

description

February 23, 2024

1

```
[ ]: metric_vars = ['year', 'price', 'owner_count', 'engine_volume_lites',  
    ↪ 'power_in_hp', 'mileage', 'mileage_per_year']  
import pandas as pd  
  
data = pd.read_csv('extended_data.csv', sep = ';', index_col=0)
```

```
[ ]: # -  
from scipy import stats  
  
vars_normal_dict = dict()  
for x in metric_vars:  
    stat, p_value = stats.kstest(data.dropna(subset=[x])[x], 'norm',  
    ↪ args=(data[x].mean(), data[x].std()))  
    vars_normal_dict[x] = {"stat": stat, "p_value": p_value}  
  
pd.DataFrame.from_dict(vars_normal_dict)
```

```
[ ]:      year      price  owner_count  engine_volume_lites  power_in_hp  \  
stat      0.169712  0.193456      0.202267          0.247702      0.167191  
p_value    0.000000  0.000000      0.000000          0.000000      0.000000  
  
      mileage  mileage_per_year  
stat      1.293014e-01      6.916734e-02  
p_value    2.386274e-216      4.679091e-62
```

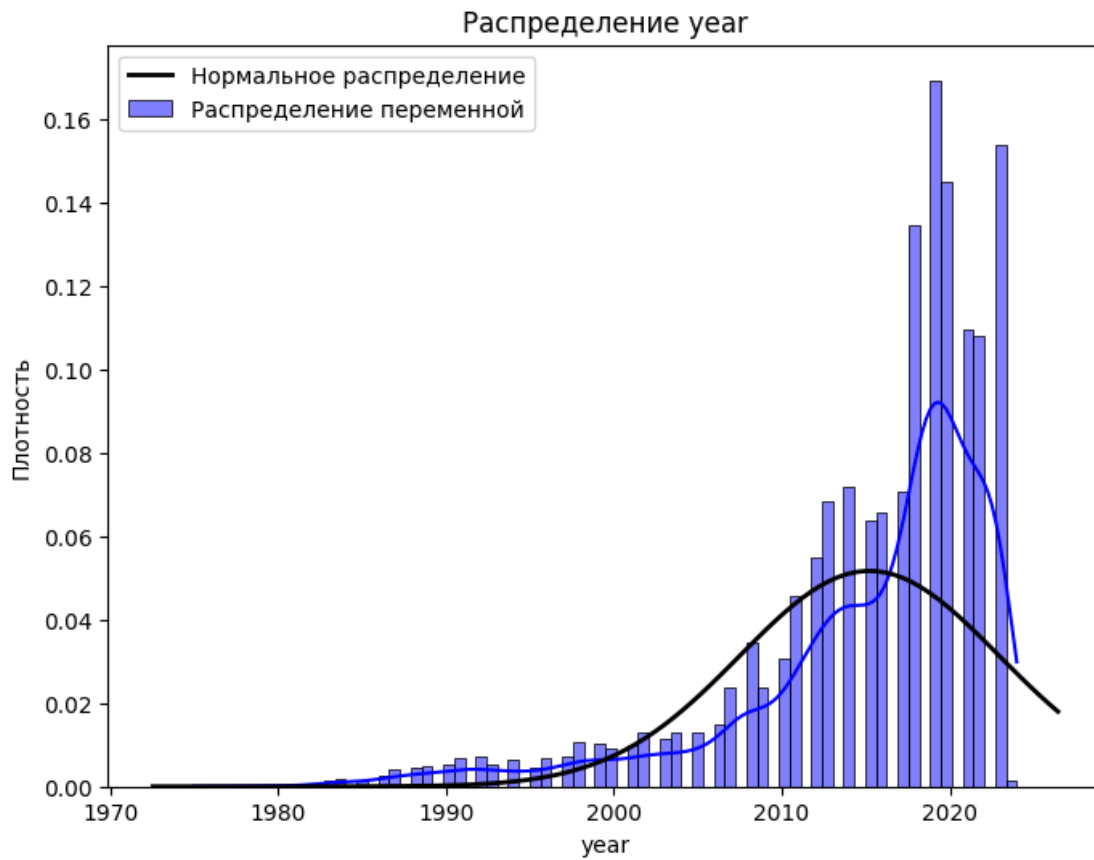
```
[ ]: #  
from matplotlib import pyplot as plt  
import seaborn as sns  
import numpy as np  
  
for var in metric_vars:  
    plt.figure(figsize=(8, 6))
```

