Data Science and Business Analytics

Data Science Case Studies
Lecture 4

Moscow
2025

Analytics in Retail & CPG. Demand Forecasting Problem

Alexey Romanenko

alexromsput@gmail.com

Agenda

- Course Intro
- Retail Chain Analytics
- E2E Demand Forecasting System Overview
  - Input data overview
  - Data preprocessing
  - Models training
  - Forecast postprocessing

What is inside Analytics in Retail & CPG

- 3 lectures
  - ➢ Predictive analytics in Supply Chain
  - ➢ Prescriptive analytics in Supply Chain
  - ➢ Optimization problems in Supply Chain
- 3 practical seminars

- HW (for 2 weeks)
  - ➢ Kaggle Competition about demand forecasting

- Graduated MIPT at 2011
- 15+ years in teaching
- Scientific publications
- International Conferences
- 15 years in Retail and CPG Analytics
- 10+ Clients
- 9 Big Projects
- 2 Years in Marketplace Analytics


LET'S GET
STARTED

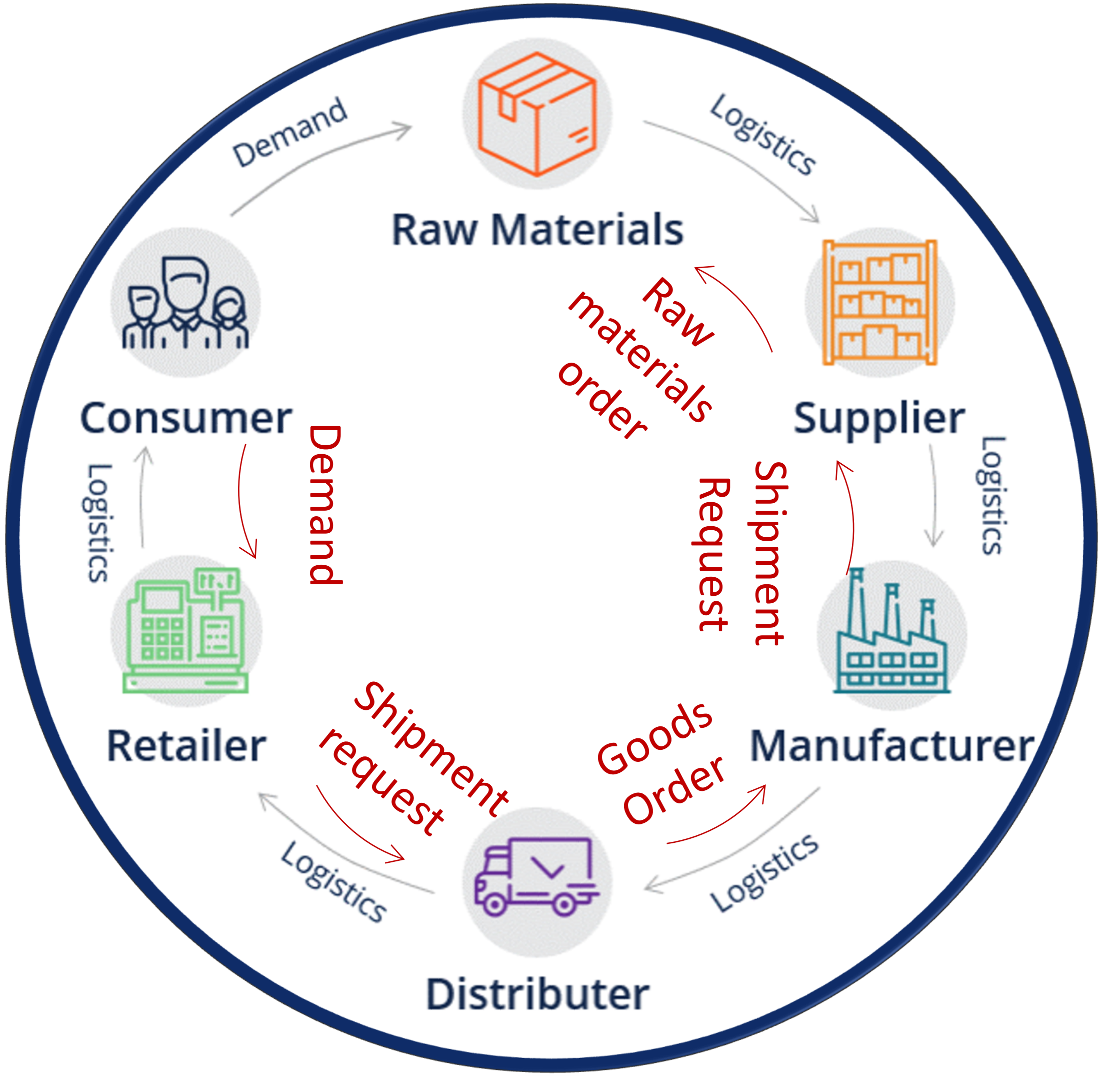
What is the Supply Chain?

What is the Supply Chain?

- Products are moved clockwise

What is the Supply Chain?

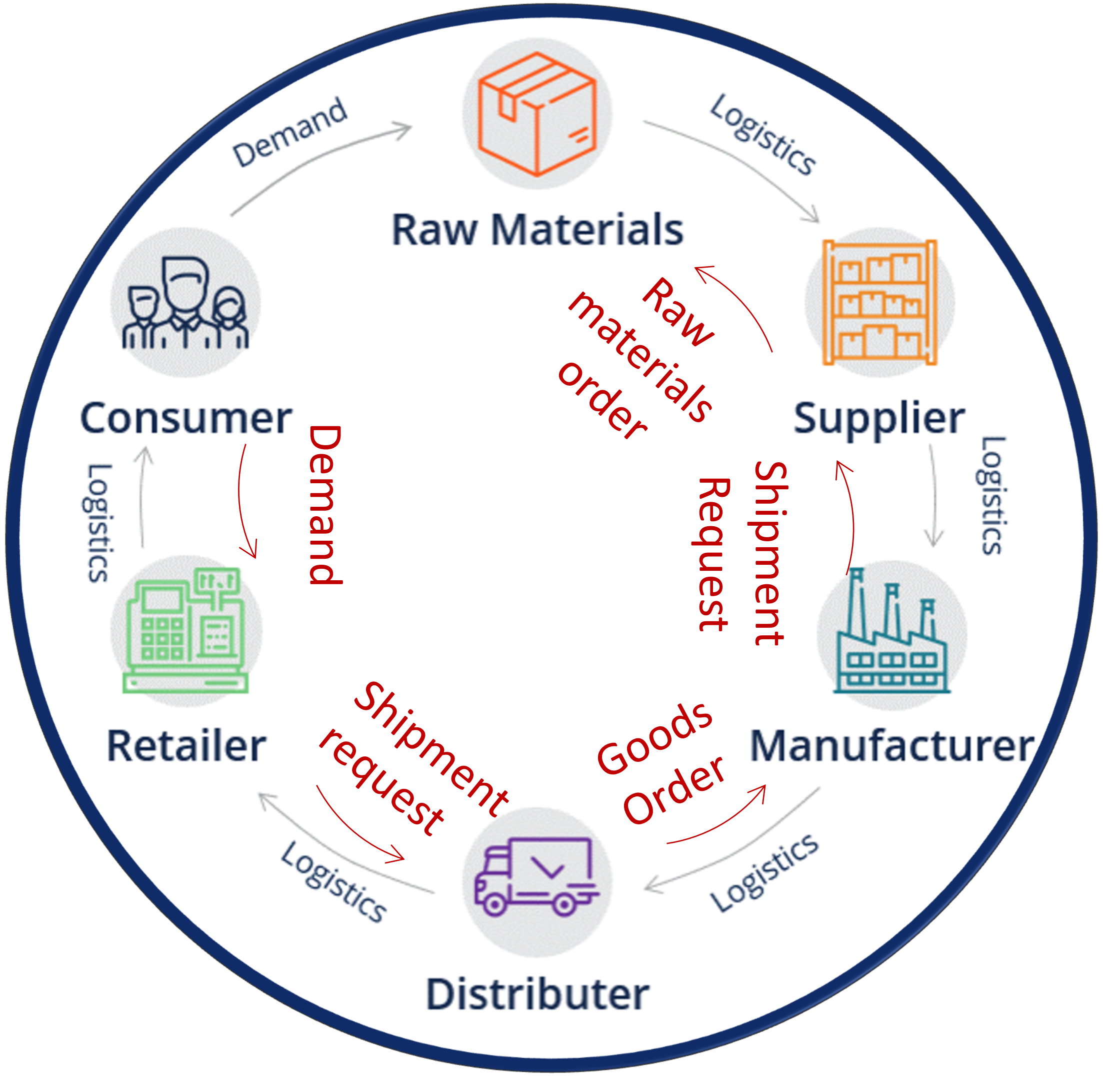
➢ Products are moved clockwise



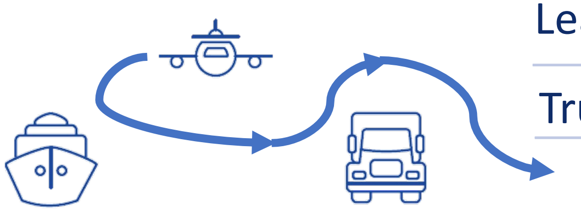
What is the Supply Chain?

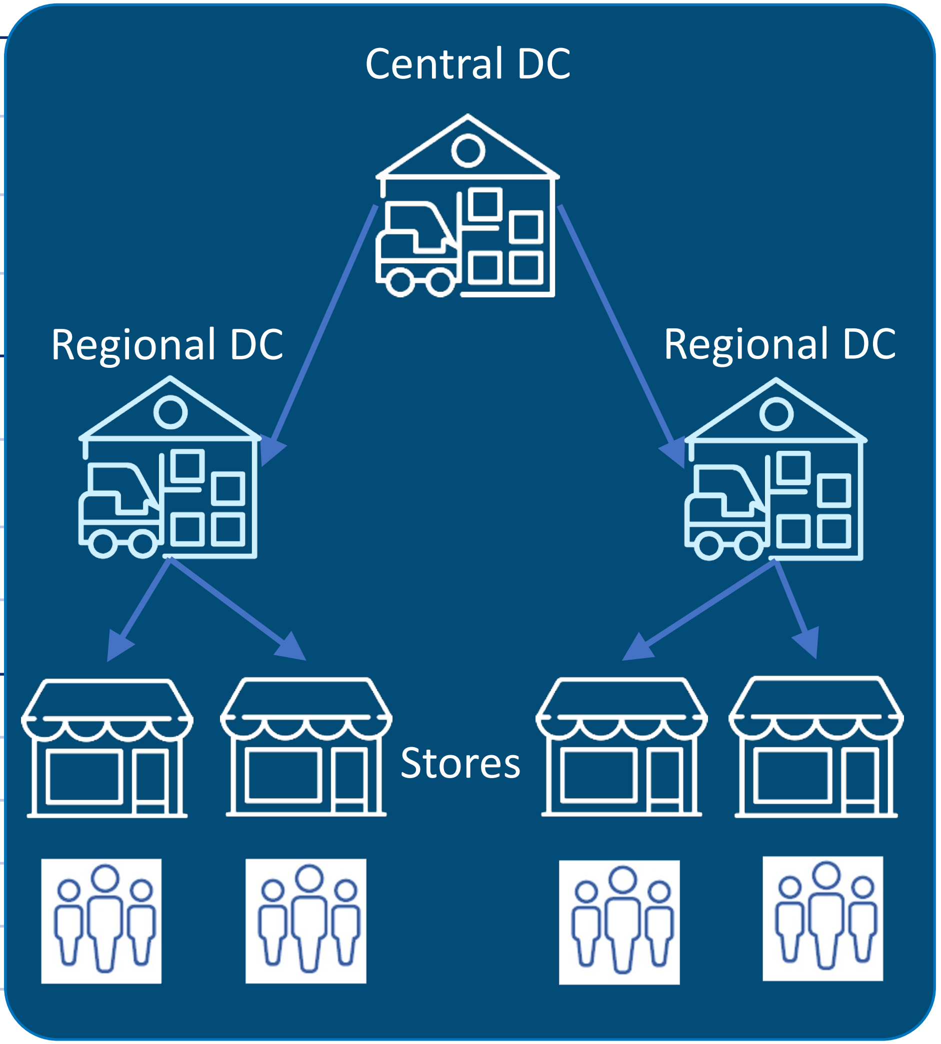
- Products are moved clockwise

- Decisions are made counterclockwise



Objects	Goods, Docs, Distribution Centers (DC),
	Warehouses, Cross-Docs
	Stores, Outlets
Logistics	Transportation routes
	Lead times, transportation cost
	Truck, Ship, Train,
Management	Buy quantity
	Merchandise Prices
	Contracts





Key questions of Retail Chain (7-R questions)

Key questions of Retail Chain

This could be 7-dimensional optimization problem!

How can people solve it?

Key questions of Retail Chain

It is splitted to 7 independent questions:

Categorical Management | Inventory Management | Buyers | Store Management | Logistic s | Marketing | Price Management

