   Location         791 non-null    object
 2   Price            791 non-null    object
 3   Metrage          791 non-null    float64
 4   Transmission     791 non-null    object
 5   engine_type      791 non-null    object
 6   brand            791 non-null    object
 7   model            791 non-null    object
 8   production_year  791 non-null    object
```

```
dtypes: float64(1), object(8)
memory usage: 61.8+ KB
```

[95]:
```
df
```

[95]:
```
                     Brand         Location   Price   Metrage Transmission  \
0      2022 Renault Captur  ad-Dar-al-Bayda     N/A   41713.0       Manual
1        2014 Dacia Dokker       Casablanca     N/A  192000.0       Manual
15        2018 Peugeot 308  ad-Dar-al-Bayda     N/A  112789.0    Automatic
18    2009 Porsche Cayenne  ad-Dar-al-Bayda     N/A  116242.0    Automatic
19         2020 Toyota CHR  ad-Dar-al-Bayda     N/A   59641.0    Automatic
…                      …              …     …       …          …
1547   2010 Hyundai Tucson  ad-Dar-al-Bayda     N/A  230000.0       Manual
1548      2013 Dacia Logan  ad-Dar-al-Bayda  50 000  600000.0       Manual
1553   2009 Toyota Corolla  ad-Dar-al-Bayda  70 000  162000.0       Manual
1554   2009 Toyota Corolla  ad-Dar-al-Bayda  70 000  162000.0       Manual
1556     2002 BMW 3-Series  ad-Dar-al-Bayda  53 000  260000.0       Manual

     engine_type    brand     model production_year
0         Diesel  Renault    Captur            2022
1         Diesel    Dacia    Dokker            2014
15        Petrol  Peugeot       308            2018
18        Petrol  Porsche   Cayenne            2009
19        Hybrid   Toyota       CHR            2020
…           …       …        …              …
1547      Petrol  Hyundai    Tucson            2010
1548      Diesel    Dacia     Logan            2013
1553      Petrol   Toyota   Corolla            2009
1554      Petrol   Toyota   Corolla            2009
1556      Petrol      BMW  3-Series            2002

[791 rows x 9 columns]
```

[96]:
```
df['production_year'] = df['production_year'].astype(int)
```

[97]:
```
df['production_year']=df['production_year'].apply(lambda x : x if x < 2024 else␣
↪np.nan)
```

[98]:
```
df.dropna(inplace=True,axis=0)
```

[99]:
```
df.isna().sum()
```

[99]:
```
Brand              0
Location           0
Price              0
Metrage            0
Transmission       0
engine_type        0
```

```
brand              0
model              0
production_year    0
dtype: int64
```

[100]:
```
df.to_excel('cars_sales.xlsx')
```