

time_series_correlation

September 12, 2021

```
[1]: #importing required libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from pandas.plotting import register_matplotlib_converters
register_matplotlib_converters()
```

```
[2]: #the dataset was renamed to cement_data
#Date/Time was renamed to date_time to follow python rules
#the dataset was saved as CSV and imported using pandas

data = pd.read_csv('cement_data.csv')
data
```

```
[2]:
```

	date_time	Cement	Type	MDP	MDC	FFS	\
0	7/16/2021 10:26		42.5	3298.291862	178.211069	864.566667	
1	7/16/2021 10:27		42.5	3291.276123	179.438870	864.433333	
2	7/16/2021 10:28		42.5	3274.168245	181.262520	864.450000	
3	7/16/2021 10:29		42.5	3228.772392	177.037553	864.466667	
4	7/16/2021 10:30		42.5	3351.213582	185.036202	864.583333	
...
25326	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25327	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25328	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25329	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25330	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	SDS	SDC	BED	MD	MIP1	...	\
0	851.001191	187.445892	58.763833	51.893810	-4.708457	...	
1	850.986004	186.991837	59.547584	51.856921	-4.653129	...	
2	850.984485	188.002238	57.997499	52.635440	-3.945597	...	
3	850.990560	187.675466	60.853833	52.424708	-4.158661	...	
4	850.952591	187.599559	57.858166	52.527366	-4.365102	...	
...

25326	NaN	NaN	NaN	NaN	NaN	...
25327	NaN	NaN	NaN	NaN	NaN	...
25328	NaN	NaN	NaN	NaN	NaN	...
25329	NaN	NaN	NaN	NaN	NaN	...
25330	NaN	NaN	NaN	NaN	NaN	...

	Feed	PSS	RP3	RP2	RP1	CP \
0	279.060143	61.705166	72.422198	71.253243	53.315950	21.021806
1	306.984365	61.704105	72.791336	72.006428	54.056215	20.981683
2	283.658207	61.693496	71.848385	71.934312	52.972109	21.040324
3	303.102896	61.700392	73.003317	71.926304	53.619336	21.008689
4	303.862520	61.704105	74.351956	73.144215	56.527814	20.869032
...
25326	NaN	NaN	NaN	NaN	NaN	NaN
25327	NaN	NaN	NaN	NaN	NaN	NaN
25328	NaN	NaN	NaN	NaN	NaN	NaN
25329	NaN	NaN	NaN	NaN	NaN	NaN
25330	NaN	NaN	NaN	NaN	NaN	NaN

	WP	FCT	Blaine	Residue
0	68.209569	233.592009	NaN	NaN
1	68.342281	233.516286	NaN	NaN
2	68.326076	233.665709	NaN	NaN
3	68.382401	233.670993	NaN	NaN
4	68.534401	233.791611	NaN	NaN
...
25326	NaN	NaN	NaN	NaN
25327	NaN	NaN	NaN	NaN
25328	NaN	NaN	NaN	NaN
25329	NaN	NaN	NaN	NaN
25330	NaN	NaN	NaN	NaN

[25331 rows x 25 columns]

```
[3]: #date_time column converted to pandas time series
data['date_time'] = pd.to_datetime(data['date_time'])

#date_time column declared as index
data.set_index('date_time', inplace=True, drop=False)

#dataset grouped by 30minutes and average of each 30 minutes taken
by30mins = data.resample('30min', on='date_time').mean()
by30mins
```

[3]:	Cement Type	MDP	MDC	FFS \
date_time				
2021-07-16 10:00:00	42.5	3273.127155	178.987503	864.479167

