Excercise 1

Review of data wrangling and visualization in Python In this excercise we will familiarize ourselves with pandas, and matplotlib package.

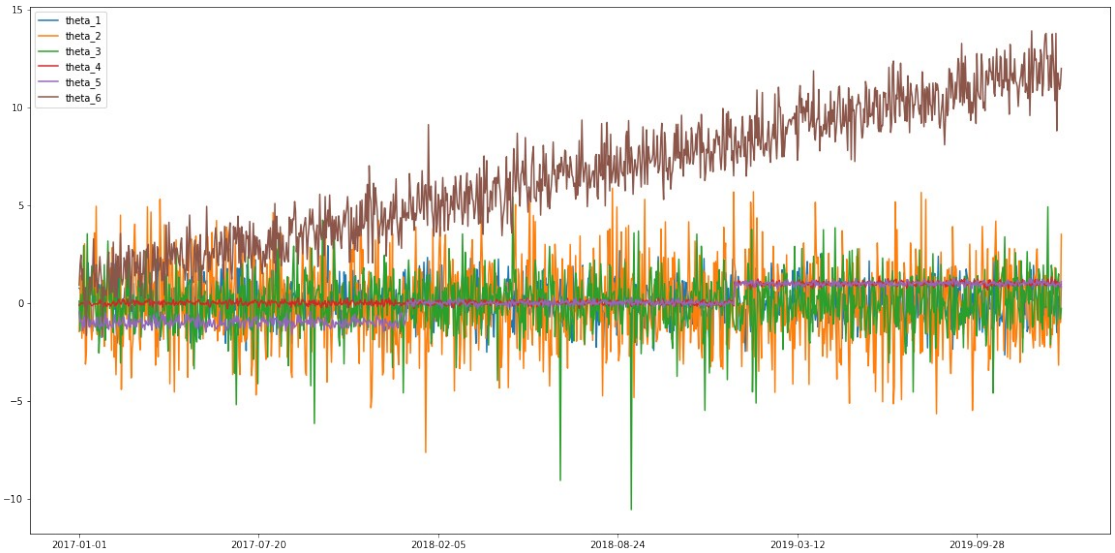
Course of excercise:

- Import Data1.csv file to python.
- Set first column as the index.
- Plot all columns as time series.
- Plot histograms of all columns, verify bin size. Plot all on a single, faceted plot.
- Plot KDE-s (Kernel Denisty Estimators) for all columns.
- Repeat analysis for columns θ_1-θ_4 in 2018.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

data = pd.read_csv("data1.csv", index_col=0)
data.head()
```

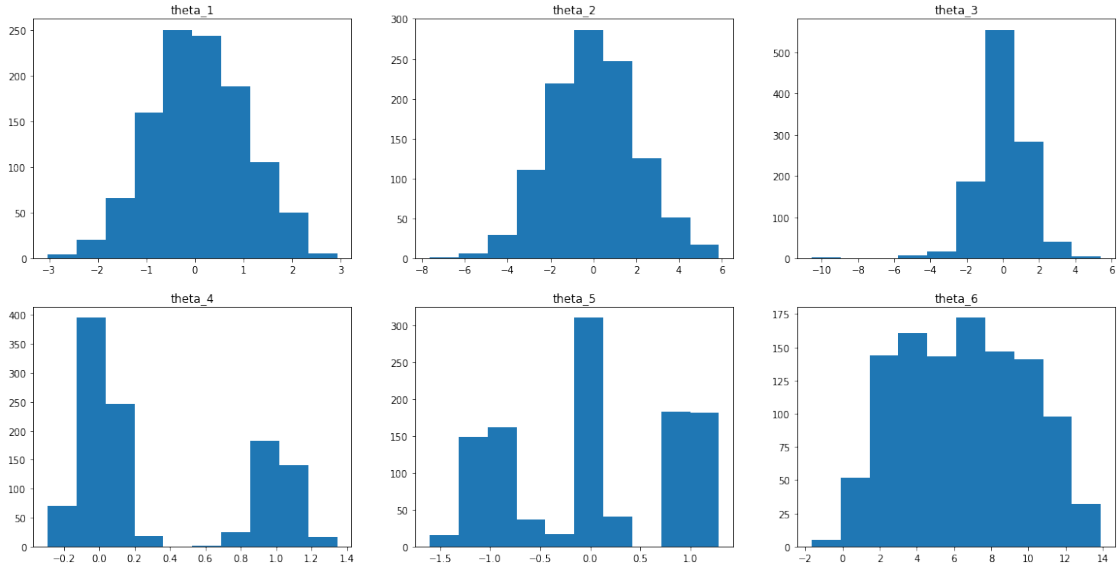
	theta_1	theta_2	theta_3	theta_4	theta_5	theta_6
2017-01-01	0.756936	-1.467790	0.096136	-0.115306	-0.447908	0.902579
2017-01-02	0.767089	0.185797	-1.428536	-0.086443	-0.954288	1.930909
2017-01-03	0.404544	1.415887	0.443466	0.000200	-0.892351	2.449691
2017-01-04	1.313957	-1.804471	-0.836986	0.011785	-1.012518	1.182085
2017-01-05	0.209862	1.315868	0.140993	-0.046473	-1.417092	1.742433

