Window functions in Tableau

TIME SERIES ANALYSIS IN TABLEAU

Chris Hui
VP of Product, Tracked

datacamp

What is a window?

- Windows refer to specific partitions you want to analyze
- Used specifically as an argument for window functions
- Window size is determined based off the start window and end window
- Provided no start or end window is supplied, the entire measures range of values is treated as one window

Let's select a **2 hour window** below.

Hour	Value
12 AM	100
1 AM	200
2 AM	300
3 AM	400
4 AM	500
5 AM	600
6 AM	700

The values from **12 AM – 1 AM** would make up your first window.

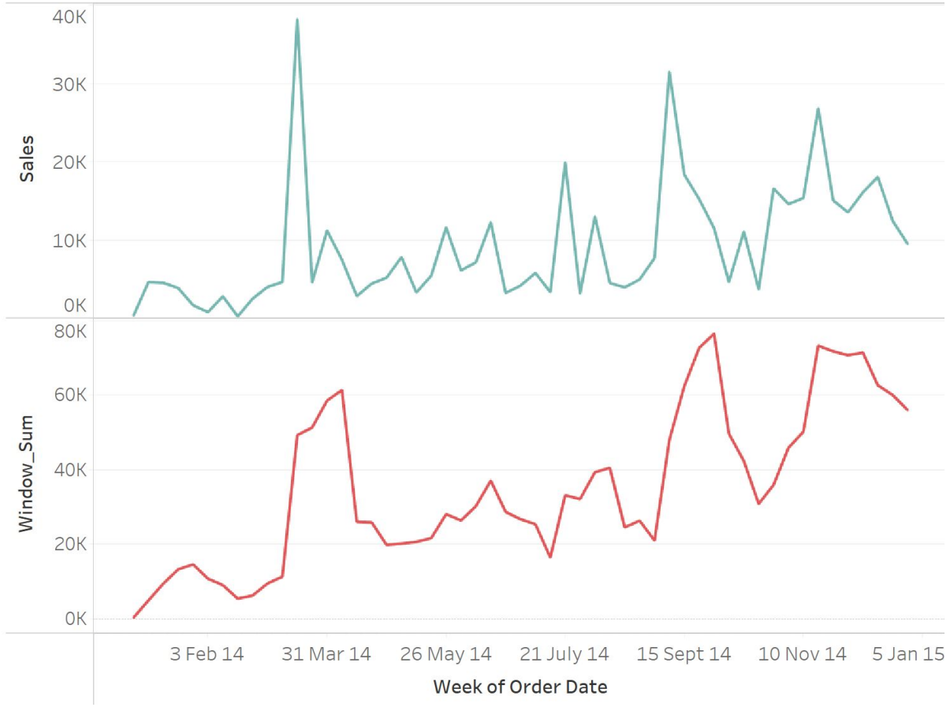
Hour	Value
12 AM	100
1 AM	200
2 AM	300
3 AM	400
4 AM	500
5 AM	600
6 AM	700

If we don't explicitly state the size of our windows, it would contain **all the values from 100 – 700.**

What is a window function?

Known as moving calculations that smooth data over specified time windows

- Examples include: window sums, moving averages and moving standard deviations
- Composed of a measure, start size, and end size



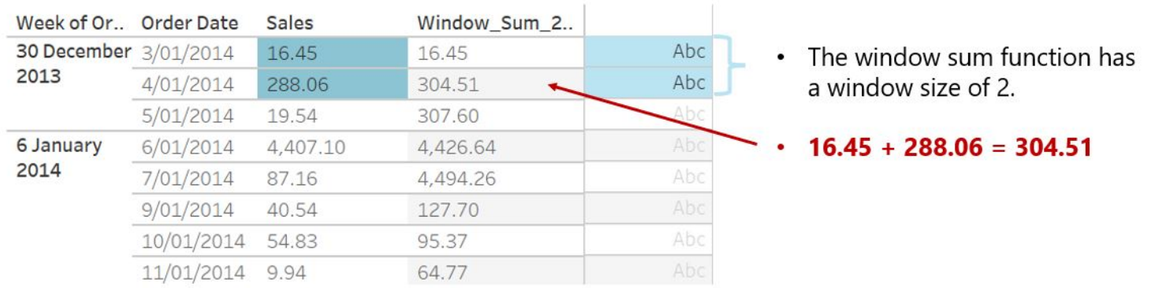
Window_Function(Measure, Window_Start, Window_End)

1 2 3

Measure – This is the respective measure you want to apply your Window Function against.
Window_Start – This is the **start** of your Window function, with **respect to your granularity**.
Window_End – This is the **end** of your Window function, with **respect to your granularity.**

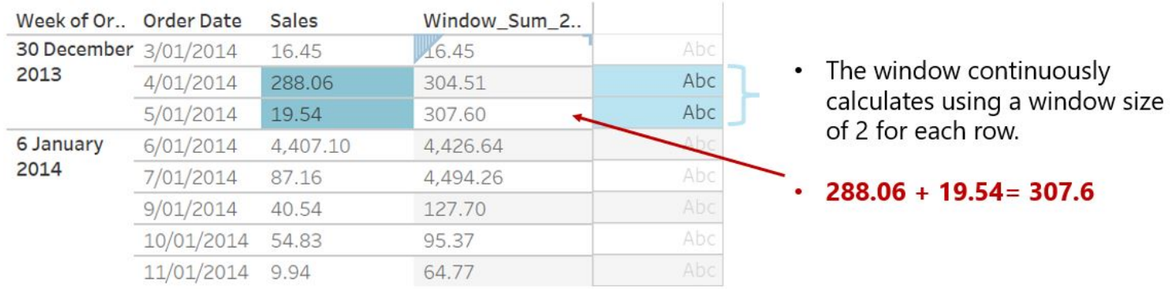
How does window sizing work?