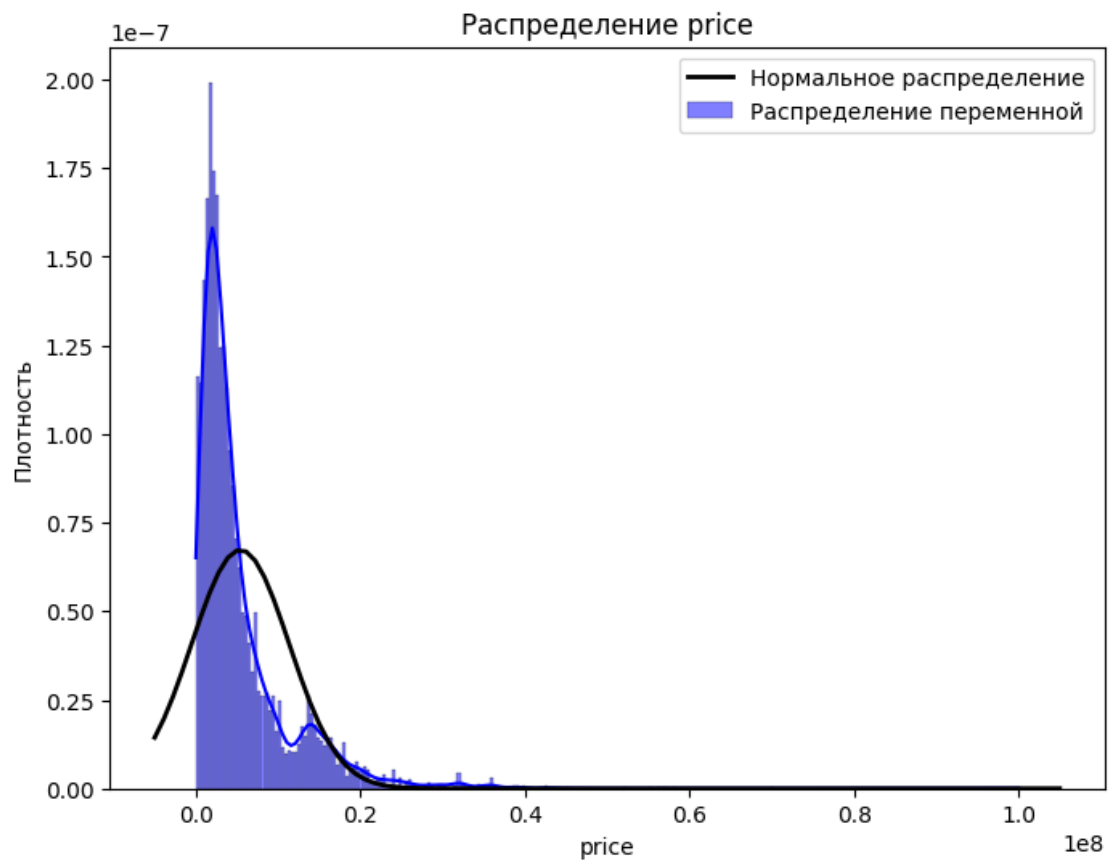
```

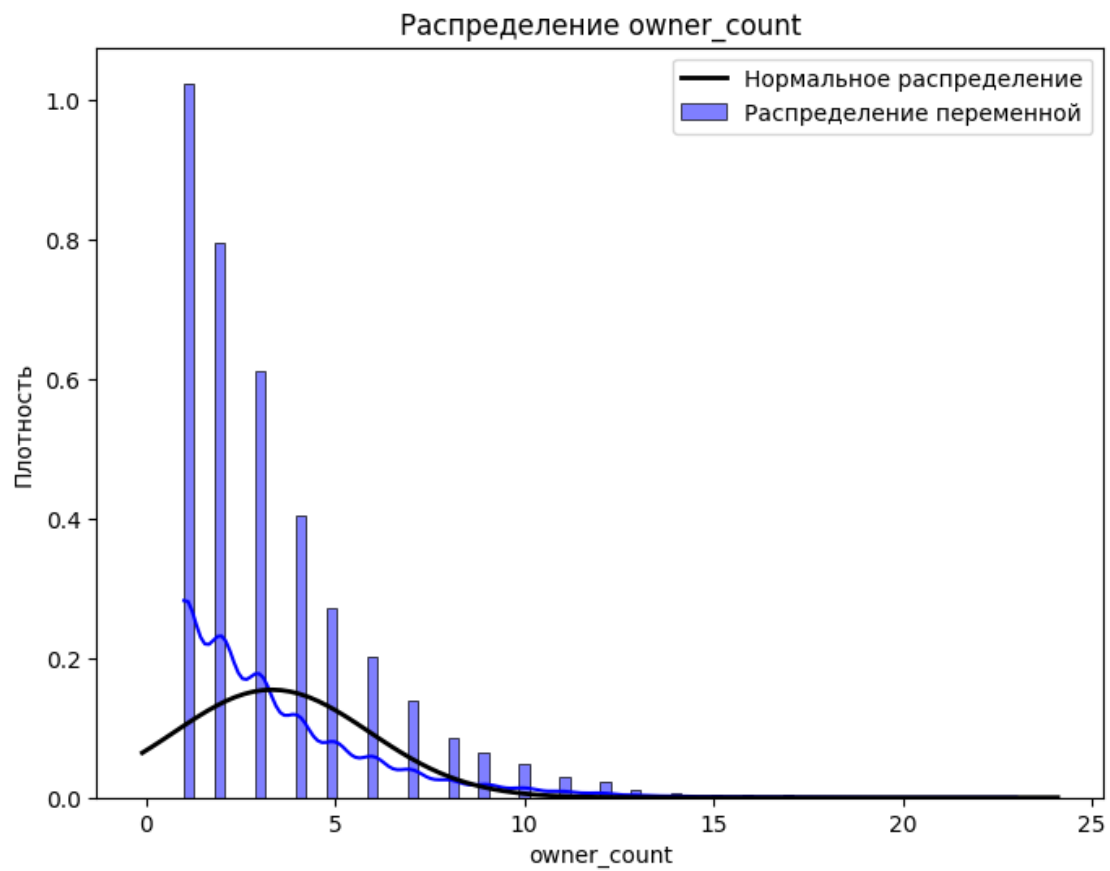
sns.histplot(data[var], kde=True, color='blue', label='year',
stat='density')
