Data Science in Retail - Exploring Demand Forecasting

Miguel Cabrera
Senior Data Scientist at NewYorker

@mfcabrera

Photo by Daniel Seßler on Unsplash

HELLO!

I’m Miguel Cabrera
Senior Data Scientist at NewYorker

@mfcabrera

THE NEW YORKER

NEW YORKER

THE AGENDA FOR TODAY

INTRODUCTION

Basic concepts

REQUIREMENTS

What do we need to take into account

MODELS

What models can help us solve this problem

PRACTICE

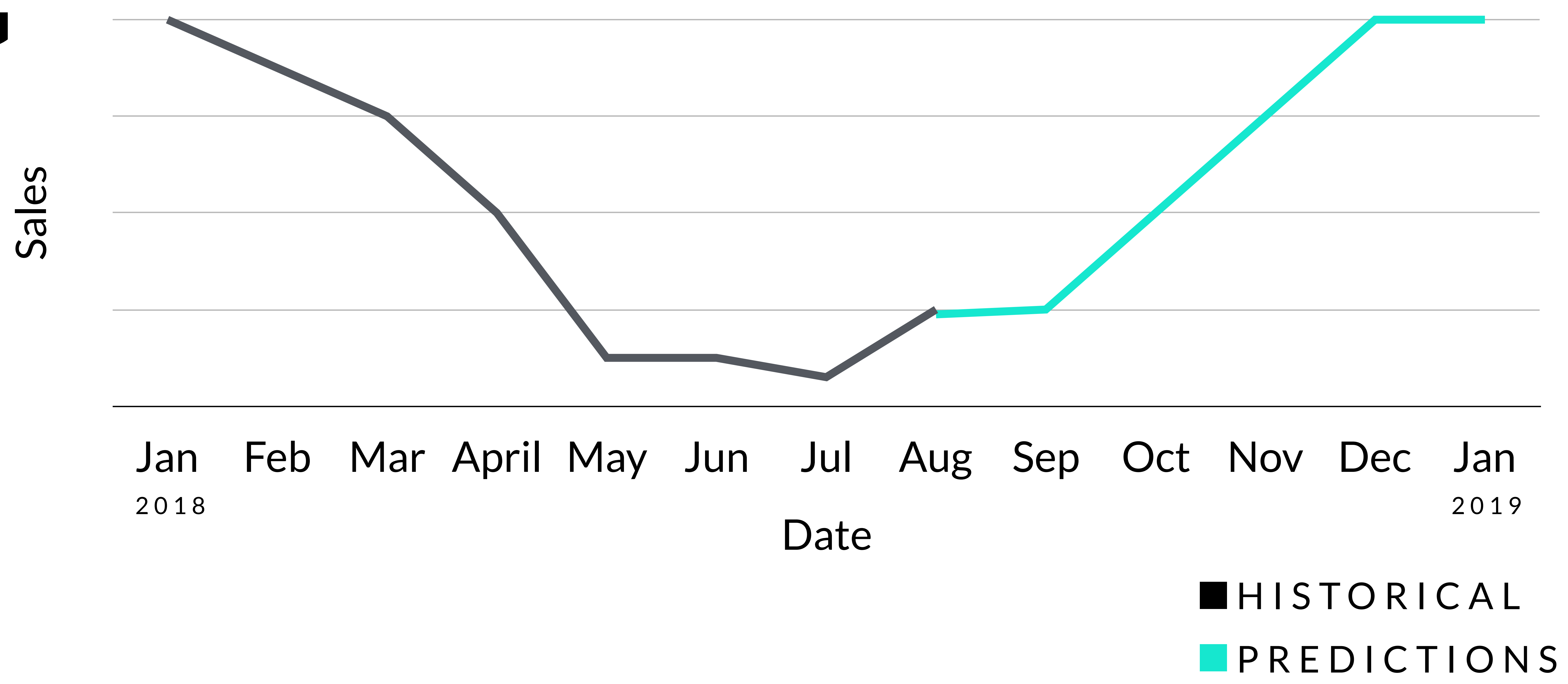
Common techniques and pattern to tackle this problem

INTRODUCTION

DEMAND FORECASTING

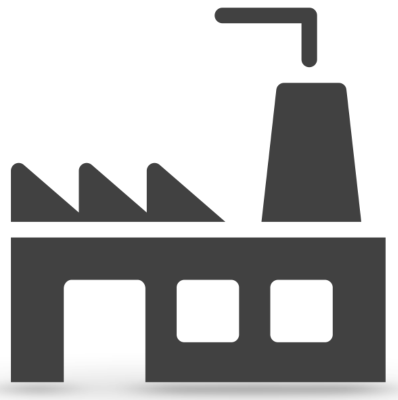
Demand Forecasting refers to predicting future **demand** (or **sales**), assuming that the factors which affected demand in the past and are affecting the present and will still have an influence in the future. [1]

Date	Sales
Feb 2018	3500
Mar 2018	3000
April 2018	2000
May 2018	500
Jun 2018	500
...	...
T	1000
T+1	??
T+2	??
T+3	??
...	??
T+n	??



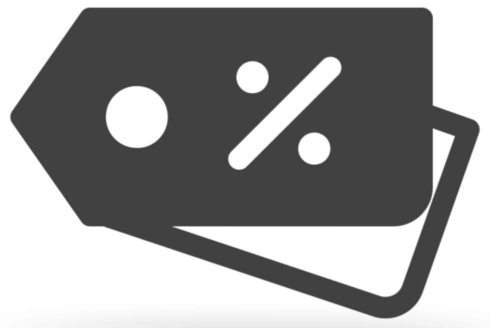
APPLICATIONS

PRODUCTION PLANNING

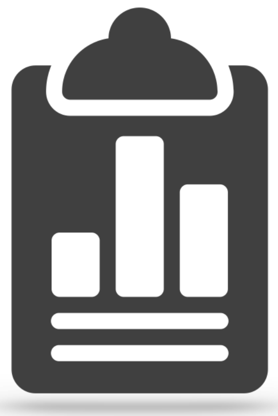


REPLENISHMENT

DISCOUNT & PROMOTIONS



FINANCIAL PLANNING



CONSTRAINTS

- Strong **relationship** between **garments** and **weather** make sales seasonal and prone to **unpredictability**
- Sales are **disturbed** by **exogenous variables** like end-of-season sales, promotions, competition, marketing and purchasing power of consumers.
- **Fashion** trends create **volatility** in consumer **demands**, the design and style have to be up to date
- **High** product **variety**. Many colours alternatives and various sizes.
- Most of the items **are not renewed** for the next collection and even basic products might change slightly due to fashion trends.
- Consumers are very **unfaithful** and generally their selection is based on the price of the product.

REQUIREMENTS

MULTI-HORIZON

Many decisions are based on sales forecasting and should be considered in a sufficient time based on lead times