- Sizing of your window refers to the number of data points to be included
- Once the window sizing is set, it applies this for each subsequent row wise calculation



Week of Or..	Order Date	Sales	Window_Sum_2..	
30 December 2013	3/01/2014	16.45	16.45	Abc
	4/01/2014	288.06	304.51	Abc
	5/01/2014	19.54	307.60	Abc
6 January 2014	6/01/2014	4,407.10	4,426.64	Abc
	7/01/2014	87.16	4,494.26	Abc
	9/01/2014	40.54	127.70	Abc
	10/01/2014	54.83	95.37	Abc
	11/01/2014	9.94	64.77	Abc

- The window sum function has a window size of 2.
- **16.45 + 288.06 = 304.51**

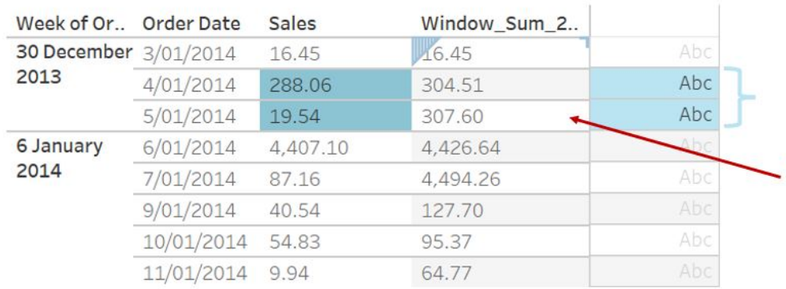


Week of Or..	Order Date	Sales	Window_Sum_2..	
30 December 2013	3/01/2014	16.45	16.45	Abc
	4/01/2014	288.06	304.51	Abc
	5/01/2014	19.54	307.60	Abc
6 January 2014	6/01/2014	4,407.10	4,426.64	Abc
	7/01/2014	87.16	4,494.26	Abc
	9/01/2014	40.54	127.70	Abc
	10/01/2014	54.83	95.37	Abc
	11/01/2014	9.94	64.77	Abc

- The window continuously calculates using a window size of 2 for each row.
- **288.06 + 19.54= 307.6**

How does granularity impact window sizing?

- Granularity directly impacts the result of window aggregation function
- Aggregations are computed first, **before** window size



Week of Or..	Order Date	Sales	Window_Sum_2..	
30 December 2013	3/01/2014	16.45	16.45	Abc
	4/01/2014	288.06	304.51	Abc
	5/01/2014	19.54	307.60	Abc
6 January 2014	6/01/2014	4,407.10	4,426.64	Abc
	7/01/2014	87.16	4,494.26	Abc
	9/01/2014	40.54	127.70	Abc
	10/01/2014	54.83	95.37	Abc
	11/01/2014	9.94	64.77	Abc

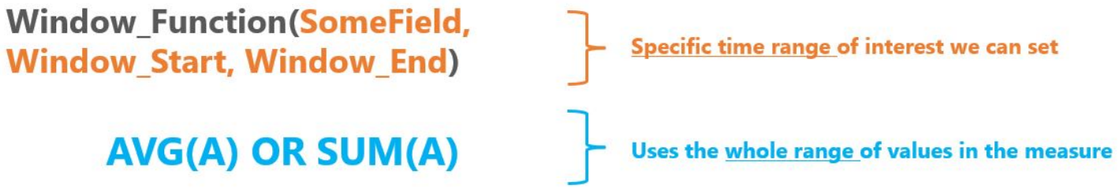
- Here we have a **daily** granularity.
- **Each window is composed of 2 days of sales.**

Week of Order Date	Sales	Window_Sum_2..	
30 December 2013	324.04	324.04	Abc
6 January 2014	4,599.57	4,923.62	Abc
13 January 2014	4,509.13	9,108.70	Abc
20 January 2014	3,842.39	8,351.52	Abc

- Here we have a **weekly** granularity.
- **Each window is composed of 2 weeks of sales.**
- **The window expression hasn't changed, but the granularity has.**

What's the purpose of a window function?

- Creates **aggregate** views of your data based on **specific** time windows
- Normal aggregation functions use the whole range of your data (i.e. no specification)

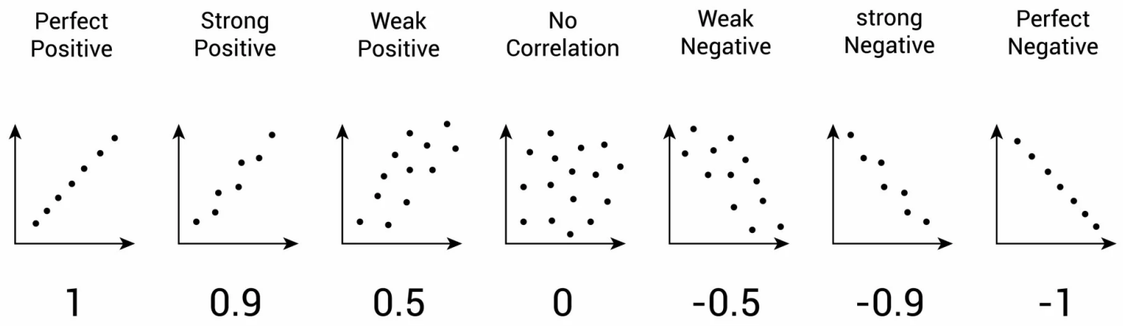


Example window functions:

- `WINDOW_SUM()`
- `WINDOW_AVG()`
- `WINDOW_STDEV()`
- `WINDOW_CORR()`
- `WINDOW_MEDIAN()`

What's correlation?