2021-07-16 10:30:00	42.5	3507.799957	193.703640	864.502222
2021-07-16 11:00:00	42.5	3318.101181	183.367252	864.495556
2021-07-16 11:30:00	42.5	3515.477154	193.253288	864.481818
2021-07-16 12:00:00	NaN	NaN	NaN	NaN
...
2021-08-06 08:00:00	32.5	2982.467770	164.671961	867.142949
2021-08-06 08:30:00	32.5	3445.416474	188.501743	879.490000
2021-08-06 09:00:00	32.5	3238.479048	177.946894	881.966111
2021-08-06 09:30:00	32.5	3231.000119	178.288727	884.483889
2021-08-06 10:00:00	32.5	3120.038407	171.787697	884.500000

	SDS	SDC	BED	MD	MIP1 \
date_time					
2021-07-16 10:00:00	850.990560	187.528858	59.290687	52.202720	-4.366461
2021-07-16 10:30:00	850.969955	187.898081	60.154844	53.313380	-4.194578
2021-07-16 11:00:00	850.976233	187.464129	59.594608	52.648417	-4.476093
2021-07-16 11:30:00	850.972887	188.228193	59.984583	53.070231	-4.228303
2021-07-16 12:00:00	NaN	NaN	NaN	NaN	NaN
...
2021-08-06 08:00:00	830.085380	183.827438	61.529734	54.608658	-4.289459
2021-08-06 08:30:00	821.083225	183.857108	65.303791	57.838527	-4.305306
2021-08-06 09:00:00	821.085807	183.622474	65.400744	59.078356	-4.068506
2021-08-06 09:30:00	821.086567	183.429574	63.484330	58.571520	-4.239002
2021-08-06 10:00:00	821.085927	183.637737	63.657916	58.597963	-4.274498

	FOP ...	Feed	PSS	RP3 \
date_time	...			
2021-07-16 10:00:00	-67.871152 ...	293.201403	61.700789	72.516309
2021-07-16 10:30:00	-68.346510 ...	304.567006	61.700975	72.826423
2021-07-16 11:00:00	-68.268760 ...	300.371180	61.701647	76.051705
2021-07-16 11:30:00	-68.348599 ...	302.167349	61.702128	74.807484
2021-07-16 12:00:00	NaN ...	NaN	NaN	NaN
...
2021-08-06 08:00:00	-68.038632 ...	289.619629	60.192140	92.945202
2021-08-06 08:30:00	-71.602344 ...	320.209908	59.538397	100.954869
2021-08-06 09:00:00	-72.303393 ...	326.533309	59.536452	105.234724
2021-08-06 09:30:00	-72.253094 ...	311.045140	59.536364	103.799383
2021-08-06 10:00:00	-72.359485 ...	311.940744	59.538883	103.517958

	RP2	RP1	CP	WP	FCT \
date_time					
2021-07-16 10:00:00	71.780072	53.490903	21.013126	68.315082	233.611249
2021-07-16 10:30:00	71.608468	53.677047	21.012135	68.481934	233.631143
2021-07-16 11:00:00	74.622544	58.365673	20.769549	68.477434	233.625162
2021-07-16 11:30:00	73.460715	56.185080	20.891478	68.153596	233.634699
2021-07-16 12:00:00	NaN	NaN	NaN	NaN	NaN
...

2021-08-06 08:00:00	90.650592	75.434913	24.673327	70.179987	230.198320
2021-08-06 08:30:00	95.297246	86.824267	21.293043	69.969552	233.401861
2021-08-06 09:00:00	98.045869	90.542572	21.028853	72.003085	232.110957
2021-08-06 09:30:00	96.095112	91.324450	20.921397	76.631251	223.433903
2021-08-06 10:00:00	96.284976	92.403691	20.871781	77.936011	223.390304

	Blaine	Residue
date_time		
2021-07-16 10:00:00	NaN	NaN
2021-07-16 10:30:00	NaN	NaN
2021-07-16 11:00:00	NaN	NaN
2021-07-16 11:30:00	NaN	NaN
2021-07-16 12:00:00	NaN	NaN
...
2021-08-06 08:00:00	3914.0	14.6
2021-08-06 08:30:00	3934.0	16.6
2021-08-06 09:00:00	3701.0	17.6
2021-08-06 09:30:00	3545.0	18.0
2021-08-06 10:00:00	3809.0	16.7

