energy_factor_analysis_by_region

December 12, 2021

[40]:
```
import pandas as pd
import os
import statsmodels.api as sm
from sklearn import linear_model
import numpy as np
```

0.1 This notebooks explores the relationship between a state's number of vehicle registrations, population, GDP per capita, GDP per capita by industry, C02 emissions, average yearly tempature, average yearly windspeed, minimum yearly tempature, maximim yearly tempature, total yearly precipitation , and total yearly snowfall on it's energy consumption within different regions of the US.

0.1.1 The goal is to model energy consuption for state's within different regions of the US by using the data listed above. With this model we can make energy consuption predictions and understand what leads to high energy consuption.

0.1.2 The contents of the notebook include

- **Data Gathering**
  - read in the dataframes that have been cleaned by data_gathering_and_cleaning notebook
- **Data analysis**
  - create a multiple linear regression model for energy consuption
- **Conclusion**
  - Discuss what we discovered and draw conclusions

Note: If there are no files in the Data/cleaned diretory, you will need to run the 'data_gathering_and_cleaning" notebook to clwan and write out the files to that directory.

0.1.3 Data Gathering

This section of the notebooks reads in the data files and stores them im pandas dataframes. The dataframes frames in this section all have columns of represting years ranging from [1967-2020] and rows for each state.

[41]:
```
csv_path = os.path.join(os.getcwd(), "data/cleaned/csv")
excel_path = os.path.join(os.getcwd(), "data/cleaned/excel")
```

[42]:
```
#Read in all datasets here

vehicle_registration_df = pd.read_csv(os.path.join(csv_path,␣
 ↪"vehicle_registrations_by_state.csv"))
energy_consumption_per_real_gdp_df = pd.read_csv(os.path.join(csv_path,␣
 ↪"energy_consumption_per_real_gdp.csv"))
current_dollar_gdp_df = pd.read_csv(os.path.join(csv_path, "Current_dollar_GDP.
 ↪csv")) #in millions
total_consuption_df = pd.read_csv(os.path.join(csv_path, "total_consuption.
 ↪csv")) #in billion Btu
industy_gdp_by_state_df = pd.read_csv(os.path.join(csv_path,␣
 ↪"industy_gdp_by_state.csv"))
total_population_df = pd.read_csv(os.path.join(csv_path, "total_population.
 ↪csv"))
real_gdp_df = pd.read_csv(os.path.join(csv_path, "real_GDP.csv")) #in millions
co2_emissions_df = pd.read_excel(os.path.join(excel_path, "co2_emissions.xlsx"))
tavg_df =  pd.read_csv(os.path.join(csv_path + '/NOA', "TAVG.csv"))
wind_df =  pd.read_csv(os.path.join(csv_path + '/NOA', "AWND.csv"))
tmax_df =  pd.read_csv(os.path.join(csv_path + '/NOA', "TMAX.csv"))
tmin_df =  pd.read_csv(os.path.join(csv_path + '/NOA', "TMIN.csv"))
precip_df = pd.read_csv(os.path.join(csv_path + '/NOA', "PRCP.csv"))
snow_df = pd.read_csv(os.path.join(csv_path + '/NOA', "SNOW.csv"))
```

[43]:
```
#Use the columns that are in each dataframe after columns with empty values␣
 ↪have been dropped.
columns_to_evaluate = list(set(vehicle_registration_df.columns).
 ↪intersection(total_population_df.columns).intersection(total_consuption_df.
 ↪columns).intersection(real_gdp_df.columns).
 ↪intersection(industy_gdp_by_state_df.columns).intersection(co2_emissions_df.
 ↪columns).intersection(tavg_df.columns).intersection(wind_df.columns).
 ↪intersection(tmax_df.columns).intersection(tmin_df.columns).
 ↪intersection(precip_df.columns).intersection(snow_df.columns))
columns_to_evaluate
```

[43]:
```
['2016',
 '2015',
 '2010',
 '2008',
 '2011',
 '2018',
```

```
'2007',
'2014',
'Unnamed: 0',
'2009',
'2012',
'2013',
'2017']
```

[44]:
```
#ensure each column we are going to evaluate has the same number of values
for col in columns_to_evaluate:
    if(not (len(vehicle_registration_df[col]) == len(total_consuption_df[col])␣
 ↪== len(total_population_df[col]) == len(real_gdp_df[col])==␣
 ↪len(industy_gdp_by_state_df[col]) == len(co2_emissions_df[col]) ==␣
 ↪len(tavg_df[col])== len(wind_df[col])== len(tmax_df[col])==␣
 ↪len(tmin_df[col])== len(precip_df[col])== len(snow_df[col]))):
        print("unequal entries for column:" + col)
```