- Measures the extent to which two variables may be related
- **Correlation coefficient (r)** measures the **strength** and **direction** of the relationship
- Values range from -1 to 1
- A positive value means that as one variable increase, the other also increases
- Inversely, a negative direction means when one variable increases, the other decreases



How do window correlations work?

- Calculate the pearson correlation coefficient for the whole view between two aggregated variables
- In contrast to the `CORR()` function that requires non-aggregated variables

Category	Sales	Quantity	Correlation	Window_Corr
Furniture	741,999.795300001	8,028	0.437463606	-0.931733339
Office Supplies	719,047.032000002	22,906	0.159607503	-0.931733339
Technology	836,154.032999996	6,939	0.206949464	-0.931733339

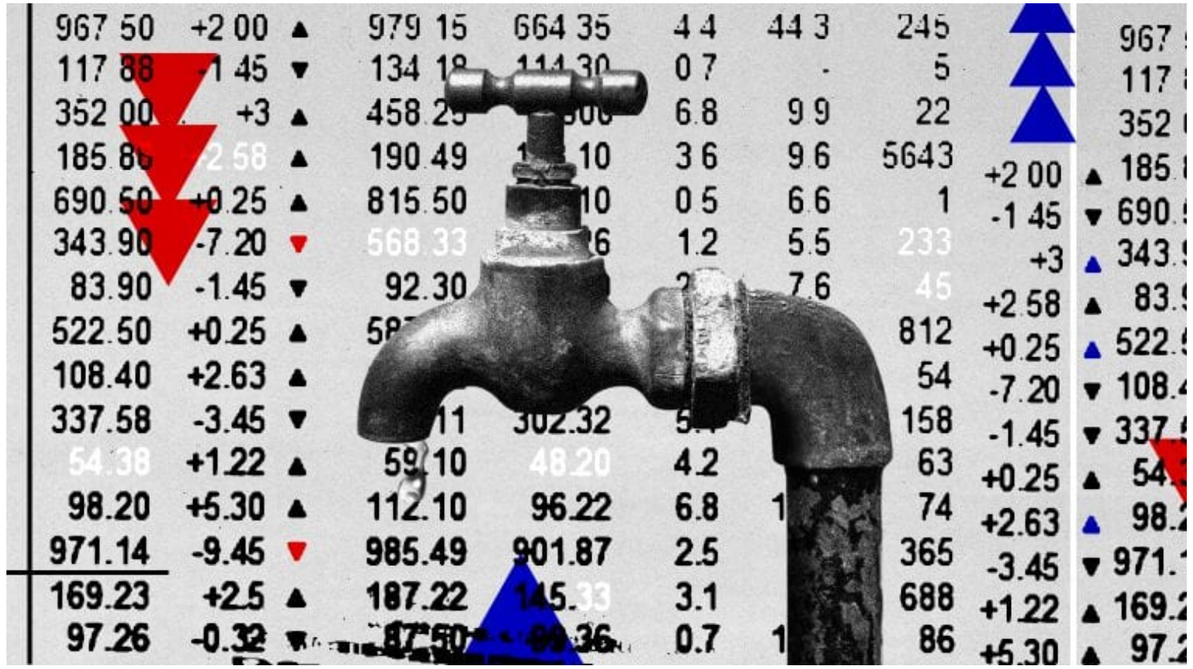
- **WINDOW_CORR()** is finding the correlation using a window containing all the values in sales and quantity at an aggregated granularity.

Category	Sales	Quantity	Correlation	
Furniture	741,999.795300001	8,028	0.437463606	Abc
Office Supplies	719,047.032000002	22,906	0.159607503	Abc
Technology	836,154.032999996	6,939	0.206949464	Abc

- **CORR()** is finding the correlation on a row-by-row basis, giving us a distinct value for each category.

Introducing the dataset

- Analyzing water trading activity and the seasonal transactions that occur
- Time series techniques to identify abnormal pricing patterns:
  - Moving averages
  - Window correlations



[1] https://theswaddle.com/water-is-now-a-traded-commodity-can-it-still-be-a-human-right-too/

Let's practice!

TIME SERIES ANALYSIS IN TABLEAU

Calculations with window functions

TIME SERIES ANALYSIS IN TABLEAU

Chris Hui
VP, Tracked

datacamp

Let's practice!