```
cols = data.columns

for col in cols:
  data[col].plot(subplots=False, figsize=(20,10), legend=True)
```



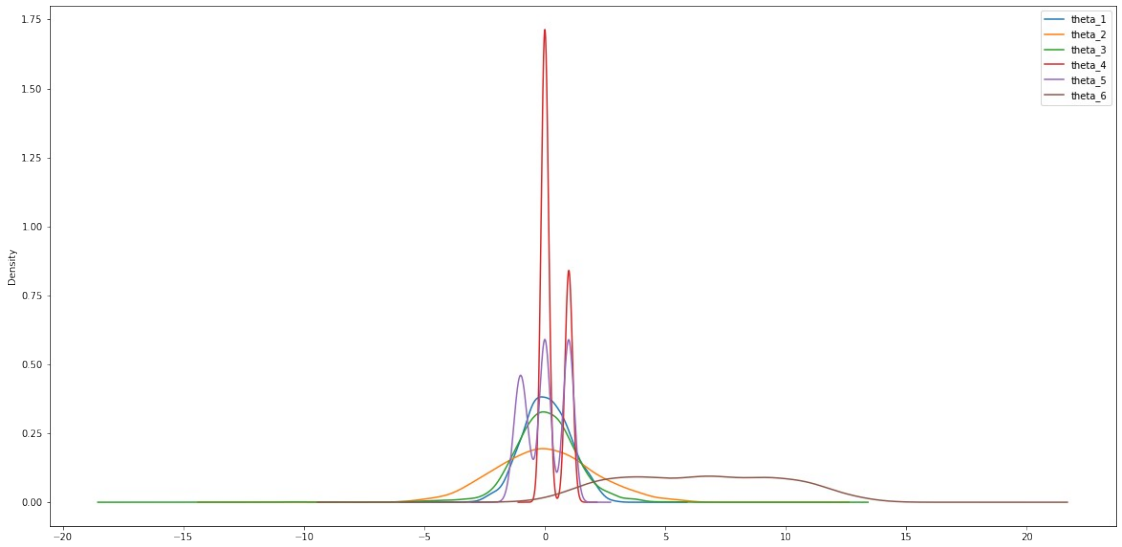
```
fig, axes = plt.subplots(2, 3, figsize=(20, 10))

axes[0, 0].hist(data['theta_1'])
axes[0, 0].set_title('theta_1')

axes[0, 1].hist(data['theta_2'])
axes[0, 1].set_title('theta_2')

axes[0, 2].hist(data['theta_3'])
axes[0, 2].set_title('theta_3')

axes[1, 0].hist(data['theta_4'])
axes[1, 0].set_title('theta_4')

axes[1, 1].hist(data['theta_5'])
axes[1, 1].set_title('theta_5')