[45]:
```
west = ["California", "Hawaii", "Nevada", "Colorado", "Idaho", "Montana",␣
 ↪"Utah", "Wyoming", "Oregon", "Washington", "Alaska"]
south_west = ["New Mexico", "Arizona", "Texas", "Oklahoma"]
mid_west = ["Iowa", "Kansas", "Missouri", "Nebraska", "North Dakota", "South␣
 ↪Dakota", "Illinois", "Indiana", "Michigan", "Minnesota", "Ohio", "Wisconsin"]
south_east = ["Alabama", "Florida", "Georgia", "Mississippi", "South Carolina",␣
 ↪"Arkansas", "Louisiana", "Delaware", "Kentucky", "Maryland", "North␣
 ↪Carolina", "Tennessee", "Virginia", "West Virginia"]
north_east = ["New Jersey", "New York", "Pennsylvania", "Connecticut", "Maine",␣
 ↪"Massachusetts", "New Hampshire", "Rhode Island", "Vermont"]
```

[46]:
```
west_abr = ["CA", "HI", "NV", "CO", "ID", "MT", "UT", "WY", "OR", "WA", "AK"]
southwest_abr = ["NM", "AZ", "TX", "OK"]
midwest_abr = ["IA", "KS", "MO", "NE", "ND", "SD", "IL", "IN", "MI", "MN",␣
 ↪"OH", "WI"]
southeast_abr = ["AL", "FL", "GA", "MS", "SC", "AR", "LA", "DE", "KY", "MD",␣
 ↪"NC", "TN", "VA", "WV"]
northeast_abr = ["NJ", "NY", "PA", "CT", "ME", "MA", "NH", "RI", "VT"]
```

0.1.4 Data Analysis

This section of the notebooks creates a multiple linear regression model for a state's energy consuption.

In the model summary each variable is represented by the following

- x1: Vehicle regisrations
- x2: Population
- x3: GDP per capita
- x4: Industry GDP per capita
- x5: C02 emissions

- x6: Average tempature
- x7: Average wind speed
- x8: Maximum tempature
- x9: Minimum tempature
- x10: Total precipitation
- x11: Total snow fall

There are some other values in the summary that give us a good indication as to how well our model fits energy consuption such at the r squared value and F statistic.

[47]:
```
# loop through the data frames and add each value to data_point_pairs array.
# The data_point_pairs array will be the
# [vehicle registration, population, GDP, Industry GDP, CO2 emissions, average␣
↪tempature, average wind speed, max temperature, min tempature, total␣
↪precipitation, total snowfall]
# value for each year and each state
# The total_consumption_vals will be the cooresponding energy consuption value
# for the data point pairs item
west_data_point_pairs = []
west_total_consumption_vals = []

southwest_data_point_pairs = []
southwest_total_consumption_vals = []

midwest_data_point_pairs = []
midwest_total_consumption_vals = []

southeast_data_point_pairs = []
southeast_total_consumption_vals = []

northeast_data_point_pairs = []
northeast_total_consumption_vals = []
for col in columns_to_evaluate:
    for i in range(0,50):
        pair = [vehicle_registration_df.iloc[i][col], total_population_df.
↪iloc[i][col], real_gdp_df.iloc[i][col], industy_gdp_by_state_df.
↪iloc[i][col], co2_emissions_df.iloc[i][col], tavg_df.iloc[i][col],wind_df.
↪iloc[i][col],tmax_df.iloc[i][col],tmin_df.iloc[i][col],precip_df.
↪iloc[i][col],snow_df.iloc[i][col]]
        if(total_consuption_df.iloc[i]['State'] in west_abr) :
            west_data_point_pairs.append(pair)
            west_total_consumption_vals.append(total_consuption_df.iloc[i][col])

        if(total_consuption_df.iloc[i]['State'] in southwest_abr) :
            southwest_data_point_pairs.append(pair)
            southwest_total_consumption_vals.append(total_consuption_df.
↪iloc[i][col])
```

```
        if(total_consuption_df.iloc[i]['State'] in midwest_abr) :
            midwest_data_point_pairs.append(pair)
            midwest_total_consumption_vals.append(total_consuption_df.
↪iloc[i][col])

        if(total_consuption_df.iloc[i]['State'] in southeast_abr) :
            southeast_data_point_pairs.append(pair)
            southeast_total_consumption_vals.append(total_consuption_df.
↪iloc[i][col])

        if(total_consuption_df.iloc[i]['State'] in northeast_abr) :
            northeast_data_point_pairs.append(pair)
            northeast_total_consumption_vals.append(total_consuption_df.
↪iloc[i][col])
```

