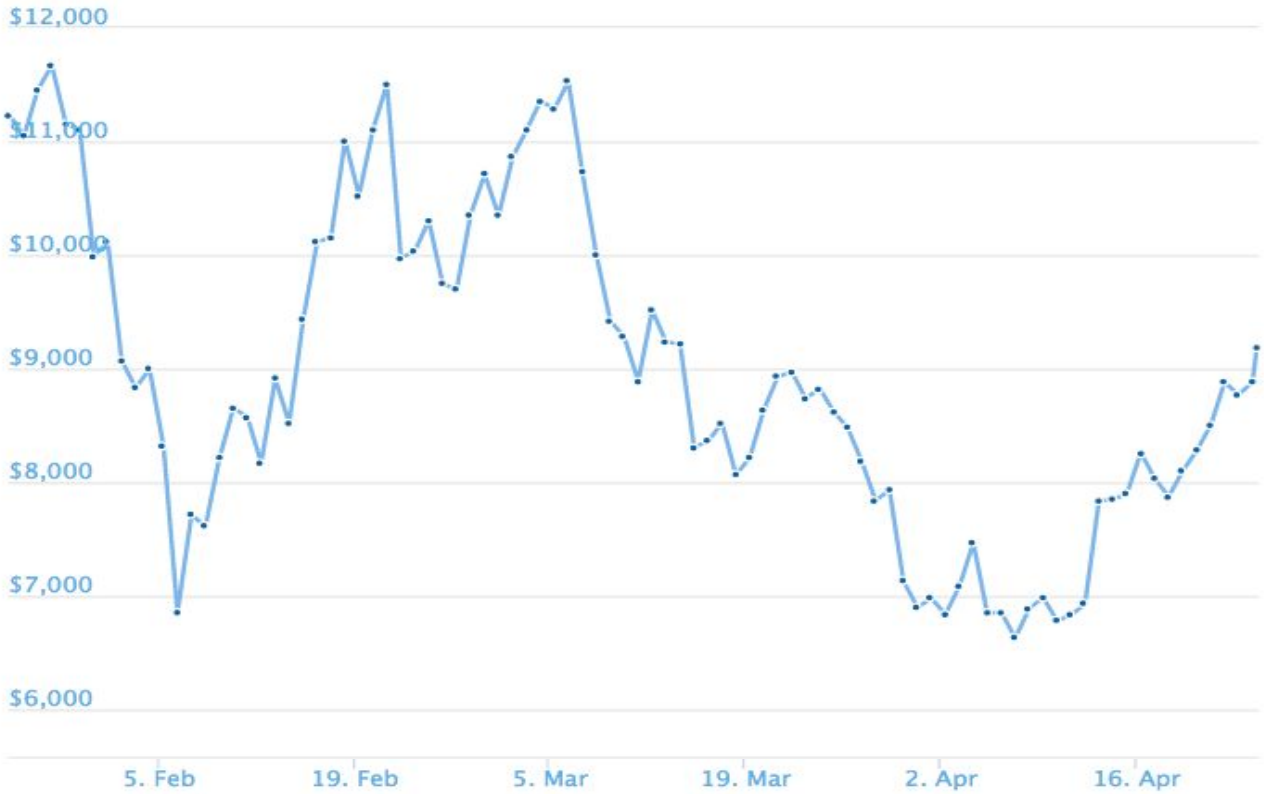
Time Series Intro

Week 09 - Day 01

What is a Time Series?


$12,000
$11,000
$10,000
$9,000
$8,000
$7,000
$6,000
5. Feb
19. Feb
5. Mar
19. Mar
2. Apr
16. Apr

What is x?

What is y?

Time series = time + numerical values

Time = **regular time**

(e.g. days, months, seconds, etc.)

Main Goals

Identify Patterns

Forecast

Time Series
Basic operation

Analyse the values

(mean, median, max, distribution, etc.)

Group by day/week/month/etc.

+

Sum, mean, median, max, min, etc.

