

```
In [ ]: DATA DOSYASININ ASLINI ZİP OLARAK AYRICA SAKLA.  
JUPYTER'DE FILE MAKECOPY YAPARAK DENEMELERİ BU DOSYADA YAPIP ÇALIŞTIĞINI GÖRÜNCE DİĞER DOSYADA İLERLE.  
ÇALIŞMANIN NOTEBOOK ARA VERSİYONLARINI CHECKPOINT OLARAK KAYDET.  
DATA ÜZERİNDE OYNAMALARI ARA VERSİYONALAR HALİNDE KAYDET.  
SÜTUN ÜZERİNDEKİ HER TÜRLÜ İŞLEMİ ÖNCE FARKLI İSİMLİ YENİ BİR SÜTUNA ATA.  
NANLARI GÖREBİLMEK İÇİN VALUE_COUNTS DROPNA=False İLE YAPILMALI.    value_counts(dropna=False  
EXTRACT BULAMADIĞI DEĞERLERİ NAN YAPAR.  
NAN'IN TYPE'I FLOAT'TUR.
```

```
In [ ]: `\\d` Any numeric digit from `0` to `9`.  
`\\D` Matches any character which is not a decimal digit.  
      This is the opposite of `\\d`.  
`\\w` Any letter, numeric digit, or the underscore character.  
      (Think of this as matching "word" characters.)  
`\\W` Any character that is not a letter, numeric digit, or the underscore character.  
`\\s` Any space, tab, or newline character.  
      (Think of this as matching white-space characters.)  
`\\S` Any character that is not a space, tab, or newline.
```

```
In [ ]: "[" A set of characters "[a-m]"  
"\\" Signals a special sequence (can also be used to escape special characters)  
"." Any character (except newline character)  
"^" Starts with "^hello"  
"$" Ends with "world$"  
"*" Zero or more occurrences  
"+" One or more occurrences  
"{}" Exactly the specified number of occurrences  
"|" Either or "falls|stays"  
"()" Capture and group
```

```
In [ ]: # LİSTELERDE TÜM  '[' ile BAŞLAYAN BÖLÜMLERİ TOPLU SİLME:  
def clean_all_1(df):  
  
    dell_strings = lambda x: x.replace("[ '\n", "")  if isinstance(x, str) else x  
    return df.applymap(dell_strings)  
df=clean_all_1(df)
```

```
In [ ]: # LİSTELERDE TÜM  '\n]' İLE BİTEN BÖLÜMLERİ TOPLU SİLME:  
def clean_all_2(df):  
  
    dell_strings = lambda x: x.replace("\n']", "")  if isinstance(x, str) else x  
    return df.applymap(dell_strings)  
df=clean_all_2(df)
```

```
In [ ]: df.Fuel[13]    # Fuel sütununun 13.satırında hangi değer var?  'Benzine'
```

PART- 1 (Data Cleaning)

In [2]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

import warnings
warnings.filterwarnings("ignore")
warnings.warn("this will not show")

%matplotlib inline
# %matplotlib notebook

plt.rcParams["figure.figsize"] = (9,5)
# plt.rcParams['figure.dpi'] = 100

sns.set_style("whitegrid")
pd.set_option('display.float_format', lambda x: '%.2f' % x)

pd.options.display.max_rows = 100
pd.options.display.max_columns = 100
```

In [3]:

```
df = pd.read_json("scout_car.json", lines=True)
```

In [4]:

```
df.head(3).T
```

Out[4]:

| | 0 | 1 |
|--------------------------|--|---|
| url | https://www.autoscout24.com/offers/audi-a1-sp... | https://www.autoscout24.com/offers/audi-a1-1... |
| make_model | Audi A1 | Audi A1 |
| short_description | Sportback 1.4 TDI S-tronic Xenon Navi Klima | 1.8 TFSI sport |
| body_type | Sedans | Sedans |
| price | 15770 | 14500 |
| vat | VAT deductible | Price negotiable |
| km | 56,013 km | 80,000 km |
| registration | 01/2016 | 03/2017 |
| prev_owner | 2 previous owners | None |
| kW | NaN | NaN |
| hp | 66 kW | 141 kW |
| Type | [, Used, , Diesel (Particulate Filter)] | [, Used, , Gasoline] |
| Previous Owners | \n2\n | NaN |
| Next Inspection | \n06/2021\n, \n99 g CO2/km (comb)\n] | NaN |
| Inspection new | \nYes\n, \nEuro 6\n] | NaN |
| Warranty | \n, \n, \n4 (Green)\n] | NaN |
| Full Service | \n, \n] | NaN |

| | 0 | 1 | |
|--|--|---|--|
| Non-smoking Vehicle | [\\n, \\n] | NaN | |
| null | [] | [] | |
| Make | \\nAudi\\n | \\nAudi\\n | \\nA |
| Model | [\\n, A1, \\n] | [\\n, A1, \\n] | [\\n, A |
| Offer Number | [\\nLR-062483\\n] | NaN | [\\nAM-953 |
| First Registration | [\\n, 2016, \\n] | [\\n, 2017, \\n] | [\\n, 201 |
| Body Color | [\\n, Black, \\n] | [\\n, Red, \\n] | [\\n, Blac |
| Paint Type | [\\nMetallic\\n] | NaN | [\\nMetal |
| Body Color Original | [\\nMythosSchwarz\\n] | NaN | [\\nmythosSchwarz metal |
| Upholstery | [\\nCloth, Black\\n] | [\\nCloth, Grey\\n] | [\\nCloth, Bla |
| Body | [\\n, Sedans, \\n] | [\\n, Sedans, \\n] | [\\n, Sedan |
| Nr. of Doors | [\\n5\\n] | [\\n3\\n] | [\\n |
| Nr. of Seats | [\\n5\\n] | [\\n4\\n] | [\\n |
| Model Code | [\\n0588/BDF\\n] | [\\n0588/BCY\\n] | |
| Gearing Type | [\\n, Automatic, \\n] | [\\n, Automatic, \\n] | [\\n, Automati |
| Displacement | [\\n1,422 cc\\n] | [\\n1,798 cc\\n] | [\\n1,598 |
| Cylinders | [\\n3\\n] | [\\n4\\n] | |
| Weight | [\\n1,220 kg\\n] | [\\n1,255 kg\\n] | |
| Drive chain | [\\nfront\\n] | [\\nfront\\n] | [\\nfrro |
| Fuel | [\\n, Diesel (Particulate Filter), \\n] | [\\n, Gasoline, \\n] | [\\n, Diesel (Particulate Filter) |
| Consumption | [[3.8 l/100 km (comb)], [4.3 l/100 km (city)],...] | [[5.6 l/100 km (comb)], [7.1 l/100 km (city)],...] | [[3.8 l/100 km (comb)], [4.4 l/10 (ci |
| CO2 Emission | [\\n99 g CO2/km (comb)\\n] | [\\n129 g CO2/km (comb)\\n] | [\\n99 g CO2/km (com |
| Emission Class | [\\nEuro 6\\n] | [\\nEuro 6\\n] | [\\nEurc |
| \\nComfort & Convenience\\n | [Air conditioning, Armrest, Automatic climate ...] | [Air conditioning, Automatic climate control, ...] | [Air conditioning, Cruise control, Elec |
| \\nEntertainment & Media\\n | [Bluetooth, Hands-free equipment, On-board com...] | [Bluetooth, Hands-free equipment, On-board com...] | [MP3, On-board comp |
| \\nExtras\\n | [Alloy wheels, Catalytic Converter, Voice Cont...] | [Alloy wheels, Sport seats, Sport suspension, ...] | [Alloy wheels, Voice Con |
| \\nSafety & Security\\n | [ABS, Central door lock, Daytime running light...] | [ABS, Central door lock, Central door lock wit...] | [ABS, Central door lock, Daytime runn li |
| description | [\\n, Sicherheit:, , Deaktivierung f\\u00fcr Beifahr...] | [\\nLangstreckenfahrzeug daher die hohe Kilomet...] | [\\n, Fahrzeug-Nummer: AM-95365, , E UF |
| Emission Label | NaN | [\\n4 (Green)\\n] | [\\n4 (Gre |
| Gears | NaN | [\\n7\\n] | |
| Country version | NaN | NaN | |
| Electricity consumption | NaN | NaN | |
| Last Service Date | NaN | NaN | |
| Other Fuel Types | NaN | NaN | |
| Availability | NaN | NaN | |
| Last Timing Belt Service Date | NaN | NaN | |
| Available from | NaN | NaN | |

In [5]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15919 entries, 0 to 15918
Data columns (total 54 columns):
 #   Column           Non-Null Count Dtype  
 ---  -- 
 0   url              15919 non-null  object  
 1   make_model        15919 non-null  object  
 2   short_description 15873 non-null  object  
 3   body_type         15859 non-null  object  
 4   price             15919 non-null  int64  
 5   vat               11406 non-null  object  
 6   km                15919 non-null  object  
 7   registration      15919 non-null  object  
 8   prev_owner        9091 non-null  object  
 9   kW                0 non-null     float64 
 10  hp                15919 non-null  object  
 11  Type              15917 non-null  object  
 12  Previous Owners  9279 non-null  object  
 13  Next Inspection   3535 non-null  object  
 14  Inspection new    3932 non-null  object  
 15  Warranty          10499 non-null object  
 16  Full Service      8215 non-null  object  
 17  Non-smoking Vehicle 7177 non-null object  
 18  null              15919 non-null  object  
 19  Make              15919 non-null  object  
 20  Model             15919 non-null  object  
 21  Offer Number      12744 non-null object  
 22  First Registration 14322 non-null object  
 23  Body Color         15322 non-null object  
 24  Paint Type         10147 non-null object  
 25  Body Color Original 12160 non-null object  
 26  Upholstery        12199 non-null object  
 27  Body               15859 non-null object  
 28  Nr. of Doors       15707 non-null object  
 29  Nr. of Seats        14942 non-null object  
 30  Model Code         4978 non-null  object  
 31  Gearing Type       15919 non-null object  
 32  Displacement       15423 non-null object  
 33  Cylinders          10239 non-null object  
 34  Weight             8945 non-null  object  
 35  Drive chain        9061 non-null  object  
 36  Fuel               15919 non-null object  
 37  Consumption         14013 non-null object  
 38  CO2 Emission        14111 non-null object  
 39  Emission Class      12898 non-null object  
 40
Comfort & Convenience
 14999 non-null  object
41
Entertainment & Media
 14545 non-null  object
42
Extras
 12957 non-null  object
43
Safety & Security
 14937 non-null  object
44  description         15919 non-null  object  
45  Emission Label      3985 non-null  object  
46  Gears               11207 non-null  object  
47  Country version     7586 non-null  object  
48  Electricity consumption 137 non-null object  
49  Last Service Date   566 non-null   object  
50  Other Fuel Types    880 non-null   object  
51  Availability         635 non-null   object  
52  Last Timing Belt Service Date 16 non-null object  
53  Available from       272 non-null   object  
dtypes: float64(1), int64(1), object(52)
memory usage: 6.6+ MB
```

In [6]:

df.columns

```
Out[6]: Index(['url', 'make_model', 'short_description', 'body_type', 'price', 'vat',
 'km', 'registration', 'prev_owner', 'kW', 'hp', 'Type',
 'Previous Owners', 'Next Inspection', 'Inspection new', 'Warranty',
 'Full Service', 'Non-smoking Vehicle', 'null', 'Make', 'Model',
 'Offer Number', 'First Registration', 'Body Color', 'Paint Type',
 'Body Color Original', 'Upholstery', 'Body', 'Nr. of Doors',
 'Nr. of Seats', 'Model Code', 'Gearing Type', 'Displacement',
 'Cylinders', 'Weight', 'Drive chain', 'Fuel', 'Consumption',
 'CO2 Emission', 'Emission Class', '\nComfort & Convenience\n',
```

```
'\nEntertainment & Media\n', '\nExtras\n', '\nSafety & Security\n',
'description', 'Emission Label', 'Gears', 'Country version',
'Electricity consumption', 'Last Service Date', 'Other Fuel Types',
'Availability', 'Last Timing Belt Service Date', 'Available from'],
dtype='object')
```

DF.COLUMNS= [...] İLE SÜTUN SIRASINI DEĞİŞTİREBİLİR, SÜTUNLARI ELEYEBİLİRİZ.

In [7]: # df.columns=['make_model', 'short_description', ... 'body_type', 'price']

In [8]: df.select_dtypes(include='object').describe().T
İŞLEM YAPILABİLMESİ İÇİN ÖNCELİKLE KÖŞELİ PARANTEZ İÇİNDEKİ İFADELERİN LİSTE DIŞINA ÇIKARILMASI GEREKECEK.

| | count | unique | | top | freq |
|----------------------------|-------|--------|---|---------------------------------------|-------|
| url | 15919 | 15919 | https://www.autoscout24.com//offers/audi-a1-sp... | 1 | |
| make_model | 15919 | 9 | | Audi A3 | 3097 |
| short_description | 15873 | 10001 | SPB 1.6 TDI 116 CV S tronic Sport | 64 | |
| body_type | 15859 | 9 | | Sedans | 7903 |
| vat | 11406 | 2 | | VAT deductible | 10980 |
| km | 15919 | 6690 | | 10 km | 1045 |
| registration | 15919 | 48 | | -/- | 1597 |
| prev_owner | 9091 | 4 | | 1 previous owner | 8294 |
| hp | 15919 | 81 | | 85 kW | 2542 |
| Type | 15917 | 169 | [, Used, , Diesel (Particulate Filter)] | 3475 | |
| Previous Owners | 9279 | 102 | | \n1\n | 8101 |
| Next Inspection | 3535 | 1399 | | \n04/2022\n | 62 |
| Inspection new | 3932 | 202 | | [\nYes\n, \nEuro 6\n] | 523 |
| Warranty | 10499 | 515 | | [\n, \n, \nEuro 6\n] | 1868 |
| Full Service | 8215 | 122 | | [\n, \n, \n4 (Green)\n] | 2235 |
| Non-smoking Vehicle | 7177 | 94 | | [\n, \n] | 3647 |
| null | 15919 | 1 | | [] | 15919 |
| Make | 15919 | 3 | | \nOpel\n | 7343 |
| Model | 15919 | 9 | | [\n, A3, \n] | 3097 |
| Offer Number | 12744 | 11440 | | [\nTP62881\n] | 27 |
| First Registration | 14322 | 4 | | [\n, 2018, \n] | 4522 |
| Body Color | 15322 | 14 | | [\n, Black, \n] | 3745 |
| Paint Type | 10147 | 3 | | [\nMetallic\n] | 9794 |
| Body Color Original | 12160 | 1927 | | [\nOnyx Schwarz\n] | 338 |
| Upholstery | 12199 | 46 | | [\nCloth, Black\n] | 5821 |
| Body | 15859 | 9 | | [\n, Sedans, \n] | 7903 |
| Nr. of Doors | 15707 | 6 | | [\n5\n] | 11575 |
| Nr. of Seats | 14942 | 6 | | [\n5\n] | 13336 |
| Model Code | 4978 | 232 | | [\n0035/BCB\n] | 268 |
| Gearing Type | 15919 | 3 | | [\n, Manual, \n] | 8153 |
| Displacement | 15423 | 77 | | [\n1,598 cc\n] | 4761 |
| Cylinders | 10239 | 7 | | [\n4\n] | 8105 |
| Weight | 8945 | 434 | | [\n1,163 kg\n] | 574 |
| Drive chain | 9061 | 3 | | [\nfront\n] | 8886 |
| Fuel | 15919 | 77 | | [\n, Diesel (Particulate Filter), \n] | 4315 |

| | count | unique | | top | freq |
|--|-------|--------|---|-------|------|
| Consumption | 14013 | 881 | [[3.9 l/100 km (comb)], [4.1 l/100 km (city)], ...] | 304 | |
| CO2 Emission | 14111 | 123 | [\\n120 g CO2/km (comb)\\n] | 740 | |
| Emission Class | 12898 | 7 | [\\nEuro 6\\n] | 10139 | |
| \\nComfort & Convenience\\n | 14999 | 6198 | [Air conditioning, Electrical side mirrors, Hi... | 216 | |
| \\nEntertainment & Media\\n | 14545 | 346 | [Bluetooth, Hands-free equipment, On-board com... | 1282 | |
| \\nExtras\\n | 12957 | 659 | [Alloy wheels] | 3245 | |
| \\nSafety & Security\\n | 14937 | 4443 | [ABS, Central door lock, Daytime running light...] | 538 | |
| description | 15919 | 12874 | [] | 636 | |
| Emission Label | 3985 | 6 | [\\n4 (Green)\\n] | 3553 | |
| Gears | 11207 | 10 | [\\n6\\n] | 5822 | |
| Country version | 7586 | 23 | [\\nGermany\\n] | 4502 | |
| Electricity consumption | 137 | 1 | [\\n0 kWh/100 km (comb)\\n] | 137 | |
| Last Service Date | 566 | 266 | [\\n02/2019\\n, \\nEuro 6\\n] | 23 | |
| Other Fuel Types | 880 | 1 | [[], [], []] | 880 | |
| Availability | 635 | 15 | [\\nin 90 days from ordering\\n] | 196 | |
| Last Timing Belt Service Date | 16 | 15 | [\\n04/2016\\n] | 2 | |
| Available from | 272 | 46 | [\\n31/08/19\\n] | 98 | |

İSTENİLEN SÜTUNLARI GÖRÜNTÜLEME

In [9]: df.iloc[:,10:20] # DF'de indeks sırası 10-20 arasında olan sütunları getirir.