[1009 rows x 24 columns]

```
[4]: #rows with incomplete data are removed,
by30mins_data = by30mins.dropna()

#export dataset to create new clean dataset
by30mins_data.to_csv('clean_cement_data.csv')

#importing the new dataset
df = pd.read_csv('clean_cement_data.csv')
df
```

```
[4]:
```

	date_time	Cement	Type	MDP	MDC	FFS	\
0	2021-07-17 01:00:00		32.5	3379.448385	185.134306	884.500556	
1	2021-07-17 01:30:00		32.5	3505.942832	192.545055	884.496667	
2	2021-07-17 02:00:00		32.5	3415.357422	187.220392	884.497778	
3	2021-07-17 03:30:00		32.5	3560.460656	195.390083	884.510556	
4	2021-07-17 05:30:00		32.5	3314.776902	181.307640	884.497222	
..	
630	2021-08-06 08:00:00		32.5	2982.467770	164.671961	867.142949	
631	2021-08-06 08:30:00		32.5	3445.416474	188.501743	879.490000	
632	2021-08-06 09:00:00		32.5	3238.479048	177.946894	881.966111	
633	2021-08-06 09:30:00		32.5	3231.000119	178.288727	884.483889	
634	2021-08-06 10:00:00		32.5	3120.038407	171.787697	884.500000	

	SDS	SDC	BED	MD	MIP1	...	Feed	\
0	865.920661	191.034568	63.684622	56.959591	-3.810432	...	323.697747	

1	865.917421	189.949864	65.836741	57.340028	-3.782796	...	333.696425
2	865.920155	190.257382	65.317725	56.789716	-3.909147	...	332.584217
3	873.127776	191.992946	68.341839	58.596666	-3.537462	...	359.106143
4	865.914991	191.032408	62.048616	56.670620	-3.791760	...	313.935933
..
630	830.085380	183.827438	61.529734	54.608658	-4.289459	...	289.619629
631	821.083225	183.857108	65.303791	57.838527	-4.305306	...	320.209908
632	821.085807	183.622474	65.400744	59.078356	-4.068506	...	326.533309
633	821.086567	183.429574	63.484330	58.571520	-4.239002	...	311.045140
634	821.085927	183.637737	63.657916	58.597963	-4.274498	...	311.940744

	PSS	RP3	RP2	RP1	CP	WP \
0	62.776615	74.701360	73.877237	57.485579	20.709726	62.341182
1	62.777605	75.708271	75.330591	59.937359	20.544170	62.366644
2	62.777304	77.190922	77.538733	60.665759	20.839069	62.025887
3	63.304850	66.174477	63.275886	47.898112	21.622998	60.564019
4	62.778560	72.109829	71.365686	54.941299	20.988602	62.557380
..
630	60.192140	92.945202	90.650592	75.434913	24.673327	70.179987
631	59.538397	100.954869	95.297246	86.824267	21.293043	69.969552
632	59.536452	105.234724	98.045869	90.542572	21.028853	72.003085
633	59.536364	103.799383	96.095112	91.324450	20.921397	76.631251
634	59.538883	103.517958	96.284976	92.403691	20.871781	77.936011