TIME SERIES ANALYSIS IN TABLEAU

Anomaly detection with window functions

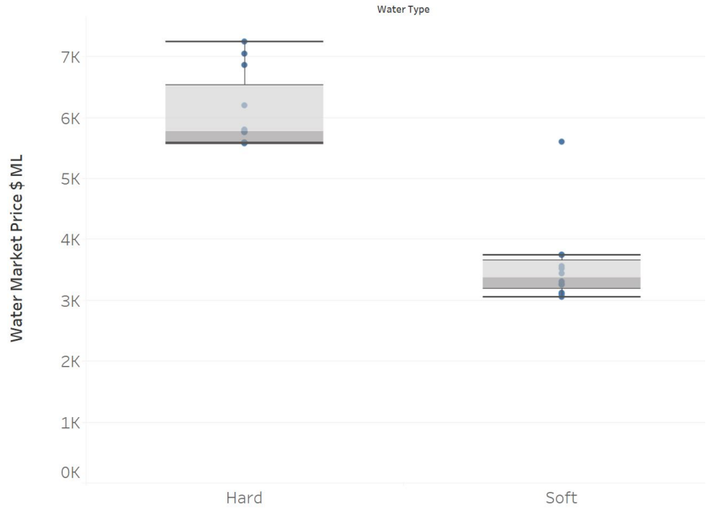
TIME SERIES ANALYSIS IN TABLEAU

Chris Hui
VP, Tracked

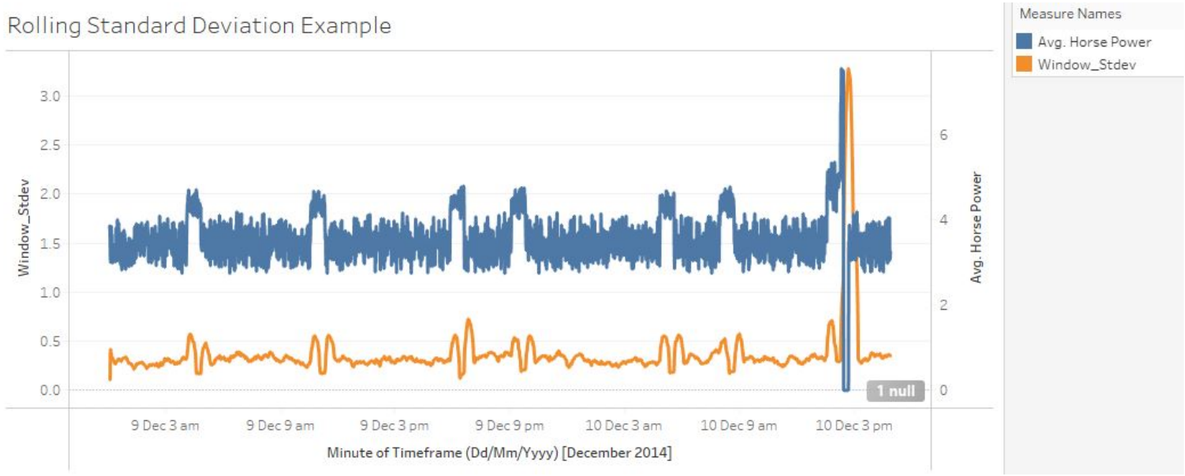
datacamp

Standard deviation versus rolling standard deviation

- Standard deviation measures the degree of dispersion in a set of values
  - **High standard deviation** = high variance
  - **Low standard deviation** = low variance

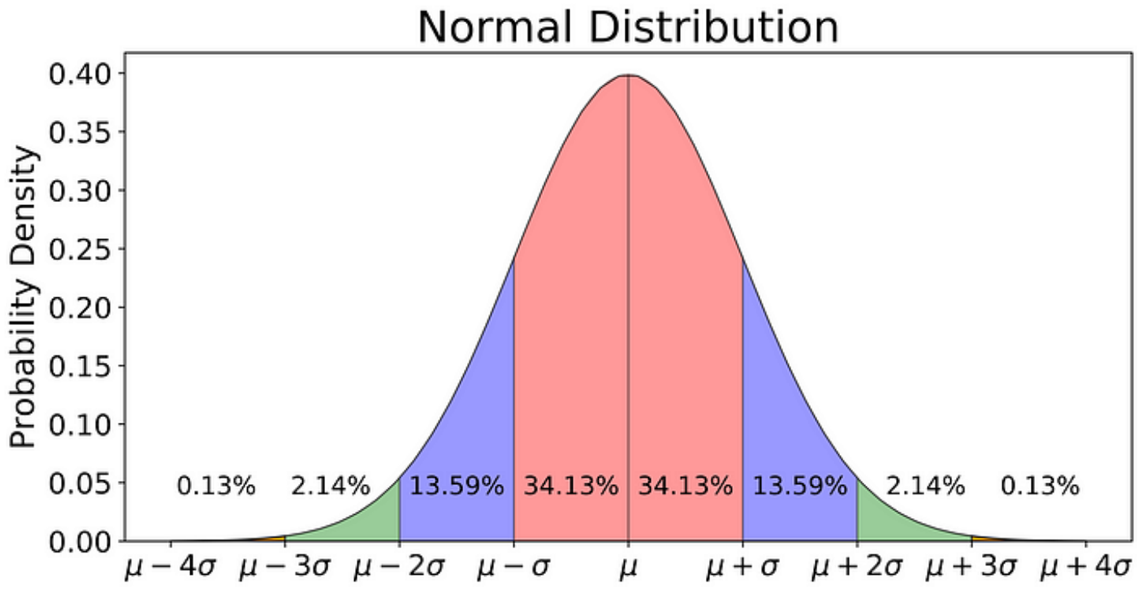


- Rolling standard deviation, calculated on a window subset, is useful to **identify variance inflation with respect to time**
- As the variance grows larger, this may signal an anomaly for analysis purposes