Out[9]:

| | hp | Type | Previous Owners | Next Inspection | Inspection new | Warranty | Full Service | Non-smoking Vehicle | null | Make |
|-------|--------|--|---|--|--|--------------------------------------|--------------------------------------|-----------------------------|------|---------------|
| 0 | 66 kW | [, Used, , Diesel (Particulate Filter)] | \\n2\\n | [\\n06/2021\\n, \\n99 g CO2/km (comb)\\n] | [\\nYes\\n, \\nEuro 6\\n] | [\\n, \\n, \\n4 (Green)\\n] | [\\n, \\n] | [\\n, \\n] | [] | \\nAudi\\n |
| 1 | 141 kW | [, Used, , Gasoline] | NaN | NaN | NaN | NaN | NaN | NaN | [] | \\nAudi\\n |
| 2 | 85 kW | [, Used, , Diesel (Particulate Filter)] | \\n1\\n | NaN | NaN | [\\n, \\n, \\n99 g CO2/km (comb)\\n] | NaN | NaN | [] | \\nAudi\\n |
| 3 | 66 kW | [, Used, , Diesel (Particulate Filter)] | \\n1\\n | NaN | NaN | NaN | [\\n, \\n, \\n99 g CO2/km (comb)\\n] | [\\n, \\n, \\nEuro 6\\n] | [] | \\nAudi\\n |
| 4 | 66 kW | [, Used, , Diesel (Particulate Filter)] | \\n1\\n | NaN | [\\nYes\\n, \\n109 g CO2/km (comb)\\n] | [\\n, \\n, \\nEuro 6\\n] | [\\n, \\n, \\n4 (Green)\\n] | [\\n, \\n] | [] | \\nAudi\\n |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 15914 | 147 kW | [, New, , Diesel (Particulate Filter)] | NaN | NaN | NaN | [\\n24 months\\n] | NaN | NaN | [] | \\nRenault\\n |
| 15915 | 165 kW | [, Used, , Super 95 / Super Plus 98 (Particula...] | [\\n1\\n, \\n, 7.4 l/100 km (comb), \\n, 9.2 l/100... | [\\n01/2022\\n, \\n168 g CO2/km (comb)\\n] | NaN | [\\n, \\n, \\nEuro 6\\n] | NaN | [\\n, \\n, \\n4 (Green)\\n] | [] | \\nRenault\\n |
| 15916 | 146 kW | [, Pre-registered, , Diesel] | [\\n1\\n, \\n139 g CO2/km (comb)\\n] | NaN | [\\nYes\\n, \\nEuro 6d-TEMP\\n] | [\\n, \\n, \\n4 (Green)\\n] | [\\n, \\n] | [\\n, \\n] | [] | \\nRenault\\n |
| 15917 | 147 kW | [, Pre-registered, , Diesel] | NaN | NaN | NaN | [\\n] | NaN | NaN | [] | \\nRenault\\n |

| hp | Type | Previous Owners | Next Inspection | Inspection new | Warranty | Full Service | Non-smoking Vehicle | null | Make |
|-------|-------------------------------------|-----------------|-----------------|--|----------|--------------|---------------------|------|---------------|
| 15918 | 165 kW [Demonstration, , Super 95] | \n1\n | NaN | [\\nYes\\n, \\n153 g CO2/km (comb)\\n] | NaN | NaN | NaN | [] | \\nRenault\\n |

15919 rows × 10 columns

```
In [10]: # ÖNCELİKLE SÜTUN İSİMLERİ DÜZELTİLDİ
df["Comfort_Convenience"] = df["\nComfort & Convenience\n"]
df["Entertainment_Media"] = df["\nEntertainment & Media\n"]
df["Extras"] = df["\nExtras\n"]
df["Safety_Security"] = df["\nSafety & Security\n"]

In [11]: drop_columns = ["\nComfort & Convenience\n", "\nEntertainment & Media\n", "\nExtras\n", "\nSafety & Security\n"]
df.drop(drop_columns, axis = 1, inplace = True)

In [12]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15919 entries, 0 to 15918
Data columns (total 54 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   url              15919 non-null   object  
 1   make_model        15919 non-null   object  
 2   short_description 15873 non-null   object  
 3   body_type         15859 non-null   object  
 4   price             15919 non-null   int64  
 5   vat               11406 non-null   object  
 6   km                15919 non-null   object  
 7   registration      15919 non-null   object  
 8   prev_owner        9091 non-null   object  
 9   kW                0 non-null      float64 
 10  hp               15919 non-null   object  
 11  Type              15917 non-null   object  
 12  Previous Owners  9279 non-null   object  
 13  Next Inspection   3535 non-null   object  
 14  Inspection new    3932 non-null   object  
 15  Warranty           10499 non-null  object  
 16  Full Service       8215 non-null   object  
 17  Non-smoking Vehicle 7177 non-null  object  
 18  null              15919 non-null   object  
 19  Make              15919 non-null   object  
 20  Model              15919 non-null   object  
 21  Offer Number       12744 non-null  object  
 22  First Registration 14322 non-null  object  
 23  Body Color          15322 non-null  object  
 24  Paint Type          10147 non-null  object  
 25  Body Color Original 12160 non-null  object  
 26  Upholstery          12199 non-null  object  
 27  Body                15859 non-null  object  
 28  Nr. of Doors        15707 non-null  object  
 29  Nr. of Seats         14942 non-null  object  
 30  Model Code          4978 non-null   object  
 31  Gearing Type         15919 non-null  object  
 32  Displacement          15423 non-null  object  
 33  Cylinders            10239 non-null  object  
 34  Weight              8945 non-null   object  
 35  Drive chain          9061 non-null   object  
 36  Fuel                 15919 non-null  object  
 37  Consumption           14013 non-null  object  
 38  CO2 Emission          14111 non-null  object  
 39  Emission Class        12898 non-null  object  
 40  description           15919 non-null  object  
 41  Emission Label        3985 non-null   object  
 42  Gears                 11207 non-null  object  
 43  Country version       7586 non-null   object  
 44  Electricity consumption 137 non-null   object  
 45  Last Service Date     566 non-null   object  
 46  Other Fuel Types       880 non-null   object  
 47  Availability           635 non-null   object  
 48  Last Timing Belt Service Date 16 non-null   object  
 49  Available from         272 non-null   object  
 50  Comfort_Convenience   14999 non-null  object
```

```
51 Entertainment_Media           14545 non-null object
52 Extras                      12957 non-null object
53 Safety_Security              14937 non-null object
dtypes: float64(1), int64(1), object(52)
memory usage: 6.6+ MB
```

```
In [13]: df.shape
```

```
Out[13]: (15919, 54)
```

SÜTUNLARIN NAN DEĞERLERİNİN YÜZDESİNİ BULMA

```
In [14]: df.isnull().sum()*100/df.shape[0]
```

```
Out[14]: url                  0.00
make_model            0.00
short_description      0.29
body_type              0.38
price                  0.00
vat                     28.35
km                     0.00
registration            0.00
prev_owner              42.89
kW                     100.00
hp                     0.00
Type                   0.01
Previous Owners        41.71
Next Inspection         77.79
Inspection new          75.30
Warranty                34.05
Full Service             48.39
Non-smoking Vehicle     54.92
null                    0.00
Make                   0.00
Model                  0.00
Offer Number            19.94
First Registration       10.03
Body Color               3.75
Paint Type              36.26
Body Color Original      23.61
Upholstery              23.37
Body                   0.38
Nr. of Doors             1.33
Nr. of Seats              6.14
Model Code              68.73
Gearing Type             0.00
Displacement             3.12
Cylinders                35.68
Weight                  43.81
Drive chain              43.08
Fuel                     0.00
Consumption              11.97
CO2 Emission              11.36
Emission Class            18.98
description              0.00
Emission Label            74.97
Gears                   29.60
Country version            52.35
Electricity consumption      99.14
Last Service Date          96.44
Other Fuel Types            94.47
Availability                96.01
Last Timing Belt Service Date 99.90
Available from              98.29
Comfort_Convenience          5.78
Entertainment_Media          8.63
Extras                     18.61
Safety_Security              6.17
dtype: float64
```

Dropping columns that have %90 percent and higher of missing values.

```
In [15]: def show_nans(df, limit):
    missing = df.isnull().sum()*100/df.shape[0]
    return missing.loc[lambda x : x >= limit]

def perc_nans(serial):
```

```
# display percentage of nans in a Series
return serial.isnull().sum()/serial.shape[0]*100
```

```
In [16]: show_nans(df,90)
```

```
Out[16]: kW          100.00
Electricity consumption    99.14
Last Service Date        96.44
Other Fuel Types         94.47
Availability             96.01
Last Timing Belt Service Date 99.90
Available from           98.29
dtype: float64
```

```
In [17]: drop_columns = show_nans(df,90).index
drop_columns
```

```
Out[17]: Index(['kW', 'Electricity consumption', 'Last Service Date',
       'Other Fuel Types', 'Availability', 'Last Timing Belt Service Date',
       'Available from'],
      dtype='object')
```

```
In [18]: df.drop(drop_columns, axis = 1, inplace = True)
```

```
In [19]: df.drop("null", axis = 1, inplace = True)
```

```
In [20]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15919 entries, 0 to 15918
Data columns (total 46 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   url               15919 non-null   object 
 1   make_model        15919 non-null   object 
 2   short_description 15873 non-null   object 
 3   body_type          15859 non-null   object 
 4   price              15919 non-null   int64  
 5   vat                11406 non-null   object 
 6   km                 15919 non-null   object 
 7   registration       15919 non-null   object 
 8   prev_owner         9091 non-null   object 
 9   hp                 15919 non-null   object 
 10  Type               15917 non-null   object 
 11  Previous Owners   9279 non-null   object 
 12  Next Inspection    3535 non-null   object 
 13  Inspection new     3932 non-null   object 
 14  Warranty            10499 non-null  object 
 15  Full Service       8215 non-null   object 
 16  Non-smoking Vehicle 7177 non-null   object 
 17  Make               15919 non-null   object 
 18  Model               15919 non-null   object 
 19  Offer Number        12744 non-null  object 
 20  First Registration 14322 non-null   object 
 21  Body Color          15322 non-null   object 
 22  Paint Type          10147 non-null   object 
 23  Body Color Original 12160 non-null  object 
 24  Upholstery          12199 non-null   object 
 25  Body                15859 non-null   object 
 26  Nr. of Doors        15707 non-null   object 
 27  Nr. of Seats        14942 non-null   object 
 28  Model Code          4978 non-null   object 
 29  Gearing Type        15919 non-null   object 
 30  Displacement        15423 non-null   object 
 31  Cylinders           10239 non-null   object 
 32  Weight              8945 non-null   object 
 33  Drive chain         9061 non-null   object 
 34  Fuel                15919 non-null   object 
 35  Consumption          14013 non-null   object 
 36  CO2 Emission         14111 non-null   object 
 37  Emission Class       12898 non-null   object 
 38  description          15919 non-null   object 
 39  Emission Label       3985 non-null   object 
 40  Gears                11207 non-null   object 
 41  Country version      7586 non-null   object 
 42  Comfort_Convenience 14999 non-null   object 
 43  Entertainment_Media 14545 non-null   object
```

```
44 Extras           12957 non-null object
45 Safety_Security 14937 non-null object
dtypes: int64(1), object(45)
memory usage: 5.6+ MB
```

Let's examine and clean all the columns/features one by one

url

```
In [21]: df.url.value_counts(dropna=False)
```

```
Out[21]: https://www.autoscout24.com//offers/audi-a1-sportback-1-4-tdi-adrenalin-90-diesel-silver-5a63129e-ed55-4e80-b05e-3aada2cd17e9          1
https://www.autoscout24.com//offers/renault-espace-initiale-paris-dci-160-cv-4-control-techo-diesel-grey-d8295716-b020-474f-a642-139735abbefb 1
https://www.autoscout24.com//offers/renault-clio-intens-dci-90-edc-wltp-diesel-blue-9dcde57e-488c-45d8-9c3c-984938ce8787      1
https://www.autoscout24.com//offers/renault-clio-iv-1-2-tce-120-intens-energy-klima-navi-gasoline-blue-ad49d196-2e24-4061-81bf-4d30f9bf1e01 1
https://www.autoscout24.com//offers/audi-a3-spb-1-6tdi-116cv-sport-navi-xenon-pdc-manuale-diesel-black-be4e0c6a-8235-4831-82cd-a087cbc4ecc6 1
..
https://www.autoscout24.com//offers/audi-a3-sportback-30-tdi-design-led-navi-neu-diesel-black-cf73aa7f-60af-4a84-bec8-18bb04578126      1
https://www.autoscout24.com//offers/renault-clio-dci-75-business-energy-eco2-euro6-55-kw-75-c-diesel-white-61d03ccf-50f7-4025-a13f-9155f7f127b4 1
https://www.autoscout24.com//offers/audi-a3-sportback-1-6-tdi-sport-business-plus-pack-diesel-white-8f5c6bf9-ced1-41f5-b7a4-35efdb9ae026 1
https://www.autoscout24.com//offers/audi-a3-sportback-tfsi-150-s-tronic-klima-licht-regen-gasoline-white-a9193b1d-c4ff-4220-a8b7-ca32a222511a 1
https://www.autoscout24.com//offers/opel-insignia-selective-automatico-1-6-cdti-136cv-diesel-black-f7b91903-9868-45a4-97e0-348a495707b2 1
Name: url, Length: 15919, dtype: int64
```

```
In [22]: df.drop("url", axis = 1, inplace = True)
```

make_model

```
In [23]: df.make_model.value_counts(dropna=False)
```

```
Out[23]: Audi A3        3097
Audi A1        2614
Opel Insignia  2598
Opel Astra     2526
Opel Corsa     2219
Renault Clio    1839
Renault Espace   991
Renault Duster   34
Audi A2         1
Name: make_model, dtype: int64
```

Make

```
In [24]: df.Make.value_counts(dropna=False)
```

```
Out[24]: \nOpel\n      7343
\nAudi\n      5712
\nRenault\n    2864
Name: Make, dtype: int64
```

```
In [25]: df["Make"] = df.Make.str.strip('\n')
```

```
In [26]: df.Make.value_counts(dropna=False)
```

```
Out[26]: Opel      7343
Audi      5712
Renault   2864
Name: Make, dtype: int64
```

```
In [27]: df.drop("Make", axis=1, inplace=True)
```

Model

In [28]:

```
df.Model.value_counts(dropna=False)
```

```
-----  
TypeError                                                 Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

Out[28]:

```
[\n, A3, \n]      3097  
[\n, A1, \n]      2614  
[\n, Insignia, \n] 2598  
[\n, Astra, \n]   2526  
[\n, Corsa, \n]   2219  
[\n, Clio, \n]    1839  
[\n, Espace, \n]   991  
[\n, Duster, \n]   34  
[\n, A2, \n]        1  
Name: Model, dtype: int64
```

LİSTE HALİNDEKİ SÜTUNDAN İNDEKS SIRASIYLA VERİ ÇEKME

In [29]:

```
df["Model"] = df.Model.str[1]      # 1. İNDEKSTEKİ CLIO'YU DEĞER OLARAK ATAR [\n, Clio, \n]
```

In [30]:

```
df.Model.value_counts(dropna=False)
```

Out[30]:

```
A3      3097  
A1      2614  
Insignia 2598  
Astra   2526  
Corsa   2219  
Clio    1839  
Espace  991  
Duster  34  
A2      1  
Name: Model, dtype: int64
```

In [31]:

```
df.drop("Model", axis=1, inplace=True)
```

short_description

In [32]:

```
df.short_description.value_counts(dropna=False)
```

Out[32]:

```
SPB 1.6 TDI 116 CV S tronic Sport          64  
NaN                                         46  
1.4 66kW (90CV) Selective                 40  
MOVE KLIMA CD USB ALLWETTER BLUETOOTH     38  
SPB 30 TFSI S tronic Admired                35  
..  
Sportback Ambition, 1.6 TDI, 6-Gang, S-Line Exter 1  
Limousine Sport 1.0 TFSI 7-G.Autom.*AHK*Navi*  1  
1.4 TDI Sport PDC 16 MEDIA-PAKET           1  
GS [A](B) Dynamic 1.5 Turbo Komfortpaket Allwetter 1  
1,6 CDTi ST Cosmo NAVI+KLIMAAUTOMAT+LM+PDC    1  
Name: short_description, Length: 10002, dtype: int64
```

In [33]:

```
df.drop("short_description", axis = 1, inplace = True)
```

description

In [34]:

```
df.description.value_counts(dropna=False)
```

```
TypeError Traceback (most recent call last)
pandas\_libs\hashtable\_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()
```

```
TypeError: unhashable type: 'list'
```

```
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
```

```
Traceback (most recent call last):
```

```
  File "pandas\_libs\hashtable\_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
```

```
TypeError: unhashable type: 'list'
```

```
Out[34]: []
```

```
636
```

```
[\n, \n]
```

```
30
```

```
[\nSichern Sie sich diesen Opel Astra K 1.4 SIDI Turbo Innovation Start/Stop zum Top-Preis inklusive dem, AUTOHE RO-RUNDUM-SORGLOS-PAKET, , Infos:, 2.Hand, deutsche Auslieferung, Unfallfrei, scheckheftgepflegt, mindestens 1 2 Monate gültige HU und AU, , Highlights, , Rückfahrkamera, , Voll-LED-Matrix-Scheinwerfer (IntelliLux), , A udio-Navigationssystem Navi 900 Europa Touch, , Spurassistent, , Leichtmetallfelgen, Komfort, , Sitzheizung vorn, , Klimaanomatik 2-Zonen, , Parkpilotensystem vorn und hinten, , 6-Stufen Automatikgetriebe, , Tempomat (Geschwindigkeits-Regelanlage) mit Begrenzer, , Berganfahrassistent, , Servolenkung, , Lendenwirbelstütze e lektropneumatisch für Fahrer, , Kopfstützen vorn und hinten, , Multifunktionslenkrad in Leder, beheizbar, , Mittelarmlehne vorn verschiebbar, , Zentralverriegelung mit Fernbedienung, , Fensterheber elektrisch vorn + hinten, , Berganfahr-Assistent (HSA, Hill Start Assist), Multimedia, , 8 Zoll Touchscreen-Farbdisplay, , Verk ehrszeichenerkennung, , Audiostreaming, , Apple CarPlay und Android Auto, , 6 Lautsprecher, , Bluetooth-Fr eisprecheinrichtung, , Sprachsteuerung, , Bordcomputer, , Multimedia-Schnittstelle (USB / AUX-IN), Licht und Sicht, , Außenspiegel elektrisch verstell-, heiz- und anklappbar, , Abbiegelicht, , LED Tagfahrlicht, , Nebe lscheinwerfer, , Ambientebeleuchtung, , Innenrückspiegel automatisch abblendend, , Fernlichtassistent und Leu chtweitenregulierung, , Licht und Regensor, , Wärmeschutzverglasung hinten abgedunkelt (Solar-Protect), , LED Heckleuchten, Sicherheit, , Abstandswarner, , Auffahrwarnsystem, , Isofix-Aufnahmen für Kindersitz an Rücksitz, , Frontkamera, , Airbag für Fahrer- und Beifahrerseite, , Wegfahrsperre, , ...]
```

