Date transformations and visualizations

TIME SERIES ANALYSIS IN TABLEAU

Chris Hui
VP of Product, Tracked

Splitting up data in Tableau

- Data can be encoded via delimiters that can be split apart into their own fields (e.g. AUS-001)
- Tableau has two distinct splitting options: *Split* or *Custom split*
- **Split:** separates the data based off the most common delimiter

Splitting up data in Tableau

- Data can be encoded via delimiters that can be split apart into their own fields (e.g. AUS-001)
- Tableau has two distinct splitting options: *Split* or *Custom split*
- **Split:** separates the data based off the most common delimiter

Cust Order Key	Split 1	Split 2	Split 3
AA-B1-C1	AA	B1	C1
AA-B2-C2	AA	B2	C2
AA-B3-C3	AA	B3	C3

Splitting up data in Tableau

- Data can be encoded via delimiters that can be split apart into their own fields (e.g. AUS-001)
- Tableau has two distinct splitting options: *Split* or *Custom split*
- **Split:** separates the data based off the most common delimiter
- **Custom split:** separates the data based off custom delimiters specified

Cust Order Key	Split 1	Split 2	Split 3
AA-B1-C1	AA	B1	C1
AA-B2-C2	AA	B2	C2
AA-B3-C3	AA	B3	C3

Cust Order Key	Split 1	Split 2	Split 3
AA-B1-C1.DD	AA	B1	C1.DD
AA-B2-C2.DD	AA	B2	C2.DD
AA-B3-C3.DD	AA	B3	C3.DD

Calculated splits in Tableau

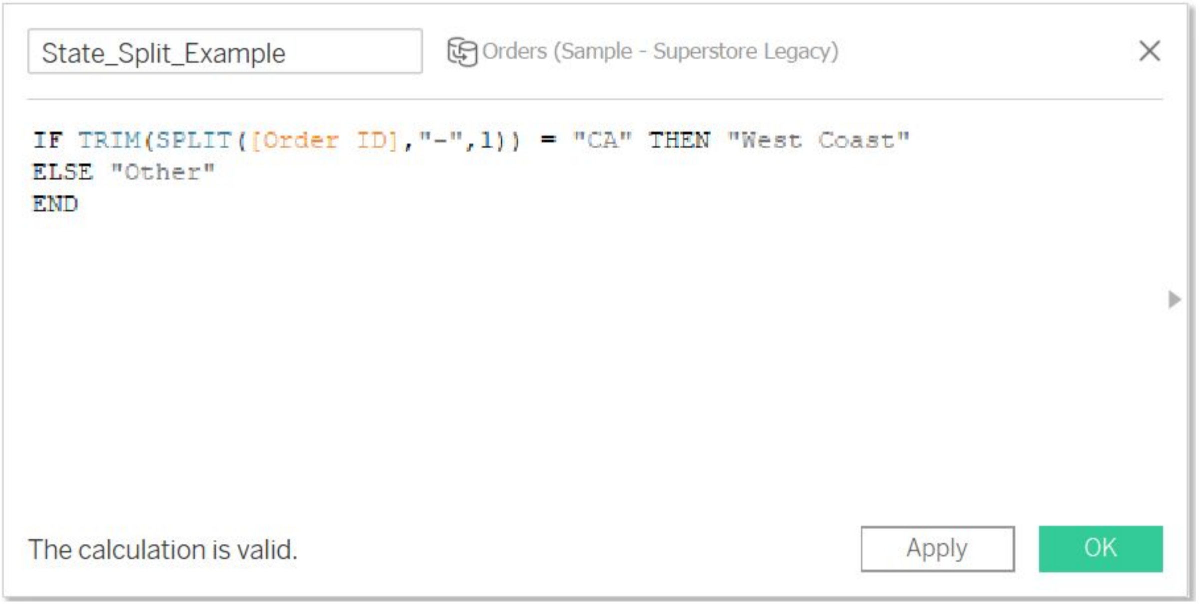
- `SPLIT()` is helpful when you need to split up one or **more** delimiters in a textual field
- The tokens (text) are returned either left to right (positive), or right to left (negative) dependent on position
- Splitting functions are generally nested and combined with functions like:
  - `IF()` / `TRIM()` / `CASE`

SPLIT("Alpha-Beta-Gamma","-",3)

Token Position	**1**	**2**	**3**
Token Position	**-3**	**-2**	**-1**

Positive numbers indicate a **Left to Right** sequence.
Negative numbers indicate a **Right to Left** sequence.

