toyota06-02-2025

February 13, 2025

[ ]:
```
import pandas as pd
from google.colab import files
uploaded = files.upload()
file_name = list(uploaded.keys())[0]
cars = pd.read_csv(file_name)
```

```
<IPython.core.display.HTML object>
```

```
Saving Toyoto_Corrola.csv to Toyoto_Corrola.csv
```

[ ]:
```
df = pd.DataFrame(cars)
```

[ ]:
```
from statsmodels.stats.outliers_influence import variance_inflation_factor
import statsmodels.formula.api as smf
import numpy as np
```

[ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.stats.outliers_influence import variance_inflation_factor
import statsmodels.formula.api as smf
import numpy as np
```

[ ]:
```
df.head()
```

[ ]:

	Id	Model	Price	Age_08_04	KM	HP	Doors	Cylinders	Gears	Weight
0	1	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	13500	23	46986	90	3	4	5	1165
1	2	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	13750	23	72937	90	3	4	5	1165
2	3	ÊTOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	13950	24	41711	90	3	4	5	1165
3	4	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	14950	26	48000	90	3	4	5	1165
4	5	TOYOTA Corolla 2.0 D4D HATCHB SOL 2/3-Doors	13750	30	38500	90	3	4	5	1170

[ ]:
```
cars.shape
```

[ ]:
```
(1436, 10)
```

[ ]:
```
rsq_df = pd.DataFrame()
vif_df = pd.DataFrame()
```

[ ]:
```
rsq_df['variable'] = df.columns
vif_df['variable'] = df.columns
```

[ ]:
```
isna = cars.isna().sum()
isna
```

[ ]:
```
Id           0
Model        0
Price        0
Age_08_04    0
KM           0
HP           0
Doors        0
Cylinders    0
Gears        0
Weight       0
dtype: int64
```

[ ]:
```
cars_new = cars.dropna()
```

[ ]:
```
cars_new.head()
```

[ ]:
```
   Id                                          Model  Price  Age_08_04  \
0   1   TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors  13500         23
1   2   TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors  13750         23
2   3  ÊTOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors  13950         24
3   4   TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors  14950         26
4   5     TOYOTA Corolla 2.0 D4D HATCHB SOL 2/3-Doors  13750         30

      KM  HP  Doors  Cylinders  Gears  Weight
0  46986  90      3          4      5    1165
1  72937  90      3          4      5    1165
2  41711  90      3          4      5    1165
3  48000  90      3          4      5    1165
4  38500  90      3          4      5    1170
```

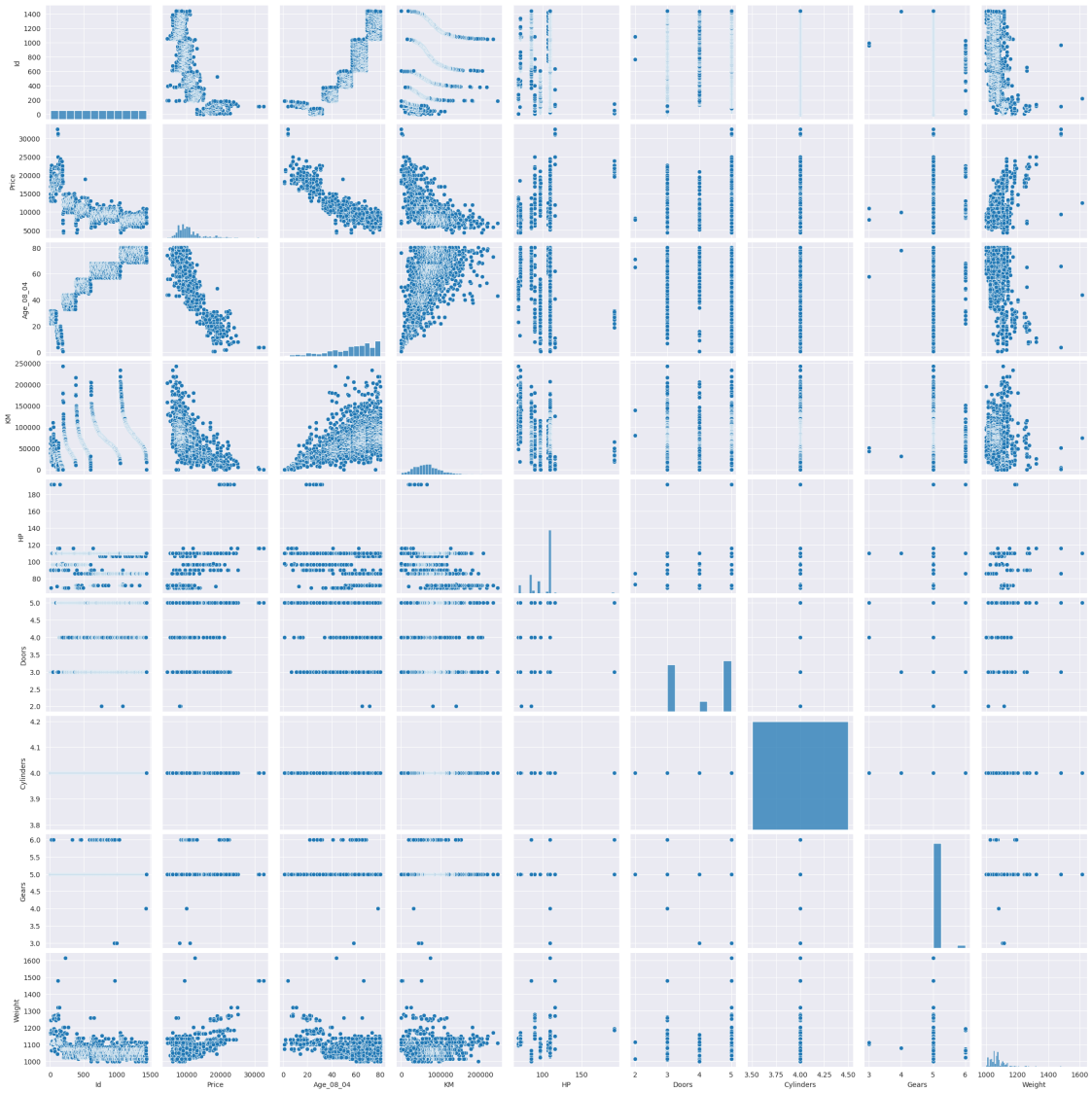
[ ]:
```
cars.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1436 entries, 0 to 1435
```