```
27
```

```
[\n, Fahrzeugbeschreibung, Sichern Sie sich diesen Renault Clio 1.6 TCE Energy Sport Trophy zum Top-Preis inklusive dem, AUTOHERO-RUNDUM-SORGLOS-PAKET, , Infos:, 1.Hand, deutsche Auslieferung, Unfallfrei, scheckheftgepflegt, , Highlights, , Audio-Navigationssystem R-Link Evolution mit Touchscreen und Carminat TomTom, , Sportsitze, , Sportfahrwerk, , Scheinwerfer Full-LED, Komfort, , Geschwindigkeits-Regelanlage (Tempomat), , Klimaanlage, , Einparkhilfe hinten, , Multifunktionslenkrad (Sport/Leder) mit Schaltwippen/-tasten, , Fensterheber elektr. mit Impulsgeber links/rechts vorn, , Zentralverriegelung / Startanlage Handsfree Entry & Drive, , Berganfahr-Assistent (HSA), Multimedia, , Audiosystem: Radio R & GO (Bluetooth, DAB-Tuner), , Bordcomputer, , Freisprecheinrichtung Bluetooth, , USB-Anschluss inkl. AUX-IN-Anschluss (Plug & Music), Licht und Sicht, , Außenspiegel elektr. verstell- und heizbar, , Licht- und Regensor, , Tagfahrlicht LED, , Nebelscheinwerfer LED, , Coming-Home-Lichtfunktion, , Heckleuchten LED, Sicherheit, , Isofix-Aufnahmen für Kindersitz, , Seitenairbag vorn, , Wegfahrsperre, , Differentialsperre, , Bremsassistent, , Elektron. Stabilitäts-Pro gramm (ESP), , Anti-Blockier-System (ABS), , Anti-Submarining-Airbag im Fond, , Airbag Fahrer-/Beifahrerseite, , Differentialsperre, , Reifendruck-Kontrollsyste, , Untersteuerungskontrolle (USC), , Antischlupfregelung (ASR), Weiteres, , Start/Stop-Anlage, , Servolenkung geschwindigkeitsabhängig, , R.S. Drive (Fahrmodusschalter), , Gepäckraumabdeckung / Rollo, , Heckdiffusor, , LM-Felgen, , Hecksspoiler (Renault Sport), , Pedale Aluminium, , Energierückgewinnung (Energy Smart Management), ...]
```

```
27
```

```
[\n, Fahrzeugbeschreibung, Sichern Sie sich diesen Audi A1 1.4 TDI Design zum Top-Preis inklusive dem, AUTOHERO -RUNDUM-SORGLOS-PAKET, , Infos:, 2. Hand, deutsche Auslieferung, Unfallfrei, scheckheftgepflegt, , Highlights, , Leichtmetallfelgen, , Sitzheizung vorn, , Sportsitze vorn, , Dynamik-Fahrwerk, , Multi-Media-Interface MMI Navigation, , Einparkhilfe hinten (APS), , CO2-optimiert (Ausführung e, Ultra), , Komfort, , Klimaanomatik, , Mittelarmlehne vorn mit Fach, , Fensterheber elektrisch vorn, , Zentralverriegelung mit Fernbedienung, , Berganfahr-Assistent, , Multifunktionslederlenkrad, , Rücksitzlehne geteilt/klappbar, , Sitze vorn höhe nverstellbar, , Lendenwirbelstützen verstellbar, , Multimedia, , Audi music interface, , Fahrer-Informatio ns-System (FIS), , Bluetooth-Freisprecheinrichtung, , Licht und Sicht, , Außenspiegel elektrisch verstell- un d heizbar, beide, , Tagfahrlicht, , Scheibenwaschdüsen heizbar, , Licht- und Regensor, , Nebelscheinwerfer integriert, , Wärmeschutzverglasung grün getönt, , Heckleuchten LED, , Sicherheit, , Isofix-Aufnahmen für Kindersitz, , Bremsassistent, , Seitenairbag vorn, , Reifenkontroll-Anzeige, , Anti-Blockier-System (ABS), , Wegfahrsperre (elektronisch), , Airbag Fahrer-/Beifahrerseite, , Antriebs-Schlupfregelung (ASR), , Kopf-Air bag-System (Sideguard), , Elektronische Differentialsperre (EDS), , Elektronisches Stabilitäts-Programm (ESP), , Weiteres, , Start/Stop-Anlage, , Wagenheber, , Servolenkung, , ...]
```

```
27
```

```
...
```

```
[\n, Ausstattungspakete:, Ablagepaket, Assistenzsysteme:, Sprachbedienung, Einparkhilfe hinten, Licht und Sicht:, Tagfahrlicht, Audio & Kommunikation:, Multi Media Interface MMI, Audio-Schnittstelle: Audi music interface, Räder und Reifen:, Leichtmetallfelgen: 16 Zoll Aluminium Gussräder, Ganzjahresreifen, Komfort:, Climatroni c, Fahrer-/Beifahrersitz höhenverstellbar, Lordosenstütze Fahrer/Beifahrer, Sitzheizung Fahrer/Beifahrer, Easy-E ntry Einstiegshilfe, Armauflage Fahrer/Beifahrer, Servotronic, Zentralver. mit Fernbedienung, Interieur:, Sitze Stoff schwarz-schwarz-felsgrau, Innenfarbe schwarz, Multifunktions-Sport-/Lederlenkrad, Rücksitze klappbar, Exte rieur:, Sportfahrwerk, Außenspiegel elekt. und beheizt, Start-Stopp System, Sicherheit:, 6 x Airbag, Antriebss chlupfregelung ASR, Reifendruckkontrolle, Weiteres:, 3 Jahre Neuwagenanschlussgarantie bis max. 100.000km, Lehn e geteilt umlegbar, geschwindigkeitsabhängig, Nichtraucherfahrzeug, Scheckheft gepflegtes Fahrzeug, Bestand Inza hlungnahme, Weiteres: ehemalige unverbindliche Preisempfehlung des Herstellers 20.974 EUR. TOP Angebot vom Vw/Au di/Skoda - Vertragshändler. Besuchen Sie uns auf über 100.000 qm Ausstellungsfläche und sprechen Sie persönlich mit einem/einer unserer 46 Verkaufsberater/-innen. Unsere Öffnungszeiten im Verkauf: Mo.-Fr. 8:00-20:00 - Sa. 9: 00-16:00 Willkommen zur Probefahrt. Finanzierung zu Sonderkonditionen, auch ohne Anzahlung möglich. Gerne nehmen wir Ihren Gebrauchtwagen in Zahlung. Irrtümer und Änderungen vorbehalten. Weitere Fahrzeuge finden Sie auf www.p otti.de, Audi GW:plus, Kontakt:, Ihr Direktruf 023819089002, \n]
```

```
1
```

```
[\n, Highlights:, Herstellereigene Garantie 24 Monate, Technik + Sicherheit: , ESP; ASR; Kopfairbags; Hecksche ibenwischer; Innenspiegel automatisch abblendend; Außentemperatur-Anzeige; Dynamiklenkung; Elektr. Bremssystem (EBS); Sonstiges Technik & Sicherheit, Multimedia: , Multimedia; Bluetooth; Radio, Assistenzsysteme: , Berganfahrassistent; Reifendruck-Kontrolle; Sonstiges Assistenzsysteme, Innen: , Klima; Bordcomputer; Fensterheber
```

elektr. 2*; Fußmatten; Sonstiges Innenausstattung, Sitze + Polster: , Komfortsitz(e) vorn; Einstieghilfe ("Easy Entry"); Fahrersitz höhenverstellbar; Rücksitzbank umlegbar; Rückbank geteilt; Sonstiges Sitze & Polster; Iso-Fix, Außen: , elektr. Außenspiegel; Zentralverriegelung: Keyless Entry; getönte Scheiben; Dämmglas, Scheinwerfer + Leuchten: , Leuchtweitenregulierung; Tagfahrlicht; Halogenscheinwerfer; Sonstiges Scheinwerfer & Leuchten, Räder + Reifen: , Leichtmetallfelgen: 14"; Sommerreifen, Sonstiges: , CO2-Effizienzklasse: D, "Wir akzeptieren die Bezahlung mit sicherbezahlen.de", Fahrzeugstandort: Autohaus Am Ruhrdeich GmbH, Daimlerstrasse 1 in 47167 Duisburg, Änderungen, Zwischenverkauf und Irrtümer vorbehalten!, \n]

1

[\\n, Fahrzeugnummer für Anfragen: LA027224, KL, Serienausstattung, ABS mit elektronischem Bremskraftverteiler (EBV), Airbag, Fahrer- und Beifahrer (abschaltbar), Airbag, Kopf vorn und hinten, Airbag, Seiten für Fahrer- und Beifahrer, Antriebs-Schlupf-Regelung (ASR), Audi connect Notruf ("eCall"), Audi pre sense front, Blinker (LED) in den Außenspiegeln integriert, Bremsassistent Folgekollision, Dachkantenspoiler, Drehzahlmesser, Elektrisch einstellbare Außenspiegel, Elektrische Fensterheber vorn und hinten, Elektronische Differenzialsperre (EDS), Elektronische Stabilisierungskontrolle (ESC), Fahrer- und Beifahrersitz, höhenverstellbar, Fußmatten vorn und hinten, Gepäckraumabdeckung und ?Beleuchtung, Geschwindigkeitsbegrenzer, Geteilt umklappbare Rücksitzlehne (60:40), Gurtanlegekontrolle, Halogenscheinwerfer in Freiformtechnik mit Klarglasabdeckung, Heckscheibe zwischer und -heizung, Karosserie, verzinkt, Kindersitzvorbereitung ??Isofix?? und Top Tether für die äußeren Fondsitze, Klimaanlage, manuell, Kopfstützen vorn und hinten, Lenkrad, höhenverstellbar, Leuchtweitenregulierung, Licht- und Regensensor, Make-up-Spiegel in den Sonnenblenden, Multifunktions-Leder-Lenkrad im 3-Speichen-Design mit bedientasten für das MMI, Nicht-Raucher-Ausführung, Pollen- und Staubfilter, Radioanlage MMI digitales Kombiinstrument 10,25 Zoll, ohne zusätzliches MMI touch-Display, 1 x SDXC-Kartenleser, 1 x USB-Typ-A-Schnittstelle zum Laden externer Geräte, Bluetooth-Schnittstelle (Freisprechen und Audiostreaming), Sprachbedienung für Telefonie und Medien, 4 Passivlautsprecher vorn mit 40 Watt, AM-/FM-Phasendiversity, geschwindigkeitsabhängige Lautstärkeanpassung, Radschrauben mit Diebstahlsicherung, Reifen-Reparatur-Set, Reifendruck-Kontrollanzeige, Scheibenwischer mit Intervallschaltung, Service-Intervall-Anzeige, Servolenkung, Spiegelgehäuse außen in schwarz, Spurverlassenswarnung (Beim Verlassen der Spur warnt das System über Vibration im Lenkrad den Fahrer), Stahlräder mit Bereifung 185/65 R15, Start-Stopp-System mit Rekuperation, Stoßstangen in Wagenfarbe, Tagfahrlicht, Wärmeschutzverglasung, Warndreieck, Zentralverriegelung mit Fernbedienung, Sonderausstattung, Diebstahlwarnanlage, Geschwindigkeitsregelanlage (Tempomat) mit -Begrenzer, Innenspiegel automatisch abblendend, Komfortklimaautomatik 2-Zonen, Leder-Multifunktions-Lenkrad plus im 3-Speichen Design, LED-Heckleuchten, Leichtmetallfelgen 15 Zoll im 5-Arm-Design mit Bereifung 185/65 R15, Mittelarmlehne vorn, Raucherpaket beinhaltet: Aschenbecher und Zigarettenanzünder, Sitzheizung für die Vordersitze, Gern nehmen wir Ihren Gebrauchtwagen in Zahlung. Auf Wunsch bieten wir Ihnen eine maßgeschneiderte Finanzierung mit oder ohne Anzahlung! Eine Garantieverlängerung über CarGarantie von zusätzlichen 60 Monaten ist auch möglich, Preis auf Anfrage (abhängig von Motor- und Laufleistung). Unser Team berät Sie in Deutsch, russisch, polnisch, französisch, englisch, und hessisch. Zwischenverkauf und Irrtümer vorbehalten.\n] 1

[\\n, Sitzbezug / Polsterung: Velours / Leder, Audio-Navigationssystem R-Link Evolution mit Touchscreen und Carmat TomTom, Tagfahrlicht LED, Anti-Submarining-Airbag im Fond, Airbag Beifahrerseite abschaltbar, Seitenairbag vorn, Klimaautomatik, Anti-Blockier-System (ABS), Elektron. Stabilitäts-Programm (ESP), Einparkhilfe hinten, Zentralverriegelung / Startanlage Handsfree Entry & Drive, Fensterheber elektr. mit Impulsgeber links vorn, Sitzheizung vorn, Servolenkung geschwindigkeitsabhängig, Sound-System BOSE, Geschwindigkeits-Regelanlage (Tempomat), Start/Stop-Anlage, Fahrassistenz-System: Berganfahr-Assistent (HSA), Wegfahrsperrre, Licht- und Regensensor, LM-Felgen, Bordcomputer, Außenspiegel elektr. verstell- und heizbar, Innenspiegel mit Abblendautomatik, Lenkrad (Leder), Mittelarmlehne vorn mit Ablagefach, Gepäckraumabdeckung / Rollo, Automatik-Getriebe, Rücksitzlehne geteilt/klappbar, Coming-Home-Lichtfunktion, Reifendruck-Kontrollsyste, Komfort-Paket Plus, Isofix-Aufnahmen für Kindersitz, Isofix-Aufnahmen für Kindersitz an Beifahrersitz, Reserverad in Fahrbereifung, Nebelscheinwerfer LED, AH Standard, Modellpflege, Komfort-Paket, Sitz vorn links höhenverstellbar, Sitz vorn rechts höhenverstellbar, Modularitäts-Paket, Lenksäule (Lenkrad) höhenverstellbar, Lenksäule (Lenkrad) längsverstellbar, Techno-Paket, Look-Paket Chrom (außen), Kühlergrill verchromt, Seitenschutzleisten mit Chromeinlage, Heckzierleiste Chrom, Look-Paket Klavierlack schwarz (innen), Sondermodell Bose Edition, Außenspiegel schwarz hochglänzend, Scheinwerfer Full-LED, Heckleuchten LED, Nebelscheinwerfer, Fensterzierleisten verchromt, Stoßfänger Wagenfarbe, Verglasung hinten abgedunkelt, Subwoofer, Schaltpunktanzeige, Airbag Fahrer-/Beifahrerseite, Kopfstützen verstellbar, Fensterheber elektrisch hinten, Türgriffe außen Wagenfarbe, Gurtstraffer, Leseleuchte Beifahrerseite, Elektr. Bremskraftverteilung, Bremsassistent, Untersteuerungskontrolle (USC), Antischlupfregelung (ASR), Getriebe 6-Gang - Doppelkulplungsgetriebe, Schadstofffarm nach Abgasnorm Euro 6, Energierückgewinnung (Energy Smart Management), Finanzierungsbispiel, *, Barzahlungspreis: 18.990,00 Euro, Anzahlung: 3.798,00 Euro, Laufzeit: 72 Monate, Nettokreditbetrag: 15.192,00 Euro, Effektiver Jahreszins: 5,99 %, Sollzinssatz (gebunden) p.a.: 5,83 %, Bruttokreditbetrag: 18.041,00 Euro, Monatliche Rate: 251,00 Euro, *Repräsentatives Finanzierungsbeispiel der Renault Bank / Renault Leasing, Geschäftsbereich der RCI Banque S.A. Niederlassung Deutschland, Jagenbergstraße 1, 41468 Neuss für Privatkunden, für die das Autohaus als Darlehnsvermittler beratend tätig ist. Bonität vorausgesetzt. Angaben entsprechen zugeleich dem 2/3 Beispiel nach § 6a Abs. 3 PAngV.\n]

1

[\\nEuro6, Rückfahrkamera, Klimaautomatik 2-fach, Fahrerairbag, Beifahrerairbag, Fenster-/Kopfairbags vorne, Fenster-/Kopfairbags im Fond, Einparkhilfe Kamera, Einparkhilfe Sensoren vorne, Einparkhilfe Sensoren hinten, Traktionskontrolle, Fensterheber elektrisch 4-fach, Sitz links orthopaedisch, Bluetooth Freisprecheinrichtung, Außenspiegel beheizbar, Bremsassistent, Tempomat, ISOFIX Kindersitzhalterung, Leichtmetallfelgen, Lenksäule einstellbar, Gepäckraumabdeckung, weitere Ausstattungen vorhanden, Airbag Fahrer-/Beifahrerseite, Audiosystem R 4.0 Intel liLink, Außenspiegel Wagenfarbe, Außenspiegel elektr. verstell- und heizbar, elektr. anklappbar, Bordcomputer, Bremsassistent, Dach, Design-Ausführung, Drehzahlmesser, Ecoflex, Einstiegsleisten mit Schriftzug, Elektron. Stabilitäts-Programm Plus (ESP), Fahrassistenz-System: Berganfahr-Assistent (HSA, Hill Start Assist), Fahrassistenz-System: Notbrems-Assistent, Fahrzeugschlüssel (2) beide klappbar, Fensterheber elektrisch vorn + hinten, Fensterzierleisten Edelstahl / Chrom, Feststellbremse elektrisch, Frontkamera, Geschwindigkeits-Regelanlage (Tempomat), Isofix-Aufnahmen für Kindersitz an Rücksitz, Karosserie: 5-türig, Klimaautomatik 2-Zonen, Kopf-Airbag-System, Laderaumabdeckung, Lendenwirbelstütze Sitz vorn links, elektr. verstellbar, Lenkrad mit Multifunktion, Lenksäule (Lenkrad) verstellbar, Licht-Paket, Motor 1,0 Ltr. - 77 kW 12V, Nebelscheinwerfer, OnStar SOS Service, Parkpilot system vorn und hinten, Reifendruck-Kontrollsyste, Rückfahrkamera, Schadstofffarm nach Abgasnorm Euro 6, Seitenairbag vorn, Servolenkung geschwindigkeitsabhängig, Start/Stop-Anlage, Steckdose (12V-Anschluß) in Mittelkonsole, Tagfahrlicht LED, Türgriffe außen Wagenfarbe, Wegfahrsperrre, Zentralverriegelung mit Fernbedienung, ecoFLEX-Paket, incl. Winterkompletträder , \n]

1

Name: description, Length: 12874, dtype: int64

In [35]:

```
df.drop("description", axis = 1, inplace = True)
```

body_type

```
In [36]: df.body_type.value_counts(dropna=False)
```

```
Out[36]: Sedans      7903  
Station wagon    3553  
Compact        3153  
Van            783  
Other          290  
Transporter     88  
NaN             60  
Off-Road        56  
Coupe           25  
Convertible       8  
Name: body_type, dtype: int64
```

BELLİ ŞARTI SAĞLAYAN SATIRLARIN İNDEKSİNİ ALMA:

```
In [37]: # body_type_nanlarının_satır_indeks_numaraları = df[df.body_type.isnull()].index
```

Body