0.1.5 Multiple linear regression model for energy consuption of states in the western region of US

California, Hawaii, Nevada, Colorado, Idaho, Montana, Utah, Wyoming, Oregon, Washington, Alaska

[48]:
```
X = west_data_point_pairs
y = west_total_consumption_vals
lm = linear_model.LinearRegression()
model = lm.fit(X,y)


# predict energy consuption for vehicle registration = 4610845 , population␣
↪=699 (10,000), GDP = 55911,
# Industry GDP = 9717, CO2 emissions = 121, Average tempature = 6.7, Average␣
↪Wind Speed = 2.5
# Maximim tempature = 14.07, Mimimum tempature = -0.44, Total Precipitation =␣
↪47, Total snowfall: 190
predictions = lm.predict([[4610845, 699, 55911, 9717, 121,  6.7, 2.5, 14.07, -0.
↪44, 47, 190]])
print("Predicted energy consumpion: ", predictions )

model = sm.OLS(y, X).fit()
model.summary()
```

```
Predicted energy consumpion:  [643108.49639605]
```

[48]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                      y   R-squared (uncentered):
```

```
0.996
Model:                            OLS   Adj. R-squared (uncentered):
0.996
Method:                 Least Squares   F-statistic:
3277.
Date:                Mon, 06 Dec 2021   Prob (F-statistic):
4.38e-155
Time:                        17:36:14   Log-Likelihood:
-1903.3
No. Observations:                 143   AIC:
3829.
Df Residuals:                     132   BIC:
3861.
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.0221      0.008      2.608      0.010       0.005       0.039
x2           177.5311     20.097      8.834      0.000     137.778     217.285
x3             0.5744      0.389      1.476      0.142      -0.195       1.344
x4           -11.7354      2.895     -4.054      0.000     -17.462      -6.009
x5          1392.3846    663.136      2.100      0.038      80.635    2704.134
x6          1.571e+05   3.05e+05      0.515      0.607   -4.46e+05    7.61e+05
x7          2.164e+04   6353.207      3.405      0.001    9067.925    3.42e+04
x8         -5.722e+04   1.52e+05     -0.375      0.708   -3.59e+05    2.44e+05
x9         -1.237e+05   1.53e+05     -0.809      0.420   -4.26e+05    1.79e+05
x10          551.0159    584.241      0.943      0.347    -604.671    1706.703
x11         -261.4483    194.667     -1.343      0.182    -646.520     123.623
==============================================================================
Omnibus:                       19.394   Durbin-Watson:                   2.547
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               22.659
Skew:                           0.930   Prob(JB):                     1.20e-05
Kurtosis:                       3.587   Cond. No.                     2.97e+08
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.97e+08. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

0.1.6 Multiple linear regression model for energy consuption of states in the south western region of US

New Mexico, Arizona, Texas, Oklahoma

[36]:
```python
X = southwest_data_point_pairs
y = southwest_total_consumption_vals
lm = linear_model.LinearRegression()
model = lm.fit(X,y)

# predict energy consuption for vehicle registration = 4610845 , population␣
↪=699 (10,000), GDP = 55911,
# Industry GDP = 9717, CO2 emissions = 121, Average tempature = 6.7, Average␣
↪Wind Speed = 2.5
# Maximim tempature = 14.07, Mimimum tempature = -0.44, Total Precipitation =␣
↪47, Total snowfall: 190
predictions = lm.predict([[4610845, 699, 55911, 9717, 121,  6.7, 2.5, 14.07, -0.
↪44, 47, 190]])
print("Predicted energy consumpion: ", predictions )

model = sm.OLS(y, X).fit()
model.summary()
```

```
Predicted energy consumpion:  [559550.25847271]
```

