HW3

March 5, 2018

```
In [1]: import pandas as pd
        import numpy as np
        import warnings
        warnings.filterwarnings('ignore')
        import fancyimpute as fi
        from sklearn.linear_model import Ridge
        from sklearn.preprocessing import StandardScaler
        from sklearn.model_selection import train_test_split

        from datetime import datetime
```

```
Using TensorFlow backend.
```

```
In [ ]:
```

1 Task 1: Linear Model and Data Cleaning

For Tasks 1-3 we used the following features: 'Model Year', 'Index (Model Type Index)', 'Range1 - Model Type Driving Range - Conventional Fuel', '2Dr Pass Vol', '4Dr Pass Vol', '4Dr Lugg Vol', 'Htchbk Pass Vol', 'Htchbk Lugg Vol', 'Fuel2 Annual Fuel Cost - Alternative Fuel', 'Carline Class', 'Release Date', '$ You Save over 5 years (amount saved in fuel costs over 5 years - on label) ',

'Mfr Name', 'Division', 'Verify Mfr Cd', '# Cyl', 'Transmission', 'Air Aspir Method', 'Trans', '# Gears', 'Lockup Torque Converter', 'Trans Creeper Gear', 'Drive Sys', 'Max Ethanol % - Gasoline', 'Fuel Usage - Conventional Fuel', 'Gas Guzzler Exempt (Where Truck = 1975 NHTSA truck definition)', ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel', ' Fuel2 Usage - Alternative Fuel', 'Exhaust Valves Per Cyl', 'Car/Truck Category - Cash for Clunkers Bill.', 'Unique Label?', 'Label Recalc?', 'Cyl Deact?', 'Var Valve Timing?', 'Var Valve Lift?', 'Fuel Metering Sys Cd', 'Off Board Charge Capable (Y or N)', 'Camless Valvetrain (Y or N)', 'Stop/Start System (Engine Management System) Code'

Our system for choosing these features is based on domain knowledge of cars, removing the list of features that give away the target and mapping correlations.

```
In [2]: d15 = pd.read_excel("2015 FE Guide-for DOE-Mobility Ventures only-OK to release-no-sales
        d16 = pd.read_excel("2016 FE Guide for DOE-OK to release-no-sales-4-27-2017Mercedesforpu
        d17 = pd.read_excel("2017 FE Guide for DOE-release dates before 9-20-2017-no sales-9-19-
```

```
In [3]: d18 = pd.read_excel("2018 FE Guide for DOE-release dates before 2-17-2018-no-sales-2-15-
In [4]: d18.shape

Out[4]: (1220, 162)

In [5]: d_test = d18

In [6]: frames = [d15, d16, d17]
        all_years_frame = [d15, d16, d17, d18]

        d_all = pd.concat(all_years_frame)
        d_train = pd.concat(frames)

In [65]: y_tr = d_train['Comb Unrd Adj FE - Conventional Fuel']
         y_te = d18['Comb Unrd Adj FE - Conventional Fuel']

         d_train.shape

Out[65]: (3701, 162)

In [8]: #Finding number of numerical categories
        num_cols = d_train._get_numeric_data().columns

In [9]: num_cols.shape

Out[9]: (86,)

In [10]: d_all.shape

Out[10]: (4921, 162)

In [11]: d_temp = d_all[['Model Year',
             'Index (Model Type Index)',
          'Range1 - Model Type Driving Range - Conventional Fuel',
          '2Dr Pass Vol',
          '4Dr Pass Vol',
          '4Dr Lugg Vol',
          'Htchbk Pass Vol',
          'Htchbk Lugg Vol',
          'Fuel2 Annual Fuel Cost - Alternative Fuel',
          'Carline Class',
          'Release Date',
          '$ You Save over 5 years (amount saved in fuel costs over 5 years - on label) ',

          'Mfr Name',
          'Division',
          'Verify Mfr Cd',
          '# Cyl',
          'Transmission',
          'Air Aspir Method',
          'Trans',
          '# Gears',
          'Lockup Torque Converter',
          'Trans Creeper Gear',
          'Drive Sys',
          'Max Ethanol % - Gasoline',
          'Fuel Usage  - Conventional Fuel',
          'Gas Guzzler Exempt (Where Truck = 1975 NHTSA truck definition)',
          ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel',
          ' Fuel2 Usage - Alternative Fuel',
          'Exhaust Valves Per Cyl',
          'Car/Truck Category - Cash for Clunkers Bill.',
          'Unique Label?',
          'Label Recalc?',
          'Cyl Deact?',
          'Var Valve Timing?',
          'Var Valve Lift?',
          'Fuel Metering Sys Cd',
          'Off Board Charge Capable (Y or N)',
          'Camless Valvetrain (Y or N)',
          'Stop/Start System (Engine Management System) Code']]

In [12]: d_tr = d_train[['Model Year',
             'Index (Model Type Index)',
          'Range1 - Model Type Driving Range - Conventional Fuel',
          '2Dr Pass Vol',
          '4Dr Pass Vol',
          '4Dr Lugg Vol',
          'Htchbk Pass Vol',
          'Htchbk Lugg Vol',
          'Fuel2 Annual Fuel Cost - Alternative Fuel',
          'Carline Class',
          'Release Date',
          '$ You Save over 5 years (amount saved in fuel costs over 5 years - on label) ',
          'Mfr Name',
          'Division',
          'Verify Mfr Cd',
          '# Cyl',
          'Transmission',
          'Air Aspir Method',
          'Trans',
          '# Gears',
          'Lockup Torque Converter',
          'Trans Creeper Gear',
          'Drive Sys',
          'Max Ethanol % - Gasoline',
          'Fuel Usage  - Conventional Fuel',
          'Eng Displ',
          ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel',
          ' Fuel2 Usage - Alternative Fuel',
          'Exhaust Valves Per Cyl',
          'Car/Truck Category - Cash for Clunkers Bill.',
          'Unique Label?',
          'Label Recalc?',
          'Cyl Deact?',
          'Var Valve Timing?',
          'Var Valve Lift?',
          'Fuel Metering Sys Cd',
          'Off Board Charge Capable (Y or N)',
          'Camless Valvetrain (Y or N)',
          'Stop/Start System (Engine Management System) Code']]
```

```
In [13]: d_te = d_test[['Model Year',
             'Index (Model Type Index)',
          'Range1 - Model Type Driving Range - Conventional Fuel',
          '2Dr Pass Vol',
          '4Dr Pass Vol',
          '4Dr Lugg Vol',
          'Htchbk Pass Vol',
          'Htchbk Lugg Vol',
          'Fuel2 Annual Fuel Cost - Alternative Fuel',
          'Carline Class',
          'Release Date',
          '$ You Save over 5 years (amount saved in fuel costs over 5 years - on label) ',
          'Mfr Name',
          'Division',
          'Verify Mfr Cd',
          '# Cyl',
          'Transmission',
          'Air Aspir Method',
          'Trans',
          '# Gears',
          'Lockup Torque Converter',
          'Trans Creeper Gear',
          'Drive Sys',
          'Max Ethanol % - Gasoline',
          'Fuel Usage  - Conventional Fuel',
          'Eng Displ',
          ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel',
          ' Fuel2 Usage - Alternative Fuel',
          'Exhaust Valves Per Cyl',
          'Car/Truck Category - Cash for Clunkers Bill.',
          'Unique Label?',
          'Label Recalc?',
          'Cyl Deact?',
          'Var Valve Timing?',
          'Var Valve Lift?',
           'Fuel Metering Sys Cd',
           'Off Board Charge Capable (Y or N)',
           'Camless Valvetrain (Y or N)',
           'Stop/Start System (Engine Management System) Code']]
```

```
In [14]: d_tr.shape
         d_tr_cat = d_tr.select_dtypes(include = ['object'])
         d_tr_num = d_tr.select_dtypes(exclude = ['object'])
```

```
In [15]: d_te.shape
         d_te_cat = d_te.select_dtypes(include = ['object'])
         d_te_num = d_te.select_dtypes(exclude = ['object'])
```

```
In [16]: d_tr_num = d_tr_num.drop(['Model Year','Carline Class','# Cyl','# Gears',
                                  'Max Ethanol % - Gasoline','Exhaust Valves Per Cyl','Release Date
```

```
In [17]: d_tr_num.head()
```

```
Out[17]:    Index (Model Type Index)  2Dr Pass Vol  4Dr Pass Vol  4Dr Lugg Vol  \
         0                       264           NaN           NaN           NaN
         1                         8           NaN           NaN           NaN
         2                         4           NaN           NaN           NaN
         3                         1           NaN           NaN           NaN
         4                         5           NaN           NaN           NaN

            Htchbk Pass Vol  Htchbk Lugg Vol  \
         0              NaN              NaN
         1              NaN              NaN
         2              NaN              NaN
         3              NaN              NaN
         4              NaN              NaN

            Fuel2 Annual Fuel Cost - Alternative Fuel  \
         0                                        NaN
         1                                        NaN
         2                                        NaN
         3                                        NaN
         4                                        NaN

            $ You Save over 5 years (amount saved in fuel costs over 5 years - on label)   \
         0                                              750.0
         1                                                NaN
         2                                                NaN
         3                                                NaN
         4                                                NaN

            Eng Displ
         0        1.8
         1        6.0
```

```
2        4.7
3        4.7
4        4.7
```

```
In [18]: d_te_num = d_te_num.drop(['Model Year','Carline Class','# Cyl','# Gears',
                                   'Max Ethanol % - Gasoline','Exhaust Valves Per Cyl','Release Date
```

```
In [19]: col_list = ['Model Year','Carline Class','# Cyl','# Gears',
                                   'Max Ethanol % - Gasoline','Exhaust Valves Per Cyl']
```

```
In [20]: for i in col_list:
             d_tr_cat[i]=d_tr[i]
```

```
In [21]: for i in col_list:
             d_te_cat[i]=d_te[i]
```

```
In [22]: d_tr_cat.head()
```

```
Out[22]:   Range1 - Model Type Driving Range - Conventional Fuel      Mfr Name  \
         0                                                NaN        FCA Italy
         1                                                NaN     aston martin
         2                                                NaN     aston martin
         3                                                NaN     aston martin
         4                                                NaN     aston martin

                            Division Verify Mfr Cd Transmission Air Aspir Method Trans  \
         0                Alfa Romeo           FTG    Auto(AM6)              TC    AM
         1  Aston Martin Lagonda Ltd           ASX    Auto(AM7)             NaN    AM
         2  Aston Martin Lagonda Ltd           ASX    Auto(AM7)             NaN    AM
         3  Aston Martin Lagonda Ltd           ASX  Manual(M6)              NaN     M
         4  Aston Martin Lagonda Ltd           ASX    Auto(AM7)             NaN    AM

           Lockup Torque Converter Trans Creeper Gear Drive Sys         ...             \
         0                       Y                  N         R         ...
         1                       N                  N         R         ...
         2                       N                  N         R         ...
         3                       N                  N         R         ...
         4                       N                  N         R         ...

           Fuel Metering Sys Cd Off Board Charge Capable (Y or N)  \
         0                  GDI                                NaN
         1                  MFI                                NaN
         2                  MFI                                NaN
         3                  MFI                                NaN
         4                  MFI                                NaN

           Camless Valvetrain (Y or N)  \
         0                          N
         1                          N
```

```
2                           N
3                           N
4                           N

  Stop/Start System (Engine Management System) Code Model Year Carline Class  \
0                                                 N       2015             1
1                                                 N       2015             1
2                                                 N       2015             1
3                                                 N       2015             1
4                                                 N       2015             1

  # Cyl # Gears Max Ethanol % - Gasoline Exhaust Valves Per Cyl
0     4       6                     10.0                      2
1    12       7                     10.0                      2
2     8       7                     10.0                      2
3     8       6                     10.0                      2
4     8       7                     10.0                      2

[5 rows x 29 columns]
```

```
In [23]: d_te_cat.head()
```

Out[23]:

```
  Range1 - Model Type Driving Range - Conventional Fuel             Mfr Name  \
0                                                NaN                   Honda
1                                                NaN              FCA US LLC
2                                                NaN     Volkswagen Group of
3                                                NaN     Volkswagen Group of
4                                                NaN     Volkswagen Group of

     Division Verify Mfr Cd Transmission Air Aspir Method Trans  \
0       Acura           HNX  Auto(AM-S9)               TC   AMS
1  ALFA ROMEO           CRX    Auto(AM6)               TC    AM
2        Audi           VGA  Auto(AM-S7)              NaN   AMS
3        Audi           VGA  Auto(AM-S7)              NaN   AMS
4        Audi           VGA  Auto(AM-S7)              NaN   AMS

  Lockup Torque Converter Trans Creeper Gear Drive Sys          ...            \
0                       Y                  N         A          ...
1                       Y                  N         R          ...
2                       Y                  N         A          ...
3                       Y                  N         R          ...
4                       Y                  N         A          ...

  Fuel Metering Sys Cd Off Board Charge Capable (Y or N)  \
0                  GDI                                 N
1                  GDI                               NaN
2                 GDPI                               NaN
3                 GDPI                               NaN
```

```
4                 GDPI                               NaN

  Camless Valvetrain (Y or N)  \
0                           N
1                           N
2                           N
3                           N
4                           N

  Stop/Start System (Engine Management System) Code Model Year Carline Class  \
0                                                 Y       2018             1
1                                                 N       2018             1
2                                                 N       2018             1
3                                                 N       2018             1
4                                                 N       2018             1

  # Cyl # Gears Max Ethanol % - Gasoline Exhaust Valves Per Cyl
0     6       9                     10.0                      2
1     4       6                     10.0                      2
2    10       7                     15.0                      2
3    10       7                     15.0                      2
4    10       7                     15.0                      2

[5 rows x 29 columns]
```

```
In [24]: d_tr_num.dtypes
```

```
Out[24]: Index (Model Type Index)                                                           int6
         2Dr Pass Vol                                                                     float6
         4Dr Pass Vol                                                                     float6
         4Dr Lugg Vol                                                                     float6
         Htchbk Pass Vol                                                                  float6
         Htchbk Lugg Vol                                                                  float6
         Fuel2 Annual Fuel Cost - Alternative Fuel                                        float6
         $ You Save over 5 years (amount saved in fuel costs over 5 years - on label)     float6
         Eng Displ                                                                        float6
         dtype: object
```

```
In [25]: d_te_num.dtypes
```

```
Out[25]: Index (Model Type Index)                                                           int6
         2Dr Pass Vol                                                                     float6
         4Dr Pass Vol                                                                     float6
         4Dr Lugg Vol                                                                     float6
         Htchbk Pass Vol                                                                  float6
         Htchbk Lugg Vol                                                                  float6
         Fuel2 Annual Fuel Cost - Alternative Fuel                                        float6
         $ You Save over 5 years (amount saved in fuel costs over 5 years - on label)     float6
```

```
Eng Displ                                                                  float6
dtype: object
```

In [26]:
```
d_tr_num = d_tr_num.reset_index(drop=True)
```

In [27]:
```
d_te_num = d_te_num.reset_index(drop=True)
```

In [28]:
```
#d_tr_num = d_tr_num.drop(["Release Date"], axis = 1)
```

In [29]:
```
# for i in range(len(d_tr_num['Release Date'])):
#     d_tr_num['Release Date'][i] = datetime.strptime(str(d_tr_num['Release Date'][i]).
#     print(i)
```