```
In [38]: df.Body.value_counts(dropna=False)
```

```
-----  
TypeError                                                 Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

```
Out[38]: [\n, Sedans, \n]      7903  
[\n, Station wagon, \n]    3553  
[\n, Compact, \n]        3153  
[\n, Van, \n]            783  
[\n, Other, \n]          290  
[\n, Transporter, \n]    88  
NaN                      60  
[\n, Off-Road, \n]        56  
[\n, Coupe, \n]           25  
[\n, Convertible, \n]     8  
Name: Body, dtype: int64
```

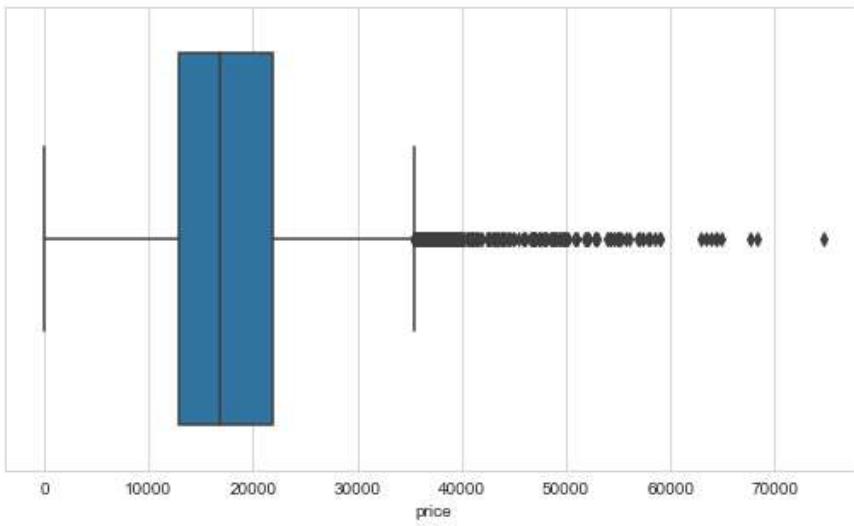
```
In [39]: df.drop("Body", axis = 1, inplace = True)
```

price (target column)

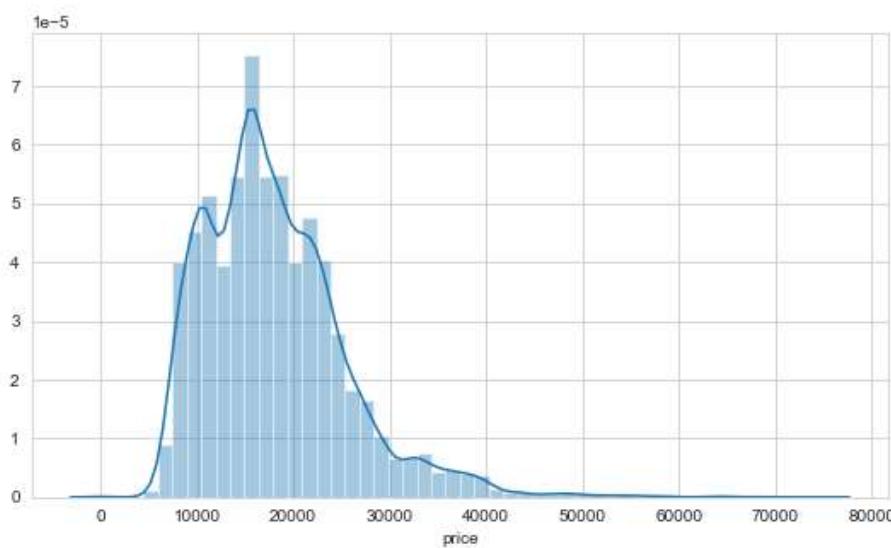
```
In [40]: df.price.isnull().sum()
```

```
Out[40]: 0
```

```
In [41]: sns.boxplot(df.price);
```



```
In [42]: sns.distplot(df.price);
```



vat

The Value Added Tax, or VAT, in the European Union is a general, broadly based consumption tax assessed on the value added to goods and services.

```
In [43]: df.vat.value_counts(dropna=False)
```

```
Out[43]: VAT deductible      10980
NaN                      4513
Price negotiable        426
Name: vat, dtype: int64
```

km

```
In [44]: df.km.value_counts(dropna=False)
```

```
Out[44]: 10 km       1045
- km        1024
1 km        367
5 km        170
50 km       148
...
52,811 km      1
116,593 km     1
25,976 km      1
59,217 km      1
30,348 km      1
Name: km, Length: 6690, dtype: int64
```

```
In [45]: df["km"] = pd.Categorical(df["km"], categories=["- km"], ordered=True)
```

```
In [46]: df.km
```

```
Out[46]: 0      NaN  
1      NaN  
2      NaN  
3      NaN  
4      NaN  
...  
15914  - km  
15915  NaN  
15916  NaN  
15917  NaN  
15918  - km  
Name: km, Length: 15919, dtype: category  
Categories (1, object): ['- km']
```

```
In [47]: df.km.str.replace(',', '').str.extract('(\d{1,8})')
```

```
Out[47]: 0      0  
1      NaN  
2      NaN  
3      NaN  
4      NaN  
...    ...  
15914  NaN  
15915  NaN  
15916  NaN  
15917  NaN  
15918  NaN
```

15919 rows × 1 columns

```
In [48]: df["km"] = df.km.str.replace(',', '').str.extract('(\d{1,8})')[0].astype('float')  
#Function of extract returns a dataframe, that's why we have to use slicing method before changing the type of
```

EXTRACT DF DÖNDÜRÜR. SERİ DÖNMESİ İÇİN SONUNA [0] KOYDUK. .str.extract('(\d{1,8})')[0] SERİ YAPMAZSAK astype çalışmaz .str.extract('(\d{1,8})')[0].astype('float') FLOAT'A ÇEVİRMEZKSEK OBJECT KALIR. YENİ DEĞER GİBİ ALGILAR.

```
In [49]: df.km.value_counts(dropna=False)
```

```
Out[49]: NaN      15919  
Name: km, dtype: int64
```

registration & First Registration

```
In [50]: df.registration.value_counts(dropna=False)
```

```
Out[50]: --      1597  
03/2018    695  
02/2019    585  
05/2018    572  
03/2019    543  
01/2019    541  
04/2018    541  
02/2018    539  
03/2016    536  
06/2018    532  
04/2016    532  
01/2018    511  
04/2019    506  
02/2016    472
```

```
03/2017    471
05/2016    459
06/2016    452
05/2019    440
06/2017    409
05/2017    404
07/2018    396
04/2017    380
01/2016    376
02/2017    368
01/2017    306
08/2018    285
06/2019    224
07/2017    215
11/2017    180
07/2016    176
10/2016    160
10/2017    154
09/2017    149
11/2016    142
09/2018    141
09/2016    141
12/2016    134
12/2017    123
08/2017    114
11/2018    110
12/2018    103
10/2018    97
08/2016    94
07/2019    6
09/2019    5
08/2019    1
12/2019    1
11/2019    1
Name: registration, dtype: int64
```

```
In [51]: df.registration.unique()
```

```
Out[51]: array(['01/2016', '03/2017', '02/2016', '08/2016', '05/2016', '03/2016',
               '06/2017', '05/2018', '09/2016', '06/2016', '10/2016', '04/2016',
               '06/2018', '11/2017', '07/2016', '12/2016', '09/2018', '04/2018',
               '07/2017', '02/2017', '11/2016', '01/2018', '05/2017', '04/2017',
               '07/2018', '01/2017', '02/2018', '03/2018', '---', '08/2017',
               '10/2017', '08/2018', '09/2017', '12/2017', '04/2019', '12/2018',
               '03/2019', '06/2019', '05/2019', '02/2019', '01/2019', '10/2018',
               '11/2018', '09/2019', '07/2019', '08/2019', '11/2019', '12/2019'],
              dtype=object)
```

```
In [52]: df.registration.sort_values()
```

```
Out[52]: 8394      -/
 11117     -/
 2305      -/
 1188      -/
 2310      -/
 ...
 4780      12/2018
 1560      12/2018
 15437     12/2018
 4500      12/2018
 12975     12/2019
Name: registration, Length: 15919, dtype: object
```

ÖNCELİKLE NAN DEĞERLERİ GÖRMEK İÇİN

```
In [53]: df.sort_values(by="registration", na_position="first")
```

```
Out[53]:
```

| | make_model | body_type | price | VAT | km | registration | prev_owner | hp | Type | Previous Owners | Next Inspection |
|------|------------|-----------|-------|----------------|-----|--------------|------------|-------|------------------------------|-----------------|-----------------|
| 8394 | Opel Corsa | Compact | 12788 | VAT deductible | NaN | --- | None | 66 kW | [, New,, Regular/Benzine 91] | NaN | NaN |

| | make_model | body_type | price | vat | km | registration | prev_owner | hp | Type | Previous Owners | Next Inspection |
|-------|----------------|---------------|-------|----------------|-----|--------------|------------------|--------|--|-----------------|--|
| 11117 | Opel Insignia | Station wagon | 11980 | None | NaN | -/- | 1 previous owner | 195 kW | [, Used, , Diesel (Particulate Filter)] | \n1\n | NaN |
| 2305 | Audi A1 | Compact | 19990 | VAT deductible | NaN | -/- | None | 70 kW | [, New, , Super 95] | NaN | NaN |
| 1188 | Audi A1 | Compact | 15900 | VAT deductible | NaN | -/- | 1 previous owner | 70 kW | [, Demonstration, , Super 95 / Super E10 95] | \n1\n | [, \n09/2021\n, \n, 4.2 l/100 km (comb), \n, 5 l...] |
| 2310 | Audi A1 | Compact | 19990 | VAT deductible | NaN | -/- | None | 70 kW | [, New, , Super E10 95 / Super 95] | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 4780 | Audi A3 | Sedans | 21490 | VAT deductible | NaN | 12/2018 | 1 previous owner | 85 kW | [, Demonstration, , Super 95] | \n1\n | [, \n06/2021\n, \n116 g CO2/km (comb)\n] |
| 1560 | Audi A1 | Sedans | 23500 | VAT deductible | NaN | 12/2018 | None | 85 kW | [, Pre-registered, , Super 95 (Particulate Fil...] | NaN | NaN |
| 15437 | Renault Espace | None | 33000 | VAT deductible | NaN | 12/2018 | None | 118 kW | [, Pre-registered, , Diesel] | NaN | NaN |
| 4500 | Audi A3 | Sedans | 23750 | VAT deductible | NaN | 12/2018 | 1 previous owner | 110 kW | [, Used, , Gasoline] | \n1\n | [, \n12/2021\n, \n116 g CO2/km (comb)\n] |
| 12975 | Opel Insignia | Compact | 21900 | None | NaN | 12/2019 | None | 100 kW | [, Used, , Diesel] | NaN | \n12/2022\n |

15919 rows × 40 columns

In [54]:

```
df.registration.isna().sum()
```

Out[54]: 0

In [55]:

```
df["registration"] = df["registration"].str.replace('^-/-', 'NaN', regex=False)
```

In [56]:

```
df["registration"] = df["registration"].str.replace('/', ' ', regex=False)
```

```
In [57]: df["registration"] = pd.to_datetime(df["registration"], format='%m%Y')      # format="%Y-%d-%m %H:%M:%S"
# TARİH FORMATINDA %m/%Y ARAYA "/" KOYDUĞUMUZDA İŞLEM YAPTI ANCAK YILIN BÜTÜN AYLARINI OCAK AYIYMIŞ GİBİ GÖSTER
```

```
In [58]: df["First Registration"].value_counts(dropna=False)
```

```
-----
TypeError                                                 Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'
```

```
Out[58]: [\n, 2018, \n]    4522
[\n, 2016, \n]    3674
[\n, 2017, \n]    3273
[\n, 2019, \n]    2853
NaN              1597
Name: First Registration, dtype: int64
```

I prefer "First Registration" column because I don't need month values

```
In [59]: df['First Registration'] = df['First Registration'].str[1].astype('float')
```

```
In [60]: df["First Registration"].value_counts(dropna=False)
```

```
Out[60]: 2018.00    4522
2016.00    3674
2017.00    3273
2019.00    2853
NaN        1597
Name: First Registration, dtype: int64
```

```
In [61]: df['age'] = 2019 - df['First Registration']
```

```
In [62]: df.age.value_counts(dropna=False)
```

```
Out[62]: 1.00    4522
3.00    3674
2.00    3273
0.00    2853
NaN     1597
Name: age, dtype: int64
```

```
In [63]: df.drop(["First Registration"], axis=1, inplace=True)
```

prev_owner & Previous Owners

```
In [64]: df.prev_owner.value_counts(dropna=False)
```

```
Out[64]: 1 previous owner    8294
NaN            6828
2 previous owners    778
3 previous owners     17
4 previous owners      2
Name: prev_owner, dtype: int64
```

```
In [65]: df["Previous Owners"].value_counts(dropna=False)
```

```
-----
TypeError                                                 Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_loc
```

ations
TypeError: unhashable type: 'list'

```
Out[65]: \n1\n          8101
NaN           6640
\n2\n          766
\n0\n          163
\n3\n          17
...
[\n2\n, \n127 g CO2/km (comb)\n]      1
[\n1\n, \n159 g CO2/km (comb)\n]      1
[\n0\n, \n117 g CO2/km (comb)\n]      1
[\n0\n, \n187 g CO2/km (comb)\n]      1
[\n1\n, \n102 g CO2/km (comb)\n]      1
Name: Previous Owners, Length: 103, dtype: int64
```

I prefer "Previous Owners" column because it has less NaN values

```
In [66]: [item[0] if type(item) == list else item for item in df["Previous Owners"]]
```


'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
nan,
'\n2\n',
'\n2\n',
'\n2\n',
'\n1\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
nan,
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
nan,
'\n2\n',
'\n2\n',
nan,
'\n1\n',
nan,
'\n1\n',
nan,
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n1\n',
nan,
'\n1\n',
'\n2\n',
'\n2\n',
'\n2\n',
'\n2\n',
nan,
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n2\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,

nan,
nan,
nan,
nan,
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n2\n',
'\n2\n',
'\n1\n',
'\n1\n',
'\n2\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n2\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
nan,
nan,
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
nan,
nan,
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
nan,
nan,
nan,
'\n2\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
nan,
nan,
'\n1\n',
'\n1\n',
nan,
nan,
nan,
nan,
'\n1\n',

nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n2\n',
'\n2\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n2\n',
nan,
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n2\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
nan,
nan,
nan,
nan,
'\n2\n',
'\n1\n',
nan,
nan,
'\n1\n',
'\n1\n',
'\n1\n',

'\n1\n',
'\n1\n',
'\n1\n',
nan,
nan,
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
nan,
'\n2\n',
'\n0\n',
'\n1\n',
'\n2\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n2\n',
nan,
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n2\n',
nan,
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n2\n',
'\n1\n',
'\n2\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n2\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n2\n',
'\n2\n',
nan,
'\n1\n',
nan,
nan,
'\n1\n',
'\n1\n',
nan,
'\n0\n',
'\n1\n',
'\n1\n',
'\n2\n',
'\n1\n',
'\n2\n',
'\n1\n',
nan,
nan,
'\n1\n',
'\n1\n',

'\n1\n',
'\n1\n',
nan,
nan,
'\n2\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
nan,
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n3\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
nan,
'\n1\n',
nan,
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
nan,
'\n0\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n0\n',

'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n2\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
nan,
'\n2\n',
nan,
nan,
'\n1\n',
'\n1\n',
nan,
nan,
nan,
nan,
nan,
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
nan,
nan,
nan,
'\n2\n',
'\n1\n',
'\n2\n',
'\n1\n',
'\n1\n',
nan,
nan,
nan,
'\n1\n',
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n4\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
'\n0\n',
'\n1\n',
nan,
'\n1\n',
nan,
'\n1\n',
'\n1\n',
nan,
'\n1\n',
'\n1\n',
nan,
nan,
nan,
nan,
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',
'\n1\n',

```
'\n1\n',
...]
```

```
In [67]: df["Previous_Owners"] = [item[0] if type(item) == list else item for item in df["Previous_Owners"]]
df["Previous_Owners"]
```

```
Out[67]: 0      \n2\n
1      NaN
2      \n1\n
3      \n1\n
4      \n1\n
...
15914    NaN
15915    \n1\n
15916    \n1\n
15917    NaN
15918    \n1\n
Name: Previous_Owners, Length: 15919, dtype: object
```

```
In [68]: df["Previous_Owners2"] = df["Previous_Owners"].apply(lambda item: item[0] if type(item) == list else item)
df["Previous_Owners2"]
```

```
Out[68]: 0      \n2\n
1      NaN
2      \n1\n
3      \n1\n
4      \n1\n
...
15914    NaN
15915    \n1\n
15916    \n1\n
15917    NaN
15918    \n1\n
Name: Previous_Owners2, Length: 15919, dtype: object
```

```
In [69]: df["Previous_Owners"] = df["Previous_Owners"].str.strip("\n").astype('float')
```

```
In [70]: df["Previous_Owners"].value_counts(dropna=False)
```

```
Out[70]: 1.00    8294
NaN      6640
2.00     778
0.00     188
3.00      17
4.00       2
Name: Previous_Owners, dtype: int64
```

```
In [71]: df.drop(["prev_owner", "Previous_Owners", "Previous_Owners2"], axis=1, inplace=True)
# Prev Own'da daha az nan olduğundan bu sütun tercih edildi.
# SÜTUNUN DATA TYPE'I INTEGER ANCAK İÇİNDEKİ DEĞERLER FLOAT
```

hp

```
In [72]: df.hp.value_counts(dropna=False).tail(15)
```

```
Out[72]: 167 kW    2
150 kW    2
44 kW     1
132 kW    1
163 kW    1
4 kW      1
195 kW    1
75 kW     1
137 kW    1
133 kW    1
9 kW      1
239 kW    1
84 kW     1
123 kW    1
115 kW    1
Name: hp, dtype: int64
```

```
In [73]: df["hp_kw"] = df.hp.str.extract('(\d{1,4})')[0].astype('float')
```

```
In [74]:
```

```
#Alternative method  
#df.replace({"hp" : {"kW" : ""}}, regex = True, inplace = True)  
#df["hp_kw"] = pd.to_numeric(df.hp)
```

```
dfj["hp"] = dfj["hp"].str.replace(' kW', '', regex=False) dfj["hp"] = dfj["hp"].str.replace('-kW', 'NaN', regex=False)
```

```
In [75]:
```

```
df.drop('hp', axis=1, inplace=True)
```

Type

```
In [76]:
```

```
df.Type.value_counts(dropna=False)
```

```
-----  
TypeError Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