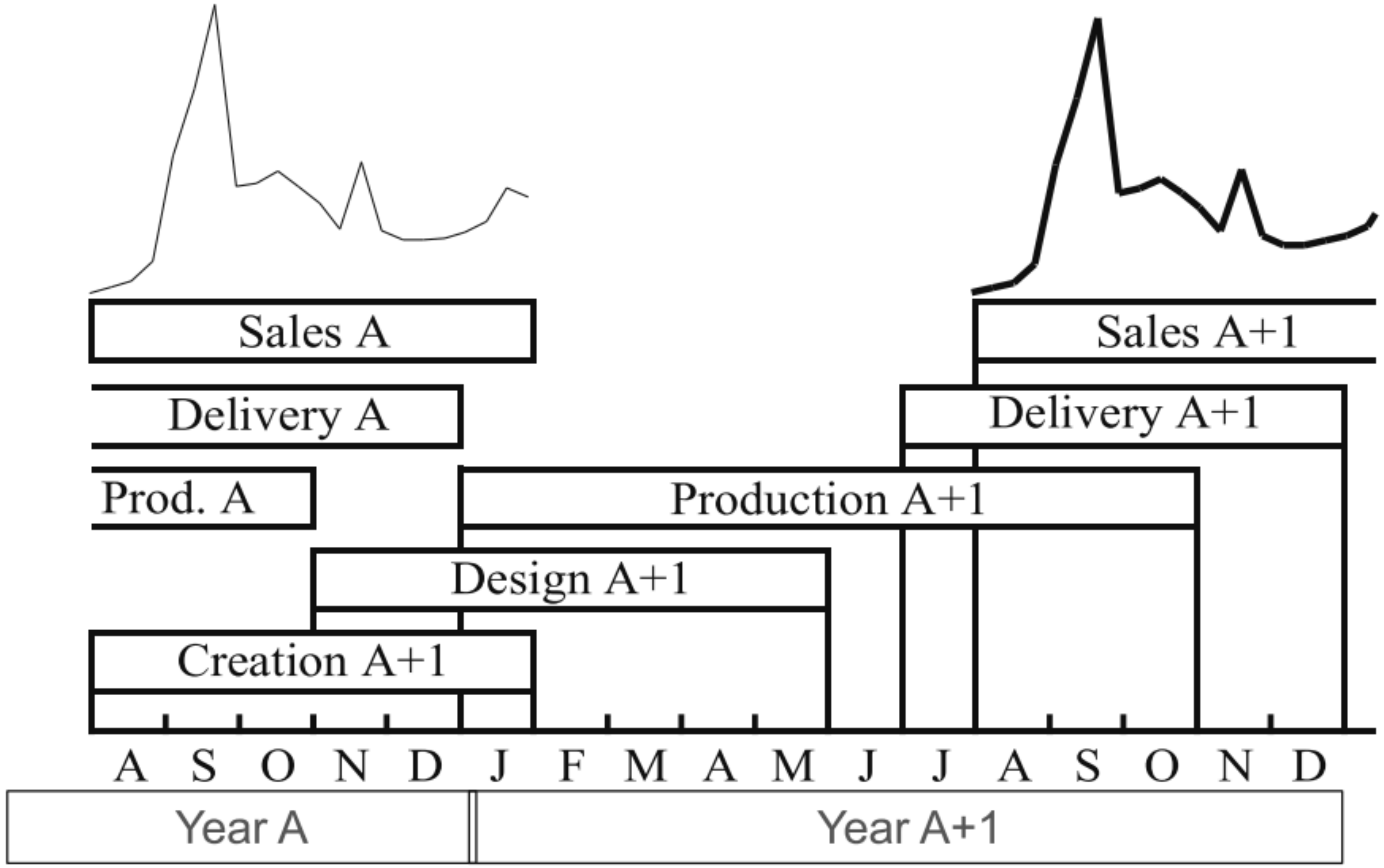
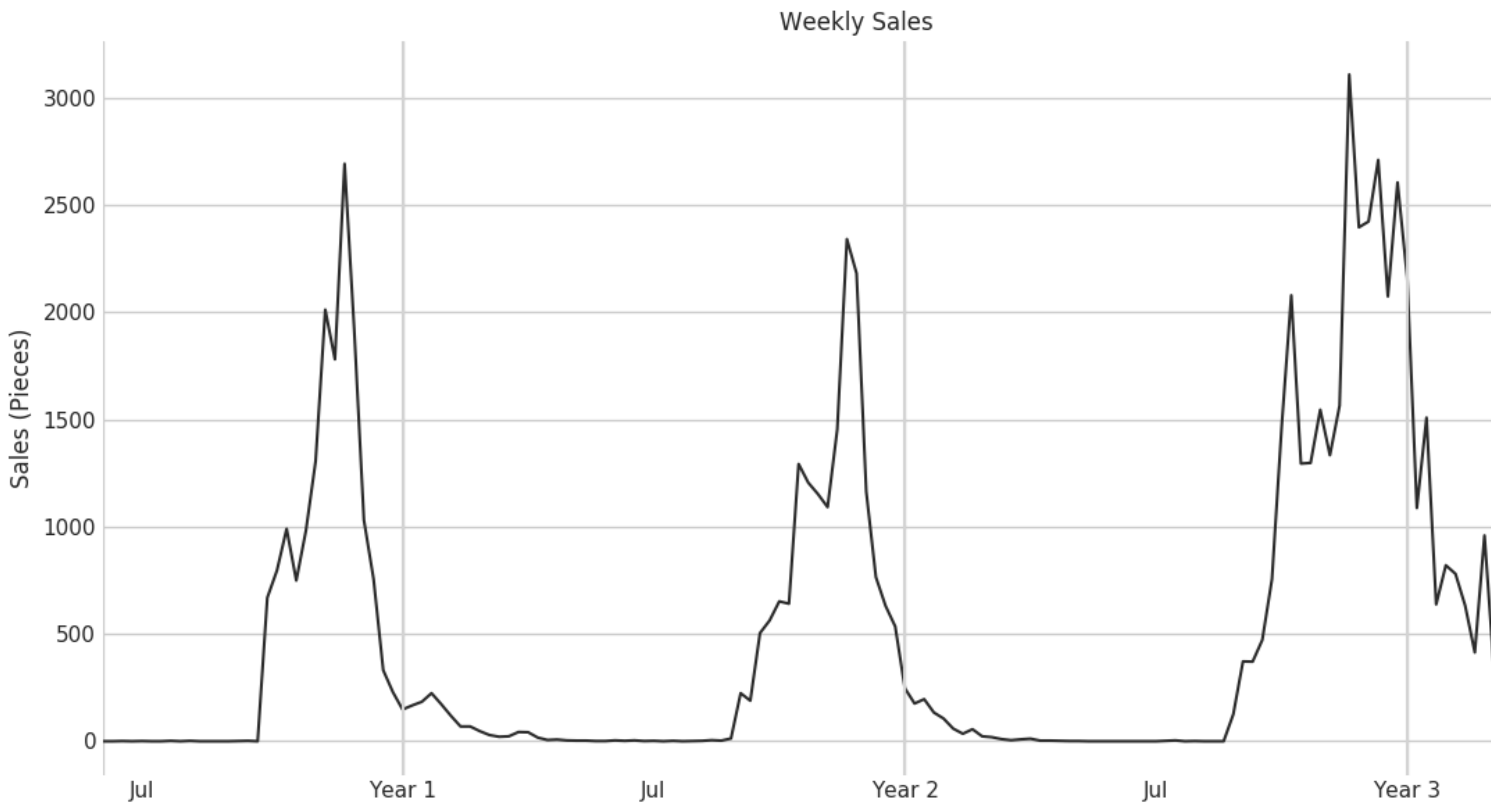