plt.xlabel(var)
plt.ylabel('year')
plt.title(f'year')

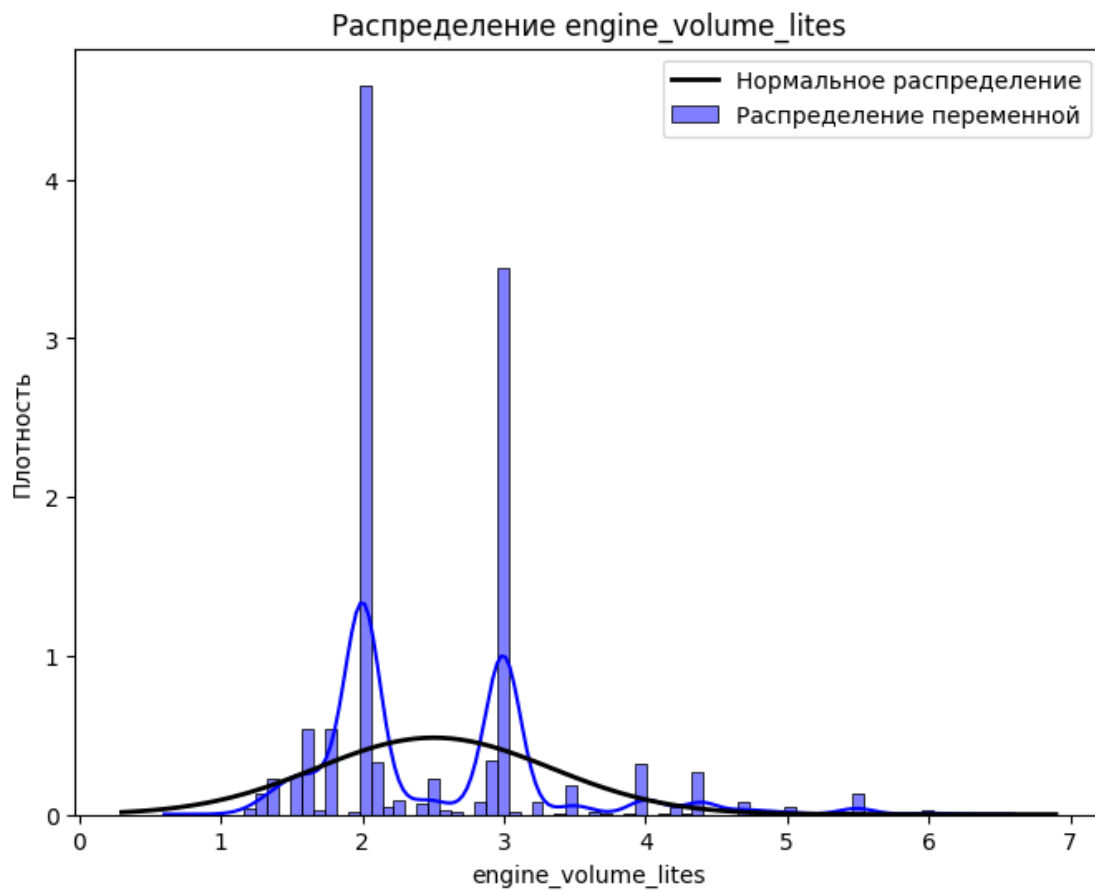
#
xmin, xmax = plt.xlim()
x = np.linspace(xmin, xmax, 100)
p = stats.norm.pdf(x, data[var].mean(), data[var].std())
plt.plot(x, p, 'k', linewidth=2, label='Normal distribution')
plt.legend()
plt.show()

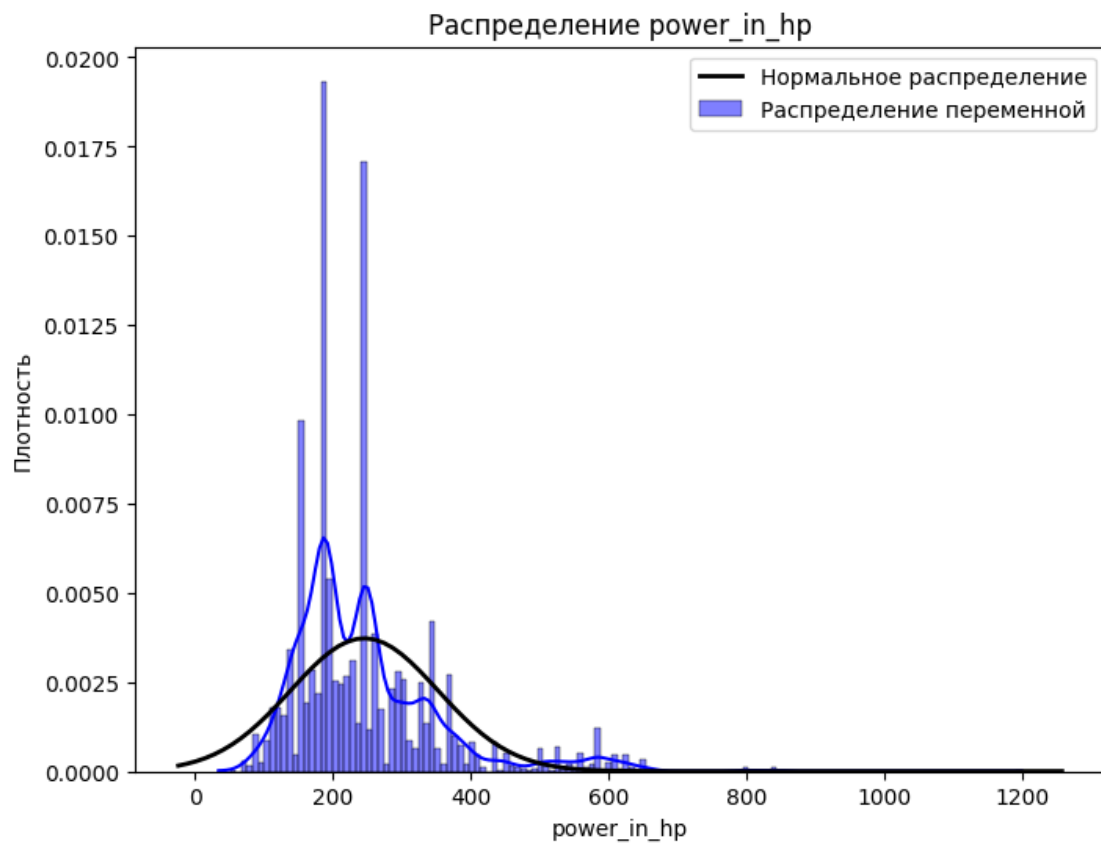
```