In [192]:
```
d_tr_num.head
```

Out[192]:
```
<bound method NDFrame.head of       Index (Model Type Index)  2Dr Pass Vol  4Dr Pass V
0                          264          NaN          NaN          NaN
1                            8          NaN          NaN          NaN
2                            4          NaN          NaN          NaN
3                            1          NaN          NaN          NaN
4                            5          NaN          NaN          NaN
5                            2          NaN          NaN          NaN
6                            6          NaN          NaN          NaN
7                            3          NaN          NaN          NaN
8                           27          NaN          NaN          NaN
9                           29          NaN          NaN          NaN
10                          35          NaN          NaN          NaN
11                          33          NaN          NaN          NaN
12                          26          NaN          NaN          NaN
13                          28          NaN          NaN          NaN
14                          34          NaN          NaN          NaN
15                          32          NaN          NaN          NaN
16                           3          NaN          NaN          NaN
17                         110          NaN          NaN          NaN
18                         428          NaN          NaN          NaN
19                         429          NaN          NaN          NaN
20                         436          NaN          NaN          NaN
21                         438          NaN          NaN          NaN
22                          60          NaN          NaN          NaN
23                          59          NaN          NaN          NaN
24                          67          NaN          NaN          NaN
25                          60          NaN          NaN          NaN
26                          68          NaN          NaN          NaN
27                         185          NaN          NaN          NaN
28                         143          NaN          NaN          NaN
29                         142          NaN          NaN          NaN
...                        ...          ...          ...          ...
3671                       402          NaN          NaN          NaN
3672                       408          NaN          NaN          NaN
```

```
3673                       421           NaN           NaN           NaN
3674                       423           NaN           NaN           NaN
3675                       283           NaN           NaN           NaN
3676                       401           NaN           NaN           NaN
3677                       412           NaN           NaN           NaN
3678                       402           NaN           NaN           NaN
3679                       411           NaN           NaN           NaN
3680                       421           NaN           NaN           NaN
3681                       422           NaN           NaN           NaN
3682                        69           NaN           NaN           NaN
3683                       207           NaN           NaN           NaN
3684                       103           NaN           NaN           NaN
3685                       104           NaN           NaN           NaN
3686                       114           NaN           NaN           NaN
3687                       105           NaN           NaN           NaN
3688                       106           NaN           NaN           NaN
3689                        37           NaN           NaN           NaN
3690                        12           NaN           NaN           NaN
3691                         8           NaN           NaN           NaN
3692                        46           NaN           NaN           NaN
3693                        53           NaN           NaN           NaN
3694                        52           NaN           NaN           NaN
3695                       348           NaN           NaN           NaN
3696                       293           NaN           NaN           NaN
3697                       212           NaN           NaN           NaN
3698                       211           NaN           NaN           NaN
3699                       232           NaN           NaN           NaN
3700                       231           NaN           NaN           NaN

      Htchbk Pass Vol  Htchbk Lugg Vol  \
0                 NaN              NaN
1                 NaN              NaN
2                 NaN              NaN
3                 NaN              NaN
4                 NaN              NaN
5                 NaN              NaN
6                 NaN              NaN
7                 NaN              NaN
8                 NaN              NaN
9                 NaN              NaN
10                NaN              NaN
11                NaN              NaN
12                NaN              NaN
13                NaN              NaN
14                NaN              NaN
15                NaN              NaN
16                NaN              NaN
17                NaN              NaN
```

```
18                NaN              NaN
19                NaN              NaN
20                NaN              NaN
21                NaN              NaN
22                NaN              NaN
23                NaN              NaN
24                NaN              NaN
25                NaN              NaN
26                NaN              NaN
27                NaN              NaN
28                NaN              NaN
29                NaN              NaN
...               ...              ...
3671              NaN              NaN
3672              NaN              NaN
3673              NaN              NaN
3674              NaN              NaN
3675              NaN              NaN
3676              NaN              NaN
3677              NaN              NaN
3678              NaN              NaN
3679              NaN              NaN
3680              NaN              NaN
3681              NaN              NaN
3682              NaN              NaN
3683              NaN              NaN
3684              NaN              NaN
3685              NaN              NaN
3686              NaN              NaN
3687              NaN              NaN
3688              NaN              NaN
3689              NaN              NaN
3690              NaN              NaN
3691              NaN              NaN
3692              NaN              NaN
3693              NaN              NaN
3694              NaN              NaN
3695              NaN              NaN
3696              NaN              NaN
3697              NaN              NaN
3698              NaN              NaN
3699              NaN              NaN
3700              NaN              NaN

      Fuel2 Annual Fuel Cost - Alternative Fuel  \
0                                           NaN
1                                           NaN
2                                           NaN
```

```
3                                            NaN
4                                            NaN
5                                            NaN
6                                            NaN
7                                            NaN
8                                            NaN
9                                            NaN
10                                           NaN
11                                           NaN
12                                           NaN
13                                           NaN
14                                           NaN
15                                           NaN
16                                           NaN
17                                           NaN
18                                           NaN
19                                           NaN
20                                           NaN
21                                           NaN
22                                           NaN
23                                           NaN
24                                           NaN
25                                           NaN
26                                           NaN
27                                           NaN
28                                           NaN
29                                           NaN
...                                          ...
3671                                         NaN
3672                                         NaN
3673                                         NaN
3674                                         NaN
3675                                         NaN
3676                                         NaN
3677                                         NaN
3678                                         NaN
3679                                         NaN
3680                                         NaN
3681                                         NaN
3682                                         NaN
3683                                         NaN
3684                                         NaN
3685                                         NaN
3686                                         NaN
3687                                         NaN
3688                                         NaN
3689                                         NaN
3690                                         NaN
```

```
3691                                          3100.0
3692                                             NaN
3693                                             NaN
3694                                             NaN
3695                                             NaN
3696                                             NaN
3697                                             NaN
3698                                             NaN
3699                                             NaN
3700                                             NaN

      $ You Save over 5 years (amount saved in fuel costs over 5 years - on label)   \
0                                                 750.0
1                                                   NaN
2                                                   NaN
3                                                   NaN
4                                                   NaN
5                                                   NaN
6                                                   NaN
7                                                   NaN
8                                                   NaN
9                                                   NaN
10                                                  NaN
11                                                  NaN
12                                                  NaN
13                                                  NaN
14                                                  NaN
15                                                  NaN
16                                                  0.0
17                                                  NaN
18                                                  0.0
19                                                  0.0
20                                                  NaN
21                                                  NaN
22                                                  NaN
23                                                  NaN
24                                                  NaN
25                                                  NaN
26                                                  NaN
27                                                  NaN
28                                                  NaN
29                                                  NaN
...                                                 ...
3671                                                NaN
3672                                                NaN
3673                                                NaN
3674                                                NaN
3675                                                NaN
```

```
3676                                                 NaN
3677                                                 NaN
3678                                                 NaN
3679                                                 NaN
3680                                                 NaN
3681                                                 NaN
3682                                                 NaN
3683                                                 NaN
3684                                                 NaN
3685                                                 NaN
3686                                                 NaN
3687                                               500.0
3688                                               750.0
3689                                                 NaN
3690                                                 NaN
3691                                                 NaN
3692                                                 NaN
3693                                                 NaN
3694                                                 NaN
3695                                                 NaN
3696                                                 NaN
3697                                                 NaN
3698                                                 NaN
3699                                                 NaN
3700                                                 NaN

      Eng Displ
0           1.8
1           6.0
2           4.7
3           4.7
4           4.7
5           4.7
6           4.7
7           4.7
8           4.2
9           4.2
10          5.2
11          5.2
12          4.2
13          4.2
14          5.2
15          5.2
16          2.0
17          4.0
18          2.0
19          2.0
20          3.0
```

```
21          3.0
22          8.0
23          6.2
24          6.2
25          6.2
26          6.2
27          8.4
28          4.5
29          4.5
...         ...
3671        3.5
3672        3.0
3673        3.0
3674        4.7
3675        5.6
3676        3.6
3677        3.6
3678        3.6
3679        3.6
3680        4.8
3681        4.8
3682        4.0
3683        4.0
3684        3.5
3685        3.5
3686        3.5
3687        3.5
3688        3.5
3689        5.7
3690        5.7
3691        5.7
3692        3.6
3693        2.0
3694        2.0
3695        3.5
3696        3.5
3697        2.0
3698        2.0
3699        2.5
3700        2.5

[3701 rows x 9 columns]>
```

```
In [193]: d_te_num.head
```

```
Out[193]: <bound method NDFrame.head of       Index (Model Type Index)  2Dr Pass Vol  4Dr Pass V
          0                            57            NaN           NaN           NaN
          1                           410            NaN           NaN           NaN
```

```
2                           65           NaN           NaN           NaN
3                           71           NaN           NaN           NaN
4                           66           NaN           NaN           NaN
5                           72           NaN           NaN           NaN
6                           46           NaN           NaN           NaN
7                          488           NaN           NaN           NaN
8                           38           NaN           NaN           NaN
9                          278           NaN           NaN           NaN
10                         223           NaN           NaN           NaN
11                         285           NaN           NaN           NaN
12                         276           NaN           NaN           NaN
13                         142           NaN           NaN           NaN
14                         143           NaN           NaN           NaN
15                         145           NaN           NaN           NaN
16                         144           NaN           NaN           NaN
17                         154           NaN           NaN           NaN
18                         405           NaN           NaN           NaN
19                         406           NaN           NaN           NaN
20                         322          43.0           NaN           NaN
21                         185           NaN           NaN           NaN
22                         161           NaN           NaN           NaN
23                         171           NaN           NaN           NaN
24                         184           NaN           NaN           NaN
25                         160           NaN           NaN           NaN
26                         170           NaN           NaN           NaN
27                         159           NaN           NaN           NaN
28                         158           NaN           NaN           NaN
29                         165           NaN           NaN           NaN
...                        ...           ...           ...           ...
1190                       271           NaN           NaN           NaN
1191                       406           NaN           NaN           NaN
1192                       272           NaN           NaN           NaN
1193                       274           NaN           NaN           NaN
1194                       424           NaN           NaN           NaN
1195                       435           NaN           NaN           NaN
1196                       436           NaN           NaN           NaN
1197                       821           NaN           NaN           NaN
1198                       402           NaN           NaN           NaN
1199                       421           NaN           NaN           NaN
1200                       423           NaN           NaN           NaN
1201                       283           NaN           NaN           NaN
1202                       401           NaN           NaN           NaN
1203                       412           NaN           NaN           NaN
1204                       402           NaN           NaN           NaN
1205                       411           NaN           NaN           NaN
1206                       421           NaN           NaN           NaN
1207                       422           NaN           NaN           NaN
1208                        58           NaN           NaN           NaN
```

```
1209                      207          NaN          NaN          NaN
1210                      102          NaN          NaN          NaN
1211                      103          NaN          NaN          NaN
1212                      108          NaN          NaN          NaN
1213                      111          NaN          NaN          NaN
1214                      114          NaN          NaN          NaN
1215                       86          NaN          NaN          NaN
1216                       37          NaN          NaN          NaN
1217                       33          NaN          NaN          NaN
1218                       53          NaN          NaN          NaN
1219                       52          NaN          NaN          NaN

      Htchbk Pass Vol  Htchbk Lugg Vol  \
0                 NaN              NaN
1                 NaN              NaN
2                 NaN              NaN
3                 NaN              NaN
4                 NaN              NaN
5                 NaN              NaN
6                 NaN              NaN
7                 NaN              NaN
8                 NaN              NaN
9                 NaN              NaN
10                NaN              NaN
11                NaN              NaN
12                NaN              NaN
13                NaN              NaN
14                NaN              NaN
15                NaN              NaN
16                NaN              NaN
17                NaN              NaN
18                NaN              NaN
19                NaN              NaN
20                NaN              NaN
21                NaN              NaN
22                NaN              NaN
23                NaN              NaN
24                NaN              NaN
25                NaN              NaN
26                NaN              NaN
27                NaN              NaN
28                NaN              NaN
29                NaN              NaN
...               ...              ...
1190              NaN              NaN
1191              NaN              NaN
1192              NaN              NaN
1193              NaN              NaN
```

```
1194               NaN               NaN
1195               NaN               NaN
1196               NaN               NaN
1197               NaN               NaN
1198               NaN               NaN
1199               NaN               NaN
1200               NaN               NaN
1201               NaN               NaN
1202               NaN               NaN
1203               NaN               NaN
1204               NaN               NaN
1205               NaN               NaN
1206               NaN               NaN
1207               NaN               NaN
1208               NaN               NaN
1209               NaN               NaN
1210               NaN               NaN
1211               NaN               NaN
1212               NaN               NaN
1213               NaN               NaN
1214               NaN               NaN
1215               NaN               NaN
1216               NaN               NaN
1217               NaN               NaN
1218               NaN               NaN
1219               NaN               NaN

      Fuel2 Annual Fuel Cost - Alternative Fuel  \
0                                           NaN
1                                           NaN
2                                           NaN
3                                           NaN
4                                           NaN
5                                           NaN
6                                           NaN
7                                           NaN
8                                           NaN
9                                           NaN
10                                          NaN
11                                          NaN
12                                          NaN
13                                          NaN
14                                          NaN
15                                          NaN
16                                          NaN
17                                          NaN
18                                          NaN
19                                          NaN
```

```
20                                          NaN
21                                          NaN
22                                          NaN
23                                          NaN
24                                          NaN
25                                          NaN
26                                          NaN
27                                          NaN
28                                          NaN
29                                          NaN
...                                         ...
1190                                        NaN
1191                                        NaN
1192                                        NaN
1193                                        NaN
1194                                        NaN
1195                                        NaN
1196                                        NaN
1197                                     2100.0
1198                                        NaN
1199                                        NaN
1200                                        NaN
1201                                        NaN
1202                                        NaN
1203                                        NaN
1204                                        NaN
1205                                        NaN
1206                                        NaN
1207                                        NaN
1208                                        NaN
1209                                        NaN
1210                                        NaN
1211                                        NaN
1212                                        NaN
1213                                        NaN
1214                                        NaN
1215                                        NaN
1216                                        NaN
1217                                     2900.0
1218                                        NaN
1219                                        NaN

      $ You Save over 5 years (amount saved in fuel costs over 5 years - on label)   \
0                                                   NaN
1                                                   NaN
2                                                   NaN
3                                                   NaN
4                                                   NaN
```

```
5                                                     NaN
6                                                     NaN
7                                                     NaN
8                                                     NaN
9                                                     NaN
10                                                    NaN
11                                                    NaN
12                                                    NaN
13                                                    NaN
14                                                    NaN
15                                                    NaN
16                                                    NaN
17                                                    NaN
18                                                    NaN
19                                                    NaN
20                                                    NaN
21                                                    NaN
22                                                    NaN
23                                                    NaN
24                                                    NaN
25                                                    NaN
26                                                    NaN
27                                                    NaN
28                                                    NaN
29                                                    NaN
...                                                   ...
1190                                                  NaN
1191                                                  NaN
1192                                                  NaN
1193                                                  NaN
1194                                                  NaN
1195                                                  NaN
1196                                                  NaN
1197                                                  NaN
1198                                                  NaN
1199                                                  NaN
1200                                                  NaN
1201                                                  NaN
1202                                                  NaN
1203                                                  NaN
1204                                                  NaN
1205                                                  NaN
1206                                                  NaN
1207                                                  NaN
1208                                                  NaN
1209                                                  NaN
1210                                                  NaN
1211                                                  NaN
```

```
1212                                                 NaN
1213                                               250.0
1214                                               500.0
1215                                                 NaN
1216                                                 NaN
1217                                                 NaN
1218                                                 NaN
1219                                                 NaN

      Eng Displ
0           3.5
1           1.8
2           5.2
3           5.2
4           5.2
5           5.2
6           2.0
7           3.0
8           8.0
9           6.2
10          6.2
11          6.2
12          6.2
13          3.9
14          3.9
15          3.9
16          3.9
17          6.5
18          1.4
19          1.4
20          3.5
21          2.0
22          3.0
23          3.0
24          2.0
25          3.0
26          3.0
27          5.0
28          5.0
29          3.0
...         ...
1190        3.0
1191        5.5
1192        5.5
1193        5.5
1194        5.5
1195        4.0
1196        4.0
```

```
1197        3.5
1198        3.5
1199        3.0
1200        4.7
1201        5.6
1202        3.6
1203        3.6
1204        3.6
1205        3.6
1206        4.8
1207        4.8
1208        4.0
1209        4.0
1210        3.5
1211        3.5
1212        3.5
1213        3.5
1214        3.5
1215        5.7
1216        5.7
1217        5.7
1218        2.0
1219        2.0

[1220 rows x 9 columns]>
```

```
In [32]: from sklearn.preprocessing import Imputer

         imp_num = Imputer(strategy = 'median').fit(d_tr_num)
```

```
In [33]: from sklearn.preprocessing import Imputer

         imp_num_t = Imputer(strategy = 'median').fit(d_te_num)
```

```
In [34]: X_tr_imp = imp_num.transform(d_tr_num)
```

```
In [35]: X_te_imp = imp_num_t.transform(d_te_num)
```

```
In [36]: X_tr_num = pd.DataFrame(X_tr_imp, columns=d_tr_num.columns)
```

```
In [37]: X_te_num = pd.DataFrame(X_te_imp, columns=d_te_num.columns)
```

```
In [194]: X_tr_num.head
```

```
Out[194]: <bound method NDFrame.head of       Index (Model Type Index)  2Dr Pass Vol  4Dr Pass V
          0                         264.0        83.0          98.0          14.0
          1                           8.0        83.0          98.0          14.0
          2                           4.0        83.0          98.0          14.0
          3                           1.0        83.0          98.0          14.0
```

```
4                         5.0          83.0          98.0          14.0
5                         2.0          83.0          98.0          14.0
6                         6.0          83.0          98.0          14.0
7                         3.0          83.0          98.0          14.0
8                        27.0          83.0          98.0          14.0
9                        29.0          83.0          98.0          14.0
10                       35.0          83.0          98.0          14.0
11                       33.0          83.0          98.0          14.0
12                       26.0          83.0          98.0          14.0
13                       28.0          83.0          98.0          14.0
14                       34.0          83.0          98.0          14.0
15                       32.0          83.0          98.0          14.0
16                        3.0          83.0          98.0          14.0
17                      110.0          83.0          98.0          14.0
18                      428.0          83.0          98.0          14.0
19                      429.0          83.0          98.0          14.0
20                      436.0          83.0          98.0          14.0
21                      438.0          83.0          98.0          14.0
22                       60.0          83.0          98.0          14.0
23                       59.0          83.0          98.0          14.0
24                       67.0          83.0          98.0          14.0
25                       60.0          83.0          98.0          14.0
26                       68.0          83.0          98.0          14.0
27                      185.0          83.0          98.0          14.0
28                      143.0          83.0          98.0          14.0
29                      142.0          83.0          98.0          14.0
...                       ...           ...           ...           ...
3671                    402.0          83.0          98.0          14.0
3672                    408.0          83.0          98.0          14.0
3673                    421.0          83.0          98.0          14.0
3674                    423.0          83.0          98.0          14.0
3675                    283.0          83.0          98.0          14.0
3676                    401.0          83.0          98.0          14.0
3677                    412.0          83.0          98.0          14.0
3678                    402.0          83.0          98.0          14.0
3679                    411.0          83.0          98.0          14.0
3680                    421.0          83.0          98.0          14.0
3681                    422.0          83.0          98.0          14.0
3682                     69.0          83.0          98.0          14.0
3683                    207.0          83.0          98.0          14.0
3684                    103.0          83.0          98.0          14.0
3685                    104.0          83.0          98.0          14.0
3686                    114.0          83.0          98.0          14.0
3687                    105.0          83.0          98.0          14.0
3688                    106.0          83.0          98.0          14.0
3689                     37.0          83.0          98.0          14.0
3690                     12.0          83.0          98.0          14.0
3691                      8.0          83.0          98.0          14.0
```

```
3692                     46.0          83.0          98.0          14.0
3693                     53.0          83.0          98.0          14.0
3694                     52.0          83.0          98.0          14.0
3695                    348.0          83.0          98.0          14.0
3696                    293.0          83.0          98.0          14.0
3697                    212.0          83.0          98.0          14.0
3698                    211.0          83.0          98.0          14.0