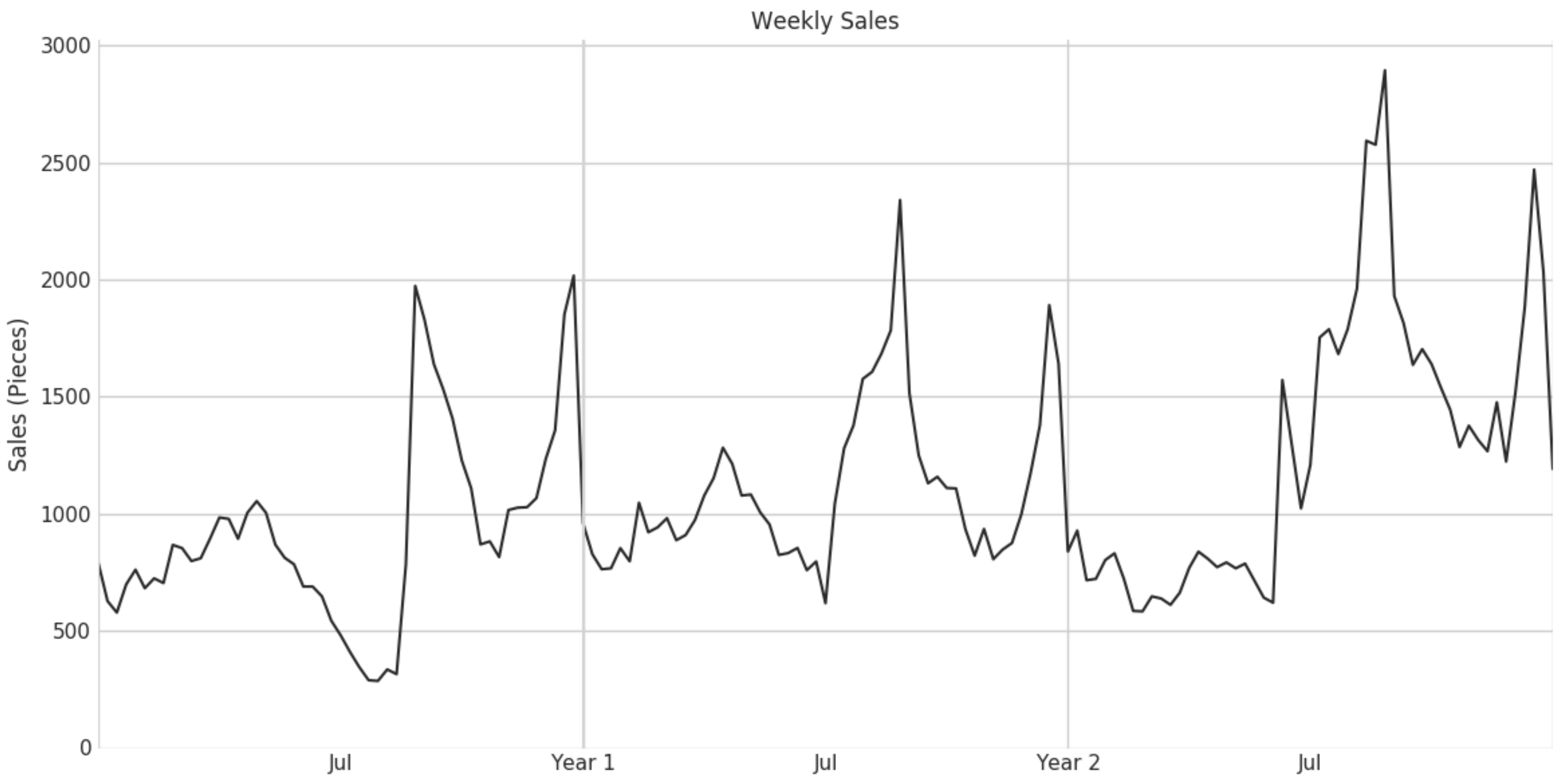