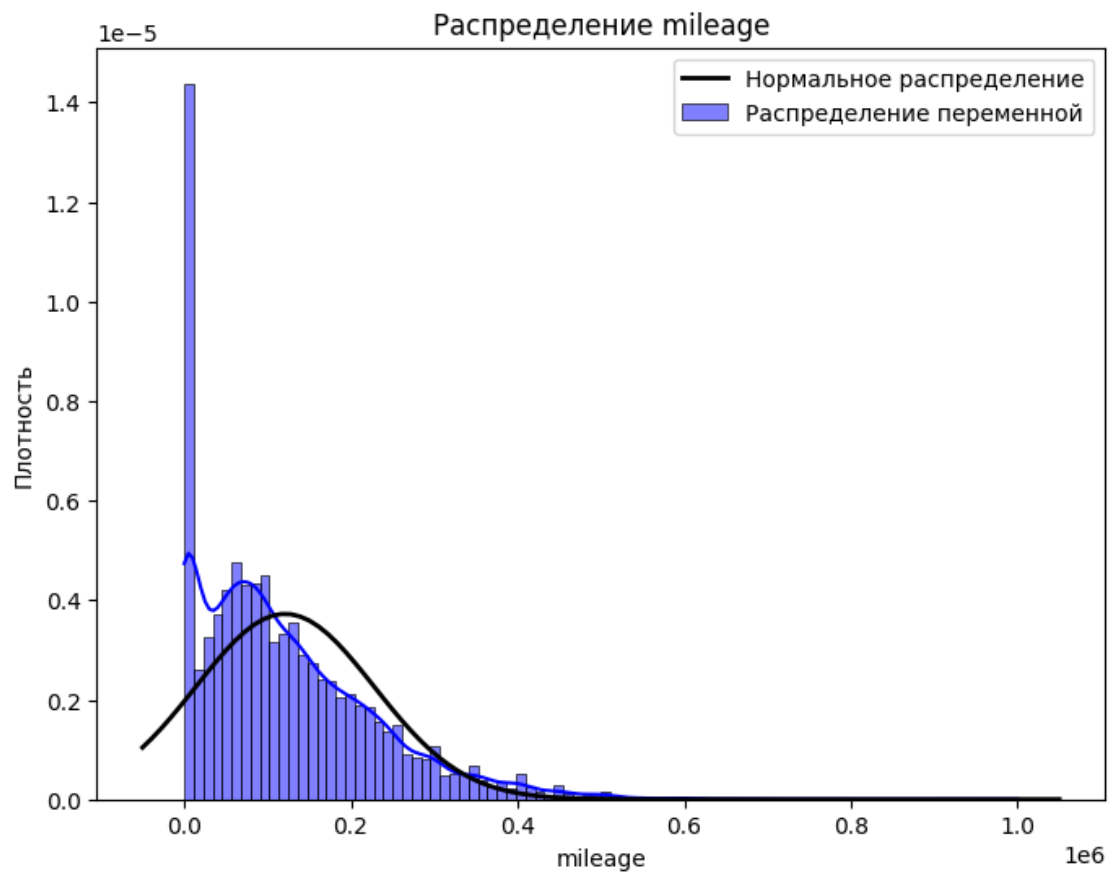


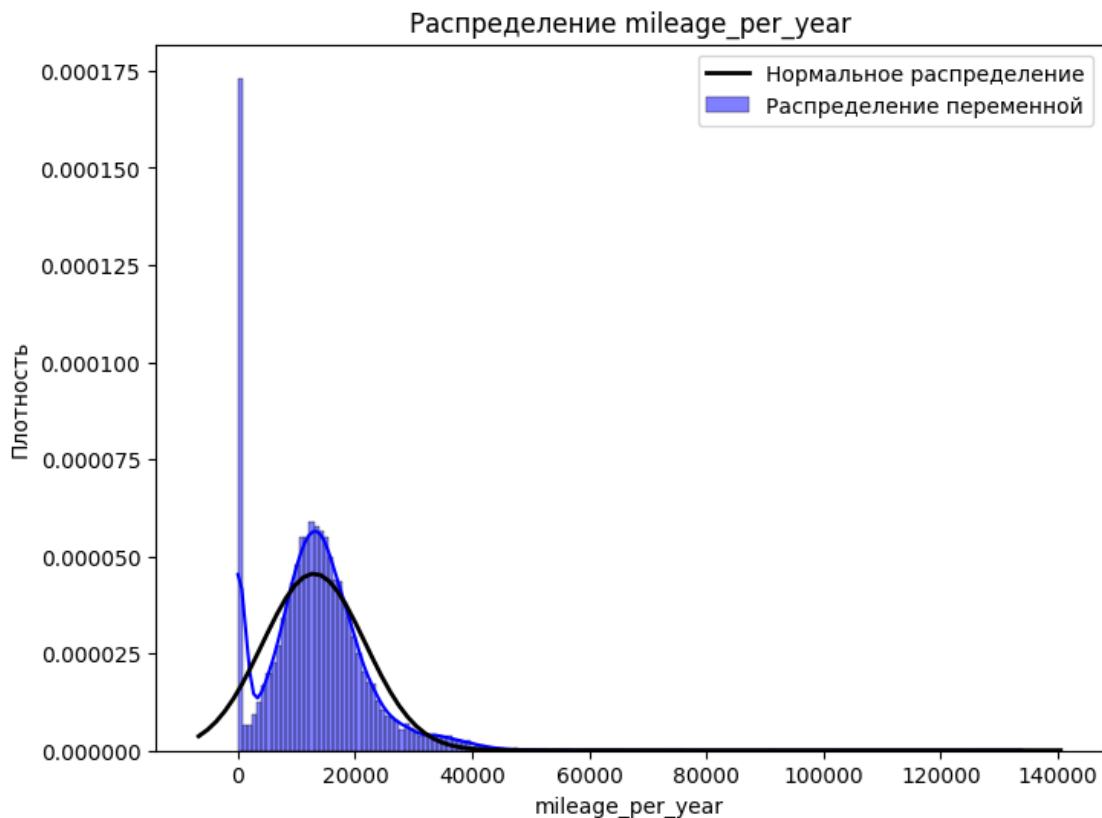












```
[ ]: for i in metric_vars:
      for i2 in metric_vars:
          if i != i2:
              data2 = data.dropna(subset=[i, i2])
              print(f'                {i} {i2}:')
              display(stats.pearsonr(data2[i], data2[i2]))
              print(f'                {i} {i2}:')
              display(stats.spearmanr(data2[i], data2[i2]))
              data2.plot(i, i2, kind = 'scatter')
```

year price:

PearsonRResult(statistic=0.5525368474565608, pvalue=0.0)

year price:

SignificanceResult(statistic=0.9050411832043884, pvalue=0.0)

year owner_count:

PearsonRResult(statistic=-0.6172569685516448, pvalue=0.0)

year owner_count:

SignificanceResult(statistic=-0.6828460506877803, pvalue=0.0)


```

        year    engine_volume_lites:
PearsonRResult(statistic=0.005815596876976406, pvalue=0.47939449804327067)

        year    engine_volume_lites:
SignificanceResult(statistic=0.0896295836490386, pvalue=9.008615904505249e-28)

        year    power_in_hp:
PearsonRResult(statistic=0.3447754255884183, pvalue=0.0)

        year    power_in_hp:
SignificanceResult(statistic=0.405756341051799, pvalue=0.0)

        year    mileage:
PearsonRResult(statistic=-0.7829249617899781, pvalue=0.0)

        year    mileage:
SignificanceResult(statistic=-0.8670809379021445, pvalue=0.0)

        year    mileage_per_year:
PearsonRResult(statistic=-0.10723260939338589, pvalue=3.912333499214111e-39)

        year    mileage_per_year:
SignificanceResult(statistic=-0.3291754154136661, pvalue=0.0)

        price   year:
PearsonRResult(statistic=0.5525368474565608, pvalue=0.0)

        price   year:
SignificanceResult(statistic=0.9050411832043884, pvalue=0.0)

        price   owner_count:
PearsonRResult(statistic=-0.40462729177728524, pvalue=0.0)

        price   owner_count:
SignificanceResult(statistic=-0.6003620562432049, pvalue=0.0)

        price   engine_volume_lites:
PearsonRResult(statistic=0.39323640951055316, pvalue=0.0)

        price   engine_volume_lites:
SignificanceResult(statistic=0.36355601074591776, pvalue=0.0)

        price   power_in_hp:
PearsonRResult(statistic=0.6728708652271249, pvalue=0.0)

        price   power_in_hp:
SignificanceResult(statistic=0.6317606699106313, pvalue=0.0)