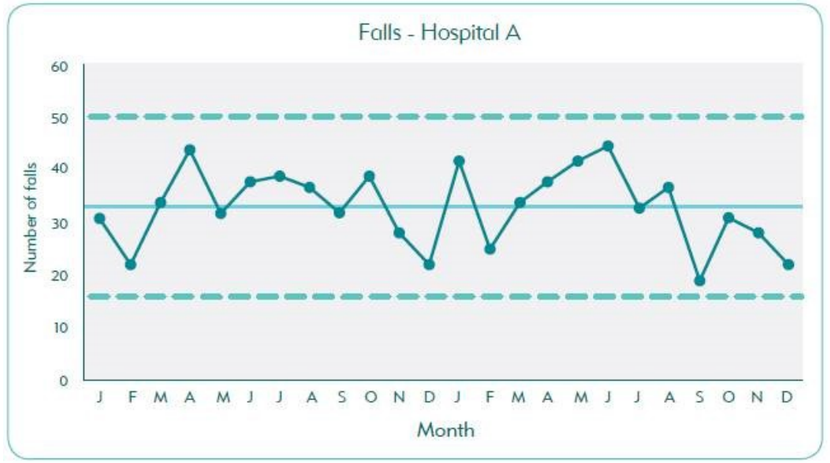
Standard deviation and anomaly detection

- Anomaly detection for time series data generally follows the 68, 95, 99 rule
- ~ 68% of all values = 1 standard deviation away from the mean
- 95% of all values = 2 standard deviation away from the mean
- ~ 99.7% of all values = 3 standard deviation away from the mean
- Any value > 3 standard deviations away from the mean is anomalous

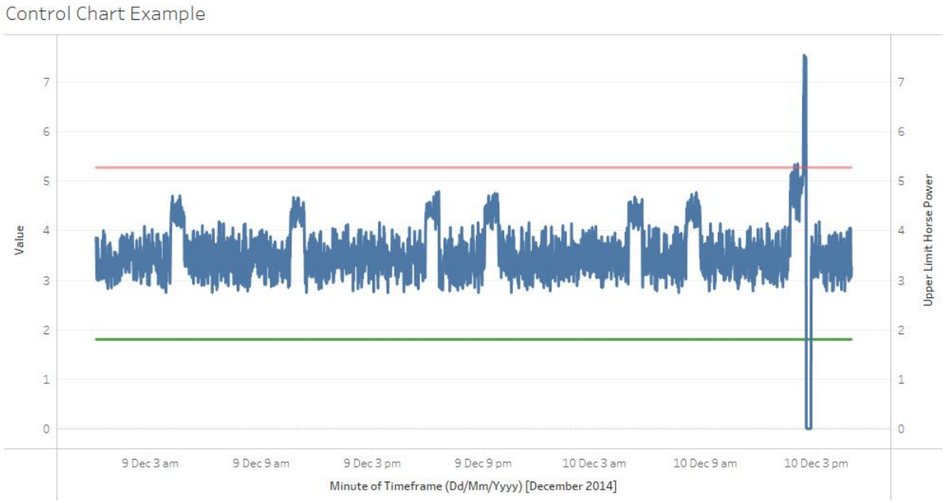


Upper and lower control limits

- Primarily utilized for univariate time series analysis as opposed to multivariate
- Control charts are an effective visual way of identifying the upper and lower bounds of what are *acceptable* values

