US TREASURY YIELDS DATA ANALYSIS

BY KEERTHI PRAKASH

LANTERN INSTITUTE

Overview:

- Introduction
- About Data : Data exploration, Data preprocessing
- Time Series of Yields, 1 Day & 25 Day Increments
- Histogram and Density plots Yields, 1 Day & 25 Day Increments
- Distribution Fitting
- Term Structure of Yield
- Standard Deviation : 1-day and 25-day increments
- Correlation Matrix
- SARIMA Model
- Conclusion

What is the Treasury Yield?

- The Treasury yield is the interest rate that the U.S. government pays to borrow money for different lengths of time.

- Each of the Treasury securities (T-bonds, T-bills, and T-notes) has a different yield; longer-term Treasury securities usually have a higher yield than shorter-term Treasury securities.

- Treasury yields reflect how investors feel about the economy; the higher the yields on long-term instruments, the more optimistic their outlook.

The formula for calculating the Treasury yield on notes and bonds held to maturity is:

Treasury Yield = [C + ((FV - PP) / T)] ÷ [(FV + PP)/2]

where C= coupon rate

FV = face value

PP = purchase price

T = time to maturity

Data Exploration

- The trend of 55 years(1962 to 2017) of US treasury yields. insights: From the Fred's website it is clear that
- As we can see the yields around 1980 for all maturity lengths was significantly high (approx 16%) compared to those of present(approx 2-3%).
- We can see great recession between 2007 and 2009.
- Most of the public holiday and weekend data is missing.
- Some short terms yields are missing during most of the years.

Data Preprocessing

- Missing weekend and public holiday data because markets close.
- Considering 1 year, 3 year, 5 year and 10 year yields.
- Yield rates are most correlated with the values from their immediate past e.g. yield on Monday = yield on Friday.
- For weekends and public holiday instead of imputing values, we modified the time series such that there are no breaks after we ignore the weekend & public holiday timestamps.

Data

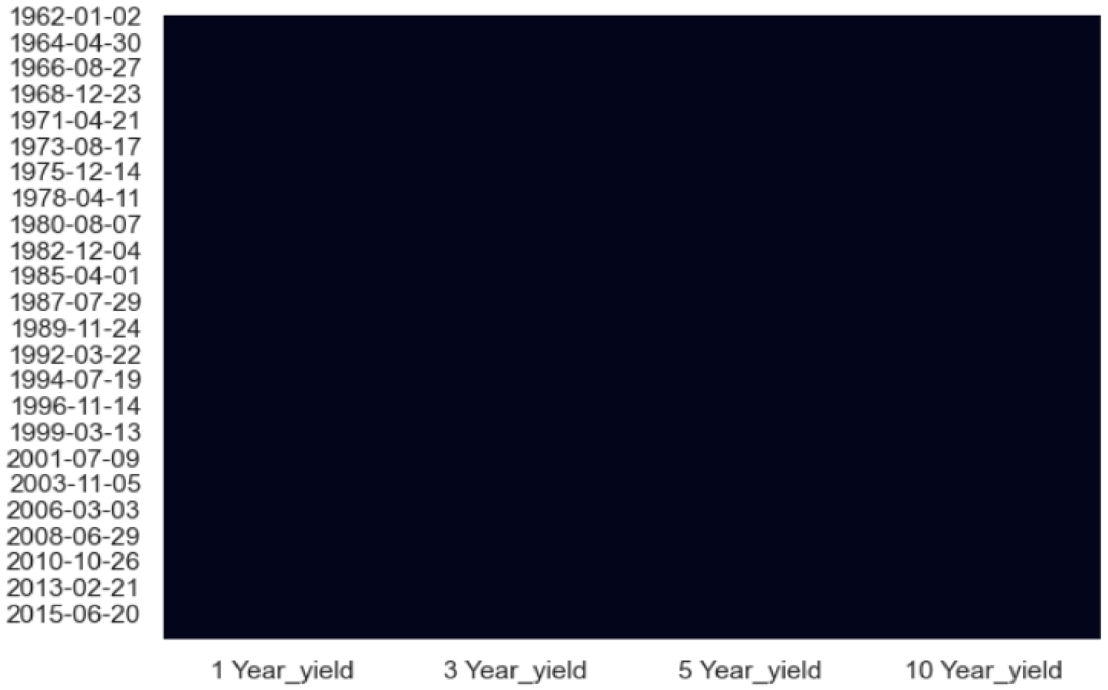
```
yield_data.head()
```

	1 Year_yield	3 Year_yield	5 Year_yield	10 Year_yield
1962-01-02	3.22	3.70	3.88	4.06
1962-01-03	3.24	3.70	3.87	4.03
1962-01-04	3.24	3.69	3.86	3.99
1962-01-05	3.26	3.71	3.89	4.02
1962-01-06	3.26	3.71	3.89	4.02

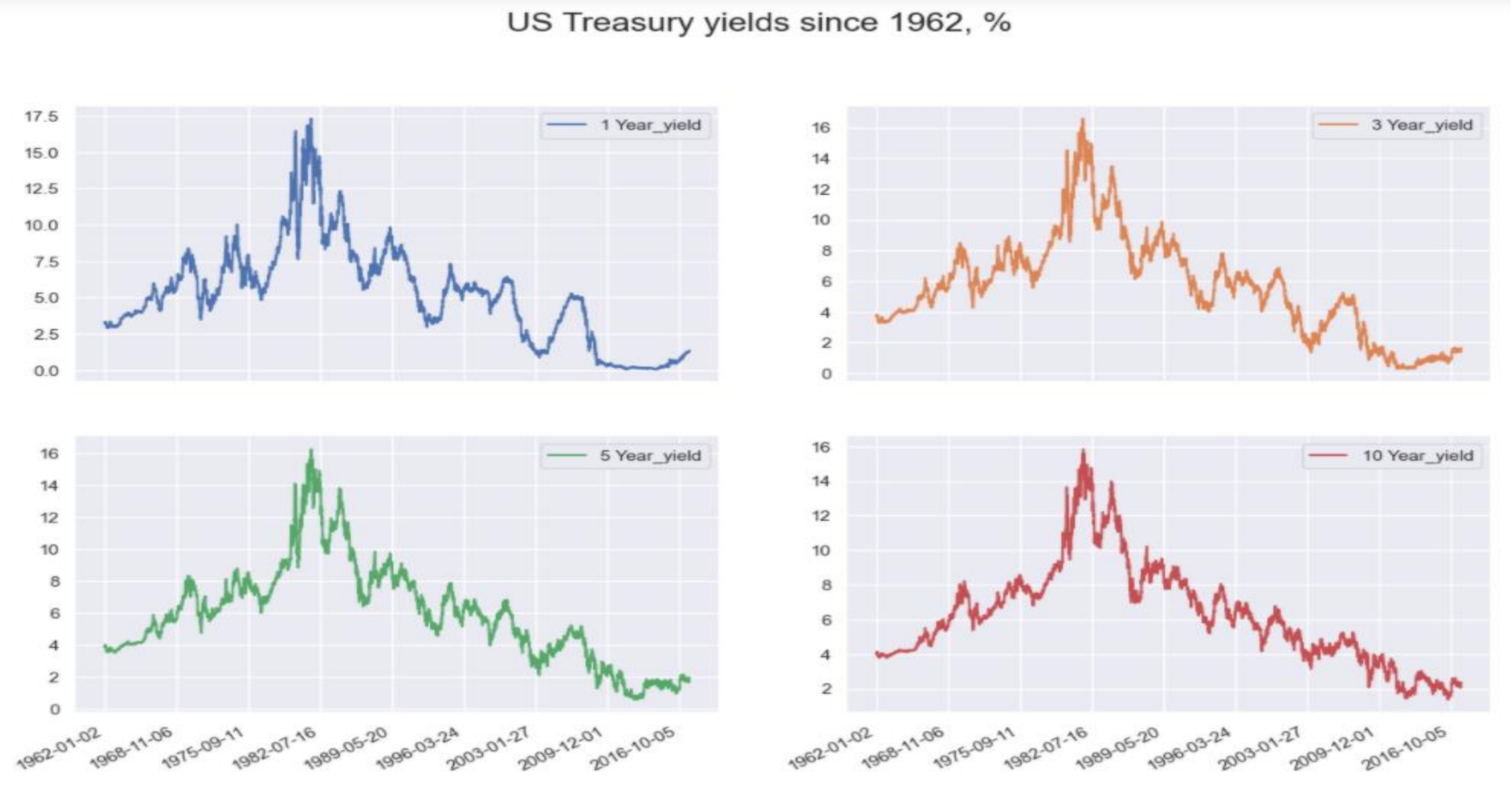
	1 Year_yield	3 Year_yield	5 Year_yield	10 Year_yield
count	20366.000000	20366.000000	20366.000000	20366.000000
mean	5.219524	5.649806	5.921859	6.264055
std	3.411029	3.258169	3.098484	2.861167
min	0.080000	0.280000	0.560000	1.370000
25%	3.060000	3.530000	3.860000	4.190000
50%	5.280000	5.650000	5.820000	6.020000
75%	7.090000	7.500000	7.720000	7.840000
max	17.310000	16.590000	16.270000	15.840000

```
sns.heatmap(yield_data.isnull(), cbar=False)
```

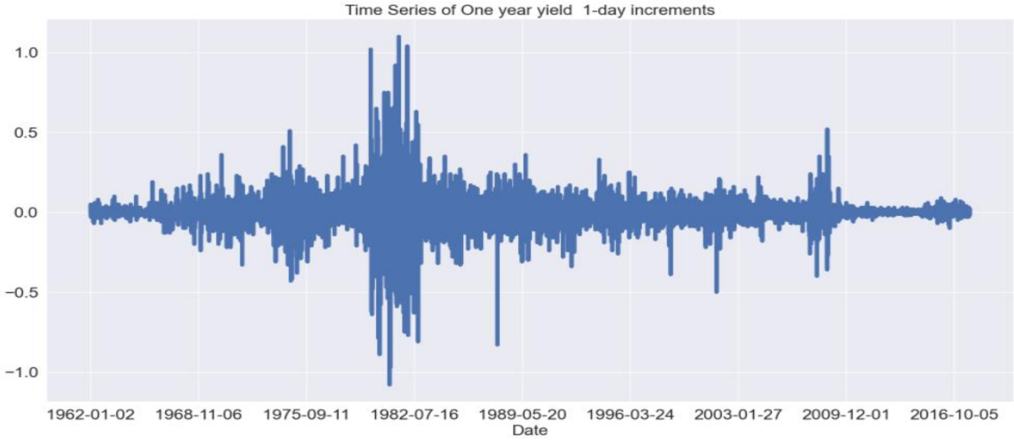
```
<matplotlib.axes._subplots.AxesSubplot at 0x1efd4425ee0>
```

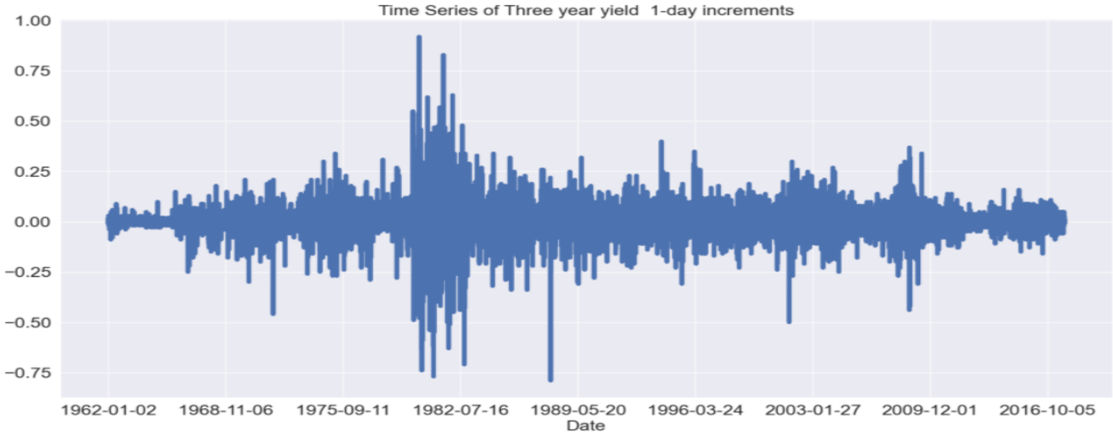


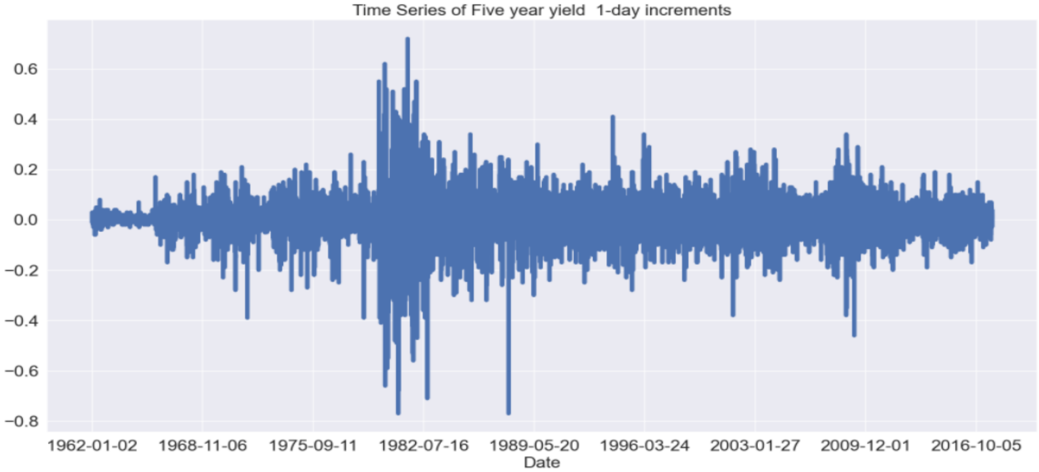
Time Series of the Yields

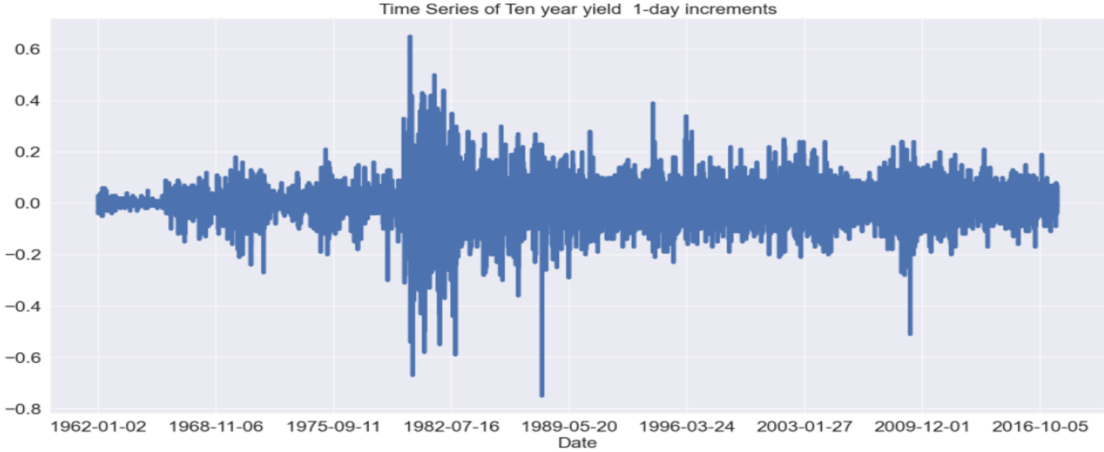


Time Series of Yields 1-day Increment

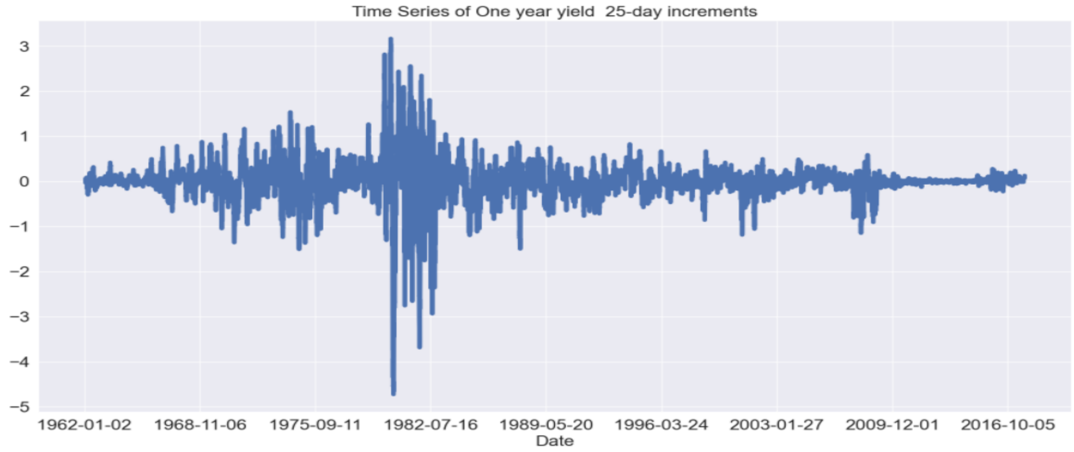


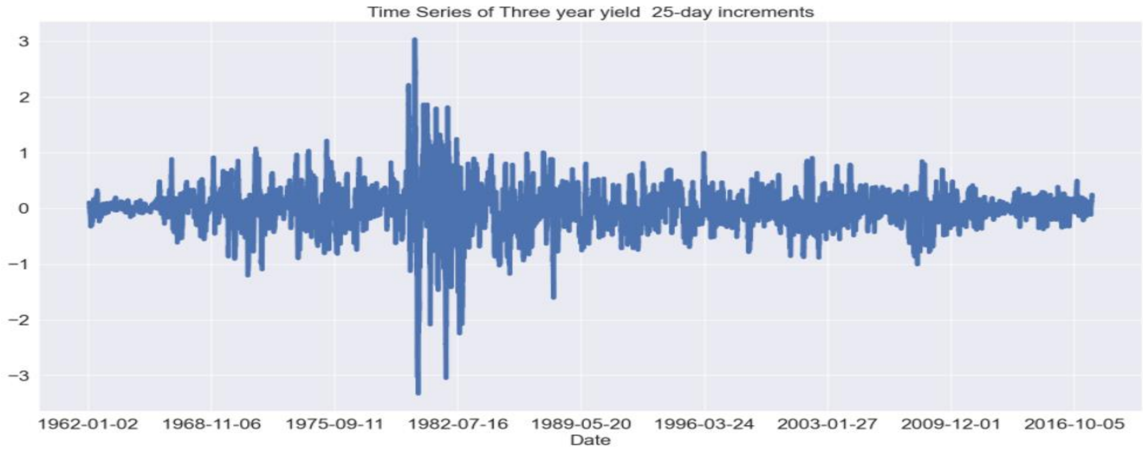


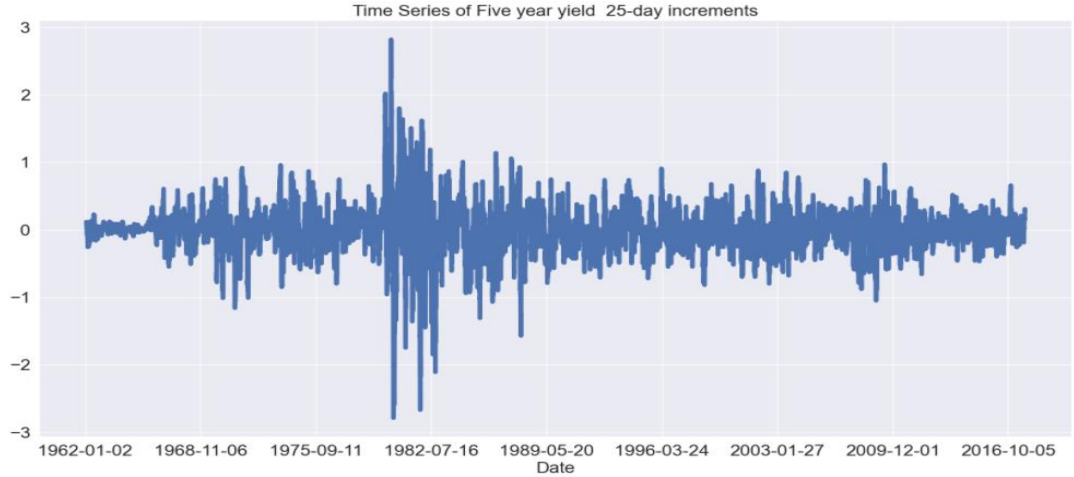


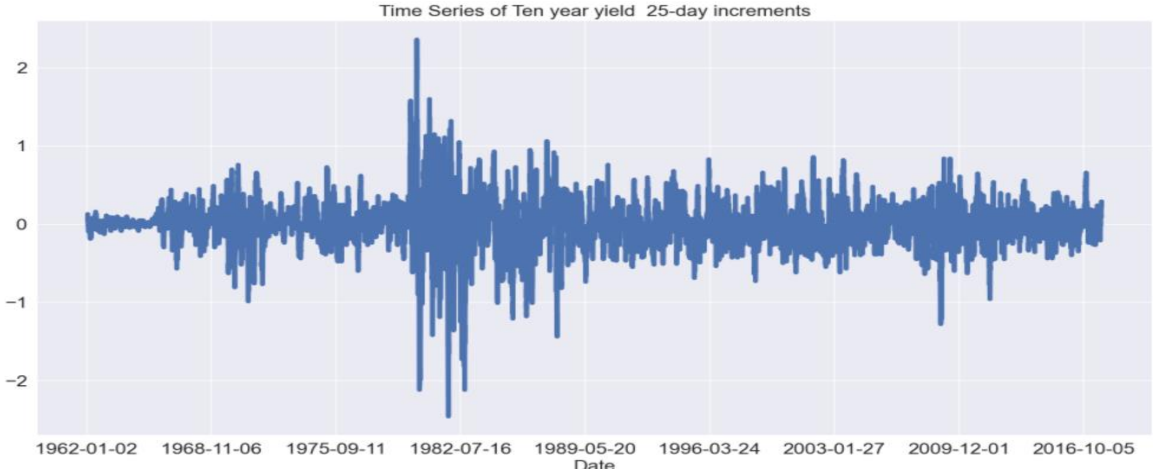


Time Series of Yield 25-day Increments



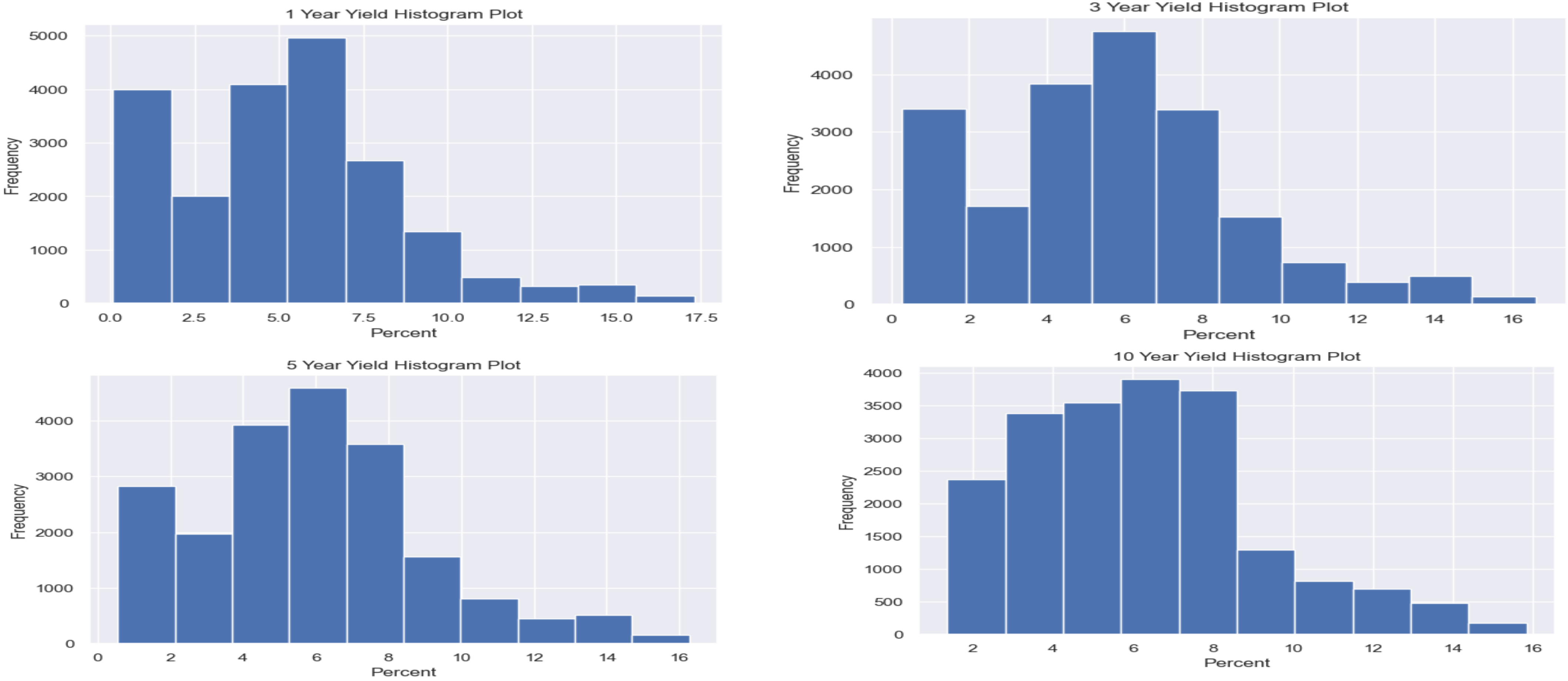




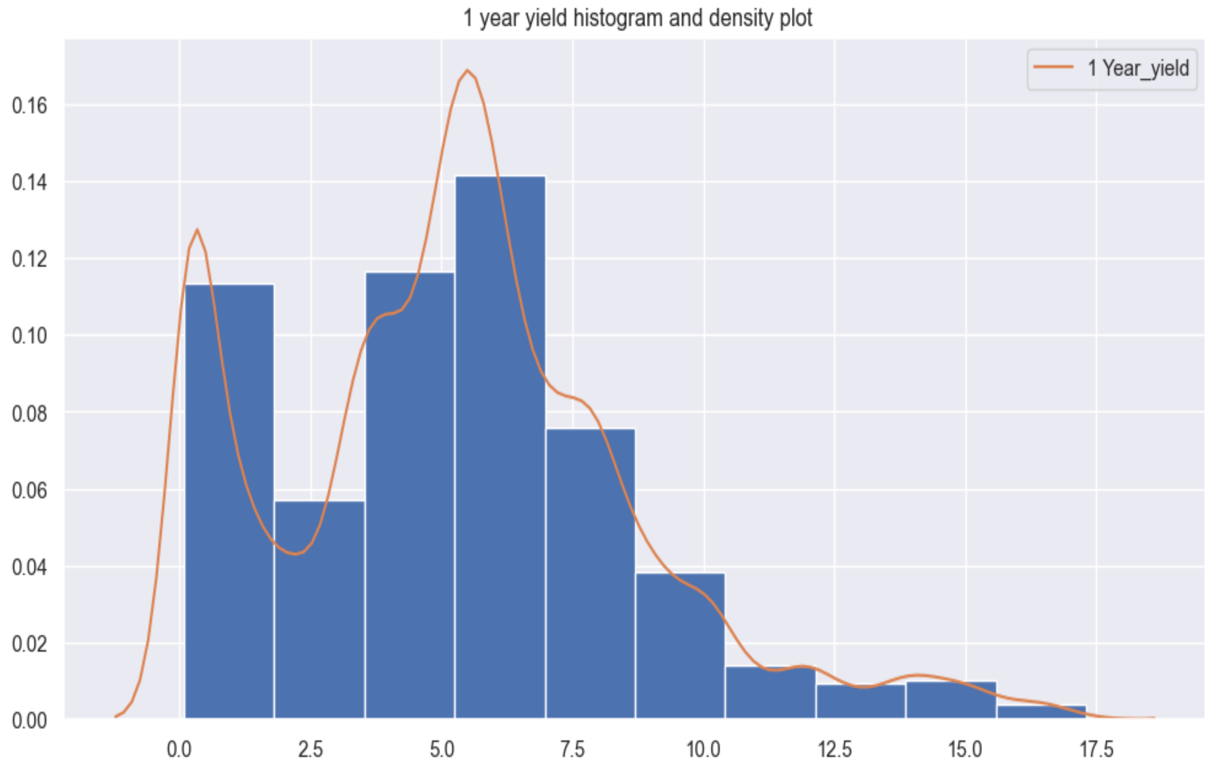


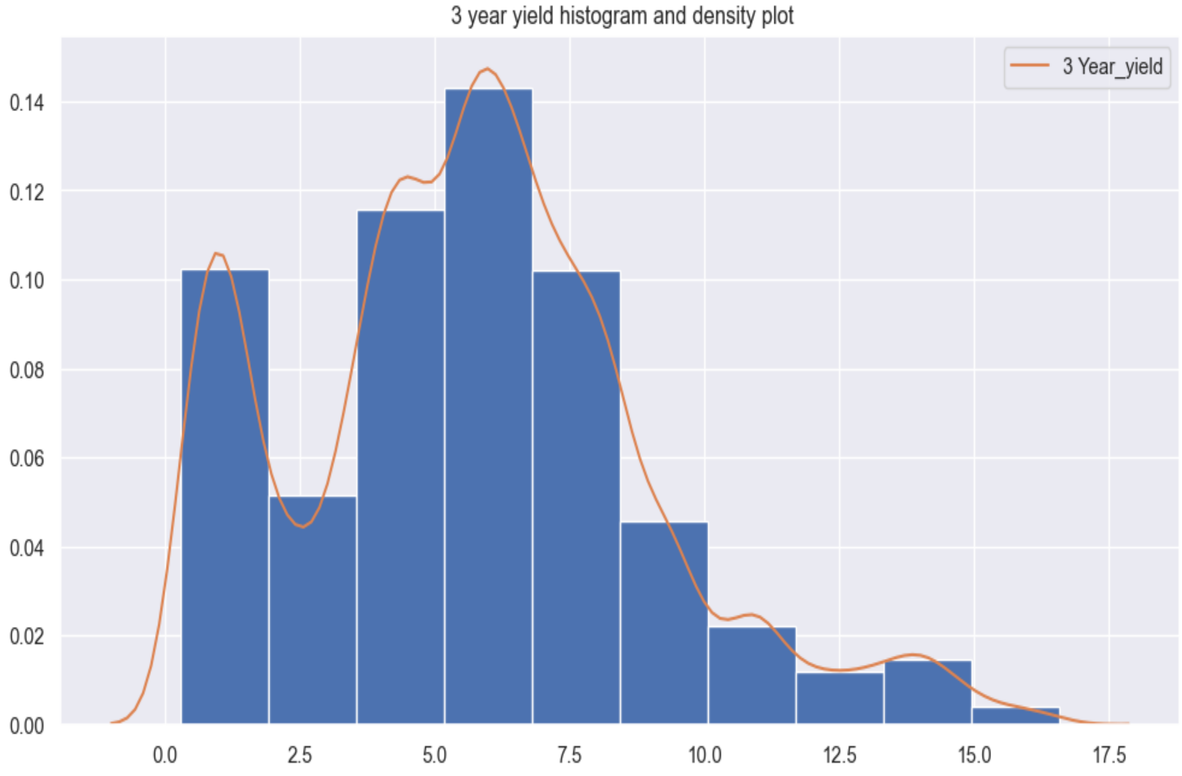
LANTERN INSTITUTE

Histogram of Yields

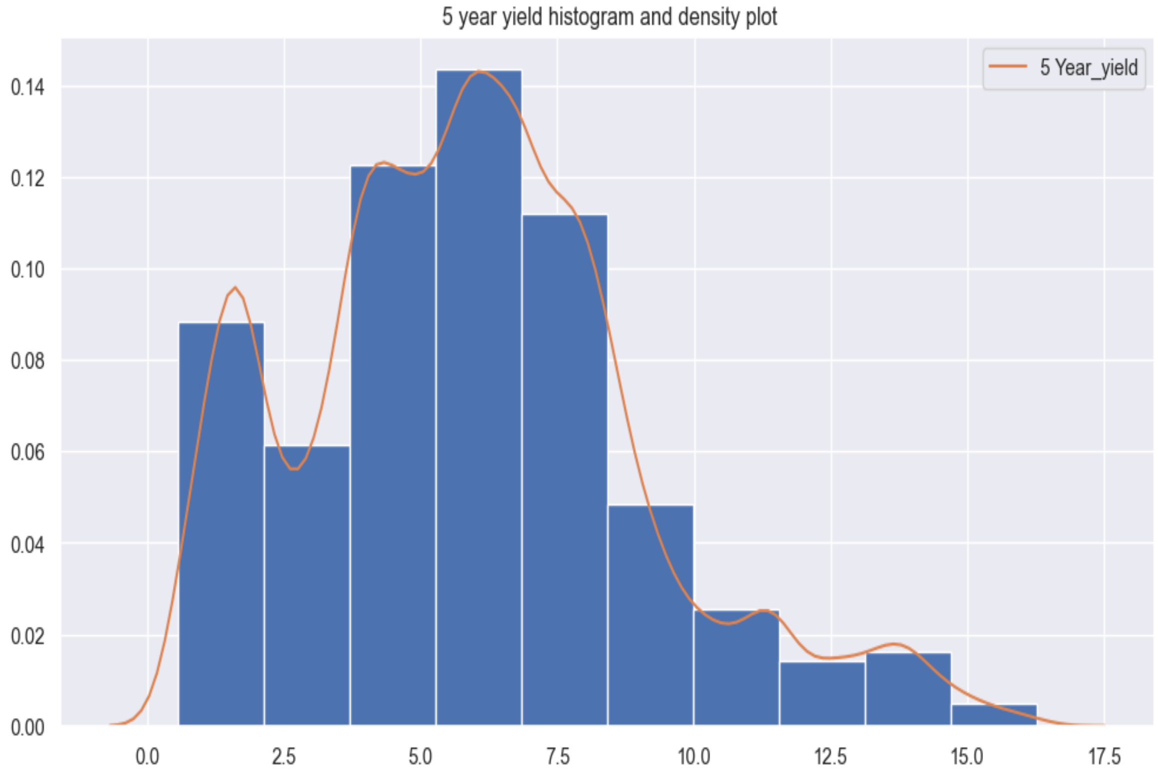


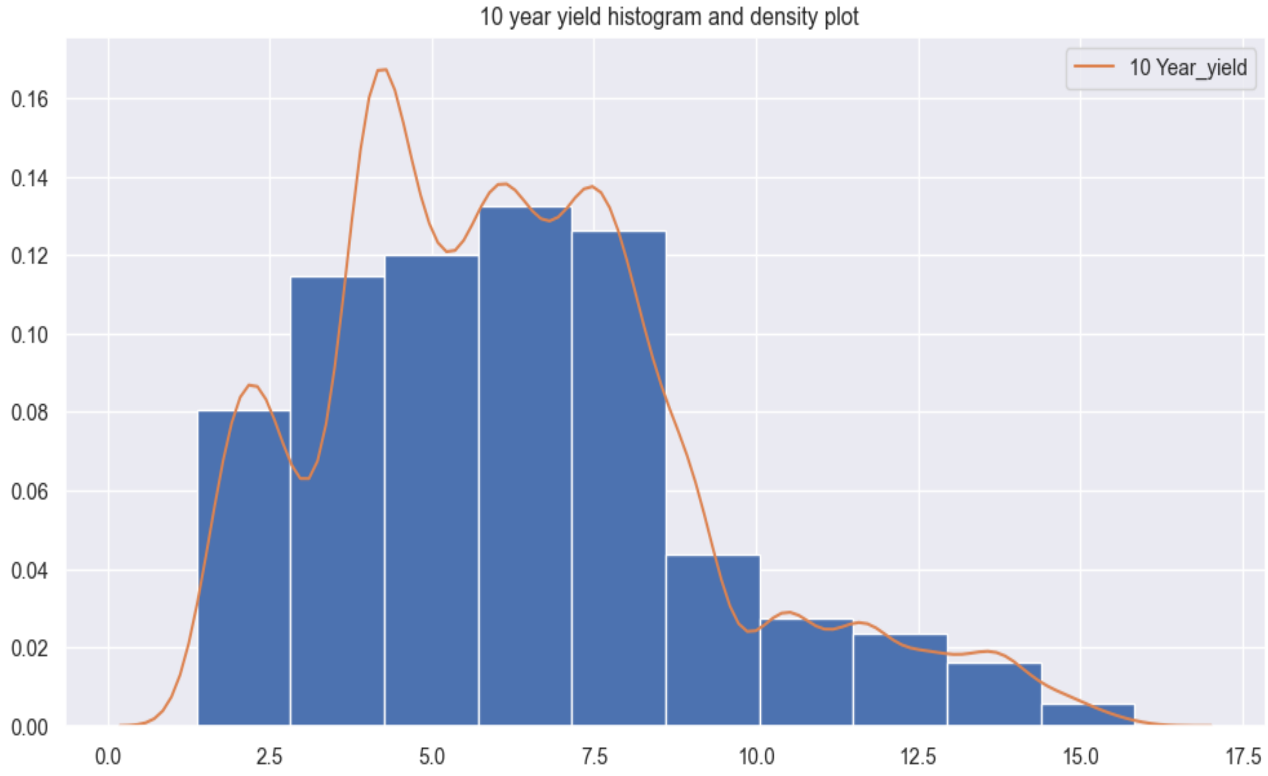
Histogram and Density plot of Yields



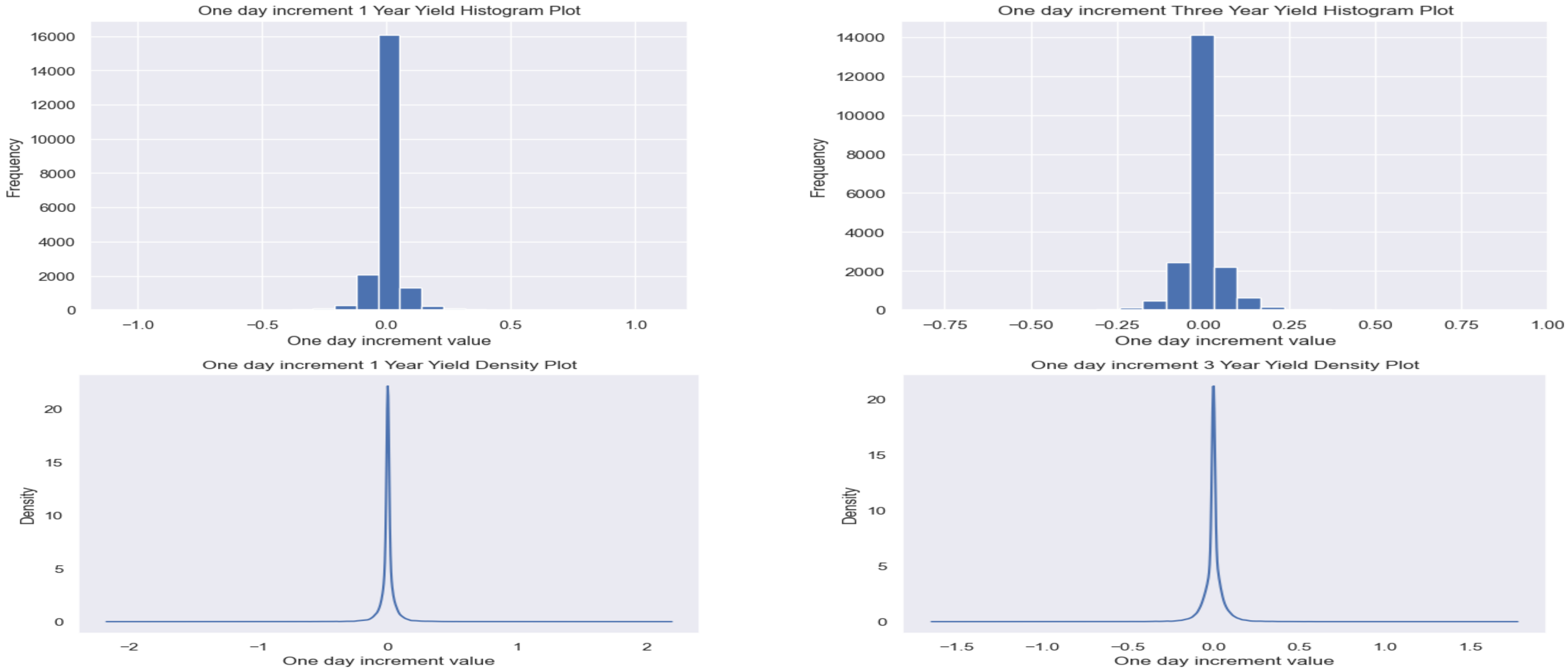


Histogram and Density plot of Yields

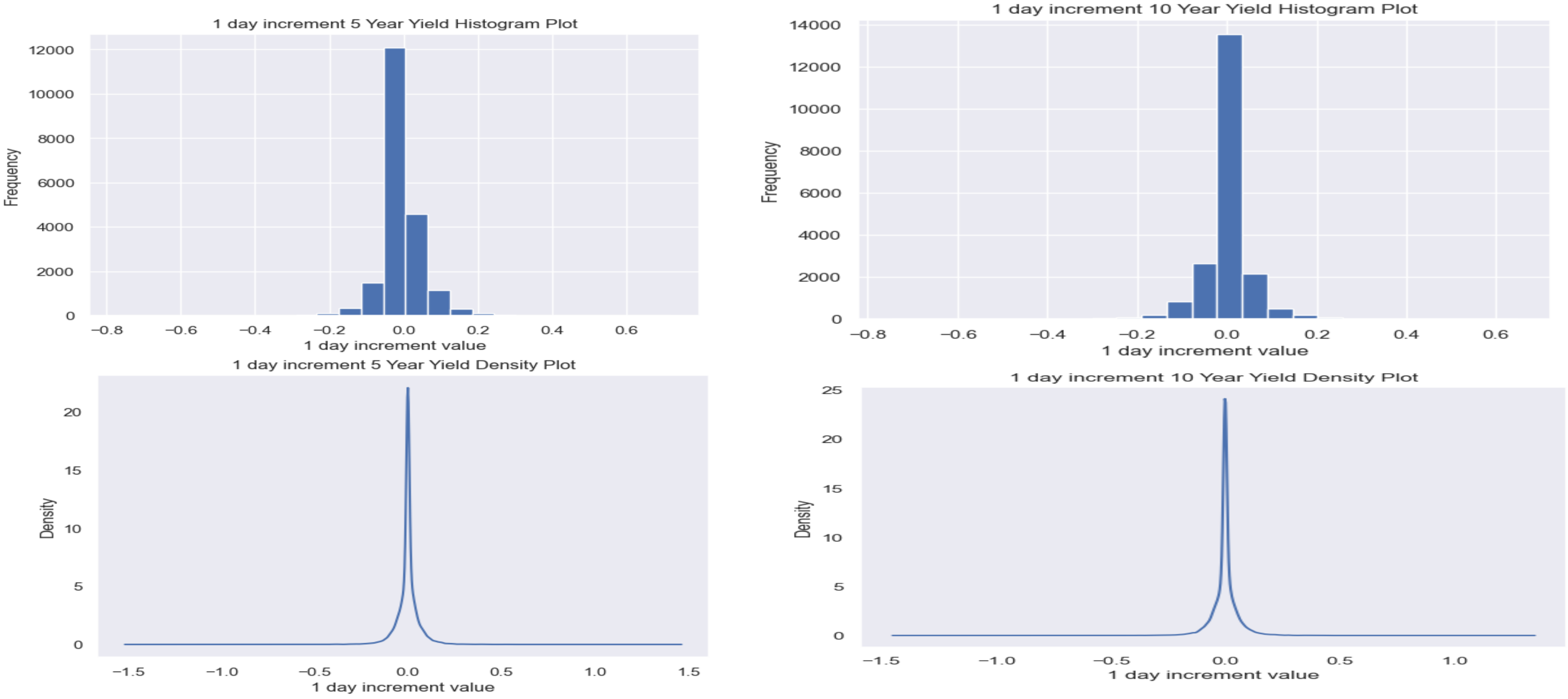




Histogram and Density plot of 1 day increments

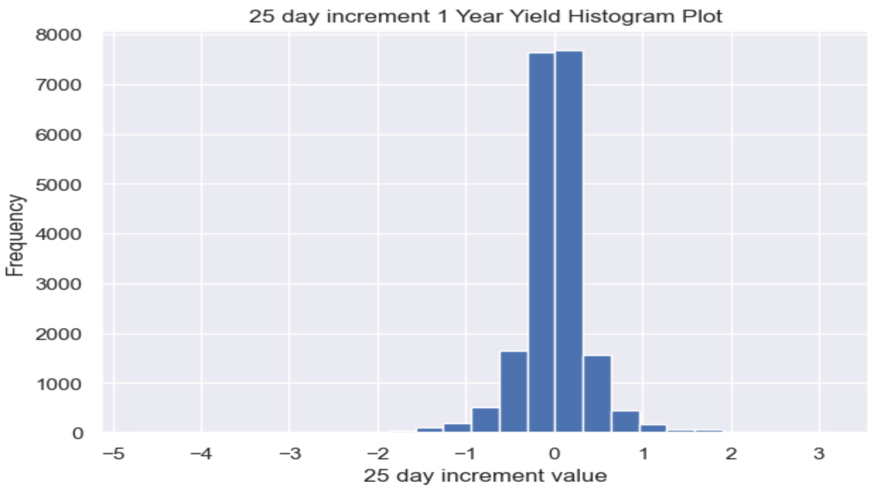


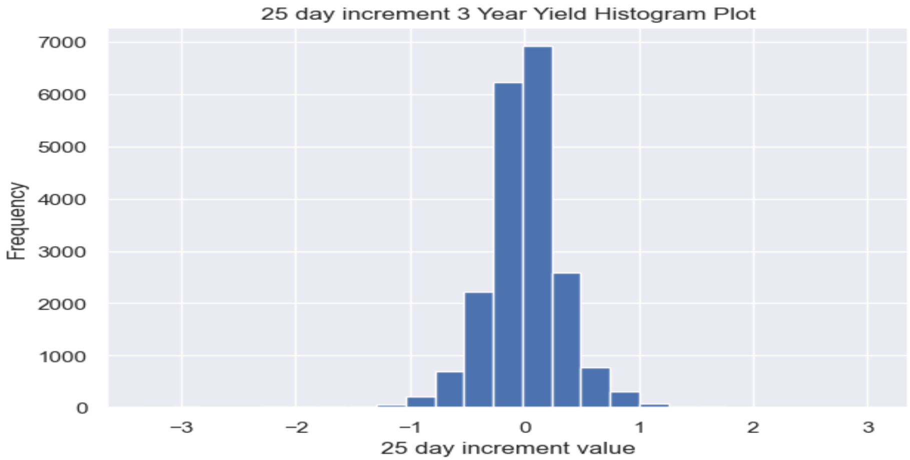
Histogram and Density plot of 1 day increments

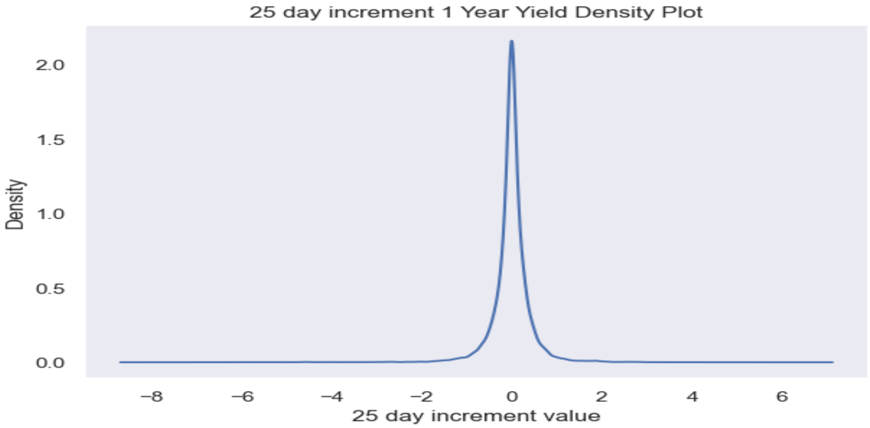


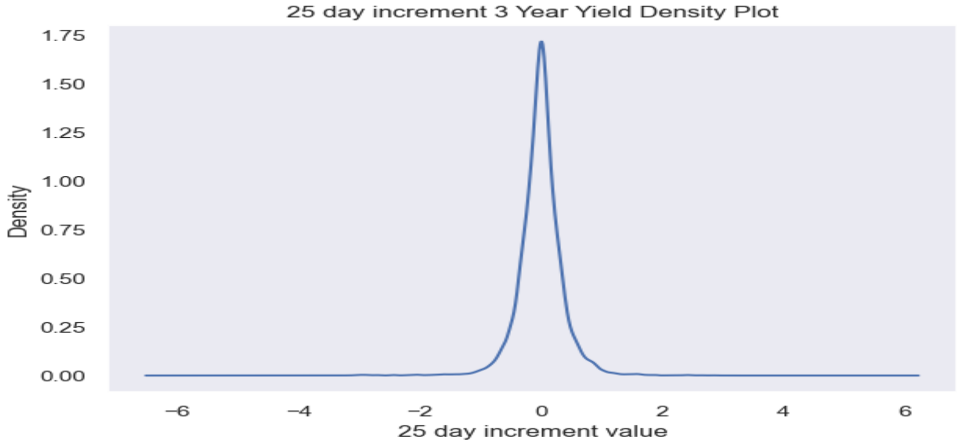
LANTERN INSTITUTE

Histogram and Density plot of 25 day increments

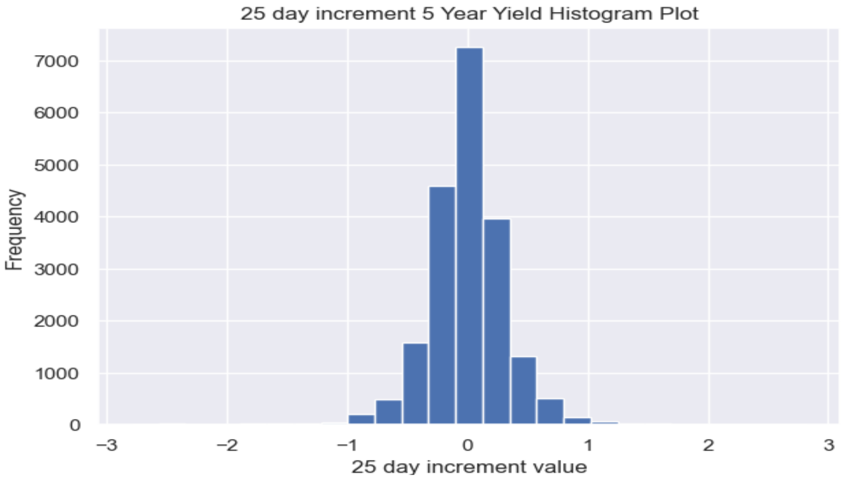


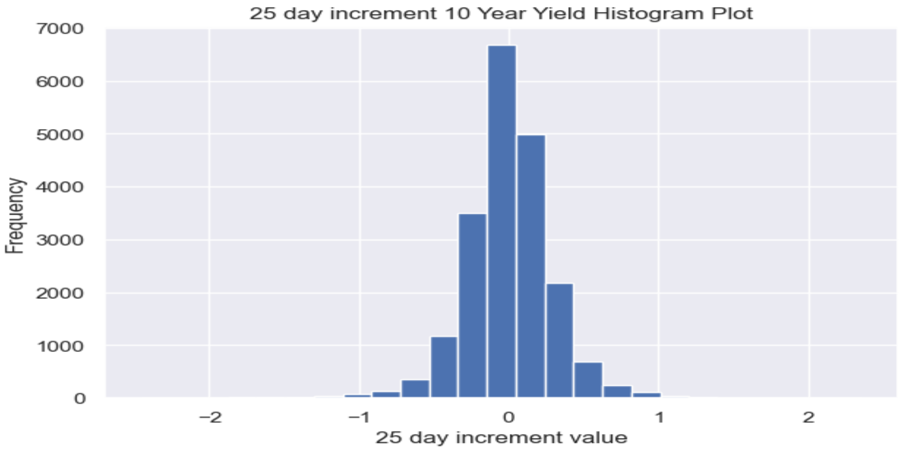


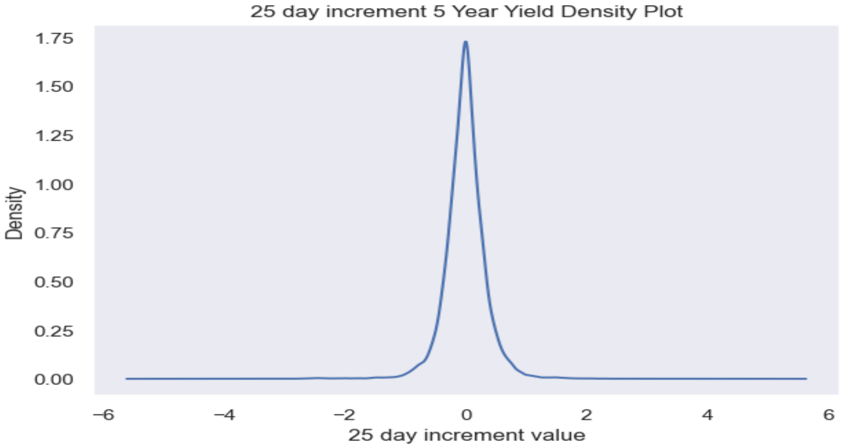


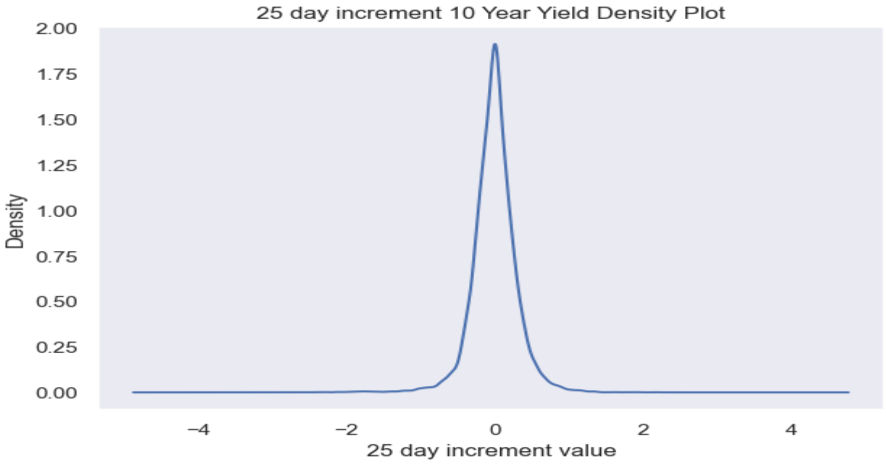


Histogram and Density plot of 25 day increments

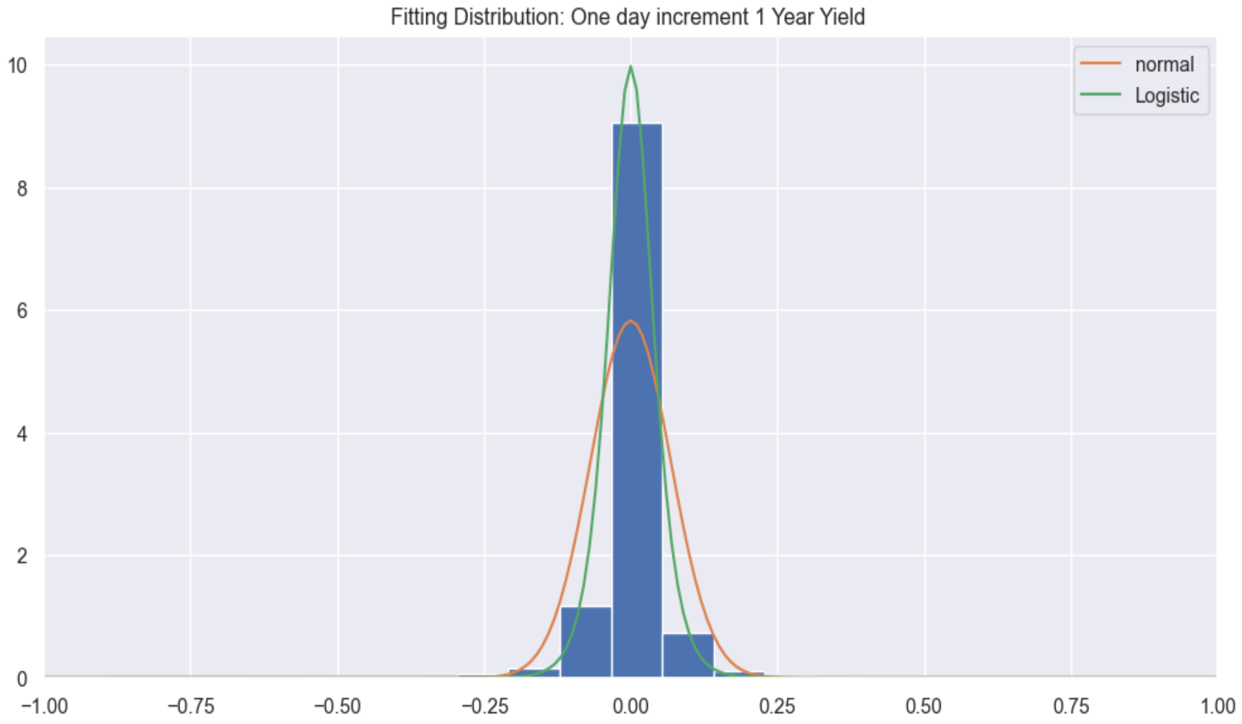








Distribution Fitting : 1 Year Yield

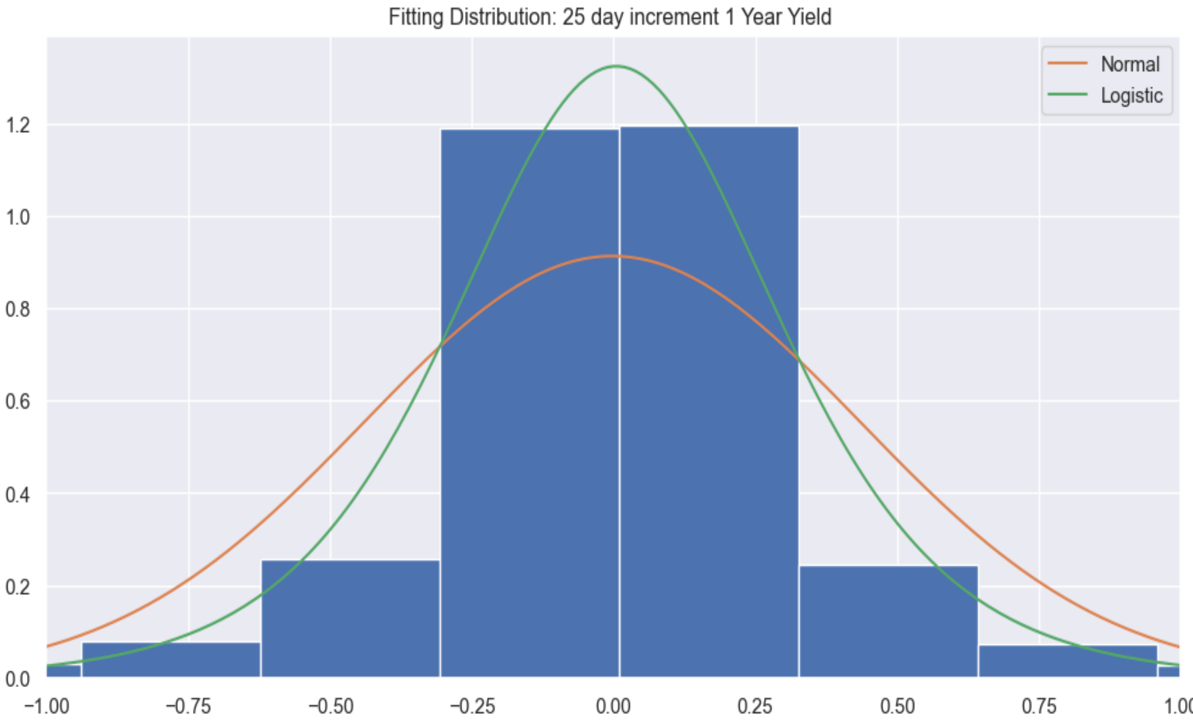



```
print(aic_n, aic_log)
```

-51415.91570046266 -64218.267404516504

```
print(bic_n, bic_log)
```

-51400.0725544225 -64202.42425847634




```
print(aic_n, aic_log)
```

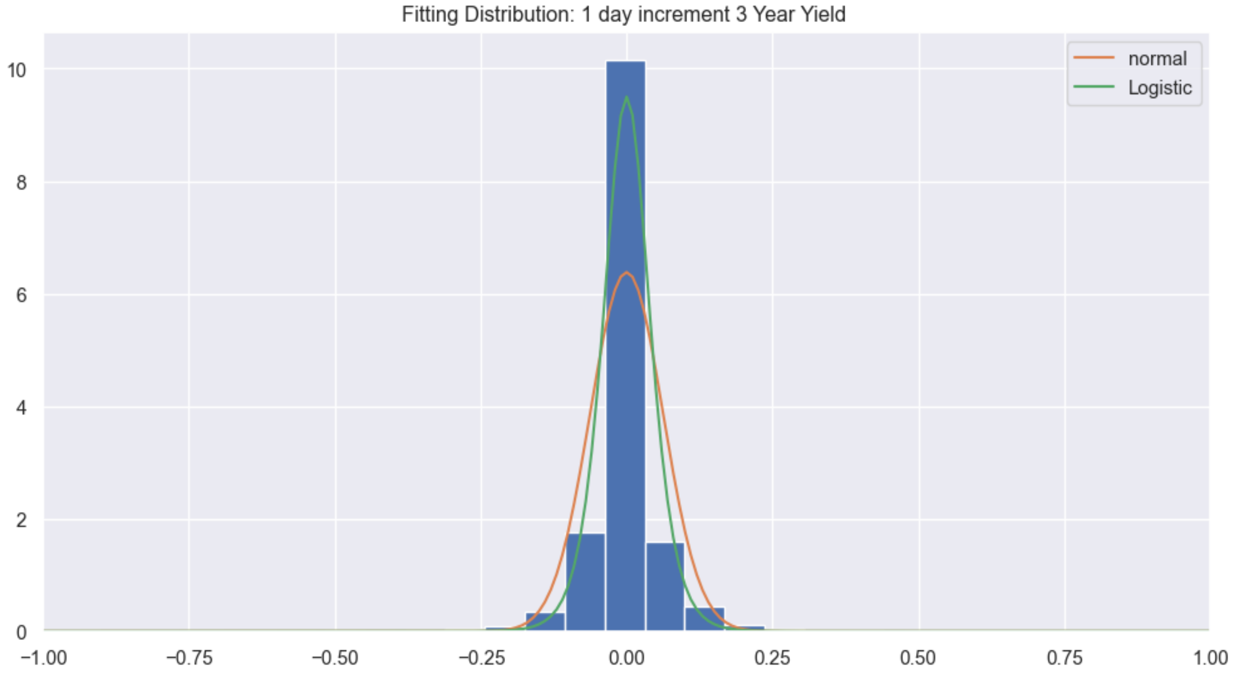
24060.970771303022 16236.427041627629

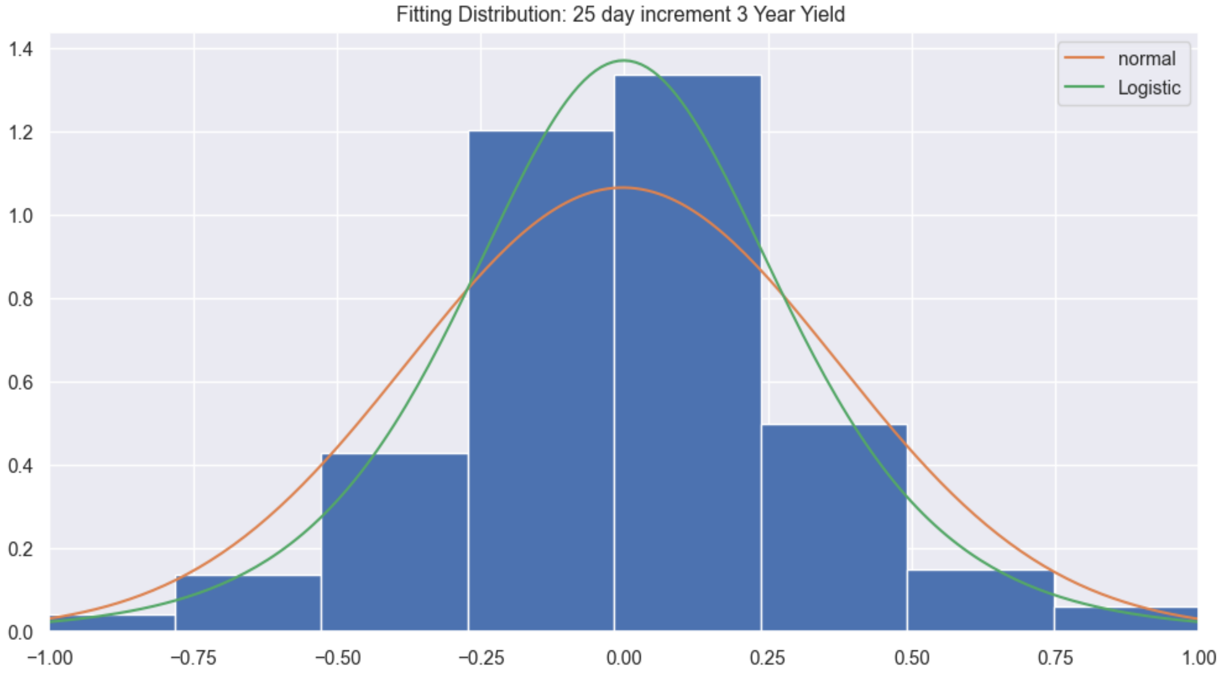
```
print(bic_n, bic_log)
```

24076.811558968227 16252.267829292834

Logistic Distribution fits better than Normal Distribution

Distribution Fitting : 3 Year Yield






```
print(aic_n, aic_log)
```

-55117.39966928548 -63511.18223845474

```
print(bic_n, bic_log)
```

-55101.556523245315 -63495.33909241458

```
print(aic_n, aic_log)
```

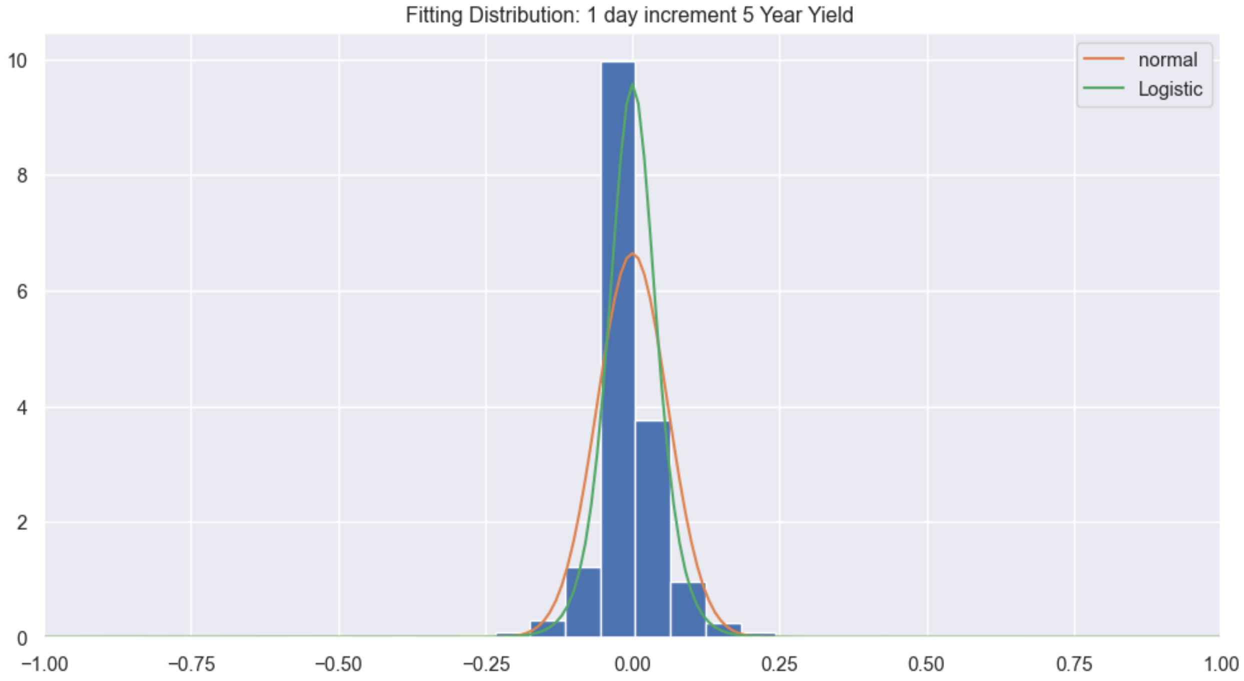
17748.86281176753 13386.133425738133

```
print(bic_n, bic_log)
```

17764.703599432734 13401.974213403339

Logistic Distribution fits better than Normal Distribution

Distribution Fitting : 5 Year Yield

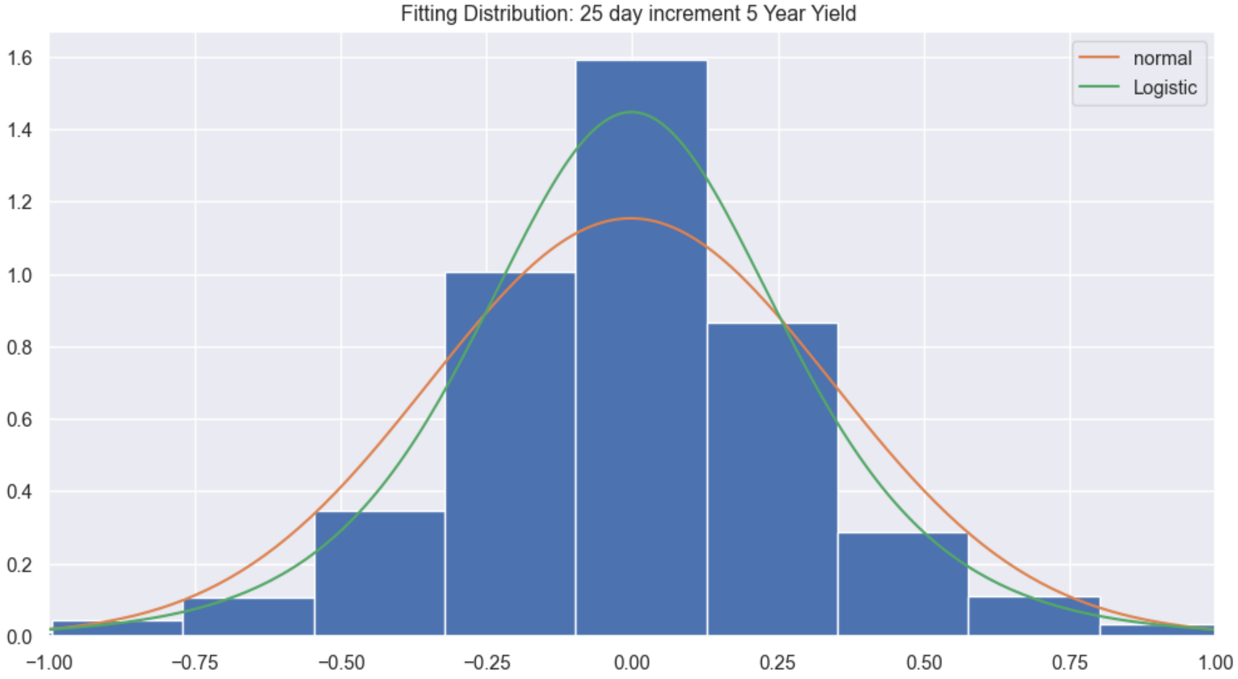



```
print(aic_n, aic_log)
```

-56782.90580082069 -64071.16348146684

```
print(bic_n, bic_log)
```

-56767.06265478053 -64055.32033542668




```
print(aic_n, aic_log)
```

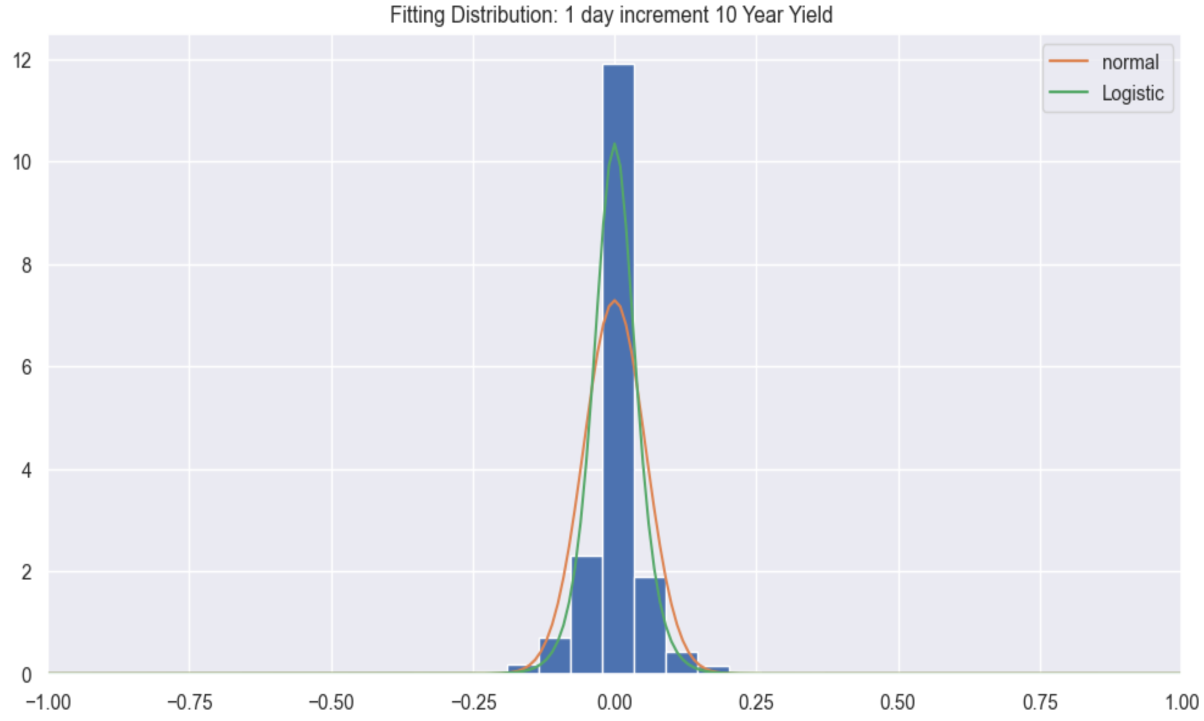
14536.907216342977 10964.133004406962

```
print(bic_n, bic_log)
```

14552.748004008183 10979.973792072167

Logistic Distribution fits better than Normal Distribution

Distribution Fitting : 10 Year Yield

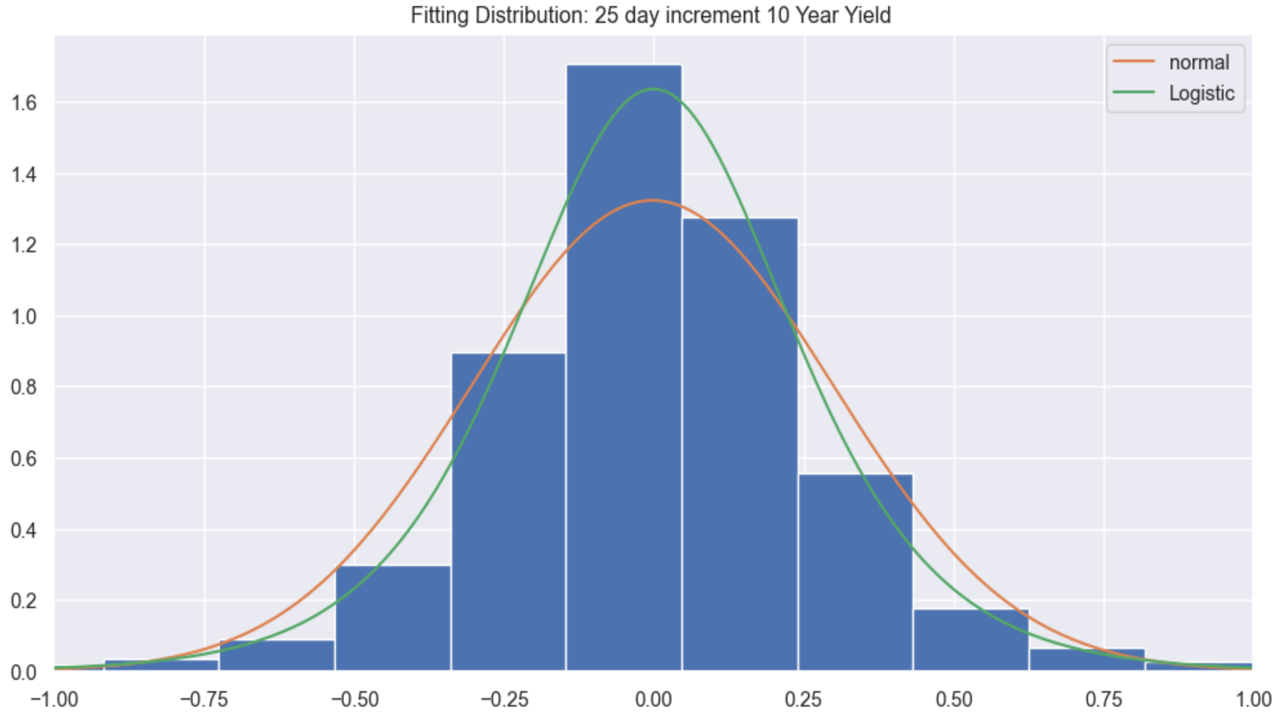



```
print(aic_n, aic_log)
```

-60540.274179628104 -67300.48230284681

```
print(bic_n, bic_log)
```

-60524.43103358794 -67284.63915680665




```
print(aic_n, aic_log)
```

8970.09258975655 5846.226514839199

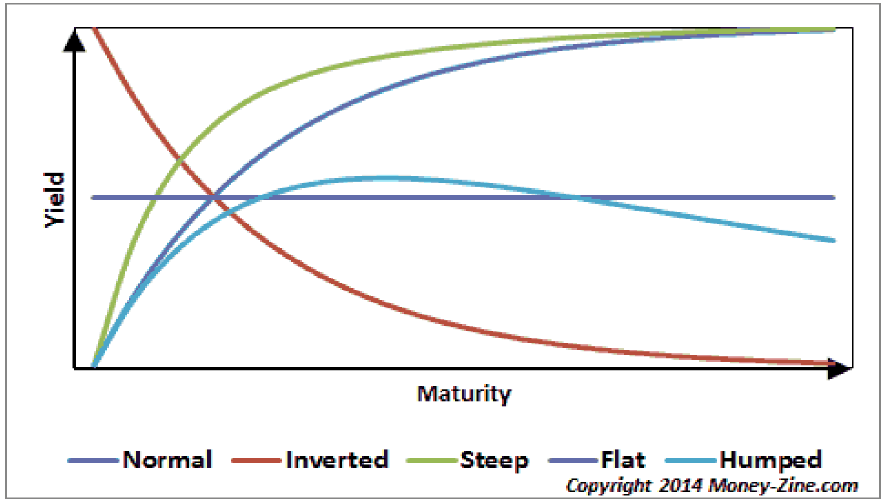
```
print(bic_n, bic_log)
```

8985.933377421756 5862.067302504404

Logistic Distribution fits better than Normal Distribution

Term Structure of Yield

- Essentially, term structure of interest rates is the relationship between interest rates or bond yields and different terms or maturities.
- When graphed, the term structure of interest rates is known as a yield curve, and it plays a crucial role in identifying the current state of an economy.
- The term structure of interest yield reflects expectations of market participants about future changes in interest rates and their assessment of monetary policy conditions.

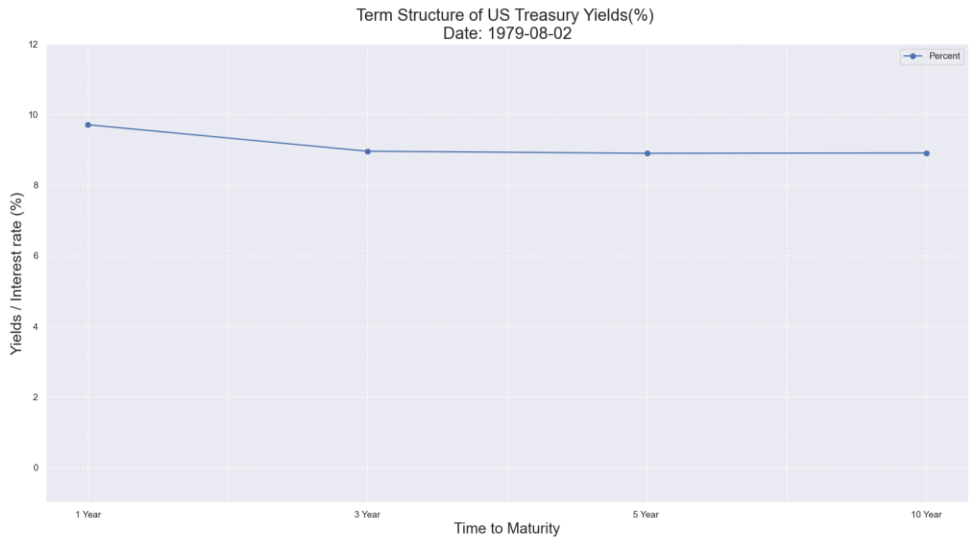


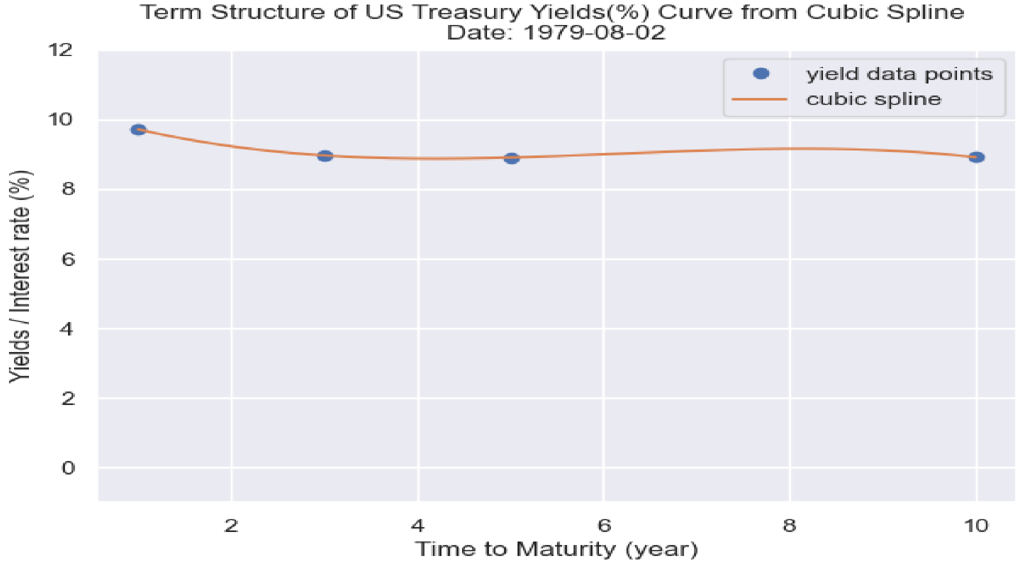
- Upward sloping/Normal—long term yields are higher than short term yields. This is considered to be the "normal" slope of the yield curve and signals that the economy is in an expansionary mode.

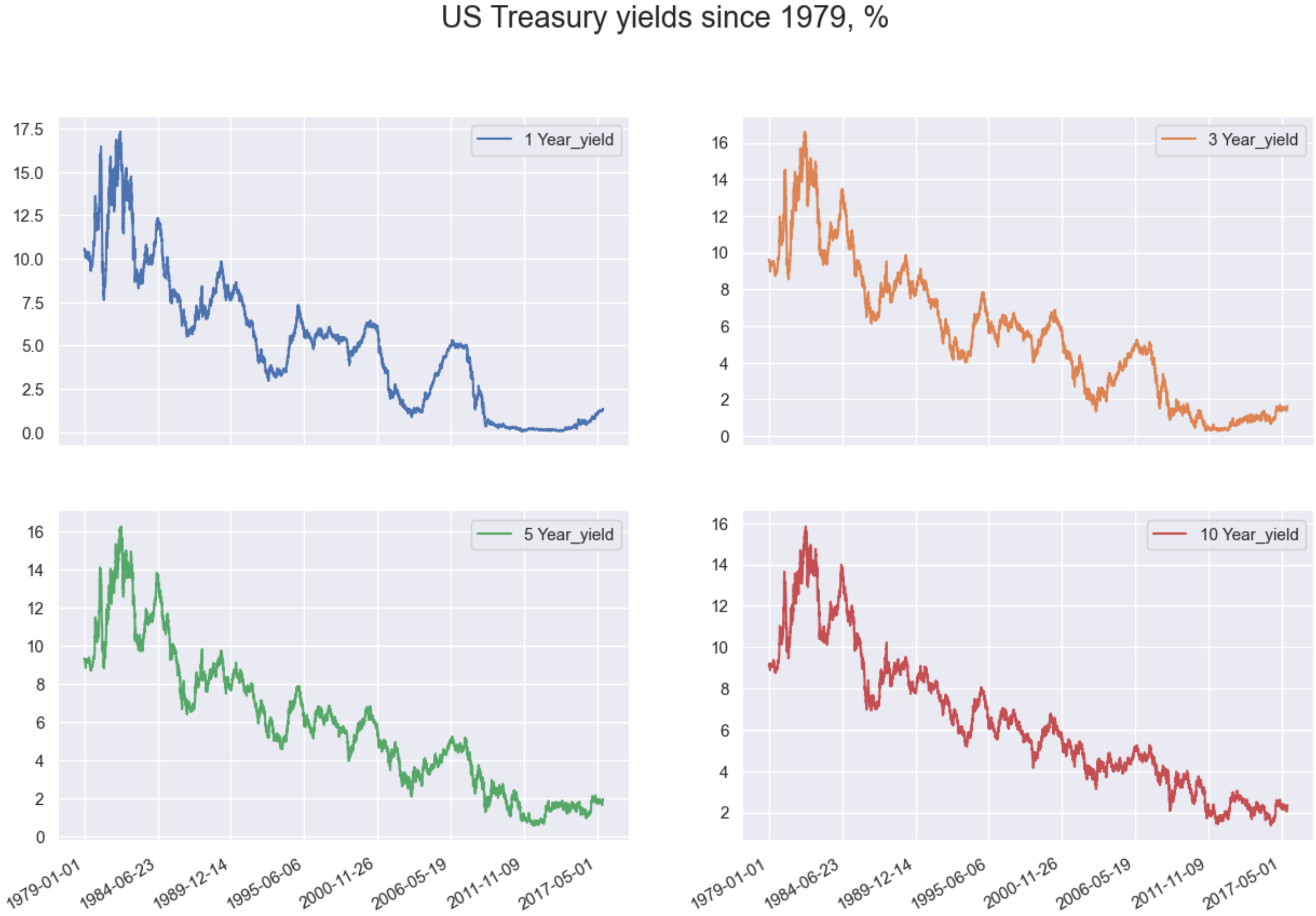
- Downward sloping/Inverted—short term yields are higher than long term yields. Dubbed as an "inverted" yield curve and signifies that the economy is in, or about to enter, a recessive period.

- Flat/Humped—very little variation between short and long term yields. Signals that the market is unsure about the future direction of the economy.

Term Structure of Yield : 1979-08-02

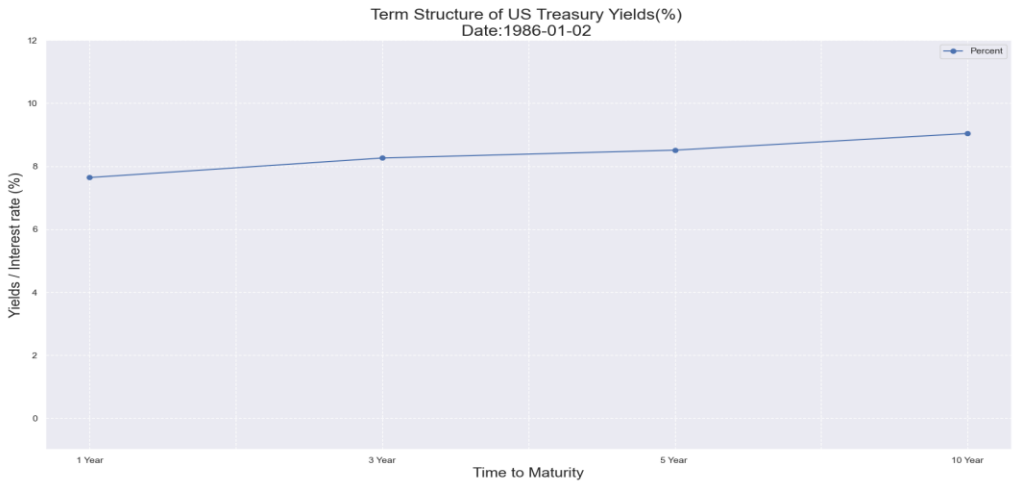


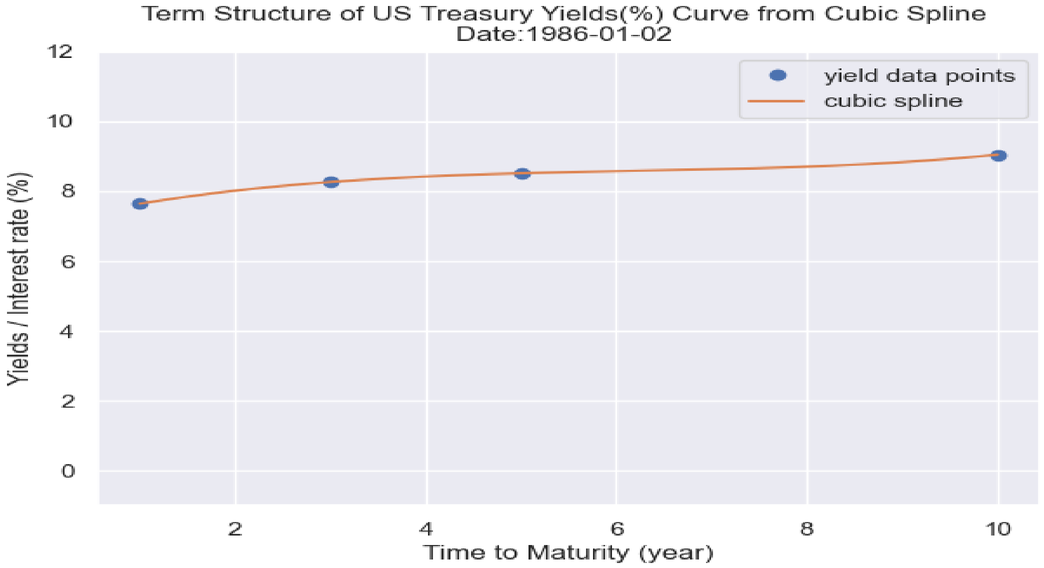


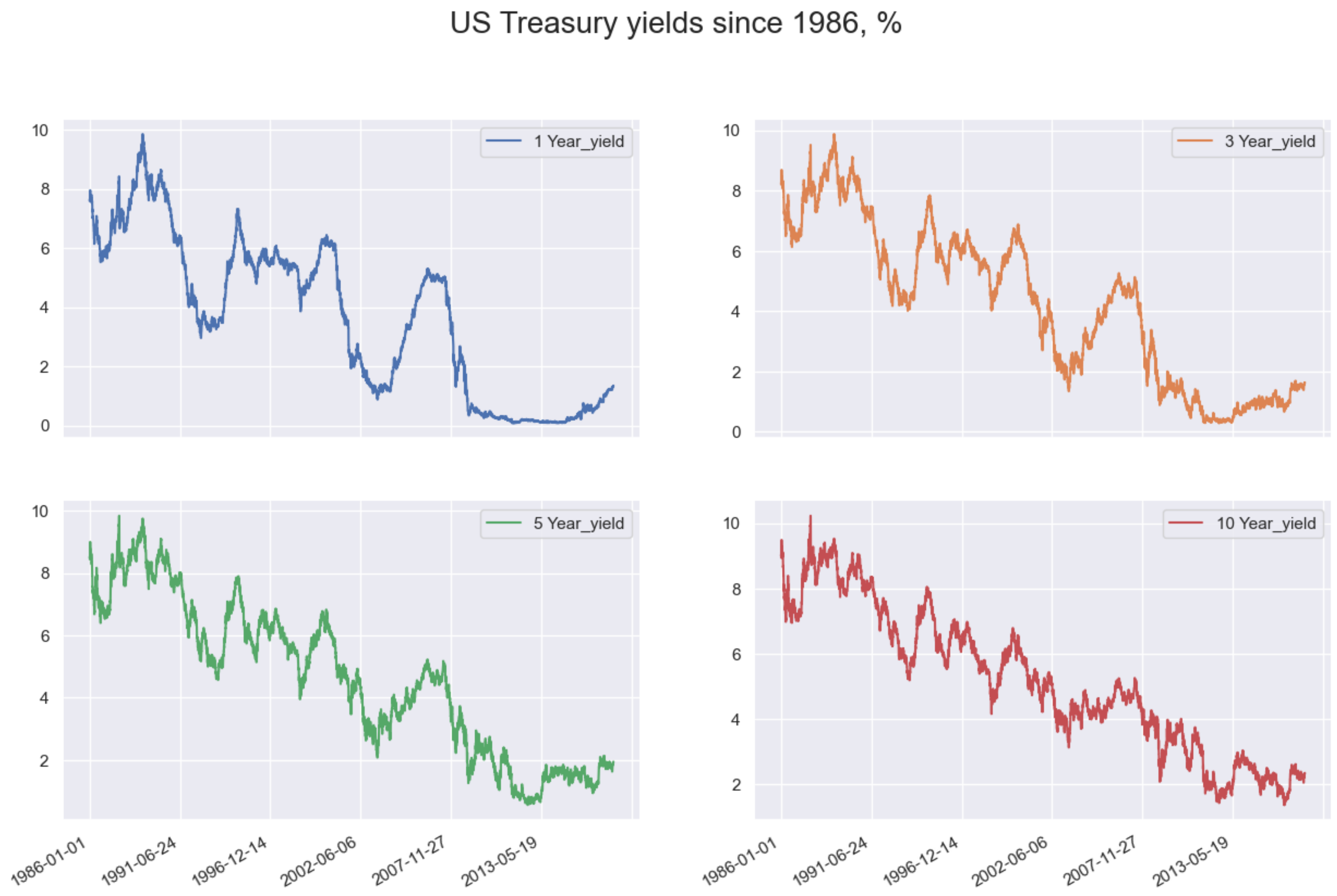


std: 0.394208

Term Structure of Yield : 1986-01-02







std: 0.58117

Standard Deviation : 1-day and 25-day increments

```
yield_data.diff(periods=1).describe()
```

	1 Year_yield	3 Year_yield	5 Year_yield	10 Year_yield
count	20365.000000	20365.000000	20365.000000	20365.000000
mean	-0.000092	-0.000102	-0.000095	-0.000084
std	0.068469	0.062521	0.060016	0.054727
min	-1.080000	-0.790000	-0.770000	-0.750000
25%	-0.010000	-0.010000	-0.010000	-0.010000
50%	0.000000	0.000000	0.000000	0.000000
75%	0.010000	0.010000	0.010000	0.010000
max	1.100000	0.920000	0.720000	0.650000

```
yield_data.diff(periods=25).describe()
```

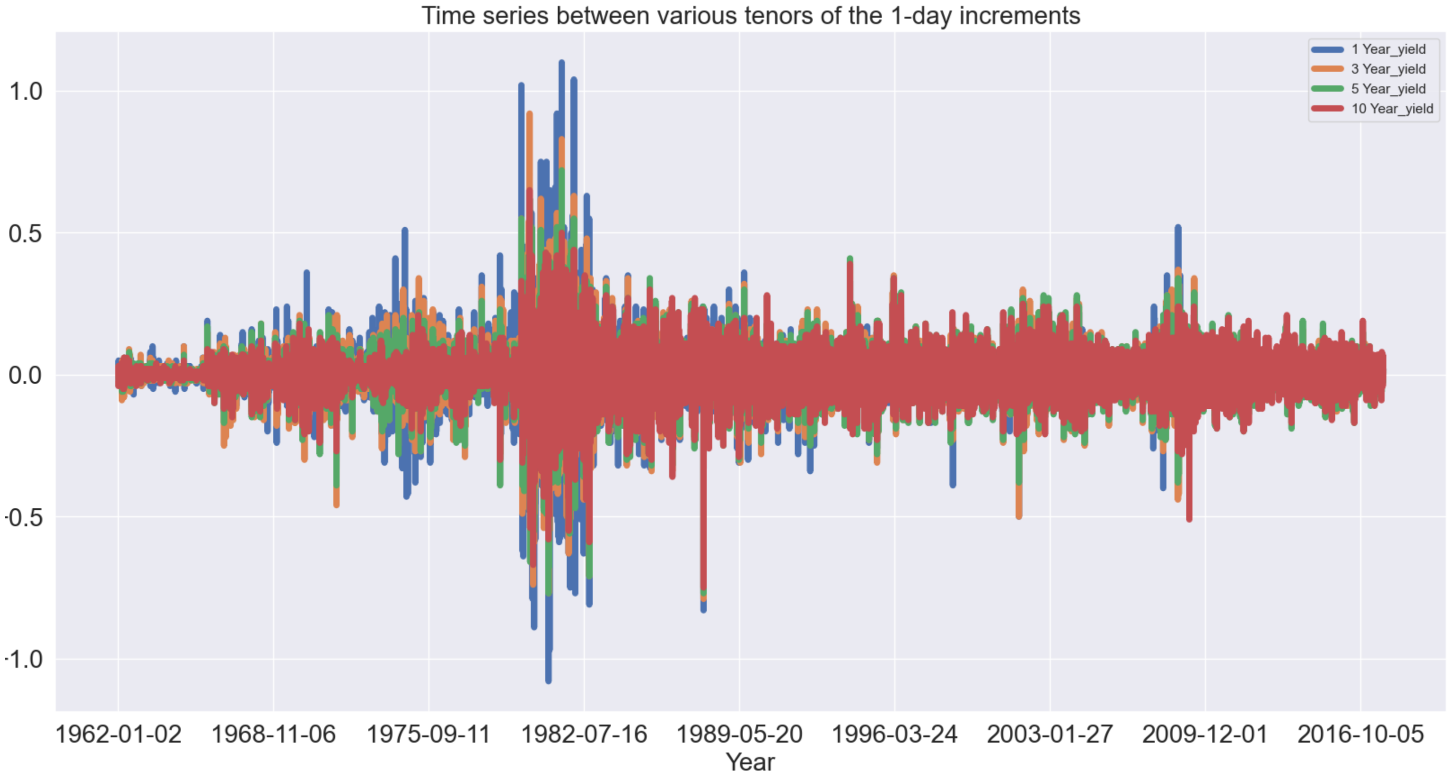
	1 Year_yield	3 Year_yield	5 Year_yield	10 Year_yield
count	20341.000000	20341.000000	20341.000000	20341.000000
mean	-0.002427	-0.002689	-0.002543	-0.002231
std	0.437109	0.374288	0.345874	0.301641
min	-4.730000	-3.330000	-2.790000	-2.460000
25%	-0.130000	-0.170000	-0.170000	-0.150000
50%	0.000000	0.000000	0.000000	0.000000
75%	0.140000	0.170000	0.170000	0.150000
max	3.170000	3.040000	2.820000	2.360000

- We can observe the probability density functions do not differ a lot in the graphs. Histograms of 1 day increment for each tenor looks similar. The reason for the similar probability density function is because the standard deviation is small and almost equal to 0.06 (for three tenor)

Correlation Matrix b/w tenors of 1 day increments

```
yield_data.diff(periods=1).corr(method ='pearson')
```

	1 Year_yield	3 Year_yield	5 Year_yield	10 Year_yield
1 Year_yield	1.000000	0.837036	0.781079	0.702266
3 Year_yield	0.837036	1.000000	0.938262	0.864023
5 Year_yield	0.781079	0.938262	1.000000	0.925168
10 Year_yield	0.702266	0.864023	0.925168	1.000000

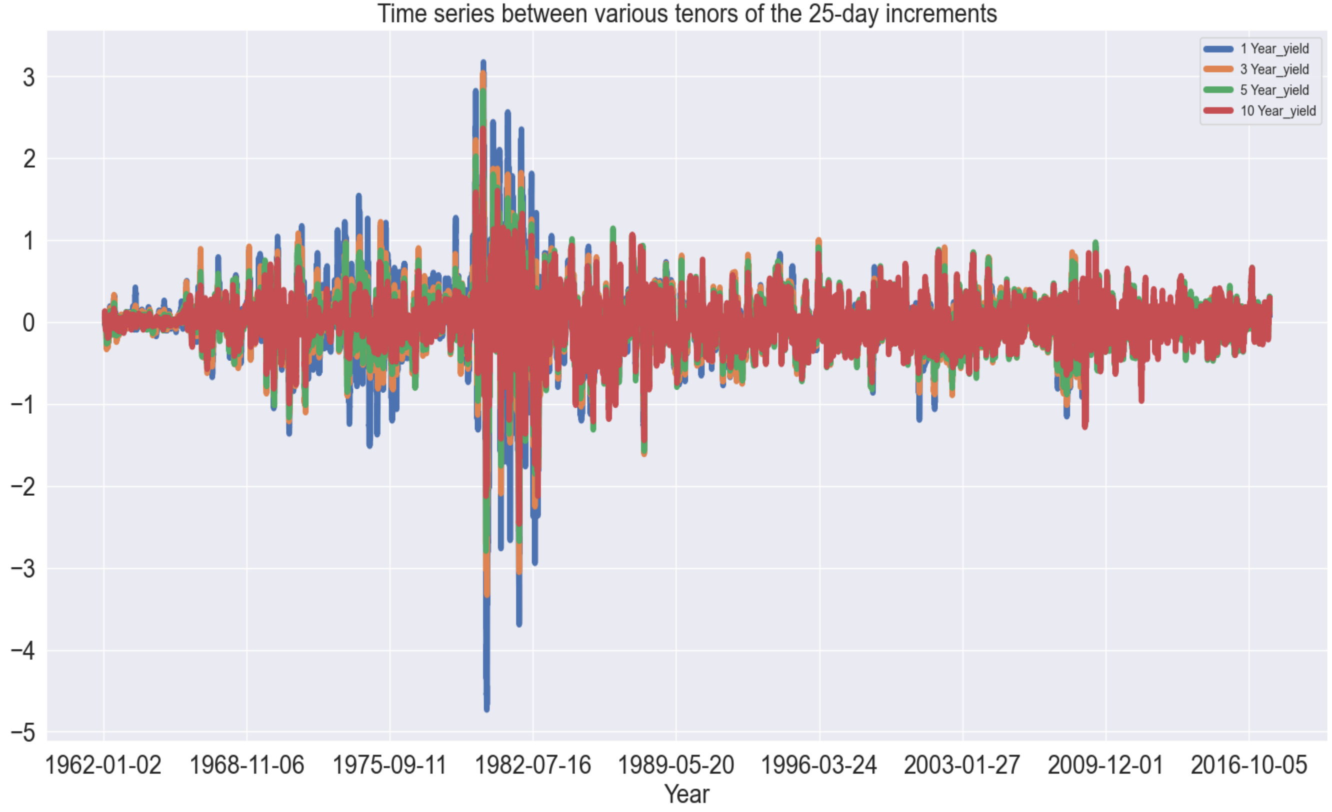


Strong Positive Correlation by pearson's method

Correlation Matrix b/w tenors of 25 day increments

```
yield_data.diff(periods=25).corr(method ='pearson')
```

	1 Year_yield	3 Year_yield	5 Year_yield	10 Year_yield
1 Year_yield	1.000000	0.915800	0.857685	0.762787
3 Year_yield	0.915800	1.000000	0.971627	0.900484
5 Year_yield	0.857685	0.971627	1.000000	0.948313
10 Year_yield	0.762787	0.900484	0.948313	1.000000



Strong Positive Correlation by pearson's method

ARIMA Model (Autoregressive Integrated Moving Average)

Stationary or not?

A **non-stationary** time series shows seasonal effects, trends, and other structures that depend on the time index.

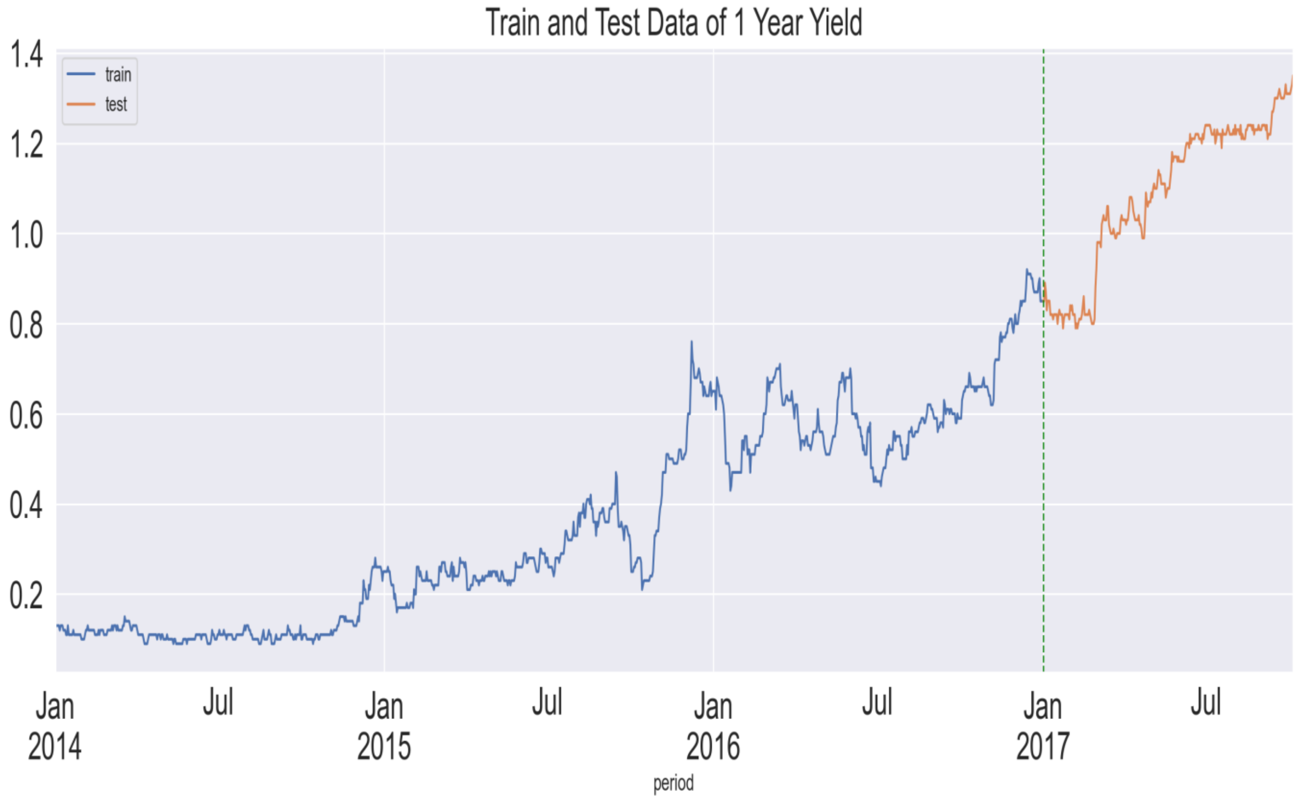
A time series is **stationary** if its statistical properties are all constant over time.

Properties of a stationary time series:

- no trend
- variations around its mean have a constant amplitude
- it wiggles in a consistent fashion (i.e., its short-term random time patterns always look the same)

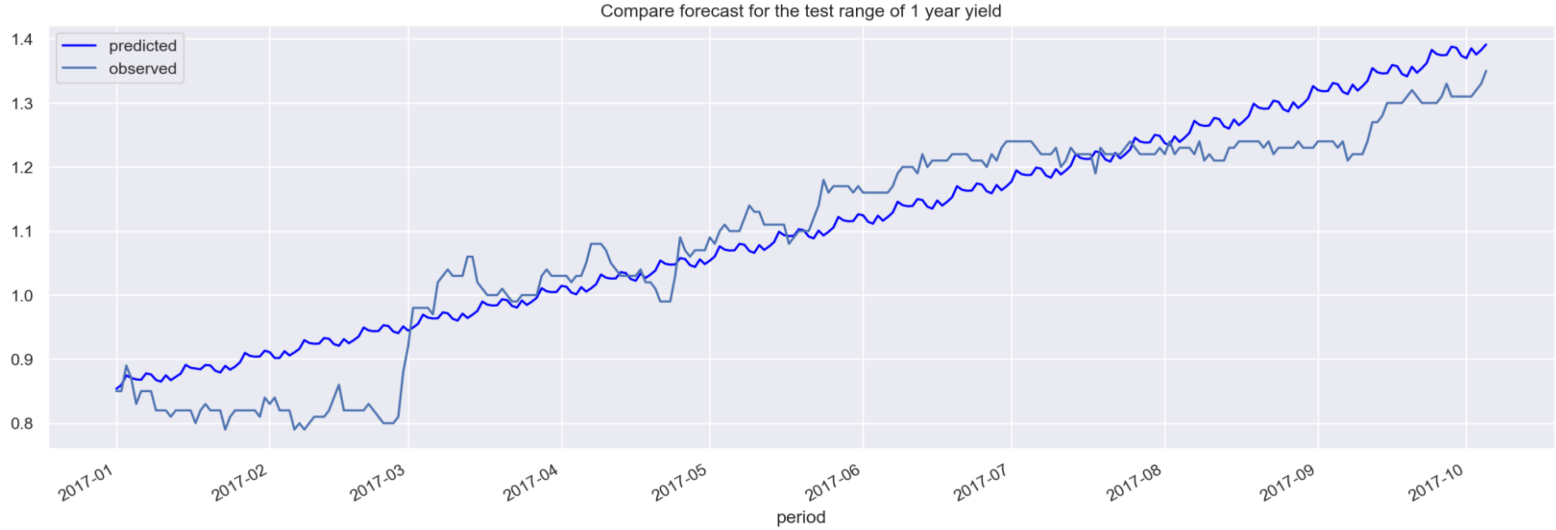
- Differencing is a method of transforming a non-stationary time series into a stationary one. This is an important step in preparing data to be used in an ARIMA model.

SARIMA Model for 1 Year Yield



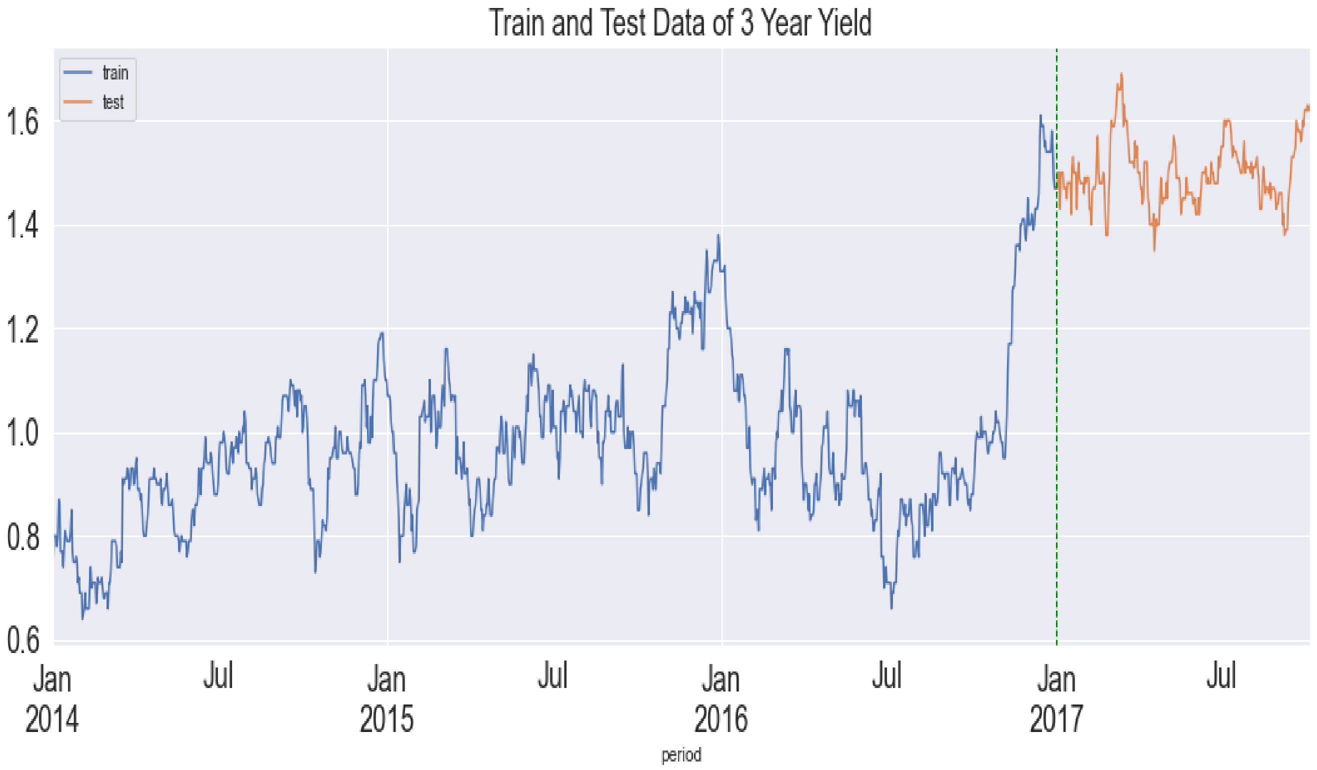
SARIMA Model for 1 Year Yield

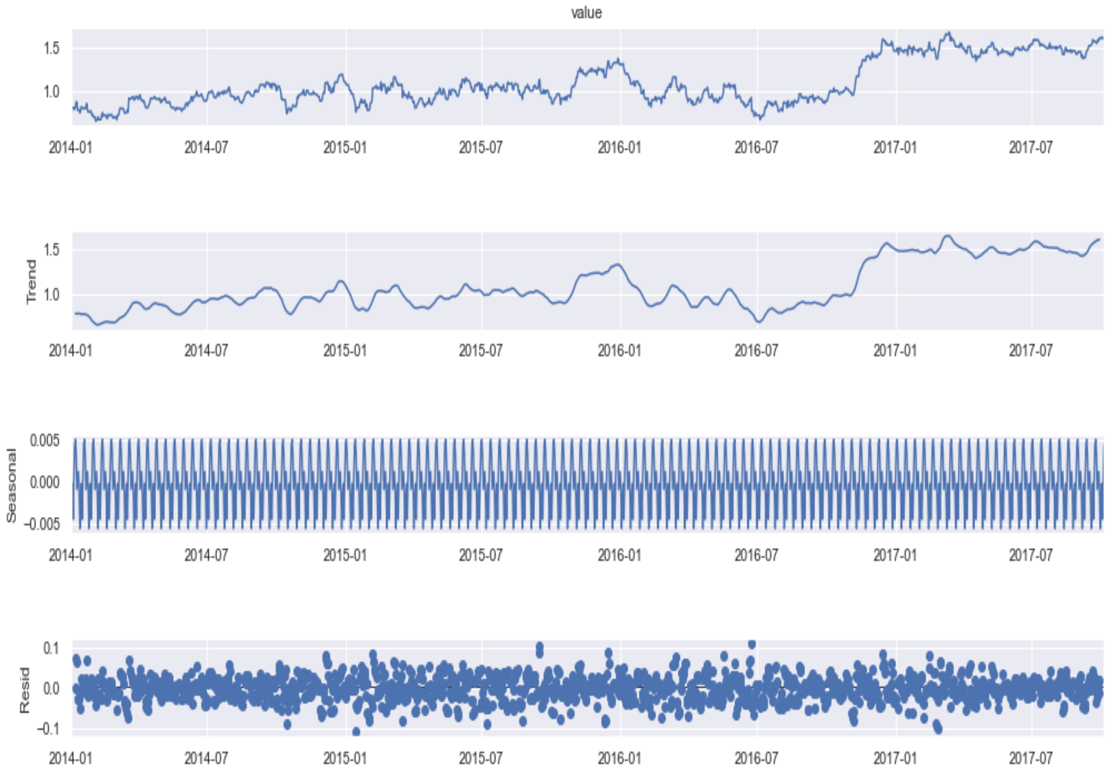
p - the number of lag observations to include in the model, or lag order. (AR)
d - the number of times that the raw observations are differenced, or the degree of differencing. (I)
q - the size of the moving average window, also called the order of moving average.(MA)
We took differences of the first order to stationaries the time series, so d=1




```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                            value   No. Observations:                  1096
Model:             SARIMAX(3, 1, 1)x(3, 1, 1, 12)   Log Likelihood                 1696.550
Date:                          Fri, 18 Dec 2020   AIC                           -3375.100
Time:                                  16:48:30   BIC                           -3330.213
Sample:                              01-01-2014   HQIC                          -3358.106
                                   - 12-31-2016
Covariance Type:                            opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.0788      2.898     -0.027      0.978      -5.759       5.601
ar.L2         -0.0116      0.307     -0.038      0.970      -0.613       0.590
ar.L3          0.0098      0.049      0.198      0.843      -0.087       0.106
ma.L1         -0.0266      2.897     -0.009      0.993      -5.705       5.652
ar.S.L12      -0.0181      0.027     -0.672      0.502      -0.071       0.035
ar.S.L24       0.0049      0.038      0.129      0.897      -0.070       0.080
ar.S.L36      -0.0505      0.030     -1.673      0.094      -0.110       0.009
ma.S.L12      -0.9931      0.059    -16.970      0.000      -1.108      -0.878
sigma2         0.0024      0.000     18.086      0.000       0.002       0.003
===================================================================================
Ljung-Box (Q):                       55.48   Jarque-Bera (JB):               807.90
Prob(Q):                              0.05   Prob(JB):                         0.00
Heteroskedasticity (H):               0.27   Skew:                             0.13
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.22
===================================================================================
```

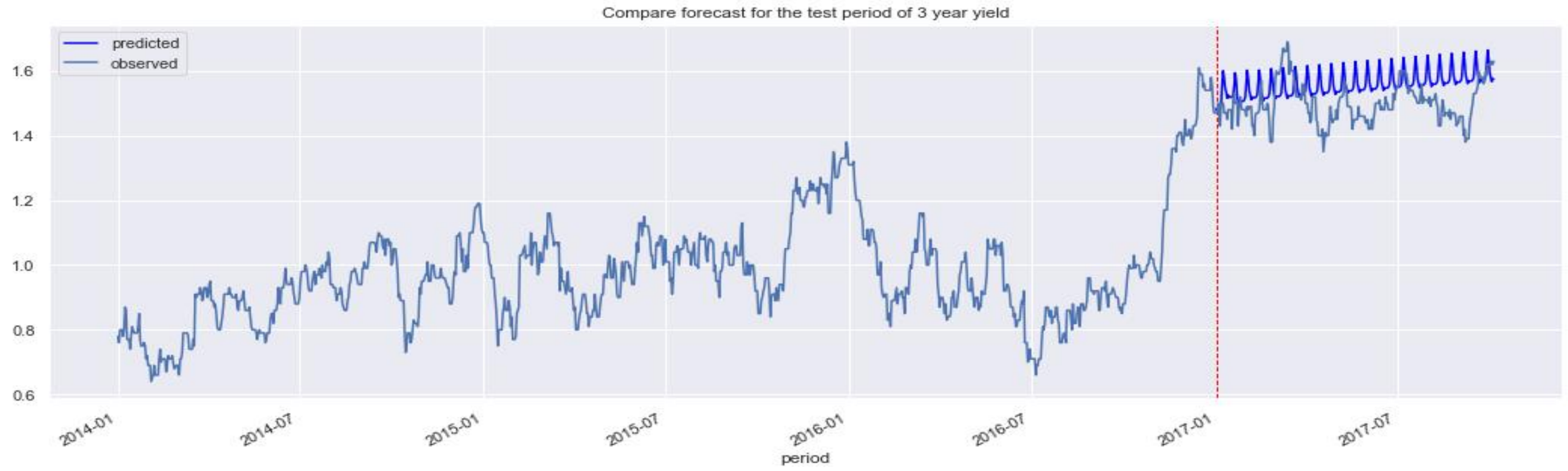
SARIMA Model for 3 Year Yield

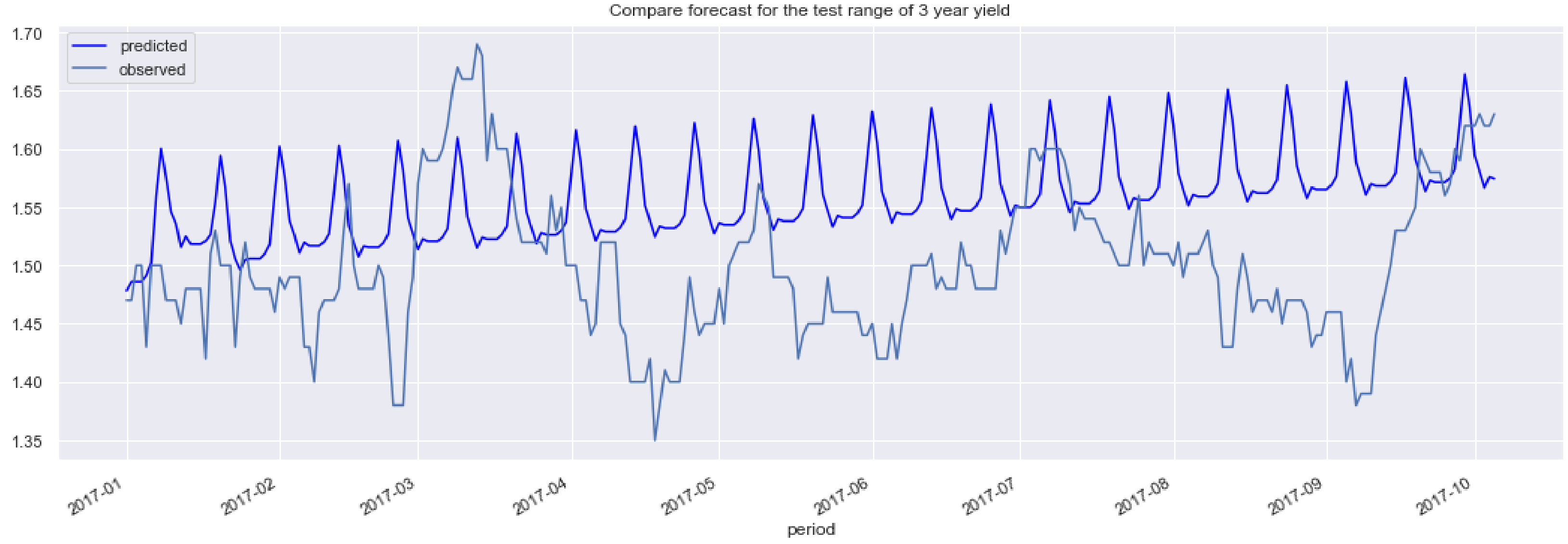




SARIMA Model for 3 Year Yield

p - the number of lag observations to include in the model, or lag order. (AR)
d - the number of times that the raw observations are differenced, or the degree of differencing. (I)
q - the size of the moving average window, also called the order of moving average.(MA)
We took differences of the first order to stationaries the time series, so d=1






```
                                  SARIMAX Results
==========================================================================================
Dep. Variable:                          value   No. Observations:                    1096
Model:           SARIMAX(1, 1, 1)x(1, 1, [], 12)   Log Likelihood                  1936.352
Date:                        Fri, 18 Dec 2020   AIC                             -3864.705
Time:                                17:03:15   BIC                             -3844.755
Sample:                            01-01-2014   HQIC                            -3857.152
                                 - 12-31-2016
Covariance Type:                          opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.2264      0.570     -0.397      0.691      -1.344       0.891
ma.L1          0.1722      0.575      0.300      0.764      -0.954       1.299
ar.S.L12      -0.5218      0.021    -25.113      0.000      -0.562      -0.481
sigma2         0.0016   5.15e-05     31.710      0.000       0.002       0.002
===================================================================================
Ljung-Box (Q):                      173.63   Jarque-Bera (JB):               173.82
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               1.13   Skew:                            -0.19
Prob(H) (two-sided):                  0.26   Kurtosis:                         4.93
===================================================================================
```

Conclusion

- During this project got an idea about US treasury yields for multiple tenors.
- In python we are mainly using six libraries- pandas, numpy, seaborn, matplotlib.pyplot, scipy, statsmodels.
- Logistic Distribution fits better than Normal Distribution for all the increments
- Identified that there are correlations between 1 day and 25 day increment for multiple tenors by Pearson's method.
- SARIMA Stochastic Model is implemented to predict Yields.
- Future Scope: To implement Stochastic model like Geometric Brownian motion for predicting yields.

LANTERN INSTITUTE