[36]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                      y   R-squared (uncentered):
1.000
Model:                            OLS   Adj. R-squared (uncentered):
1.000
Method:                 Least Squares   F-statistic:
1.013e+04
Date:                Mon, 06 Dec 2021   Prob (F-statistic):
1.04e-66
Time:                        17:09:49   Log-Likelihood:
-679.90
No. Observations:                  52   AIC:
1382.
Df Residuals:                      41   BIC:
1403.
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.0294      0.011     -2.688      0.010      -0.052      -0.007
x2            -3.2953     39.591     -0.083      0.934     -83.252      76.661
x3             5.5634      0.621      8.960      0.000       4.309       6.817
```

```
x4          -10.1936      1.599     -6.376      0.000     -13.422      -6.965
x5         8659.9045    784.661     11.036      0.000    7075.250    1.02e+04
x6         1.012e+06   6.53e+05      1.549      0.129   -3.08e+05    2.33e+06
x7        -8.015e+04   1.77e+04     -4.541      0.000   -1.16e+05   -4.45e+04
x8        -5.327e+05   3.26e+05     -1.634      0.110   -1.19e+06    1.26e+05
x9        -4.096e+05   3.32e+05     -1.234      0.224   -1.08e+06     2.6e+05
x10        -256.9868   1028.137     -0.250      0.804   -2333.352    1819.378
x11         959.7019    590.575      1.625      0.112    -232.989    2152.392
==============================================================================
Omnibus:                        1.503   Durbin-Watson:                   2.041
Prob(Omnibus):                  0.472   Jarque-Bera (JB):                0.991
Skew:                           0.333   Prob(JB):                        0.609
Kurtosis:                       3.112   Cond. No.                     4.97e+08
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 4.97e+08. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

0.1.7 Multiple linear regression model for energy consuption of states in the mid western region of US

Iowa, Kansas, Missouri, Nebraska, North Dakota, South Dakota, Illinois, Indiana, Michigan, Minnesota, Ohio, Wisconsin

[37]:
```
X = midwest_data_point_pairs
y = midwest_total_consumption_vals
lm = linear_model.LinearRegression()
model = lm.fit(X,y)

#predict energy consuption for vehicle registration = 4610845 , population =699␣
↪(10,000), GDP = 55911, Industry GDP = 9717, CO2 emissions = 121
# predict energy consuption for vehicle registration = 4610845 , population␣
↪=699 (10,000), GDP = 55911,
# Industry GDP = 9717, CO2 emissions = 121, Average tempature = 6.7, Average␣
↪Wind Speed = 2.5
# Maximim tempature = 14.07, Mimimum tempature = -0.44, Total Precipitation =␣
↪47, Total snowfall: 190
predictions = lm.predict([[4610845, 699, 55911, 9717, 121,  6.7, 2.5, 14.07, -0.
↪44, 47, 190]])
print("Predicted energy consumpion: ", predictions )

model = sm.OLS(y, X).fit()
model.summary()
```

```
Predicted energy consumpion:  [178487.45565612]
```

[37]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                      y   R-squared (uncentered):
0.991
Model:                            OLS   Adj. R-squared (uncentered):
0.991
Method:                 Least Squares   F-statistic:
1537.
Date:                Mon, 06 Dec 2021   Prob (F-statistic):
3.93e-144
Time:                        17:09:49   Log-Likelihood:
-2125.3
No. Observations:                 156   AIC:
4273.
Df Residuals:                     145   BIC:
4306.
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.0006      0.016      0.036      0.971      -0.031       0.033
x2           320.2389     30.466     10.512      0.000     260.025     380.453
x3             0.2932      0.483      0.607      0.545      -0.662       1.248
x4           -16.0495      5.748     -2.792      0.006     -27.410      -4.689
x5           -58.4310   1122.196     -0.052      0.959   -2276.406    2159.544
x6          5.673e+05   4.51e+05      1.258      0.211   -3.24e+05    1.46e+06
x7         -2.979e+04   1.17e+04     -2.543      0.012   -5.29e+04   -6638.566
x8         -2.928e+05   2.26e+05     -1.297      0.197   -7.39e+05    1.53e+05
x9         -2.481e+05   2.26e+05     -1.099      0.274   -6.94e+05    1.98e+05
x10         1532.0431    711.569      2.153      0.033     125.656    2938.430
x11         1738.9202    339.090      5.128      0.000    1068.723    2409.117
==============================================================================
Omnibus:                       33.250   Durbin-Watson:                   1.907
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               49.710
Skew:                           1.130   Prob(JB):                     1.61e-11
Kurtosis:                       4.593   Cond. No.                     1.54e+08
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

```
[2] The condition number is large, 1.54e+08. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

0.1.8 Multiple linear regression model for energy consuption of states in the south eastern region of US

Alabama, Florida, Georgia, Mississippi, South Carolina, Arkansas, Louisiana, Delaware, Kentucky, Maryland, North Carolina, Tennessee, Virginia, West Virginia

[38]:
```
X = southeast_data_point_pairs
y = southeast_total_consumption_vals
lm = linear_model.LinearRegression()
model = lm.fit(X,y)