```
Out[76]: [ , Used, , Diesel (Particulate Filter)]
```

```
3475
```

```
[ , Used, , Diesel]
```

```
2516
```

```
[ , Used, , Gasoline]
```

```
2367
```

```
[ , Used, , Super 95]
```

```
1818
```

```
[ , Pre-registered, , Super 95]
```

```
500
```

```
...
```

```
[ , Pre-registered, , CNG]
```

```
1
```

```
[ , Used, , Others (Particulate Filter)]
```

```
1
```

```
[ , New, , Super 95 / Regular/Benzine 91 / Super Plus 98 / Regular/Benzine E10 91 / Super Plus E10 98 / Super E10 95]
```

```
1
```

```
[ , Pre-registered, , Super E10 95 (Particulate Filter)]
```

```
1
```

```
[ , Employee's car, , Super E10 95 (Particulate Filter)]
```

```
1
```

```
Name: Type, Length: 170, dtype: int64
```

```
In [77]:
```

```
df["Type"] = df.Type.str[1]
```

SÜTUN DEĞERLERİNİN KATEGORİ OLARAK LİSTESİ ALMA

```
In [78]:
```

```
df["Type1"] = pd.Categorical(df["Type"], ordered=False)  
df.Type1 # Output: Categories (5, object): ['Demonstration', 'Employee's car', 'New', 'Pre-registe
```

```
Out[78]:
```

```
0           Used
```

```
1           Used
```

```
2           Used
```

```
3           Used
```

```
4           Used
```

```
...
```

```
15914          New
```

```
15915          Used
```

```
15916  Pre-registered
```

```
15917  Pre-registered
```

```
15918  Demonstration
```

```
Name: Type1, Length: 15919, dtype: category
```

```
Categories (5, object): ['Demonstration', 'Employee's car', 'New', 'Pre-registered', 'Used']
```

```
In [79]:
```

```
df['Type'].value_counts(dropna=False)
```

```
Out[79]: Used 11096
```

```
New 1650
```

```
Pre-registered 1364
```

```
Employee's car 1011
```

```
Demonstration      796
NaN                2
Name: Type, dtype: int64

df.Type = [i[1] if type(i) == list else np.nan for i in df.Type]df.loc[df['Type'].str.contains('Used', case=False, regex=False, na=False), 'Type'] = 'Used'
df.loc[df['Type'].str.contains('New', case=False, regex=False, na=False), 'Type'] = 'New' df.loc[df['Type'].str.contains("Employee's car", case=False, regex=False, na=False), 'Type'] = "Employee's car" df.loc[df['Type'].str.contains('Demonstration', case=False, regex=False, na=False), 'Type'] = 'Demonstration' df.loc[df['Type'].str.contains('Pre-registered', case=False, regex=False, na=False), 'Type'] = 'Pre-registered'
df.loc[df['Type'].str.startswith('[', na=False), 'Type'] = 'Pre-registered'
```

Next Inspection & Inspection new

```
In [80]: df["Next Inspection"].value_counts(dropna=False)
```

```
-----
TypeError                                 Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_loc
ations
TypeError: unhashable type: 'list'
```

```
Out[80]: NaN          12384
\n04/2022\n          62
\n03/2021\n          38
\n03/2022\n          36
\n06/2021\n          34
...
[\n02/2021\n, \n154 g CO2/km (comb)\n]      1
[\n04/2020\n, \n129 g CO2/km (comb)\n]      1
[\n02/2019\n, \n139 g CO2/km (comb)\n]      1
[\n11/2020\n, \n119 g CO2/km (comb)\n]      1
[\n04/2020\n, \n103 g CO2/km (comb)\n]      1
Name: Next Inspection, Length: 1400, dtype: int64
```

```
df['Next Inspection']=df['Next Inspection'].str.extract('(../201.)')# AY VE YILI ÇEKMEK İÇİN: df['Next Inspection']=df['Next
Inspection'].str.extract('(\S+)/(\S+)') df['Next Inspection']=df['Next Inspection'].str.extract("(\\d{2})/(\\d{4})")
```

```
In [81]: df["Inspection new"].value_counts(dropna=False)
```

```
-----
TypeError                                 Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_loc
ations
TypeError: unhashable type: 'list'
```

```
Out[81]: NaN          11987
\nYes\n, \nEuro 6\n          523
\nYes\n          362
[\nYes\n, \n102 g CO2/km (comb)\n]          174
[\nYes\n, \n4 (Green)\n]          166
...
[\nYes\n, \n, 4.4 l/100 km (comb), \n, 5.4 l/100 km (city), \n, 5 l/100 km (country), \n]      1
[\nYes\n, \n194 g CO2/km (comb)\n]          1
[\nYes\n, \n88 g CO2/km (comb)\n]          1
[\nYes\n, \n, 6 l/100 km (comb), \n, 7.5 l/100 km (city), \n, 5.2 l/100 km (country), \n]      1
[\nYes\n, \n, 6.2 l/100 km (comb), \n, 7.6 l/100 km (city), \n, 5.4 l/100 km (country), \n]      1
Name: Inspection new, Length: 203, dtype: int64
```

```
In [82]: df["Inspection_new"] = [item[0] if type(item) == list else item for item in df["Inspection new"]]
df["Inspection_new"] = df["Inspection_new"].str.strip("\n")
```

```
dfj['Inspection new']=dfj['Inspection new'].str.extract('Yes')
```

```
In [83]: df["Inspection_new"].value_counts(dropna=False)
```

```
Out[83]: NaN    11987
Yes    3932
Name: Inspection_new, dtype: int64
```

```
In [84]: df.drop(["Next Inspection", "Inspection new"], axis=1, inplace=True)
```

Warranty

In [85]:

```
df.Warranty.value_counts(dropna=False)
```

```
-----  
TypeError Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'  
  
Out[85]:  
NaN 5420  
[\n, \n, \nEuro 6\n] 1868  
\n12 months\n 1177  
\n 979  
\n24 months\n 566  
... 1  
[\n10 months\n, \n104 g CO2/km (comb)\n] 1  
[\n, \n, \n, 4.9 l/100 km (comb), \n, 6.2 l/100 km (city), \n, 4.2 l/100 km (country), \n] 1  
[\n60 months\n, \n118 g CO2/km (comb)\n] 1  
[\n16 months\n, \n152 g CO2/km (comb)\n] 1  
[\n24 months\n, \n141 g CO2/km (comb)\n] 1  
Name: Warranty, Length: 516, dtype: int64
```

In [86]:

```
df["Warranty"] = df.Warranty.apply(lambda x : x[0] if type(x)==list else x)  
df["Warranty"] = df.Warranty.str.strip("\n").str.extract('(\d{1,2})')[0].astype("float")
```

In [87]:

```
df.Warranty.value_counts(dropna=False).tail(15)
```

```
Out[87]:  
28.00 2  
21.00 2  
22.00 2  
9.00 2  
30.00 1  
33.00 1  
56.00 1  
40.00 1  
7.00 1  
15.00 1  
8.00 1  
10.00 1  
49.00 1  
47.00 1  
65.00 1  
Name: Warranty, dtype: int64
```

```
dfj['Warranty']=dfj['Warranty'].str.extract('...months')dfj["Warranty"] = dfj["Warranty"].str.replace('n1', '1', regex=False) dfj["Warranty"] = dfj["Warranty"].str.replace('n2', '2', regex=False) dfj["Warranty"] = dfj["Warranty"].str.replace('n3', '3', regex=False) dfj["Warranty"] = dfj["Warranty"].str.replace('n4', '4', regex=False) dfj["Warranty"] = dfj["Warranty"].str.replace('n6', '6', regex=False) dfj["Warranty"] = dfj["Warranty"].str.replace('n7', '7', regex=False)dfj["Warranty"] = dfj["Warranty"].str.replace(' months', '', regex=False)df = df.rename(columns={'Warranty': 'Warranty_months'}) df = df.rename(columns={'Next Inspection': 'Next Inspection_Date'})
```

Full Service

In [88]:

```
df['Full Service'].value_counts(dropna=False)
```

```
-----  
TypeError Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'  
  
Out[88]:
```

| | |
|----------------------------|------|
| NaN | 7704 |
| [\n, \n, \n4 (Green)\n] | 2235 |
| [\n, \n, \nEuro 6\n] | 2097 |
| [\n, \n] | 1702 |
| [\n, \n, \nEuro 6d-TEMP\n] | 399 |

```

[\n, \n, \n, 5.4 l/100 km (comb), \n, 7 l/100 km (city), \n, 4.5 l/100 km (country), \n]   1
[\n, \n, \n, 4.7 l/100 km (comb), \n, 5.4 l/100 km (city), \n, 4.2 l/100 km (country), \n]   1
[\n, \n, \n164 g CO2/km (comb)\n]   1
[\n, \n, \n, 5.4 l/100 km (comb), \n, 7.3 l/100 km (city), \n, 4.2 l/100 km (country), \n]   1
[\n, \n, \n, 6 l/100 km (comb), \n, 7.5 l/100 km (city), \n, 5.2 l/100 km (country), \n]   1
Name: Full Service, Length: 123, dtype: int64

```

In [89]: `df.drop("Full Service", axis=1, inplace=True)`

Non-smoking Vehicle

In [90]: `df['Non-smoking Vehicle'].value_counts(dropna=False)`

```

-----
TypeError                                     Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'

```

| Category | Count |
|--|-------|
| NaN | 8742 |
| [\n, \n] | 3647 |
| [\n, \n, \n4 (Green)\n] | 1240 |
| [\n, \n, \nEuro 6\n] | 1127 |
| [\n, \n, \nEuro 6d-TEMP\n] | 345 |
| \n | 70 |
| [\n, \n, \n120 g CO2/km (comb)\n] | 63 |
| [\n, \n, \n85 g CO2/km (comb)\n] | 46 |
| [\n, \n, \n107 g CO2/km (comb)\n] | 36 |
| [\n, \n, \n104 g CO2/km (comb)\n] | 36 |
| [\n, \n, \n153 g CO2/km (comb)\n] | 31 |
| [\n, \n, \n124 g CO2/km (comb)\n] | 29 |
| [\n, \n, \n1 (No sticker)\n] | 24 |
| [\n, \n, \n99 g CO2/km (comb)\n] | 23 |
| [\n, \n, \n97 g CO2/km (comb)\n] | 20 |
| [\n, \n, \n106 g CO2/km (comb)\n] | 20 |
| [\n, \n, \n108 g CO2/km (comb)\n] | 19 |
| [\n, \n, \n105 g CO2/km (comb)\n] | 18 |
| [\n, \n, \n109 g CO2/km (comb)\n] | 18 |
| [\n, \n, \n100 g CO2/km (comb)\n] | 18 |
| [\n, \n, \n113 g CO2/km (comb)\n] | 18 |
| [\n, \n, \nEuro 6c\n] | 17 |
| [\n, \n, \n98 g CO2/km (comb)\n] | 15 |
| [\n, \n, \n129 g CO2/km (comb)\n] | 14 |
| [\n, \n, \n131 g CO2/km (comb)\n] | 13 |
| [\n, \n, \n114 g CO2/km (comb)\n] | 12 |
| [\n, \n, \n103 g CO2/km (comb)\n] | 12 |
| [\n, \n, \n133 g CO2/km (comb)\n] | 12 |
| [\n, \n, \n135 g CO2/km (comb)\n] | 11 |
| [\n, \n, \n116 g CO2/km (comb)\n] | 11 |
| [\n, \n, \n102 g CO2/km (comb)\n] | 11 |
| [\n, \n, \n127 g CO2/km (comb)\n] | 10 |
| [\n, \n, \n143 g CO2/km (comb)\n] | 10 |
| [\n, \n, \n112 g CO2/km (comb)\n] | 9 |
| [\n, \n, \nEuro 6d\n] | 9 |
| [\n, \n, \n92 g CO2/km (comb)\n] | 9 |
| [\n, \n, \n5 (Blue)\n] | 8 |
| [\n, \n, \n118 g CO2/km (comb)\n] | 8 |
| [\n, \n, \n147 g CO2/km (comb)\n] | 7 |
| [\n, \n, \n139 g CO2/km (comb)\n] | 7 |
| [\n, \n, \n128 g CO2/km (comb)\n] | 7 |
| [\n, \n, \n117 g CO2/km (comb)\n] | 6 |
| [\n, \n, \n140 g CO2/km (comb)\n] | 6 |
| [\n, \n, \n126 g CO2/km (comb)\n] | 6 |
| [\n, \n, \n137 g CO2/km (comb)\n] | 6 |
| [\n, \n, \n130 g CO2/km (comb)\n] | 5 |
| [\n, \n, \n115 g CO2/km (comb)\n] | 5 |
| [\n, \n, \n123 g CO2/km (comb)\n] | 4 |
| [\n, \n, \nEuro 5\n] | 4 |
| [\n, \n, \n151 g CO2/km (comb)\n] | 4 |
| [\n, \n, \n, 5.2 l/100 km (comb), \n, 5.8 l/100 km (city), \n, 4.8 l/100 km (country), \n] | 4 |
| [\n, \n, \n149 g CO2/km (comb)\n] | 3 |
| [\n, \n, \n148 g CO2/km (comb)\n] | 3 |
| [\n, \n, \n119 g CO2/km (comb)\n] | 3 |
| [\n, \n, \n95 g CO2/km (comb)\n] | 3 |
| [\n, \n, \n157 g CO2/km (comb)\n] | 3 |

```

[\n, \n, \n136 g CO2/km (comb)\n] 3
[\n, \n, \n110 g CO2/km (comb)\n] 3
[\n, \n, \n125 g CO2/km (comb)\n] 2
[\n, \n, \n142 g CO2/km (comb)\n] 2
[\n, \n, \n, 5.2 l/100 km (comb), \n, 6.4 l/100 km (city), \n, 4.4 l/100 km (country), \n] 2
[\n, \n, \n134 g CO2/km (comb)\n] 2
[\n, \n, \n121 g CO2/km (comb)\n] 2
[\n, \n, \n138 g CO2/km (comb)\n] 2
[\n, \n, \n90 g CO2/km (comb)\n] 2
[\n, \n, \n111 g CO2/km (comb)\n] 2
[\n, \n, \n166 g CO2/km (comb)\n] 2
[\n, \n, \n, 5.6 l/100 km (comb), \n, 7.3 l/100 km (city), \n, 4.7 l/100 km (country), \n] 1
[\n, \n, \n152 g CO2/km (comb)\n] 1
[\n, \n, \n146 g CO2/km (comb)\n] 1
[\n, \n, \n175 g CO2/km (comb)\n] 1
[\n, \n, \n, 6.1 l/100 km (comb), \n, 7.7 l/100 km (city), \n, 5.2 l/100 km (country), \n] 1
[\n, \n, \n, 6.5 l/100 km (comb), \n, 7.9 l/100 km (city), \n, 5.6 l/100 km (country), \n] 1
[\n, \n, \n, 1.6 l/100 km (comb), \n, \n, \n] 1
[\n, \n, \n, 6.1 l/100 km (comb), \n, 7.8 l/100 km (city), \n, 5.1 l/100 km (country), \n] 1
[\n, \n, \n82 g CO2/km (comb)\n] 1
[\n, \n, \n159 g CO2/km (comb)\n] 1
[\n, \n, \n, 5.7 l/100 km (comb), \n, 7.3 l/100 km (city), \n, 4.7 l/100 km (country), \n] 1
[\n, \n, \n150 g CO2/km (comb)\n] 1
[\n, \n, \n, 6.2 l/100 km (comb), \n, 7.9 l/100 km (city), \n, 5.3 l/100 km (country), \n] 1
[\n, \n, \n91 g CO2/km (comb)\n] 1
[\n, \n, \n189 g CO2/km (comb)\n] 1
[\n, \n, \n154 g CO2/km (comb)\n] 1
[\n, \n, \n, 6.3 l/100 km (comb), \n, 8 l/100 km (city), \n, 5.3 l/100 km (country), \n] 1
[\n, \n, \n161 g CO2/km (comb)\n] 1
[\n, \n, \n122 g CO2/km (comb)\n] 1
[\n, \n, \n, 6.5 l/100 km (comb), \n, 8.5 l/100 km (city), \n, 5.4 l/100 km (country), \n] 1
[\n, \n, \n172 g CO2/km (comb)\n] 1
[\n, \n, \n132 g CO2/km (comb)\n] 1
[\n, \n, \n158 g CO2/km (comb)\n] 1
[\n, \n, \n101 g CO2/km (comb)\n] 1
[\n, \n, \n169 g CO2/km (comb)\n] 1
[\n, \n, \n156 g CO2/km (comb)\n] 1
[\n, \n, \n, 6.4 l/100 km (comb), \n, 8.1 l/100 km (city), \n, 5.4 l/100 km (country), \n] 1
[\n, \n, \n144 g CO2/km (comb)\n] 1
Name: Non-smoking Vehicle, dtype: int64

```

In [91]: `df.drop("Non-smoking Vehicle", axis=1, inplace=True)`

Offer Number

In [92]: `df['Offer Number'].value_counts(dropna=False)`

```

-----
TypeError                                 Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'

```

Out[92]:

| | |
|-----------------|------|
| NaN | 3175 |
| [\nXT61649\n] | 27 |
| [\nHMS53619\n] | 27 |
| [\nTP62881\n] | 27 |
| [\nJV03654\n] | 27 |
| ... | |
| [\n9277941\n] | 1 |
| [\n19155-1\n] | 1 |
| [\n3124254\n] | 1 |
| [\nJ1109546\n] | 1 |
| [\n040394-M0\n] | 1 |

Name: Offer Number, Length: 11441, dtype: int64

In [93]: `df.drop("Offer Number", axis=1, inplace=True)`

Body Color

In [94]: `df['Body Color'].value_counts(dropna=False)`

```
-----  
TypeError Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

```
Out[94]: [\n, Black, \n]      3745  
[\n, Grey, \n]      3505  
[\n, White, \n]      3406  
[\n, Silver, \n]     1647  
[\n, Blue, \n]       1431  
[\n, Red, \n]        957  
NaN                597  
[\n, Brown, \n]      289  
[\n, Green, \n]      154  
[\n, Beige, \n]      108  
[\n, Yellow, \n]     51  
[\n, Violet, \n]     18  
[\n, Bronze, \n]     6  
[\n, Orange, \n]     3  
[\n, Gold, \n]        2  
Name: Body Color, dtype: int64
```

```
In [95]: df['Body_Color'] = df['Body Color'].str[1]
```

```
In [96]: df['Body_Color'].value_counts(dropna=False)
```

```
Out[96]: Black      3745  
Grey       3505  
White      3406  
Silver     1647  
Blue       1431  
Red        957  
NaN        597  
Brown      289  
Green      154  
Beige      108  
Yellow     51  
Violet     18  
Bronze     6  
Orange     3  
Gold       2  
Name: Body_Color, dtype: int64
```

```
In [97]: df.drop("Body Color", axis=1, inplace=True)
```

Body Color Original

```
In [98]: df['Body Color Original'].value_counts(dropna=False)
```

```
-----  
TypeError Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

```
Out[98]: NaN                  3759  
[\nOnyx Schwarz\n]            338  
[\nBianco\n]                 282  
[\nMythos schwarz Metallic\n] 238  
[\nBrillantschwarz\n]          216  
...  
[\nBlu cosmo metallizzato\n]   1  
[\nBlac Ivoire D16\n]          1  
[\nkarbonschwarz met\n]         1  
[\nBelugabraun Metallic (bruin me\n] 1  
[\nSATIN STEEL\n]               1  
Name: Body Color Original, Length: 1928, dtype: int64
```

```
In [99]: df.drop("Body Color Original", axis=1, inplace=True)
```

Paint Type

```
In [100... df['Paint Type'].value_counts(dropna=False)
```

```
-----  
TypeError Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

```
Out[100... [\nMetallic\n] 9794  
NaN 5772  
[\nUni/basic\n] 347  
[\nPerl effect\n] 6  
Name: Paint Type, dtype: int64
```

```
In [101... df['Paint_Type'] = df['Paint Type'].str[0].str.strip('\n')
```

```
In [102... df['Paint_Type'].value_counts(dropna=False)
```

```
Out[102... Metallic 9794  
NaN 5772  
Uni/basic 347  
Perl effect 6  
Name: Paint_Type, dtype: int64
```

```
In [103... df.drop("Paint Type", axis=1, inplace=True)
```

Upholstery

```
In [104... df['Upholstery'].value_counts(dropna=False)
```

```
-----  
TypeError Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

```
Out[104... [\nCloth, Black\n] 5821  
NaN 3720  
[\nPart leather, Black\n] 1121  
[\nCloth\n] 1005  
[\nCloth, Grey\n] 891  
[\nCloth, Other\n] 639  
[\nFull leather, Black\n] 575  
[\nBlack\n] 491  
[\nGrey\n] 273  
[\nOther, Other\n] 182  
[\nPart leather\n] 140  
[\nFull leather\n] 139  
[\nFull leather, Brown\n] 116  
[\nPart leather, Grey\n] 116  
[\nOther, Black\n] 110  
[\nFull leather, Other\n] 72  
[\nFull leather, Grey\n] 67  
[\nPart leather, Other\n] 65  
[\nOther\n] 56  
[\nPart leather, Brown\n] 50  
[\nalcantara, Black\n] 47  
[\nFull leather, Beige\n] 36  
[\nVelour, Black\n] 36  
[\nCloth, Brown\n] 28  
[\nVelour\n] 16
```

```
[\"nOther, Grey\\n"]          15
[\"nCloth, Beige\\n"]         13
[\"nCloth, Blue\\n"]          12
[\"nBrown\\n"]                 12
[\"nCloth, White\\n"]          8
[\"nVelour, Grey\\n"]          8
[\"nalcantara, Grey\\n"]       6
[\"nCloth, Red\\n"]            5
[\"nOther, Yellow\\n"]          4
[\"nPart leather, Red\\n"]     3
[\"nBeige\\n"]                  3
[\"nalcantara\\n"]              2
[\"nFull leather, Blue\\n"]    2
[\"nBlue\\n"]                   2
[\"nPart leather, Beige\\n"]    2
[\"nalcantara, Other\\n"]       2
[\"nPart leather, White\\n"]    2
[\"nWhite\\n"]                  2
[\"nFull leather, Red\\n"]      1
[\"nOther, Brown\\n"]            1
[\"nFull leather, White\\n"]    1
[\"nCloth, Orange\\n"]           1
Name: Upholstery, dtype: int64
```

```
In [105...]: df["Upholstery"] = [item[0] if type(item) == list else item for item in df.Upholstery]
```

```
In [106...]: df["Upholstery"] = df.Upholstery.str.strip("\n").str.split(", ", )
```

```
In [107...]: df["Upholstery"].value_counts(dropna=False)
```

```
-----  
TypeError                                                 Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

```
Out[107...]: [Cloth, Black]          5821
NaN                      3720
[Part leather, Black]    1121
[Cloth]                  1005
[Cloth, Grey]             891
[Cloth, Other]            639
[Full leather, Black]    575
[Black]                   491
[Grey]                     273
[Other, Other]            182
[Part leather]            140
[Full leather]            139
[Full leather, Brown]    116
[Part leather, Grey]     116
[Other, Black]             110
[Full leather, Other]    72
[Full leather, Grey]     67
[Part leather, Other]    65
[Other]                     56
[Part leather, Brown]    50
[alcantara, Black]        47
[Full leather, Beige]     36
[Velour, Black]            36
[Cloth, Brown]             28
[Velour]                   16
[Other, Grey]               15
[Cloth, Beige]              13
[Cloth, Blue]                12
[Brown]                     12
[Cloth, White]                 8
[Velour, Grey]                 8
[alcantara, Grey]                6
[Cloth, Red]                   5
[Other, Yellow]                  4
[Part leather, Red]                 3
[Beige]                      3
[alcantara]                    2
[Full leather, Blue]                 2
[Blue]                        2
```

```
[Part leather, Beige]      2
[alcantara, Other]        2
[Part leather, White]     2
[White]                   2
[Full leather, Red]       1
[Other, Brown]            1
[Full leather, White]     1
[Cloth, Orange]           1
Name: Upholstery, dtype: int64
```

```
In [108...]: u_type = ["Cloth", "Part leather", "Full leather", "Velour", "alcantara"]
df["Upholstery_type"] = df["Upholstery"].apply(lambda x : x[0] if type(x) == list and x[0] in u_type else np.nan)
```

```
In [109...]: df["Upholstery_type"].value_counts(dropna=False)
```

```
Out[109...]: Cloth      8423
NaN        4871
Part leather    1499
Full leather    1009
Velour         60
alcantara      57
Name: Upholstery_type, dtype: int64
```

```
dfj.loc[dfj['Upholstery'].str.contains('Full leather', case=False, regex=False, na=False), 'Upholstery'] = 'Full leather'
dfj.loc[dfj['Upholstery'].str.contains('Part leather', case=False, regex=False, na=False), 'Upholstery'] = 'Part leather'
dfj.loc[dfj['Upholstery'].str.contains('Cloth', case=False, regex=False, na=False), 'Upholstery'] = 'Cloth'
dfj.loc[dfj['Upholstery'].str.contains('alcantara', case=False, regex=False, na=False), 'Upholstery'] = 'Alcantara'
dfj.loc[dfj['Upholstery'].isin(["\nBlack\n", "\nOther, Black\n", "\nOther, Other\n", "\nVelour\n", "\nGrey\n", "\nBlue\n", "\nVelour, Black\n", "\nVelour, Grey\n", "\nOther\n", "\nOther, Grey\n", "\nWhite\n", "\nOther, Yellow\n", "\nBeige\n", "\nBrown\n", "\nOther, Brown\n"]), 'Upholstery'] = 'Cloth'
```

```
In [110...]: color = ['Black', 'Grey', 'Brown', 'Beige', 'White', 'Blue', 'Red', 'Yellow', 'Orange']

def finder(x):
    if type(x) == list and len(x) == 2:
        return x[1]
    elif type(x) == list and x[0] in color:
        return x[0]
    else:
        return np.nan

df['Upholstery_color'] = df.Upholstery.apply(finder)
```

```
In [111...]: df["Upholstery_color"].value_counts(dropna=False)
```

```
Out[111...]: Black      8201
NaN        5078
Grey       1376
Other      960
Brown      207
Beige      54
Blue       16
White      13
Red        9
Yellow     4
Orange     1
Name: Upholstery_color, dtype: int64
```

```
In [112...]: df.drop("Upholstery", axis=1, inplace=True)
```

Nr. of Doors

```
In [113...]: df["Nr. of Doors"].value_counts(dropna=False)
```

```
-----  
TypeError                                Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

```
Out[113...]: {\n5\n: 11575,\n4\n: 3079,\n3\n: 832,\n2\n: 219,\nNaN: 212,\n7\n: 1,\n1\n: 1,\nName: Nr. of Doors, dtype: int64}
```

```
In [114...]: df['Nr_of_Doors'] = df['Nr. of Doors'].str[0].str.strip('\n').astype('float')
```

```
In [115...]: df["Nr_of_Doors"].value_counts(dropna=False)
```

```
Out[115...]: 5.00    11575\n4.00    3079\n3.00    832\n2.00    219\nNaN    212\n7.00    1\n1.00    1\nName: Nr_of_Doors, dtype: int64
```

```
In [116...]: df.drop("Nr. of Doors", axis=1, inplace=True)
```

Nr. of Seats

```
In [117...]: df["Nr. of Seats"].value_counts(dropna=False)
```

```
-----\nTypeError: Traceback (most recent call last)\n  pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()\n\nTypeError: unhashable type: 'list'\nException ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'\nTraceback (most recent call last):\n  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations\nTypeError: unhashable type: 'list'
```

```
Out[117...]: {\n5\n: 13336,\n4\n: 1125,\nNaN: 977,\n7\n: 362,\n2\n: 116,\n6\n: 2,\n3\n: 1,\nName: Nr. of Seats, dtype: int64}
```

```
In [118...]: df['Nr_of_Seats'] = df['Nr. of Seats'].str[0].str.strip('\n').astype('float')
```

```
In [119...]: df["Nr_of_Seats"].value_counts(dropna=False)
```

```
Out[119...]: 5.00    13336\n4.00    1125\nNaN    977\n7.00    362\n2.00    116\n6.00    2\n3.00    1\nName: Nr_of_Seats, dtype: int64
```

```
In [120...]: df.drop("Nr. of Seats", axis=1, inplace=True)
```

Model Code

```
In [121...]: df["Model Code"].value_counts(dropna=False)
```

```
-----\nTypeError: Traceback (most recent call last)\n  pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()
```

```
TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'
Out[121...]
NaN          10941
[\n0035/BCB\n]    268
[\n0588/BNO\n]    245
[\n0588/BDB\n]    206
[\n0588/BHX\n]    188
...
[\n0035/BHX\n]      1
[\n0588/AVR\n]      1
[\n0035/ASN\n]      1
[\n1844/AEG\n]      1
[\n3333/BBL\n]      1
Name: Model Code, Length: 233, dtype: int64
```

```
In [122...]: df.drop("Model Code", axis=1, inplace=True)
```

Gearing Type

```
In [123...]: df["Gearing Type"].value_counts(dropna=False)
```

```
-----
TypeError                                                 Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'
Out[123...]
[\n, Manual, \n]        8153
[\n, Automatic, \n]     7297
[\n, Semi-automatic, \n] 469
Name: Gearing Type, dtype: int64
```

```
In [124...]: df['Gearing_Type'] = df['Gearing Type'].str[1]
```

```
In [125...]: df["Gearing_Type"].value_counts(dropna=False)
```

```
Out[125...]
Manual           8153
Automatic        7297
Semi-automatic   469
Name: Gearing_Type, dtype: int64
```

```
In [126...]: df.drop("Gearing Type", axis=1, inplace=True)
```

Displacement

```
In [127...]: df["Displacement"].value_counts(dropna=False)
```

```
-----
TypeError                                                 Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'
Out[127...]
[\n1,598 cc\n]    4761
[\n999 cc\n]       2438
[\n1,398 cc\n]     1314
[\n1,399 cc\n]     749
[\n1,229 cc\n]     677
```

```

[\"n1,956 cc\\n]      670
[\"n1,461 cc\\n]      595
[\"n1,490 cc\\n]      559
NaN                  496
[\"n1,422 cc\\n]      467
[\"n1,197 cc\\n]      353
[\"n898 cc\\n]         351
[\"n1,395 cc\\n]      320
[\"n1,968 cc\\n]      301
[\"n1,149 cc\\n]      288
[\"n1,618 cc\\n]      212
[\"n1,798 cc\\n]      210
[\"n1,498 cc\\n]      196
[\"n1,600 cc\\n]      130
[\"n1,248 cc\\n]      110
[\"n1,997 cc\\n]      103
[\"n1,364 cc\\n]      102
[\"n1,400 cc\\n]      90
[\"n998 cc\\n]         72
[\"n1,500 cc\\n]      50
[\"n2,000 cc\\n]      46
[\"n1,000 cc\\n]      40
[\"n1 cc\\n]           36
[\"n1,998 cc\\n]      25
[\"n2,480 cc\\n]      20
[\"n1,984 cc\\n]      18
[\"n1,200 cc\\n]      18
[\"n1,397 cc\\n]      11
[\"n899 cc\\n]         11
[\"n160 cc\\n]          6
[\"n929 cc\\n]          5
[\"n1,499 cc\\n]      5
[\"n139 cc\\n]          4
[\"n900 cc\\n]          4
[\"n1,596 cc\\n]      4
[\"n997 cc\\n]          4
[\"n1,396 cc\\n]      3
[\"n1,199 cc\\n]      3
[\"n1,599 cc\\n]      3
[\"n995 cc\\n]          2
[\"n1,300 cc\\n]      2
[\"n1,589 cc\\n]      2
[\"n2 cc\\n]           2
[\"n1,495 cc\\n]      2
[\"n1,696 cc\\n]      1
[\"n1,896 cc\\n]      1
[\"n54 cc\\n]           1
[\"n1,995 cc\\n]      1
[\"n15,898 cc\\n]     1
[\"n1,533 cc\\n]      1
[\"n122 cc\\n]          1
[\"n1,350 cc\\n]      1
[\"n1,239 cc\\n]      1
[\"n16,000 cc\\n]     1
[\"n1,198 cc\\n]      1
[\"n996 cc\\n]          1
[\"n1,496 cc\\n]      1
[\"n1,568 cc\\n]      1
[\"n1,333 cc\\n]      1
[\"n1,390 cc\\n]      1
[\"n2,967 cc\\n]      1
[\"n890 cc\\n]          1
[\"n1,580 cc\\n]      1
[\"n1,368 cc\\n]      1
[\"n973 cc\\n]          1
[\"n1,195 cc\\n]      1
[\"n1,100 cc\\n]      1
[\"n1,856 cc\\n]      1
[\"n1,800 cc\\n]      1
[\"n1,686 cc\\n]      1
[\"n1,369 cc\\n]      1
[\"n140 cc\\n]          1
[\"n1,584 cc\\n]      1
Name: Displacement, dtype: int64

```

In [128...]

```
df[\"Displacement\"] = df.Displacement.str[0].str.strip('\\n').str.replace(',', '').str.extract('(\d{1,5})')[0].ast
```

```
dfj['Displacement'] = [i[0] if type(i) == list else i for i in dfj['Displacement']] dfj['Displacement']=dfj['Displacement'].str.replace(" cc", "", regex=False) dfj['Displacement']=dfj['Displacement'].str.replace(",","", regex=False) dfj['Displacement'] = pd.to_numeric(dfj['Displacement'], downcast='float', errors='coerce') dfj['Displacement'] = dfj['Displacement'].astype('Int64') dfj = dfj.rename(columns={'Displacement': 'cc'})
```

In [129...]

```
df[\"Displacement_cc\"] = df[\"Displacement\"]
```

```
In [130... df["Displacement_cc"].value_counts(dropna=False).tail(32)
```

```
Out[130... 1589.00      2
2.00        2
995.00      2
1568.00      1
1584.00      1
16000.00     1
1369.00      1
996.00        1
1686.00      1
1100.00      1
1496.00      1
1195.00      1
1896.00      1
1239.00      1
890.00        1
973.00        1
1198.00      1
1333.00      1
15898.00     1
1995.00      1
2967.00      1
1800.00      1
1350.00      1
1580.00      1
1696.00      1
122.00        1
140.00        1
1368.00      1
1390.00      1
54.00        1
1856.00      1
1533.00      1
Name: Displacement_cc, dtype: int64
```

```
In [131... df.drop("Displacement", axis=1, inplace=True)
```

Cylinders

```
In [132... df["Cylinders"].value_counts(dropna=False)
```

```
-----  
TypeError                                Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

```
Out[132... [\n4\n]    8105
NaN      5680
[\n3\n]    2104
[\n5\n]    22
[\n6\n]    3
[\n2\n]    2
[\n8\n]    2
[\n1\n]    1
Name: Cylinders, dtype: int64
```

```
In [133... df["Cylinders"] = df.Cylinders.str[0].str.strip(' \n').astype("float")
```

```
In [134... df["Cylinders"].value_counts(dropna=False)
```

```
Out[134... 4.00    8105
NaN      5680
3.00    2104
5.00    22
6.00    3
8.00    2
2.00    2
1.00    1
Name: Cylinders, dtype: int64
```

Weight

In [135...]

```
df["Weight"].value_counts(dropna=False)
```

```
-----  
TypeError                                                 Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

Out[135...]

```
NaN          6974  
[\n1,163 kg\n]    574  
[\n1,360 kg\n]    356  
[\n1,165 kg\n]    301  
[\n1,335 kg\n]    242  
...  
[\n1,761 kg\n]      1  
[\n1,331 kg\n]      1  
[\n1,974 kg\n]      1  
[\n1,711 kg\n]      1  
[\n1,686 kg\n]      1  
Name: Weight, Length: 435, dtype: int64
```

In [136...]

```
df["Weight_kg"] = df.Weight.str[0].str.strip('\n').str.replace(',', '').str.extract('(\d{1,6})')[0].astype('float')
```

In [137...]

```
df["Weight_kg"].value_counts(dropna=False)
```

Out[137...]

```
NaN          6974  
1163.00      574  
1360.00      356  
1165.00      301  
1335.00      242  
...  
1161.00      1  
1254.00      1  
1379.00      1  
1575.00      1  
1159.00      1  
Name: Weight_kg, Length: 435, dtype: int64
```

In [138...]

```
df.drop("Weight", axis=1, inplace=True)
```

Drive chain

In [139...]

