DEMAND PLANNING
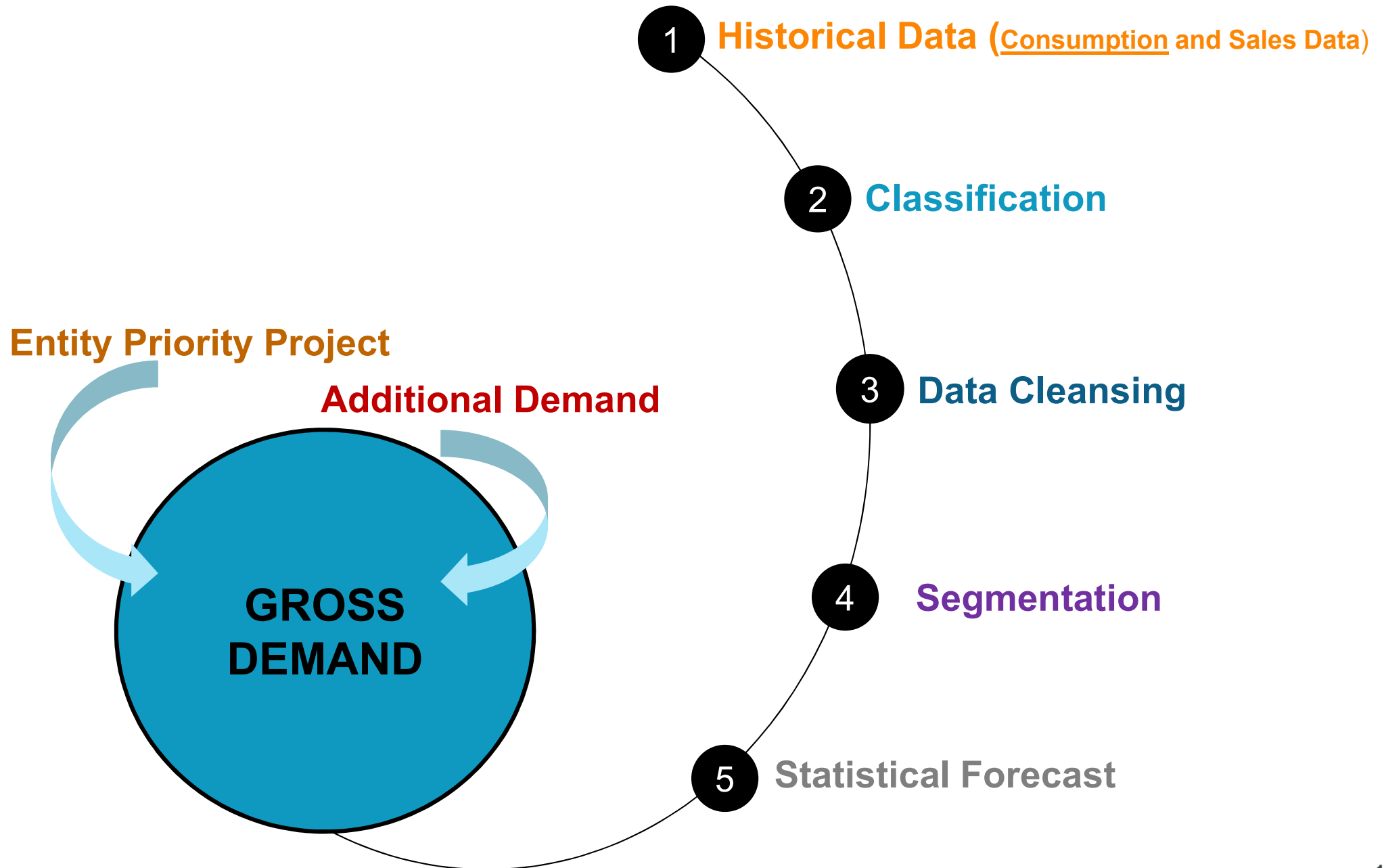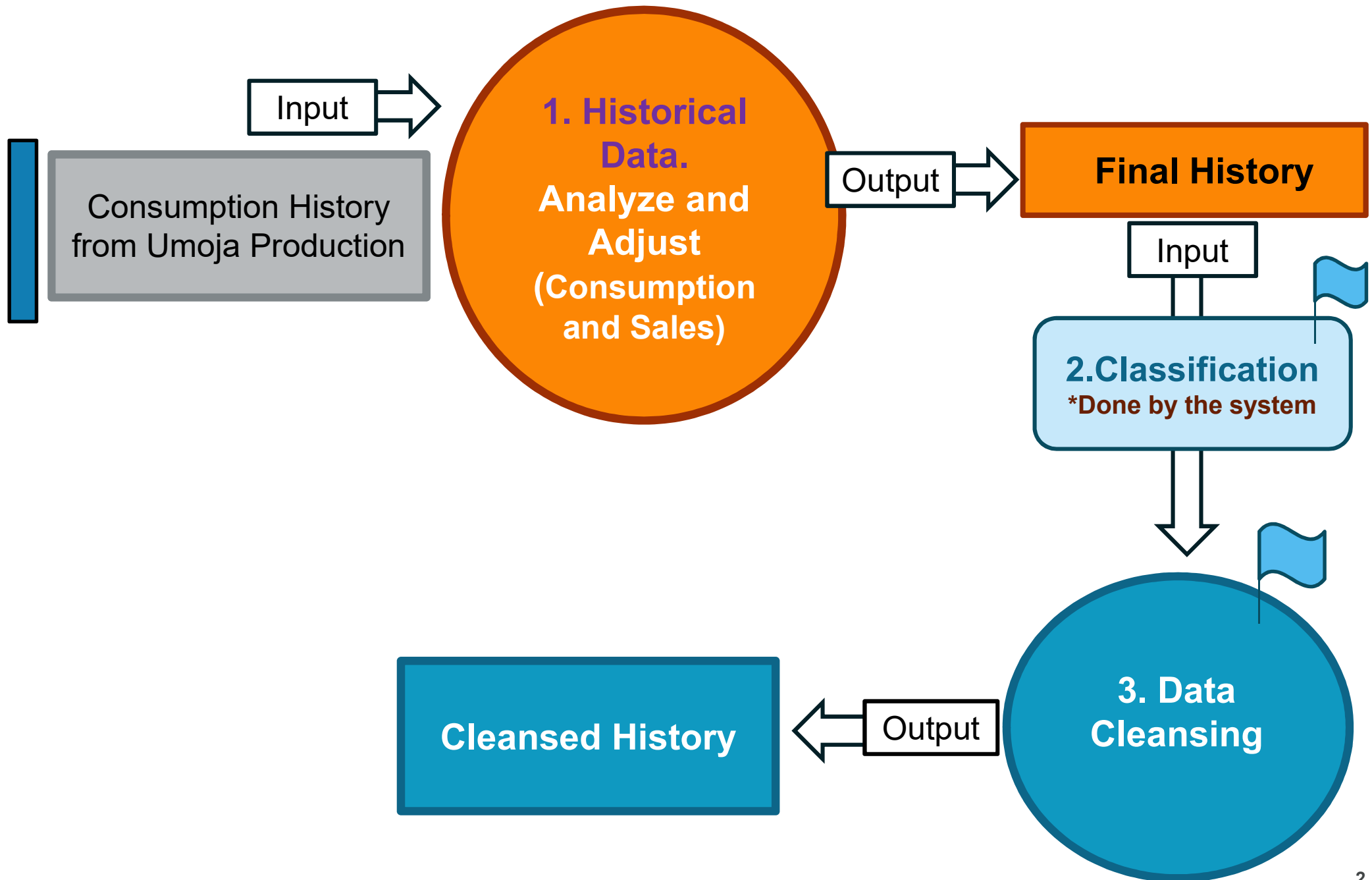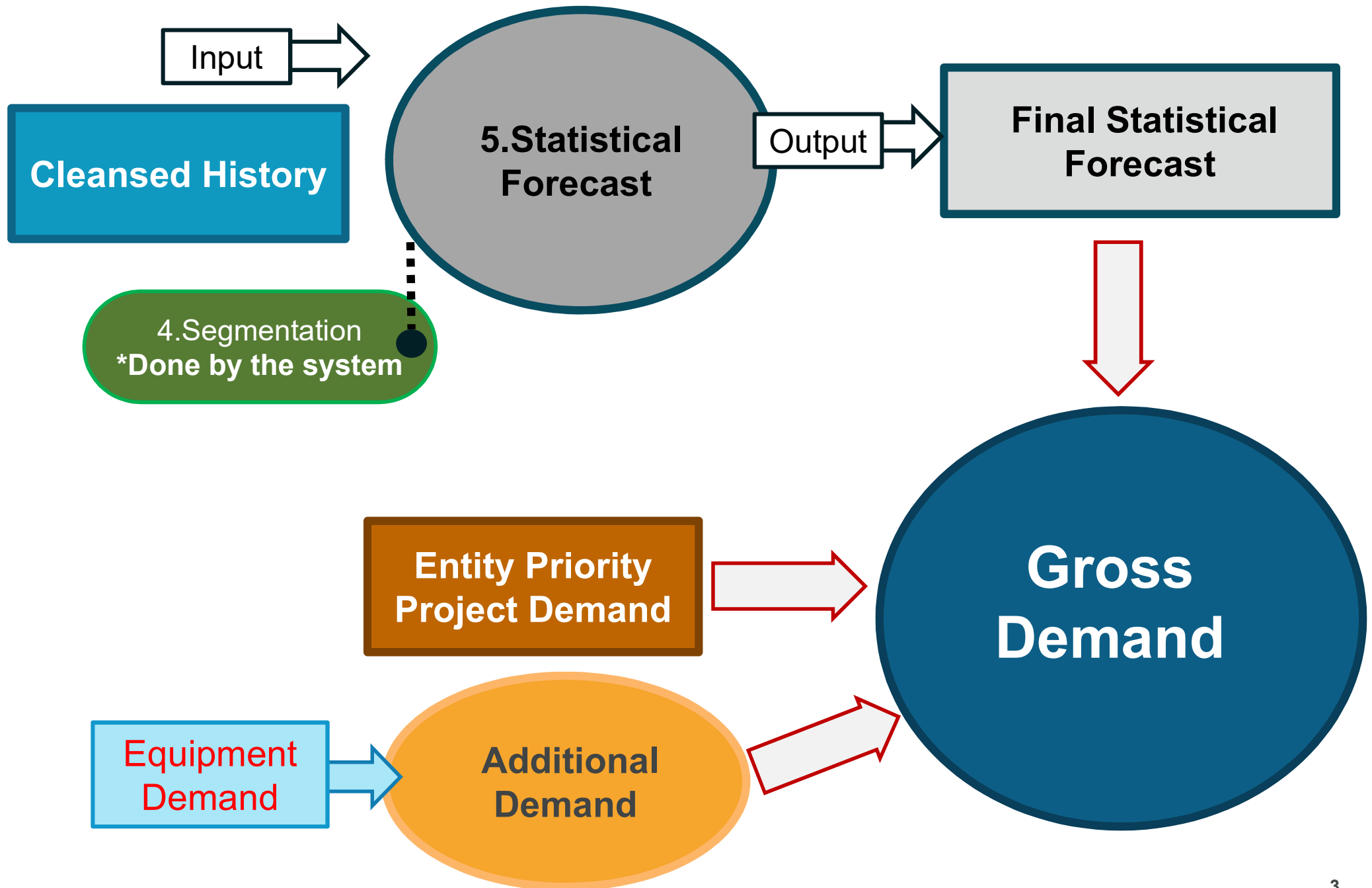
Umoja Demand Planning and Supply Network Planning Solution

Classification and Data Cleansing

UCS Training Team

DEPARTMENT OF OPERATIONAL SUPPORT

1 **Historical Data (**<u>**Consumption**</u> **and Sales Data**)

2 **Classification**

3 **Data Cleansing**

4 **Segmentation**

5 Statistical Forecast

Entity Priority Project

Additional Demand

GROSS DEMAND

DEPARTMENT OF
**OPERATIONAL
SUPPORT**

Input

Consumption History
from Umoja Production

**1. Historical Data.
Analyze and Adjust
(Consumption and Sales)**

Output

Final History

Input

2.Classification
Done by the system

3. Data Cleansing

Output

Cleansed History

2

DEPARTMENT OF OPERATIONAL SUPPORT
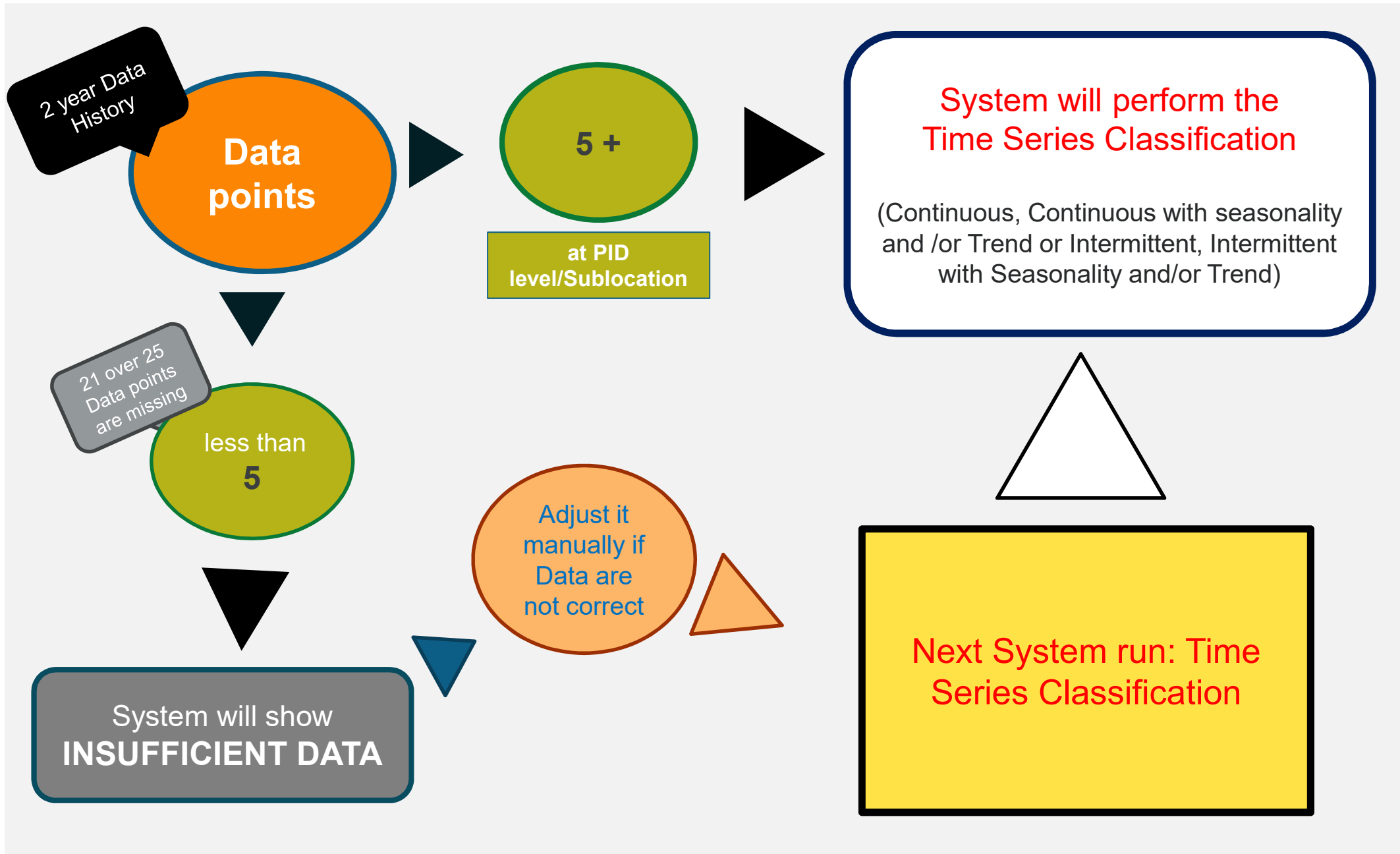
2.Classification
*Done by the system

Classifies the Final Consumption History

➢The purpose of this pre-process step is carrying out a statistical analysis on the <u>Final Consumption History</u>, and classifying the time series according to various criteria such as seasonality, trend, intermittence etc.

The characteristics, called "<u>Time Series Classification</u>" are:

❖ Continuous
❖ Continuous with seasonality
❖ Continuous with seasonality and trend
❖ Continuous with trend
❑ Intermittent
❑ Intermittent with seasonality
❑ Intermittent with seasonality and trend
❑ Intermittent with trend

DEPARTMENT OF
OPERATIONAL
SUPPORT

2 year Data History

Data points

5 +

at PID level/Sublocation

System will perform the Time Series Classification

(Continuous, Continuous with seasonality and /or Trend or Intermittent, Intermittent with Seasonality and/or Trend)

21 over 25 Data points are missing

less than **5**

Adjust it manually if Data are not correct

System will show **INSUFFICIENT DATA**

Next System run: Time Series Classification

Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18	12-18	01-19
Paper:Ptr/Copier Ream/A4	KE00_5101	Final History	**Continuous**	970	510	750	215	1260	550	775	635	470	950	755	1025	510	1205	110
Paper:Ptr/Copier Ream/A4	SS10_4101	Final History	**Continuous with seasonality**	20	37	25	18	27	30	30	21	16	47	25	37	25	34	35
Paper:Ptr/Copier Ream/A4	SS10_4111	Final History	**Continuous with seasonality and trend**	56	38	42	29	87	55	50	53	102	70	60	180	102	80	151
Paper:Ptr/Copier Ream/A4	SS10_4106	Final History	**Continuous with trend**	145	150	200	220	240	260	280	300	340	350	380	400	450	460	460
Paper:Ptr/Copier Ream/A4	LB10_4101	Final History	**Intermittent**		312	1110	665	373						60	40	55		
Paper:Ptr/Copier Ream/A4	SS10_4112	Final History	**Intermittent with seasonality**						28	80	5	10	51	8	70	55	40	100

Continuous

Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18	12-18	01-19
Paper:Ptr/Copier Ream/A4	KE00_5101	Final History	Continuous	970	510	750	215	1260	550	775	635	470	950	755	1025	510	1205	110

Representation of Continuous
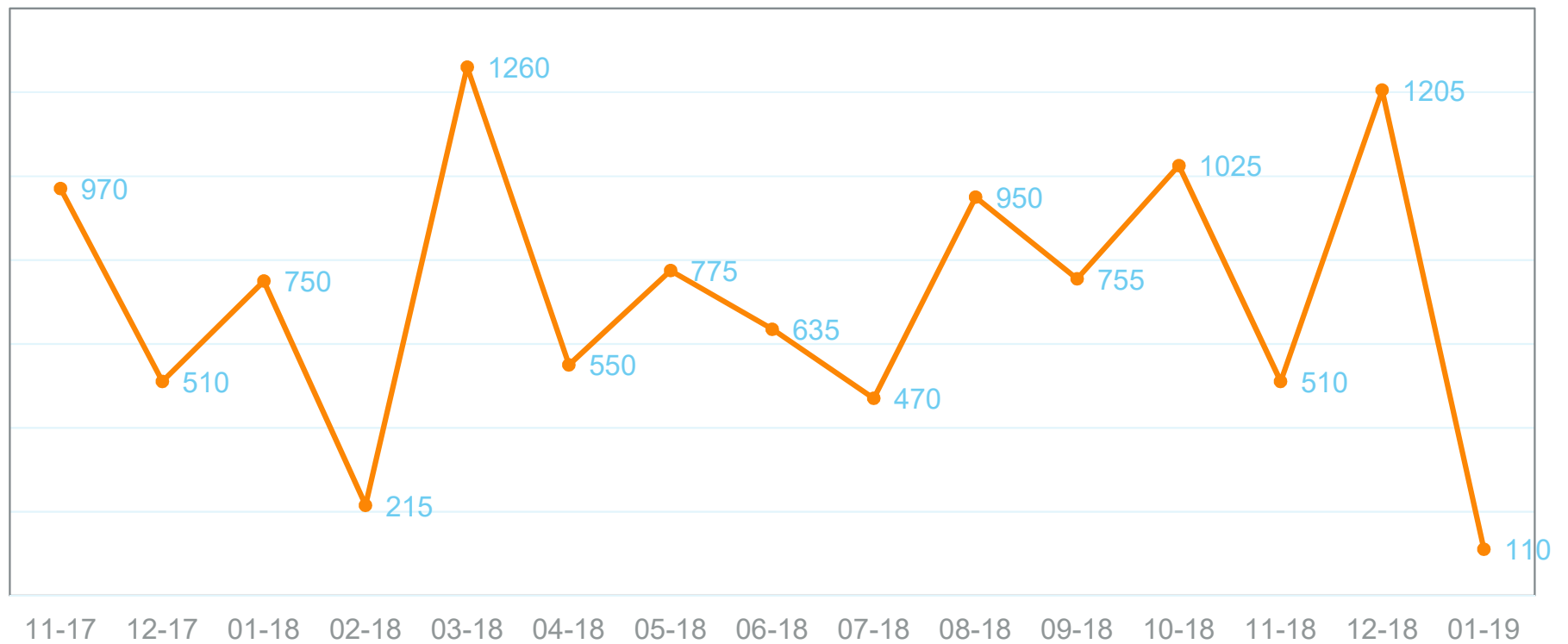
Continuous with seasonality

Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18	12-18	01-19
Paper:Ptr/Copier Ream/A4	SS10_4101	Final History	Continuous with seasonality	20	37	25	18	27	30	30	21	16	47	25	37	25	34	35

Representation of Continuous with seasonality
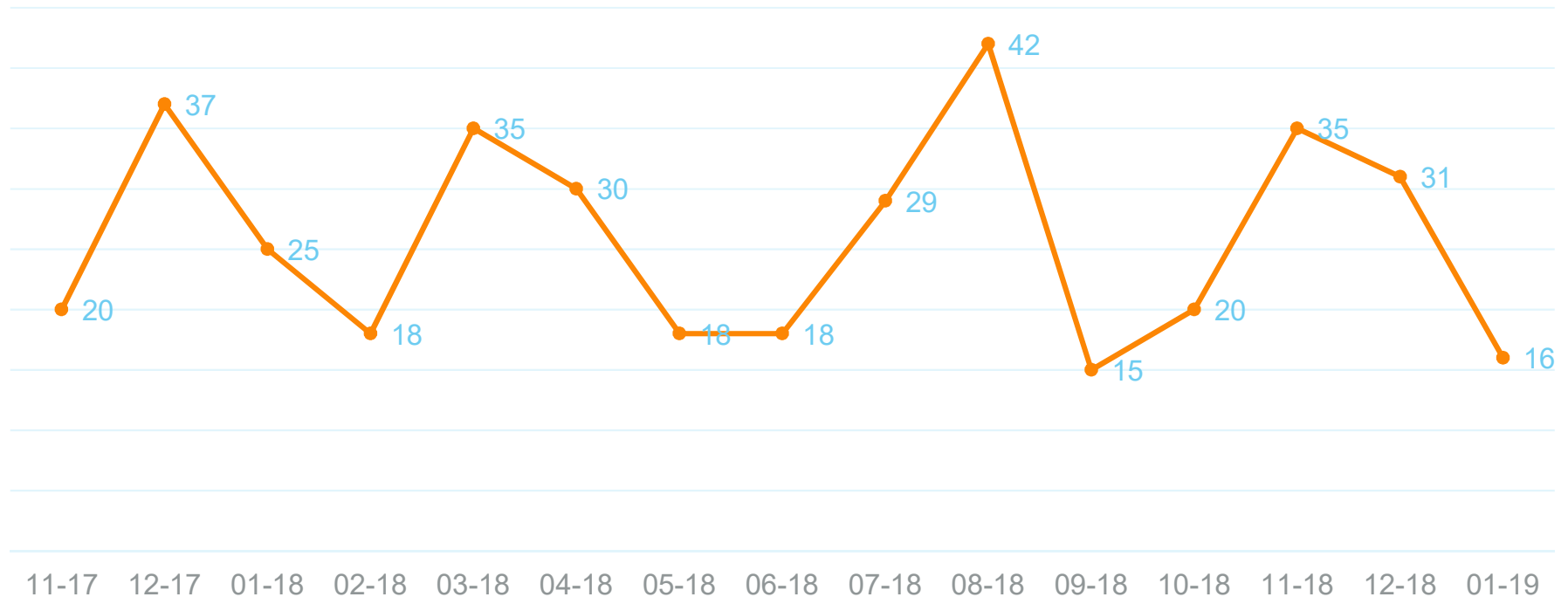
Every 2 or 3 months there is a raise or a fall

8

Continuous with trend

Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18	12-18	01-19
Paper:Ptr/Copier Ream/A4	SS10_4106	Final History	Continuous with trend	145	150	200	220	240	260	280	300	340	350	380	400	450	460	460

Representation of Continuous with trend
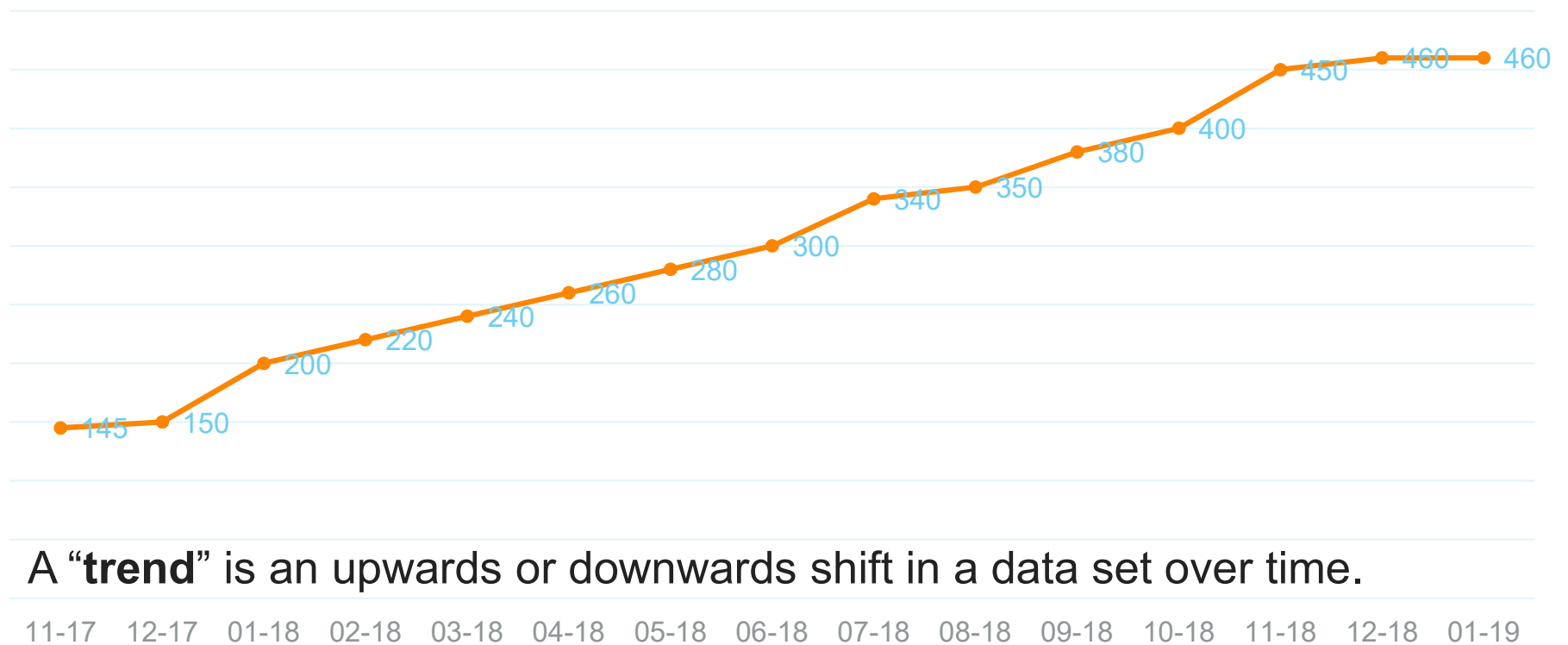
A "**trend**" is an upwards or downwards shift in a data set over time.

11-17 12-17 01-18 02-18 03-18 04-18 05-18 06-18 07-18 08-18 09-18 10-18 11-18 12-18 01-19

Continuous with seasonality and trend

Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18	12-18	01-19
Paper:Ptr/Copier Ream/A4	SS10_4111	Final History	Continuous with seasonality and trend	56	38	42	29	87	55	50	53	102	70	60	180	102	80	151

Representation of Continuous with seasonality and trend
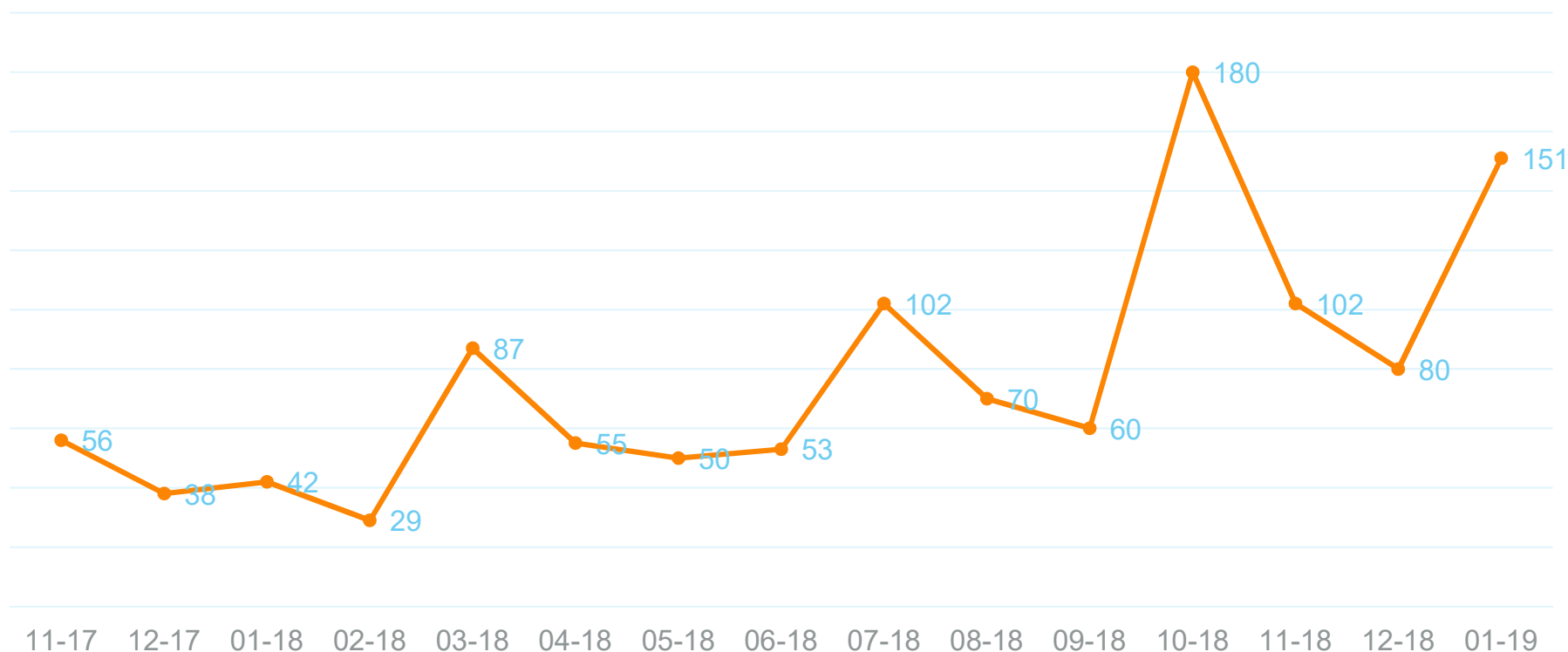
Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18	12-18	01-19
Paper:Ptr/Copier Ream/A4	LB10_4101	Final History	Intermittent		312	1110	665	373						60	40	55		

Representation of Intermittent
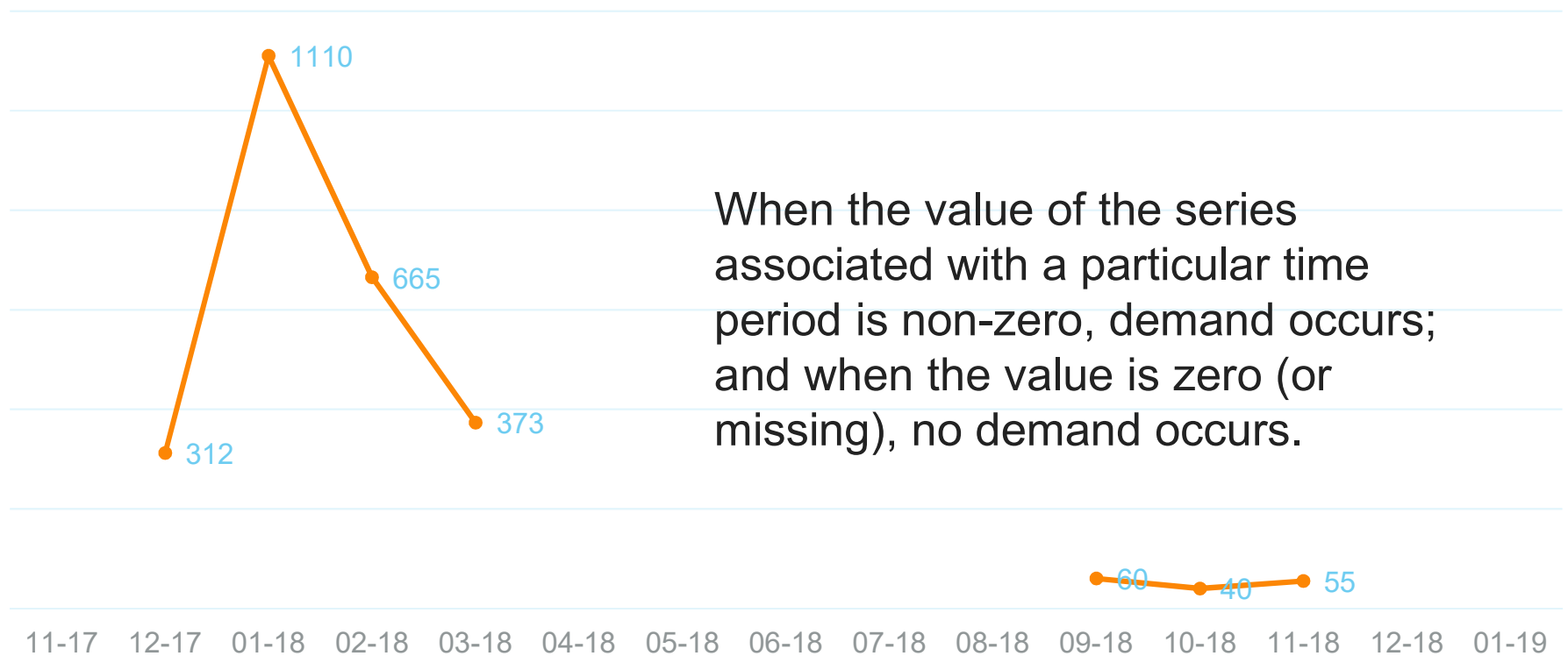
When the value of the series associated with a particular time period is non-zero, demand occurs; and when the value is zero (or missing), no demand occurs.

11-17 12-17 01-18 02-18 03-18 04-18 05-18 06-18 07-18 08-18 09-18 10-18 11-18 12-18 01-19

Intermittent with seasonality

Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18	12-18	01-19
Paper:Ptr/Copier Ream/A4	SS10_4112	Final History	Intermittent with seasonality						28	80	5	10	51	8	70	55	40	100

Representation of Intermittent with seasonality
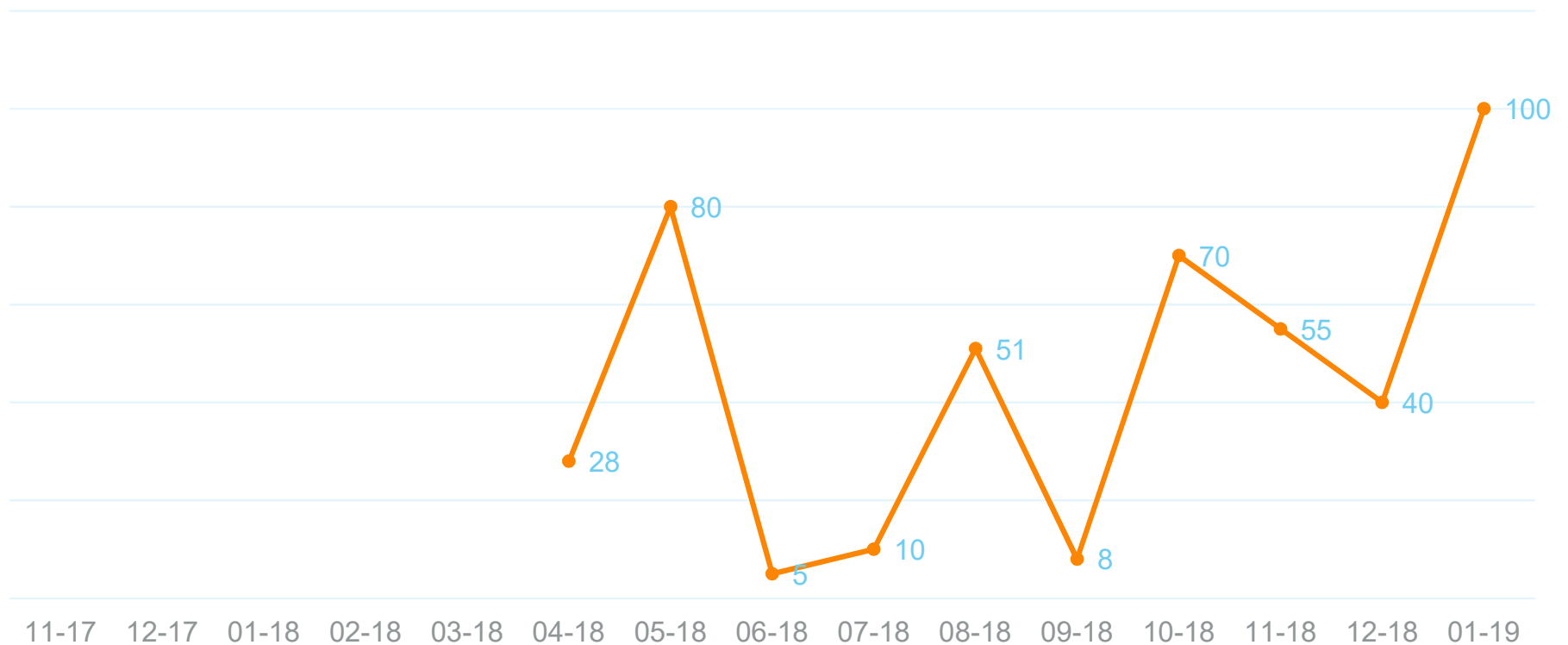
The non-zero period follow a seasonality pattern.

3.
Cleansing
Data

Final History
Only Time Series classified
as Continuous

Input →

3. Data Cleansing

1st Substitute Missing Value
(Continuous)

2nd Outlier Correction
(Continuous)

Cleansed History

Auto Cleansing template
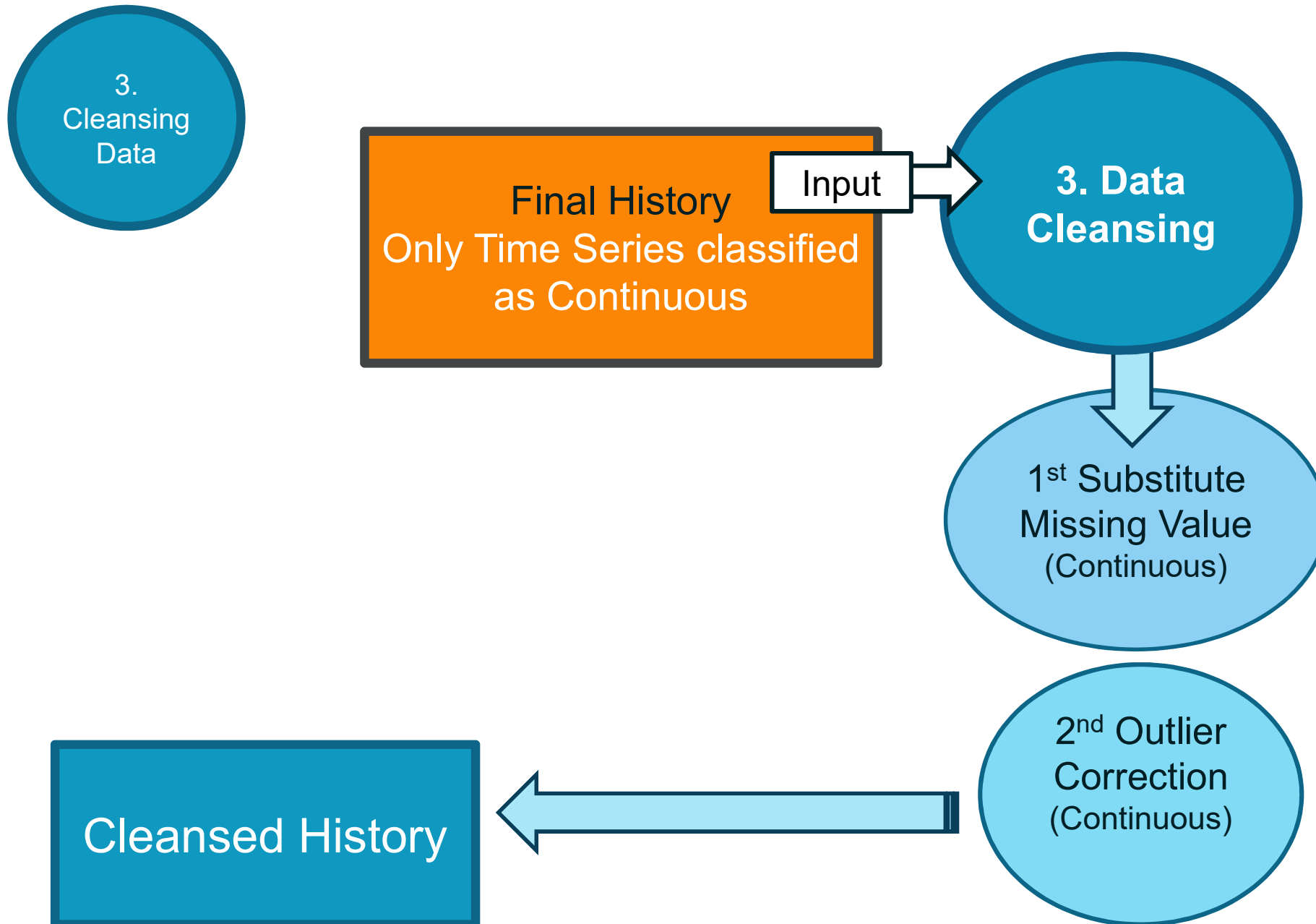
Auto Cleansing

❖ Automatically prepares the data classified as Continuous for generating Statistical Forecast.

Data Cleansing template

SCREEN SHOT WILL BE REPLACED WITH NEW ONE
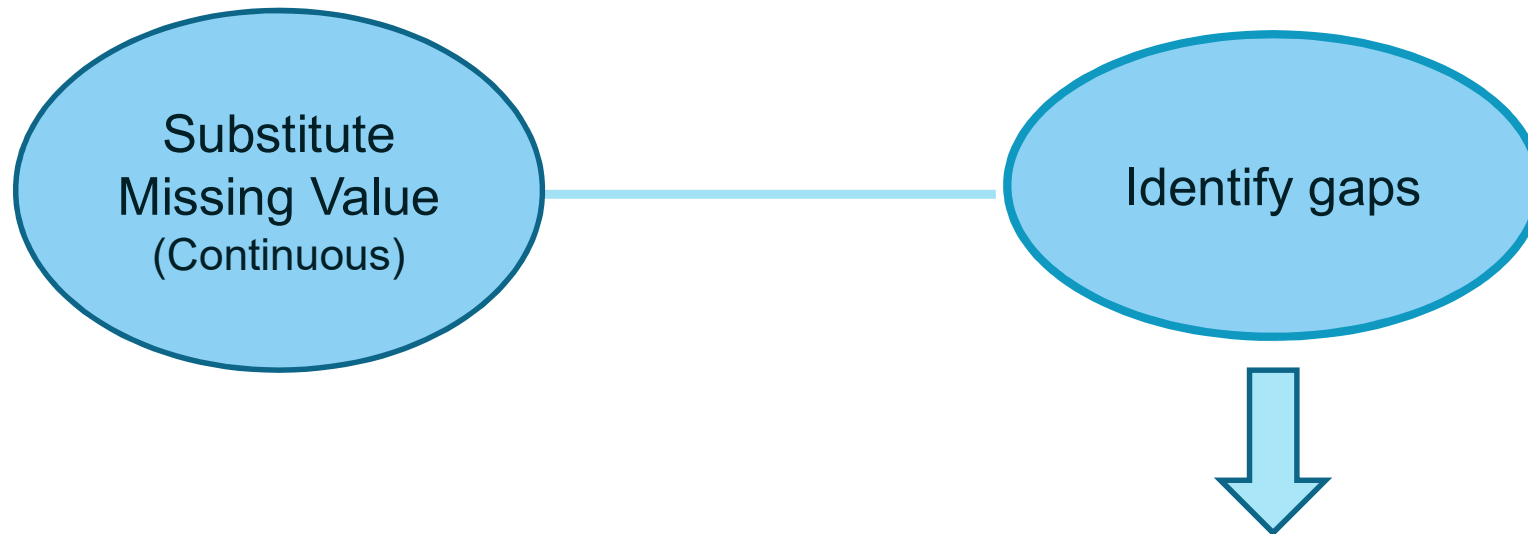
1st Substitute Missing Value (Continuous)

➤ The purpose of this pre-processing step is to identify gaps in time-series and replace them. Gaps will be replaced with:

❖ Mean - Configured

ONLY APPLICABLE TO CONTINUOUS TIME SERIES

Substitute Missing Value (Continuous) —— Identify gaps

Replace gaps with the Mean Value

Mean value: the arithmetic average

1. Sum of all the data values;
2. Divide by the number of the data values

x	μ
2	5
3	5
4	5
7	5
9	5

$$\mu = \frac{2+3+4+7+9}{5} = 5$$

Substitute missing values

Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18
Paper:Ptr/Copier Ream/A4	KE00_5101	Final History	Continuous	600	980	750	660	910	670		470	510	1025	790	620	800

Identify the gaps
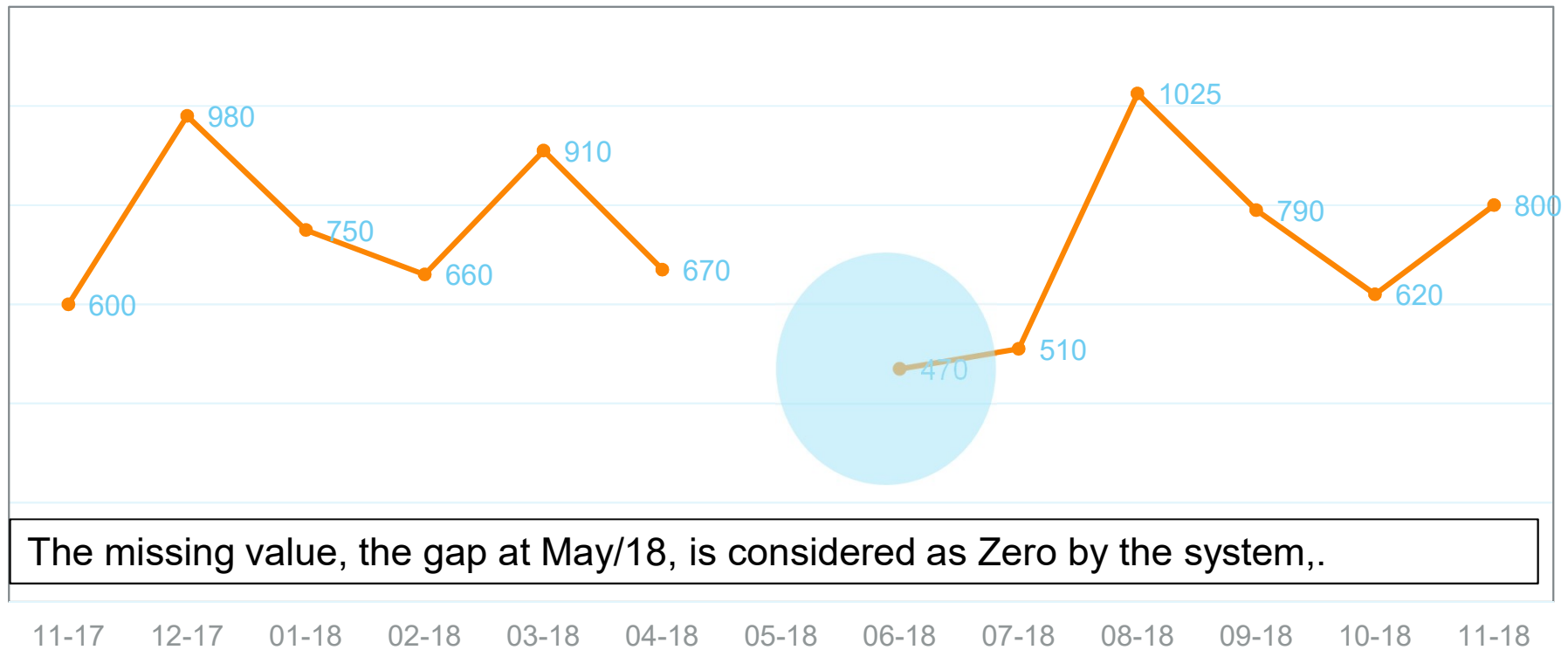
The missing value, the gap at May/18, is considered as Zero by the system,.

Substitute missing values

Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18
Paper:Ptr/Copier Ream/A4	KE00_5101	Final History	Continuous	600	980	750	660	910	670		470	510	1025	790	620	800

Identifying the mean value

Definition of mean: a number that represents the data set, the typica, the middle, the centerthe central tendency
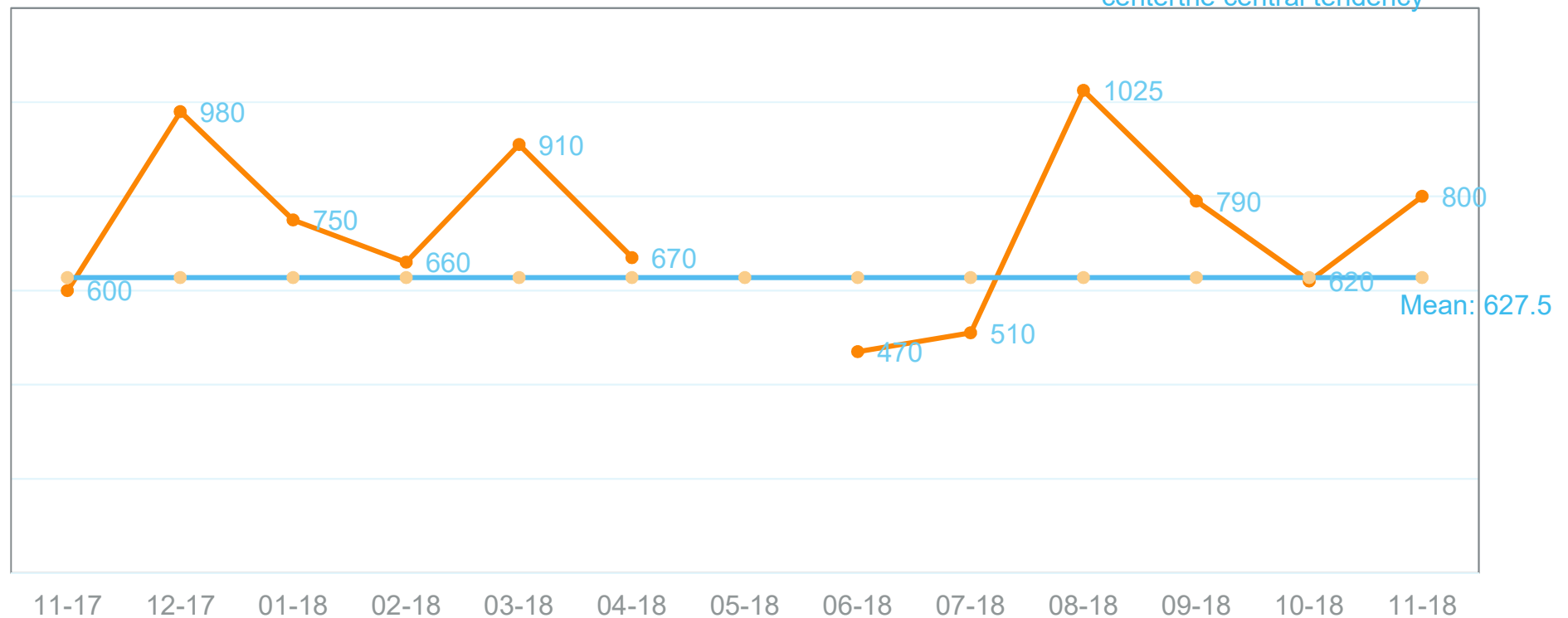
Mean: 627.5

1ˢᵗ Substitute missing values

Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18
Paper:Ptr/Copier Ream/A4	KE00_5101	Final History	Continuous	600	980	750	660	910	670	627.5	470	510	1025	790	620	800

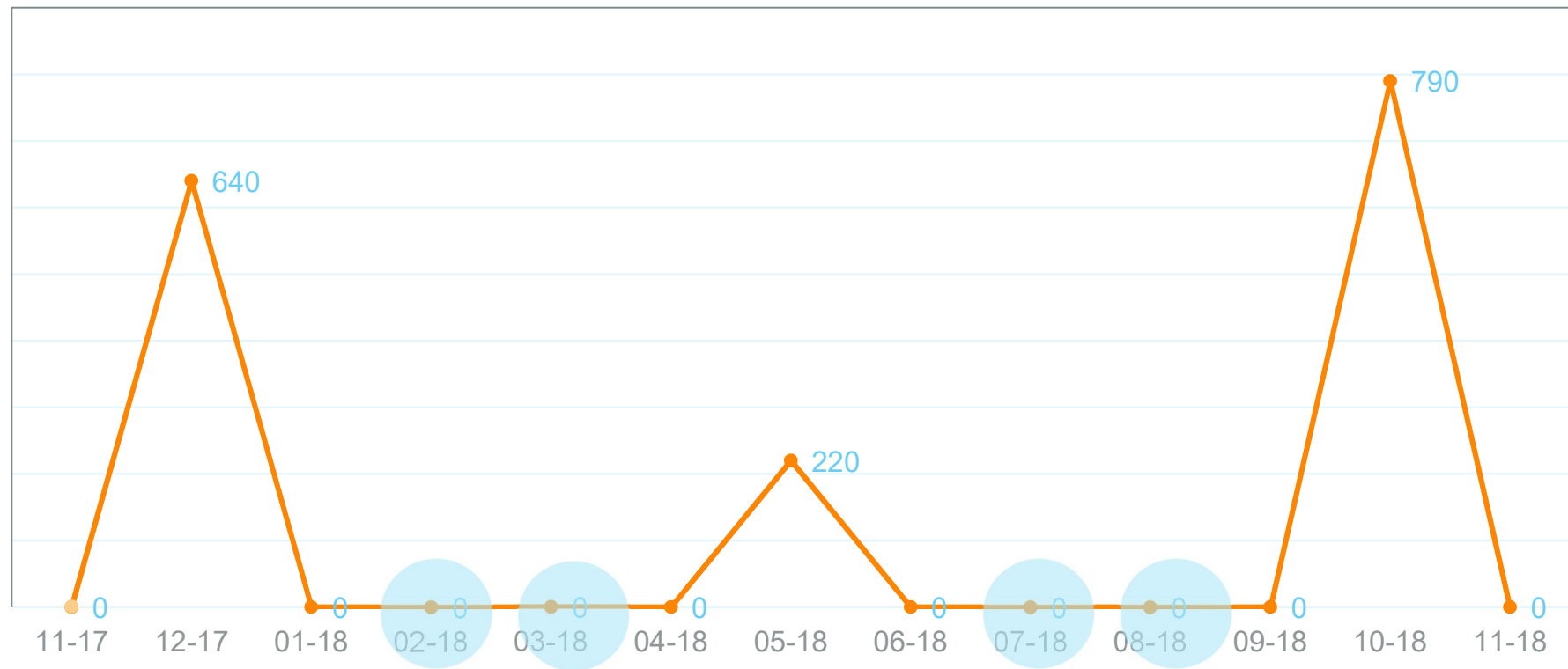## Replacing value with mean value

Mean: 627.5
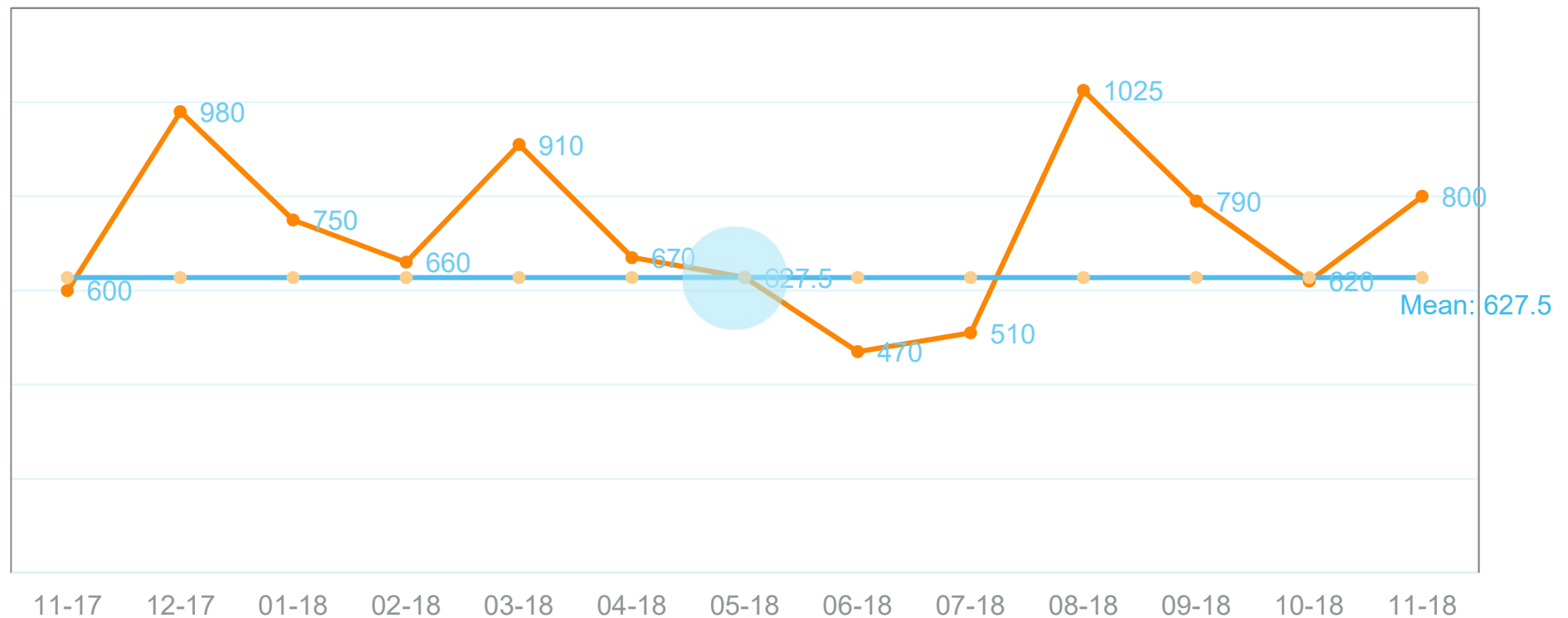
➢The process of identifying gaps and replace them applies only when Data Series are:

    o Continuous

    o Continuous with seasonality

    o Continuous with seasonality and trend
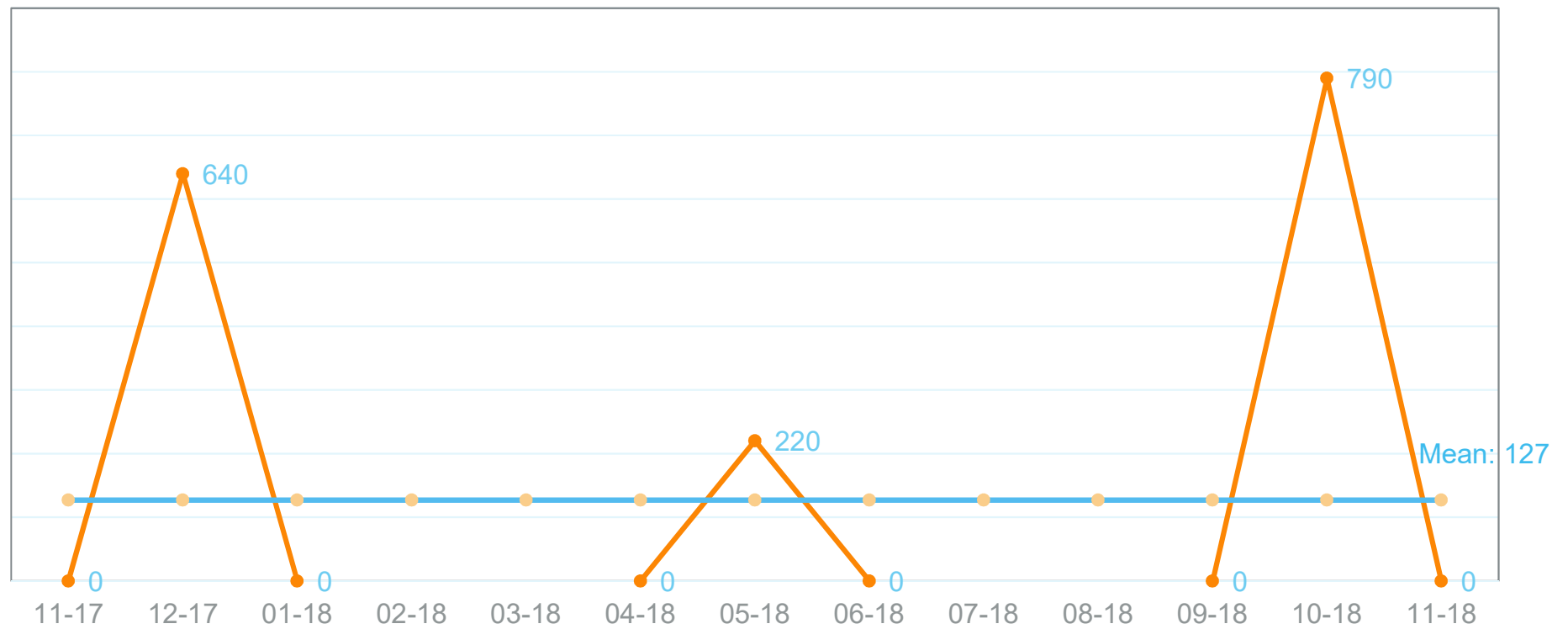
    o Continuous with trend

BUT NOT WHEN DATA CLASSIFICATION IS:
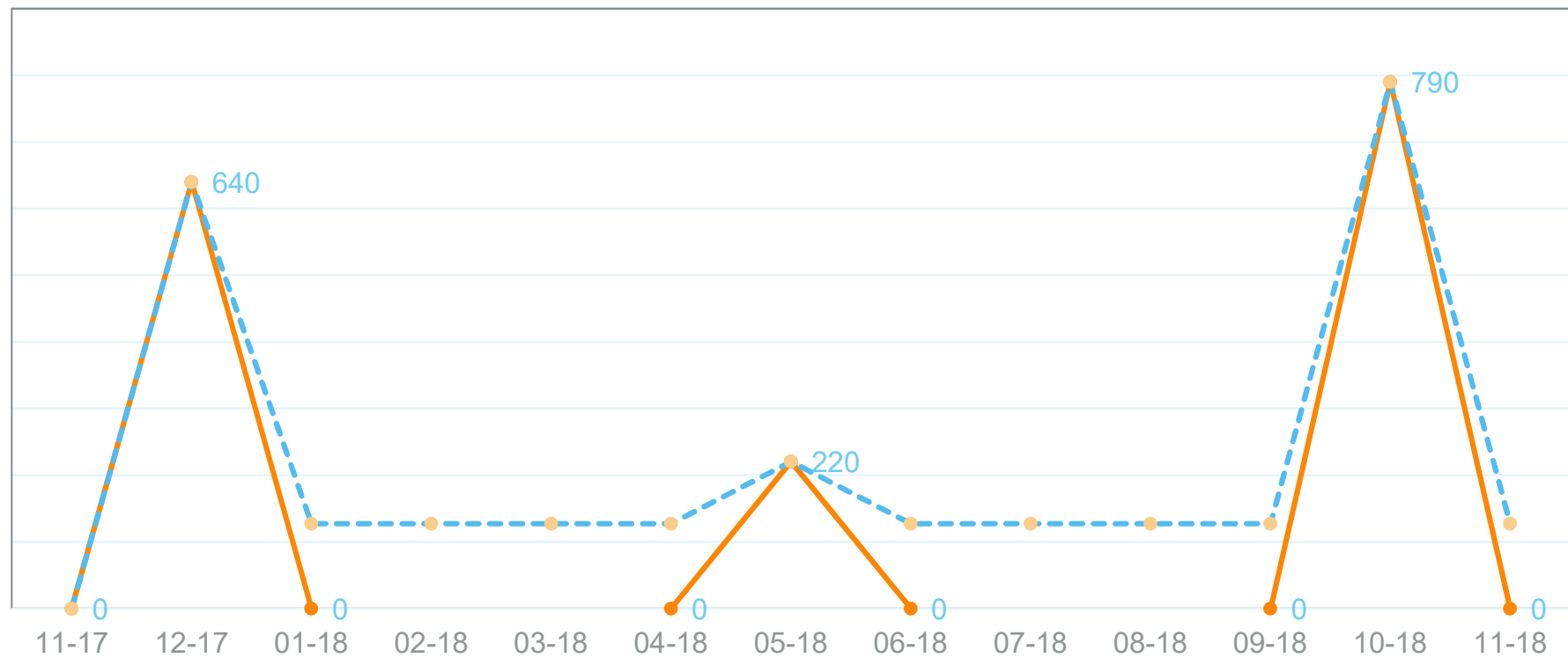
Intermittent

2nd Outlier Correction

➢ The purpose of this pre-processing step is to identify outliers in the time-series and replace them.

➢ It is done **only for Continuous Data** and not for Intermittent Data.

What is an Outlier

Value, a data point of a series, that differs significantly from the other observed values of the same series.

How we identify it

We use the Variance and Standard deviation. Variance measures how far a set of numbers is spread out from its average value (mean). Standard Deviation is the mean of the variance of all numbers in the series.

How the system acts

Replacing the outlier with a tolerance value to be able to prepare a forecast (upper/lower bounds).

Variance ð$_2$

1. Calculate the Mean Value;

2. For each of the data points, find the difference between the data point and the Mean;

3. Square each result. Square makes all the difference positive numbers, so they don't cancel them out.

4. Summing up all the differences and divide it by the number of data points.

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{n}$$

$$= \frac{(x_i - \mu)^2 + (x_i - \mu)^2 + ... + (x_n - \mu)^2}{n}$$

X$_i$ = Data points

μ = Mean Value

n = number of data points

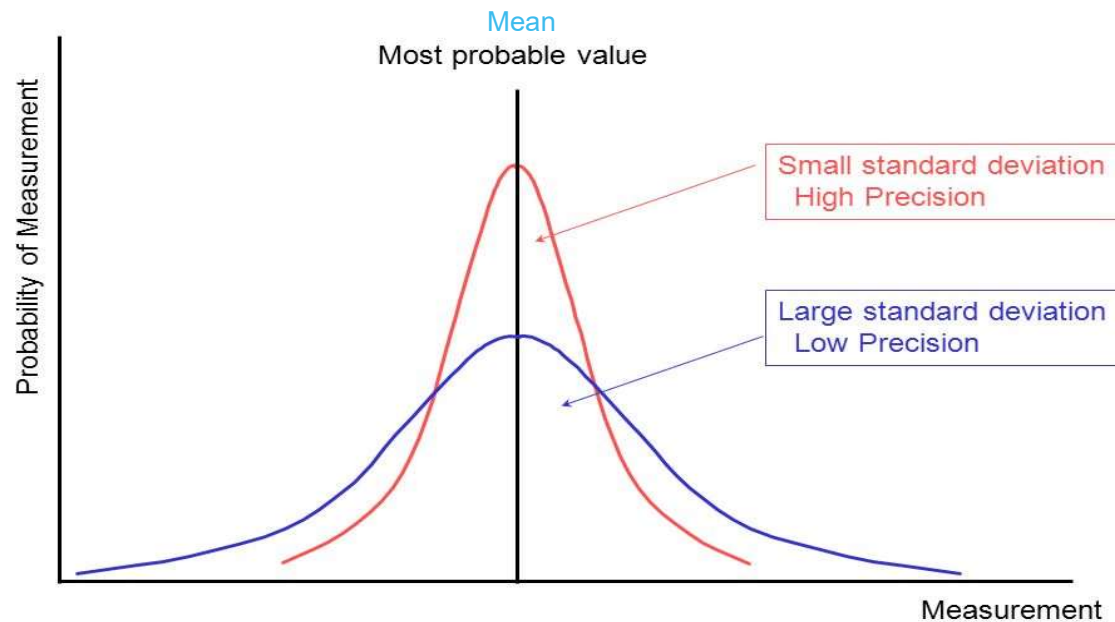| X	μ	X - μ	(X - μ)2
2	5	-3	9
3	5	-2	4
4	5	-1	1
7	5	2	4
9	5	4	16

$$\delta = \frac{9+4+1+4+16}{5}^2 = 6.8$$

ð =Standard Deviation: Square root of the variance = 2.6

26

Outlier Correction

What SD tells us?

How close the values in a data set are to the Mean value

Small Standard Deviation indicates a small variability for a data set. There are lots of values close to the Mean, which makes the distribution of data less spread out.

Large Standard Deviation indicates a big variability. Thera are lots more value that are farther from the Mean, which makes the distribution of data more spread out. The data are dispersed over a wide range of values.

Outlier Correction

Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18
Paper:Ptr/Copier Ream/A4	KE00_5101	Final History	Continuous	600	890	750	660	910	670	190	615	610	1390	790	620	800
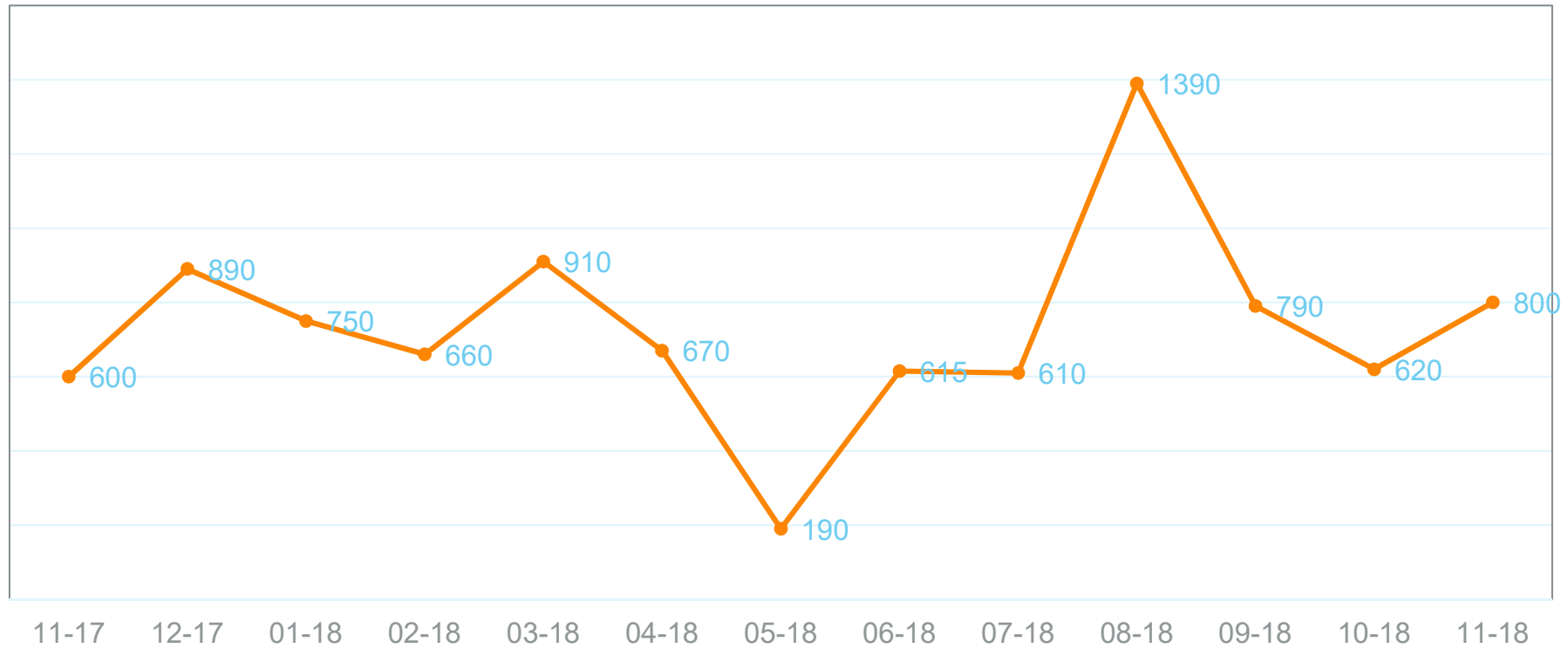
Outlier Correction

Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18
Paper:Ptr/Copier Ream/A4	KE00_5101	Final History	Continuous	600	890	750	660	910	670	90	615	610	1390	790	620	800

Identify the outlines: Mean, standard Deviation and Tolerance
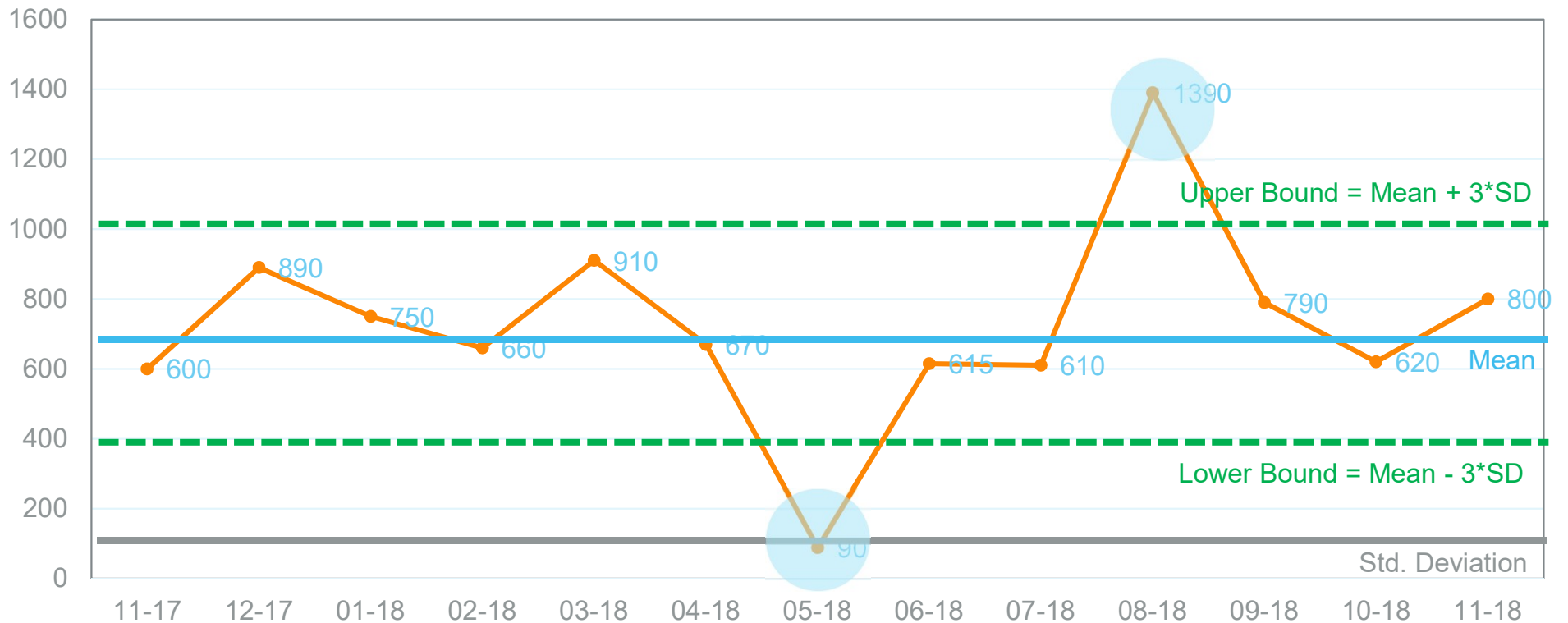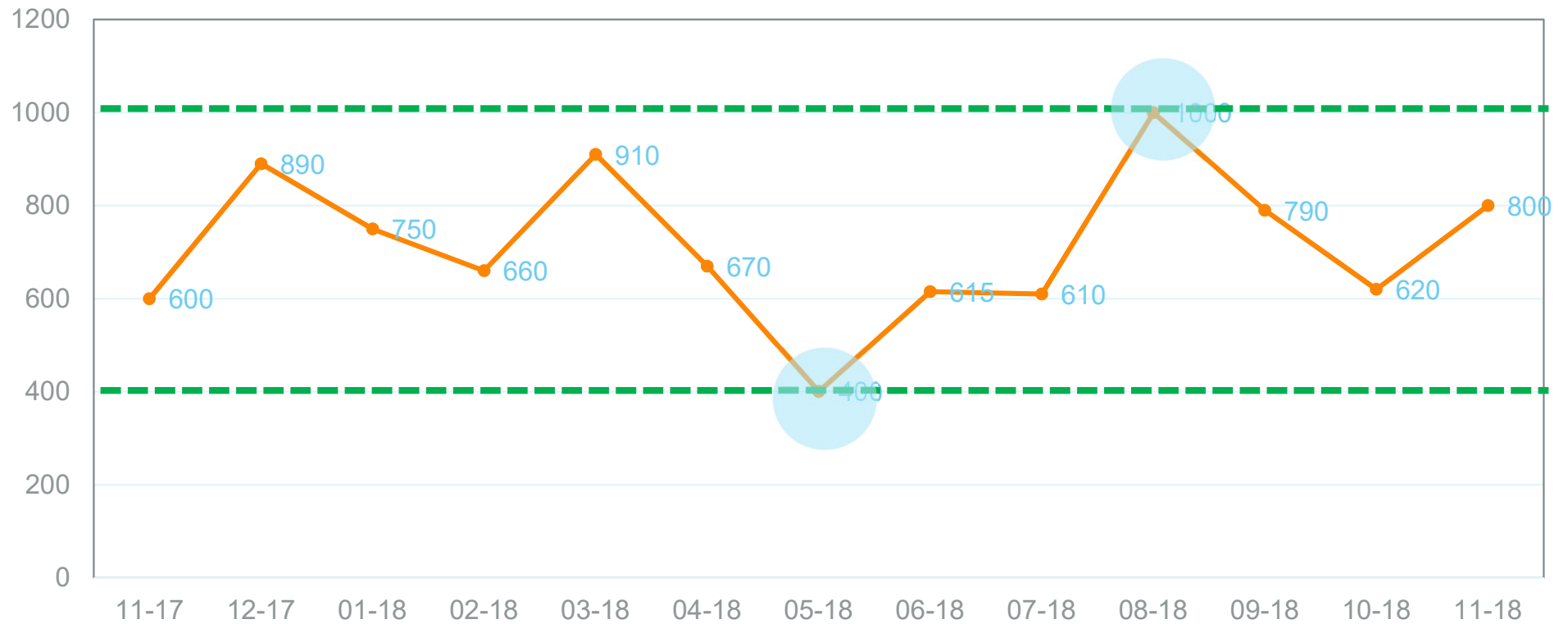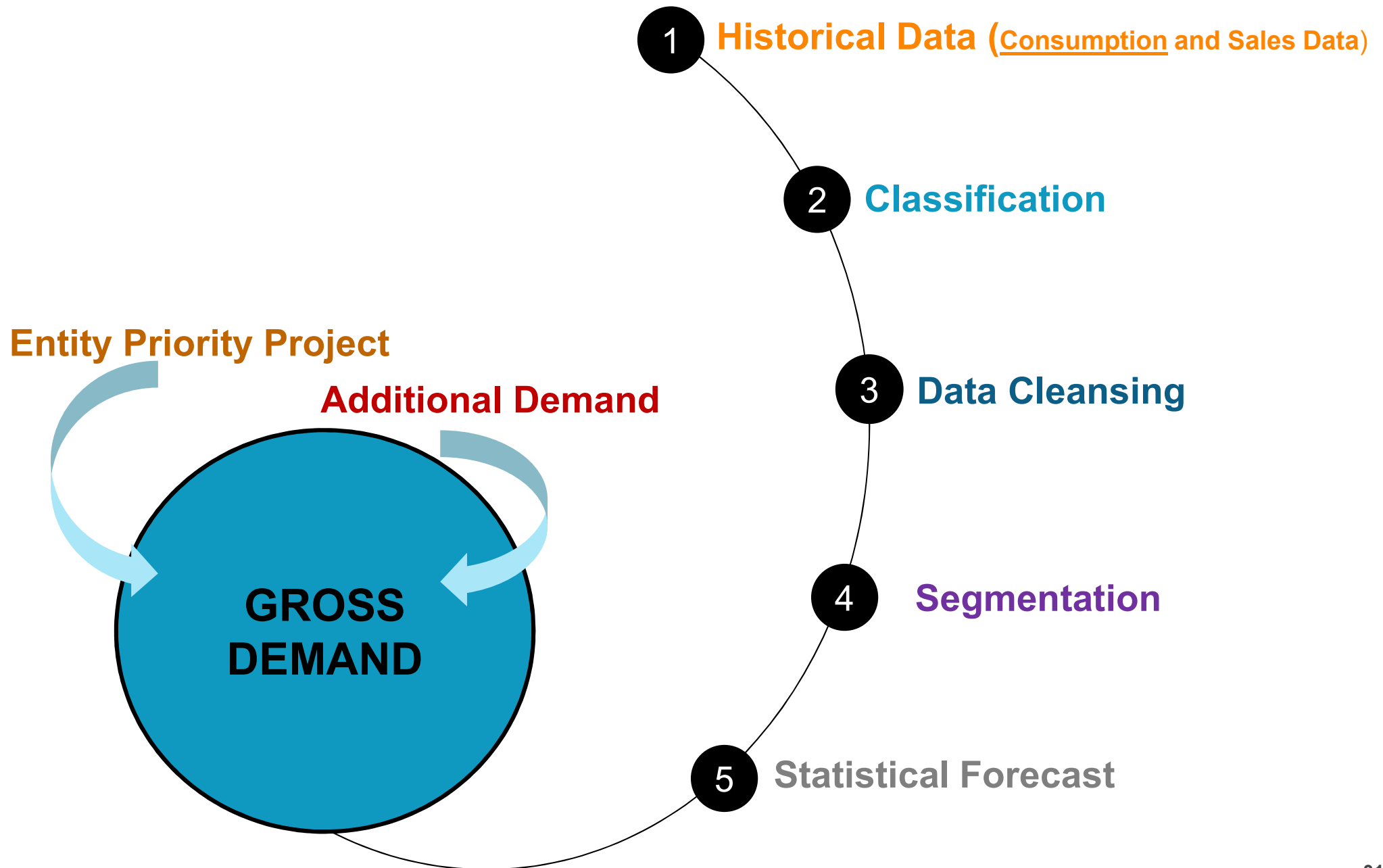

Tolerance values: Standard Deviation ± 3

Outlier Correction

Product	Sublocation	Key Figure	Time Series Classification	11-17	12-17	01-18	02-18	03-18	04-18	05-18	06-18	07-18	08-18	09-18	10-18	11-18
Paper:Ptr/Copier Ream/A4	KE00_5101	Final History	Continuous	600	890	750	660	910	670	400	615	610	1000	790	620	800

Replace the values

DEPARTMENT OF OPERATIONAL SUPPORT

1 **Historical Data (**<u>Consumption</u> **and Sales Data)**

2 **Classification**

3 **Data Cleansing**

4 **Segmentation**

5 **Statistical Forecast**

Entity Priority Project

Additional Demand

GROSS DEMAND

Thank you!