3699                    232.0          83.0          98.0          14.0
3700                    231.0          83.0          98.0          14.0

      Htchbk Pass Vol  Htchbk Lugg Vol  \
0                90.0             16.0
1                90.0             16.0
2                90.0             16.0
3                90.0             16.0
4                90.0             16.0
5                90.0             16.0
6                90.0             16.0
7                90.0             16.0
8                90.0             16.0
9                90.0             16.0
10               90.0             16.0
11               90.0             16.0
12               90.0             16.0
13               90.0             16.0
14               90.0             16.0
15               90.0             16.0
16               90.0             16.0
17               90.0             16.0
18               90.0             16.0
19               90.0             16.0
20               90.0             16.0
21               90.0             16.0
22               90.0             16.0
23               90.0             16.0
24               90.0             16.0
25               90.0             16.0
26               90.0             16.0
27               90.0             16.0
28               90.0             16.0
29               90.0             16.0
...               ...              ...
3671             90.0             16.0
3672             90.0             16.0
3673             90.0             16.0
3674             90.0             16.0
3675             90.0             16.0
3676             90.0             16.0
```

```
3677             90.0             16.0
3678             90.0             16.0
3679             90.0             16.0
3680             90.0             16.0
3681             90.0             16.0
3682             90.0             16.0
3683             90.0             16.0
3684             90.0             16.0
3685             90.0             16.0
3686             90.0             16.0
3687             90.0             16.0
3688             90.0             16.0
3689             90.0             16.0
3690             90.0             16.0
3691             90.0             16.0
3692             90.0             16.0
3693             90.0             16.0
3694             90.0             16.0
3695             90.0             16.0
3696             90.0             16.0
3697             90.0             16.0
3698             90.0             16.0
3699             90.0             16.0
3700             90.0             16.0

      Fuel2 Annual Fuel Cost - Alternative Fuel  \
0                                        2750.0
1                                        2750.0
2                                        2750.0
3                                        2750.0
4                                        2750.0
5                                        2750.0
6                                        2750.0
7                                        2750.0
8                                        2750.0
9                                        2750.0
10                                       2750.0
11                                       2750.0
12                                       2750.0
13                                       2750.0
14                                       2750.0
15                                       2750.0
16                                       2750.0
17                                       2750.0
18                                       2750.0
19                                       2750.0
20                                       2750.0
21                                       2750.0
```

```
22                                       2750.0
23                                       2750.0
24                                       2750.0
25                                       2750.0
26                                       2750.0
27                                       2750.0
28                                       2750.0
29                                       2750.0
...                                         ...
3671                                     2750.0
3672                                     2750.0
3673                                     2750.0
3674                                     2750.0
3675                                     2750.0
3676                                     2750.0
3677                                     2750.0
3678                                     2750.0
3679                                     2750.0
3680                                     2750.0
3681                                     2750.0
3682                                     2750.0
3683                                     2750.0
3684                                     2750.0
3685                                     2750.0
3686                                     2750.0
3687                                     2750.0
3688                                     2750.0
3689                                     2750.0
3690                                     2750.0
3691                                     3100.0
3692                                     2750.0
3693                                     2750.0
3694                                     2750.0
3695                                     2750.0
3696                                     2750.0
3697                                     2750.0
3698                                     2750.0
3699                                     2750.0
3700                                     2750.0

      $ You Save over 5 years (amount saved in fuel costs over 5 years - on label)   \
0                                                  750.0
1                                                 1000.0
2                                                 1000.0
3                                                 1000.0
4                                                 1000.0
5                                                 1000.0
6                                                 1000.0
```

```
7                                                 1000.0
8                                                 1000.0
9                                                 1000.0
10                                                1000.0
11                                                1000.0
12                                                1000.0
13                                                1000.0
14                                                1000.0
15                                                1000.0
16                                                   0.0
17                                                1000.0
18                                                   0.0
19                                                   0.0
20                                                1000.0
21                                                1000.0
22                                                1000.0
23                                                1000.0
24                                                1000.0
25                                                1000.0
26                                                1000.0
27                                                1000.0
28                                                1000.0
29                                                1000.0
...                                                  ...
3671                                              1000.0
3672                                              1000.0
3673                                              1000.0
3674                                              1000.0
3675                                              1000.0
3676                                              1000.0
3677                                              1000.0
3678                                              1000.0
3679                                              1000.0
3680                                              1000.0
3681                                              1000.0
3682                                              1000.0
3683                                              1000.0
3684                                              1000.0
3685                                              1000.0
3686                                              1000.0
3687                                               500.0
3688                                               750.0
3689                                              1000.0
3690                                              1000.0
3691                                              1000.0
3692                                              1000.0
3693                                              1000.0
3694                                              1000.0
```

```
3695                                              1000.0
3696                                              1000.0
3697                                              1000.0
3698                                              1000.0
3699                                              1000.0
3700                                              1000.0

      Eng Displ
0           1.8
1           6.0
2           4.7
3           4.7
4           4.7
5           4.7
6           4.7
7           4.7
8           4.2
9           4.2
10          5.2
11          5.2
12          4.2
13          4.2
14          5.2
15          5.2
16          2.0
17          4.0
18          2.0
19          2.0
20          3.0
21          3.0
22          8.0
23          6.2
24          6.2
25          6.2
26          6.2
27          8.4
28          4.5
29          4.5
...         ...
3671        3.5
3672        3.0
3673        3.0
3674        4.7
3675        5.6
3676        3.6
3677        3.6
3678        3.6
3679        3.6
```

```
3680        4.8
3681        4.8
3682        4.0
3683        4.0
3684        3.5
3685        3.5
3686        3.5
3687        3.5
3688        3.5
3689        5.7
3690        5.7
3691        5.7
3692        3.6
3693        2.0
3694        2.0
3695        3.5
3696        3.5
3697        2.0
3698        2.0
3699        2.5
3700        2.5

[3701 rows x 9 columns]>
```

```
In [195]: X_te_num.head
```

```
Out[195]: <bound method NDFrame.head of       Index (Model Type Index)  2Dr Pass Vol  4Dr Pass V
0                         57.0          83.0          99.0          14.0
1                        410.0          83.0          99.0          14.0
2                         65.0          83.0          99.0          14.0
3                         71.0          83.0          99.0          14.0
4                         66.0          83.0          99.0          14.0
5                         72.0          83.0          99.0          14.0
6                         46.0          83.0          99.0          14.0
7                        488.0          83.0          99.0          14.0
8                         38.0          83.0          99.0          14.0
9                        278.0          83.0          99.0          14.0
10                       223.0          83.0          99.0          14.0
11                       285.0          83.0          99.0          14.0
12                       276.0          83.0          99.0          14.0
13                       142.0          83.0          99.0          14.0
14                       143.0          83.0          99.0          14.0
15                       145.0          83.0          99.0          14.0
16                       144.0          83.0          99.0          14.0
17                       154.0          83.0          99.0          14.0
18                       405.0          83.0          99.0          14.0
19                       406.0          83.0          99.0          14.0
20                       322.0          43.0          99.0          14.0
```

```
21                       185.0         83.0          99.0          14.0
22                       161.0         83.0          99.0          14.0
23                       171.0         83.0          99.0          14.0
24                       184.0         83.0          99.0          14.0
25                       160.0         83.0          99.0          14.0
26                       170.0         83.0          99.0          14.0
27                       159.0         83.0          99.0          14.0
28                       158.0         83.0          99.0          14.0
29                       165.0         83.0          99.0          14.0
...                        ...          ...           ...           ...
1190                     271.0         83.0          99.0          14.0
1191                     406.0         83.0          99.0          14.0
1192                     272.0         83.0          99.0          14.0
1193                     274.0         83.0          99.0          14.0
1194                     424.0         83.0          99.0          14.0
1195                     435.0         83.0          99.0          14.0
1196                     436.0         83.0          99.0          14.0
1197                     821.0         83.0          99.0          14.0
1198                     402.0         83.0          99.0          14.0
1199                     421.0         83.0          99.0          14.0
1200                     423.0         83.0          99.0          14.0
1201                     283.0         83.0          99.0          14.0
1202                     401.0         83.0          99.0          14.0
1203                     412.0         83.0          99.0          14.0
1204                     402.0         83.0          99.0          14.0
1205                     411.0         83.0          99.0          14.0
1206                     421.0         83.0          99.0          14.0
1207                     422.0         83.0          99.0          14.0
1208                      58.0         83.0          99.0          14.0
1209                     207.0         83.0          99.0          14.0
1210                     102.0         83.0          99.0          14.0
1211                     103.0         83.0          99.0          14.0
1212                     108.0         83.0          99.0          14.0
1213                     111.0         83.0          99.0          14.0
1214                     114.0         83.0          99.0          14.0
1215                      86.0         83.0          99.0          14.0
1216                      37.0         83.0          99.0          14.0
1217                      33.0         83.0          99.0          14.0
1218                      53.0         83.0          99.0          14.0
1219                      52.0         83.0          99.0          14.0

      Htchbk Pass Vol  Htchbk Lugg Vol  \
0                92.0             19.0
1                92.0             19.0
2                92.0             19.0
3                92.0             19.0
4                92.0             19.0
5                92.0             19.0
```

```
6                92.0              19.0
7                92.0              19.0
8                92.0              19.0
9                92.0              19.0
10               92.0              19.0
11               92.0              19.0
12               92.0              19.0
13               92.0              19.0
14               92.0              19.0
15               92.0              19.0
16               92.0              19.0
17               92.0              19.0
18               92.0              19.0
19               92.0              19.0
20               92.0              19.0
21               92.0              19.0
22               92.0              19.0
23               92.0              19.0
24               92.0              19.0
25               92.0              19.0
26               92.0              19.0
27               92.0              19.0
28               92.0              19.0
29               92.0              19.0
...               ...               ...
1190             92.0              19.0
1191             92.0              19.0
1192             92.0              19.0
1193             92.0              19.0
1194             92.0              19.0
1195             92.0              19.0
1196             92.0              19.0
1197             92.0              19.0
1198             92.0              19.0
1199             92.0              19.0
1200             92.0              19.0
1201             92.0              19.0
1202             92.0              19.0
1203             92.0              19.0
1204             92.0              19.0
1205             92.0              19.0
1206             92.0              19.0
1207             92.0              19.0
1208             92.0              19.0
1209             92.0              19.0
1210             92.0              19.0
1211             92.0              19.0
1212             92.0              19.0
```

```
1213             92.0              19.0
1214             92.0              19.0
1215             92.0              19.0
1216             92.0              19.0
1217             92.0              19.0
1218             92.0              19.0
1219             92.0              19.0

      Fuel2 Annual Fuel Cost - Alternative Fuel  \
0                                        2100.0
1                                        2100.0
2                                        2100.0
3                                        2100.0
4                                        2100.0
5                                        2100.0
6                                        2100.0
7                                        2100.0
8                                        2100.0
9                                        2100.0
10                                       2100.0
11                                       2100.0
12                                       2100.0
13                                       2100.0
14                                       2100.0
15                                       2100.0
16                                       2100.0
17                                       2100.0
18                                       2100.0
19                                       2100.0
20                                       2100.0
21                                       2100.0
22                                       2100.0
23                                       2100.0
24                                       2100.0
25                                       2100.0
26                                       2100.0
27                                       2100.0
28                                       2100.0
29                                       2100.0
...                                         ...
1190                                     2100.0
1191                                     2100.0
1192                                     2100.0
1193                                     2100.0
1194                                     2100.0
1195                                     2100.0
1196                                     2100.0
1197                                     2100.0
```

```
1198                                     2100.0
1199                                     2100.0
1200                                     2100.0
1201                                     2100.0
1202                                     2100.0
1203                                     2100.0
1204                                     2100.0
1205                                     2100.0
1206                                     2100.0
1207                                     2100.0
1208                                     2100.0
1209                                     2100.0
1210                                     2100.0
1211                                     2100.0
1212                                     2100.0
1213                                     2100.0
1214                                     2100.0
1215                                     2100.0
1216                                     2100.0
1217                                     2900.0
1218                                     2100.0
1219                                     2100.0

      $ You Save over 5 years (amount saved in fuel costs over 5 years - on label)   \
0                                                 750.0
1                                                 750.0
2                                                 750.0
3                                                 750.0
4                                                 750.0
5                                                 750.0
6                                                 750.0
7                                                 750.0
8                                                 750.0
9                                                 750.0
10                                                750.0
11                                                750.0
12                                                750.0
13                                                750.0
14                                                750.0
15                                                750.0
16                                                750.0
17                                                750.0
18                                                750.0
19                                                750.0
20                                                750.0
21                                                750.0
22                                                750.0
23                                                750.0
```

```
24                                                 750.0
25                                                 750.0
26                                                 750.0
27                                                 750.0
28                                                 750.0
29                                                 750.0
...                                                  ...
1190                                               750.0
1191                                               750.0
1192                                               750.0
1193                                               750.0
1194                                               750.0
1195                                               750.0
1196                                               750.0
1197                                               750.0
1198                                               750.0
1199                                               750.0
1200                                               750.0
1201                                               750.0
1202                                               750.0
1203                                               750.0
1204                                               750.0
1205                                               750.0
1206                                               750.0
1207                                               750.0
1208                                               750.0
1209                                               750.0
1210                                               750.0
1211                                               750.0
1212                                               750.0
1213                                               250.0
1214                                               500.0
1215                                               750.0
1216                                               750.0
1217                                               750.0
1218                                               750.0
1219                                               750.0

      Eng Displ
0           3.5
1           1.8
2           5.2
3           5.2
4           5.2
5           5.2
6           2.0
7           3.0
8           8.0
```

```
9           6.2
10          6.2
11          6.2
12          6.2
13          3.9
14          3.9
15          3.9
16          3.9
17          6.5
18          1.4
19          1.4
20          3.5
21          2.0
22          3.0
23          3.0
24          2.0
25          3.0
26          3.0
27          5.0
28          5.0
29          3.0
...         ...
1190        3.0
1191        5.5
1192        5.5
1193        5.5
1194        5.5
1195        4.0
1196        4.0
1197        3.5
1198        3.5
1199        3.0
1200        4.7
1201        5.6
1202        3.6
1203        3.6
1204        3.6
1205        3.6
1206        4.8
1207        4.8
1208        4.0
1209        4.0
1210        3.5
1211        3.5
1212        3.5
1213        3.5
1214        3.5
1215        5.7
```

```
1216        5.7
1217        5.7
1218        2.0
1219        2.0

[1220 rows x 9 columns]>
```

```
In [40]: from sklearn.linear_model import Ridge
         from sklearn.preprocessing import StandardScaler
         from sklearn.model_selection import train_test_split

         scaler = StandardScaler()

         scaler.fit(X_tr_num)

         X_tr_num_sc = scaler.transform(X_tr_num)
```

```
In [41]: X_te_num_sc = scaler.transform(X_te_num)
```

```
In [42]: X_tr_num_sc = pd.DataFrame(X_tr_num_sc, columns=X_tr_num.columns)
```

```
In [43]: X_te_num_sc = pd.DataFrame(X_te_num_sc, columns=X_te_num.columns)
```

```
In [196]: X_tr_num_sc.head #train data numerical scaled
```

```
Out[196]: <bound method NDFrame.head of       Index (Model Type Index)  2Dr Pass Vol  4Dr Pass V
          0                     0.249234      0.014368     -0.050836     -0.173528
          1                    -0.907082      0.014368     -0.050836     -0.173528
          2                    -0.925150      0.014368     -0.050836     -0.173528
          3                    -0.938700      0.014368     -0.050836     -0.173528
          4                    -0.920633      0.014368     -0.050836     -0.173528
          5                    -0.934184      0.014368     -0.050836     -0.173528
          6                    -0.916116      0.014368     -0.050836     -0.173528
          7                    -0.929667      0.014368     -0.050836     -0.173528
          8                    -0.821262      0.014368     -0.050836     -0.173528
          9                    -0.812228      0.014368     -0.050836     -0.173528
          10                   -0.785127      0.014368     -0.050836     -0.173528
          11                   -0.794161      0.014368     -0.050836     -0.173528
          12                   -0.825779      0.014368     -0.050836     -0.173528
          13                   -0.816745      0.014368     -0.050836     -0.173528
          14                   -0.789644      0.014368     -0.050836     -0.173528
          15                   -0.798678      0.014368     -0.050836     -0.173528
          16                   -0.929667      0.014368     -0.050836     -0.173528
          17                   -0.446363      0.014368     -0.050836     -0.173528
          18                    0.989999      0.014368     -0.050836     -0.173528
          19                    0.994516      0.014368     -0.050836     -0.173528
          20                    1.026134      0.014368     -0.050836     -0.173528
          21                    1.035168      0.014368     -0.050836     -0.173528
          22                   -0.672206      0.014368     -0.050836     -0.173528
```

```
23                   -0.676723      0.014368     -0.050836     -0.173528
24                   -0.640588      0.014368     -0.050836     -0.173528
25                   -0.672206      0.014368     -0.050836     -0.173528
26                   -0.636071      0.014368     -0.050836     -0.173528
27                   -0.107598      0.014368     -0.050836     -0.173528
28                   -0.297306      0.014368     -0.050836     -0.173528
29                   -0.301823      0.014368     -0.050836     -0.173528
...                        ...           ...           ...           ...
3671                  0.872561      0.014368     -0.050836     -0.173528
3672                  0.899662      0.014368     -0.050836     -0.173528
3673                  0.958381      0.014368     -0.050836     -0.173528
3674                  0.967415      0.014368     -0.050836     -0.173528
3675                  0.335054      0.014368     -0.050836     -0.173528
3676                  0.868044      0.014368     -0.050836     -0.173528
3677                  0.917730      0.014368     -0.050836     -0.173528
3678                  0.872561      0.014368     -0.050836     -0.173528
3679                  0.913213      0.014368     -0.050836     -0.173528
3680                  0.958381      0.014368     -0.050836     -0.173528
3681                  0.962898      0.014368     -0.050836     -0.173528
3682                 -0.631554      0.014368     -0.050836     -0.173528
3683                 -0.008227      0.014368     -0.050836     -0.173528
3684                 -0.477981      0.014368     -0.050836     -0.173528
3685                 -0.473464      0.014368     -0.050836     -0.173528
3686                 -0.428295      0.014368     -0.050836     -0.173528
3687                 -0.468947      0.014368     -0.050836     -0.173528
3688                 -0.464430      0.014368     -0.050836     -0.173528
3689                 -0.776093      0.014368     -0.050836     -0.173528
3690                 -0.889015      0.014368     -0.050836     -0.173528
3691                 -0.907082      0.014368     -0.050836     -0.173528
3692                 -0.735442      0.014368     -0.050836     -0.173528
3693                 -0.703824      0.014368     -0.050836     -0.173528
3694                 -0.708341      0.014368     -0.050836     -0.173528
3695                  0.628650      0.014368     -0.050836     -0.173528
3696                  0.380223      0.014368     -0.050836     -0.173528
3697                  0.014357      0.014368     -0.050836     -0.173528
3698                  0.009840      0.014368     -0.050836     -0.173528
3699                  0.104695      0.014368     -0.050836     -0.173528
3700                  0.100178      0.014368     -0.050836     -0.173528

      Htchbk Pass Vol  Htchbk Lugg Vol  \
0            0.015271        -0.066354
1            0.015271        -0.066354
2            0.015271        -0.066354
3            0.015271        -0.066354
4            0.015271        -0.066354
5            0.015271        -0.066354
6            0.015271        -0.066354
7            0.015271        -0.066354
```

```
8            0.015271        -0.066354
9            0.015271        -0.066354
10           0.015271        -0.066354
11           0.015271        -0.066354
12           0.015271        -0.066354
13           0.015271        -0.066354
14           0.015271        -0.066354
15           0.015271        -0.066354
16           0.015271        -0.066354
17           0.015271        -0.066354
18           0.015271        -0.066354
19           0.015271        -0.066354
20           0.015271        -0.066354
21           0.015271        -0.066354
22           0.015271        -0.066354
23           0.015271        -0.066354
24           0.015271        -0.066354
25           0.015271        -0.066354
26           0.015271        -0.066354
27           0.015271        -0.066354
28           0.015271        -0.066354
29           0.015271        -0.066354
...               ...              ...
3671         0.015271        -0.066354
3672         0.015271        -0.066354
3673         0.015271        -0.066354
3674         0.015271        -0.066354
3675         0.015271        -0.066354
3676         0.015271        -0.066354
3677         0.015271        -0.066354
3678         0.015271        -0.066354
3679         0.015271        -0.066354
3680         0.015271        -0.066354
3681         0.015271        -0.066354
3682         0.015271        -0.066354
3683         0.015271        -0.066354
3684         0.015271        -0.066354
3685         0.015271        -0.066354
3686         0.015271        -0.066354
3687         0.015271        -0.066354
3688         0.015271        -0.066354
3689         0.015271        -0.066354
3690         0.015271        -0.066354
3691         0.015271        -0.066354
3692         0.015271        -0.066354
3693         0.015271        -0.066354
3694         0.015271        -0.066354
3695         0.015271        -0.066354
```

```
3696         0.015271        -0.066354
3697         0.015271        -0.066354
3698         0.015271        -0.066354
3699         0.015271        -0.066354
3700         0.015271        -0.066354

      Fuel2 Annual Fuel Cost - Alternative Fuel  \
0                                     -0.015760
1                                     -0.015760
2                                     -0.015760
3                                     -0.015760
4                                     -0.015760
5                                     -0.015760
6                                     -0.015760
7                                     -0.015760
8                                     -0.015760
9                                     -0.015760
10                                    -0.015760
11                                    -0.015760
12                                    -0.015760
13                                    -0.015760
14                                    -0.015760
15                                    -0.015760
16                                    -0.015760
17                                    -0.015760
18                                    -0.015760
19                                    -0.015760
20                                    -0.015760
21                                    -0.015760
22                                    -0.015760
23                                    -0.015760
24                                    -0.015760
25                                    -0.015760
26                                    -0.015760
27                                    -0.015760
28                                    -0.015760
29                                    -0.015760
...                                         ...
3671                                  -0.015760
3672                                  -0.015760