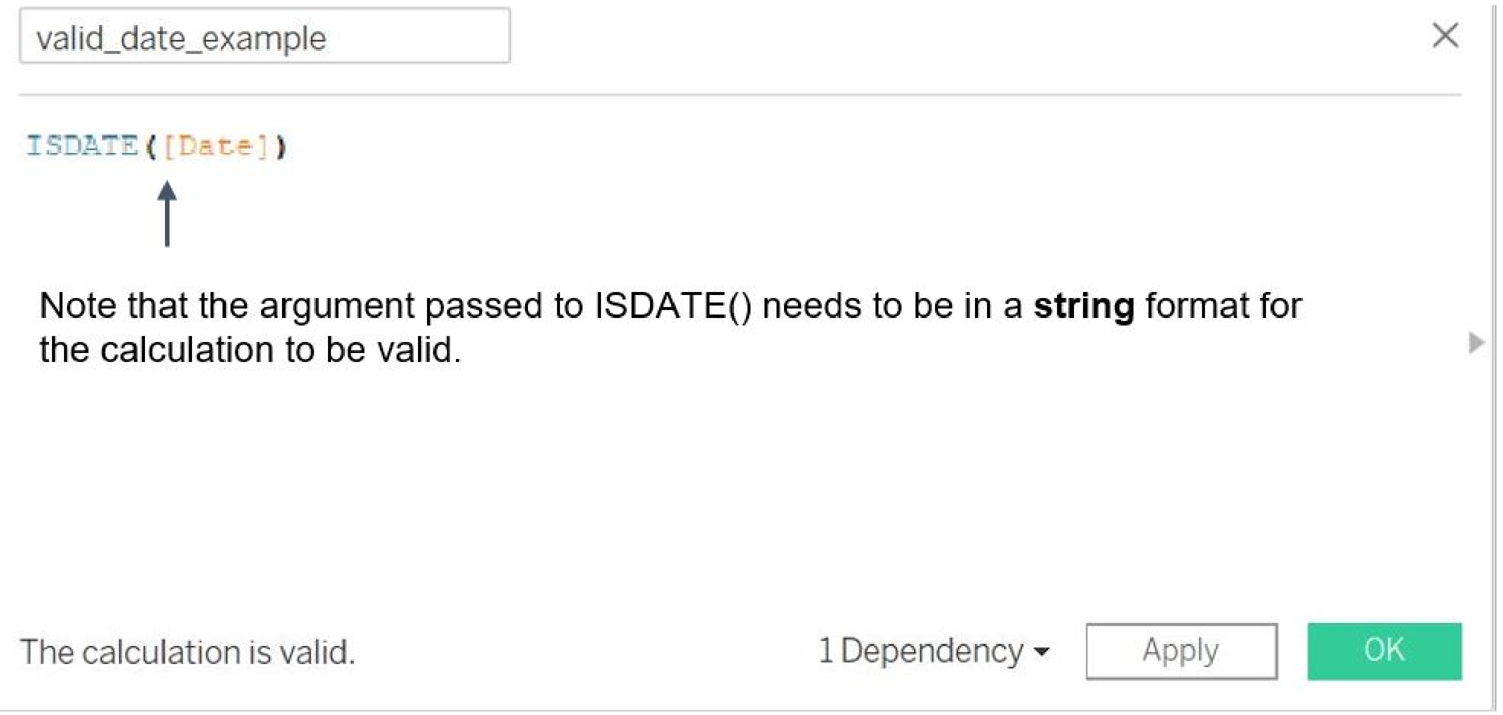
This is why Gamma, can be returned either with a 3 position, or -1 position.



Validating dates

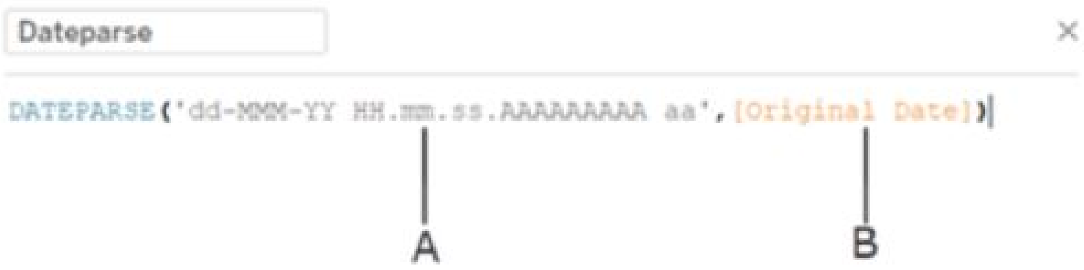
- By design, Tableau adheres to and recognizes ISO-8601 and Standard Gregorian date formats
- However dates can sometimes appear in a textual / string format
- The `ISDATE()` function can be utilized to return **True** for all entries recognized as dates and **False** otherwise

