TIME SERIES PREPROCESSING WITH THE **SUIT** METHOD - CHEAT SHEET

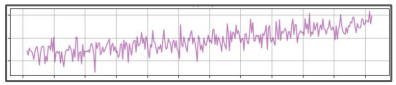
@pazpazthecoder, github.com/hip023/suit

November 2022

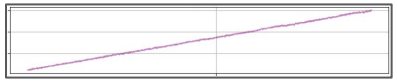
SAMPLE

Just-right frequency

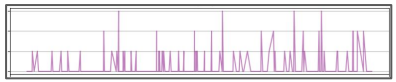
1. Delta



2. Cumulative



3. Events



PHIDS
Plot, Head, Info, Describe, Set Index

As Freq
Only for cumulative data

```
df[col].asfreq(freq='1D')
```

Resample
Events -> Sampling

```
df.resample('1T').agg(..)
```

aligner reducer

UNIVARIATE

Vanilla Feature Engineering

```
df.resample('1D').aggregate(...)
df.resample('1D').apply(lambda x: …)
def get_lambda(x):
    return …
data.resample('1D').apply(get_lambda)
```

Datetime Features

```
resampler.index.isocalendar()
resampler.index.month
```

ISOLATE

Combine resampling with other features!

```
groupers = [pd.Grouper(freq='1D'),col]
by_date_and_col = df.groupby(groupers)...
```

aligner reducer

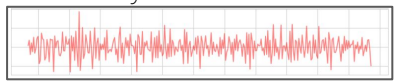
We can manipulate different samplers together!

```
by_date = df.groupby(pd.Grouper(freq='1D'))...
(by_date_and_col / by_date).unstack()
```

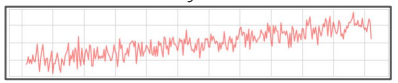
TRANSFORM

Augmented Dickey-Fuller test

stationary



non-stationary



```
from statsmodels.tsa.
stattools import adfuller
```

```
adf = adfuller(data['x'])
if adf[0] > 0.05:
print("data is not
                stationary")
```

Transformation Techniques
"Features of Features"
lags / diffs
second derivative
combined transformation
…

DataNights