3673                                  -0.015760
3674                                  -0.015760
3675                                  -0.015760
3676                                  -0.015760
3677                                  -0.015760
3678                                  -0.015760
3679                                  -0.015760
3680                                  -0.015760
```

```
3681                                  -0.015760
3682                                  -0.015760
3683                                  -0.015760
3684                                  -0.015760
3685                                  -0.015760
3686                                  -0.015760
3687                                  -0.015760
3688                                  -0.015760
3689                                  -0.015760
3690                                  -0.015760
3691                                   1.699727
3692                                  -0.015760
3693                                  -0.015760
3694                                  -0.015760
3695                                  -0.015760
3696                                  -0.015760
3697                                  -0.015760
3698                                  -0.015760
3699                                  -0.015760
3700                                  -0.015760

      $ You Save over 5 years (amount saved in fuel costs over 5 years - on label)   \
0                                             -0.553136
1                                             -0.126290
2                                             -0.126290
3                                             -0.126290
4                                             -0.126290
5                                             -0.126290
6                                             -0.126290
7                                             -0.126290
8                                             -0.126290
9                                             -0.126290
10                                            -0.126290
11                                            -0.126290
12                                            -0.126290
13                                            -0.126290
14                                            -0.126290
15                                            -0.126290
16                                            -1.833677
17                                            -0.126290
18                                            -1.833677
19                                            -1.833677
20                                            -0.126290
21                                            -0.126290
22                                            -0.126290
23                                            -0.126290
24                                            -0.126290
25                                            -0.126290
```

```
26                                            -0.126290
27                                            -0.126290
28                                            -0.126290
29                                            -0.126290
...                                                 ...
3671                                          -0.126290
3672                                          -0.126290
3673                                          -0.126290
3674                                          -0.126290
3675                                          -0.126290
3676                                          -0.126290
3677                                          -0.126290
3678                                          -0.126290
3679                                          -0.126290
3680                                          -0.126290
3681                                          -0.126290
3682                                          -0.126290
3683                                          -0.126290
3684                                          -0.126290
3685                                          -0.126290
3686                                          -0.126290
3687                                          -0.979983
3688                                          -0.553136
3689                                          -0.126290
3690                                          -0.126290
3691                                          -0.126290
3692                                          -0.126290
3693                                          -0.126290
3694                                          -0.126290
3695                                          -0.126290
3696                                          -0.126290
3697                                          -0.126290
3698                                          -0.126290
3699                                          -0.126290
3700                                          -0.126290

      Eng Displ
0     -1.008939
1      2.074229
2      1.119915
3      1.119915
4      1.119915
5      1.119915
6      1.119915
7      1.119915
8      0.752871
9      0.752871
10     1.486959
```

```
11      1.486959
12      0.752871
13      0.752871
14      1.486959
15      1.486959
16     -0.862122
17      0.606054
18     -0.862122
19     -0.862122
20     -0.128034
21     -0.128034
22      3.542405
23      2.221047
24      2.221047
25      2.221047
26      2.221047
27      3.836040
28      0.973098
29      0.973098
...          ...
3671    0.239010
3672   -0.128034
3673   -0.128034
3674    1.119915
3675    1.780594
3676    0.312419
3677    0.312419
3678    0.312419
3679    0.312419
3680    1.193324
3681    1.193324
3682    0.606054
3683    0.606054
3684    0.239010
3685    0.239010
3686    0.239010
3687    0.239010
3688    0.239010
3689    1.854003
3690    1.854003
3691    1.854003
3692    0.312419
3693   -0.862122
3694   -0.862122
3695    0.239010
3696    0.239010
3697   -0.862122
3698   -0.862122
```

```
          3699  -0.495078
          3700  -0.495078

          [3701 rows x 9 columns]>
```

```
In [197]: X_te_num_sc.head
```

```
Out[197]: <bound method NDFrame.head of       Index (Model Type Index)  2Dr Pass Vol  4Dr Pass V
          0                    -0.685756      0.014368      0.174215     -0.173528
          1                     0.908696      0.014368      0.174215     -0.173528
          2                    -0.649621      0.014368      0.174215     -0.173528
          3                    -0.622520      0.014368      0.174215     -0.173528
          4                    -0.645104      0.014368      0.174215     -0.173528
          5                    -0.618003      0.014368      0.174215     -0.173528
          6                    -0.735442      0.014368      0.174215     -0.173528
          7                     1.261011      0.014368      0.174215     -0.173528
          8                    -0.771577      0.014368      0.174215     -0.173528
          9                     0.312470      0.014368      0.174215     -0.173528
          10                    0.064043      0.014368      0.174215     -0.173528
          11                    0.344088      0.014368      0.174215     -0.173528
          12                    0.303436      0.014368      0.174215     -0.173528
          13                   -0.301823      0.014368      0.174215     -0.173528
          14                   -0.297306      0.014368      0.174215     -0.173528
          15                   -0.288272      0.014368      0.174215     -0.173528
          16                   -0.292789      0.014368      0.174215     -0.173528
          17                   -0.247621      0.014368      0.174215     -0.173528
          18                    0.886112      0.014368      0.174215     -0.173528
          19                    0.890628      0.014368      0.174215     -0.173528
          20                    0.511212    -10.361108      0.174215     -0.173528
          21                   -0.107598      0.014368      0.174215     -0.173528
          22                   -0.216003      0.014368      0.174215     -0.173528
          23                   -0.170834      0.014368      0.174215     -0.173528
          24                   -0.112115      0.014368      0.174215     -0.173528
          25                   -0.220519      0.014368      0.174215     -0.173528
          26                   -0.175351      0.014368      0.174215     -0.173528
          27                   -0.225036      0.014368      0.174215     -0.173528
          28                   -0.229553      0.014368      0.174215     -0.173528
          29                   -0.197935      0.014368      0.174215     -0.173528
          ...                        ...           ...           ...           ...
          1190                  0.280852      0.014368      0.174215     -0.173528
          1191                  0.890628      0.014368      0.174215     -0.173528
          1192                  0.285369      0.014368      0.174215     -0.173528
          1193                  0.294403      0.014368      0.174215     -0.173528
          1194                  0.971932      0.014368      0.174215     -0.173528
          1195                  1.021617      0.014368      0.174215     -0.173528
          1196                  1.026134      0.014368      0.174215     -0.173528
          1197                  2.765126      0.014368      0.174215     -0.173528
          1198                  0.872561      0.014368      0.174215     -0.173528
```

```
1199                   0.958381      0.014368       0.174215     -0.173528
1200                   0.967415      0.014368       0.174215     -0.173528
1201                   0.335054      0.014368       0.174215     -0.173528
1202                   0.868044      0.014368       0.174215     -0.173528
1203                   0.917730      0.014368       0.174215     -0.173528
1204                   0.872561      0.014368       0.174215     -0.173528
1205                   0.913213      0.014368       0.174215     -0.173528
1206                   0.958381      0.014368       0.174215     -0.173528
1207                   0.962898      0.014368       0.174215     -0.173528
1208                  -0.681239      0.014368       0.174215     -0.173528
1209                  -0.008227      0.014368       0.174215     -0.173528
1210                  -0.482497      0.014368       0.174215     -0.173528
1211                  -0.477981      0.014368       0.174215     -0.173528
1212                  -0.455396      0.014368       0.174215     -0.173528
1213                  -0.441846      0.014368       0.174215     -0.173528
1214                  -0.428295      0.014368       0.174215     -0.173528
1215                  -0.554767      0.014368       0.174215     -0.173528
1216                  -0.776093      0.014368       0.174215     -0.173528
1217                  -0.794161      0.014368       0.174215     -0.173528
1218                  -0.703824      0.014368       0.174215     -0.173528
1219                  -0.708341      0.014368       0.174215     -0.173528

      Htchbk Pass Vol  Htchbk Lugg Vol  \
0            1.052295          1.68359
1            1.052295          1.68359
2            1.052295          1.68359
3            1.052295          1.68359
4            1.052295          1.68359
5            1.052295          1.68359
6            1.052295          1.68359
7            1.052295          1.68359
8            1.052295          1.68359
9            1.052295          1.68359
10           1.052295          1.68359
11           1.052295          1.68359
12           1.052295          1.68359
13           1.052295          1.68359
14           1.052295          1.68359
15           1.052295          1.68359
16           1.052295          1.68359
17           1.052295          1.68359
18           1.052295          1.68359
19           1.052295          1.68359
20           1.052295          1.68359
21           1.052295          1.68359
22           1.052295          1.68359
23           1.052295          1.68359
24           1.052295          1.68359
```

```
25           1.052295          1.68359
26           1.052295          1.68359
27           1.052295          1.68359
28           1.052295          1.68359
29           1.052295          1.68359
...               ...              ...
1190         1.052295          1.68359
1191         1.052295          1.68359
1192         1.052295          1.68359
1193         1.052295          1.68359
1194         1.052295          1.68359
1195         1.052295          1.68359
1196         1.052295          1.68359
1197         1.052295          1.68359
1198         1.052295          1.68359
1199         1.052295          1.68359
1200         1.052295          1.68359
1201         1.052295          1.68359
1202         1.052295          1.68359
1203         1.052295          1.68359
1204         1.052295          1.68359
1205         1.052295          1.68359
1206         1.052295          1.68359
1207         1.052295          1.68359
1208         1.052295          1.68359
1209         1.052295          1.68359
1210         1.052295          1.68359
1211         1.052295          1.68359
1212         1.052295          1.68359
1213         1.052295          1.68359
1214         1.052295          1.68359
1215         1.052295          1.68359
1216         1.052295          1.68359
1217         1.052295          1.68359
1218         1.052295          1.68359
1219         1.052295          1.68359

      Fuel2 Annual Fuel Cost - Alternative Fuel  \
0                                     -3.201663
1                                     -3.201663
2                                     -3.201663
3                                     -3.201663
4                                     -3.201663
5                                     -3.201663
6                                     -3.201663
7                                     -3.201663
8                                     -3.201663
9                                     -3.201663
```

```
10                                     -3.201663
11                                     -3.201663
12                                     -3.201663
13                                     -3.201663
14                                     -3.201663
15                                     -3.201663
16                                     -3.201663
17                                     -3.201663
18                                     -3.201663
19                                     -3.201663
20                                     -3.201663
21                                     -3.201663
22                                     -3.201663
23                                     -3.201663
24                                     -3.201663
25                                     -3.201663
26                                     -3.201663
27                                     -3.201663
28                                     -3.201663
29                                     -3.201663
...                                          ...
1190                                   -3.201663
1191                                   -3.201663
1192                                   -3.201663
1193                                   -3.201663
1194                                   -3.201663
1195                                   -3.201663
1196                                   -3.201663
1197                                   -3.201663
1198                                   -3.201663
1199                                   -3.201663
1200                                   -3.201663
1201                                   -3.201663
1202                                   -3.201663
1203                                   -3.201663
1204                                   -3.201663
1205                                   -3.201663
1206                                   -3.201663
1207                                   -3.201663
1208                                   -3.201663
1209                                   -3.201663
1210                                   -3.201663
1211                                   -3.201663
1212                                   -3.201663
1213                                   -3.201663
1214                                   -3.201663
1215                                   -3.201663
1216                                   -3.201663
```

```
1217                                   0.719449
1218                                  -3.201663
1219                                  -3.201663

      $ You Save over 5 years (amount saved in fuel costs over 5 years - on label)   \
0                                             -0.553136
1                                             -0.553136
2                                             -0.553136
3                                             -0.553136
4                                             -0.553136
5                                             -0.553136
6                                             -0.553136
7                                             -0.553136
8                                             -0.553136
9                                             -0.553136
10                                            -0.553136
11                                            -0.553136
12                                            -0.553136
13                                            -0.553136
14                                            -0.553136
15                                            -0.553136
16                                            -0.553136
17                                            -0.553136
18                                            -0.553136
19                                            -0.553136
20                                            -0.553136
21                                            -0.553136
22                                            -0.553136
23                                            -0.553136
24                                            -0.553136
25                                            -0.553136
26                                            -0.553136
27                                            -0.553136
28                                            -0.553136
29                                            -0.553136
...                                                 ...
1190                                          -0.553136
1191                                          -0.553136
1192                                          -0.553136
1193                                          -0.553136
1194                                          -0.553136
1195                                          -0.553136
1196                                          -0.553136
1197                                          -0.553136
1198                                          -0.553136
1199                                          -0.553136
1200                                          -0.553136
1201                                          -0.553136
```

```
1202                                          -0.553136
1203                                          -0.553136
1204                                          -0.553136
1205                                          -0.553136
1206                                          -0.553136
1207                                          -0.553136
1208                                          -0.553136
1209                                          -0.553136
1210                                          -0.553136
1211                                          -0.553136
1212                                          -0.553136
1213                                          -1.406830
1214                                          -0.979983
1215                                          -0.553136
1216                                          -0.553136
1217                                          -0.553136
1218                                          -0.553136
1219                                          -0.553136

      Eng Displ
0      0.239010
1     -1.008939
2      1.486959
3      1.486959
4      1.486959
5      1.486959
6     -0.862122
7     -0.128034
8      3.542405
9      2.221047
10     2.221047
11     2.221047
12     2.221047
13     0.532645
14     0.532645
15     0.532645
16     0.532645
17     2.441273
18    -1.302574
19    -1.302574
20     0.239010
21    -0.862122
22    -0.128034
23    -0.128034
24    -0.862122
25    -0.128034
26    -0.128034
27     1.340142
```

```
          28     1.340142
          29    -0.128034
          ...         ...
          1190  -0.128034
          1191   1.707185
          1192   1.707185
          1193   1.707185
          1194   1.707185
          1195   0.606054
          1196   0.606054
          1197   0.239010
          1198   0.239010
          1199  -0.128034
          1200   1.119915
          1201   1.780594
          1202   0.312419
          1203   0.312419
          1204   0.312419
          1205   0.312419
          1206   1.193324
          1207   1.193324
          1208   0.606054
          1209   0.606054
          1210   0.239010
          1211   0.239010
          1212   0.239010
          1213   0.239010
          1214   0.239010
          1215   1.854003
          1216   1.854003
          1217   1.854003
          1218  -0.862122
          1219  -0.862122

          [1220 rows x 9 columns]>
```

```
In [46]: d_tr_cat.head()
```

```
Out[46]:   Range1 - Model Type Driving Range - Conventional Fuel      Mfr Name  \
         0                                                NaN        FCA Italy
         1                                                NaN     aston martin
         2                                                NaN     aston martin
         3                                                NaN     aston martin
         4                                                NaN     aston martin

                            Division Verify Mfr Cd Transmission Air Aspir Method Trans  \
         0                Alfa Romeo           FTG    Auto(AM6)                TC    AM
         1  Aston Martin Lagonda Ltd           ASX    Auto(AM7)               NaN    AM
```

```
2  Aston Martin Lagonda Ltd           ASX    Auto(AM7)              NaN    AM
3  Aston Martin Lagonda Ltd           ASX   Manual(M6)              NaN     M
4  Aston Martin Lagonda Ltd           ASX    Auto(AM7)              NaN    AM

  Lockup Torque Converter Trans Creeper Gear Drive Sys          ...            \
0                       Y                  N         R          ...
1                       N                  N         R          ...
2                       N                  N         R          ...
3                       N                  N         R          ...
4                       N                  N         R          ...

  Fuel Metering Sys Cd Off Board Charge Capable (Y or N)  \
0                  GDI                               NaN
1                  MFI                               NaN
2                  MFI                               NaN
3                  MFI                               NaN
4                  MFI                               NaN

  Camless Valvetrain (Y or N)  \
0                           N
1                           N
2                           N
3                           N
4                           N

  Stop/Start System (Engine Management System) Code Model Year Carline Class  \
0                                                 N       2015             1
1                                                 N       2015             1
2                                                 N       2015             1
3                                                 N       2015             1
4                                                 N       2015             1

  # Cyl # Gears Max Ethanol % - Gasoline Exhaust Valves Per Cyl
0     4       6                     10.0                      2
1    12       7                     10.0                      2
2     8       7                     10.0                      2
3     8       6                     10.0                      2
4     8       7                     10.0                      2

[5 rows x 29 columns]
```

```
In [47]: d_te_cat.head()
```

```
Out[47]:   Range1 - Model Type Driving Range - Conventional Fuel             Mfr Name  \
         0                                                NaN                   Honda
         1                                                NaN              FCA US LLC
         2                                                NaN     Volkswagen Group of
         3                                                NaN     Volkswagen Group of
```

```
4                                                NaN     Volkswagen Group of

     Division Verify Mfr Cd Transmission Air Aspir Method Trans  \
0       Acura           HNX  Auto(AM-S9)               TC   AMS
1  ALFA ROMEO           CRX    Auto(AM6)               TC    AM
2        Audi           VGA  Auto(AM-S7)              NaN   AMS
3        Audi           VGA  Auto(AM-S7)              NaN   AMS
4        Audi           VGA  Auto(AM-S7)              NaN   AMS

  Lockup Torque Converter Trans Creeper Gear Drive Sys          ...             \
0                       Y                  N         A          ...
1                       Y                  N         R          ...
2                       Y                  N         A          ...
3                       Y                  N         R          ...
4                       Y                  N         A          ...

  Fuel Metering Sys Cd Off Board Charge Capable (Y or N)  \
0                  GDI                                  N
1                  GDI                                NaN
2                 GDPI                                NaN
3                 GDPI                                NaN
4                 GDPI                                NaN

  Camless Valvetrain (Y or N)  \
0                           N
1                           N
2                           N
3                           N
4                           N

  Stop/Start System (Engine Management System) Code Model Year Carline Class  \
0                                                 Y       2018             1
1                                                 N       2018             1
2                                                 N       2018             1
3                                                 N       2018             1
4                                                 N       2018             1

  # Cyl # Gears Max Ethanol % - Gasoline Exhaust Valves Per Cyl
0     6       9                     10.0                      2
1     4       6                     10.0                      2
2    10       7                     15.0                      2
3    10       7                     15.0                      2
4    10       7                     15.0                      2

[5 rows x 29 columns]
```

```
In [48]: d_tr_cat = d_tr_cat.apply(lambda x:x.fillna(x.value_counts().index[0]))
```

```
In [49]: d_te_cat = d_te_cat.apply(lambda x:x.fillna(x.value_counts().index[0]))
In [50]: d_tr_cat.isna().any()
```

```
Out[50]: Range1 - Model Type Driving Range - Conventional Fuel            False
         Mfr Name                                                         False
         Division                                                         False
         Verify Mfr Cd                                                    False
         Transmission                                                     False
         Air Aspir Method                                                 False
         Trans                                                            False
         Lockup Torque Converter                                          False
         Trans Creeper Gear                                               False
         Drive Sys                                                        False
         Fuel Usage  - Conventional Fuel                                  False
          Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel    False
          Fuel2 Usage - Alternative Fuel                                  False
         Car/Truck Category - Cash for Clunkers Bill.                     False
         Unique Label?                                                    False
         Label Recalc?                                                    False
         Cyl Deact?                                                       False
         Var Valve Timing?                                                False
         Var Valve Lift?                                                  False
         Fuel Metering Sys Cd                                             False
         Off Board Charge Capable (Y or N)                                False
         Camless Valvetrain (Y or N)                                      False
         Stop/Start System (Engine Management System) Code                False
         Model Year                                                       False
         Carline Class                                                    False
         # Cyl                                                            False
         # Gears                                                          False
         Max Ethanol % - Gasoline                                         False
         Exhaust Valves Per Cyl                                           False
         dtype: bool
```

```
In [51]: d_te_cat.isna().any()
```

```
Out[51]: Range1 - Model Type Driving Range - Conventional Fuel            False
         Mfr Name                                                         False
         Division                                                         False
         Verify Mfr Cd                                                    False
         Transmission                                                     False
         Air Aspir Method                                                 False
         Trans                                                            False
         Lockup Torque Converter                                          False
         Trans Creeper Gear                                               False
         Drive Sys                                                        False
         Fuel Usage  - Conventional Fuel                                  False
          Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel    False
          Fuel2 Usage - Alternative Fuel                                  False
         Car/Truck Category - Cash for Clunkers Bill.                     False
         Unique Label?                                                    False
         Label Recalc?                                                    False
         Cyl Deact?                                                       False
         Var Valve Timing?                                                False
         Var Valve Lift?                                                  False
         Fuel Metering Sys Cd                                             False
         Off Board Charge Capable (Y or N)                                False
         Camless Valvetrain (Y or N)                                      False
         Stop/Start System (Engine Management System) Code                False
         Model Year                                                       False
         Carline Class                                                    False
         # Cyl                                                            False
         # Gears                                                          False
         Max Ethanol % - Gasoline                                         False
         Exhaust Valves Per Cyl                                           False
         dtype: bool
```

```
In [52]: d_tr_cat.isnull().any()
```

```
Out[52]: Range1 - Model Type Driving Range - Conventional Fuel            False
         Mfr Name                                                         False
         Division                                                         False
         Verify Mfr Cd                                                    False
         Transmission                                                     False
         Air Aspir Method                                                 False
         Trans                                                            False
         Lockup Torque Converter                                          False
         Trans Creeper Gear                                               False
         Drive Sys                                                        False
         Fuel Usage  - Conventional Fuel                                  False
          Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel    False
          Fuel2 Usage - Alternative Fuel                                  False
         Car/Truck Category - Cash for Clunkers Bill.                     False
         Unique Label?                                                    False
         Label Recalc?                                                    False
         Cyl Deact?                                                       False
         Var Valve Timing?                                                False
         Var Valve Lift?                                                  False
         Fuel Metering Sys Cd                                             False
         Off Board Charge Capable (Y or N)                                False
         Camless Valvetrain (Y or N)                                      False
         Stop/Start System (Engine Management System) Code                False
         Model Year                                                       False
         Carline Class                                                    False
         # Cyl                                                            False
         # Gears                                                          False
         Max Ethanol % - Gasoline                                         False
```