Format according to ISO 8601	Value ranges
Year (Y)	YYYY, four-digit, abbreviated to two-digit
Month (M)	MM, 01 to 12
Week (W)	WW, 01 to 53
Day (D)	D, day of the week, 1 to 7
Hour (h)	hh, 00 to 23, 24:00:00 as the end time
Minute (m)	mm, 00 to 59
Second (s)	ss, 00 to 59
Decimal fraction (f)	Fractions of seconds, any degree of accuracy



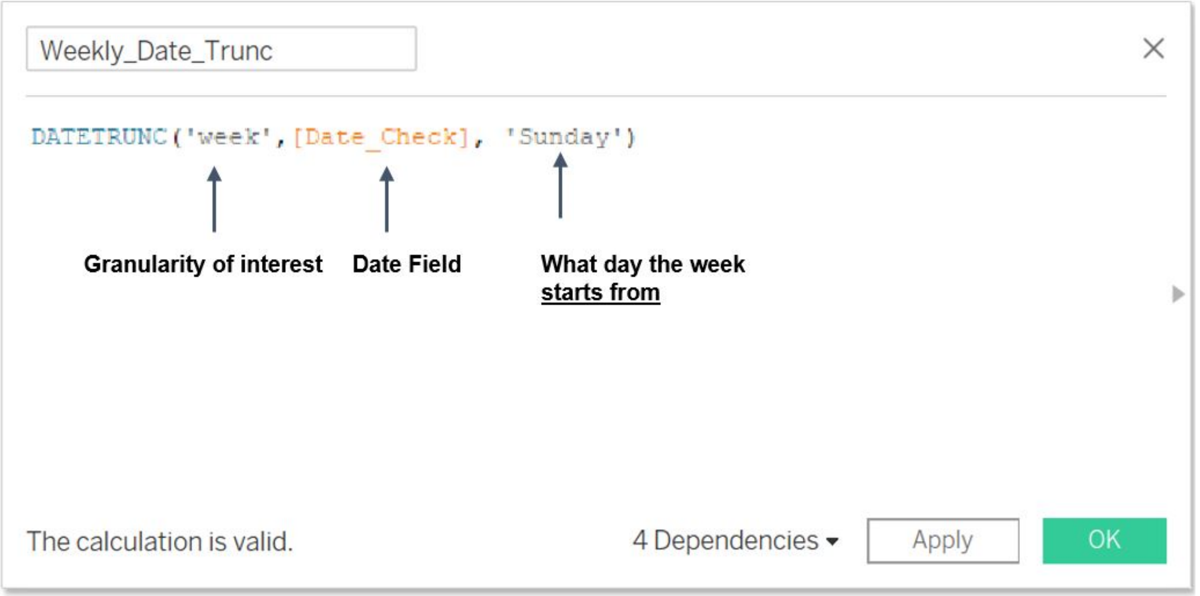
To parse or not to parse?

- `DATEPARSE()` converts non-standard string fields to date time objects through **explicit** parsing of the date format
- It can be utilized for **all** date strings that are recognized as strings as opposed to dates
- *MM-YY-DD* might makes sense to us, but not to Tableau, **unless** you specify this with `DATEPARSE()`



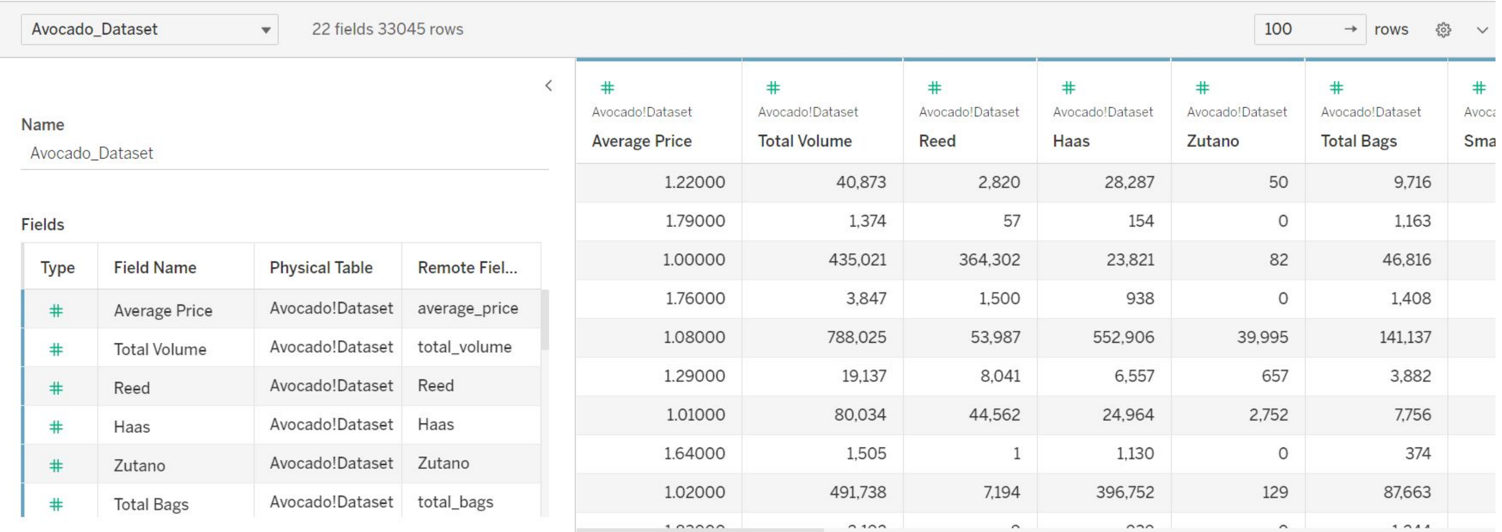
Cleansing with date truncation

- `ISDATE()` and `DATEPARSE()` are useful to identify invalid dates, but **not** correcting these dates
- `DATETRUNC()` returns the **lowest** value for each date partition, dependent of the day specification (i.e. Week start on Monday)
- `DATETRUNC()` is useful for verifying re-occurring reporting periods for errors (e.g. Weekly)



Date	reporting_week
2015/9/32	27/09/2015 12:00:00 am
2015/12/32	27/12/2015 12:00:00 am

The dataset



Avocado_Dataset 22 fields 33045 rows 100 rows

Name
Avocado_Dataset

Fields

Type	Field Name	Physical Table	Remote Fiel...
#	Average Price	Avocado!Dataset	average_price
#	Total Volume	Avocado!Dataset	total_volume
#	Reed	Avocado!Dataset	Reed
#	Haas	Avocado!Dataset	Haas
#	Zutano	Avocado!Dataset	Zutano
#	Total Bags	Avocado!Dataset	total_bags

Avocado!Dataset Average Price	Avocado!Dataset Total Volume	Avocado!Dataset Reed	Avocado!Dataset Haas	Avocado!Dataset Zutano	Avocado!Dataset Total Bags	Avoca Sma
1.22000	40,873	2,820	28,287	50	9,716	
1.79000	1,374	57	154	0	1,163	
1.00000	435,021	364,302	23,821	82	46,816	
1.76000	3,847	1,500	938	0	1,408	
1.08000	788,025	53,987	552,906	39,995	141,137	
1.29000	19,137	8,041	6,557	657	3,882	
1.01000	80,034	44,562	24,964	2,752	7,756	
1.64000	1,505	1	1,130	0	374	
1.02000	491,738	7,194	396,752	129	87,663	

- Univariate and multivariate analysis of Avocado Varieties
- Seasonal variations and pricing distributions
- Trend analysis & percentiles

Let's practice!

TIME SERIES ANALYSIS IN TABLEAU

Reforming dates in practice

TIME SERIES ANALYSIS IN TABLEAU

Chris Hui
VP Product

Let's practice!

TIME SERIES ANALYSIS IN TABLEAU

Date transformations and visualizations

TIME SERIES ANALYSIS IN TABLEAU

Chris Hui
VP of Product, Tracked

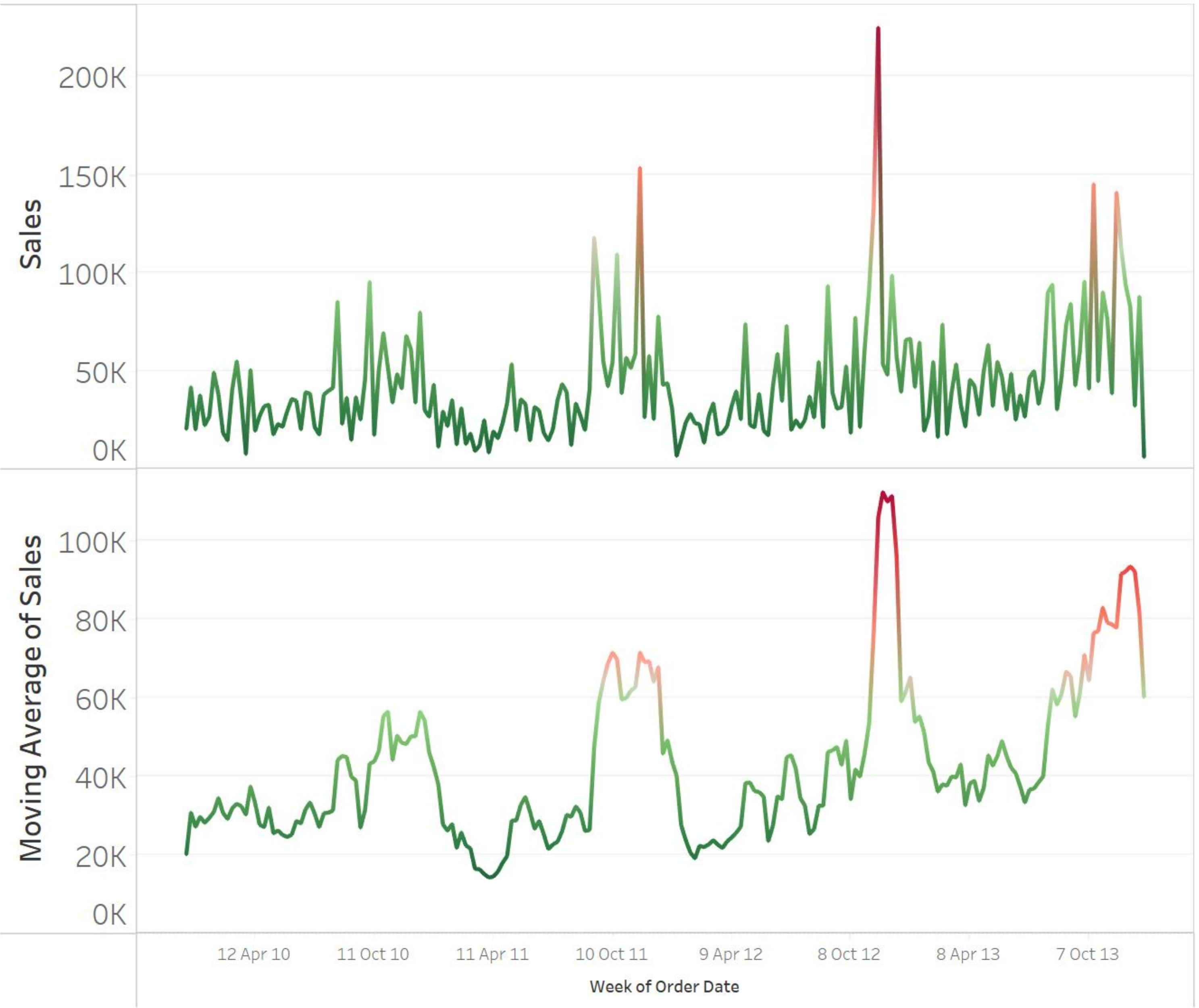
What's seasonality?

- Seasonality is when time series data experiences **regular and predictable changes** that recur every calendar year
- Examples include tourism or fruiting seasons that have variable prices based off timing
- Seasonal behaviour allows business to effectively plan around peaks and troughs to optimize their business

Treating seasonality with moving averages

Common methods to treat seasonality:

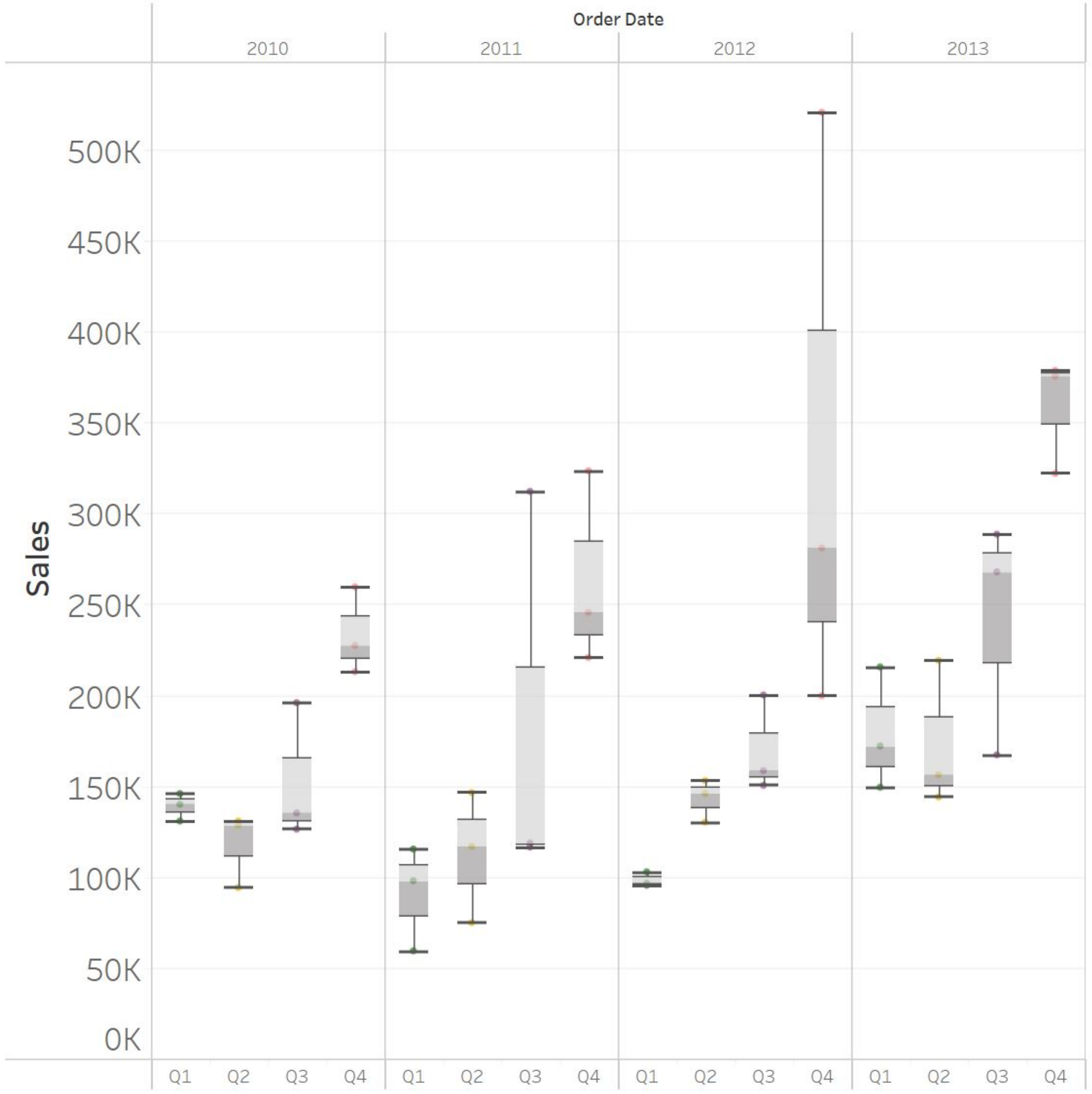
- **Moving averages**
  - Technique to smooth out short term fluctuations (peaks/troughs) in the data over a specific time window
  - Used to filter out noise while preserving the underlying signal



Identifying seasonality with seasonal boxplots

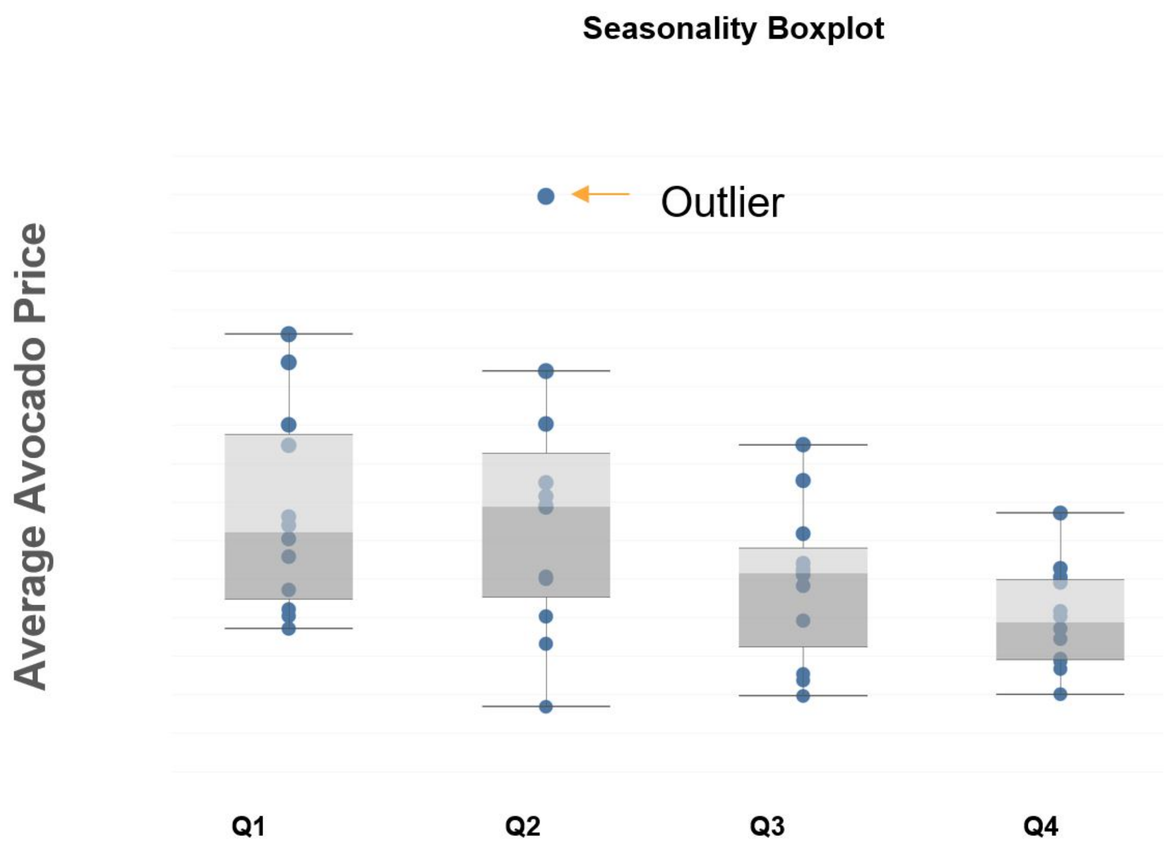
Common methods to identify seasonality:

- **Seasonal boxplots (quarterly analysis)**
  - Segments the data quarterly to enable visualization of seasonal (quarterly) fluctuations
  - Consistent volatility across quarters are an indicator of seasonal behavior



What's an anomaly?

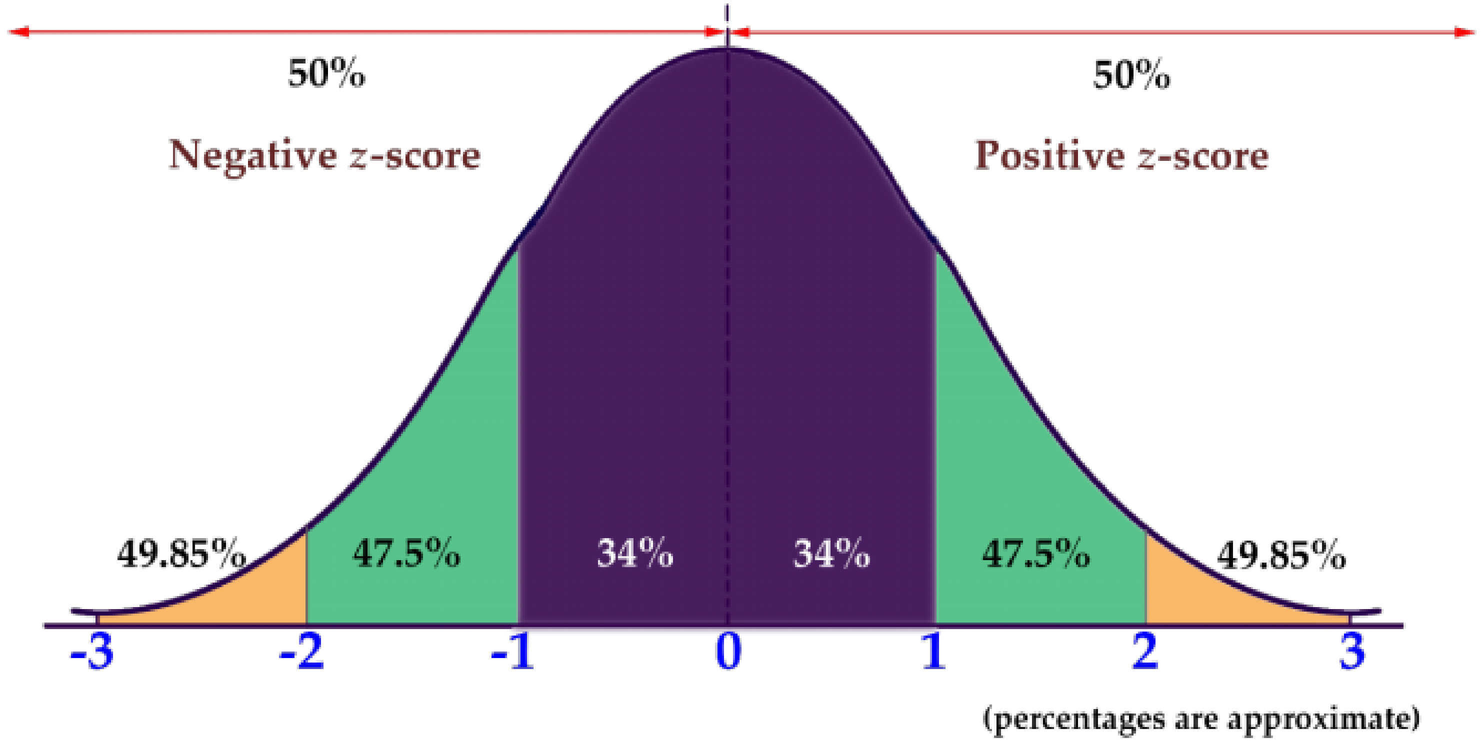
- Anomalous values (outliers), are values that deviate outside the normal distribution
- Outliers can be considered to be:
  - Any value outside +- 3 standard deviations away from the mean
  - More than 1.5 x IQR below Q1 or more than 1.5 x IQR above Q3