```

```

    price    mileage:
PearsonRResult(statistic=-0.5890870461586594, pvalue=0.0)
    price    mileage:
SignificanceResult(statistic=-0.8351638977792649, pvalue=0.0)
    price    mileage_per_year:
PearsonRResult(statistic=-0.4704577297092854, pvalue=0.0)
    price    mileage_per_year:
SignificanceResult(statistic=-0.3654517173159595, pvalue=0.0)
    owner_count    year:
PearsonRResult(statistic=-0.6172569685516448, pvalue=0.0)
    owner_count    year:
SignificanceResult(statistic=-0.6828460506877801, pvalue=0.0)
    owner_count    price:
PearsonRResult(statistic=-0.40462729177728524, pvalue=0.0)
    owner_count    price:
SignificanceResult(statistic=-0.6003620562432049, pvalue=0.0)
    owner_count    engine_volume_lites:
PearsonRResult(statistic=0.13681224671237682, pvalue=1.8088408024098546e-46)
    owner_count    engine_volume_lites:
SignificanceResult(statistic=0.07871559518054935, pvalue=2.2269689452801645e-16)
    owner_count    power_in_hp:
PearsonRResult(statistic=-0.09185765668817861, pvalue=1.0318423537673197e-21)
    owner_count    power_in_hp:
SignificanceResult(statistic=-0.13304917529096702, pvalue=6.39936482883275e-44)
    owner_count    mileage:
PearsonRResult(statistic=0.5473685673957609, pvalue=0.0)
    owner_count    mileage:
SignificanceResult(statistic=0.5669884569703822, pvalue=0.0)
    owner_count    mileage_per_year:
PearsonRResult(statistic=-0.05269191251210629, pvalue=3.94001499303432e-08)
    owner_count    mileage_per_year:
SignificanceResult(statistic=0.01912934807524108, pvalue=0.046242595607773516)

```

```

        engine_volume_lites    year:
PearsonRRResult(statistic=0.005815596876976406, pvalue=0.47939449804327067)

        engine_volume_lites    year:
SignificanceResult(statistic=0.08962958364903859, pvalue=9.00861590450538e-28)

        engine_volume_lites    price:
PearsonRRResult(statistic=0.39323640951055316, pvalue=0.0)

        engine_volume_lites    price:
SignificanceResult(statistic=0.3635560107459178, pvalue=0.0)

        engine_volume_lites    owner_count:
PearsonRRResult(statistic=0.13681224671237682, pvalue=1.8088408024098546e-46)

        engine_volume_lites    owner_count:
SignificanceResult(statistic=0.07871559518054935, pvalue=2.2269689452801645e-16)

        engine_volume_lites    power_in_hp:
/home/sgwps/.local/lib/python3.10/site-
packages/pandas/plotting/_matplotlib/core.py:513: RuntimeWarning: More than 20
figures have been opened. Figures created through the pyplot interface
(`matplotlib.pyplot.figure`) are retained until explicitly closed and may
consume too much memory. (To control this warning, see the rcParam
`figure.max_open_warning`). Consider using `matplotlib.pyplot.close()`.
    fig = self.plt.figure(figsize=self.figsize)
PearsonRRResult(statistic=0.8084067010144129, pvalue=0.0)

        engine_volume_lites    power_in_hp:
SignificanceResult(statistic=0.8164171978895333, pvalue=0.0)

        engine_volume_lites    mileage:
PearsonRRResult(statistic=-0.03417740616927616, pvalue=3.6574302782651306e-05)

        engine_volume_lites    mileage:
SignificanceResult(statistic=-0.08429362735561263, pvalue=2.0536857508840307e-24)

        engine_volume_lites    mileage_per_year:
PearsonRRResult(statistic=-0.15417134890647083, pvalue=2.915033752889662e-78)

        engine_volume_lites    mileage_per_year:
SignificanceResult(statistic=-0.11101297153220153, pvalue=3.233802427090711e-41)

        power_in_hp    year:
PearsonRRResult(statistic=0.3447754255884183, pvalue=0.0)

        power_in_hp    year:

```

```

SignificanceResult(statistic=0.4057563410517989, pvalue=0.0)
    power_in_hp    price:
PearsonRResult(statistic=0.6728708652271249, pvalue=0.0)
    power_in_hp    price:
SignificanceResult(statistic=0.6317606699106313, pvalue=0.0)
    power_in_hp    owner_count:
PearsonRResult(statistic=-0.09185765668817861, pvalue=1.0318423537673197e-21)
    power_in_hp    owner_count:
SignificanceResult(statistic=-0.13304917529096702, pvalue=6.39936482883275e-44)
    power_in_hp    engine_volume_lites:
PearsonRResult(statistic=0.8084067010144129, pvalue=0.0)
    power_in_hp    engine_volume_lites:
SignificanceResult(statistic=0.8164171978895333, pvalue=0.0)
    power_in_hp    mileage:
PearsonRResult(statistic=-0.3412024791238298, pvalue=0.0)
    power_in_hp    mileage:
SignificanceResult(statistic=-0.38749243927840565, pvalue=0.0)
    power_in_hp    mileage_per_year:
PearsonRResult(statistic=-0.26412995084403945, pvalue=1.1136320411499086e-231)
    power_in_hp    mileage_per_year:
SignificanceResult(statistic=-0.22122827918089122, pvalue=2.143221059778846e-161)
    mileage    year:
PearsonRResult(statistic=-0.7829249617899781, pvalue=0.0)
    mileage    year:
SignificanceResult(statistic=-0.8670809379021445, pvalue=0.0)
    mileage    price:
PearsonRResult(statistic=-0.5890870461586594, pvalue=0.0)
    mileage    price:
SignificanceResult(statistic=-0.8351638977792649, pvalue=0.0)
    mileage    owner_count:
PearsonRResult(statistic=0.5473685673957609, pvalue=0.0)
    mileage    owner_count:

```

```

SignificanceResult(statistic=0.5669884569703822, pvalue=0.0)
    mileage    engine_volume_lites:
PearsonRResult(statistic=-0.03417740616927616, pvalue=3.6574302782651306e-05)
    mileage    engine_volume_lites:
SignificanceResult(statistic=-0.08429362735561263, pvalue=2.0536857508840307e-24)
    mileage    power_in_hp:
PearsonRResult(statistic=-0.3412024791238298, pvalue=0.0)
    mileage    power_in_hp:
SignificanceResult(statistic=-0.38749243927840565, pvalue=0.0)
    mileage    mileage_per_year:
PearsonRResult(statistic=0.5506505264321213, pvalue=0.0)
    mileage    mileage_per_year:
SignificanceResult(statistic=0.6812994915190222, pvalue=0.0)
    mileage_per_year    year:
PearsonRResult(statistic=-0.10723260939338589, pvalue=3.912333499214111e-39)
    mileage_per_year    year:
SignificanceResult(statistic=-0.3291754154136661, pvalue=0.0)
    mileage_per_year    price:
PearsonRResult(statistic=-0.4704577297092854, pvalue=0.0)
    mileage_per_year    price:
SignificanceResult(statistic=-0.3654517173159595, pvalue=0.0)
    mileage_per_year    owner_count:
PearsonRResult(statistic=-0.05269191251210629, pvalue=3.94001499303432e-08)
    mileage_per_year    owner_count:
SignificanceResult(statistic=0.019129348075241085, pvalue=0.04624259560777347)
    mileage_per_year    engine_volume_lites:
PearsonRResult(statistic=-0.15417134890647083, pvalue=2.915033752889662e-78)
    mileage_per_year    engine_volume_lites:
SignificanceResult(statistic=-0.11101297153220153, pvalue=3.233802427090711e-41)
    mileage_per_year    power_in_hp:
PearsonRResult(statistic=-0.26412995084403945, pvalue=1.1136320411499086e-231)
    mileage_per_year    power_in_hp:

```

```
SignificanceResult(statistic=-0.22122827918089125, pvalue=2.143221059778846e-161)
```

```
mileage_per_year    mileage:
```

```
PearsonRResult(statistic=0.5506505264321213, pvalue=0.0)
```

```
mileage_per_year    mileage:
```

```
SignificanceResult(statistic=0.6812994915190222, pvalue=0.0)
```