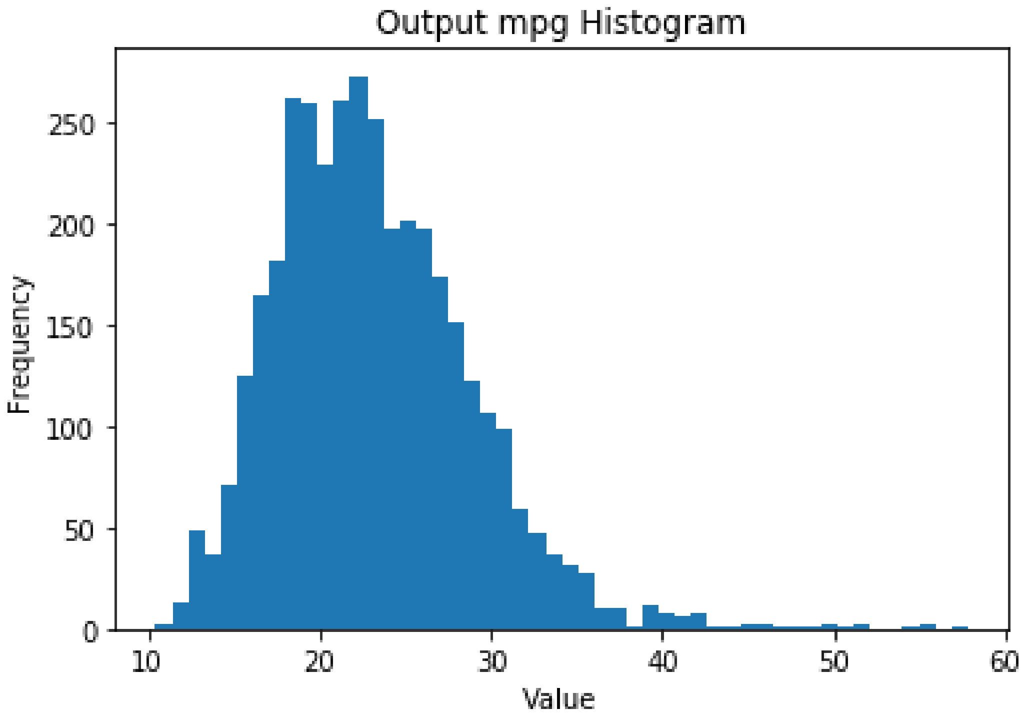
```
        Exhaust Valves Per Cyl                                         False
        dtype: bool

In [53]: d_tr_cat_dummied = pd.get_dummies(d_tr_cat, columns = list(d_tr_cat))

In [54]: d_te_cat_dummied = pd.get_dummies(d_te_cat, columns = list(d_te_cat))

In [55]: print(X_tr_num_sc.shape)
         d_tr_cat_dummied.shape

(3701, 9)


Out[55]: (3701, 373)

In [56]: print(X_te_num_sc.shape)
         d_te_cat_dummied.shape

(1220, 9)


Out[56]: (1220, 274)

In [57]: X_tr_num_sc = pd.DataFrame(X_tr_num_sc, columns=X_tr_num.columns)
         missing_cols = set( d_tr_cat_dummied.columns ) - set( d_te_cat_dummied.columns )
         # Add a missing column in test set with default value equal to 0
         for c in missing_cols:
             d_te_cat_dummied[c] = 0
         # Ensure the order of column in the test set is in the same order than in train set
         d_te_cat_dummied2 = d_te_cat_dummied[pd.DataFrame(d_tr_cat_dummied).columns]

In [58]: d_te_cat_dummied2.shape

Out[58]: (1220, 373)

In [59]: X_tr_complete = np.append(X_tr_num_sc, d_tr_cat_dummied, axis = 1)

In [60]: X_te_complete = np.append(X_te_num_sc, d_te_cat_dummied2, axis = 1)

In [61]: X_tr_complete.shape

Out[61]: (3701, 382)

In [62]: X_te_complete.shape

Out[62]: (1220, 382)
```