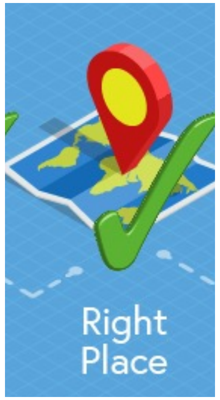
Idea of decomposition. Example

Assortment Optimization

Product List

Price Optimization

Place Optimization

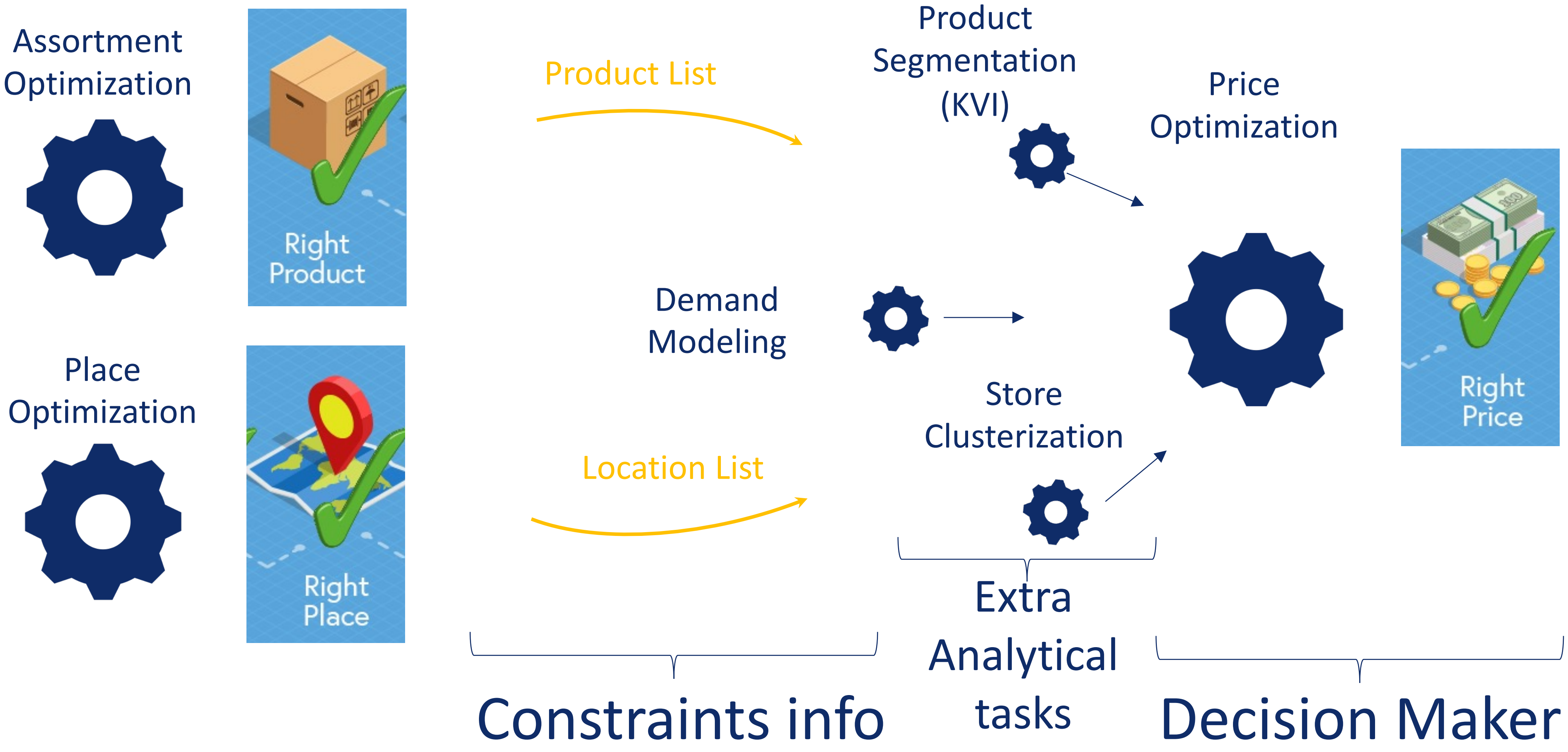


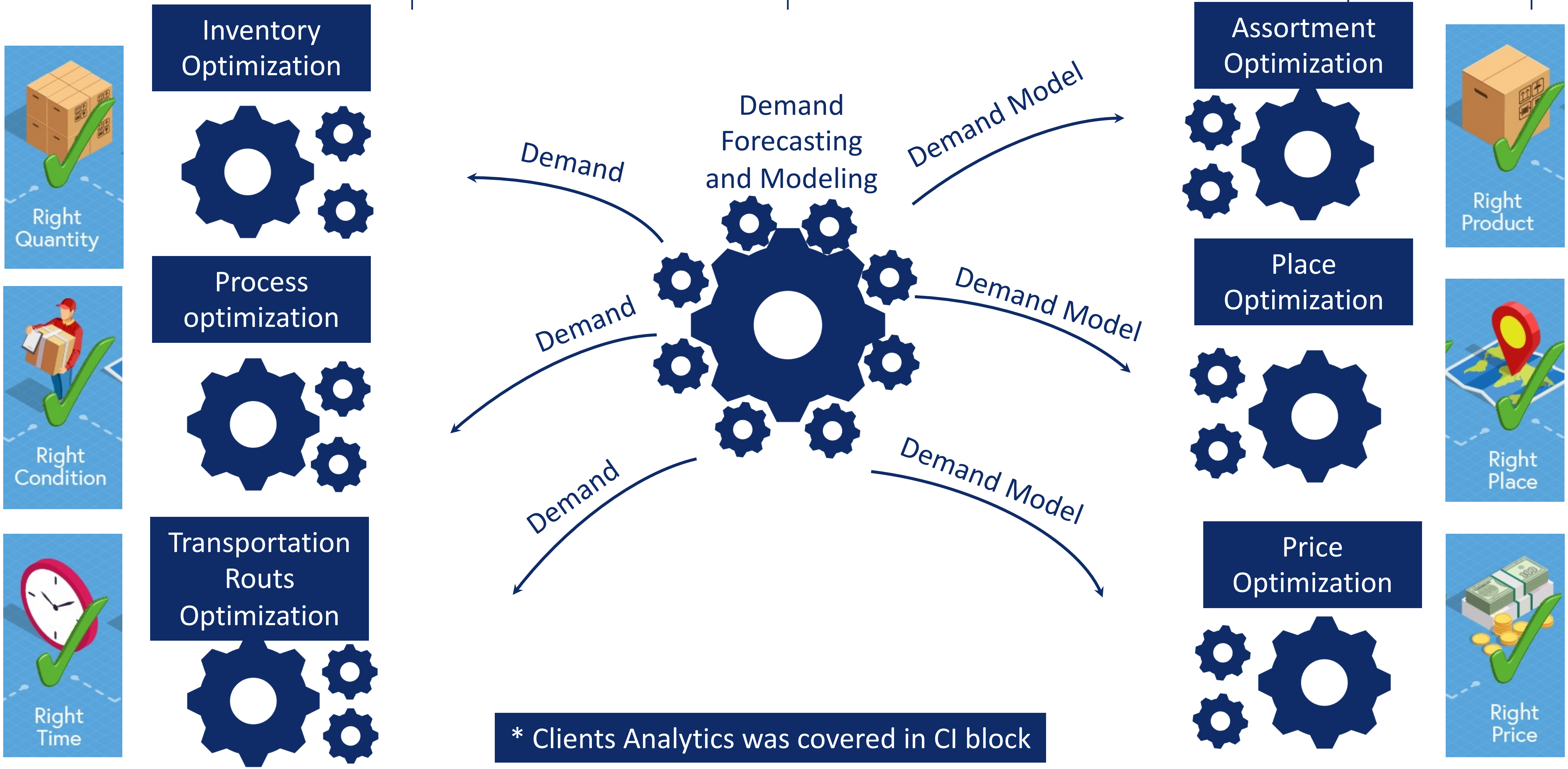
Location List

Constraints info

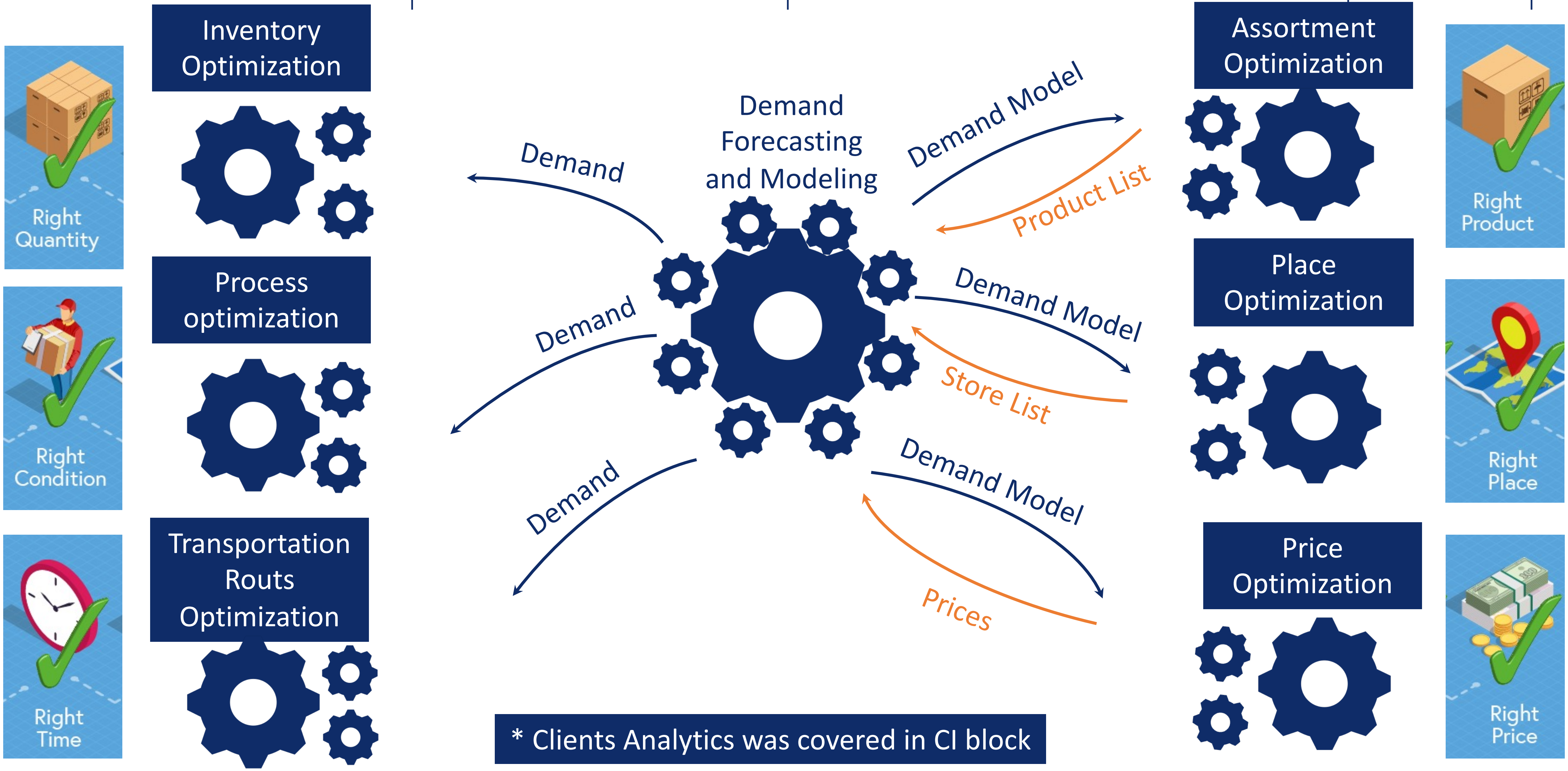
Decision Maker

Idea of decomposition. Example

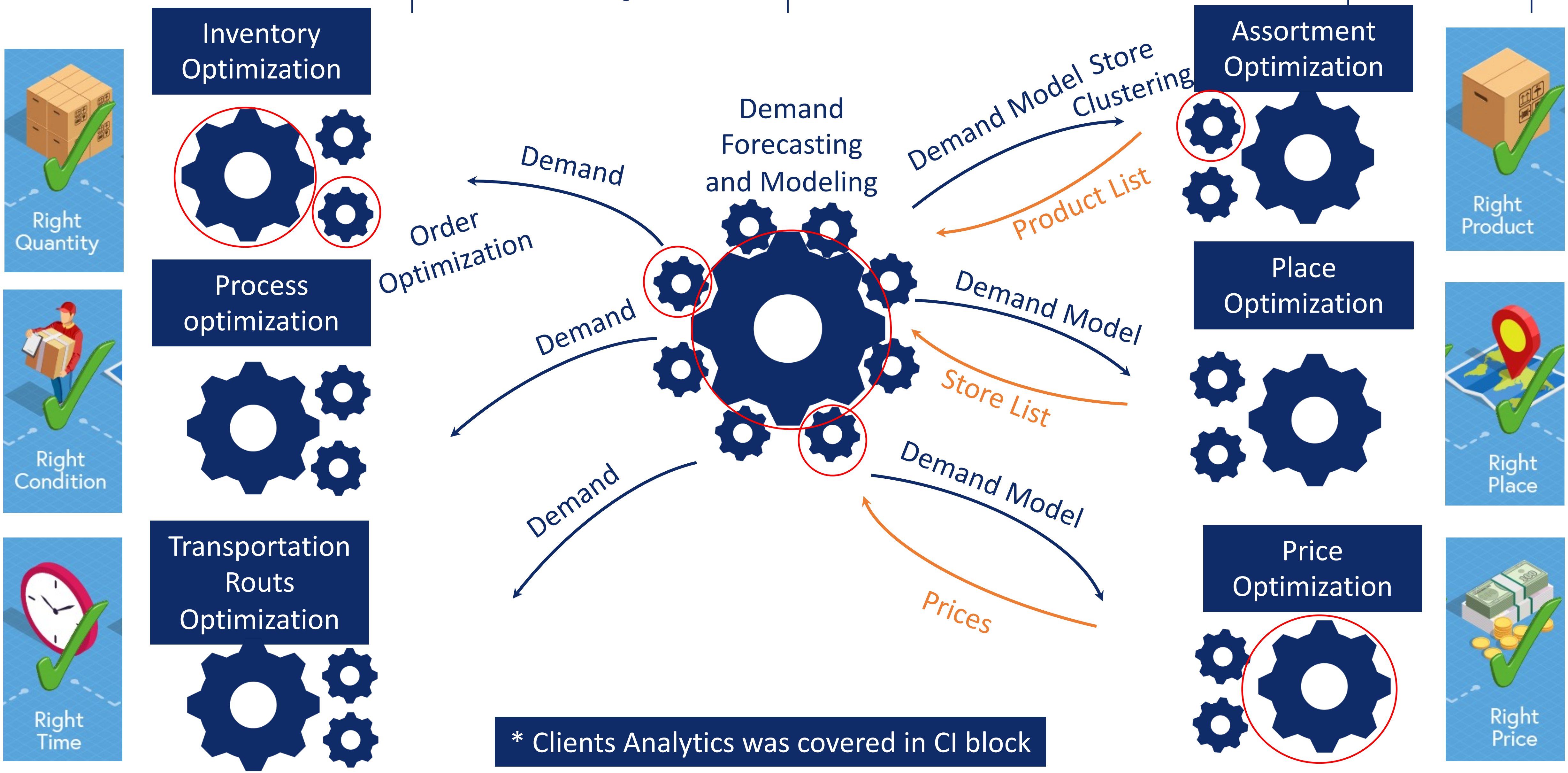




* Clients Analytics was covered in CI block



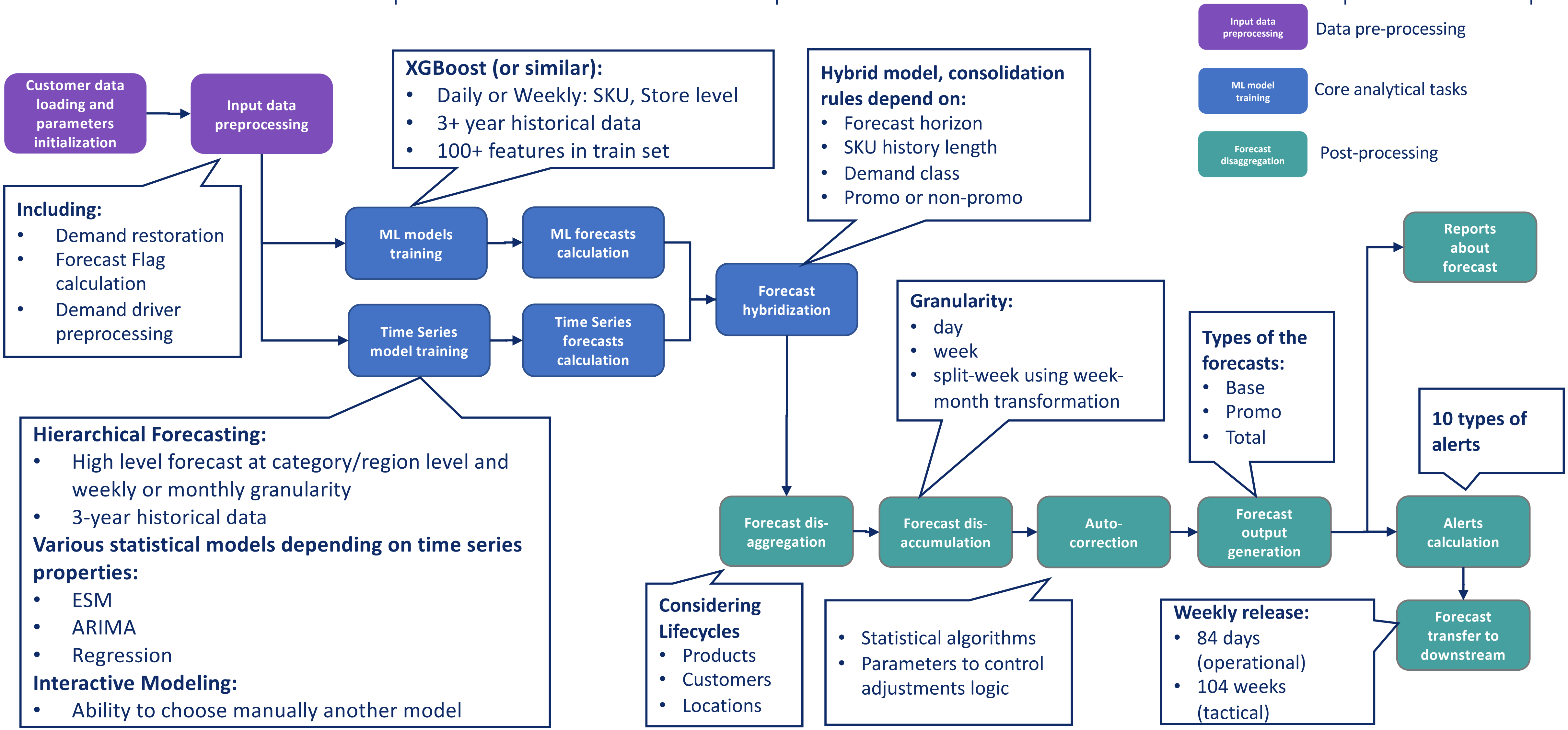
* Clients Analytics was covered in CI block



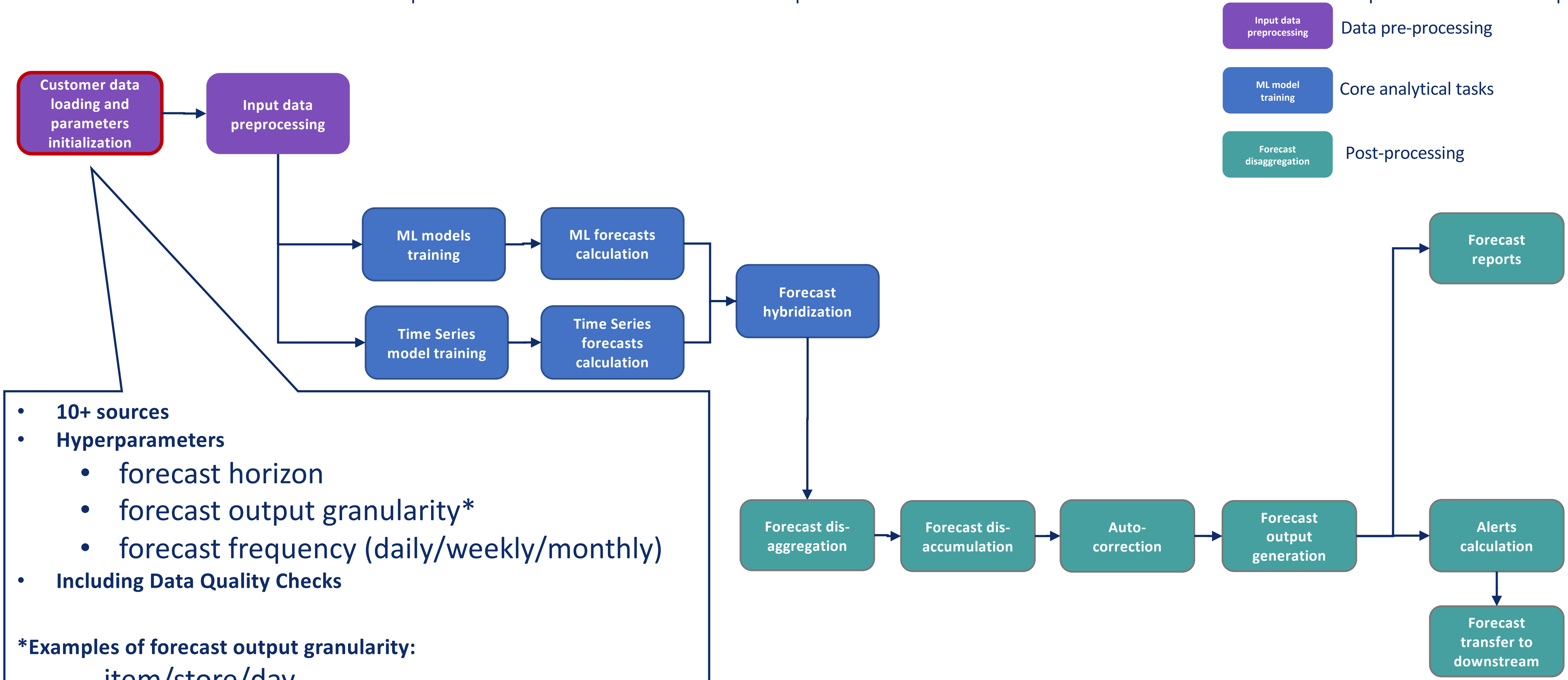
* Clients Analytics was covered in CI block

Demand Forecasting in Retail
https://www.mcompetitions.unic.ac.cy/

E2E forecasting flow overview



E2E forecasting flow overview



- **10+ sources**
- **Hyperparameters**
  - forecast horizon
  - forecast output granularity*
  - forecast frequency (daily/weekly/monthly)
- **Including Data Quality Checks**

***Examples of forecast output granularity:**

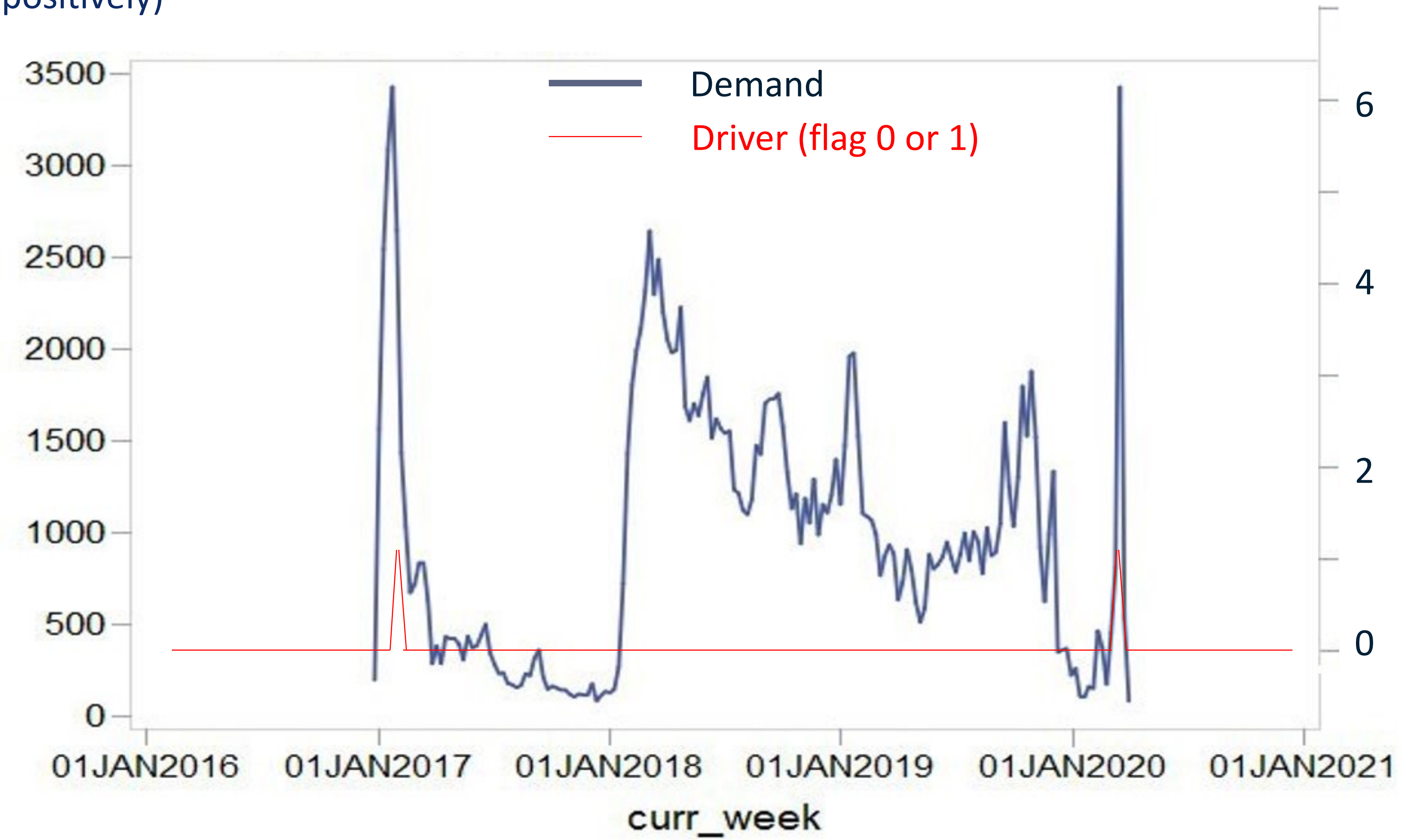
- item/store/day
- category/region/week
- catstream/country/month

store_id	product_id	Date	s_qty
1	1	01JAN2016	3
1	1	02JAN2016	9
1	1	03JAN2016	6
1	1	04JAN2016	3
1	1	05JAN2016	10
1	1	06JAN2016	4
1	1	07JAN2016	6
1	1	08JAN2016	8
1	1	09JAN2016	10
1	1	10JAN2016	2
1	1	11JAN2016	9
1	1	12JAN2016	3
1	1	13JAN2016	6
1	1	14JAN2016	9
1	1	15JAN2016	4
1	1	16JAN2016	10
1	1	17JAN2016	5
1	1	18JAN2016	3
1	1	19JAN2016	10
1	1	20JAN2016	10
1	1	21JAN2016	4
1	1013	22JAN2016	0.0

Original data, s_qty = sales quantity

What else is needed for forecasting?

