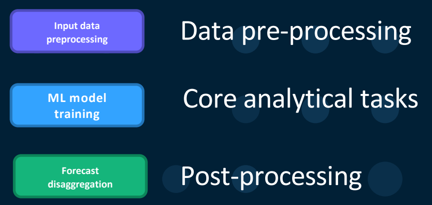
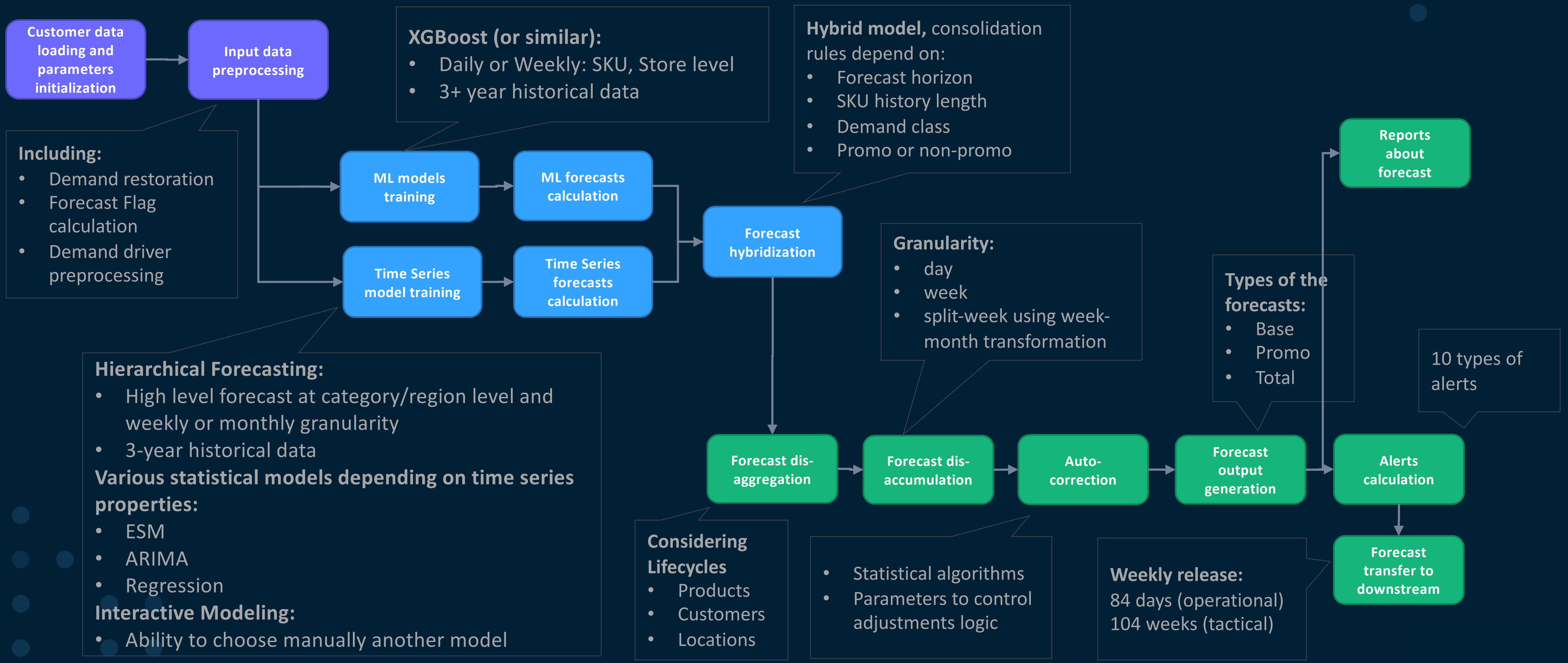
E2E forecasting flow overview





Definition of data entities

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**

DQ examples

Check values range

```
def check_val_range(data_quality_output, table, target_col, th=0):
    result = table[table[target_col] < th]
    if not result.empty:
        data_quality_output = pd.concat([data_quality_output, result])

    return data_quality_output
```

```
data_quality_output = check_val_range(data_quality_output, promo, 'PROMO_PRICE')
```

PRODUCT_ID	LOCATION_ID	CUSTOMER_ID	DISTR_CHANNEL_ID	PERIOD_START_DT	PERIOD_END_DT	PRICE	PRICE_TYPE	M

Check cross consistency

```
def check_cross_consistency(data_quality_output, df1, df2):
    common_cols = df1.columns.intersection(df2.columns)
    common_cols = list(common_cols[common_cols.str.contains('ID')])

    if common_cols == []:
        return data_quality_output

    df_merged = df1.drop_duplicates(common_cols).merge(df2.drop_duplicates(common_cols), on=common_cols,
                                   how='left', indicator=True)

    result = df_merged[df_merged['_merge'] == 'left_only']

    if not result.empty:
        data_quality_output = pd.concat([data_quality_output, result])

    return data_quality_output
```

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_forecast_attributes
FORECAST_ID
DIM1_LVL_ID
…
DIM5_LVL_ID
ASSORTMENT_TYPE
FORECAST_VARIANCE

out_dimension
DIMENSION_ID
DIMENSION_NM

Demand Restoration

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: $\lambda = 1$

		+Stock							+Stock							+Stock
Demand	...	1	0	1	2	0	2	0	1	1	0	3	2	0	1	...