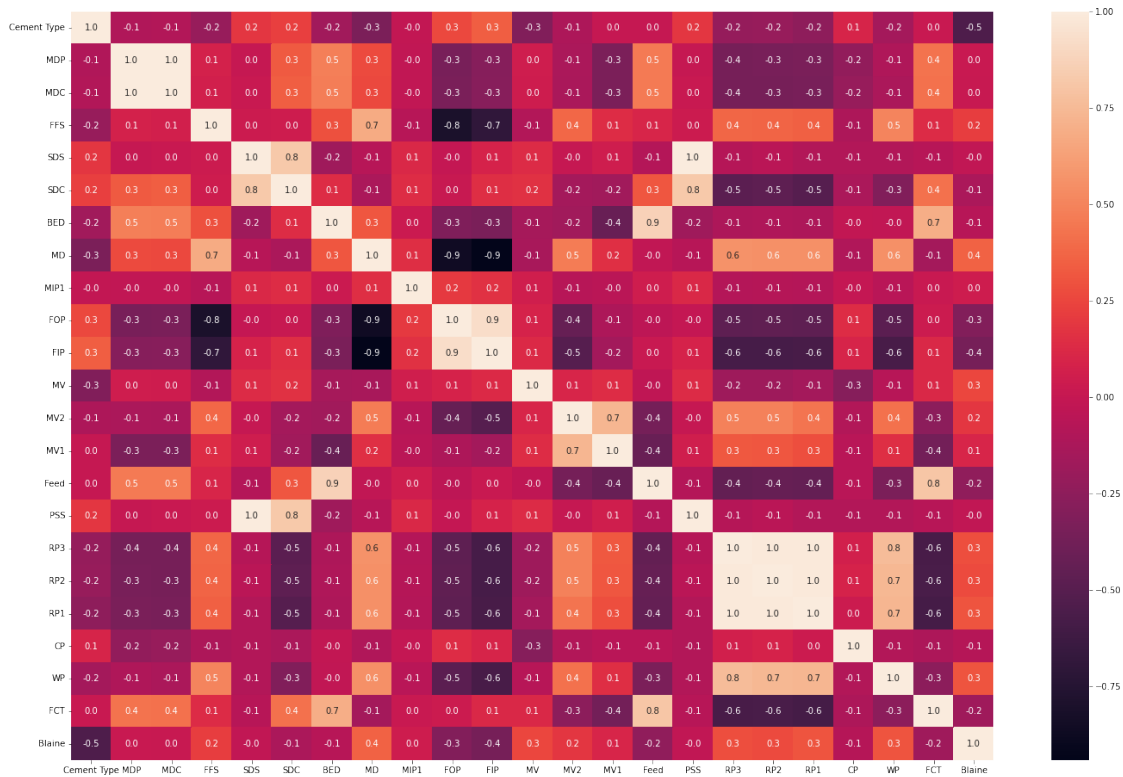
	FCT	Blaine	Residue
0	252.237037	3810.0	15.3
1	252.401972	3317.0	17.9
2	252.390025	3625.0	17.6
3	251.487817	3667.0	15.4
4	252.250273	3928.0	14.5
..
630	230.198320	3914.0	14.6
631	233.401861	3934.0	16.6
632	232.110957	3701.0	17.6
633	223.433903	3545.0	18.0
634	223.390304	3809.0	16.7

[635 rows x 25 columns]