```
df["Drive chain"].value_counts(dropna=False)
```

```
-----  
TypeError                                                 Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

Out[139...]

```
[\nfront\n]     8886  
NaN           6858  
[\n4WD\n]        171  
[\nrear\n]        4  
Name: Drive chain, dtype: int64
```

In [140...]

```
df['Drive_chain'] = df['Drive chain'].str[0].str.strip('\n')
```

In [141...]

```
df["Drive_chain"].value_counts(dropna=False)
```

Out[141...]

```
front     8886
```

```
NaN      6858
4WD      171
rear      4
Name: Drive_chain, dtype: int64
```

```
In [142...]: df.drop("Drive chain", axis=1, inplace=True)
```

Fuel

```
In [143...]: df["Fuel"].value_counts(dropna=False)
```

```
-----
TypeError                                 Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'
```

```
Out[143...]: [\n, Diesel (Particulate Filter), \n]
4315
[\n, Super 95, \n]
3338
[\n, Gasoline, \n]
3175
[\n, Diesel, \n]
2984
[\n, Super 95 / Regular/Benzine 91, \n]
424
[\n, Regular/Benzine 91, \n]
354
[\n, Super E10 95, \n]
331
[\n, Super 95 (Particulate Filter), \n]
268
[\n, Super 95 / Super E10 95, \n]
131
[\n, Regular/Benzine 91 (Particulate Filter), \n]
100
[\n, Super 95 / Super Plus 98, \n]
85
[\n, Gasoline (Particulate Filter), \n]
77
[\n, Super 95 / Super Plus 98 / Super E10 95 / Super Plus E10 98, \n]
40
[\n, Super E10 95 / Super 95, \n]
38
[\n, CNG, \n]
33
[\n, Super 95 / Super E10 95 (Particulate Filter), \n]
25
[\n, Super E10 95 / Super 95 / Super Plus 98 / Super Plus E10 98, \n]
19
[\n, LPG, \n]
16
[\n, Super 95 / Super Plus 98 (Particulate Filter), \n]
14
[\n, Regular/Benzine 91 / Super 95 / Super Plus 98 / Regular/Benzine E10 91 / Super E10 95 / Super Plus E10 98,
 \n]          13
[\n, Regular/Benzine E10 91, \n]
11
[\n, Liquid petroleum gas (LPG), \n]
10
[\n, Super 95 / Super Plus 98 / Super E10 95, \n]
8
[\n, Super E10 95 (Particulate Filter), \n]
7
[\n, Super 95 / Regular/Benzine 91 / Super Plus 98 / Regular/Benzine E10 91 / Super E10 95 / Super Plus E10 98,
 \n]          7
[\n, Regular/Benzine 91 / Regular/Benzine E10 91, \n]
7
[\n, Super Plus 98, \n]
7
[\n, Super 95 / Super Plus 98 / Super Plus E10 98 / Super E10 95, \n]
5
[\n, Others, \n]
5
[\n, Super 95 / Super Plus 98 / Super E10 95 / Super Plus E10 98 (Particulate Filter), \n]
```

4
[\n, Super 95 / Super E10 95 / Super Plus E10 98 / Super Plus 98, \n]
4
[\n, Regular/Benzine 91 / Super 95 / Super Plus 98 / Regular/Benzine E10 91 / Super Plus E10 98 / Super E10 95,
\n] 4
[\n, Electric/Gasoline, \n]
4
[\n, Super E10 95 / Super Plus E10 98, \n]
4
[\n, CNG (Particulate Filter), \n]
3
[\n, Super E10 95 / Super 95 / Super Plus 98, \n]
3
[\n, Super E10 95 / Super 95 / Super Plus 98 / Super Plus E10 98 (Particulate Filter), \n]
3
[\n, Regular/Benzine 91 / Super 95 / Super Plus 98 / Super Plus E10 98 / Super E10 95 / Regular/Benzine E10 91,
\n] 2
[\n, Super E10 95 / Regular/Benzine 91, \n]
2
[\n, Super 95 / Regular/Benzine 91 / Super Plus 98, \n]
2
[\n, Super Plus E10 98 / Super 95 / Super Plus 98 / Super E10 95, \n]
1
[\n, Regular/Benzine 91 / Super 95, \n]
1
[\n, Regular/Benzine 91 / Super 95 / Super E10 95 / Super Plus E10 98 / Super Plus 98 / Regular/Benzine E10 91,
\n] 1
[\n, Super 95 / Super Plus 98 / Super Plus E10 98, \n]
1
[\n, Super 95 / Regular/Benzine 91 / Super E10 95 / Super Plus E10 98 / Super Plus 98, \n]
1
[\n, Super 95 / Regular/Benzine E10 91 / Super Plus E10 98 / Super E10 95 / Regular/Benzine 91,
\n] 1
[\n, Super 95 / Regular/Benzine 91 / Super E10 95 / Super Plus E10 98 / Super Plus 98 / Regular/Benzine E10 91,
\n] 1
[\n, Super 95 / Regular/Benzine E10 91 / Super E10 95, \n]
1
[\n, Regular/Benzine 91 / Super 95 / Super Plus E10 98 / Super E10 95 / Regular/Benzine E10 91 / Super Plus 98,
\n] 1
[\n, Regular/Benzine E10 91 / Regular/Benzine 91 / Super 95 / Super Plus 98 / Super E10 95 / Super Plus E10 98,
\n] 1
[\n, Super 95 / Super E10 95 / Regular/Benzine 91, \n]
1
[\n, Super 95 / Regular/Benzine 91 / Super E10 95 / Super Plus 98 / Super Plus E10 98 / Regular/Benzine E10 91,
\n] 1
[\n, Domestic gas H, \n]
1
[\n, Regular/Benzine 91 / Super 95 / Regular/Benzine E10 91 / Super E10 95 / Super Plus 98 / Super Plus E10 98,
\n] 1
[\n, Super Plus 98 / Super 95 / Super E10 95, \n]
1
[\n, Others (Particulate Filter), \n]
1
[\n, Super Plus E10 98, \n]
1
[\n, Super 95 / Super E10 95 / Super Plus E10 98, \n]
1
[\n, Super 95 / Biodiesel, \n]
1
[\n, Electric, \n]
1
[\n, Regular/Benzine 91 / Super 95 / Super Plus 98 / Super Plus E10 98 / Super E10 95 / Regular/Benzine E10 91
(Particulate Filter), \n] 1
[\n, Regular/Benzine 91 / Super 95 / Regular/Benzine E10 91, \n]
1
[\n, Regular/Benzine E10 91 / Super E10 95 / Super Plus E10 98 / Super Plus 98 / Super 95 / Regular/Benzine 91,
\n] 1
[\n, Super 95 / Regular/Benzine 91 (Particulate Filter), \n]
1
[\n, Super E10 95 / Regular/Benzine 91 / Regular/Benzine E10 91 / Super 95, \n]
1
[\n, Super 95 / Regular/Benzine E10 91, \n]
1
[\n, Super Plus 98 / Regular/Benzine 91, \n]
1
[\n, Regular/Benzine 91 / Super 95 / Super E10 95, \n]
1
[\n, Super Plus 98 / Super 95 / Super Plus E10 98, \n]
1
[\n, Biogas, \n]
1
[\n, Super Plus 98 / Super E10 95, \n]
1
[\n, Super E10 95 / Regular/Benzine 91 / Super 95 / Super Plus 98 / Regular/Benzine E10 91 / Super Plus E10 98,
\n] 1

```
[\\n, Super 95 / Regular/Benzine 91 / Super Plus 98 / Regular/Benzine E10 91 / Super Plus E10 98 / Super E10 95,  
\\n]  
1  
[\\n, Regular/Benzine 91 / Super Plus 98 / Regular/Benzine E10 91 / Super 95 / Super E10 95 / Super Plus E10 98,  
\\n]  
1  
[\\n, Regular/Benzine 91 / Super 95 / Super E10 95 / Regular/Benzine E10 91, \\n]  
1  
[\\n, Super 95 / Super E10 95 / Super Plus 98, \\n]  
1  
Name: Fuel, dtype: int64
```

```
In [144... df["Fuel"] = df.Fuel.str[1].str.split("/").str[0].str.strip()
```

```
In [145... df["Fuel"].value_counts(dropna=False)
```

```
Out[145... Diesel (Particulate Filter)      4315  
Super 95                          4100  
Gasoline                         3175  
Diesel                           2984  
Regular                          503  
Super E10 95                      402  
Super 95 (Particulate Filter)    268  
Gasoline (Particulate Filter)    77  
CNG                             33  
LPG                            16  
Super Plus 98                     11  
Liquid petroleum gas (LPG)       10  
Super E10 95 (Particulate Filter) 7  
Electric                         5  
Others                           5  
CNG (Particulate Filter)        3  
Super Plus E10 98                2  
Domestic gas H                   1  
Biogas                           1  
Others (Particulate Filter)     1  
Name: Fuel, dtype: int64
```

```
In [146... df["Fuel"] = df.Fuel.str.split("(").str[0].str.strip()
```

```
In [147... df["Fuel"].value_counts(dropna=False)
```

```
Out[147... Diesel              7299  
Super 95            4368  
Gasoline           3252  
Regular            503  
Super E10 95       409  
CNG                36  
LPG                16  
Super Plus 98      11  
Liquid petroleum gas 10  
Others              6  
Electric            5  
Super Plus E10 98   2  
Domestic gas H     1  
Biogas              1  
Name: Fuel, dtype: int64
```

```
In [148... benzine = ["Gasoline", "Super 95", "Regular", "Super E10 95", "Super Plus 98", "Super Plus E10 98", "Others"]  
lpg = ["LPG", "Liquid petroleum gas", "CNG", "Biogas", "Domestic gas H"]  
def fueltype(x):  
    if x in benzine:  
        return "Benzine"  
    elif x in lpg:  
        return "LPG/CNG"  
    else:  
        return x  
df["Fuel"] = df.Fuel.apply(fueltype)
```

```
In [149... df.Fuel.value_counts(dropna=False)
```

```
Out[149... Benzine      8551  
Diesel       7299  
LPG/CNG     64
```

```
Electric      5  
Name: Fuel, dtype: int64
```

In [150...]

```
#alternative method  
#df["fuel_new"] = df.Fuel.str[1]  
#diesel_bool = df["fuel_new"].str.contains("diesel", case = False, regex = True)  
#lpg_bool = df["fuel_new"].str.contains("lpg/cng/bio/domestic/electric", case = False, regex = True)  
#df.loc[diesel_bool, "fuel_new"] = "Diesel"  
#df.loc[lpg_bool, "fuel_new"] = "LPG/CNG"  
#benz = list(df.fuel_new.loc[lambda x : x != "Diesel"][lambda x : x != "LPG/CNG"].index)  
#df.fuel_new.iloc[benz] = "Benzine"
```

LİSTE İÇİNDEKİ DİZİNİN SIFIRINCI İNDEKSİNDEN KURTUKMAK İÇİN AŞAĞIDAKİ KOMUT GİRİLDİ. MUHTEMEL HATALARA KARŞI GERİ DÖNÜŞ YAPABİLMEK İÇİN FARKLI BİR SÜTUN OLUŞTURULDU. [\n, Regular/Benzine 91 / Super 95, \n] ==> Regular/Benzine 91 / Super 95, \n
dfj['Fuel2'] = [i[1:] if type(i) == list else i[1:] for i in dfj['Fuel']] dfj['Fuel2'] = [i[:-1] if type(i) == list else i[:-1] for i in dfj['Fuel2']] # Regular/Benzine 91 / Super 95, \n son indeksinden \n kurtulduk. dfj['Fuel2'] = [i[0] if type(i) == list else i[0] for i in dfj['Fuel2']] # Artık sıfırıncı indeks içinde kalmış olan tüm değerler liste dışına çıkarılmış oldu i[0] iledfj.loc[dfj['Fuel2']].isin(['Electric/Gasoline']), 'Fuel2'] = 'Electric/Benzine'
dfj.loc[dfj['Fuel2']].isin(['Domestic gas H', 'Liquid petroleum gas (LPG)', 'CNG', 'CNG (Particulate Filter)'), 'Fuel2'] = 'LPG/CNG'
dfj.loc[dfj['Fuel2']].isin(['Super 95 / Biodiesel']), 'Fuel2'] = "Benzine/Biodiesel" dfj.loc[dfj['Fuel2']].isin(['Super', 'Regular', '91', '95', '98', 'E10', 'Others', 'Gasoline']), 'Fuel2'] = 'Benzine' dfj.loc[dfj['Fuel2']].isin(['Diesel (Particulate Filter)', 'Diesel']), 'Fuel2'] = 'Diesel'# SPLİT YERİNE REGEX İLE REPLACE YAPILARAK UNIQUE DEĞERLERE ULAŞILDI # Fuel sütunu değerleri 6 kategoride toplandı: Benzine/Biodiesel, Electric, Electric/Benzine, LPG/CNG, Diesel ve Benzine dfj['Fuel2'] = dfj['Fuel2'].str.replace(' ', '', regex=True) dfj.loc[(dfj['Fuel2'].str.startswith('Super', na=False)) | (dfj['Fuel2'].str.startswith('Regular', na=False)) | (dfj['Fuel2'].str.startswith('Gasoline', na=False)) | (dfj['Fuel2'].str.startswith('Others', na=False)), 'Fuel2'] = 'Benzine'

Consumption

In [151...]

```
df["Consumption"].value_counts(dropna=False)
```

TypeError

Traceback (most recent call last)

pandas_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'

Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'

Traceback (most recent call last):

File "pandas_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'

Out[151...]

| NaN | 1906 |
|--|------|
| [[3.9 1/100 km (comb)], [4.1 1/100 km (city)], [3.7 1/100 km (country)]] | 304 |
| [[4.2 1/100 km (comb)], [5 1/100 km (city)], [3.7 1/100 km (country)]] | 276 |
| [[5.4 1/100 km (comb)], [6.8 1/100 km (city)], [4.5 1/100 km (country)]] | 257 |
| [[3.8 1/100 km (comb)], [4.3 1/100 km (city)], [3.5 1/100 km (country)]] | 253 |
| ... | |
| [[3.1 1/100 km (comb)], [], [3.5 1/100 km (country)]] | 1 |
| [[5.2 1/100 km (comb)], [6 1/100 km (city)], [4.8 1/100 km (country)]] | 1 |
| [[5.3 1/100 km (comb)], [6 1/100 km (city)], [5 1/100 km (country)]] | 1 |
| [\n, 6.5 1/100 km (comb), \n, 8.1 1/100 km (city), \n, 5.5 1/100 km (country), \n] | 1 |
| [[50 1/100 km (comb)], [64 1/100 km (city)], [42 1/100 km (country)]] | 1 |

Name: Consumption, Length: 882, dtype: int64

In [152...]

```
def parser1(x):  
    if type(x) == float:          # FLOAT TIPI NAN OLDUĞUNDAN NAN DEĞERLERİ NP.NAN DÖNDÜR DEMEK  
        return np.nan  
    elif type(x[0]) == list:  
        if x[0] != []:  
            return x[0][0]  
        else:  
            return np.nan  
    else:  
        return x[1]  
  
def parser2(x):  
    if type(x) == float:  
        return np.nan  
    elif type(x[0]) == list:  
        if x[1] != []:  
            return x[1][0]  
        else:  
            return np.nan  
    elif x[3].endswith(')':  
        return x[3]  
    else:  
        return np.nan
```

```

def parser3(x):
    if type(x) == float:
        return np.nan
    elif type(x[0]) == list:
        if x[2] != []:
            return x[2][0]
        else:
            return np.nan
    elif type(x[0]) != list and x[3].endswith(')'):
        return x[5]
    else:
        return np.nan

```

In [153...]

```

df['cons_comb'] = df['Consumption'].apply(parser1).str.extract('(\d{1,2}.\d|\d{1,3})')[0].astype("float")
df['cons_city'] = df['Consumption'].apply(parser2).str.extract('(\d{1,2}.\d|\d{1,3})')[0].astype("float")
df['cons_country'] = df['Consumption'].apply(parser3).str.extract('(\d{1,2}.\d|\d{1,3})')[0].astype("float")

```

İLK İKİ RAKAMI ÇEKMEK İÇİN: df["Consumption"] = df["Consumption"].str.extract("(\\d\\d|\\d.\\d|\\d)") # İLK DÖRT RAKAMI ÇEKMEK İÇİN: df["Consumption"] = df["Consumption"].str.extract("(\\d\\d|\\d.\\d|\\d)+(\\d\\d|\\d)") df["Consumption"] = df["Consumption"].str.extract("(^\\d*.\\d*)\\w*/(\\d*)") df["Consumption_1"] = df["Consumption_1"].str.extract('(\f{1,4})')[0].astype('float')

In [154...]

```
df["cons_comb"].value_counts(dropna=False).head()
```

Out[154...]

| | |
|------|------|
| NaN | 2033 |
| 5.40 | 770 |
| 3.90 | 733 |
| 4.00 | 713 |
| 5.10 | 657 |

Name: cons_comb, dtype: int64

In [155...]

```
df.drop('Consumption', axis=1, inplace=True)
```

```

split_df = df['Consumption'].str.split('', expand=True)
split_df.columns = ['Consumption' + f"_{id_}" for id_ in range(len(split_df.columns))]

df = pd.merge(df, split_df, how="left", left_index=True, right_index=True)
df['Consumption_0'] = df['Consumption_0'].str.extract('([...]')
df['Consumption_1'] = df['Consumption_1'].str.extract('([...]')
df['Consumption_2'] = df['Consumption_2'].str.extract('(\f{1,4})')[0].astype('float')

```

CO2 Emission

In [156...]

```
df["CO2_Emission"].value_counts(dropna=False)
```

```

-----
TypeError                                         Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'

```

Out[156...]

| | |
|--------------------------------|------|
| NaN | 1808 |
| [\\n120 g CO2/km (comb)\\n] | 740 |
| [[], [], []] | 628 |
| [\\n99 g CO2/km (comb)\\n] | 545 |
| [\\n97 g CO2/km (comb)\\n] | 537 |
| ... | |
| [\\n990 g CO2/km (comb)\\n] | 1 |
| [\\n51 g CO2/km (comb)\\n] | 1 |
| [\\n14,457 g CO2/km (comb)\\n] | 1 |
| [\\n193 g CO2/km (comb)\\n] | 1 |
| [\\n253 g CO2/km (comb)\\n] | 1 |

Name: CO2_Emission, Length: 124, dtype: int64

In [157...]

```

df["CO2_Emission"] = [item[0] if type(item) == list else item for item in df["CO2_Emission"]]
df["CO2_Emission"] = df["CO2_Emission"].str.strip("\\n").str.rstrip(" g CO2/km (comb)").str.replace(", .").as

```

dfj["CO2_Emission"] = dfj["CO2_Emission"].str.extract('(\S{1,5})')

In [158...]

```
df["CO2_Emission"].value_counts(dropna=False)
```

Out[158...]

| | |
|--------|------|
| NaN | 2436 |
| 120.00 | 740 |

```
99.00      545
97.00      537
104.00     501
...
160.00      1
193.00      1
165.00      1
239.00      1
45.00       1
Name: CO2_Emission, Length: 121, dtype: int64
```

```
In [159... df.drop("CO2_Emission", axis=1, inplace=True)
```

Emission Class

```
In [160... df["Emission Class"].value_counts(dropna=False)
```

```
-----
TypeError                                 Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'
```

```
Out[160... [\nEuro 6\n]          10139
NaN                 3021
[\nEuro 6d-TEMP\n]    1845
[], [], []           607
[\nEuro 6c\n]         127
[\nEuro 5\n]          78
[\nEuro 6d\n]         62
[\nEuro 4\n]          40
Name: Emission Class, dtype: int64
```

```
In [161... df["Emission_Class"] = [item[0] if type(item) == list else item for item in df["Emission Class"]]
df["Emission_Class"] = df["Emission_Class"].str.strip("\n")
```

```
dfj['Emission Class'] = dfj['Emission Class'].str.extract('(\d)')
```

```
In [162... df["Emission_Class"].value_counts(dropna=False)
```

```
Out[162... Euro 6          10139
NaN             3628
Euro 6d-TEMP    1845
Euro 6c          127
Euro 5           78
Euro 6d          62
Euro 4           40
Name: Emission_Class, dtype: int64
```