(**Standard deviation** is a measure of how far any value is from the population mean)

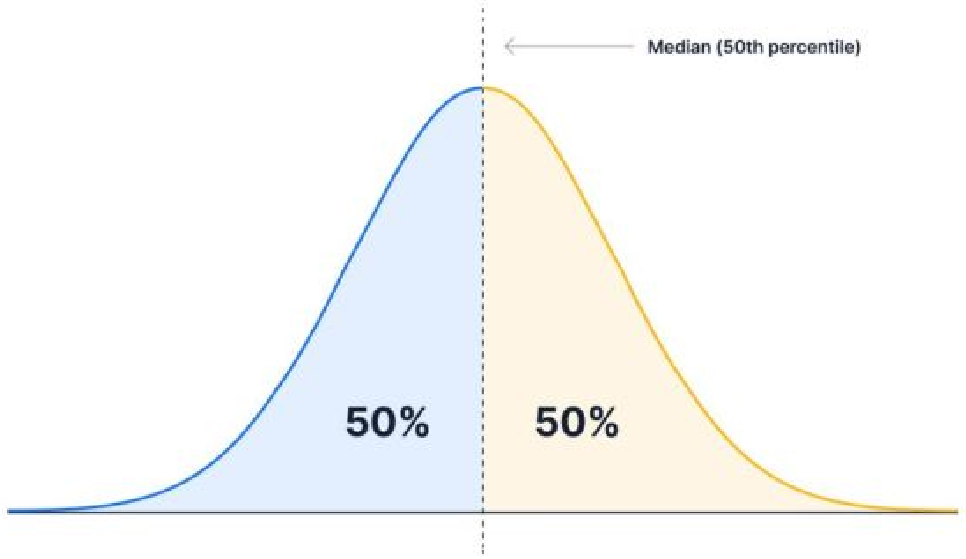
Z-score and the normal distribution

- The Z-score is the number of standard deviations a given data point lies above or below mean
- Z-scores within +-3 means ~99.7% of the population values lies within this range
- Subsequently, any Z-score outside the +-3 range can be considered an outlier

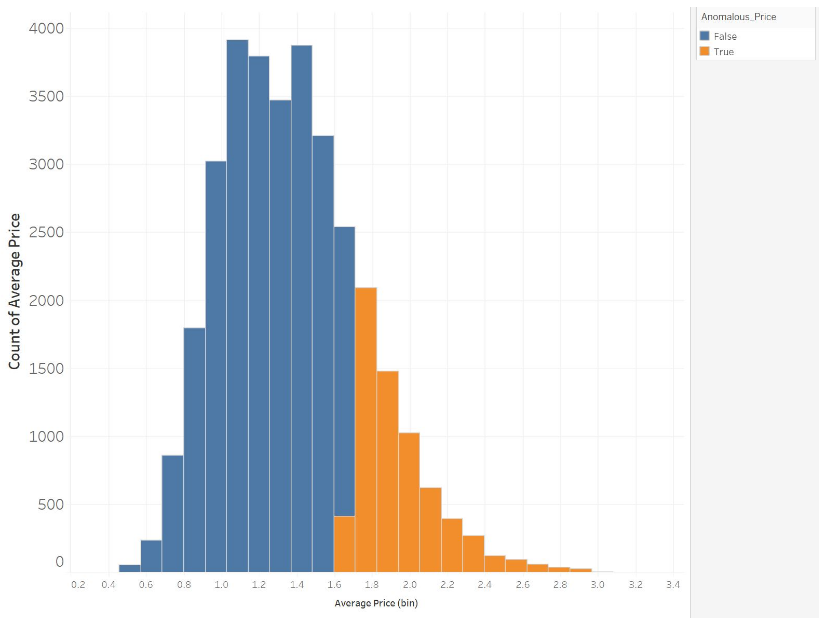


Unpacking percentiles in Tableau

- Percentiles determine where a value stands *relative* to other values
  - Median (50th Percentile)
  - Upper & Lower Quartiles (75th / 25th percentiles)



- Provides a flexible approach to outlier detection beyond the traditional Z-score methodology



Let's practice!

TIME SERIES ANALYSIS IN TABLEAU

Visualizing seasonality and percentiles

TIME SERIES ANALYSIS IN TABLEAU

Chris Hui
VP Tracked

Let's practice!

TIME SERIES ANALYSIS IN TABLEAU