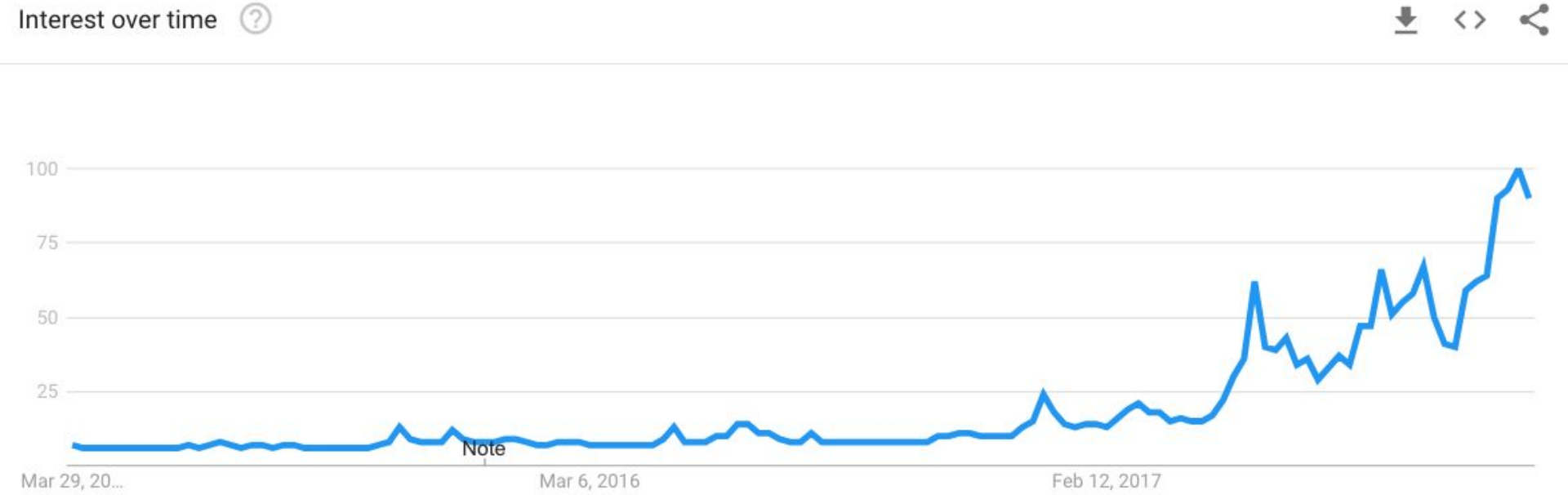
Moving average

(effect?)

