EN.553.753
Junzhou Lin

Project 1[Detail is in the coding provided]

Problem 1:

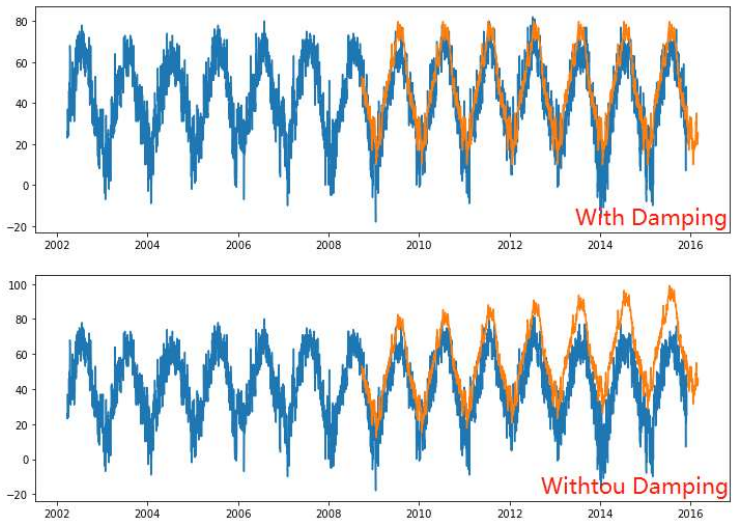
Step 1) Getting the Data

Step 2) Using ADF (df['TMin'],autolag='AIC') to check whether the data, the seasonally lagged data and the data minus it's seasonally MA is stationary. All test statistic is below the thresholds, therefore minimum temperature series is stationary.

Step 3) Considering "Global warming" and "Seasonality", I choose to use exponential smoothing methods to build our model instead of ARIMA.

```
model1 = ExponentialSmoothing(train['TMin'], trend="add", seasonal="add", damped=True,seasonal_periods=365)
```

Damped=True just because the model with damping seems to perform better.



Step 4) Using the predicted data and a 'for' loop to calculate payout, I have Expected Payout as 13186.81$, and 99th percentile shortfall as 100000$.

Our pricing structure is

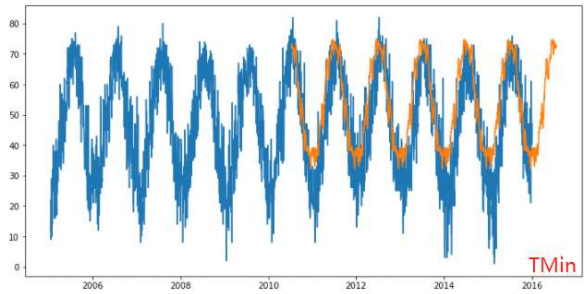
	TMin	payout
2015-12-01	38.775650	0.0
2015-12-02	35.926563	100000.0
2015-12-03	34.257054	100000.0
2015-12-04	35.151517	100000.0
2015-12-05	34.175969	100000.0
...	...	...
2016-02-25	27.182206	0.0
2016-02-26	27.560962	0.0
2016-02-27	28.267086	0.0
2016-02-28	25.680232	0.0
2016-02-29	27.561327	0.0

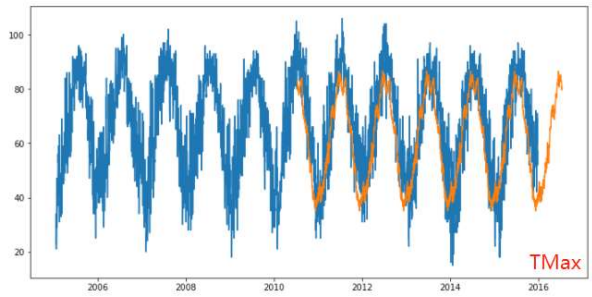
Problem 2:

Step 1) Getting the data and splitting it into two data frames, spot price and temperature.

Step 2) Repeating Step 2,3 from problem 1 for both TMax and TMin. Proven both series are stationary.

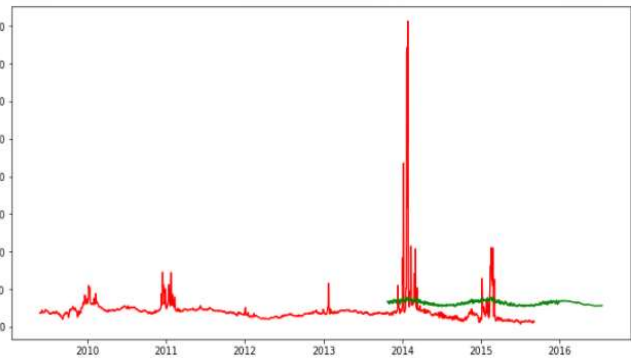
```
modelmin = ExponentialSmoothing(train['TMin'], trend="add", seasonal="add", damped=True,seasonal_periods=365)
modelmax= ExponentialSmoothing(train['TMax'], trend="add", seasonal="add", damped=True,seasonal_periods=365)
fitmin = modelmin.fit()
fitmax = modelmax.fit()
```





Step 3) Using Linear Regression model to predict the Spot price. Our test data range from 2013-10-24 to 2016-07-18.

	TMax	TMin	Spot Price	PredSpot
2013-10-24	53.000000	37.000000	3.82	6.670779
2013-10-25	56.000000	36.000000	3.78	6.641298
2013-10-26	56.000000	30.000000	3.72	6.745224
2013-10-27	60.000000	35.000000	3.72	6.596215
2013-10-28	65.000000	33.000000	3.72	6.552852
...	...	...	...	...
2016-07-14	86.023879	61.004969	NaN	5.739787
2016-07-15	83.676660	60.681487	NaN	5.782009
2016-07-16	84.251467	60.017429	NaN	5.784543
2016-07-17	84.545660	60.963205	NaN	5.763572
2016-07-18	83.333384	59.836917	NaN	5.801993



Step 4) At the last, simply put the data into the formula, following is our pricing structure:

	Month	TStrike	PStrike	Tavg	Pavg	Monthly Payout
0	2015-11	50	1.802	50.933333	6.478826	0.000000
1	2015-12	41	2.999	48.519939	6.557157	0.000000
2	2016-01	37	5.295	34.864947	7.008445	182914.835295
3	2016-02	38	5.134	38.031097	6.904917	0.000000
4	2016-03	48	2.525	46.807092	6.615283	243966.481753

Therefore, the expected five month payout is 85376.26$, and the 99th percentile shortfall is 241524.42$