Correlation for blaine output

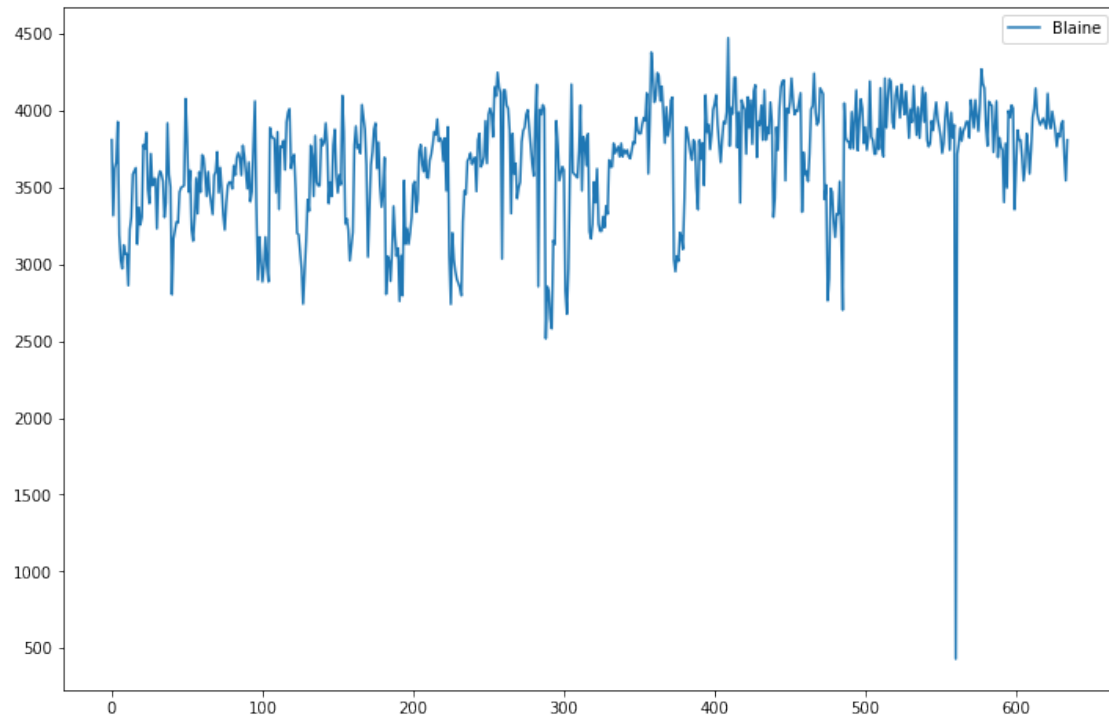
```
[5]: #plotting correlation matrix for blaine output
blaine_c = df.drop('Residue', axis=1)
plt.figure(figsize=(24,16))

sns.heatmap(blaine_c.corr(),annot=True,fmt='.1f')
plt.show()
```



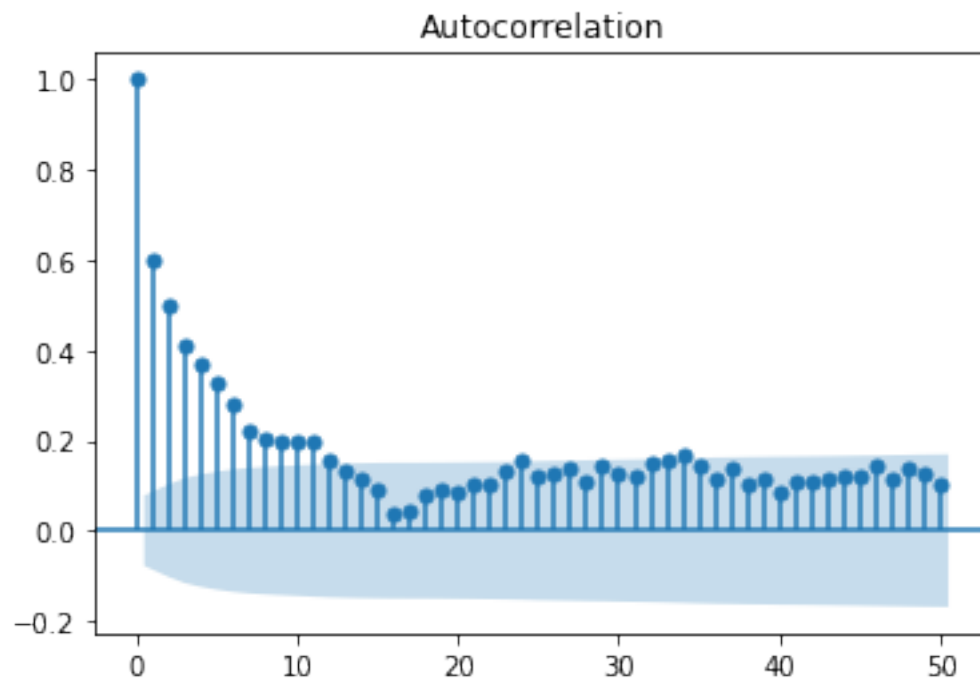
```
[6]: #plotting time series by 30 mins against blaine output
f = (12, 8)
blaine = df[['Blaine']]
blaine.plot(figsize = f)
```

```
[6]: <AxesSubplot:>
```