Plotting the distribution of the target

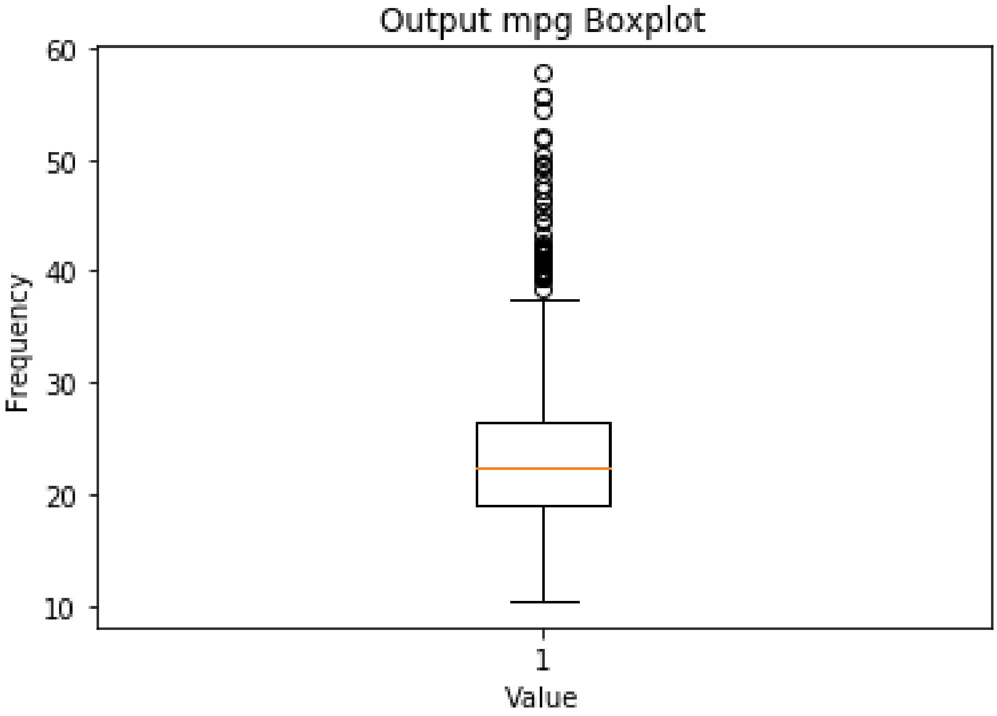
```
In [63]: import matplotlib.pyplot as plt
         %matplotlib
```

```
Using matplotlib backend: TkAgg
```