Demand Driver is an independent (causal) variable that affects on demand quantity (negatively or positively)

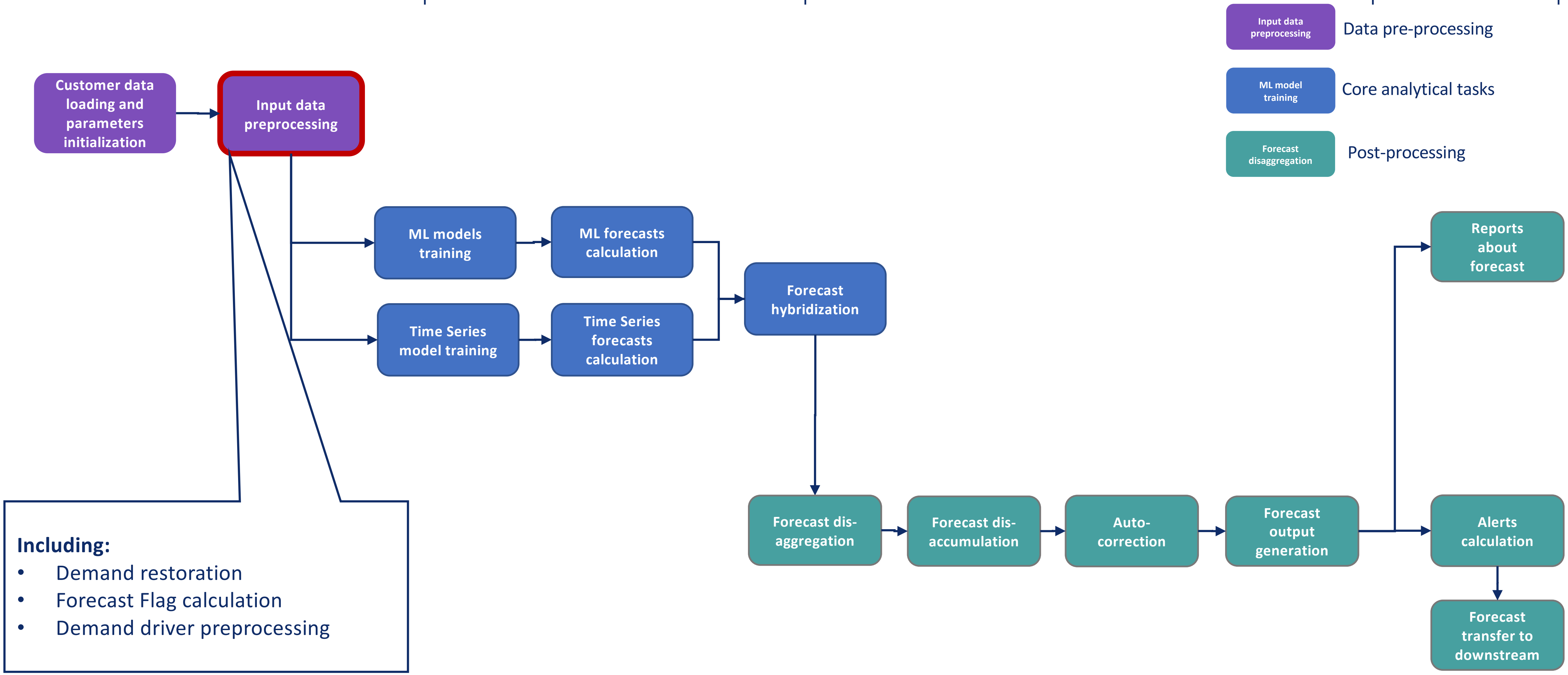


- Price info
- Promo events (and promo mechanics)
- Quantity of sales consultant in Store
- Product Attributes
- Store Attributes
- Calendar Events

What are the most Significant Calendar Events in Russia?

- New Year
- 23 February, 8 March
- Easter
- 1-2 May, 9 May
- 1 September

IA Table name	Description	Type
PRODUCT	Stores basic information about products, including product hierarchy.	Mandatory
LOCATION	Stores information about locations (location hierarchy).	Mandatory
CUSTOMER	Stores basic information about customers (distributors, retailers etc.), including customer hierarchy.	Mandatory
DISTR_CHANNEL	Stores basic information about distribution channel, including distribution channels hierarchy.	Mandatory
SELL_IN	Stores information about historical sales (e.g. shipment data).	Mandatory
SELL_OUT	Stores information about historical sales of the customers (POS data).	Mandatory
ASSORT_MATRIX	Stores information about product matrix, to define assortments that should be forecasted by the solution.	Critical
PRODUCT_LIFE	Stores information about products lifecycle. Aimed at covering the situations when the history of sales should be inherited by one product from the other.	Critical
LOCATION_LIFE	Stores information about locations lifecycle. Aimed at covering the situations when the history of sales should be inherited by one location from the other.	Critical
CUSTOMER_LIFE	Stores information about customers lifecycle. Aimed at covering the situations when the history of sales should be inherited by one customer from the other.	Critical
EVENTS	Stores information about calendar events.	Optional
PRODUCT_ATTR	Stores information about product attributes.	Optional
LOCATION_ATTR	Stores information about location attributes.	Optional
CUSTOMER_ATTR	Stores information about customer attributes.	Optional
DISTR_CHANNEL_ATTR	Stores information about distribution channel attributes.	Optional
STOCK	Stores historical information about inventories.	Optional
PROMO_TYPE	Stores information about types of promo events.	Optional
PROMO	Stores information about planned and historical promotions events.	Optional
PROMO_ATTR	Stores information about promo attributes.	Optional
PRICE	Stores information about planned and historical prices.	Optional
CROSS_ATTRIBUTES	Stores information about cross attributes. Aimed at covering the cases when the same product can have country or region or location specific attributes.	Optional


Input data preprocessing
Data pre-processing
ML model training
Core analytical tasks
Forecast disaggregation
Post-processing
Customer data loading and parameters initialization
Input data preprocessing
ML models training
ML forecasts calculation
Time Series model training
Time Series forecasts calculation
Forecast hybridization
Forecast dis-aggregation
Forecast dis-accumulation
Auto-correction
Forecast output generation
Reports about forecast
Alerts calculation
Forecast transfer to downstream
Including:
• Demand restoration
• Forecast Flag calculation
• Demand driver preprocessing


1000111
0110101
0101101
0101010

Sales VS Demand

Item 1

		+Stock							+Stock							+Stock
Daily Sales	…	1	0	0	0	0	0	0	1	0	0	0	0	0	0	…

Item 2

		+Stock							+Stock							+Stock
Daily Sales	…	0	0	0	1	0	0	0	0	0	0	1	0	0	0	…

Sales VS Demand

Item 1

		+Stock							+Stock							+Stock
Daily Sales	…	1	0	0	0	0	0	0	1	0	0	0	0	0	0	…
Daily Stock	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0

Item 2

		+Stock							+Stock							+Stock
Daily Sales	…	0	0	0	1	0	0	0	0	0	0	1	0	0	0	…
Daily Stock	0	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0

Poisson Model

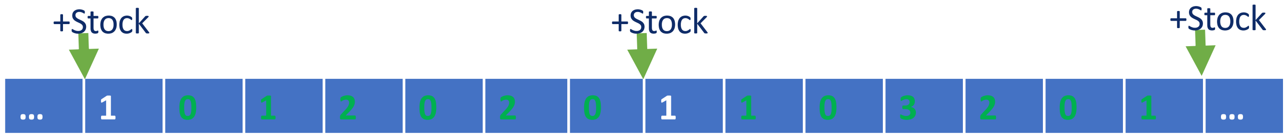
Key assumptions:

The number of sales in different periods of time is a random variable

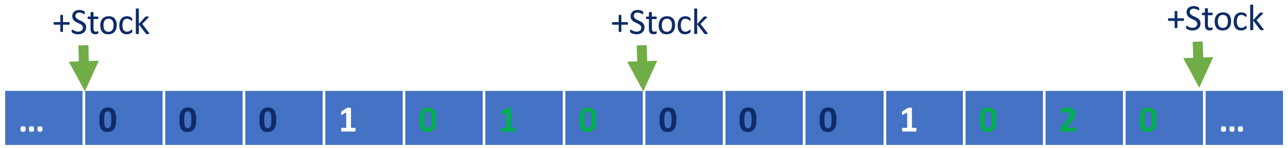
Probability to sell k pieces in t days (in case of infinite stock) is Poisson distribution

$$P_{\infty}(x = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Item 1: $\lambda = 1$



Item 2: $\lambda = 0.4231$



Poisson model for Demand restoration

Model of birth and death

m – number of pieces in stock

k – sales amount

$P_m(x = k)$ - probability to sell k pieces having m pieces in stock

$$P_m(x = k) = \begin{cases} 1, & k = m = 0, \\ 0, & k > m \text{ or } k < 0 \end{cases}$$

Poisson model for Demand restoration

Model of birth and death

m – number of pieces in stock

k – sales amount

$P_m(x = k)$ - probability to sell k pieces having m pieces in stock

$$P_m(x = k) = \begin{cases} 1, & k = m = 0, \\ 0, & k > m \text{ or } k < 0 \\ \dfrac{(\lambda)^k}{k!} e^{-\lambda}, & 0 \leq k < m, \end{cases}$$

Poisson model for Demand restoration

Model of birth and death

m – number of pieces in stock

k – sales amount

$P_m(x = k)$ - probability to sell k pieces having m pieces in stock

$$P_m(x = k) = \begin{cases} 1, & k = m = 0, \\ 0, & k > m \text{ or } k < 0 \\ \dfrac{(\lambda)^k}{k!} e^{-\lambda}, & 0 \le k < m, \\ 1 - \sum_{l=0}^{m-1} \dfrac{(\lambda)^l}{l!} e^{-\lambda}, & k = m > 0 \end{cases}$$

How to estimate λ for each item:

- Long history of data
- Find proper estimation method

Process of birth and death http://sci.alnam.ru/book_otsp.php?id=58

Poisson model for Demand restoration

Estimation of λ

- There is a sample realization of a random variable for a two parameters

$$(k_1, m_1), (k_2, m_2), \dots, (k_T, m_T)$$

- Likelihood

$$L = P_{m_1}(x = k_1) \cdot P_{m_2}(x = k_2) \cdot \dots \cdot P_{m_T}(x = k_{m_T}) =$$
$$= \prod_{i:0 \le k_i < m_i} \frac{(\lambda)^{k_i}}{k_i!} e^{-\lambda} \cdot \prod_{i:k_i = m_i > 0} \left(1 - \sum_{l=0}^{m_i - 1} \frac{(\lambda)^l}{l!} e^{-\lambda}\right) \cdot \prod_{i:k_i = m_i = 0} 1$$

Poisson model for Demand restoration

Estimation of λ

- Log likelihood

$$\ln L = \ln\left(\prod_{i:0\le k_i<m_i} \frac{(\lambda)^{k_i}}{k_i!} e^{-\lambda} \cdot \prod_{i:k_i=m_i>0} \left(1 - \sum_{l=0}^{m_i-1} \frac{(\lambda)^l}{l!} e^{-\lambda}\right) \cdot \prod_{i:k_i=m_i=0} 1\right) =$$

$$= \sum_{i:0\le k_i<m_i} \ln\left(\frac{(\lambda)^{k_i}}{k_i!} e^{-\lambda}\right) + \sum_{i:k_i=m_i>0} \ln\left(1 - \sum_{l=0}^{m_i-1} \frac{(\lambda)^l}{l!} e^{-\lambda}\right) =$$

$$= \ln\lambda \cdot \sum_{i:0\le k_i<m_i} k_i - \lambda \cdot \underbrace{\sum_{i:0\le k_i<m_i} 1}_{n_{0\le k<m}} + \sum_{i:\, k_i=m_i>0} \ln\left(1 - \sum_{l=1}^{m_i} \frac{(\lambda)^l}{l!} e^{-\lambda}\right)$$

Poisson model for Demand restoration

Estimation of λ

m_i - stock in a day i

$\sum_{i:k_i>0} k_i$ - sum of sales within observed days

$n_{0\leq k<m}$- number of days, when sales less than stock

$n_{k=m>0}$- number of days, when sales are equal to stock

$$\ln L = \ln\lambda \cdot \sum_{i:0\leq k_i<m_i} k_i - \lambda \cdot n_{0\leq k<m} + \sum_{i:\, k_i=m_i>0} \ln\left(1 - \sum_{l=1}^{m_i} \frac{(\lambda)^l}{l!} e^{-\lambda}\right) \to \max_{\lambda>0}$$

Poisson model for Demand restoration

MLE for λ estimation

Taylor series for exponent

$$1-\sum_{l=0}^{m_i-1}\frac{(\lambda)^l}{l!}e^{-\lambda}=e^{-\lambda}\left(e^{\lambda}-\sum_{l=0}^{m_i-1}\frac{(\lambda)^l}{l!}\right)=e^{-\lambda}\left(\sum_{l=0}^{\infty}\frac{(\lambda)^l}{l!}-\sum_{l=0}^{m_i-1}\frac{(\lambda)^l}{l!}\right)=e^{-\lambda}\sum_{l=m_i}^{\infty}\frac{(\lambda)^l}{l!}=\frac{(\lambda)^{m_i}}{m_i!}e^{\theta_{m_i,\lambda}\cdot\lambda}e^{-\lambda}$$

Log likelihood

$$\ln L=\ln\lambda\cdot\sum\nolimits_{i:0\le k_i<m_i}k_i-\lambda\cdot n_{0\le k<m}+\sum\nolimits_{i:\,k_i=m_i>0}\ln\left(\frac{(\lambda)^{m_i}}{m_i!}e^{\theta_{m_i,\lambda}\cdot\lambda}e^{-\lambda}\right)=$$

$$=\ln\lambda\cdot\sum_{i:0\le k_i<m_i}k_i-\lambda\cdot n_{0\le k_i<m_i}+\sum_{i:k_i=m_i>0}\left(k_i\cdot\ln\lambda+\left(\theta_{m_i,\lambda}-1\right)\cdot\lambda\right)+C=$$

$$=\ln\lambda\cdot\sum_{i:0\le k_i\le m_i}k_i-\lambda\cdot\left(n_{0\le k<m}+\sum_{i:k_i=m_i>0}\left(1-\theta_{m_i,\lambda}\right)\right)+C$$

Poisson model for Demand restoration

MLE for λ estimation

Necessary condition:

$$\frac{\partial \ln L}{\partial \lambda} = \frac{\sum_{i:0\le k_i \le m_i} k_i}{\lambda} - \left(n_{0\le k<m} + \sum_{i:k_i = m_i>0} \left(1 - \theta_{m_i,\lambda}\right) \right) = 0$$

$$\lambda = \frac{\sum_{i:0\le k_i \le m_i} k_i}{n_{0\le k<m} + \sum_{i:k_i = m_i>0} \left(1 - \theta_{m_i,\lambda}\right)} = \frac{\sum k_i}{n_{0\le k<m} + \alpha \cdot n_{k=m>0}},$$

где $\alpha \in [0,1]$

Poisson model for Demand restoration

MLE for λ estimation

m_i - stock in a day i
$\sum_{i:k_i>0} k_i$ - sum of sales within observed days
$n_{0\le k<m}$- number of days, when sales less than stock
$n_{k=m>0}$- number of days, when sales are equal to stock

$$\ln L = \ln \lambda \cdot \sum_{i:0\le k_i<m_i} k_i - \lambda \cdot n_{0\le k<m} + \sum_{i:\, k_i=m_i>0} \ln\left(1 - \sum_{l=1}^{m_i} \frac{(\lambda)^l}{l!} e^{-\lambda}\right) \to \max_{\lambda>0}$$