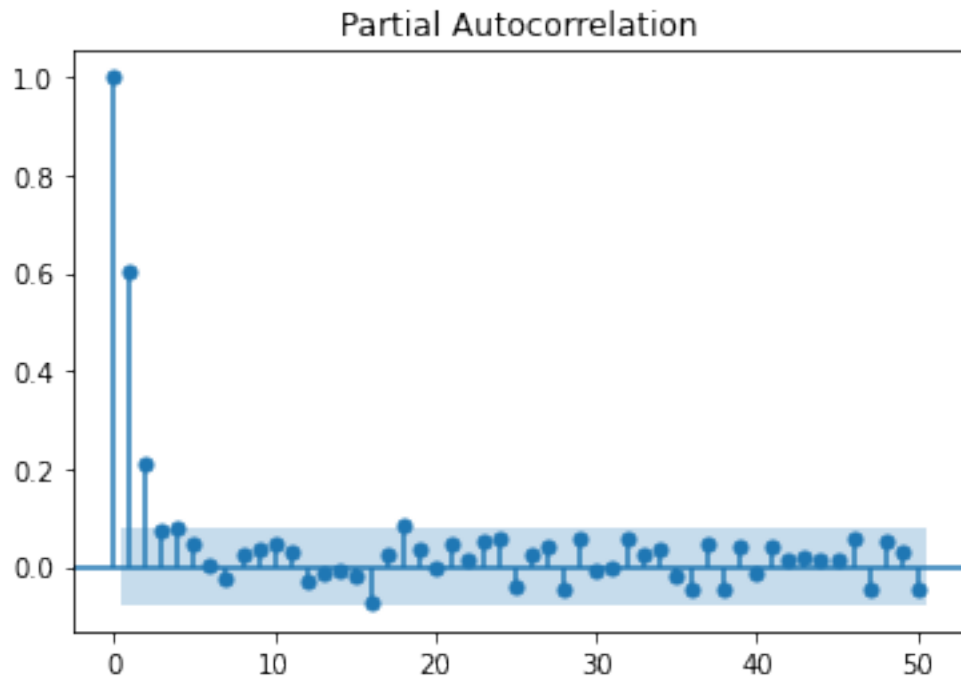
```
[7]: #Plotting time series correlation by 50 lags for blaine ouput  
plt.figure(figsize=(24,16))  
blaine_time_corr = plot_acf(blaine, lags=50)  
plt.show()
```

<Figure size 1728x1152 with 0 Axes>



```
[8]: #Plotting time series partial correlation by 50 lags for blaine ouput
plt.figure(figsize=(24,16))
plot_pacf(blaine, lags=50)
plt.show()
```