# predict energy consuption for vehicle registration = 4610845 , population␣
↪=699 (10,000), GDP = 55911,
# Industry GDP = 9717, CO2 emissions = 121, Average tempature = 6.7, Average␣
↪Wind Speed = 2.5
# Maximim tempature = 14.07, Mimimum tempature = -0.44, Total Precipitation =␣
↪47, Total snowfall: 190
predictions = lm.predict([[4610845, 699, 55911, 9717, 121,  6.7, 2.5, 14.07, -0.
↪44, 47, 190]])
print("Predicted energy consumpion: ", predictions )

model = sm.OLS(y, X).fit()
model.summary()
```

```
Predicted energy consumpion:  [1732923.0468554]
```

[38]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                      y   R-squared (uncentered):
0.983
Model:                            OLS   Adj. R-squared (uncentered):
0.982
Method:                 Least Squares   F-statistic:
924.4
Date:                Mon, 06 Dec 2021   Prob (F-statistic):
4.84e-146
Time:                        17:09:49   Log-Likelihood:
-2546.7
No. Observations:                 182   AIC:
5115.
Df Residuals:                     171   BIC:
```

```
5151.
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.0502      0.014     -3.497      0.001      -0.079      -0.022
x2           -36.4637     34.030     -1.072      0.285    -103.636      30.708
x3             5.6552      0.707      7.994      0.000       4.259       7.052
x4           -61.8024      5.321    -11.614      0.000     -72.306     -51.298
x5          1.591e+04    649.516     24.492      0.000    1.46e+04    1.72e+04
x6         -5.726e+05   4.04e+05     -1.416      0.159   -1.37e+06    2.26e+05
x7          1.511e+04   7778.081      1.943      0.054    -243.133    3.05e+04
x8          3.136e+05   2.03e+05      1.542      0.125   -8.78e+04    7.15e+05
x9          2.234e+05      2e+05      1.115      0.266   -1.72e+05    6.19e+05
x10         3010.8222   1099.419      2.739      0.007     840.642    5181.002
x11        -1647.5511    341.342     -4.827      0.000   -2321.338    -973.764
==============================================================================
Omnibus:                        1.088   Durbin-Watson:                   2.096
Prob(Omnibus):                  0.580   Jarque-Bera (JB):                1.190
Skew:                           0.145   Prob(JB):                        0.552
Kurtosis:                       2.730   Cond. No.                     1.28e+08
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.28e+08. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

0.1.9 Multiple linear regression model for energy consuption of states in the north eastern region of US

New Jersey, New York, Pennsylvania, Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont

[39]:
```
X = northeast_data_point_pairs
y = northeast_total_consumption_vals
lm = linear_model.LinearRegression()
model = lm.fit(X,y)

# predict energy consuption for vehicle registration = 4610845 , population
 ↪=699 (10,000), GDP = 55911,
# Industry GDP = 9717, C02 emissions = 121, Average tempature = 6.7, Average
 ↪Wind Speed = 2.5
```

```
# Maximim tempature = 14.07, Mimimum tempature = -0.44, Total Precipitation =␣
↪47, Total snowfall: 190
predictions = lm.predict([[4610845, 699, 55911, 9717, 121,  6.7, 2.5, 14.07, -0.
↪44, 47, 190]])
print("Predicted energy consumpion: ", predictions )

model = sm.OLS(y, X).fit()
model.summary()
```

```
Predicted energy consumpion:  [482508.1651461]
```

[39]:

```
<class 'statsmodels.iolib.summary.Summary'>
"""
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                      y   R-squared (uncentered):
0.993
Model:                            OLS   Adj. R-squared (uncentered):
0.992
Method:                 Least Squares   F-statistic:
1277.
Date:                Mon, 06 Dec 2021   Prob (F-statistic):
3.01e-107
Time:                        17:09:49   Log-Likelihood:
-1574.9
No. Observations:                 117   AIC:
3172.
Df Residuals:                     106   BIC:
3202.
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.0050      0.015     -0.330      0.742      -0.035       0.025
x2           384.6247     37.273     10.319      0.000     310.728     458.521
x3            -2.5845      0.551     -4.693      0.000      -3.676      -1.493
x4            -7.3340      3.530     -2.078      0.040     -14.332      -0.336
x5          2048.7594   1083.579      1.891      0.061     -99.541    4197.060
x6          2.801e+05   3.97e+05      0.706      0.482   -5.07e+05    1.07e+06
x7          9053.3934   6674.243      1.356      0.178   -4178.942    2.23e+04
x8         -1.179e+05   1.98e+05     -0.596      0.553    -5.1e+05    2.74e+05
x9         -1.718e+05      2e+05     -0.861      0.391   -5.68e+05    2.24e+05
x10        -2021.9385    727.881     -2.778      0.006   -3465.034    -578.843
x11         -242.9752    271.314     -0.896      0.373    -780.881     294.931
==============================================================================
```

```
Omnibus:                       17.403   Durbin-Watson:                   1.860
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               40.280
Skew:                           0.529   Prob(JB):                     1.79e-09
Kurtosis:                       5.673   Cond. No.                     1.46e+08
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.46e+08. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

0.1.10 Conclusion

This section of the notebooks discusses the results The average yearly tempature has a large effect on energy consumption for every region. All regions except the south eastern region, had a large, positive relationship between the two. The effects ranged from 157,100 - 1,012,000 increase in million british thermal units (BTU) per increase of degree celcius. It makes sense that higher tempatures would require more energy for cooling. However, for the South Eastern region, there was a strong negative relationship between energy consuption and average yearly tempature with an expected drop of 572,600 million BTU per increase of degree celcius. It is likely we are missing something to accountfor this change, but a possible reason could be that the south east is used to high tempatures, so low tempatures end up consuming more energy than high tempatures.

For every region, the yearly maximum and minimum tempature either both has a large negative effect or large postive effect on predicted energy consumption. Similar to average yearly tempature, all the regions except the south eastern region saw a large positive effect of minimum and maximim yearly tempature on energy consuption. The south easten region saw a large negative effect of minimum and maximim yearly tempature on energy consuption.

All the regions saw a large effect of wind speed on energy consuption, but whether that effect was positive or negative was mixed. The Western, South Eastern, and North Eastern regions all saw a positive effect on energy ranging from an increase of 9,053 - 21,640 million BTU per increase in mile per hour of average wind speed. The Mid Western and South Western regions saw a negative effect on energy consuption with a decrease of 80,150 and 29,790 million BTU per increase in mile per hour of average wind speed, respectively. What may account for this inconsistency is that Texas, Iowa, Oklahoma, Kansas, and Illinois make up the top 5 states with the most electricity generation from wind i.e. the most energy consuption from wind. All these states reside in the Mid Western and South Western regions. The more energy consumption from wind may lead to a significant decrease in energy consuption from other sources decreasing the overall energy consuption.

All regions except the Mid Western region saw a moderatly positive effect of CO2 emissions on energy consuption with an increase of BTU ranging between 1,392 - 15,910 million per every increase in 1 million metric ton of CO2 emissions. This makes sense since the more energy that is consumed, the more CO2 will be released. The Mid Western region saw a small negative effect of -58 million BTU per increase in 1 million metric ton of C02 emissions. We are likley missing something to explain this inconsistency, but the Mid Western region's large consumption of wind energy which doesn't produce CO2 may explain some of this result.

The total snowfall and precipitation had a moderate effect on energy consumption that varied on whether it was a positive or negative based on the region.

GDP per capita and industry GDP per capital have a small effect on energy consumption for every region. Also for every region, the industry GDP per capital has a small negative effect on energy consumption. For every increase in GDP made from industry, energy consumption decreased in BTU ranging from 61-7 million. This is surprising since GDP made from industry would require more energy consumption to run those industries. However, the effect that we saw was small, and we may be missing something to explain this.

Population has a small positive effect on energy consumption for every region except the south eastern and south western regions. This does not seem to make sense since more population should mean more energy consumption. We are most likely missing soemthing to explain this.

Lastly, for all regions, the number of vehicles registered had little to no effect on energy consumption.

[ ]:
```

```