DOĞRUDAN DF'YE REPLACE UYGULAMAYA ÖRNEK:

```
In [163... df.replace({"Emission_Class" : {"Euro 6d-TEMP": "Euro 6", "Euro 6c": "Euro 6", "Euro 6d": "Euro 6"}}, regex = True)
```

```
df.loc[df['Emission Class'].isin(['Euro 6', 'Euro 6d-TEMP', 'Euro 6c', 'Euro 6d']), 'Emission Class'] = '6'df['Emission Class'] =
pd.to_numeric(df['Emission Class'], downcast='float', errors='coerce')
```

```
In [164... df["Emission_Class"].value_counts(dropna=False)
```

```
Out[164... Euro 6          12173
NaN             3628
Euro 5           78
Euro 4           40
Name: Emission_Class, dtype: int64
```

```
In [165... df.drop("Emission Class", axis=1, inplace=True)
```

Emission Label

```
In [166... df["Emission Label"].value_counts(dropna=False)

-----
TypeError Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'

Out[166... NaN 11934
[\n4 (Green)\n] 3553
[\n1 (No sticker)\n] 381
[[], [], []] 40
[\n5 (Blue)\n] 8
[\n3 (Yellow)\n] 2
[\n2 (Red)\n] 1
Name: Emission Label, dtype: int64

In [167... df['Emission Label'] = [i[0] if type(i) == list else i for i in df['Emission Label']]

In [168... df['Emission Label'] = df['Emission Label'].str.extract('(\d)')

In [169... df['Emission Label'].value_counts()

Out[169... 4 3553
1 381
5 8
3 2
2 1
Name: Emission Label, dtype: int64

In [170... df.loc[df['Emission Label'].isin(['4 (Green)']), 'Emission Label'] = '4'
df.loc[df['Emission Label'].isin(['5 (Blue)']), 'Emission Label'] = '5'
df.loc[df['Emission Label'].isin(['3 (Yellow)']), 'Emission Label'] = '3'
df.loc[df['Emission Label'].isin(['2 (Red)']), 'Emission Label'] = '2'
df.loc[df['Emission Label'].isin(['1 (No sticker)']), 'Emission Label'] = '1'

In [171... df['Emission Label'].value_counts()

Out[171... 4 3553
1 381
5 8
3 2
2 1
Name: Emission Label, dtype: int64

In [172... # df.drop("Emission Label", axis=1, inplace=True)      # BU DÜŞÜRÜLMEDEN DE İSTİFADE EDİLEBİLİR BENCE.
```

Gears

```
In [173... df["Gears"].value_counts(dropna=False)

-----
TypeError Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'

Out[173... [\n6\n] 5822
NaN 4712
[\n5\n] 3239
[\n7\n] 1908
[\n8\n] 224
[\n9\n] 6
```

```
[\n4\n]      2
[\n3\n]      2
[\n1\n]      2
[\n50\n]     1
[\n2\n]      1
Name: Gears, dtype: int64
```

```
In [174...]: df["Gears"] = df.Gears.str[0].str.strip('\n')
```

```
In [175...]: df["Gears"].value_counts(dropna=False)
```

```
Out[175...]: 6      5822
NaN    4712
5      3239
7      1908
8      224
9      6
1      2
3      2
4      2
50     1
2      1
Name: Gears, dtype: int64
```

Country version

```
In [176...]: df["Country version"].value_counts(dropna=False)
```

```
-----
TypeError                                         Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'
```

```
Out[176...]: NaN          8333
[\nGermany\n]      4502
[\nItaly\n]        1038
[\nEuropean Union\n] 507
[\nNetherlands\n]   464
[\nSpain\n]         325
[\nBelgium\n]       314
[\nAustria\n]       208
[\nCzech Republic\n] 52
[\nPoland\n]        49
[\nFrance\n]        38
[\nDenmark\n]       33
[\nHungary\n]        28
[\nJapan\n]          8
[\nSlovakia\n]      4
[\nCroatia\n]       4
[\nSweden\n]        3
[\nBulgaria\n]      2
[\nRomania\n]       2
[\nEgypt\n]          1
[\nSerbia\n]          1
[\nLuxembourg\n]     1
[\nSwitzerland\n]    1
[\nSlovenia\n]       1
Name: Country version, dtype: int64
```

```
In [177...]: df.drop("Country version", axis=1, inplace=True)
```

Comfort_Convenience

```
In [178...]: df["Comfort_Convenience"].value_counts(dropna=False).head()
```

```
-----
TypeError                                         Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()
```

```
TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'
Out[178...]
NaN
920
[Air conditioning, Electrical side mirrors, Hill Holder, Power windows]
216
[Air conditioning, Electrical side mirrors, Power windows]
134
[Air conditioning, Power windows]
130
[Air conditioning, Armrest, Automatic climate control, Cruise control, Electrical side mirrors, Leather steering wheel, Light sensor, Lumbar support, Multi-function steering wheel, Navigation system, Park Distance Control, Parking assist system sensors front, Parking assist system sensors rear, Power windows, Rain sensor, Seat heating, Start-stop system]    105
Name: Comfort_Convenience, dtype: int64
```

","join() ile ÇOK SAYIDAKİ DEĞİŞKENİ TEKE DÜŞÜRMEK

```
In [179...]
df["Comfort_Convenience"] = [",".join(item) if type(item) == list else item for item in df["Comfort_Convenience"]]

In [180...]
df["Comfort_Convenience"].value_counts(dropna=False).head()
```

```
Out[180...]
NaN
920
Air conditioning,Electrical side mirrors,Hill Holder,Power windows
216
Air conditioning,Electrical side mirrors,Power windows
134
Air conditioning,Power windows
130
Air conditioning,Armrest,Automatic climate control,Cruise control,Electrical side mirrors,Leather steering wheel,Light sensor,Lumbar support,Multi-function steering wheel,Navigation system,Park Distance Control,Parking assist system sensors front,Parking assist system sensors rear,Power windows,Rain sensor,Seat heating,Start-stop system    105
Name: Comfort_Convenience, dtype: int64
```

Entertainment_Media

```
In [181...]
df["Entertainment_Media"].value_counts(dropna=False).head()

-----
TypeError                                 Traceback (most recent call last)
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()

TypeError: unhashable type: 'list'
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'
Traceback (most recent call last):
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations
TypeError: unhashable type: 'list'
Out[181...]
NaN                                         1374
[Bluetooth, Hands-free equipment, On-board computer, Radio, USB]      1282
[Bluetooth, Hands-free equipment, MP3, On-board computer, Radio, USB]    982
[Bluetooth, CD player, Hands-free equipment, MP3, On-board computer, Radio, USB] 783
[On-board computer, Radio]                                         487
Name: Entertainment_Media, dtype: int64

In [182...]
df["Entertainment_Media"] = [",".join(item) if type(item) == list else item for item in df["Entertainment_Media"]]

In [183...]
df["Entertainment_Media"].value_counts(dropna=False).head()

Out[183...]
NaN                                         1374
Bluetooth,Hands-free equipment,On-board computer,Radio,USB            1282
Bluetooth,Hands-free equipment,MP3,On-board computer,Radio,USB          982
Bluetooth,CD player,Hands-free equipment,MP3,On-board computer,Radio,USB 783
On-board computer,Radio                                         487
Name: Entertainment_Media, dtype: int64
```

Extras

In [184...]

```
df["Extras"].value_counts(dropna=False).head()
```

```
-----  
TypeError                                                 Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

Out[184...]

```
[Alloy wheels]           3245  
NaN                   2962  
[Alloy wheels, Touch screen] 697  
[Alloy wheels, Voice Control] 577  
[Alloy wheels, Touch screen, Voice Control] 541  
Name: Extras, dtype: int64
```

In [185...]

```
df["Extras"] = [",".join(item) if type(item) == list else item for item in df["Extras"]]
```

In [186...]

```
df["Extras"].value_counts(dropna=False).head()
```

Out[186...]

```
Alloy wheels           3245  
NaN                   2962  
Alloy wheels,Touch screen 697  
Alloy wheels,Voice Control 577  
Alloy wheels,Touch screen,Voice Control 541  
Name: Extras, dtype: int64
```

Safety_Security

In [187...]

```
df["Safety_Security"].value_counts(dropna=False).head()
```

```
-----  
TypeError                                                 Traceback (most recent call last)  
pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.map_locations()  
  
TypeError: unhashable type: 'list'  
Exception ignored in: 'pandas._libs.index.IndexEngine._call_map_locations'  
Traceback (most recent call last):  
  File "pandas\_libs\hashtable_class_helper.pxi", line 4588, in pandas._libs.hashtable.PyObjectHashTable.map_locations  
TypeError: unhashable type: 'list'
```

Out[187...]

```
NaN          982  
[ABS, Central door lock, Daytime running lights, Driver-side airbag, Electronic stability control, Fog lights, Immobilizer, Isofix, Passenger-side airbag, Power steering, Side airbag, Tire pressure monitoring system, Traction control] 538  
[ABS, Central door lock, Daytime running lights, Driver-side airbag, Electronic stability control, Immobilizer, Isofix, Passenger-side airbag, Power steering, Side airbag, Tire pressure monitoring system, Traction control] 480  
[ABS, Central door lock, Daytime running lights, Driver-side airbag, Electronic stability control, Immobilizer, Isofix, Passenger-side airbag, Power steering, Side airbag, Tire pressure monitoring system, Traction control, Xenon headlights] 275  
[ABS, Central door lock, Daytime running lights, Driver-side airbag, Electronic stability control, Fog lights, Immobilizer, Isofix, LED Daytime Running Lights, Passenger-side airbag, Power steering, Side airbag, Tire pressure monitoring system, Traction control] 272  
Name: Safety_Security, dtype: int64
```

In [188...]

```
df["Safety_Security"] = [",".join(item) if type(item) == list else item for item in df["Safety_Security"]]
```

In [189...]

```
df["Safety_Security"].value_counts(dropna=False).head()
```

Out[189...]

```
NaN          982  
ABS,Central door lock,Daytime running lights,Driver-side airbag,Electronic stability control,Fog lights,Immobilizer,Isofix,Passenger-side airbag,Power steering,Side airbag,Tire pressure monitoring system,Traction control 538  
ABS,Central door lock,Daytime running lights,Driver-side airbag,Electronic stability control,Immobilizer,Isofix,Passenger-side airbag,Power steering,Side airbag,Tire pressure monitoring system,Traction control
```

480

ABS,Central door lock,Daytime running lights,Driver-side airbag,Electronic stability control,Immobilizer,Isofix,Passenger-side airbag,Power steering,Side airbag,Tire pressure monitoring system,Traction control,Xenon headlights

275

ABS,Central door lock,Daytime running lights,Driver-side airbag,Electronic stability control,Fog lights,Immobilizer,Isofix,LED Daytime Running Lights,Passenger-side airbag,Power steering,Side airbag,Tire pressure monitoring system,Traction control

272

Name: Safety_Security, dtype: int64

The End of Part-1

In [190...]

df.shape

Out[190...](15919, 36)

In [191...]

df.head().T

Out[191...]

| | 0 | 1 | 2 | 3 |
|----------------------------|---|---|---|---|
| make_model | Audi A1 | Audi A1 | Audi A1 | Audi A1 |
| body_type | Sedans | Sedans | Sedans | Sedans |
| price | 15770 | 14500 | 14640 | 14500 |
| vat | VAT deductible | Price negotiable | VAT deductible | None |
| km | NaN | NaN | NaN | NaN |
| registration | 2016-01-01 00:00:00 | 2017-03-01 00:00:00 | 2016-02-01 00:00:00 | 2016-08-01 00:00:00 |
| Type | Used | Used | Used | Used |
| Warranty | NaN | NaN | NaN | NaN |
| Cylinders | 3.00 | 4.00 | NaN | 3.00 |
| Fuel | Diesel | Benzine | Diesel | Diesel |
| Emission Label | NaN | 4 | 4 | NaN |
| Gears | NaN | 7 | NaN | 6 |
| Comfort_Convenience | Air conditioning,Armrest,Automatic climate con... | Air conditioning,Automatic climate control,Hil... | Air conditioning,Cruise control,Electrical sid... | Air suspension,Armrest,Auxiliary heating,Elect... |
| Entertainment_Media | Bluetooth,Hands-free equipment,On-board comput... | Bluetooth,Hands-free equipment,On-board comput... | MP3,On-board computer | Bluetooth,CD player,Hands-free equipment,MP3,O... |
| Extras | Alloy wheels,Catalytic Converter,Voice Control | Alloy wheels,Sport seats,Sport suspension,Voi... | Alloy wheels,Voice Control | Alloy wheels,Sport seats,Voice Control |
| Safety_Security | ABS,Central door lock,Daytime running lights,D... | ABS,Central door lock,Central door lock with r... | ABS,Central door lock,Daytime running lights,D... | ABS,Alarm system,Central door lock with remote... |
| age | 3.00 | 2.00 | 3.00 | 3.00 |
| Previous_Owners | 2.00 | NaN | 1.00 | 1.00 |
| hp_kW | 66.00 | 141.00 | 85.00 | 66.00 |
| Type1 | Used | Used | Used | Used |
| Inspection_new | Yes | NaN | NaN | NaN |
| Body_Color | Black | Red | Black | Brown |
| Paint_Type | Metallic | NaN | Metallic | Metallic |
| Upholstery_type | Cloth | Cloth | Cloth | NaN |
| Upholstery_color | Black | Grey | Black | NaN |
| Nr_of_Doors | 5.00 | 3.00 | 4.00 | 3.00 |
| Nr_of_Seats | 5.00 | 4.00 | 4.00 | 4.00 |

| | 0 | 1 | 2 | 3 |
|------------------------|-----------|-----------|-----------|-----------|
| Gearing_Type | Automatic | Automatic | Automatic | Automatic |
| Displacement_cc | 1422.00 | 1798.00 | 1598.00 | 1422.00 |
| Weight_kg | 1220.00 | 1255.00 | NaN | 1195.00 |
| Drive_chain | front | front | front | NaN |
| cons_comb | 3.80 | 5.60 | 3.80 | 3.80 |
| cons_city | 4.30 | 7.10 | 4.40 | 4.30 |
| cons_country | 3.50 | 4.70 | 3.40 | 3.50 |
| CO2_Emission | 99.00 | 129.00 | 99.00 | 99.00 |
| Emission_Class | Euro 6 | Euro 6 | Euro 6 | Euro 6 |

In [192...]

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15919 entries, 0 to 15918
Data columns (total 36 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   make_model      15919 non-null   object  
 1   body_type       15859 non-null   object  
 2   price           15919 non-null   int64  
 3   vat              11406 non-null   object  
 4   km               0 non-null     float64 
 5   registration    14322 non-null   datetime64[ns]
 6   Type             15917 non-null   object  
 7   Warranty         4853 non-null   float64 
 8   Cylinders        10239 non-null   float64 
 9   Fuel              15919 non-null   object  
 10  Emission Label  3945 non-null   object  
 11  Gears            11207 non-null   object  
 12  Comfort_Convenience 14999 non-null   object  
 13  Entertainment_Media 14545 non-null   object  
 14  Extras           12957 non-null   object  
 15  Safety_Security 14937 non-null   object  
 16  age               14322 non-null   float64 
 17  Previous_Owners 9279 non-null   float64 
 18  hp_kw            15831 non-null   float64 
 19  Type1            15917 non-null   category 
 20  Inspection_new  3932 non-null   object  
 21  Body_Color       15322 non-null   object  
 22  Paint_Type       10147 non-null   object  
 23  Upholstery_type 11048 non-null   object  
 24  Upholstery_color 10841 non-null   object  
 25  Nr_of_Doors      15707 non-null   float64 
 26  Nr_of_Seats      14942 non-null   float64 
 27  Gearing_Type     15919 non-null   object  
 28  Displacement_cc 15423 non-null   float64 
 29  Weight_kg        8945 non-null   float64 
 30  Drive_chain      9061 non-null   object  
 31  cons_comb        13886 non-null   float64 
 32  cons_city        13483 non-null   float64 
 33  cons_country     13543 non-null   float64 
 34  CO2_Emission     13483 non-null   float64 
 35  Emission_Class   12291 non-null   object  
dtypes: category(1), datetime64[ns](1), float64(14), int64(1), object(19)
memory usage: 4.3+ MB
```

In [193...]

df.select_dtypes(include='object').describe().T

Out[193...]

| | count | unique | top | freq |
|-------------------|-------|--------|----------------|-------|
| make_model | 15919 | 9 | Audi A3 | 3097 |
| body_type | 15859 | 9 | Sedans | 7903 |
| vat | 11406 | 2 | VAT deductible | 10980 |
| Type | 15917 | 5 | Used | 11096 |
| Fuel | 15919 | 4 | Benzine | 8551 |

| | count | unique | top | freq |
|----------------------------|-------|--------|---|-------|
| Emission Label | 3945 | 5 | 4 | 3553 |
| Gears | 11207 | 10 | 6 | 5822 |
| Comfort_Convenience | 14999 | 6198 | Air conditioning,Electrical side mirrors,Hill ... | 216 |
| Entertainment_Media | 14545 | 346 | Bluetooth,Hands-free equipment,On-board comput... | 1282 |
| Extras | 12957 | 659 | Alloy wheels | 3245 |
| Safety_Security | 14937 | 4443 | ABS,Central door lock,Daytime running lights,D... | 538 |
| Inspection_new | 3932 | 1 | Yes | 3932 |
| Body_Color | 15322 | 14 | Black | 3745 |
| Paint_Type | 10147 | 3 | Metallic | 9794 |
| Upholstery_type | 11048 | 5 | Cloth | 8423 |
| Upholstery_color | 10841 | 10 | Black | 8201 |
| Gearing_Type | 15919 | 3 | Manual | 8153 |
| Drive_chain | 9061 | 3 | front | 8886 |
| Emission_Class | 12291 | 3 | Euro 6 | 12173 |

TO_CSV İÇİN INDEX=False YAZMAZSAK İLAVE İNDEKS SÜTUNU OLUŞTURUR.

```
In [196]: df.to_csv("clean_scout.csv", index=False)
```

INDEX=False YAZILMADAN KAYDEDİLMİŞ, İLAVE İNDEKSLİ DOSYAYI AÇMAK İÇİN:

```
In [195]: # pd.read_csv("clean_scout.csv", index_col=None)
```