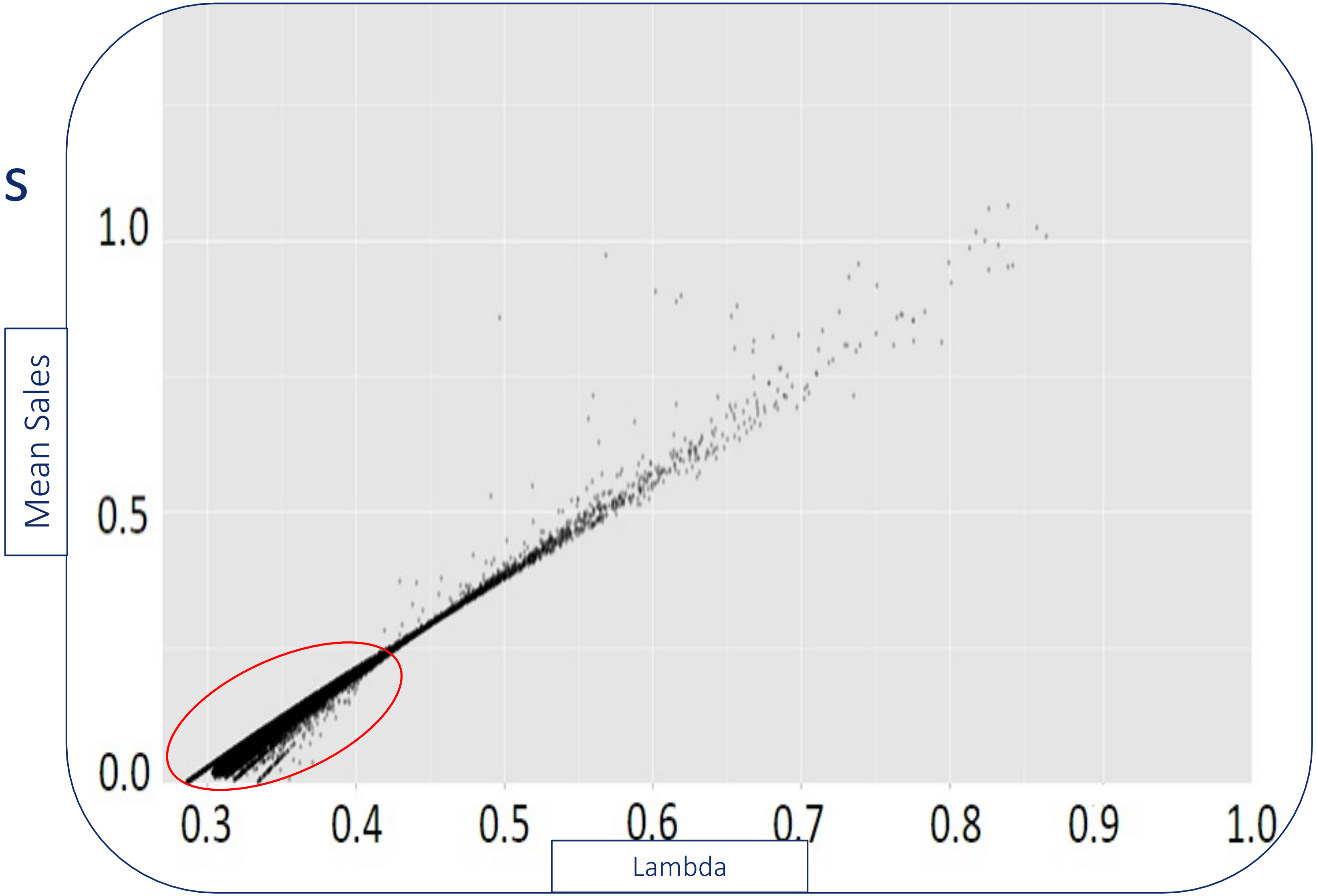
Solution:

$$\lambda = \frac{\sum k_i}{n_{0\le k<m} + \alpha \cdot n_{k=m>0}},$$

where $\alpha \in [0,1]$

Lambda vs Mean Sales

Lambda is significantly higher than Mean Sales for low-demand items



Several approaches to restore demand

10 -20

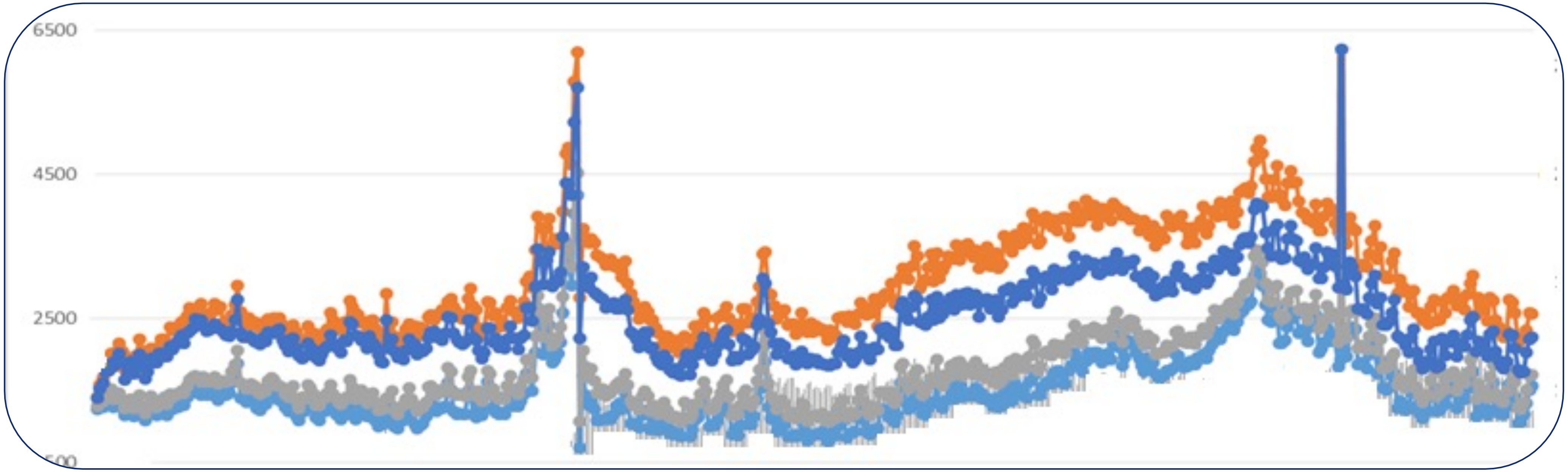
Realistic Uplift

Sales

Optimistic demand

Conservative demand

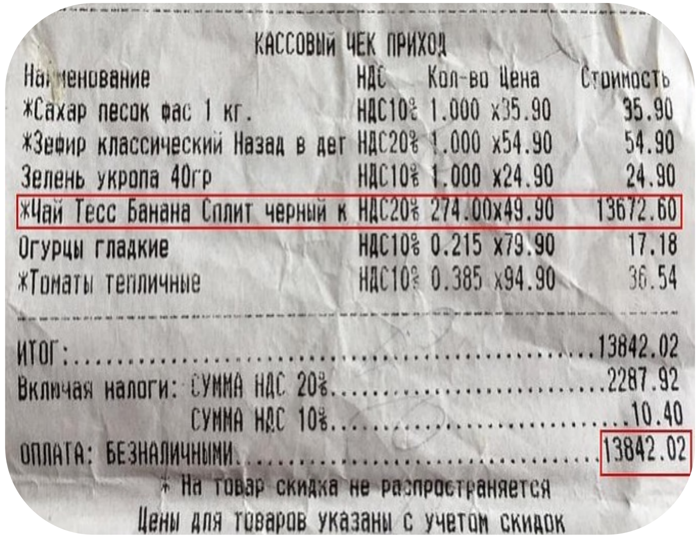
Pessimistic demand



significant under forecasting

poor quality of products or expired date (salvage)

Low shelf availability

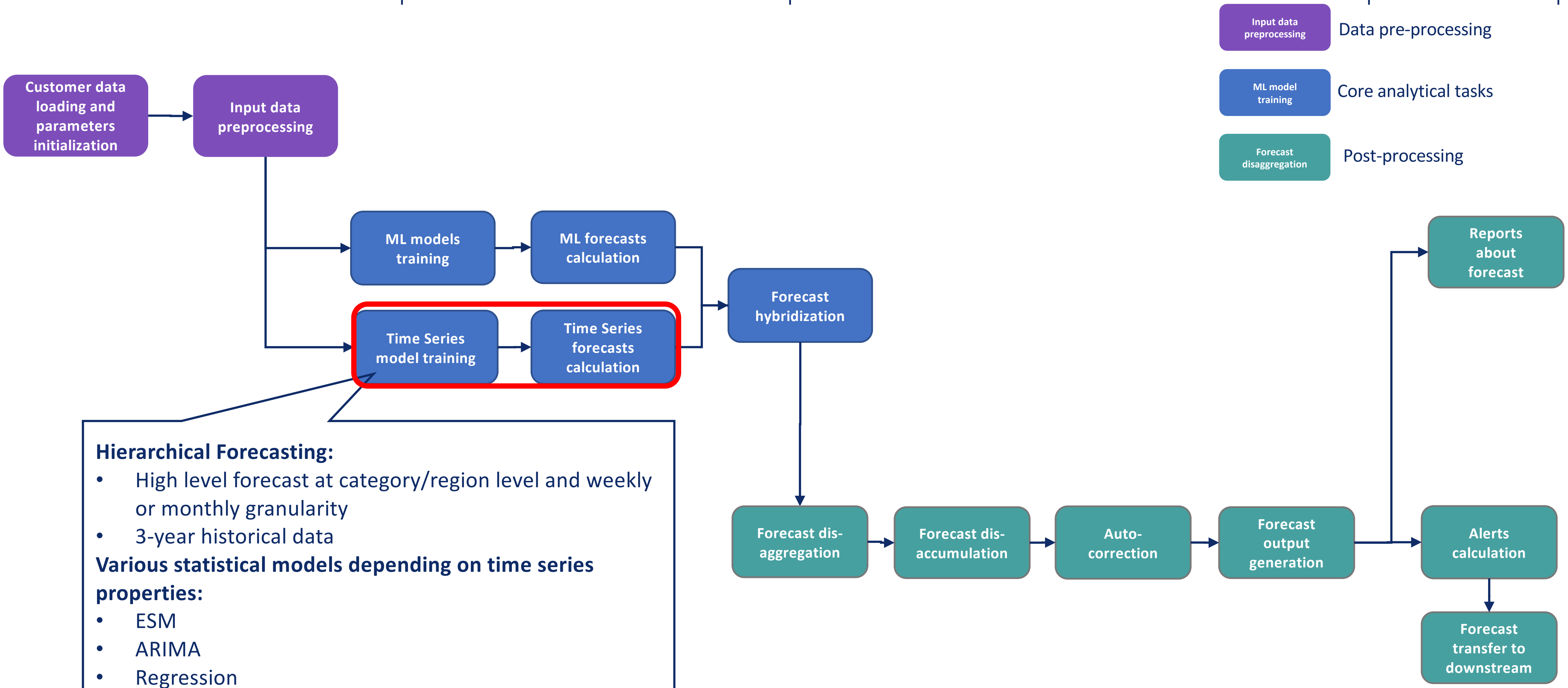


КАССОВЫЙ ЧЕК ПРИХОД

Наименование	НДС	Кол-во Цена	Стоимость
*Сахар песок фас 1 кг.	НДС10%	1.000 x35.90	35.90
*Зефир классический Назад в дет	НДС20%	1.000 x54.90	54.90
Зелень укропа 40гр	НДС10%	1.000 x24.90	24.90
*Чай Тесс Банана Сплит черный к	НДС20%	274.00x49.90	13672.60
Огурцы гладкие	НДС10%	0.215 x79.90	17.18
*Томаты тепличные	НДС10%	0.385 x94.90	36.54

ИТОГ:....................................13842.02
Включая налоги: СУММА НДС 20%.....................2287.92
СУММА НДС 10%.....................10.40
ОПЛАТА: БЕЗНАЛИЧНЫМИ.............................13842.02
* На товар скидка не распространяется
Цены для товаров указаны с учетом скидок

Technical issues



Hierarchical Forecasting:

- High level forecast at category/region level and weekly or monthly granularity
- 3-year historical data

Various statistical models depending on time series properties:

- ESM
- ARIMA
- Regression
- Croston

Train sample size:
$10^1 - 10^2$

store_id	product_id	Date	demand	promo_flg	price	event_flg
1	1	01JAN2016	3	0	54,00 ₽	1
1	1	02JAN2016	9	1	52,00 ₽	0
1	1	03JAN2016	6	1	51,00 ₽	0
1	1	04JAN2016	3	1	58,00 ₽	0
1	1	05JAN2016	10	1	58,00 ₽	0
1	1	06JAN2016	4	0	55,00 ₽	0
1	1	07JAN2016	6	1	60,00 ₽	1
1	1	08JAN2016	8	1	56,00 ₽	0
1	1	09JAN2016	10	0	56,00 ₽	0
1	1	10JAN2016	2	1	59,00 ₽	0
1	1	11JAN2016	9	0	50,00 ₽	0
1	1	12JAN2016	3	0	52,00 ₽	0
1	1	13JAN2016	6	0	57,00 ₽	0
1	1	14JAN2016	9	0	57,00 ₽	1
1	1	15JAN2016	4	0	57,00 ₽	0
1	1	16JAN2016	10	1	53,00 ₽	0
1	1	17JAN2016	5	1	58,00 ₽	0
1	1	18JAN2016	3	1	59,00 ₽	0
1	1	19JAN2016	10	0	56,00 ₽	0
1	1	20JAN2016	10	0	57,00 ₽	0
1	1	21JAN2016	4	1	56,00 ₽	0
1	1013	22JAN2016	0.0	0	51,00 ₽	

Features amount:
$10^0 - 10^1$

store_id	product_id	Date	demand	promo_flg	price	event_flg
1	1	01JAN2016	3	0	54,00 ₽	1
1	1	02JAN2016	9	1	52,00 ₽	0
1	1	03JAN2016	6	1	51,00 ₽	0
1	1	04JAN2016	3	1	58,00 ₽	0
1	1	05JAN2016	10	1	58,00 ₽	0
1	1	06JAN2016	4	0	55,00 ₽	0
1	1	07JAN2016	6	1	60,00 ₽	1
1	1	08JAN2016	8	1	56,00 ₽	0
1	1	09JAN2016	10	0	56,00 ₽	0
1	1	10JAN2016	2	1	59,00 ₽	0
1	1	11JAN2016	9	0	50,00 ₽	0
1	1	12JAN2016	3	0	52,00 ₽	0
1	1	13JAN2016	6	0	57,00 ₽	0
1	1	14JAN2016	9	0	57,00 ₽	1
1	1	15JAN2016	4	0	57,00 ₽	0
1	1	16JAN2016	10	1	53,00 ₽	0
1	1	17JAN2016	5	1	58,00 ₽	0
1	1	18JAN2016	3	1	59,00 ₽	0
1	1	19JAN2016	10	0	56,00 ₽	0
1	1	20JAN2016	10	0	57,00 ₽	0
1	1	21JAN2016	4	1	56,00 ₽	0
1	1013	22JAN2016	0.0	0	51,00 ₽	

Time Series Forecasign Models and Tools

- Exponential Smoothing
- Winters model
- Holt model
- Theil-Wage model
- (S)ARIMA(X)
- Croston Model (IDM)
- Random walk
- SSA
- (E)GARCH
- Dynamic Linear Regression
- …

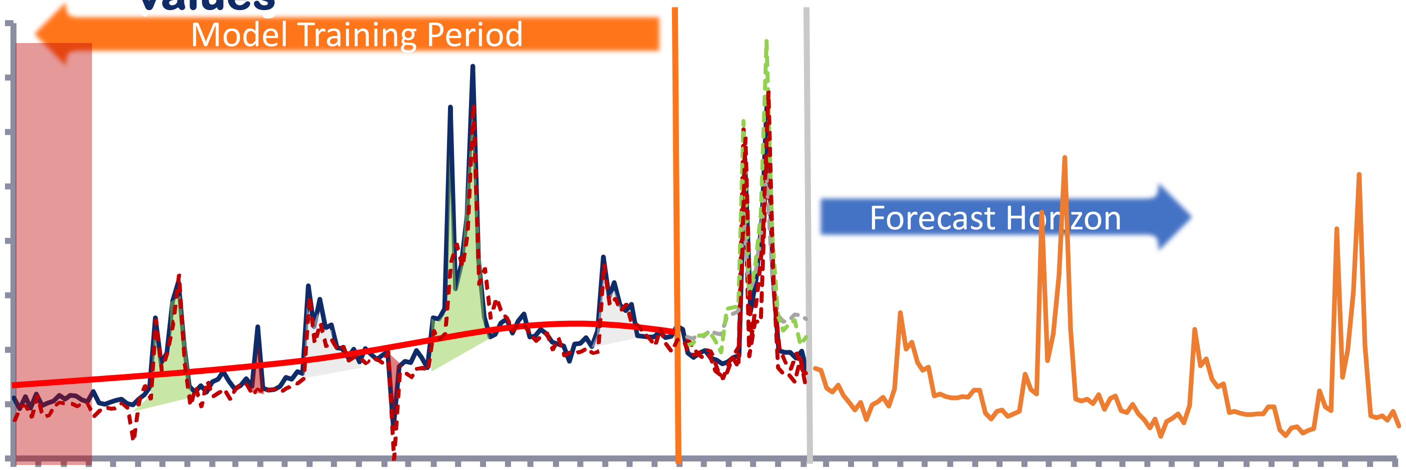
- R

- Python
  - Statsmodels
  - Pandas
  - Fbprophet
  - pmdarima
  - Sktime
  - Pyflux
  - ETSN

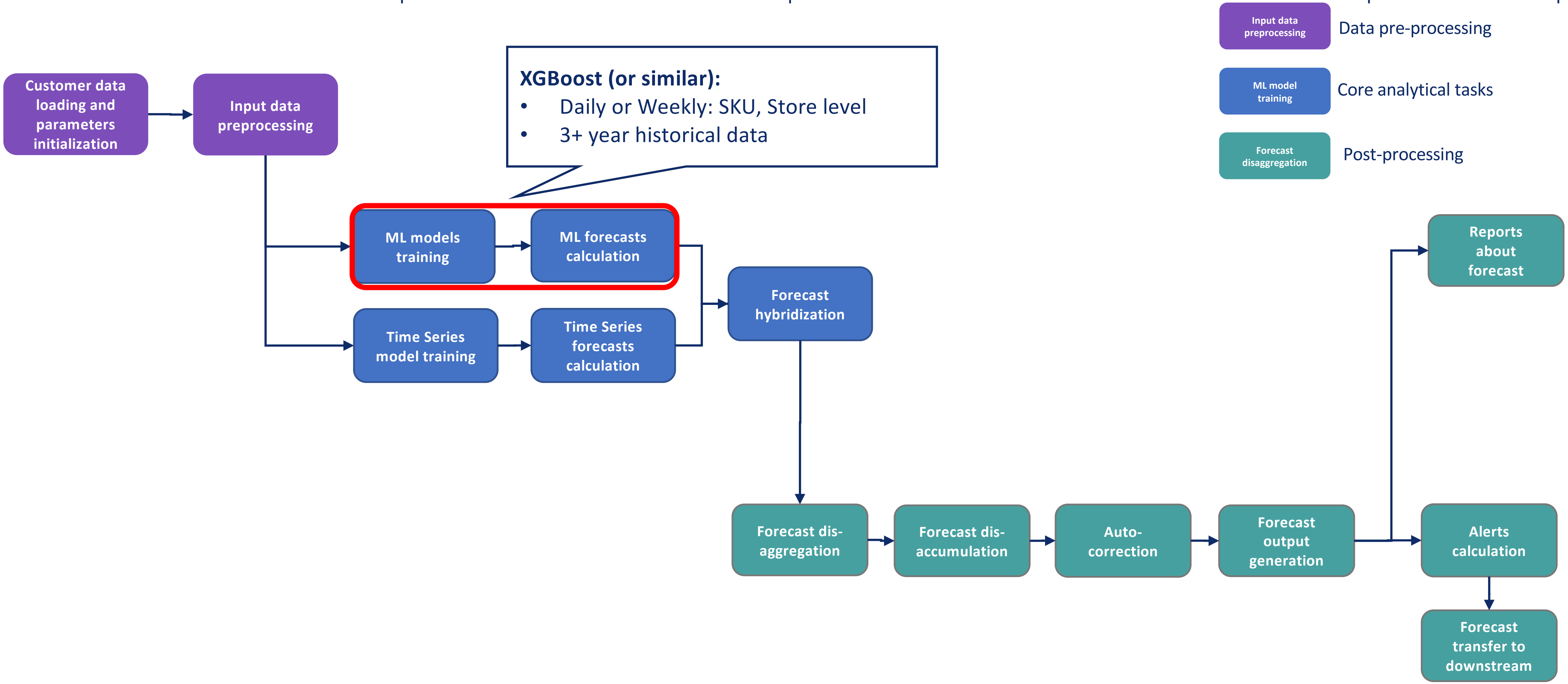
Learn more: https://github.com/aromanenko/ATSF/

Time Series forecasting approach

❻ Generating a forecast taking into account future events and factor values



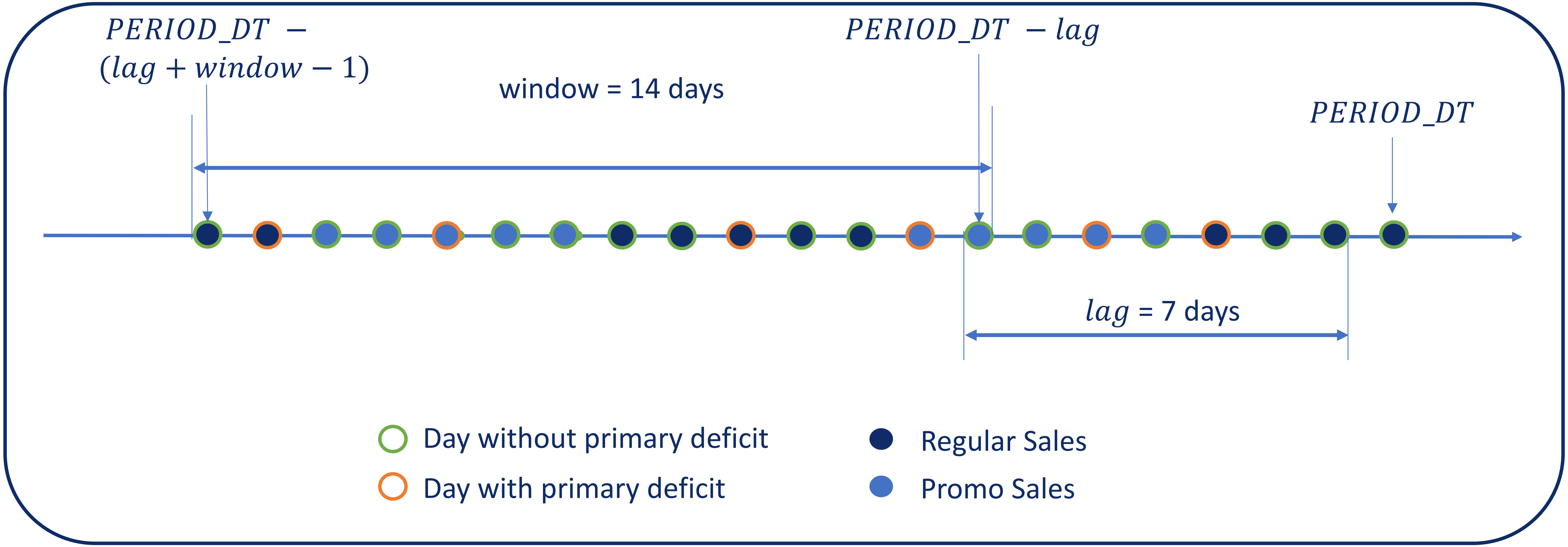
Model 1: *ARIMA*, ƒ(history, NY event, promo, price, 14Feb, cross-effects)
Model 2: *Exponentia Smoothing*, ƒ(History, Seasonality)
Model 3 3: *ARIMA*, ƒ(History, Seasonality, promo, price, NY event,…)
Other


Input data preprocessing
Data pre-processing
ML model training
Core analytical tasks
Forecast disaggregation
Post-processing
XGBoost (or similar):
• Daily or Weekly: SKU, Store level
• 3+ year historical data
Customer data loading and parameters initialization
Input data preprocessing
ML models training
ML forecasts calculation
Time Series model training
Time Series forecasts calculation
Forecast hybridization
Forecast dis-aggregation
Forecast dis-accumulation
Auto-correction
Forecast output generation
Reports about forecast
Alerts calculation
Forecast transfer to downstream

Store_id	SKU_id	Date	Sales	Promo	Regular_Price	Promo_Price
83	1	18.05.2015	91.0	NaN	131.70	NaN
14	1	09.03.2015	182.0	NaN	140.67	NaN
13	2	11.08.2015	9.0	NaN	150.73	NaN
75	1	06.04.2016	72.0	NaN	141.21	NaN
4	2	19.03.2016	69.0	1.0	138.50	124.77
21	1	11.11.2015	138.0	NaN	130.34	NaN
8	2	01.03.2015	0.0	NaN	139.86	NaN
41	2	04.10.2015	0.0	NaN	162.96	NaN
54	1	14.02.2016	402.0	1.0	141.76	130.75
2	2	28.07.2015	0.0	NaN	150.73	NaN
76	2	06.04.2016	13.0	NaN	138.36	NaN
59	2	16.01.2016	3.0	NaN	142.57	NaN
54	2	12.02.2016	82.0	1.0	138.36	108.46

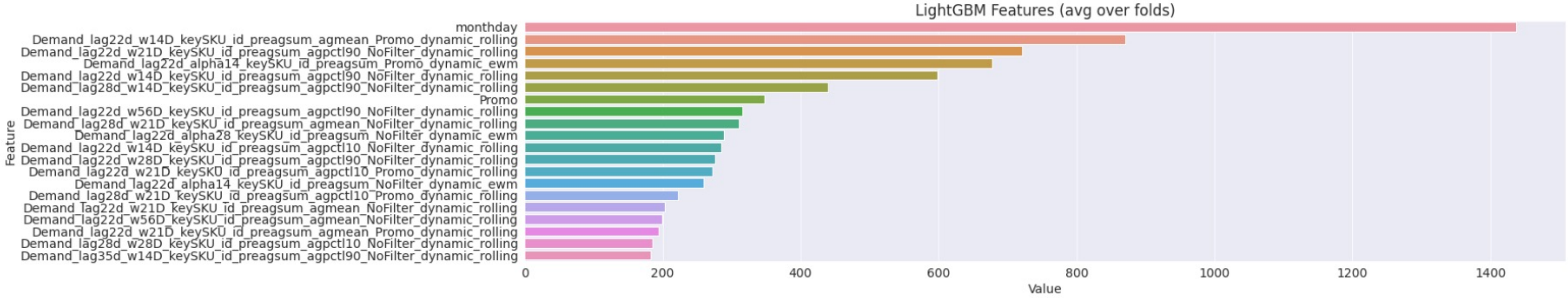
Train sample size:
$10^5 - 10^7$

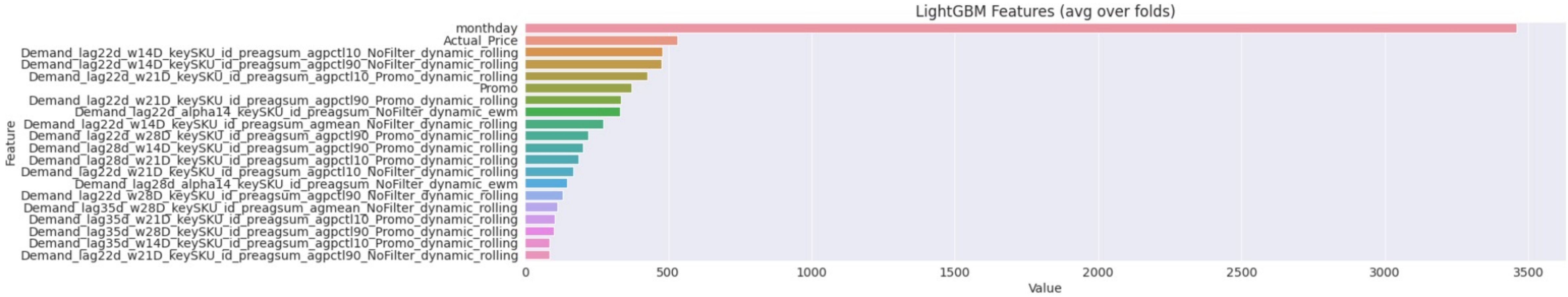
Rolling (lagged) features parameters



https://github.com/aromanenko/ATSF/blob/main/9_TSForecasting%20with%20ML.ipynb

Features in ML models





Different rolling features are to be designed to involve sales history into ML forecast

Weekday	Mon	Tue	Wed	Thr	Fri	Sat	Sun	Mon	Tue
Date	11.02.19	12.02.19	13.02.19	14.02.19	15.02.19	16.02.19	17.02.19	18.02.19	19.02.19
Demand	5	4	4	4	6	6	11	0	1
l1_w7_mean_reg	3,29	3,29	3,29	3,29	3,29	3,29	3,29	3,88	3,88

Feature name	lag, weeks	window, days	statistic	filter1: demand type	filter2: include deficit_day	filter3: weekday
l1_w7_mean_regl_...	1	7	mean	regular	yes	all
l1_w14_pct10_prm_...	1	14	percentile 1	promo	no	Monday
…	…	…	…	…	…	…
l7_w28_med_all_...	7	2	median	all	yes	Sunday
l7_w28_med_all_...	7	2	median	all	yes	rolling

Features amount: $10^2 - 10^3$

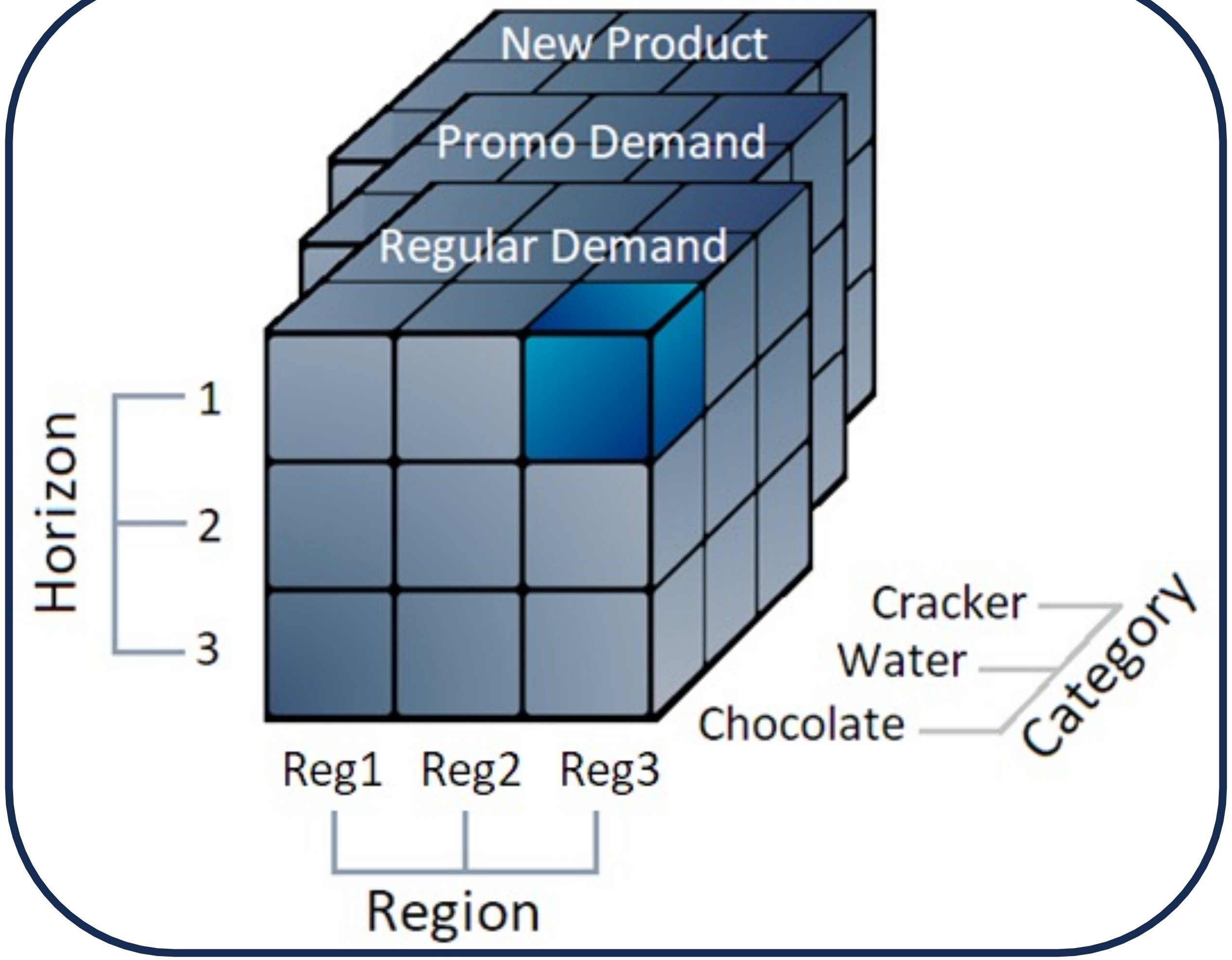
parameter	possible values	Number of values
lag	1, 2,3,7,14,91,180,365	4-8
window	7,14,21,28,91,180,365	4-7
statistic	mean, median, std, percentile (10,90)	4-5
filter1: demand type	regular, promo, all	3
filter2: include deficit_day	yes, no	2
filter3: weekday	Monday,…, Sunday, rolling, all	9

Store_id	SKU_id	Date	Sales	Promo	Regular_Price	l1_w7_mean_regl_…
83	1	18.05.2015	91.0	NaN	131.70	NaN
14	1	09.03.2015	182.0	NaN	140.67	NaN
13	2	11.08.2015	9.0	NaN	150.73	NaN
75	1	06.04.2016	72.0	NaN	141.21	NaN
4	2	19.03.2016	69.0	1.0	138.50	124.77
21	1	11.11.2015	138.0	NaN	130.34	NaN
8	2	01.03.2015	0.0	NaN	139.86	NaN
41	2	04.10.2015	0.0	NaN	162.96	NaN
54	1	14.02.2016	402.0	1.0	141.76	130.75
2	2	28.07.2015	0.0	NaN	150.73	NaN
76	2	06.04.2016	13.0	NaN	138.36	NaN
59	2	16.01.2016	3.0	NaN	142.57	NaN
54	2	12.02.2016	82.0	1.0	138.36	108.46

Dimensions of ML model

Specific ML algorithm should be train for each combination of

Product Category
X
Region
X
Horizon (1^{st}, 2^{nd} week)
X
Demand Type (Regular, Promo)
X
Assortment Type (Old, New)

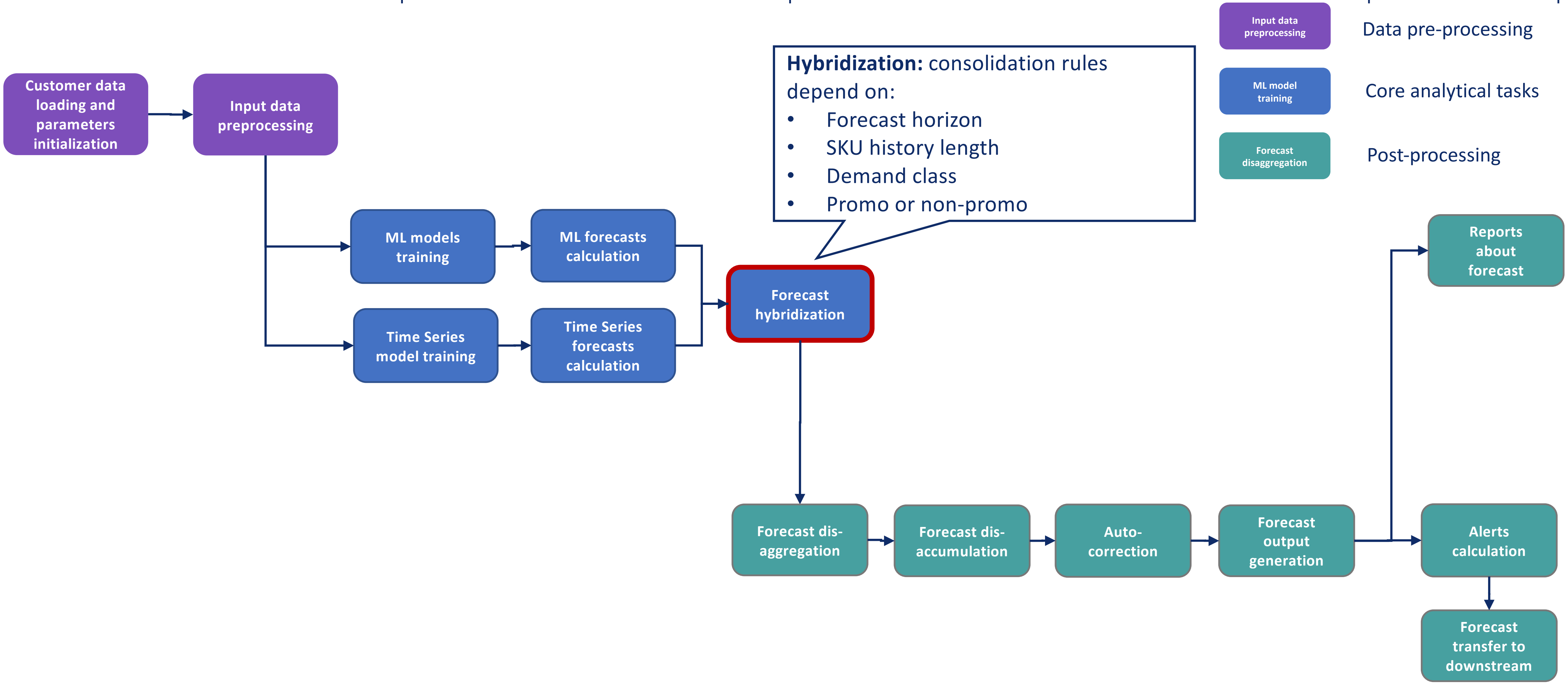


Why so many models?

For 10^6 pairs product x store you need to train

~ 10^6 classic ts models

~ $10^2 - 10^3$ ML models


Input data preprocessing
Data pre-processing
ML model training
Core analytical tasks
Forecast disaggregation
Post-processing
Hybridization: consolidation rules depend on:
• Forecast horizon
• SKU history length
• Demand class
• Promo or non-promo
Customer data loading and parameters initialization
Input data preprocessing
ML models training
ML forecasts calculation
Time Series model training
Time Series forecasts calculation
Forecast hybridization
Forecast dis-aggregation
Forecast dis-accumulation
Auto-correction
Forecast output generation
Reports about forecast
Alerts calculation
Forecast transfer to downstream

ML vs Statistics for demand forecasting

	Classic Statistics Algorithms	ML Based Algorithms
Pro	• Interpretability • Easy to consider business expertise • TS patterns are considered out-of-the-box	• Similarity between products, stores etc. is taken into account • Different types of features (inc. categorical attributes, nominal feature) can be considered
Cons	• Only continuous time series information • Unstable forecasting (often to be retrained) • A lot of manual operations to handle (demand analyst resources	• Lack of interpretability • No easy way to incorporate business expertise

How to leverage TS and ML forecasting approaches

[Makridacis competition: https://mofc.unic.ac.cy/](https://mofc.unic.ac.cy/)

2020

M4 Competition *Winner: Hybrid Approach

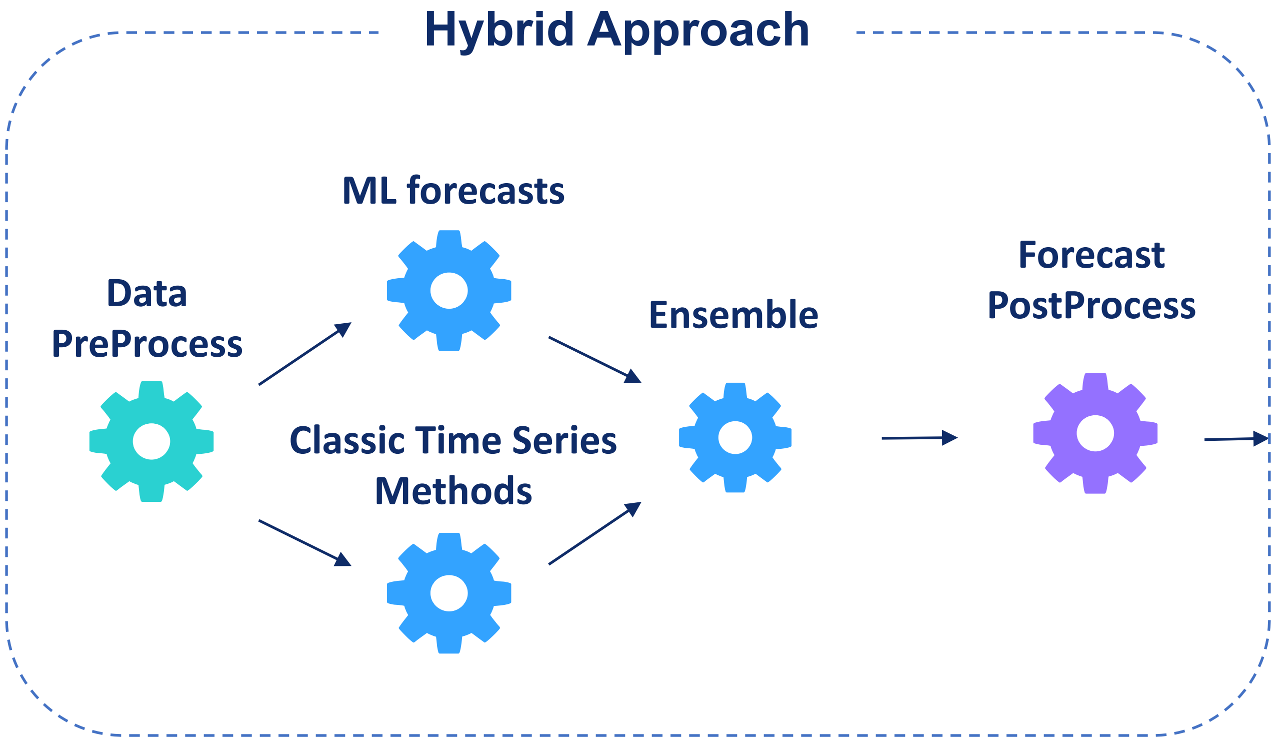
2022

M5 Competition *Winner: Hybrid Approach

2024

M6 Competition *Winner: Hybrid Approach

*along with lstm and other RNN approaches



High value
Low Forecastability

Short-History

- New products
- Life Cycle of Products
- Product-Predecessor

Key Products

- High priority products
  - Strong trend
  - Seasonal fluctuations
  - Possible cycles
  - Sales promotions
  - National marketing events
  - Advertising driven
  - Highly competitive
  - Cross-effects