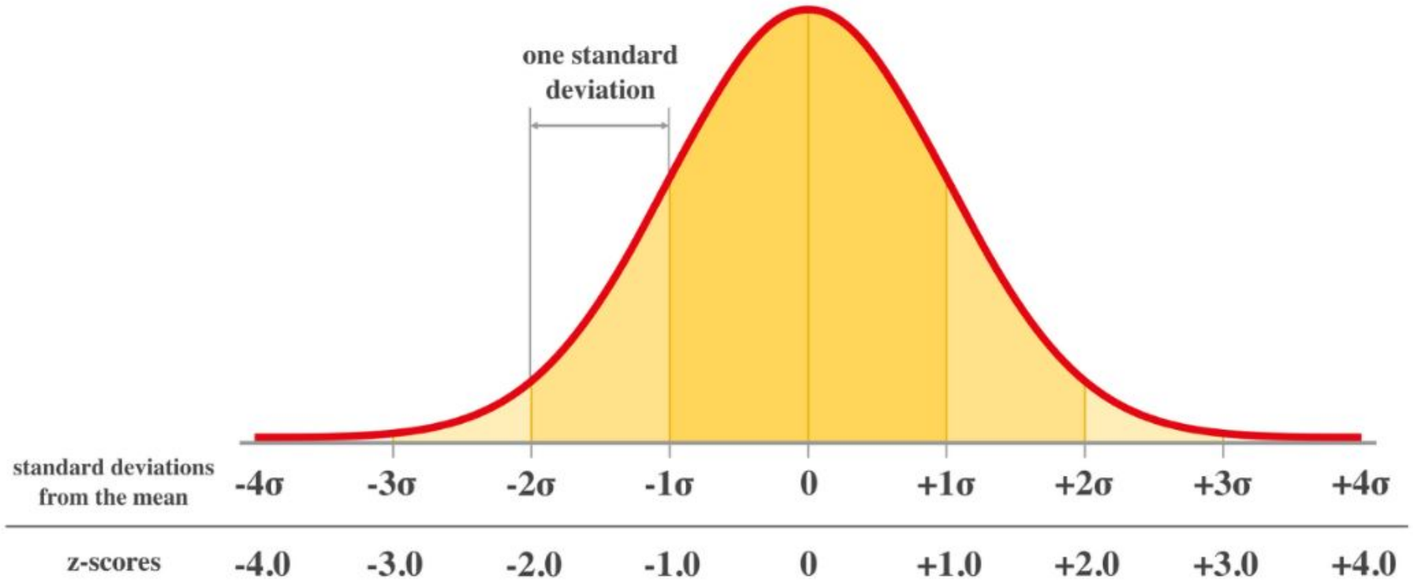


- Values that exceed the population mean +- 3 standard deviations are anomalous



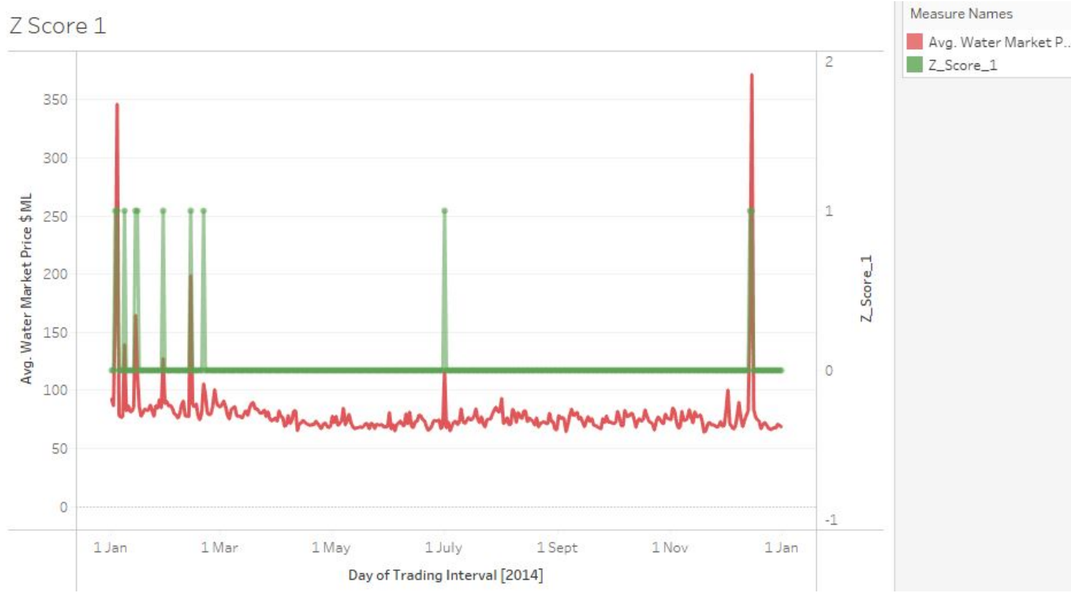
What are Z-scores?

- The Z-score is the number of standard deviations a data point lies above or below the mean
- A positive Z-score indicates the value is above the mean
- A negative Z-score indicates the value is below the mean
- Separate from standard deviation that measures distance between data points

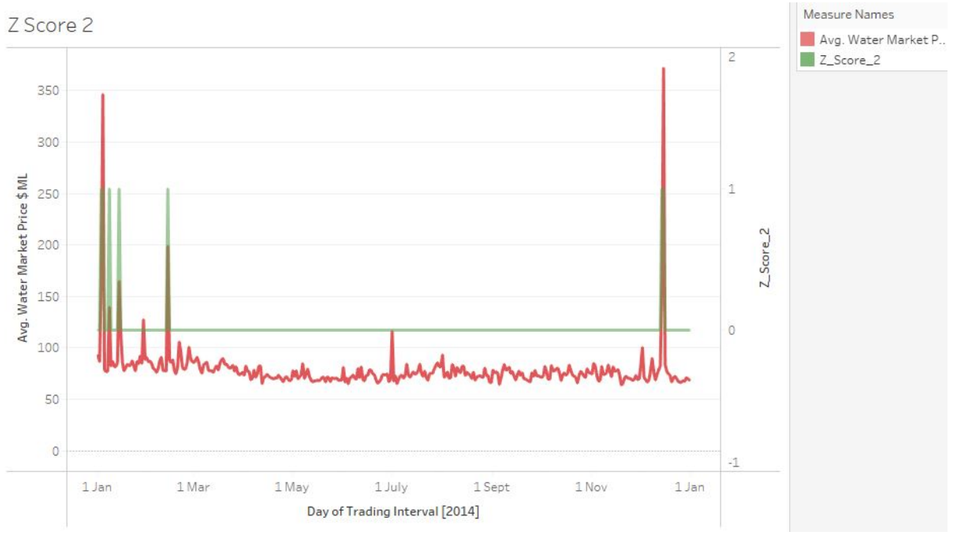


Z-scores and anomaly detection

- Z-scores +-3 are considered anomalous, but this is contextual



- Higher Z-scores mean less anomalies, but this depends on how sensitive your anomaly detection is



Let's practice!

TIME SERIES ANALYSIS IN TABLEAU

A deep dive into window calculations

TIME SERIES ANALYSIS IN TABLEAU

Christopher Hui
VP, Tracked

datacamp

Let's practice!

TIME SERIES ANALYSIS IN TABLEAU

Congratulations!