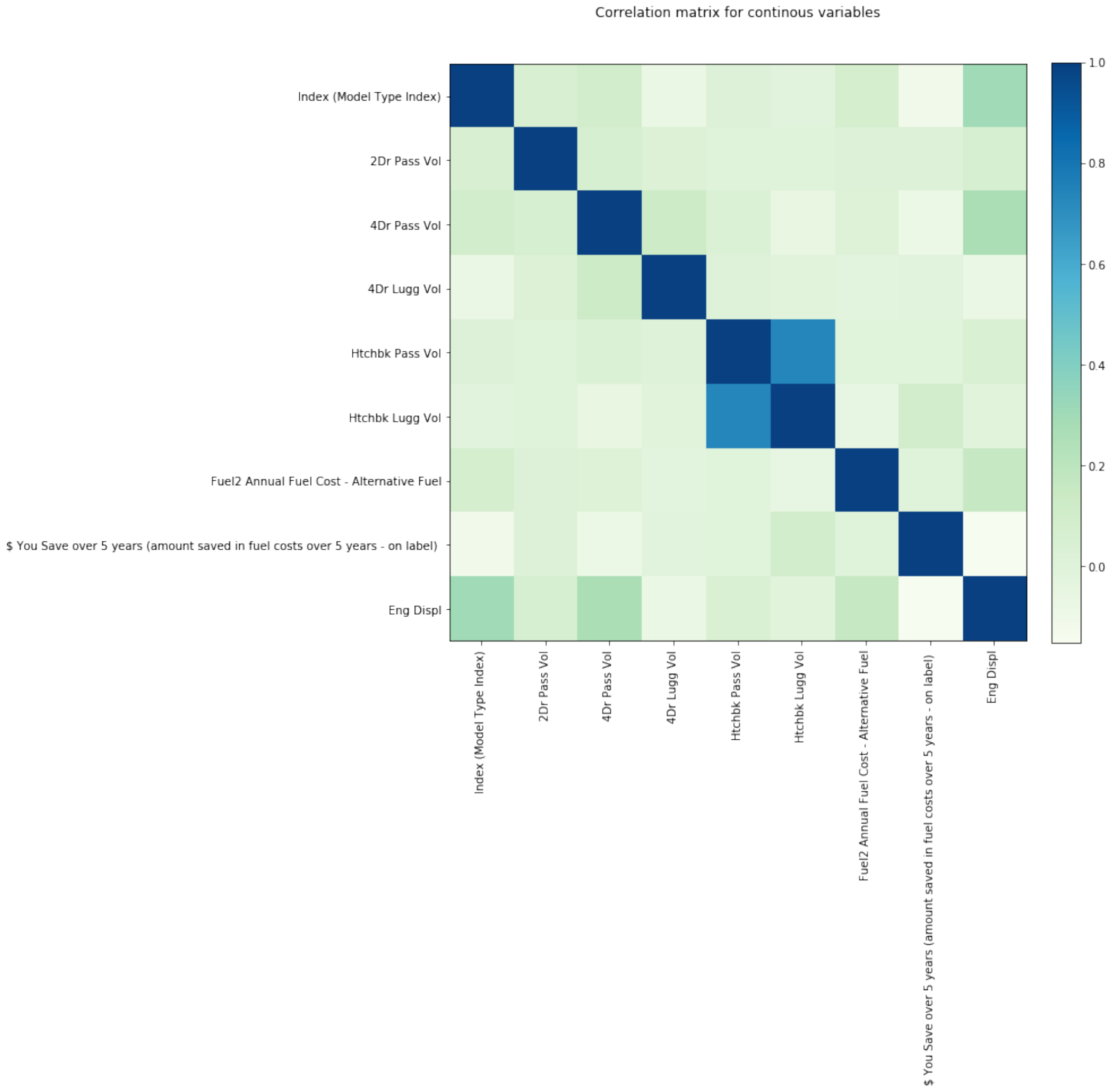
```
In [67]: #gaussian_numbers = np.random.randn(1000)
         plt.hist(y_tr, bins=50)
         plt.title("Output mpg Histogram")
         plt.xlabel("Value")
         plt.ylabel("Frequency")
         #plt.boxplot(y_train)
         fig = plt.gcf()
```



```
In [68]: plt.boxplot(y_tr)
         plt.title("Output mpg Boxplot")
         plt.xlabel("Value")
         plt.ylabel("Frequency")
         fig = plt.gcf()
```



```
In [69]: num_cols = X_tr_num.columns
         plt.matshow(X_tr_num[num_cols].corr(), cmap='GnBu')
         fig = plt.gcf()
         fig.set_size_inches(10,10)
         ax = plt.gca()
         ax.set_xticklabels(num_cols, rotation = 'vertical')
         ax.set_xticks(np.arange(len(num_cols)))
         ax.title.set_position([.5, 1.07]) #this adjusts title position. tune 1.07
         ax.set_yticklabels(num_cols)
         ax.set_yticks(np.arange(len(num_cols)))
         plt.title('Correlation matrix for continous variables')
         plt.colorbar(fraction=0.046, pad=0.04)
         ax.xaxis.set_ticks_position('bottom')
         plt.show()
```



```
In [70]: y_tr = d_train['Comb Unrd Adj FE - Conventional Fuel']
         y_te = d18['Comb Unrd Adj FE - Conventional Fuel']

In [89]: from sklearn.model_selection import GridSearchCV
         from sklearn.linear_model import Lasso

         param_ridge = {'alpha': np.logspace(-3, 3, 13)}
         grid_ridge = GridSearchCV(Ridge(), param_ridge, cv=10)
         grid_ridge.fit(X_tr_complete,y_tr)
         print(grid_ridge.best_params_)
         print(grid_ridge.best_score_)
```

```
{'alpha': 3.1622776601683795}
0.873000596902198
```

```
In [90]: print("The test score for Ridge Linear Model is "+str(grid_ridge.score(X_te_complete,y_
```

```
The test score for Ridge Linear Model is 0.8340506110223106
```

```
In [91]: from sklearn.model_selection import GridSearchCV
         from sklearn.linear_model import Lasso

         param_lasso = {'alpha': np.logspace(-3, 3, 13)}
         grid_lasso = GridSearchCV(Lasso(), param_lasso, cv=10)
         grid_lasso.fit(X_tr_complete,y_tr)
         print(grid_lasso.best_params_)
         print(grid_lasso.best_score_)
```

```
{'alpha': 0.001}
0.8720437508945863
```

```
In [92]: print("The test score for Lasso Linear Model is "+str(grid_lasso.score(X_te_complete,y_
```

```
The test score for Lasso Linear Model is 0.8517965884325657
```

Thus Lasso performs better than Ridge in linear case

2 Task 2

```
In [74]: from sklearn.preprocessing import PolynomialFeatures
```

```
In [83]: poly = PolynomialFeatures()
         X_tr_num_poly = poly.fit_transform(X_tr_num)
         X_te_num_poly = poly.transform(X_te_num)
```

```
In [84]: scaler = StandardScaler()
         X_tr_num_poly_sc = scaler.fit_transform(X_tr_num_poly)
         X_te_num_poly_sc = scaler.transform(X_te_num_poly)
```

```
In [85]: X_tr_complete_poly = np.append(X_tr_num_poly_sc, d_tr_cat_dummied, axis = 1)
         X_te_complete_poly = np.append(X_te_num_poly_sc, d_te_cat_dummied2, axis = 1)
```

```
In [86]: X_tr_complete_poly.shape
```

```
Out[86]: (3701, 428)
```

```
In [ ]:
```

```
In [87]: param_ridge = {'alpha': np.logspace(-3, 3, 15)}
         grid_ridge = GridSearchCV(Ridge(), param_ridge, cv=10)
         grid_ridge.fit(X_tr_complete_poly,y_tr)
         print(grid_ridge.best_params_)
         print(grid_ridge.best_score_)
```

```
{'alpha': 0.3727593720314938}
0.8902003054510154
```

```
In [88]: print("The test score for Ridge Linear Model is "+str(grid_ridge.score(X_te_complete_po
```

```
The test score for Ridge Linear Model is 0.8254235214509167
```

We notice that Polynomial features is causing the data to overfit

3 Task 3

```
In [93]: from sklearn.ensemble import GradientBoostingRegressor
         from sklearn.ensemble import RandomForestRegressor
         from sklearn.metrics import r2_score
```

```
In [123]: gb_grid = GridSearchCV(GradientBoostingRegressor(max_features=50, max_depth = 2,min_sa
          gb_grid.fit(X_tr_complete,y_tr)
          print(gb_grid.best_params_)
          print(gb_grid.best_score_)
```

```
{'n_estimators': 900}
0.9142360664487958
```

```
In [124]: gb_grid.score(X_te_complete,y_te)
```

```
Out[124]: 0.7286508818607985
```

Gradient Boosting seems to overfit the data a lot. We will attempt to make a grid search to find best parameters that don't underfit the model'

```
In [128]: from sklearn.model_selection import RandomizedSearchCV

          # Number of trees in random forest
          n_estimators = [int(x) for x in np.linspace(start = 200, stop = 2000, num = 10)]
          # Number of features to consider at every split
          max_features = ['auto', 'sqrt']
          # Maximum number of levels in tree
          max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
          max_depth.append(None)
          # Minimum number of samples required to split a node
          min_samples_split = [2, 5, 10]
          # Minimum number of samples required at each leaf node
          min_samples_leaf = [1, 2, 4]
          # Method of selecting samples for training each tree
          bootstrap = [True, False]
```

```
# Create the random grid
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}

print(random_grid)
```

```
{'max_features': ['auto', 'sqrt'], 'min_samples_split': [2, 5, 10], 'bootstrap': [True, False],
```

```
In [129]: # Use the random grid to search for best hyperparameters
          # First create the base model to tune
          rf = RandomForestRegressor()
          # Random search of parameters, using 3 fold cross validation,
          # search across 100 different combinations, and use all available cores
          rf_random = RandomizedSearchCV(estimator = rf, param_distributions = random_grid, n_it

          # Fit the random search model
          rf_random.fit(X_tr_complete, y_tr)
```

```
Fitting 3 folds for each of 100 candidates, totalling 300 fits
[CV] max_depth=30, min_samples_split=5, min_samples_leaf=1, n_estimators=400, max_features=sqrt,
[CV] max_depth=30, min_samples_split=5, min_samples_leaf=1, n_estimators=400, max_features=sqrt,
[CV] max_depth=30, min_samples_split=5, min_samples_leaf=1, n_estimators=400, max_features=sqrt,
[CV] max_depth=10, min_samples_split=5, min_samples_leaf=1, n_estimators=2000, max_features=sqrt
[CV]  max_depth=30, min_samples_split=5, min_samples_leaf=1, n_estimators=400, max_features=sqrt
[CV] max_depth=10, min_samples_split=5, min_samples_leaf=1, n_estimators=2000, max_features=sqrt
[CV]  max_depth=30, min_samples_split=5, min_samples_leaf=1, n_estimators=400, max_features=sqrt
[CV] max_depth=10, min_samples_split=5, min_samples_leaf=1, n_estimators=2000, max_features=sqrt
[CV]  max_depth=30, min_samples_split=5, min_samples_leaf=1, n_estimators=400, max_features=sqrt
[CV] max_depth=10, min_samples_split=5, min_samples_leaf=2, n_estimators=1200, max_features=sqrt
[CV]  max_depth=10, min_samples_split=5, min_samples_leaf=1, n_estimators=2000, max_features=sqr
[CV] max_depth=10, min_samples_split=5, min_samples_leaf=2, n_estimators=1200, max_features=sqrt
[CV]  max_depth=10, min_samples_split=5, min_samples_leaf=2, n_estimators=1200, max_features=sqr
[CV] max_depth=10, min_samples_split=5, min_samples_leaf=2, n_estimators=1200, max_features=sqrt
[CV]  max_depth=10, min_samples_split=5, min_samples_leaf=1, n_estimators=2000, max_features=sqr
[CV] max_depth=30, min_samples_split=2, min_samples_leaf=4, n_estimators=2000, max_features=auto
[CV]  max_depth=10, min_samples_split=5, min_samples_leaf=1, n_estimators=2000, max_features=sqr
[CV] max_depth=30, min_samples_split=2, min_samples_leaf=4, n_estimators=2000, max_features=auto
[CV]  max_depth=10, min_samples_split=5, min_samples_leaf=2, n_estimators=1200, max_features=sqr
[CV] max_depth=30, min_samples_split=2, min_samples_leaf=4, n_estimators=2000, max_features=auto
[CV]  max_depth=10, min_samples_split=5, min_samples_leaf=2, n_estimators=1200, max_features=sqr
[CV] max_depth=10, min_samples_split=2, min_samples_leaf=4, n_estimators=1600, max_features=sqrt
[CV]  max_depth=10, min_samples_split=2, min_samples_leaf=4, n_estimators=1600, max_features=sqr
[CV] max_depth=10, min_samples_split=2, min_samples_leaf=4, n_estimators=1600, max_features=sqrt
```

```
[CV]  max_depth=10, min_samples_split=2, min_samples_leaf=4, n_estimators=1600, max_features=sqr
[CV] max_depth=10, min_samples_split=2, min_samples_leaf=4, n_estimators=1600, max_features=sqrt
[CV]  max_depth=10, min_samples_split=2, min_samples_leaf=4, n_estimators=1600, max_features=sqr
[CV] max_depth=30, min_samples_split=5, min_samples_leaf=4, n_estimators=800, max_features=sqrt,
[CV]  max_depth=30, min_samples_split=5, min_samples_leaf=4, n_estimators=800, max_features=sqrt
[CV] max_depth=30, min_samples_split=5, min_samples_leaf=4, n_estimators=800, max_features=sqrt,
[CV]  max_depth=30, min_samples_split=5, min_samples_leaf=4, n_estimators=800, max_features=sqrt
[CV] max_depth=30, min_samples_split=5, min_samples_leaf=4, n_estimators=800, max_features=sqrt,
[CV]  max_depth=30, min_samples_split=5, min_samples_leaf=4, n_estimators=800, max_features=sqrt
[CV] max_depth=100, min_samples_split=5, min_samples_leaf=2, n_estimators=1000, max_features=sqr
[CV]  max_depth=100, min_samples_split=5, min_samples_leaf=2, n_estimators=1000, max_features=sq
[CV] max_depth=100, min_samples_split=5, min_samples_leaf=2, n_estimators=1000, max_features=sqr
[CV]  max_depth=100, min_samples_split=5, min_samples_leaf=2, n_estimators=1000, max_features=sq
[CV] max_depth=100, min_samples_split=5, min_samples_leaf=2, n_estimators=1000, max_features=sqr
[CV]  max_depth=100, min_samples_split=5, min_samples_leaf=2, n_estimators=1000, max_features=sq
[CV] max_depth=60, min_samples_split=5, min_samples_leaf=1, n_estimators=600, max_features=sqrt,
[CV]  max_depth=60, min_samples_split=5, min_samples_leaf=1, n_estimators=600, max_features=sqrt
[CV] max_depth=60, min_samples_split=5, min_samples_leaf=1, n_estimators=600, max_features=sqrt,
[CV]  max_depth=60, min_samples_split=5, min_samples_leaf=1, n_estimators=600, max_features=sqrt
[CV] max_depth=60, min_samples_split=5, min_samples_leaf=1, n_estimators=600, max_features=sqrt,
[CV]  max_depth=60, min_samples_split=5, min_samples_leaf=1, n_estimators=600, max_features=sqrt
[CV] max_depth=50, min_samples_split=2, min_samples_leaf=1, n_estimators=1000, max_features=auto
[CV]  max_depth=30, min_samples_split=2, min_samples_leaf=4, n_estimators=2000, max_features=aut
[CV] max_depth=50, min_samples_split=2, min_samples_leaf=1, n_estimators=1000, max_features=auto
[CV]  max_depth=30, min_samples_split=2, min_samples_leaf=4, n_estimators=2000, max_features=aut
[CV] max_depth=50, min_samples_split=2, min_samples_leaf=1, n_estimators=1000, max_features=auto
[CV]  max_depth=30, min_samples_split=2, min_samples_leaf=4, n_estimators=2000, max_features=aut
[CV] max_depth=10, min_samples_split=5, min_samples_leaf=4, n_estimators=1800, max_features=auto
[CV]  max_depth=50, min_samples_split=2, min_samples_leaf=1, n_estimators=1000, max_features=aut
[CV] max_depth=10, min_samples_split=5, min_samples_leaf=4, n_estimators=1800, max_features=auto
[CV]  max_depth=50, min_samples_split=2, min_samples_leaf=1, n_estimators=1000, max_features=aut
[CV] max_depth=10, min_samples_split=5, min_samples_leaf=4, n_estimators=1800, max_features=auto
[CV]  max_depth=50, min_samples_split=2, min_samples_leaf=1, n_estimators=1000, max_features=aut
[CV] max_depth=70, min_samples_split=10, min_samples_leaf=4, n_estimators=400, max_features=auto
[CV]  max_depth=70, min_samples_split=10, min_samples_leaf=4, n_estimators=400, max_features=aut
[CV] max_depth=70, min_samples_split=10, min_samples_leaf=4, n_estimators=400, max_features=auto
[CV]  max_depth=10, min_samples_split=5, min_samples_leaf=4, n_estimators=1800, max_features=aut
[CV] max_depth=70, min_samples_split=10, min_samples_leaf=4, n_estimators=400, max_features=auto
[CV]  max_depth=70, min_samples_split=10, min_samples_leaf=4, n_estimators=400, max_features=aut
[CV] max_depth=90, min_samples_split=5, min_samples_leaf=1, n_estimators=800, max_features=sqrt,
[CV]  max_depth=70, min_samples_split=10, min_samples_leaf=4, n_estimators=400, max_features=aut
[CV] max_depth=90, min_samples_split=5, min_samples_leaf=1, n_estimators=800, max_features=sqrt,
[CV]  max_depth=90, min_samples_split=5, min_samples_leaf=1, n_estimators=800, max_features=sqrt
[CV] max_depth=90, min_samples_split=5, min_samples_leaf=1, n_estimators=800, max_features=sqrt,
[CV]  max_depth=90, min_samples_split=5, min_samples_leaf=1, n_estimators=800, max_features=sqrt
[CV] max_depth=10, min_samples_split=10, min_samples_leaf=1, n_estimators=2000, max_features=sqr
```

```
[Parallel(n_jobs=-1)]: Done  33 tasks      | elapsed: 16.2min


[CV]  max_depth=90, min_samples_split=5, min_samples_leaf=1, n_estimators=800, max_features=sqrt
[CV] max_depth=10, min_samples_split=10, min_samples_leaf=1, n_estimators=2000, max_features=sqr
```

```
        ---------------------------------------------------------------------------

        KeyboardInterrupt                         Traceback (most recent call last)

        <ipython-input-129-71c0b0f18e05> in <module>()
          7
          8 # Fit the random search model
    ----> 9 rf_random.fit(X_tr_complete, y_tr)


        ~/.local/lib/python3.5/site-packages/sklearn/model_selection/_search.py in fit(self, X,
        637                                   error_score=self.error_score)
        638           for parameters, (train, test) in product(candidate_params,
    --> 639                                                    cv.split(X, y, groups)))
        640
        641         # if one choose to see train score, "out" will contain train score info


        ~/.local/lib/python3.5/site-packages/sklearn/externals/joblib/parallel.py in __call__(se
        787                 # consumption.
        788                 self._iterating = False
    --> 789             self.retrieve()
        790             # Make sure that we get a last message telling us we are done
        791             elapsed_time = time.time() - self._start_time


        ~/.local/lib/python3.5/site-packages/sklearn/externals/joblib/parallel.py in retrieve(se
        697             try:
        698                 if getattr(self._backend, 'supports_timeout', False):
    --> 699                     self._output.extend(job.get(timeout=self.timeout))
        700                 else:
        701                     self._output.extend(job.get())


        ~/.conda/envs/stuff/lib/python3.5/multiprocessing/pool.py in get(self, timeout)
        636
        637     def get(self, timeout=None):
    --> 638         self.wait(timeout)
        639         if not self.ready():
        640             raise TimeoutError
    ~/.conda/envs/stuff/lib/python3.5/multiprocessing/pool.py in wait(self, timeout)
    633
    634     def wait(self, timeout=None):
--> 635         self._event.wait(timeout)
    636
    637     def get(self, timeout=None):


    ~/.conda/envs/stuff/lib/python3.5/threading.py in wait(self, timeout)
    547             signaled = self._flag
    548             if not signaled:
--> 549                 signaled = self._cond.wait(timeout)
    550             return signaled
    551


    ~/.conda/envs/stuff/lib/python3.5/threading.py in wait(self, timeout)
    291         try:    # restore state no matter what (e.g., KeyboardInterrupt)
    292             if timeout is None:
--> 293                 waiter.acquire()
    294                 gotit = True
    295             else:


    KeyboardInterrupt:
```

```
In [177]: rf_grid = GridSearchCV(RandomForestRegressor(max_features =45), param_grid = {'n_estim
          rf_grid.fit(X_tr_complete,y_tr)
          print(rf_grid.best_params_)
          print(rf_grid.best_score_)
```

```
{'n_estimators': 90}
0.9457666197587219
```

```
In [178]: rf_grid.score(X_te_complete,y_te)
```

```
Out[178]: 0.8812460315714362
```