High value
High Forecastability

Value to Company

Low value
Low Forecastability

Niche products

- Low priority - regional specialty products
  - Some trend
  - Seasonal fluctuations
  - Irregular demand
  - Local targeted marketing events

Stable demand

- Low priority products
  - Strong trend
  - Highly seasonal
  - Possibly cycles
  - Minor sales promotions

Low value
High Forecastability

High value
Low Forecastability

Short-History

- New products
- Life Cycle of Products
- Product-Predecessor

Key Products

- High priority products
  - Strong trend
  - Seasonal fluctuations
  - Possible cycles
  - Sales promotions
  - National marketing events
  - Advertising driven
  - Highly competitive
  - Cross-effects

High value
High Forecastability

Value to Company

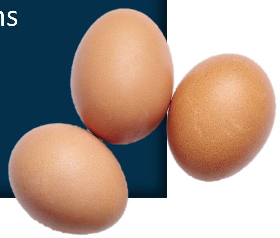
Low value
Low Forecastability

Niche products

- Low priority - regional specialty products
  - Some trend
  - Seasonal fluctuations
  - Irregular demand
  - Local targeted marketing events

Stable demand

- Low priority products
  - Strong trend
  - Highly seasonal
  - Possibly cycles
  - Minor sales promotions



Low value
High Forecastability

ML

High value to company
Low Forecastability

Short-History

Life Cycle of Products

Product-Predecessor

- Cross effects (cannibalization)

Key Products

- High priority products
  - Strong trend
  - Seasonal fluctuations
  - Possible cycles
  - Sales promotions
  - National marketing events
  - Advertising driven
  - Highly competitive
  - Cross-effects

ML+TS

High value to company
High Forecastability

Intermittent

Low value to company
Low Forecastability

Niche products

Low priority - regional specialty products
- Some trend
- Seasonal fluctuations
- Irregular demand
- Local targeted marketing events

Stable demand

- Low priority products
  - Strong trend
  - Highly seasonal
  - Possibly cycles
  - Minor sales promotions

TS

Low value to company
High Forecastability

Hybridization code example

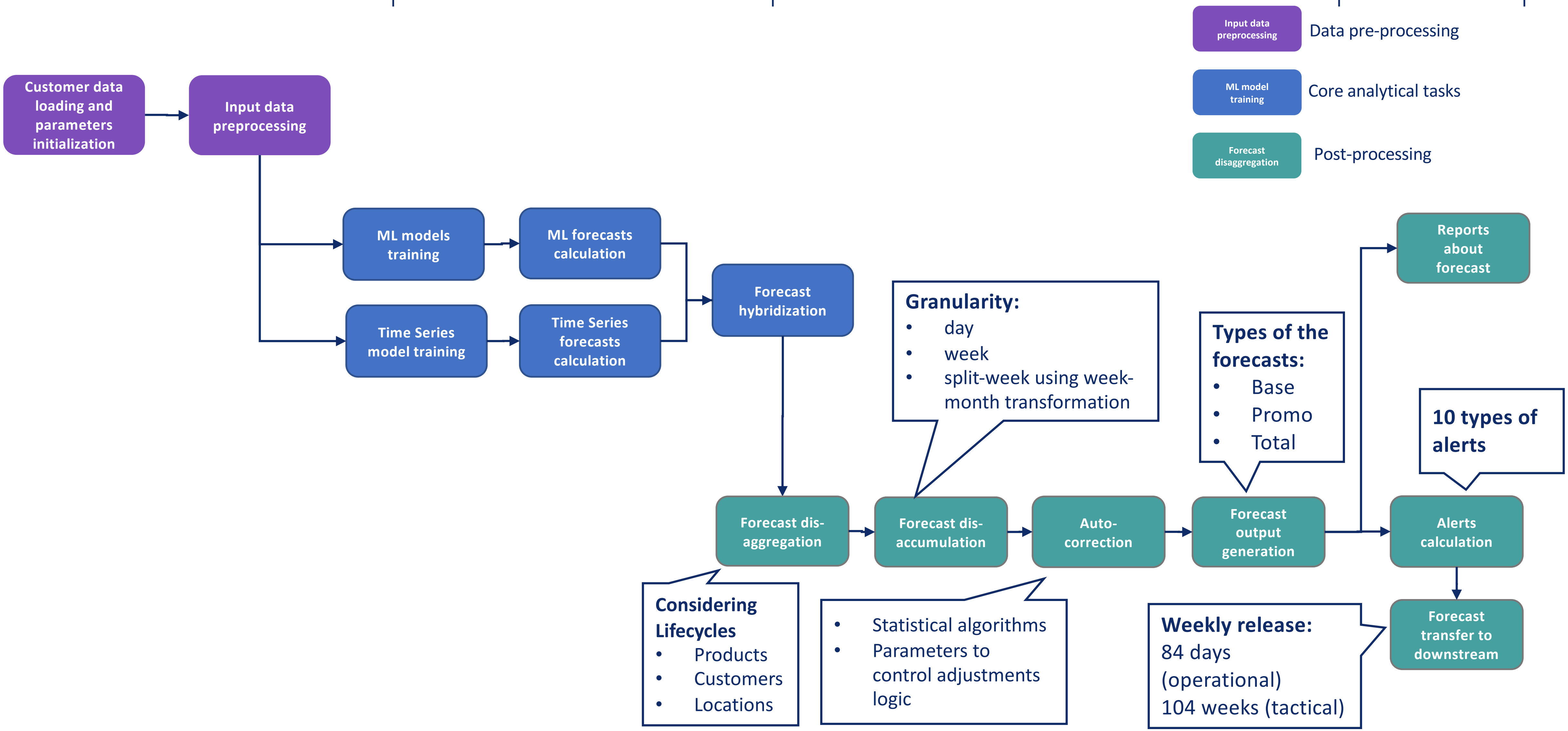
```
if 'ML_FORECAST_VALUE' not in data:
    data['ML_FORECAST_VALUE_F'] = data.VF_FORECAST_VALUE

if 'VF_FORECAST_VALUE' not in data.columns:
    data['VF_FORECAST_VALUE_F'] = data.ML_FORECAST_VALUE

HYBRID_FORECAST_VALUE = []
FORECAST_SOURCE = []
ENSEMBLE_FORECAST_VALUE = []


for _, row in data.iterrows():
    if (row.DEMAND_TYPE.lower() == 'promo' and row.SEGMENT_NAME.lower() != 'retired') or row.SEGMENT_NAME.lower() == 'short
            row.ASSORTMENT_TYPE.lower() == 'new':
        HYBRID_FORECAST_VALUE.append(row.ML_FORECAST_VALUE)
        FORECAST_SOURCE.append('ml')
        ENSEMBLE_FORECAST_VALUE.append(np.nan)
    elif row.SEGMENT_NAME.lower() == 'retired' or row.SEGMENT_NAME.lower() == 'low volume':
        VF_FORECAST_VALUE_F = IB_ZERO_DEMAND_THRESHOLD
        HYBRID_FORECAST_VALUE.append(VF_FORECAST_VALUE_F)
        FORECAST_SOURCE.append('vf')
        ENSEMBLE_FORECAST_VALUE.append(np.nan)
    else:
        HYBRID_FORECAST_VALUE.append(np.mean(row.ML_FORECAST_VALUE, row.VF_FORECAST_VALUE))
        FORECAST_SOURCE.append('ensemble')
        ENSEMBLE_FORECAST_VALUE.append(np.mean(row.ML_FORECAST_VALUE, row.VF_FORECAST_VALUE))
HYBRID_FORECAST_VALUE = pd.Series(HYBRID_FORECAST_VALUE)
FORECAST_SOURCE = pd.Series(FORECAST_SOURCE)
ENSEMBLE_FORECAST_VALUE = pd.Series(ENSEMBLE_FORECAST_VALUE)
data['HYBRID_FORECAST_VALUE'] = HYBRID_FORECAST_VALUE
data['FORECAST_SOURCE'] = FORECAST_SOURCE
data['ENSEMBLE_FORECAST_VALUE'] = ENSEMBLE_FORECAST_VALUE
```

Example see here
https://github.com/aromanenko/DFP/blob/assembling/assembling.py


Input data preprocessing
Data pre-processing
ML model training
Core analytical tasks
Forecast disaggregation
Post-processing
Customer data loading and parameters initialization
Input data preprocessing
ML models training
ML forecasts calculation
Time Series model training
Time Series forecasts calculation
Forecast hybridization
Granularity:
• day
• week
• split-week using week-month transformation
Types of the forecasts:
• Base
• Promo
• Total
Reports about forecast
10 types of alerts
Forecast dis-aggregation
Forecast dis-accumulation
Auto-correction
Forecast output generation
Alerts calculation
Forecast transfer to downstream
Considering Lifecycles
• Products
• Customers
• Locations
• Statistical algorithms
• Parameters to control adjustments logic
Weekly release:
84 days (operational)
104 weeks (tactical)

Forecast output structure (real case)

out_forecast_details
FORECAST_ID
PERIOD_DT
DIM1_LVL_ID
…
DIM5_LVL_ID
FORECAST_VALUE
FORECAST_LOWER
FORECAST_UPPER
FORECAST_SOURCE
ML_FORECAST_VALUE
VF_FORECAST_VALUE
ENSEMBLE_FORECAST_VALUE
BF_AUTOCOR_FORECAST_VALUE
DEMAND_TYPE

out_forecast
FORECAST_ID
PARENT_FORECAST_ID
TARGET_VARIABLE_NM
FORECAST_TYPE
PERIOD_TYPE
FORECAST_PERIOD_START_DT
FORECAST_PERIOD_END_DT
FORECAST_HORIZON
UOM_CD
PROCESS_START_DTTM
PROCESS_END_DTTM

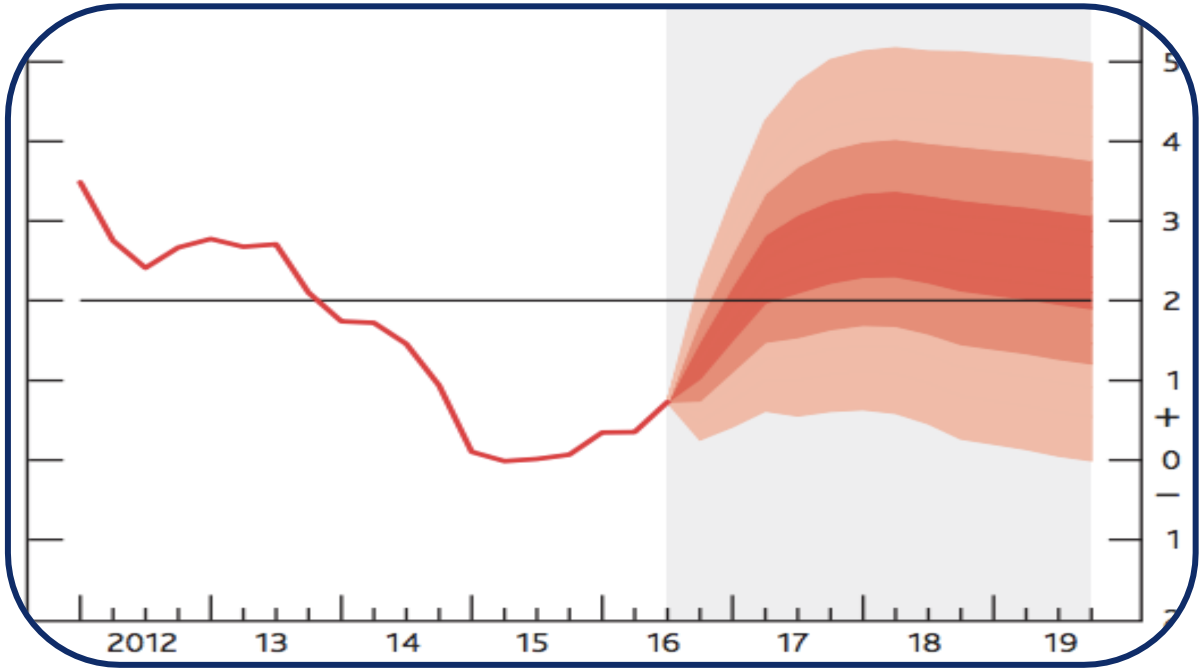
out_dimension
DIMENSION_ID
DIMENSION_NM

out_forecast_attributes
FORECAST_ID
DIM1_LVL_ID
…
DIM5_LVL_ID
ASSORTMENT_TYPE
FORECAST_VARIANCE