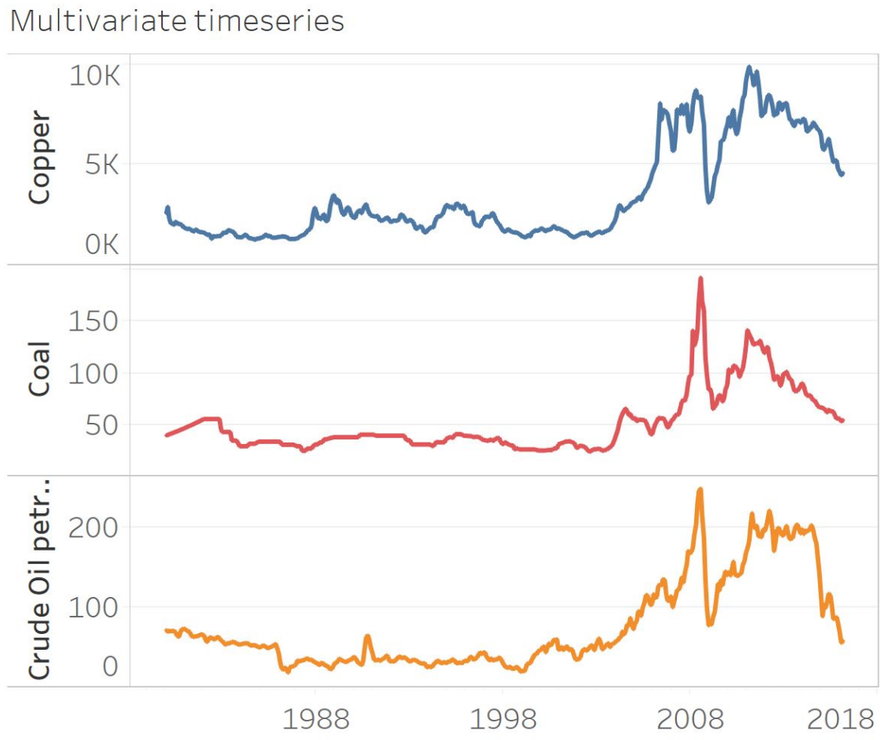
TIME SERIES ANALYSIS IN TABLEAU

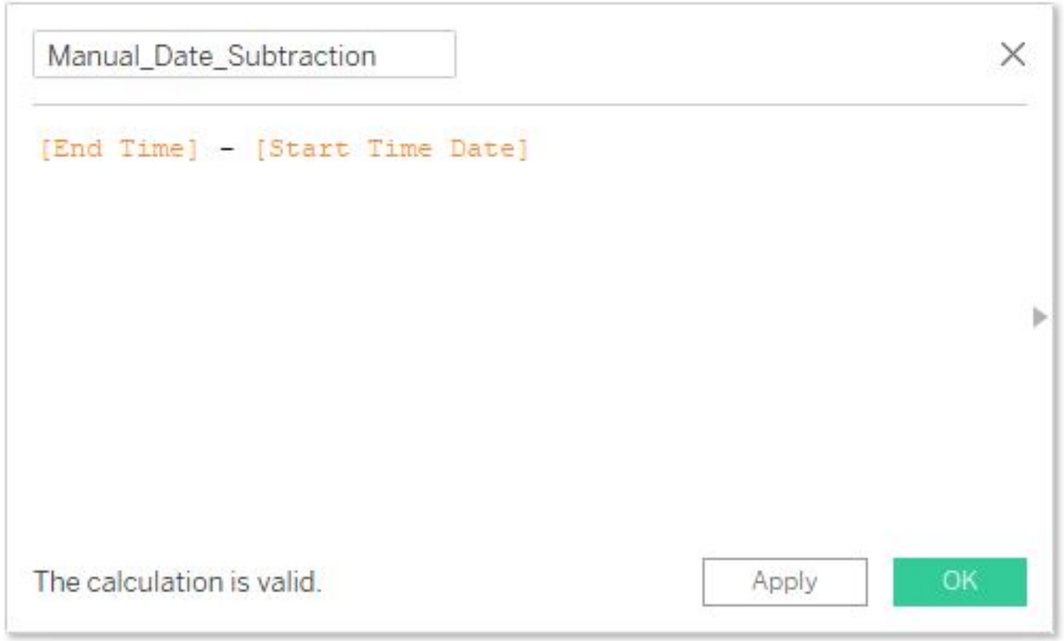
Chris Hui
VP, Tracked

datacamp

What you've learned: chapter one

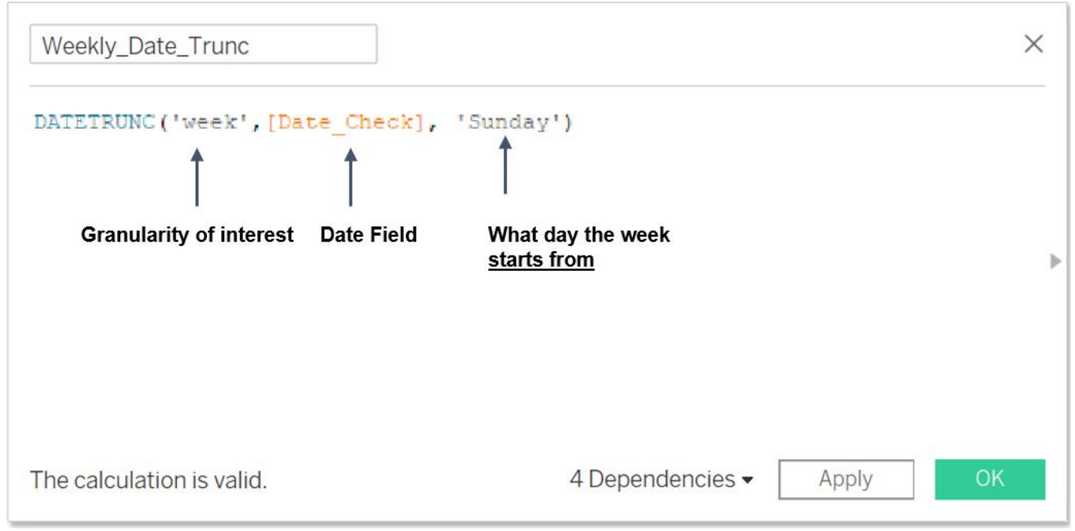
- What times series data is
- Continuous and discrete time series
- Univariate and multivariate analysis
- Data granularity
- Reforming and truncating date time stamps
- Calculated fields with dates