axes[1, 2].hist(data['theta_6'])
axes[1, 2].set_title('theta_6')
```

```
Text(0.5, 1.0, 'theta_6')
```




```
# Kernel Density Estimation
data.plot.kde(legend=True, figsize=(20,10))
```

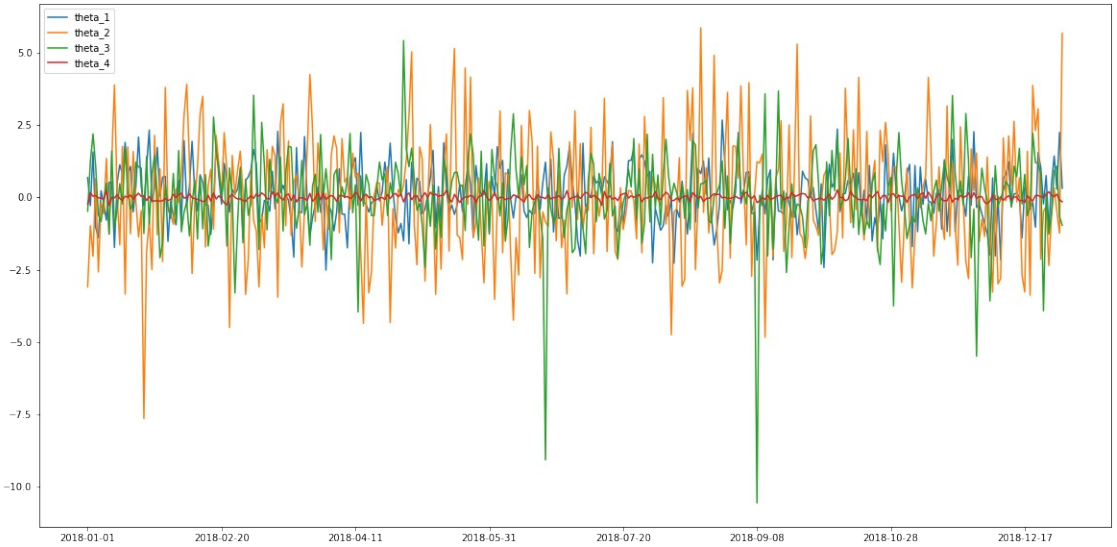
```
<matplotlib.axes._subplots.AxesSubplot at 0x7ff66f290410>
```



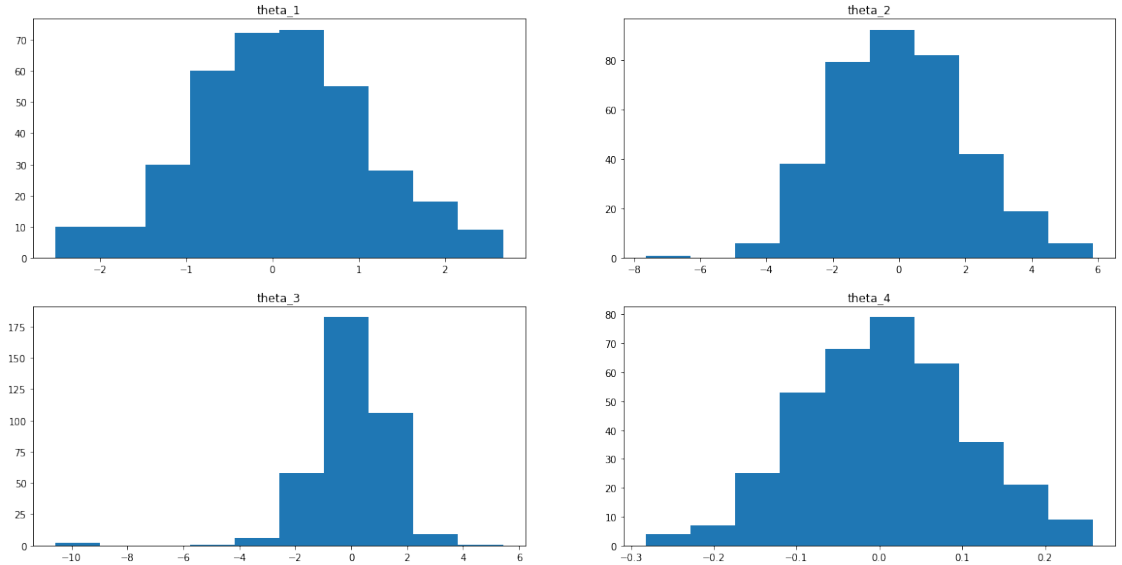

```
# Repeat analysis for columns  θ1 - θ4  in 2018
mask = (data.index >= '2018-01-01') & (data.index <= '2018-12-31')
data_2018 = data.loc[mask]
data_2018 = data_2018[['theta_1', 'theta_2', 'theta_3', 'theta_4']]

data_2018.head()
```

	theta_1	theta_2	theta_3	theta_4
2018-01-01	0.682693	-3.091767	-0.475717	-0.238530
2018-01-02	-0.283107	-0.979955	1.233933	0.158031
2018-01-03	1.572221	-2.033528	2.196317	0.041347
2018-01-04	-1.042981	0.651530	1.060125	0.064832
2018-01-05	-1.392614	-2.570905	-0.600063	-0.015025

```
cols = data_2018.columns

for col in cols:
  data_2018[col].plot(subplots=False, figsize=(20,10), legend=True)
```




```
fig, axes = plt.subplots(2, 2, figsize=(20, 10))

axes[0, 0].hist(data_2018['theta_1'])
axes[0, 0].set_title('theta_1')

axes[0, 1].hist(data_2018['theta_2'])
axes[0, 1].set_title('theta_2')

axes[1, 0].hist(data_2018['theta_3'])
axes[1, 0].set_title('theta_3')

axes[1, 1].hist(data_2018['theta_4'])
axes[1, 1].set_title('theta_4')

Text(0.5, 1.0, 'theta_4')
```



```
data_2018.plot.kde(legend=True, figsize=(20,10))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ff66eb39610>
```