Output Forecast API

Dotted forecast

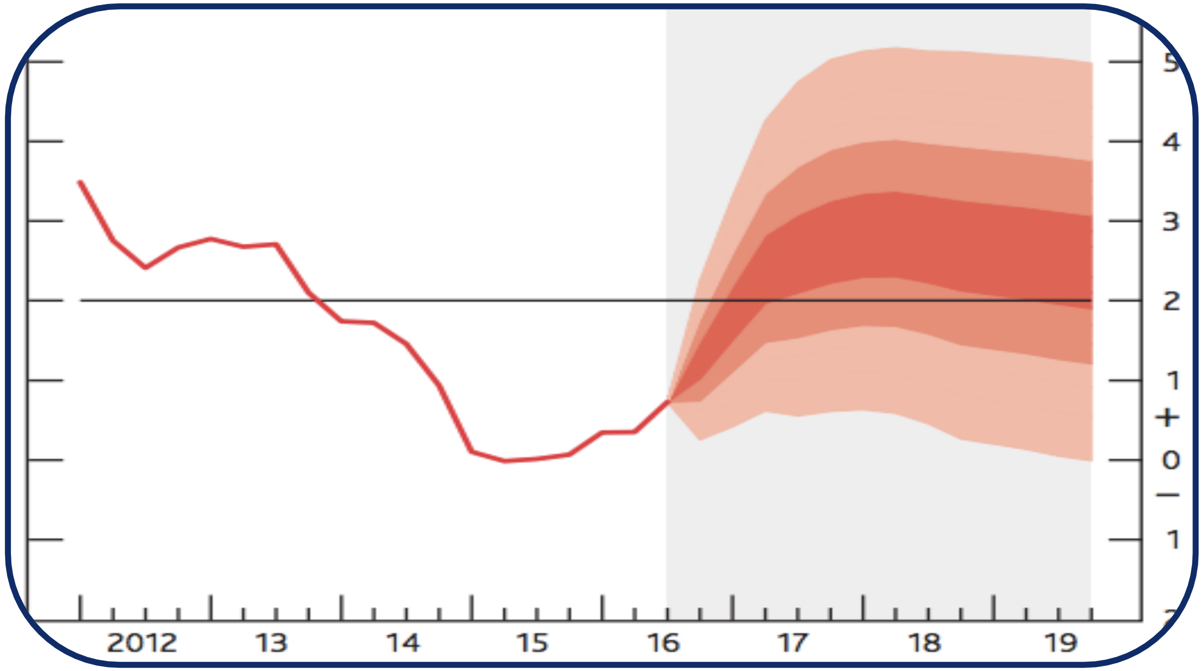
+ Confidence interval/Variance



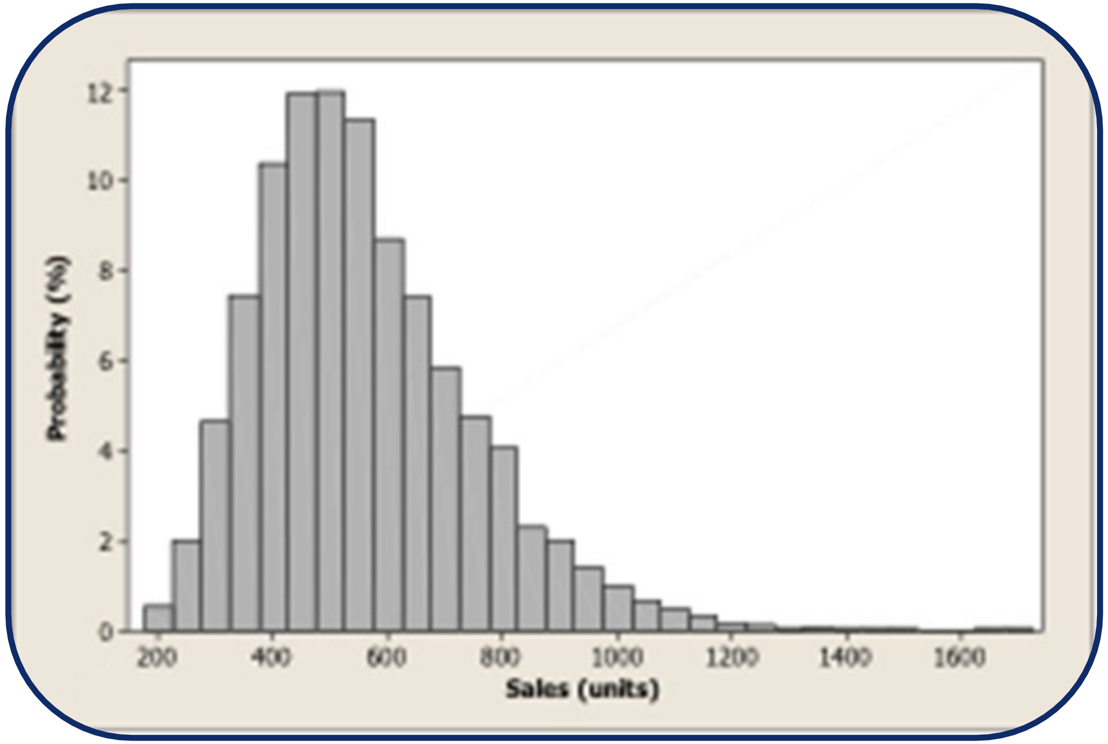
Output Forecast API

Dotted forecast

+ Confidence interval/Variance

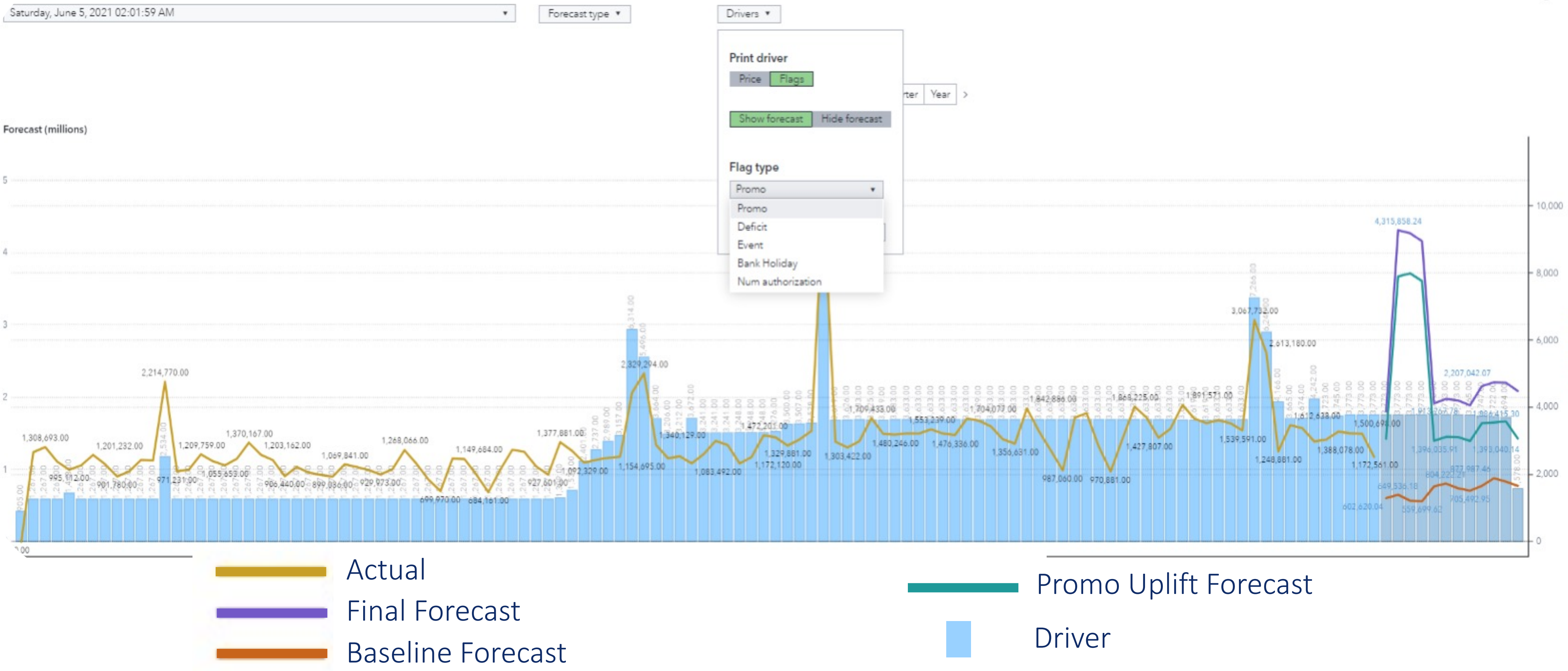


Density Forecast

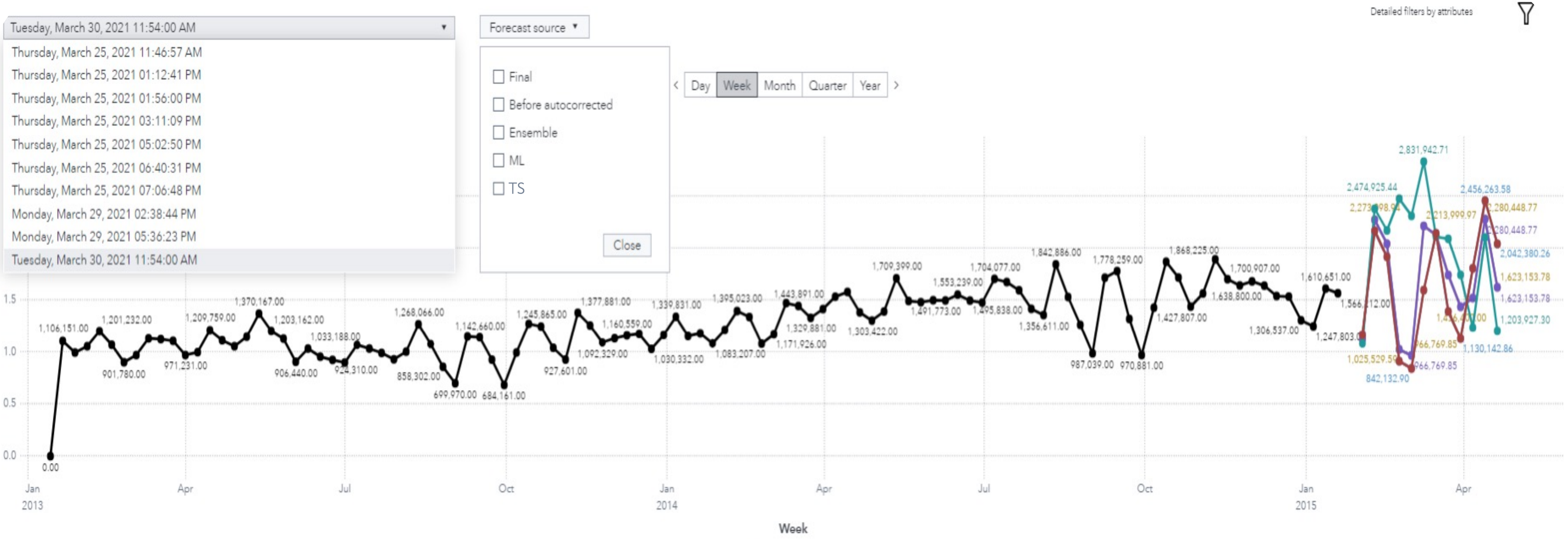


[Realization Example:
https://github.com/aromanenko/AggregatingAlgorithm/blob/1-base-aggregating-algorithm/DFA_func.ipynb](https://github.com/aromanenko/AggregatingAlgorithm/blob/1-base-aggregating-algorithm/DFA_func.ipynb)

Reports: time series overview dashboard



Report: comparing forecast sources dashboard



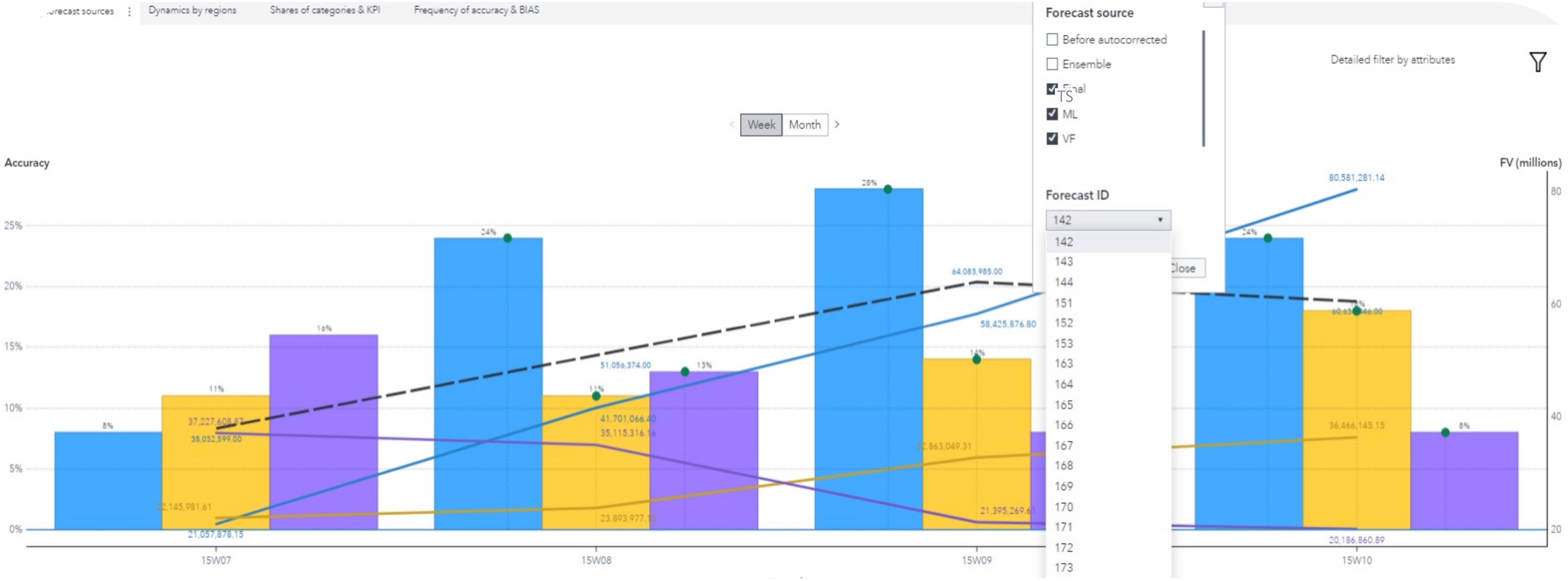
Final Forecast

ML Forecast

TS Forecast

Ensemble Forecast

Report: forecast accuracy dashboard



Final Forecast

TS Forecast

ML Forecast

Summary

- Demand forecasting and modeling is one of the key tasks for Retail companies
- Several classes of different forecasting problems can be distinguished.
- Forecasting methods vary greatly depending on the specifics of the task and data.

- Using of Hybrid (both time series and ML) approach is beneficial for forecast accuracy
- Demand restoration is necessary task to work with unconstrained demand
- Forecast postprocessing includes many vital transformations of forecast data regarding accuracy, quality and compatibility of with downstream systems

Self Check Questions

- List at least three different business processes in retail companies where demand forecast can be used.
- Provide at least 4 examples of demand driver in Retail
- Remember at least 2 advantages of time series forecasting methods and ML methods for demand forecasting problem
- What is rolling (lagged) feature?
- Why demand can differ from sales volumes?
- What is lambda parameter in demand restoration method based on Birth and Death Model?

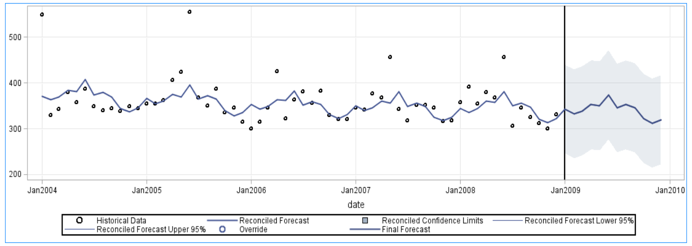
10 questions
5 minutes

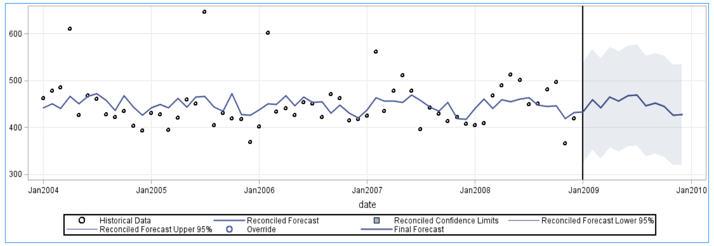
Use the link and participate in QUIZ

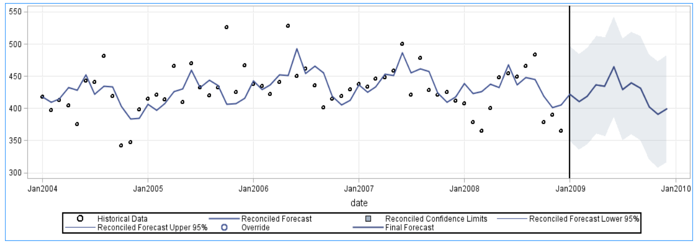
How to Improve Materials?

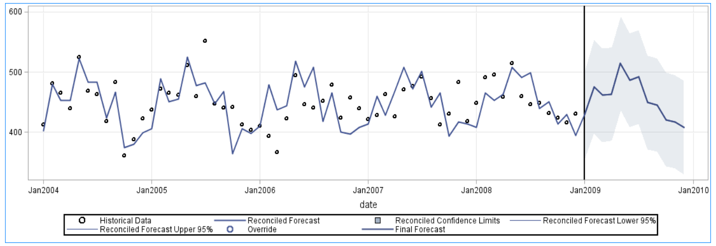
Feed back:

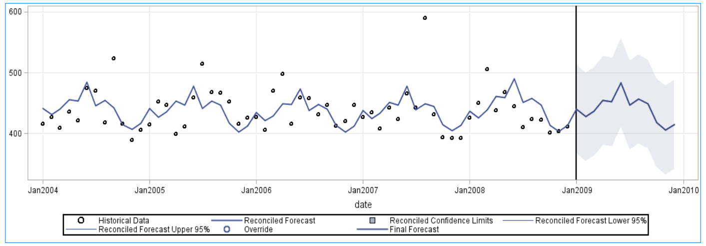
Items

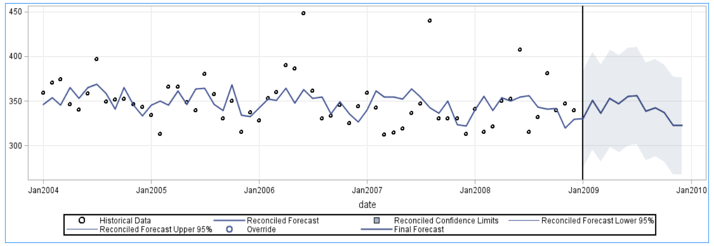

500
400
300
200
Jan2004
Jan2005
Jan2006
Jan2007
Jan2008
Jan2009
Jan2010
date
Historical Data
Reconciled Forecast
Reconciled Confidence Limits
Reconciled Forecast Lower 95%
Reconciled Forecast Upper 95%
Override
Final Forecast

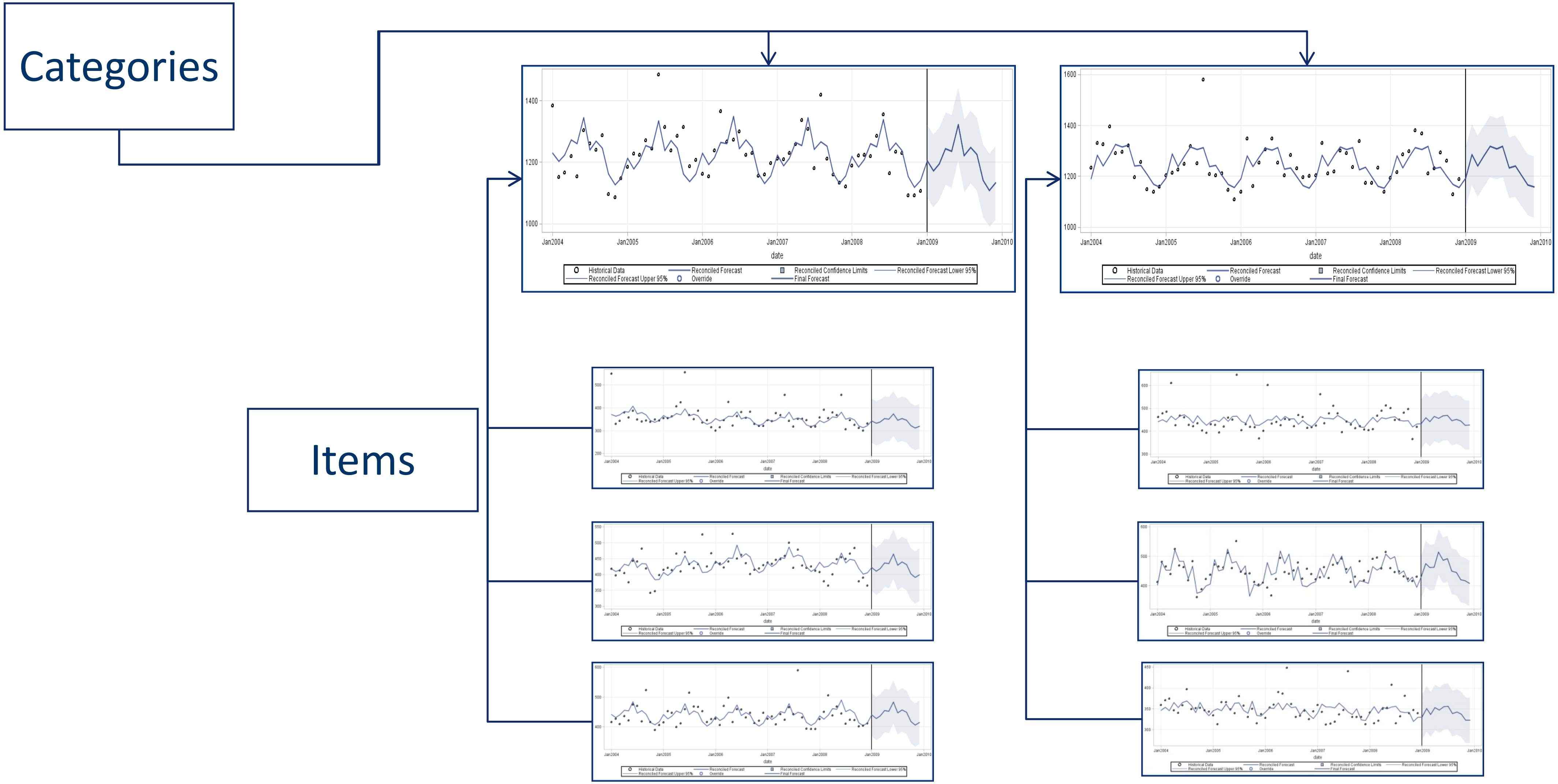

600
500
400
300
Jan2004
Jan2005
Jan2006
Jan2007
Jan2008
Jan2009
Jan2010
date
Historical Data
Reconciled Forecast
Reconciled Confidence Limits
Reconciled Forecast Lower 95%
Reconciled Forecast Upper 95%
Override
Final Forecast