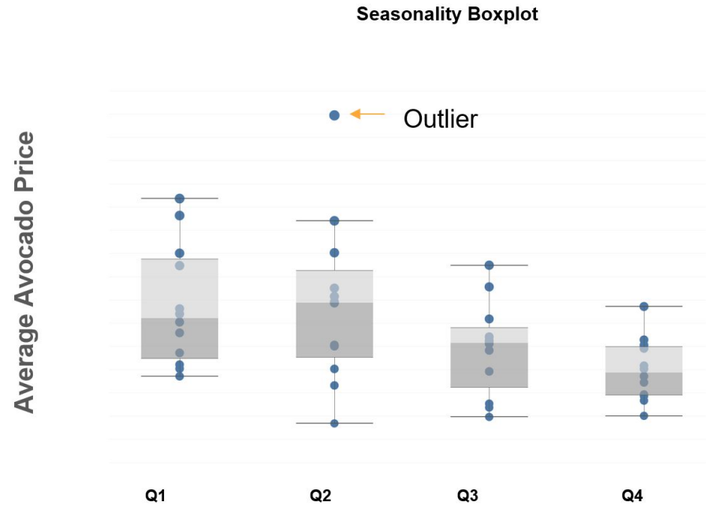


What you've learned: chapter two

- Date parsing
- Splitting up dates
- Date time stamp validation



- Seasonality
- Moving averages
- Box plots
- Percentiles



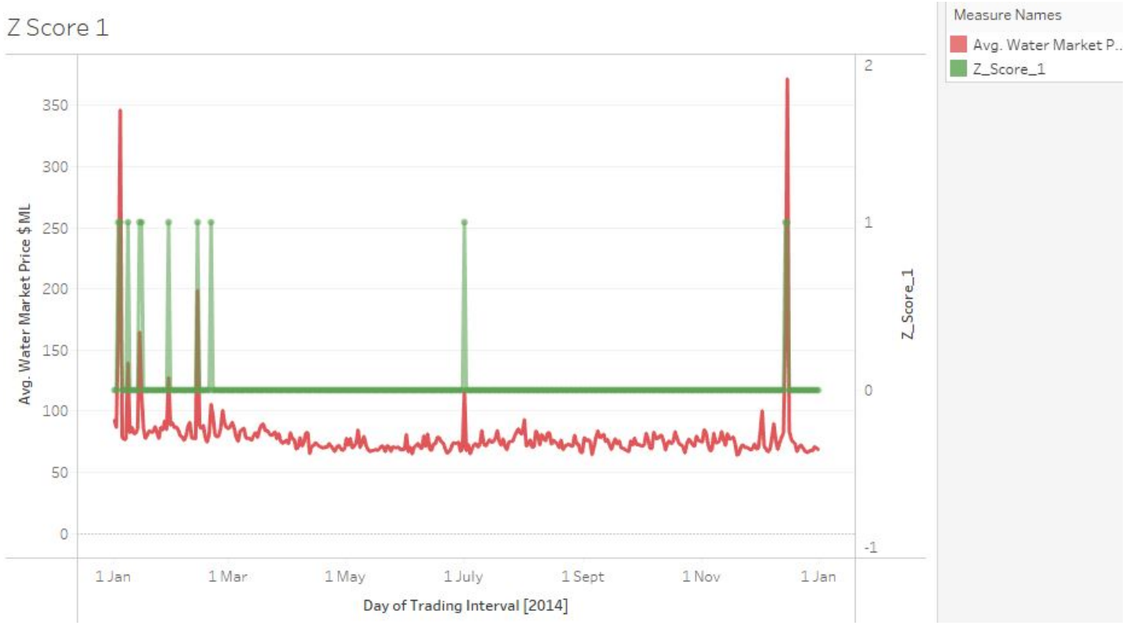
What you've learned: chapter three

- Windows
- Window sizes
- Window functions

Week of Or..	Order Date	Sales	Window_Sum_2..	
30 December 2013	3/01/2014	16.45	16.45	Abc
	4/01/2014	288.06	304.51	Abc
	5/01/2014	19.54	307.60	Abc
6 January 2014	6/01/2014	4,407.10	4,426.64	Abc
	7/01/2014	87.16	4,494.26	Abc
	9/01/2014	40.54	127.70	Abc
	10/01/2014	54.83	95.37	Abc
	11/01/2014	9.94	64.77	Abc

- Here we have a **daily** granularity.
- **Each window is composed of 2 days of sales.**

- Rolling standard deviations
- Upper and lower boundaries
- Z-scores and anomaly detection



Thank you!

TIME SERIES ANALYSIS IN TABLEAU