Item 2: $\lambda = 0.4231$

		+Stock							+Stock							+Stock
Demand	...	0	0	0	1	0	1	0	0	0	0	1	0	2	0	...

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

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

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

$$L = P_{m_1}(x = k_1) \cdot P_{m_2}(x = k_2) \cdot \dots \cdot P_{m_T}\left(x = k_{m_T}\right) =$$

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

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

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

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

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

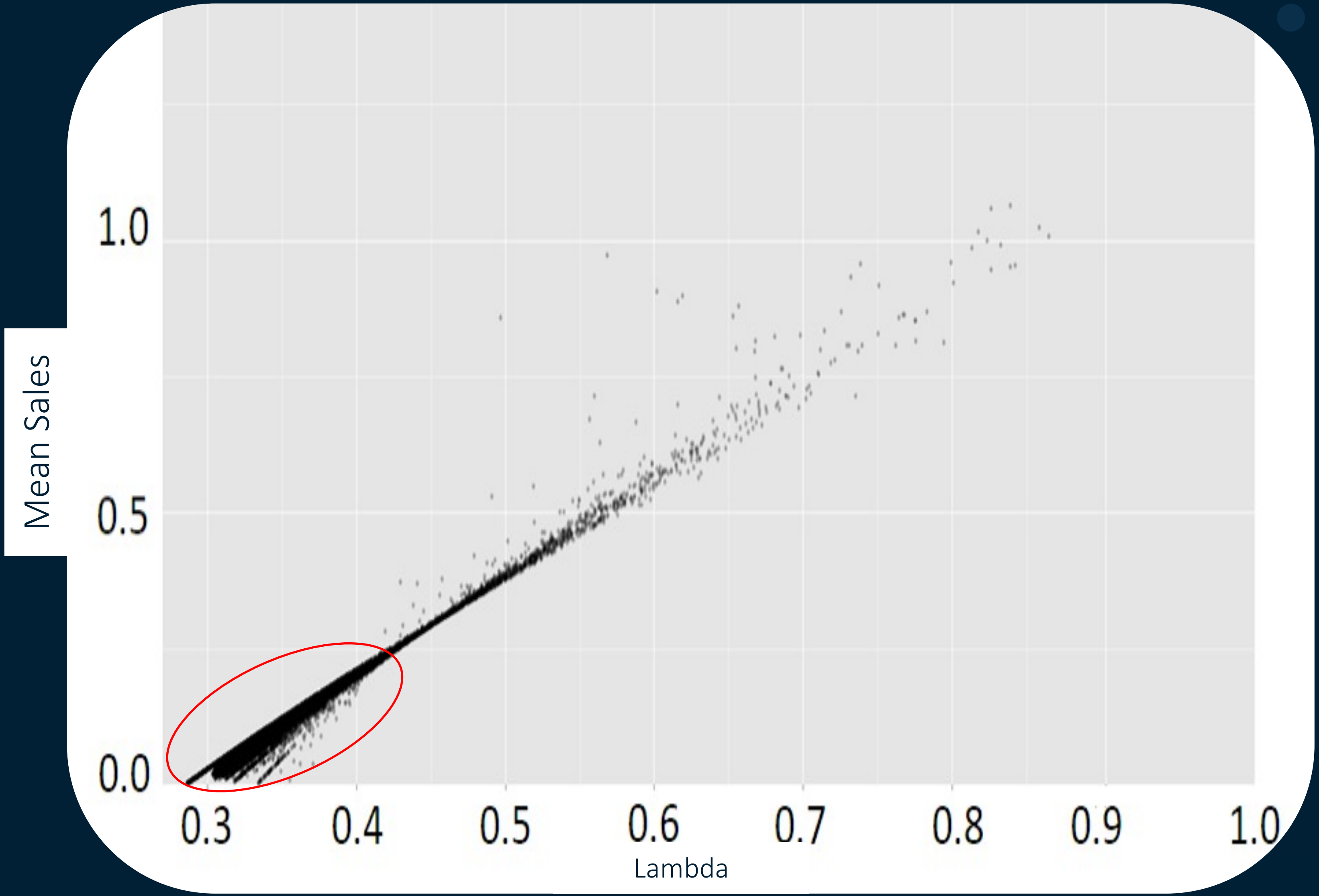
Solution:

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

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

Lambda vs Mean Sales

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



Demand Restoration results

Several approaches to restore demand

10%-20%

Realistic Uplift

Sales

Optimistic demand

Conservative demand

Pessimistic demand