Time Series Components

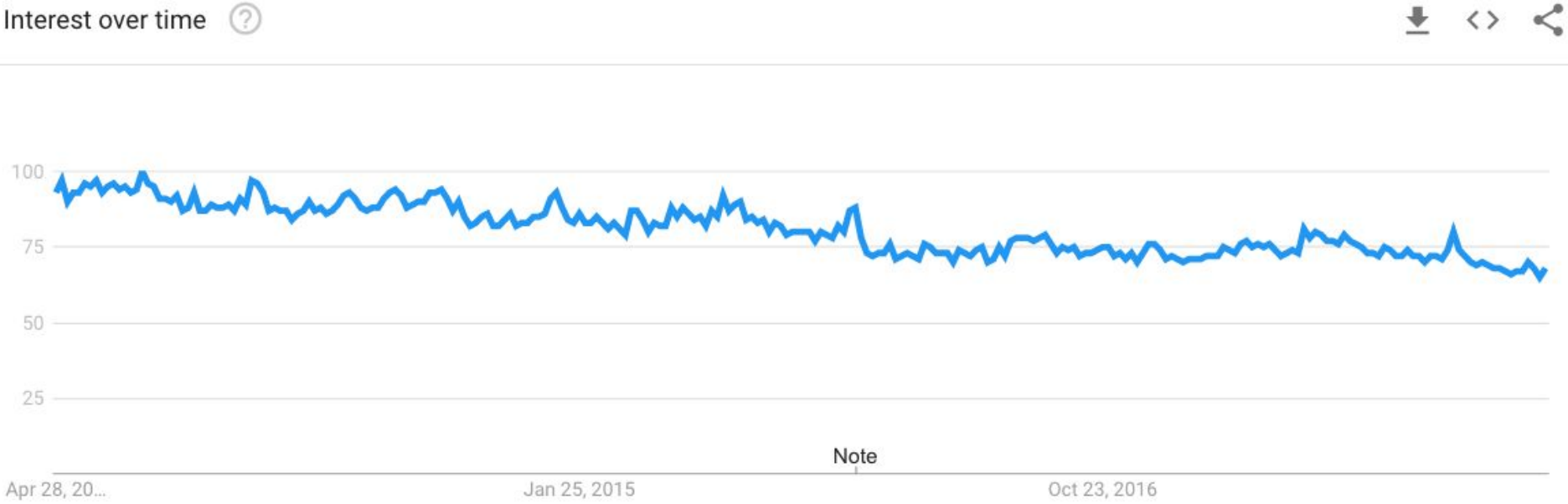
Trend

Bitcoin


Interest over time
100
75
50
25
Note
Mar 29, 20...
Mar 6, 2016
Feb 12, 2017

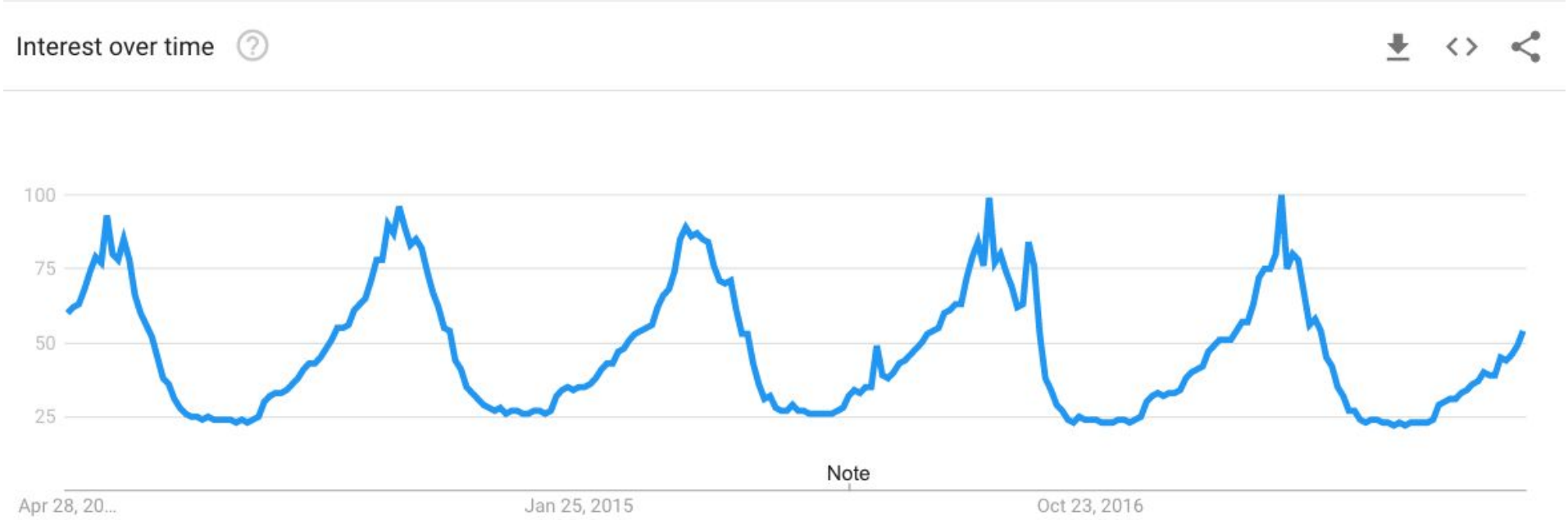
Trend can go up and down

Porn


Interest over time
100
75
50
25
Note
Apr 28, 20...
Jan 25, 2015
Oct 23, 2016

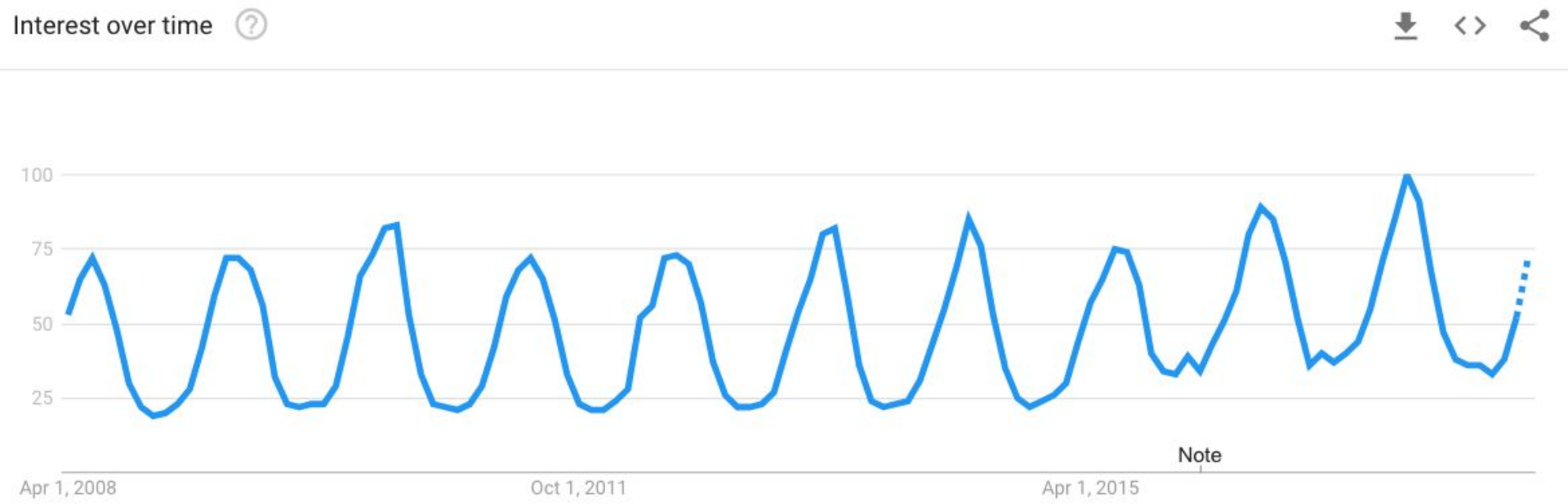
Seasonality

Flip Flops - 2y

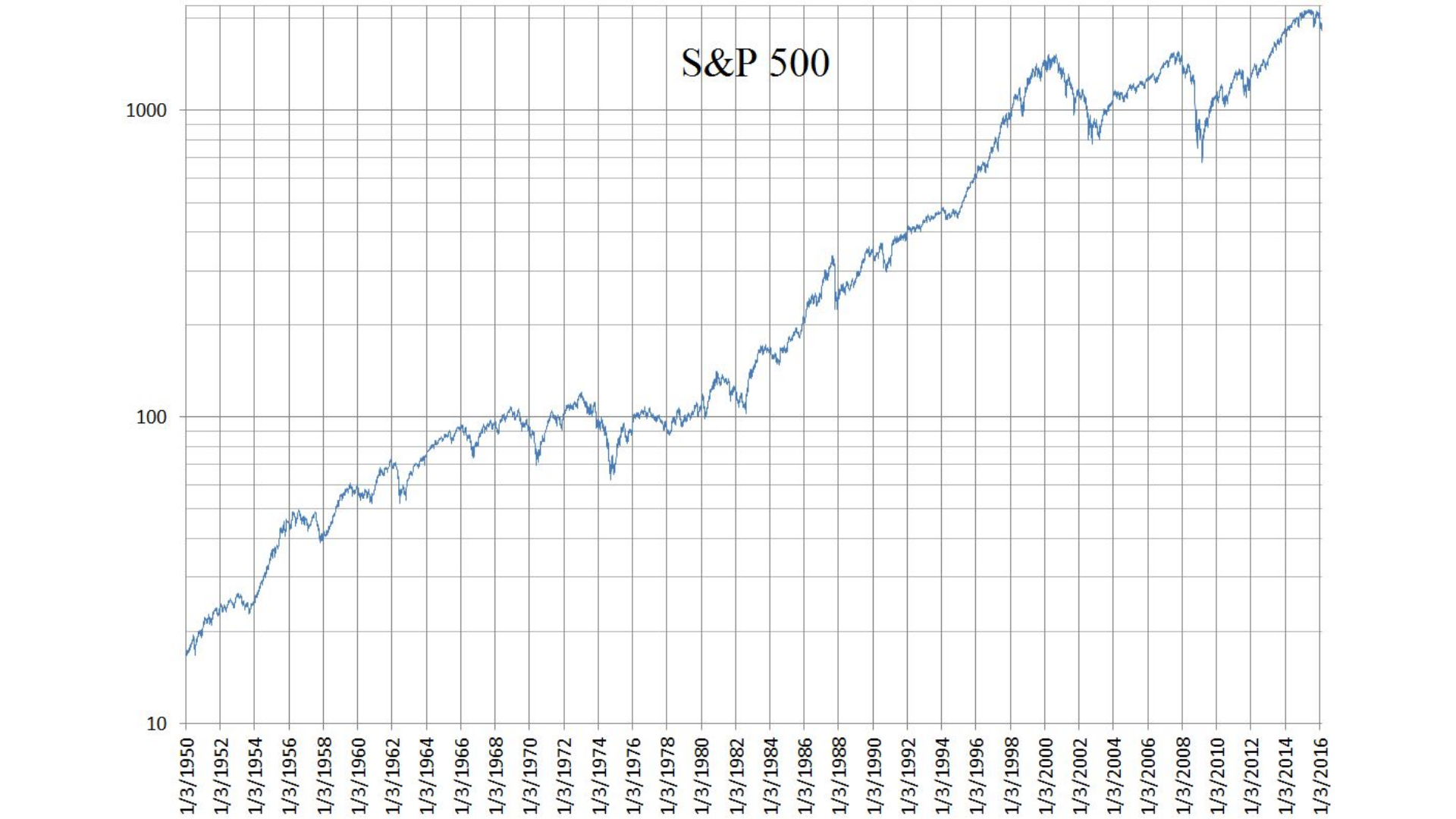