```
Data columns (total 10 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Id          1436 non-null   int64
 1   Model       1436 non-null   object
 2   Price       1436 non-null   int64
 3   Age_08_04   1436 non-null   int64
 4   KM          1436 non-null   int64
 5   HP          1436 non-null   int64
 6   Doors       1436 non-null   int64
 7   Cylinders   1436 non-null   int64
 8   Gears       1436 non-null   int64
 9   Weight      1436 non-null   int64
dtypes: int64(9), object(1)
memory usage: 112.3+ KB
```

[ ]:
```
sns.set_style(style='darkgrid')
sns.pairplot(cars)
```

[ ]:
```
<seaborn.axisgrid.PairGrid at 0x7fdea0a5ac50>
```



[ ]:
```
df.corr()
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-14-2f6f6606aa2c> in <cell line: 0>()
----> 1 df.corr()

/usr/local/lib/python3.11/dist-packages/pandas/core/frame.py in corr(self,␣
 ↪method, min_periods, numeric_only)
  11047         cols = data.columns
  11048         idx = cols.copy()
> 11049         mat = data.to_numpy(dtype=float, na_value=np.nan, copy=False)
  11050
```

```
  11051         if method == "pearson":

/usr/local/lib/python3.11/dist-packages/pandas/core/frame.py in to_numpy(self,␣
 ↪dtype, copy, na_value)
   1991         if dtype is not None:
   1992             dtype = np.dtype(dtype)
-> 1993         result = self._mgr.as_array(dtype=dtype, copy=copy,␣
 ↪na_value=na_value)
   1994         if result.dtype is not dtype:
   1995             result = np.asarray(result, dtype=dtype)

/usr/local/lib/python3.11/dist-packages/pandas/core/internals/managers.py in␣
 ↪as_array(self, dtype, copy, na_value)
   1692                 arr.flags.writeable = False
   1693         else:
-> 1694             arr = self._interleave(dtype=dtype, na_value=na_value)
   1695             # The underlying data was copied within _interleave, so no␣
 ↪need
   1696             # to further copy if copy=True or setting na_value

/usr/local/lib/python3.11/dist-packages/pandas/core/internals/managers.py in␣
 ↪_interleave(self, dtype, na_value)
   1751             else:
   1752                 arr = blk.get_values(dtype)
-> 1753             result[rl.indexer] = arr
   1754             itemmask[rl.indexer] = 1
   1755

ValueError: could not convert string to float: 'TOYOTA Corolla 2.0 D4D HATCHB␣
 ↪TERRA 2/3-Doors'
```

[ ]:
```
print(df.corr())
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-37-23236a4e6045> in <cell line: 0>()
----> 1 print(df.corr())

/usr/local/lib/python3.11/dist-packages/pandas/core/frame.py in corr(self,␣
 ↪method, min_periods, numeric_only)
  11047         cols = data.columns
  11048         idx = cols.copy()
> 11049         mat = data.to_numpy(dtype=float, na_value=np.nan, copy=False)
  11050
  11051         if method == "pearson":
```

```
/usr/local/lib/python3.11/dist-packages/pandas/core/frame.py in to_numpy(self,␣
 ↪dtype, copy, na_value)
   1991         if dtype is not None:
   1992             dtype = np.dtype(dtype)
-> 1993         result = self._mgr.as_array(dtype=dtype, copy=copy,␣
 ↪na_value=na_value)
   1994         if result.dtype is not dtype:
   1995             result = np.asarray(result, dtype=dtype)

/usr/local/lib/python3.11/dist-packages/pandas/core/internals/managers.py in␣
 ↪as_array(self, dtype, copy, na_value)
   1692                 arr.flags.writeable = False
   1693         else:
-> 1694             arr = self._interleave(dtype=dtype, na_value=na_value)
   1695             # The underlying data was copied within _interleave, so no␣
 ↪need
   1696             # to further copy if copy=True or setting na_value

/usr/local/lib/python3.11/dist-packages/pandas/core/internals/managers.py in␣
 ↪_interleave(self, dtype, na_value)
   1751             else:
   1752                 arr = blk.get_values(dtype)
-> 1753             result[rl.indexer] = arr
   1754             itemmask[rl.indexer] = 1
   1755

ValueError: could not convert string to float: 'TOYOTA Corolla 2.0 D4D HATCHB␣
 ↪TERRA 2/3-Doors'
```

[ ]: