Statistical Data Science (Module 3)

Stack Overflow Questions Count Time Series

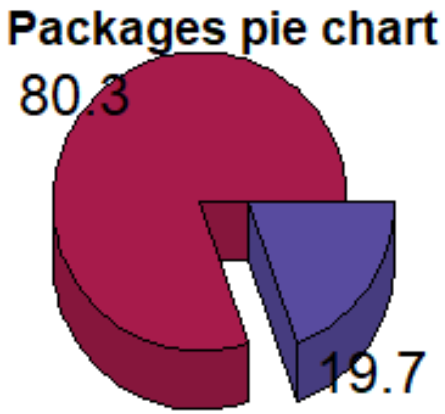
Presented by:

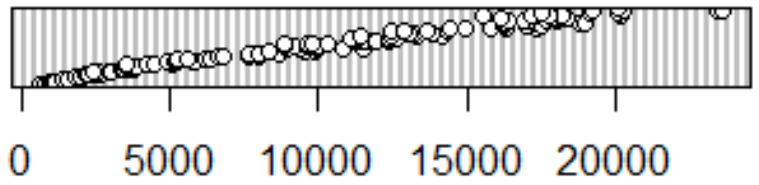
Shakshi Sharma

Dataset

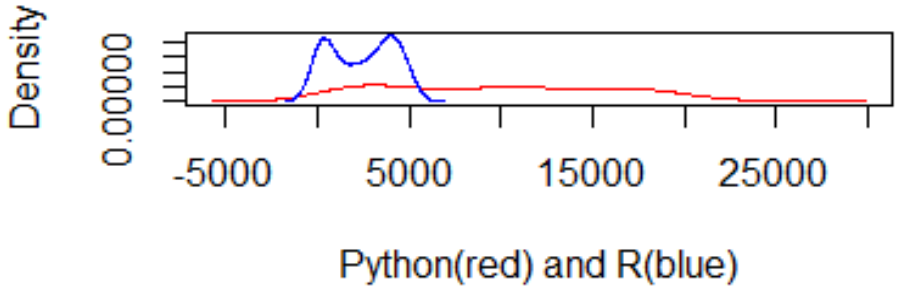
- Kaggle dataset: https://www.kaggle.com/datasets/aishu200023/stackindex?resource=download
  - Consist of various columns containing names of libraries and the rows having the count of questions in StackOverflow for each month from 2009.
- Our problem: To forecast the number of Stackoverflow questions ask in the future for python and r packages.

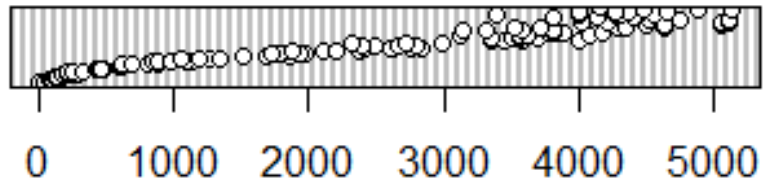
Descriptive Analysis





Python





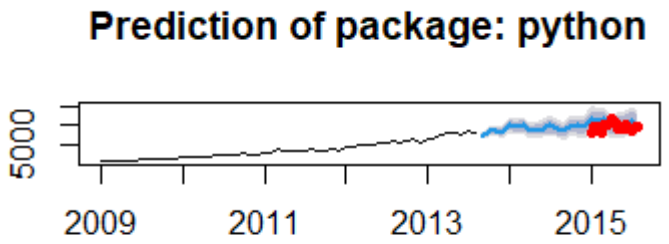
R

Statistical Methodology

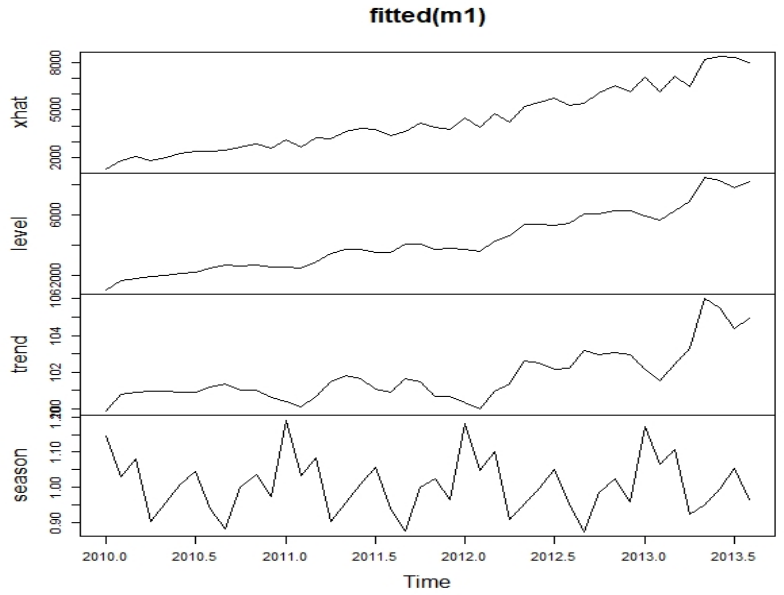
- To solve this time-series problem, will fit a Holt-Winters model.

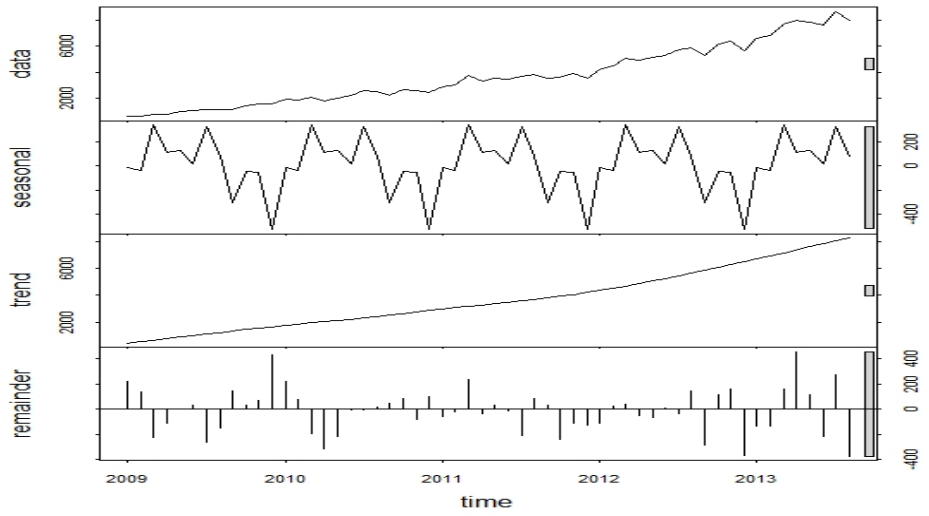
- Outcome: report the seasonality, trends for the two packages, that is, Python and R.

Results: Python

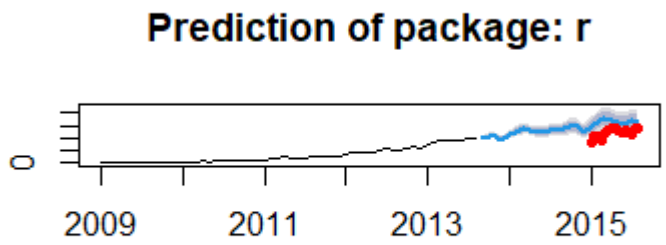



```
Call:
HoltWinters(x = y, seasonal = "mult")

Smoothing parameters:
 alpha: 0.8854003
 beta : 0.001814823
 gamma: 1

Coefficients:
            [,1]
a   8267.7356376
b    104.8894245
s1     0.8695709
s2     0.9867737
s3     1.0237435
s4     0.9504217
s5     1.1634334
s6     1.0774377
s7     1.1171790
s8     0.9427857
s9     0.9468078
s10    0.9827465
s11    1.0577661
s12    0.9644721
```

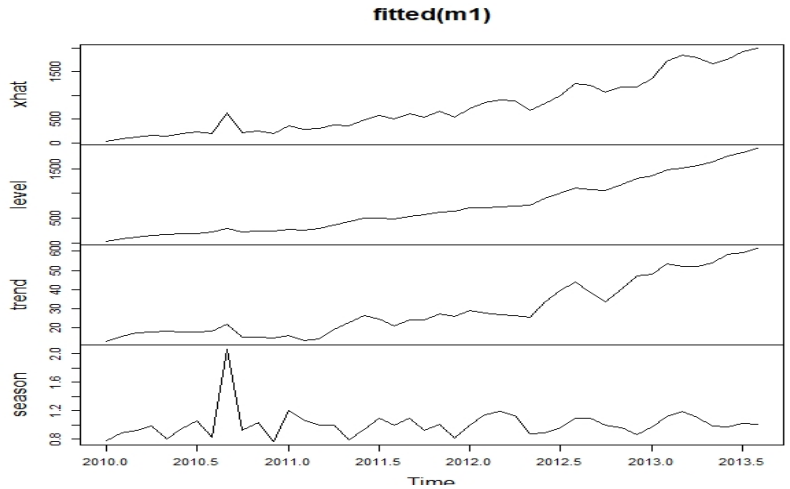


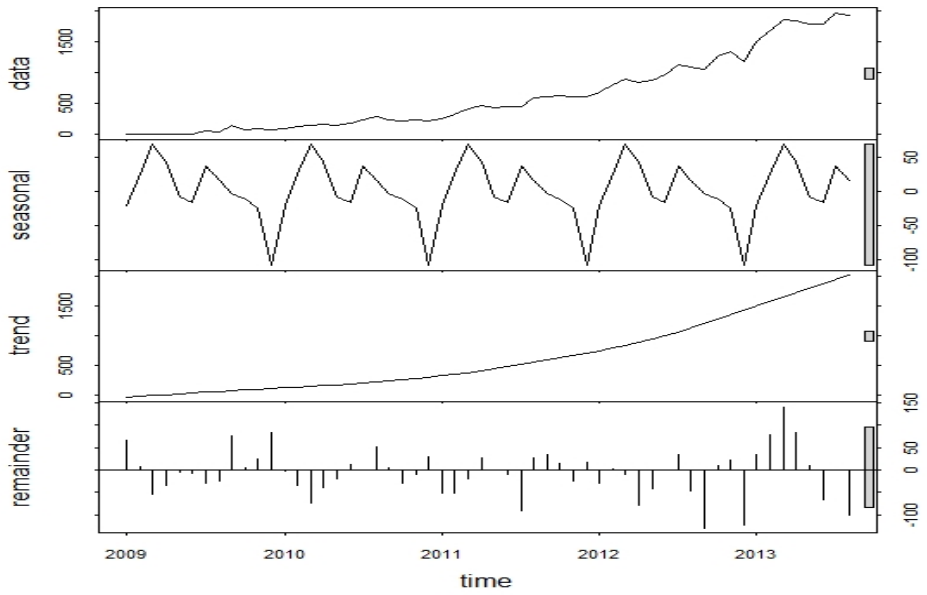


Results: R




```
Smoothing parameters:
 alpha: 0.4632221
 beta : 0.07699675
 gamma: 1

Coefficients:
              [,1]
a    1939.9169109
b      58.9418926
s1      1.0102317
s2      1.0812291
s3      1.0363733
s4      0.8777021
s5      1.0245686
s6      1.1073909
s7      1.1859390
s8      1.1259998
s9      1.0242875
s10     0.9790766
s11     1.0370620
s12     0.9948880
```





Thank you!