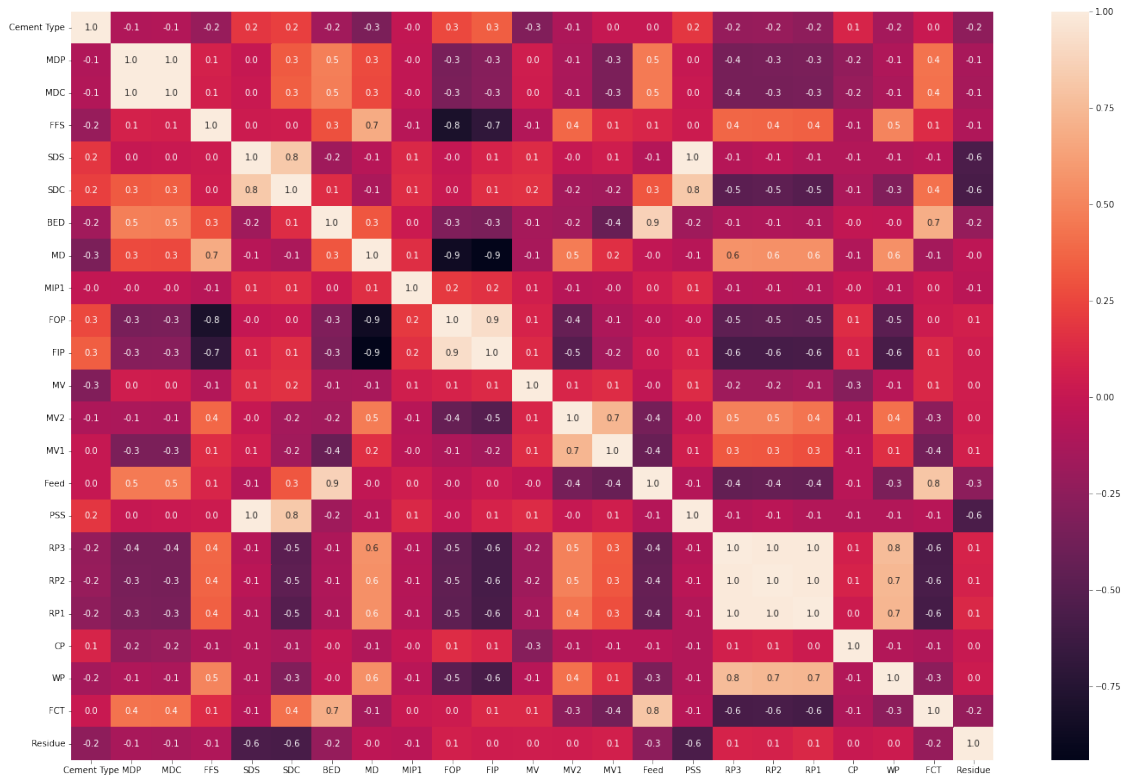
<Figure size 1728x1152 with 0 Axes>



Correlation for Residue output

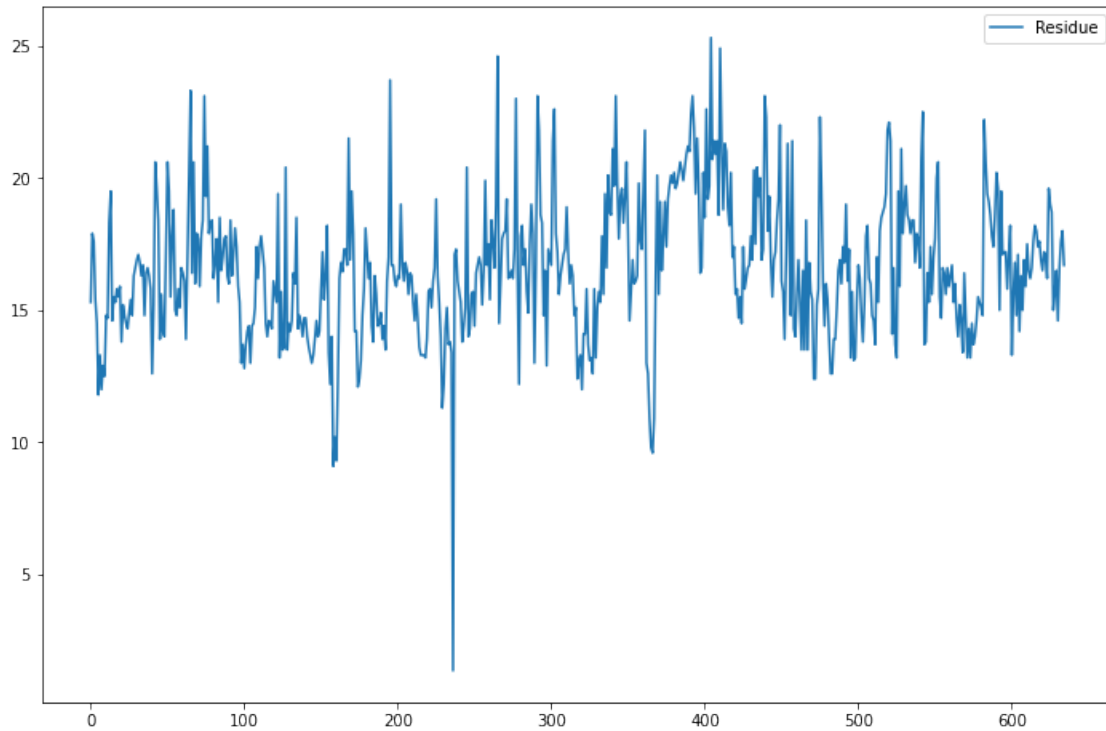
```
[9]: #plotting correlation matrix for residue output
blaine_c = df.drop('Blaine', axis=1)
plt.figure(figsize=(24,16))

sns.heatmap(blaine_c.corr(),annot=True,fmt='.1f')
plt.show()
```