4 Task 4

```
In [179]: rf_grid.best_estimator_.feature_importances_
```

```
Out[179]: array([2.41755855e-02, 2.05651041e-03, 2.13574119e-02, 1.00172349e-02,
                 7.29946323e-03, 6.49493257e-03, 5.62503254e-04, 1.47017682e-01,
1.86390106e-01, 5.75697362e-07, 5.88611925e-07, 2.09484341e-08,
6.28058894e-07, 6.66013968e-06, 1.81446732e-06, 1.42174874e-07,
4.93685447e-07, 3.29638490e-06, 7.27004696e-06, 4.32823784e-06,
6.67369396e-05, 3.63762407e-06, 8.57635666e-07, 2.07697537e-06,
2.54703136e-06, 5.25054782e-06, 4.54075069e-06, 9.63392921e-07,
1.63925125e-06, 4.21148769e-07, 3.76106265e-06, 7.14894422e-06,
1.55398935e-04, 6.97681488e-05, 1.04539321e-05, 2.41531326e-07,
3.83974228e-07, 1.43586640e-06, 1.30214396e-06, 1.09610325e-05,
1.56599779e-05, 8.78547930e-08, 5.05976513e-07, 2.02697081e-06,
3.48398227e-06, 5.51768153e-06, 4.93115083e-07, 7.84726648e-06,
4.04833558e-06, 1.16976024e-05, 3.83136477e-07, 9.64098289e-08,
2.09924738e-07, 1.85559432e-05, 1.31493832e-07, 2.75835649e-07,
6.61983471e-06, 1.18824948e-06, 5.77254541e-07, 2.72811302e-07,
2.55225506e-07, 4.77480662e-06, 1.79162754e-07, 4.33512842e-06,
1.45253671e-05, 6.36902969e-07, 1.13975468e-06, 2.33126443e-06,
3.74885931e-06, 2.54766193e-06, 2.07283367e-06, 5.09488201e-07,
8.94492310e-07, 2.15542980e-07, 2.63748313e-06, 7.84702937e-07,
5.79123520e-07, 7.87715954e-06, 2.75190896e-06, 2.78144255e-06,
1.40759667e-05, 1.86117776e-03, 9.16314940e-07, 1.08808533e-03,
7.34521812e-04, 1.87048625e-03, 1.14373171e-03, 1.30609839e-03,
3.14986061e-04, 2.66092067e-04, 3.31370652e-04, 2.87434216e-06,
8.89934477e-04, 1.83966897e-04, 5.46753207e-05, 7.37157934e-04,
4.05049836e-04, 3.60314382e-05, 2.07337261e-03, 1.20409961e-07,
4.24547296e-04, 5.89793985e-06, 7.11739036e-04, 1.26999194e-04,
6.26119828e-04, 1.34252694e-03, 7.15312585e-04, 2.00122101e-04,
4.06074648e-04, 3.47218432e-05, 2.11652585e-04, 4.25488207e-06,
2.11922948e-04, 4.23702961e-04, 7.58248233e-04, 2.31325895e-04,
5.56134695e-05, 5.30073895e-04, 2.20793068e-06, 4.18235848e-04,
4.62411470e-04, 9.45339376e-05, 3.30379998e-04, 1.82603468e-04,
1.37311170e-03, 9.27343396e-04, 7.25760381e-06, 1.75804454e-04,
8.68561938e-04, 1.23125479e-03, 7.07432304e-04, 4.70406627e-04,
5.94826062e-04, 3.33595948e-04, 8.57840541e-04, 4.74885001e-04,
2.72673290e-04, 2.48623671e-04, 2.65351430e-06, 1.76710080e-04,
9.64373906e-04, 1.98733735e-05, 1.83777955e-06, 7.13929667e-04,
5.77760968e-04, 3.39222233e-04, 3.86171709e-05, 8.04411468e-04,
1.79496713e-06, 3.01013658e-04, 5.91130757e-05, 1.75431229e-04,
3.59695353e-05, 1.00949137e-04, 5.25413855e-04, 1.78150602e-03,
2.70229921e-04, 2.41469512e-04, 5.77274762e-05, 1.71688931e-03,
9.11530440e-04, 1.15926838e-03, 2.94293029e-04, 2.05990178e-03,
3.39537655e-06, 1.18743057e-03, 1.12414313e-03, 4.34489668e-04,
3.15904058e-04, 4.12302346e-04, 1.28184582e-05, 2.02478749e-04,
5.74786222e-05, 8.82811514e-04, 1.65579698e-05, 5.78037628e-04,
1.33797830e-03, 1.02963177e-05, 2.39428188e-04, 3.65091553e-06,
4.17543967e-05, 4.72708011e-04, 9.23294766e-04, 1.81331890e-03,
7.12235310e-04, 1.63933397e-04, 7.50427856e-05, 2.26826630e-04,
1.15683915e-03, 5.27314474e-04, 3.60965717e-04, 2.60919877e-04,
9.42613506e-05, 8.73757484e-04, 6.84093779e-05, 9.88163892e-07,
1.57295632e-04, 6.95106375e-03, 9.84784660e-04, 3.22042708e-06,
       1.52197756e-02, 1.89585800e-03, 8.82639656e-04, 4.31106821e-04,
       2.87286389e-05, 9.84637762e-06, 2.83799324e-05, 2.06447836e-03,
       3.11461636e-04, 1.15071934e-03, 1.07434002e-04, 7.49404494e-04,
       1.31128498e-03, 8.70300484e-05, 5.73577437e-04, 3.73699830e-04,
       3.13120296e-05, 4.40744036e-03, 2.89151436e-03, 8.65809293e-04,
       1.05678289e-02, 1.93959094e-03, 2.59501012e-03, 3.45549360e-03,
       6.10333570e-03, 5.86326186e-03, 3.00604204e-05, 2.32991216e-05,
       1.99847690e-03, 3.53546922e-06, 5.38695682e-03, 6.31802579e-02,
       6.83528412e-04, 1.36235525e-02, 4.78714487e-06, 4.32314469e-03,
       3.42943247e-03, 3.53001541e-05, 2.04513971e-03, 2.43136848e-03,
       6.57647735e-06, 2.07816323e-07, 1.62622354e-06, 3.14519215e-07,
       2.06902754e-06, 5.88013844e-06, 8.06027524e-06, 5.45274300e-06,
       9.16457482e-07, 6.19832751e-06, 1.56544387e-07, 6.23696402e-06,
       2.95371730e-07, 2.08839553e-05, 1.76917797e-06, 3.67295685e-04,
       1.54952857e-06, 4.57109673e-06, 1.97662318e-05, 4.83832947e-06,
       6.67150366e-07, 9.38232537e-08, 1.32241561e-06, 3.01821826e-06,
       1.23713451e-05, 2.29556344e-06, 3.02354987e-06, 6.35956639e-07,
       1.70852022e-05, 6.09585837e-08, 1.02945957e-06, 2.94239235e-05,
       3.28970128e-08, 1.43526922e-05, 1.15996951e-06, 3.88412610e-08,
       2.55699045e-06, 2.94607609e-07, 9.42720907e-07, 1.06209549e-07,
       6.57586083e-07, 1.27801243e-06, 3.05573492e-06, 2.11064163e-07,
       1.64392082e-07, 4.84382060e-07, 2.41761916e-07, 1.38772527e-06,
       3.01404873e-07, 7.52214659e-06, 1.45257767e-07, 4.79366078e-07,
       1.14311235e-06, 8.76138856e-07, 8.55618419e-07, 4.55852723e-06,
       5.20703306e-08, 6.00007248e-06, 6.41321169e-07, 2.13957985e-07,
       5.63317451e-06, 1.34380631e-06, 7.34585893e-07, 8.37543624e-06,
       4.82686128e-06, 6.48402852e-07, 6.58042592e-06, 1.77412882e-05,
       4.11674644e-06, 8.01959116e-06, 1.81239140e-04, 1.33413027e-03,
       1.22236271e-03, 1.27214837e-03, 1.38172215e-03, 2.12677970e-05,
       2.25272600e-04, 1.32348669e-04, 4.96657654e-03, 3.61947940e-03,
       1.68307674e-03, 2.14359232e-03, 2.34968785e-03, 3.41405985e-03,
       3.71610570e-03, 9.91676246e-05, 3.58558693e-03, 4.82615058e-04,
       3.95851671e-03, 3.28403847e-05, 3.64523157e-05, 5.38900548e-04,
       5.08333763e-04, 1.46708688e-02, 1.50608991e-02, 2.42444599e-03,
       1.24021894e-03, 2.47111066e-03, 1.17610228e-03, 4.09032795e-04,
       7.80912709e-04, 4.08503715e-03, 6.47815556e-03, 1.50299523e-03,
       5.95127837e-04, 1.47518490e-04, 5.73945713e-04, 5.08785825e-04,
       1.68931939e-04, 7.64239514e-04, 1.29890437e-04, 9.97863543e-04,
       7.30058095e-04, 1.42837726e-05, 3.02440707e-04, 1.34763493e-04,
       2.00550703e-05, 3.23474742e-03, 3.04692490e-03, 6.09300080e-04,
       5.88340914e-03, 6.05797265e-03, 9.56926263e-02, 9.71728271e-05,
       2.86443587e-02, 6.69101449e-02, 2.21933422e-04, 1.34855156e-02,
       9.84596841e-05, 1.56668352e-02, 8.59548033e-05, 7.06921027e-04,
       2.53936785e-03, 1.94491628e-03, 1.11305682e-03, 3.84081639e-04,
       3.14532201e-05, 4.34401279e-03, 3.81822727e-03, 1.39467111e-04,
       6.39147139e-03, 1.24985395e-02])
```

```
In [181]: X_tr_com_cols = list(X_tr_num_sc) + list(d_tr_cat_dummied)
In [182]: len(X_tr_com_cols)

Out[182]: 382

In [183]: imp__feat_list = sorted(zip(map(lambda x: round(x, 4), rf_grid.best_estimator_.feature
                      reverse=True)

In [184]: imp__feat_list

Out[184]: [(0.1864, 'Eng Displ'),
           (0.147,
            '$ You Save over 5 years (amount saved in fuel costs over 5 years - on label) '),
           (0.0957, '# Cyl_4'),
           (0.0669, '# Cyl_8'),
           (0.0632, 'Drive Sys_F'),
           (0.0286, '# Cyl_6'),
           (0.0242, 'Index (Model Type Index)'),
           (0.0214, '4Dr Pass Vol'),
           (0.0157, '# Gears_1'),
           (0.0152, 'Transmission_Auto(AV)'),
           (0.0151, 'Stop/Start System (Engine Management System) Code_Y'),
           (0.0147, 'Stop/Start System (Engine Management System) Code_N'),
           (0.0136, 'Drive Sys_R'),
           (0.0135, '# Cyl_12'),
           (0.0125, 'Exhaust Valves Per Cyl_2'),
           (0.0106, 'Trans_CVT'),
           (0.01, '4Dr Lugg Vol'),
           (0.0073, 'Htchbk Pass Vol'),
           (0.007, 'Transmission_Auto(AM6)'),
           (0.0065, 'Htchbk Lugg Vol'),
           (0.0065, 'Carline Class_5'),
           (0.0064, 'Exhaust Valves Per Cyl_1'),
           (0.0061, 'Lockup Torque Converter_N'),
           (0.0061, '# Cyl_3'),
           (0.0059, 'Lockup Torque Converter_Y'),
           (0.0059, 'Carline Class_33'),
           (0.0054, 'Drive Sys_A'),
           (0.005, 'Cyl Deact?_N'),
           (0.0044, 'Trans_A'),
           (0.0043, 'Max Ethanol % - Gasoline_10.0'),
           (0.0043, 'Fuel Usage  - Conventional Fuel_DU'),
           (0.0041, 'Carline Class_4'),
           (0.004, 'Fuel Metering Sys Cd_MFI'),
           (0.0038, 'Max Ethanol % - Gasoline_15.0'),
           (0.0037, 'Fuel Metering Sys Cd_CRDI'),
           (0.0036, 'Fuel Metering Sys Cd_GDI'),
           (0.0036, 'Cyl Deact?_Y'),
           (0.0035, 'Trans_SCV'),
           (0.0034, 'Var Valve Lift?_Y'),
(0.0034, 'Fuel Usage  - Conventional Fuel_G'),
(0.0032, 'Carline Class_30'),
(0.003, 'Carline Class_31'),
(0.0029, 'Trans_AM'),
(0.0026, 'Trans_SA'),
(0.0025, 'Model Year_2017'),
(0.0025, '# Gears_6'),
(0.0024, 'Model Year_2015'),
(0.0024, 'Fuel Usage  - Conventional Fuel_GPR'),
(0.0023, 'Var Valve Lift?_N'),
(0.0021, 'Verify Mfr Cd_FMX'),
(0.0021, 'Var Valve Timing?_Y'),
(0.0021, 'Transmission_Auto(S6)'),
(0.0021, 'Mfr Name_Nissan'),
(0.0021, '2Dr Pass Vol'),
(0.002, 'Fuel Usage  - Conventional Fuel_GP'),
(0.002, 'Drive Sys_4'),
(0.0019, 'Transmission_Auto(AV-S6)'),
(0.0019, 'Trans_M'),
(0.0019, 'Mfr Name_Ford Motor Company'),
(0.0019, 'Mfr Name_BMW'),
(0.0019, '# Gears_7'),
(0.0018, 'Verify Mfr Cd_TYX'),
(0.0018, 'Division_TOYOTA'),
(0.0017, 'Verify Mfr Cd_BMX'),
(0.0017, 'Var Valve Timing?_N'),
(0.0015, 'Carline Class_6'),
(0.0014, 'Unique Label?_Y'),
(0.0014, 'Division_Ferrari North America, Inc.'),
(0.0013, 'Verify Mfr Cd_NSX'),
(0.0013, 'Unique Label?_N'),
(0.0013, 'Transmission_Manual(M6)'),
(0.0013, 'Mfr Name_Toyota'),
(0.0013, 'Mfr Name_Honda'),
(0.0013, 'Car/Truck Category - Cash for Clunkers Bill._??'),
(0.0012, 'Verify Mfr Cd_GMX'),
(0.0012, 'Verify Mfr Cd_FEX'),
(0.0012, 'Transmission_Auto(S8)'),
(0.0012, 'Transmission_Auto(A6)'),
(0.0012, 'Model Year_2016'),
(0.0012, 'Division_Honda'),
(0.0012, 'Carline Class_1'),
(0.0012, 'Car/Truck Category - Cash for Clunkers Bill._car'),
(0.0011, 'Verify Mfr Cd_HNX'),
(0.0011, 'Mfr Name_General Motors'),
(0.0011, 'Mfr Name_FCA US LLC'),
(0.0011, '# Gears_8'),
(0.001, 'Transmission_Auto(AM7)'),
(0.001, 'Division_MAZDA'),
(0.001, 'Carline Class_15'),
(0.0009, 'Verify Mfr Cd_TKX'),
(0.0009, 'Verify Mfr Cd_MBX'),
(0.0009, 'Verify Mfr Cd_CRX'),
(0.0009, 'Transmission_Auto(AV-S7)'),
(0.0009, 'Transmission_Auto(AM-S7)'),
(0.0009, 'Trans_AMS'),
(0.0009, 'Mfr Name_MAZDA'),
(0.0009, 'Division_LEXUS'),
(0.0009, 'Division_HYUNDAI MOTOR COMPANY'),
(0.0009, 'Division_Ford'),
(0.0008, 'Division_NISSAN'),
(0.0008, 'Division_BMW'),
(0.0008, 'Carline Class_3'),
(0.0008, 'Carline Class_13'),
(0.0007, 'Verify Mfr Cd_VGA'),
(0.0007, 'Transmission_Manual(M5)'),
(0.0007, 'Mfr Name_Volkswagen Group of'),
(0.0007, 'Mfr Name_Rolls-Royce'),
(0.0007, 'Mfr Name_Mercedes-Benz'),
(0.0007, 'Mfr Name_Ferrari'),
(0.0007, 'Drive Sys_P'),
(0.0007, 'Division_Mercedes-Benz'),
(0.0007, 'Division_INFINITI'),
(0.0007, 'Carline Class_17'),
(0.0007, '# Gears_5'),
(0.0006, 'Verify Mfr Cd_MTX'),
(0.0006, 'Mfr Name_Subaru'),
(0.0006, 'Fuel2 Annual Fuel Cost - Alternative Fuel'),
(0.0006, 'Division_Mini'),
(0.0006, 'Division_Jeep'),
(0.0006, 'Carline Class_7'),
(0.0006, 'Carline Class_32'),
(0.0006, 'Carline Class_10'),
(0.0006, 'Air Aspir Method_SC'),
(0.0005, 'Verify Mfr Cd_RRG'),
(0.0005, 'Transmission_Auto(A7)'),
(0.0005, 'Fuel Metering Sys Cd_GDPI'),
(0.0005, 'Division_Subaru'),
(0.0005, 'Division_Lamborghini'),
(0.0005, 'Division_Jaguar'),
(0.0005, 'Division_Chevrolet'),
(0.0005, 'Division_Buick'),
(0.0005, 'Carline Class_11'),
(0.0005, 'Camless Valvetrain (Y or N)_Y'),
(0.0005, 'Camless Valvetrain (Y or N)_N'),
(0.0004, 'Verify Mfr Cd_KMX'),
(0.0004, 'Verify Mfr Cd_HYX'),
(0.0004, 'Transmission_Auto(AV-S8)'),
(0.0004, 'Transmission_Auto(A8)'),
(0.0004, 'Mfr Name_aston martin'),
(0.0004, 'Mfr Name_Porsche'),
(0.0004, 'Mfr Name_Mitsubishi Motors Co'),
(0.0004, 'Division_Cadillac'),
(0.0004, 'Division_Audi'),
(0.0004, 'Carline Class_2'),
(0.0004, 'Air Aspir Method_TC'),
(0.0004, '# Gears_9'),
(0.0004, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_280'),
(0.0003, 'Verify Mfr Cd_JLX'),
(0.0003, 'Verify Mfr Cd_FJX'),
(0.0003, 'Transmission_Auto(S7)'),
(0.0003, 'Transmission_Auto(A9)'),
(0.0003, 'Mfr Name_Kia'),
(0.0003, 'Mfr Name_Jaguar Land Rover L'),
(0.0003, 'Mfr Name_Hyundai'),
(0.0003, 'Division_Volkswagen'),
(0.0003, 'Division_Porsche'),
(0.0003, 'Division_Mitsubishi Motors Corporation'),
(0.0003, 'Division_Land Rover'),
(0.0003, 'Division_KIA MOTORS CORPORATION'),
(0.0003, 'Division_Dodge'),
(0.0003, 'Carline Class_19'),
(0.0002, 'Verify Mfr Cd_VVX'),
(0.0002, 'Verify Mfr Cd_PRX'),
(0.0002, 'Verify Mfr Cd_MAX'),
(0.0002, 'Transmission_Auto(AM5)'),
(0.0002, 'Transmission_Auto(A5)'),
(0.0002, 'Range1 - Model Type Driving Range - Conventional Fuel_400'),
(0.0002, 'Mfr Name_Volvo'),
(0.0002, 'Mfr Name_Maserati'),
(0.0002, 'Label Recalc?_N'),
(0.0002, 'Division_Volvo Cars of North America, LLC'),
(0.0002, 'Division_Rolls-Royce Motor Cars Limited'),
(0.0002, 'Division_MASERATI'),
(0.0002, 'Division_Lincoln'),
(0.0002, 'Division_GMC'),
(0.0002, 'Division_FIAT'),
(0.0002, 'Division_Bentley'),
(0.0002, 'Division_Aston Martin Lagonda Ltd'),
(0.0002, 'Division_Acura'),
(0.0002, 'Carline Class_12'),
(0.0002, 'Car/Truck Category - Cash for Clunkers Bill._1'),
(0.0002, '# Cyl_10'),
(0.0001, 'Verify Mfr Cd_MBV'),
(0.0001, 'Verify Mfr Cd_ASX'),
(0.0001, 'Transmission_Manual(M7)'),
(0.0001, 'Transmission_Auto(S9)'),
(0.0001, 'Transmission_Auto(AM-S8)'),
(0.0001, 'Transmission_Auto(AM-S6)'),
(0.0001, 'Transmission_Auto(A4)'),
(0.0001, 'Range1 - Model Type Driving Range - Conventional Fuel_402'),
(0.0001, 'Range1 - Model Type Driving Range - Conventional Fuel_370'),
(0.0001, 'Mfr Name_Roush'),
(0.0001, 'Mfr Name_McLaren Automotive'),
(0.0001, 'Max Ethanol % - Gasoline_85.0'),
(0.0001, 'Label Recalc?_Y'),
(0.0001, 'Fuel Metering Sys Cd_DDI'),
(0.0001, 'Division_SCION'),
(0.0001, 'Division_RAM'),
(0.0001, 'Division_Chrysler'),
(0.0001, 'Division_Bugatti'),
(0.0001, 'Carline Class_8'),
(0.0001, 'Carline Class_20'),
(0.0001, 'Carline Class_14'),
(0.0001, '# Gears_4'),
(0.0001, '# Cyl_5'),
(0.0001, '# Cyl_16'),
(0.0, 'Verify Mfr Cd_RII'),
(0.0, 'Verify Mfr Cd_QTM'),
(0.0, 'Verify Mfr Cd_PGN'),
(0.0, 'Verify Mfr Cd_MLN'),
(0.0, 'Verify Mfr Cd_LTX'),
(0.0, 'Verify Mfr Cd_FTG'),
(0.0, 'Transmission_Auto(S5)'),
(0.0, 'Transmission_Auto(S4)'),
(0.0, 'Transmission_Auto(S10)'),
(0.0, 'Transmission_Auto(AM8)'),
(0.0, 'Transmission_Auto(AM-S9)'),
(0.0, 'Trans Creeper Gear_Y'),
(0.0, 'Trans Creeper Gear_N'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_610'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_604'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_580'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_572'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_570'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_558'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_520/680'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_520/640'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_494/646'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_494/608'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_492'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_490/650'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_490'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_489'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_484'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_482/716'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_476/482'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_470/610'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_470'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_468/612'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_467'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_460'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_459'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_458/680'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_455'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_452/458'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_450'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_443'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_442'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_437'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_436'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_434/644'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_432'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_430'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_428/434'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_427'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_426'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_422'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_420'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_410/609'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_410'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_409'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_408'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_406'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_405/410'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_399'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_396/570'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_395'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_391'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_390'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_389'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_388'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_386/573'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_384'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_383'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_380'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_379'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_372'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_368'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_363'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_361'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_360'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_357'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_353'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_348'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_347'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_324'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_318'),
(0.0, 'Range1 - Model Type Driving Range - Conventional Fuel_193'),
(0.0, 'Off Board Charge Capable (Y or N)_Y'),
(0.0, 'Off Board Charge Capable (Y or N)_N'),
(0.0, 'Mfr Name_Quantum Fuel System'),
(0.0, 'Mfr Name_Pagani Automobili S'),
(0.0, 'Mfr Name_Mobility Ventures L'),
(0.0, 'Mfr Name_Lotus'),
(0.0, 'Mfr Name_FCA Italy'),
(0.0, 'Label Recalc?_Mod'),
(0.0, 'Fuel Usage  - Conventional Fuel_GM'),
(0.0, 'Fuel Usage  - Conventional Fuel_CNG'),
(0.0, 'Drive Sys_4'),
(0.0, 'Division_Roush Industries, Inc.'),
(0.0, 'Division_Pagani Automobili S.p.A.'),
(0.0, 'Division_Mobility Ventures LLC'),
(0.0, 'Division_McLaren Automotive Limted'),
(0.0, 'Division_McLaren'),
(0.0, 'Division_Lotus Cars Ltd'),
(0.0, 'Division_GENESIS'),
(0.0, 'Division_CHEVROLET'),
(0.0, 'Division_Alfa Romeo'),
(0.0, 'Division_ALFA ROMEO'),
(0.0, 'Carline Class_21'),
(0.0, 'Carline Class_18'),
(0.0, 'Air Aspir Method_TS'),
(0.0, '# Gears_10'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_450'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_445'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_410'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_403'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_394'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_382'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_380'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_372'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_370'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_369'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_364/476'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_364/448'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_362/537'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_360/480'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_360'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_357/362'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_357'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_350'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_340/440'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_340'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_338/442'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_338/416'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_338'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_337/501'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_333/337'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_332'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_330'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_320'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_317'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_316'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_314'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_313/465'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_312/408'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_310'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_309/313'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_305'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_304'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_301'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_300'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_298'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_296'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_295'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_292'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_291'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_290/418'),
(0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_290'),
(0.0,
 ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_289/430'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_287'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_285'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_284'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_283'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_279'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_276'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_274'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_273'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_270'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_269'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_266'),
           (0.0,
            ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_264/380'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_264'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_262'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_260'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_253'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_250'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_241'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_234'),
           (0.0, ' Range2 - Alt Fuel Model Typ Driving Range - Alternative Fuel_119'),
           (0.0, ' Fuel2 Usage - Alternative Fuel_E'),
           (0.0, ' Fuel2 Usage - Alternative Fuel_CNG')]
```

```
In [185]: top_20 = []
          for i in range(20):
              top_20.append(imp__feat_list[i][1])
```

```
In [186]: top_20
```

```
Out[186]: ['Eng Displ',
           '$ You Save over 5 years (amount saved in fuel costs over 5 years - on label) ',
           '# Cyl_4',
           '# Cyl_8',
           'Drive Sys_F',
           '# Cyl_6',
           'Index (Model Type Index)',
           '4Dr Pass Vol',
           '# Gears_1',
           'Transmission_Auto(AV)',
           'Stop/Start System (Engine Management System) Code_Y',
           'Stop/Start System (Engine Management System) Code_N',
           'Drive Sys_R',
           '# Cyl_12',
           'Exhaust Valves Per Cyl_2',
           'Trans_CVT',
           '4Dr Lugg Vol',
           'Htchbk Pass Vol',
          'Transmission_Auto(AM6)',
          'Htchbk Lugg Vol']
```

```
In [187]: X_tr_col = pd.DataFrame(X_tr_complete, columns = X_tr_com_cols )
```

```
In [188]: X_tr_t20 = X_tr_col[top_20]
```

```
In [189]: X_te_col = pd.DataFrame(X_te_complete, columns = X_tr_com_cols )
```

```
In [165]: X_te_t20 = X_te_col[top_20]
```

```
In [190]: rf_grid2 = GridSearchCV(RandomForestRegressor(), param_grid = {'n_estimators':[90,100,
          rf_grid2.fit(X_tr_t20,y_tr)
          print(rf_grid2.best_params_)
          print(rf_grid2.best_score_)
```

```
{'n_estimators': 105}
0.8802526210741554
```

```
In [191]: rf_grid2.score(X_te_t20,y_te)
```

```
Out[191]: 0.17618987314803525
```

The top 20 features seem to grossly overfit the data

We first concatenated the data for 15,16 and 17 into a train set. Eliminated the columns that directly report the target. Next we split the data into numerical and categorical columns. Then we impute each of these splits. Next we scale the numerical one and join the 2 into one single X_train dataframe. Likewise, we used the same process for the test set. However, we fit the scaler according to the train set. Also, we use the same columns as the training the set for the test set. Finally, we use 2 different linear models, followed by a polynomial transform and finally we try the randomforest regressor and gradient boost regression.

5 In conclusion, the best test accuracy we get is with Random Forest Regressor at 88.12%