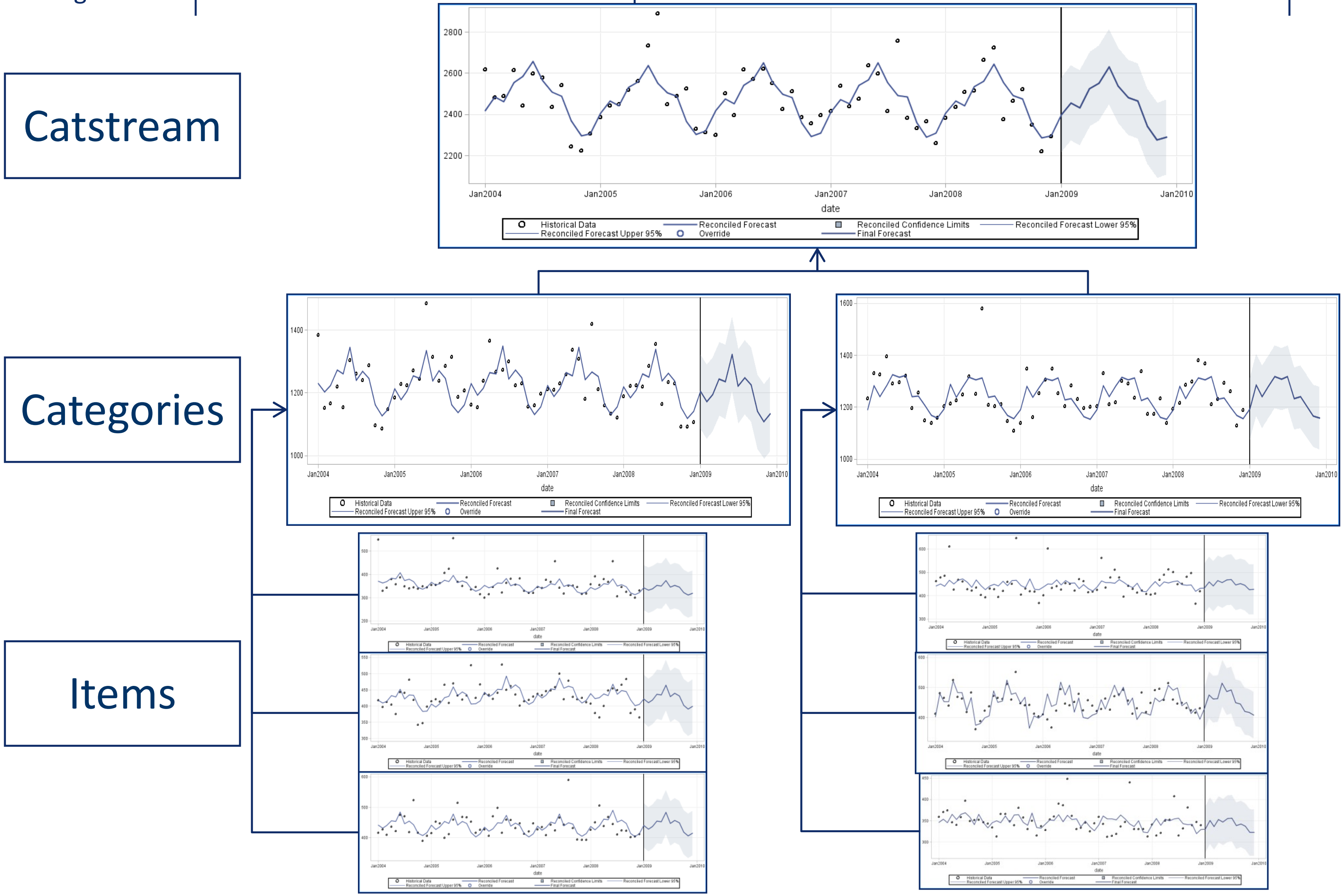

550
500
450
400
350
300
Jan2004
Jan2005
Jan2006
Jan2007
Jan2008
Jan2009
Jan2010
date
Historical Data
Reconciled Forecast
Reconciled Confidence Limits
Reconciled Forecast Lower 95%
Reconciled Forecast Upper 95%
Override
Final Forecast

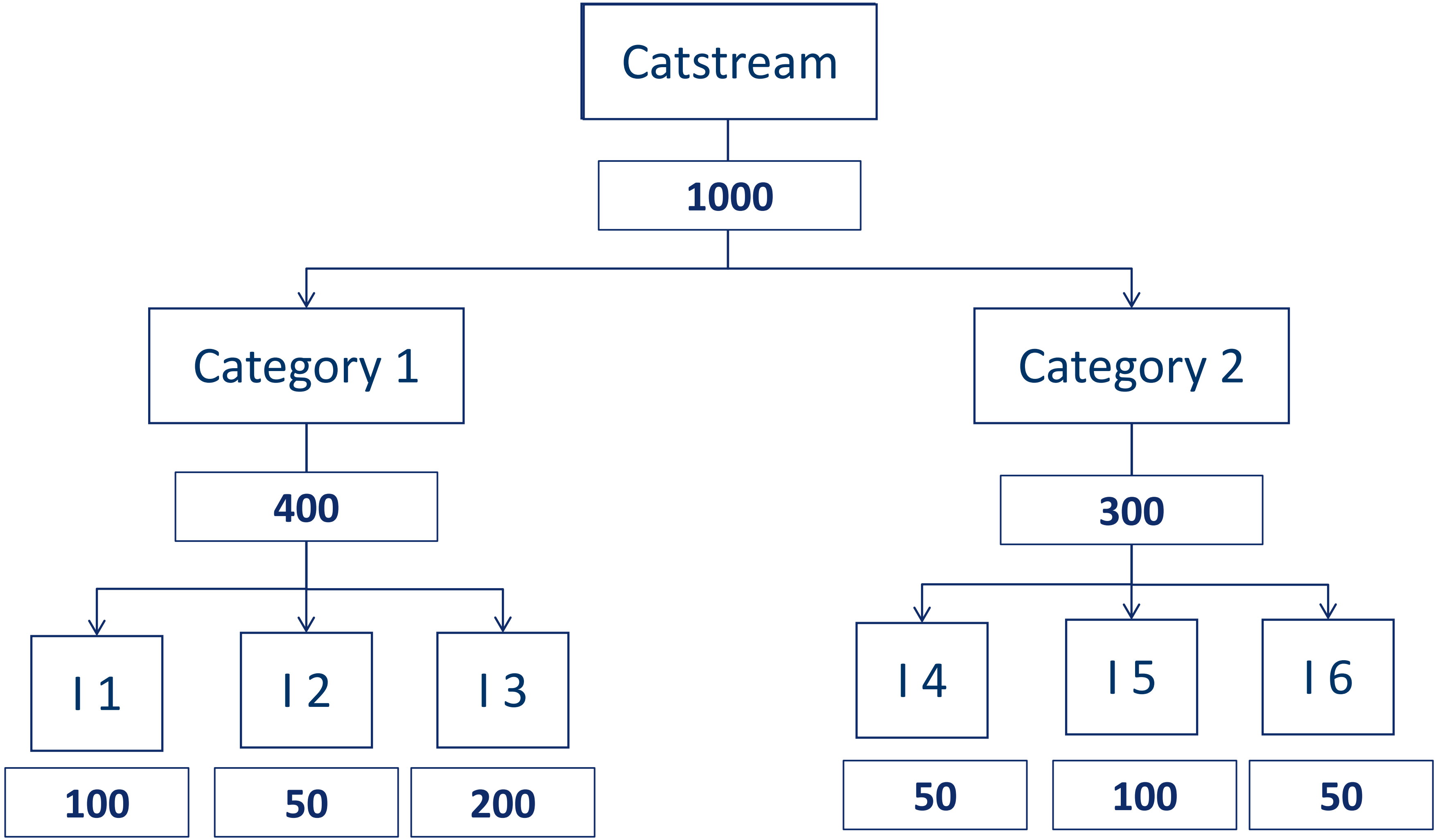

600
500
400
Jan2004
Jan2005
Jan2006
Jan2007
Jan2008
Jan2009
Jan2010
date
Historical Data
Reconciled Forecast
Reconciled Confidence Limits
Reconciled Forecast Lower 95%
Reconciled Forecast Upper 95%
Override
Final Forecast

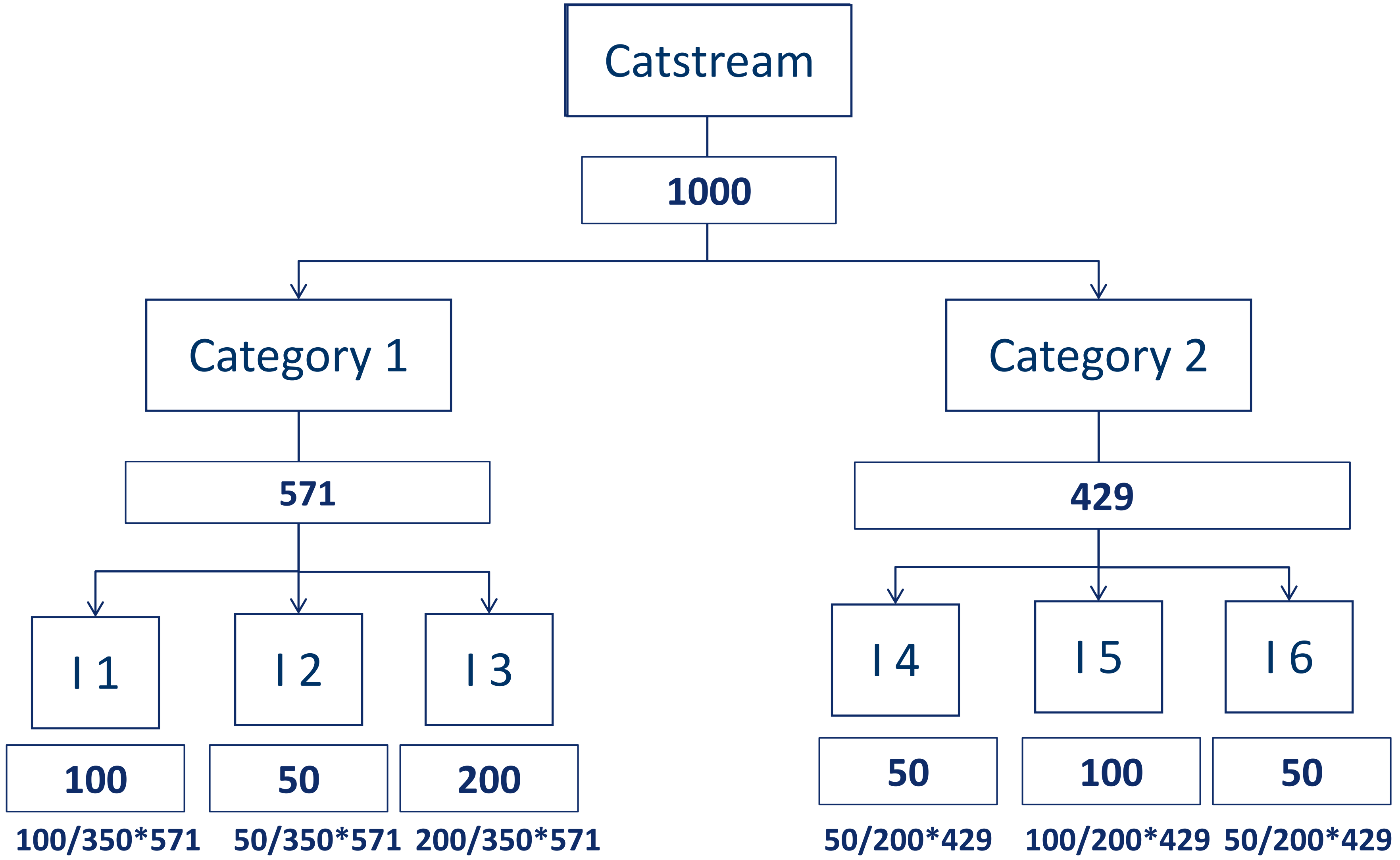

600
500
400
Jan2004
Jan2005
Jan2006
Jan2007
Jan2008
Jan2009
Jan2010
date
Historical Data
Reconciled Forecast
Reconciled Confidence Limits
Reconciled Forecast Lower 95%
Reconciled Forecast Upper 95%
Override
Final Forecast

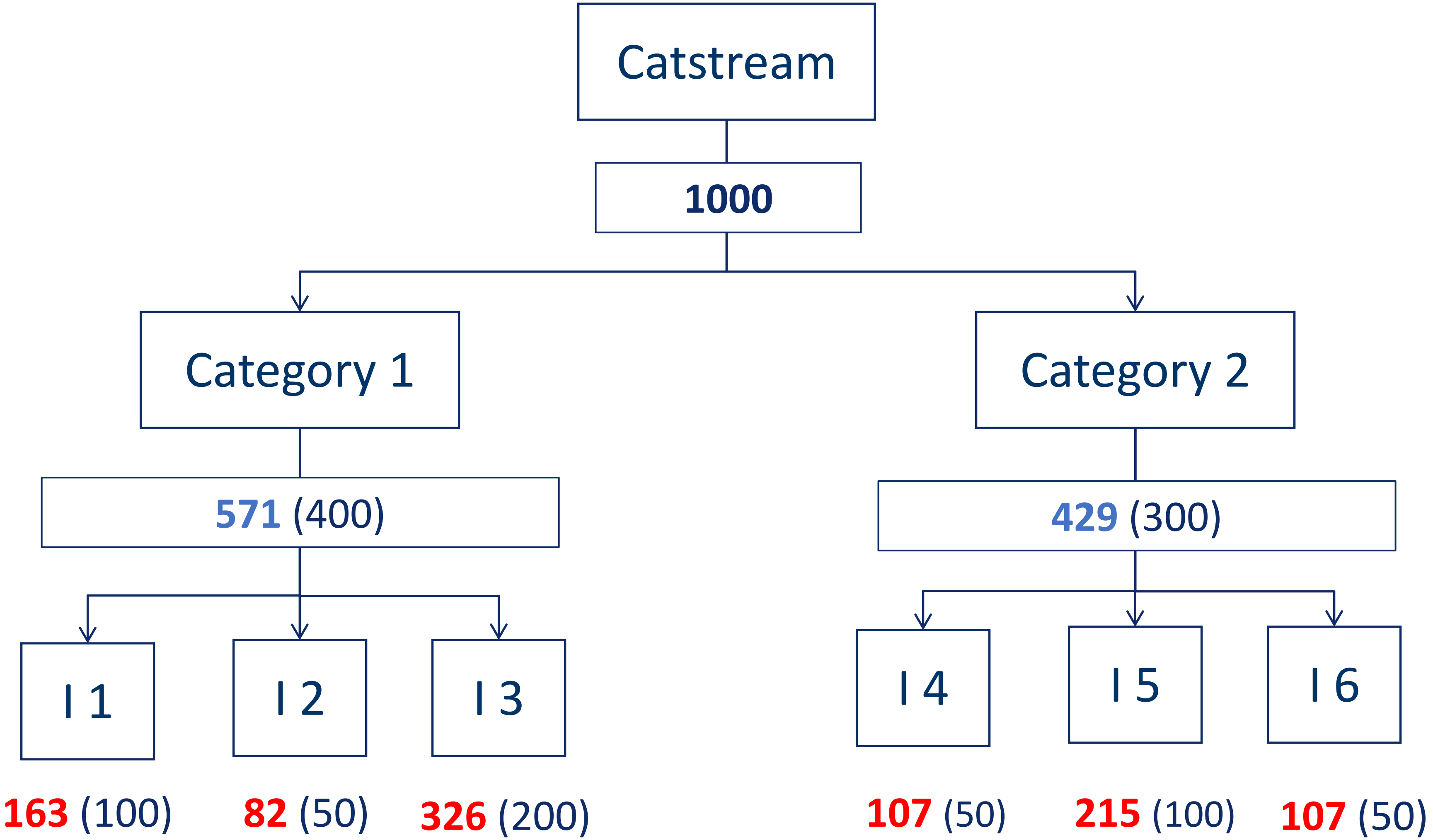

450
400
350
300
Jan2004
Jan2005
Jan2006
Jan2007
Jan2008
Jan2009
Jan2010
date
Historical Data
Reconciled Forecast
Reconciled Confidence Limits
Reconciled Forecast Lower 95%
Reconciled Forecast Upper 95%
Override
Final Forecast

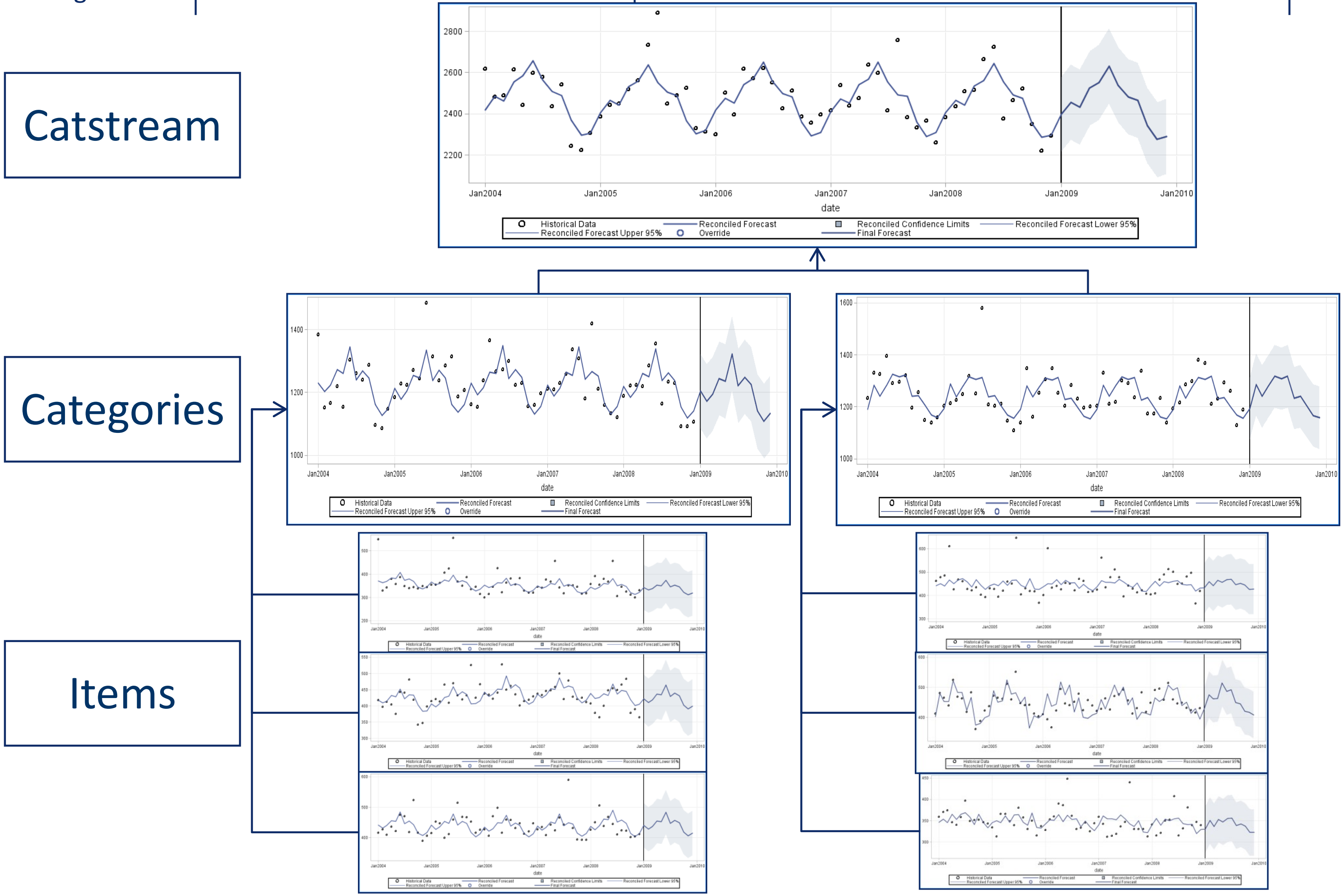

Categories
Items
date
Historical Data
Reconciled Forecast
Reconciled Confidence Limits
Reconciled Forecast Lower 95%
Reconciled Forecast Upper 95%
Override
Final Forecast


Catstream
Categories
Items
2800
2600
2400
2200
Jan2004
Jan2005
Jan2006
Jan2007
Jan2008
Jan2009
Jan2010
date
Historical Data
Reconciled Forecast
Reconciled Confidence Limits
Reconciled Forecast Lower 95%
Reconciled Forecast Upper 95%
Override
Final Forecast
1400
1200
1000
1600


Catstream
1000
Category 1
400
Category 2
300
I 1
100
I 2
50
I 3
200
I 4
50
I 5
100
I 6
50


Catstream
1000
Category 1
571
Category 2
429
I 1
I 2
I 3
I 4
I 5
I 6
100
50
200
50
100
50
100/350*571
50/350*571
200/350*571
50/200*429
100/200*429
50/200*429


Catstream
1000
Category 1
Category 2
571 (400)
429 (300)
I 1
I 2
I 3
I 4
I 5
I 6
163 (100)
82 (50)
326 (200)
107 (50)
215 (100)
107 (50)


Catstream
Categories
Items
2800
2600
2400
2200
Jan2004
Jan2005
Jan2006
Jan2007
Jan2008
Jan2009
Jan2010
date
Historical Data
Reconciled Forecast
Reconciled Confidence Limits
Reconciled Forecast Lower 95%
Reconciled Forecast Upper 95%
Override
Final Forecast

HSE UNIVERSITY • EST 1992 •