Seasonality + Trend

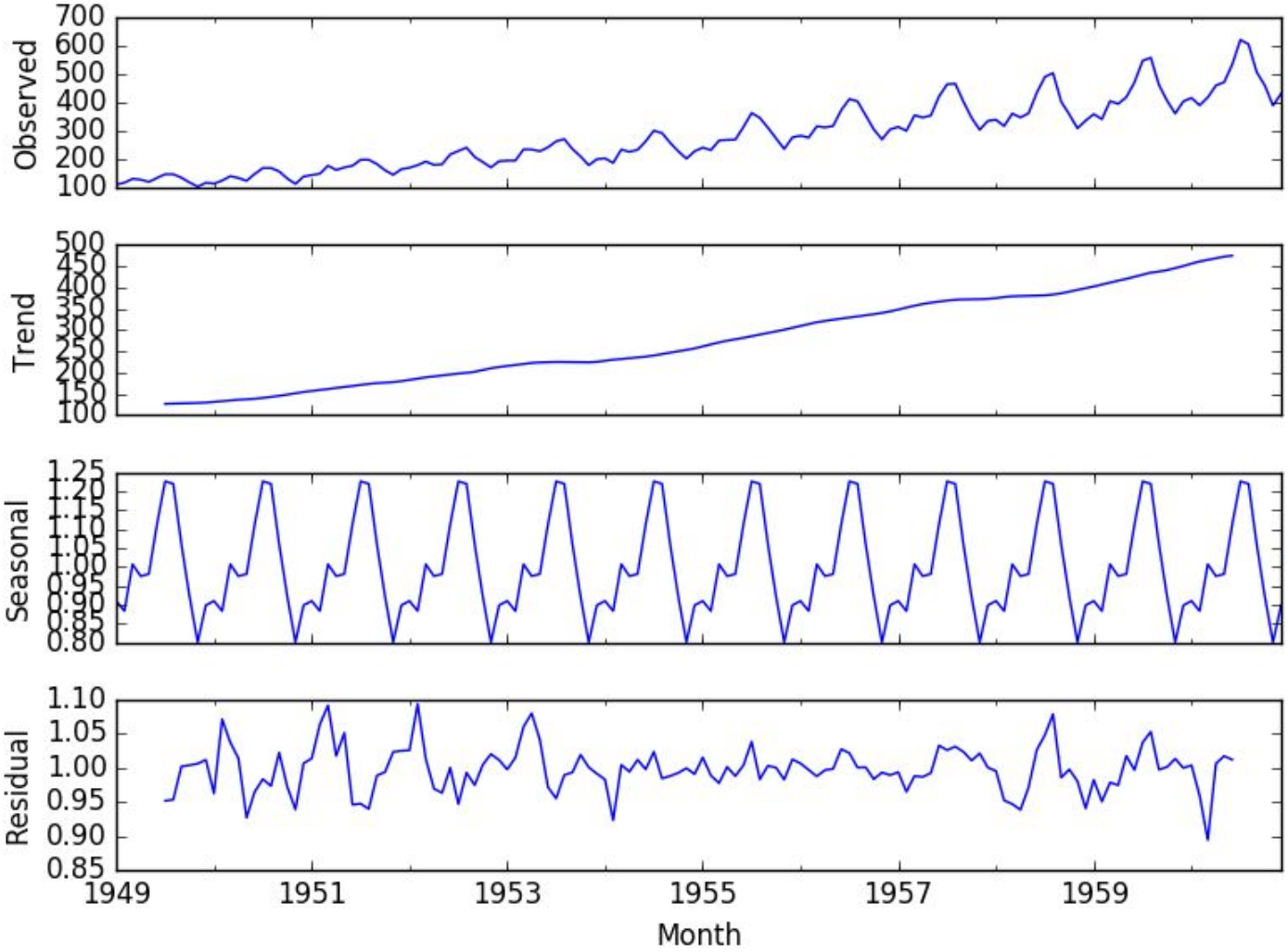
Flip Flops - 10 y



Cyclical “Non-Regular” Events


S&P 500
1000
100
10
1/3/1950
1/3/1952
1/3/1954
1/3/1956
1/3/1958
1/3/1960
1/3/1962
1/3/1964
1/3/1966
1/3/1968
1/3/1970
1/3/1972
1/3/1974
1/3/1976
1/3/1978
1/3/1980
1/3/1982
1/3/1984
1/3/1986
1/3/1988
1/3/1990
1/3/1992
1/3/1994
1/3/1996
1/3/1998
1/3/2000
1/3/2002
1/3/2004
1/3/2006
1/3/2008
1/3/2010
1/3/2012
1/3/2014
1/3/2016

Standard decomposition


Observed
700
600
500
400
300
200
100
Trend
500
450
400
350
300
250
200
150
100
Seasonal
1.25
1.20
0.80
Residual
1.10
1.05
1.00
0.95
0.90
0.85
1949
1951
1953
1955
1957
1959
Month

What's the difference between Time Series and Regression?

Regression = independent values

(e.g. each point = one customer)

TS = related values

(e.g. each point = oil price in a given day)

Time Series in Python

No Sklearn!

No grid search!

No cross-val score!

Manual approach

for test/train validation!

R is better than python

Pandas + Statsmodels

Pandas

=

DateTime index + values column

```
df['measure'].plot()
```

`datetime` library

Less documentation compared to regression, classification, clustering, etc.

Summary

1. Time Series = time+values
2. Analysis + forecasting
3. Ts = trend + seasonality + residuals
4. No sklearn, yes pandas