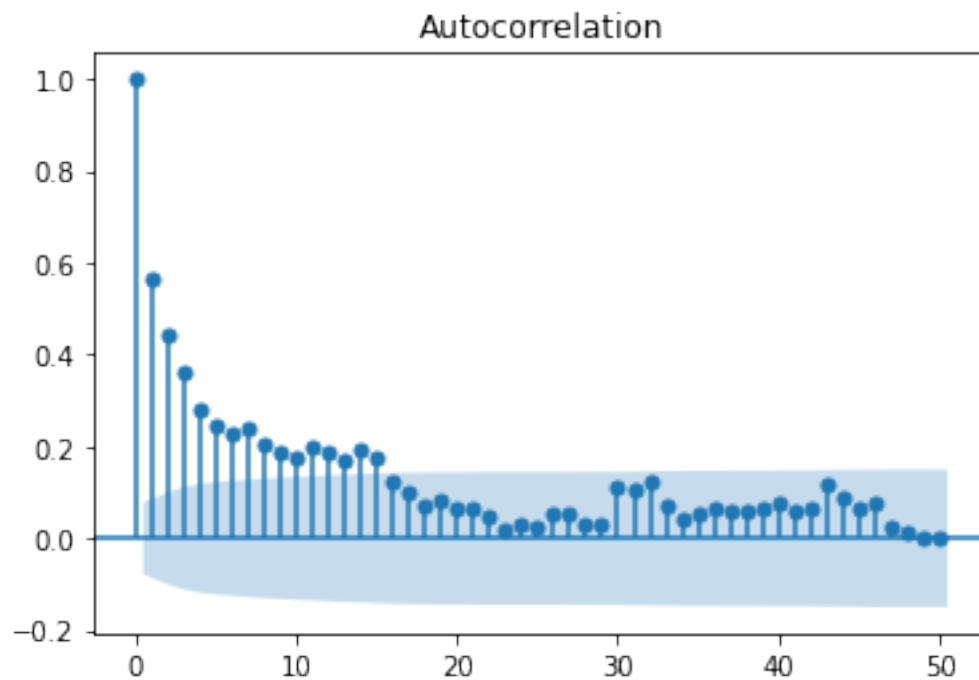
```
[10]: #plotting time series by 30 mins against residue output
f = (12, 8)
residue = df[['Residue']]
residue.plot(figsize = f)
```

```
[10]: <AxesSubplot:>
```



```
[11]: #Plotting time series correlation by 50 lags for residue ouput  
plt.figure(figsize=(24,16))  
residue_time_corr = plot_acf(residue, lags=50)  
plt.show(residue_time_corr)
```

<Figure size 1728x1152 with 0 Axes>



```
[12]: #Plotting time series partial correlation by 50 lags for residue ouput
plt.figure(figsize=(24,16))
plot_pacf(residue, lags=50)
plt.show()
```

<Figure size 1728x1152 with 0 Axes>