Image Source: Thomassey, S. (2014). Sales Forecasting in Apparel and Fashion Industry: A Review. In Intelligent Fashion Forecasting Systems: Models and Applications (pp. 9–27). https://doi.org/10.1007/978-3-642-39869-8_2

SEASONALITY

PRODUCTS ARE VERY SENSITIVE TO SEASONAL VARIATIONS





EXOGENOUS VARIABLES

- Item Features and Fashion trends
- Retailing strategy (stores, location, location in store).
- Marketing strategy
- Macroeconomic phenomena
- Calendar information (Holidays, special dates)
- Competition
- Weather

MODELS

REQUIREMENTS - IMPLICATIONS

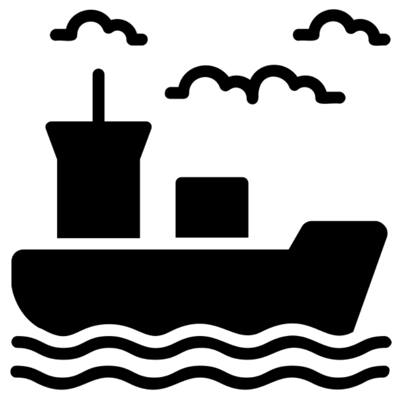
Multiple products

Multiple time series

Different product lifecycles

Highly non-stationary sales

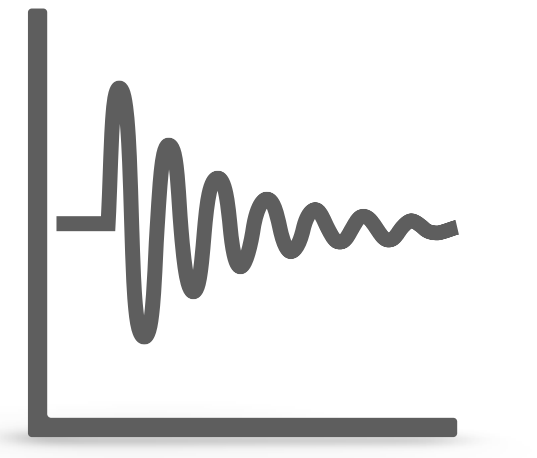
Different horizons



Multi-horizon predictions

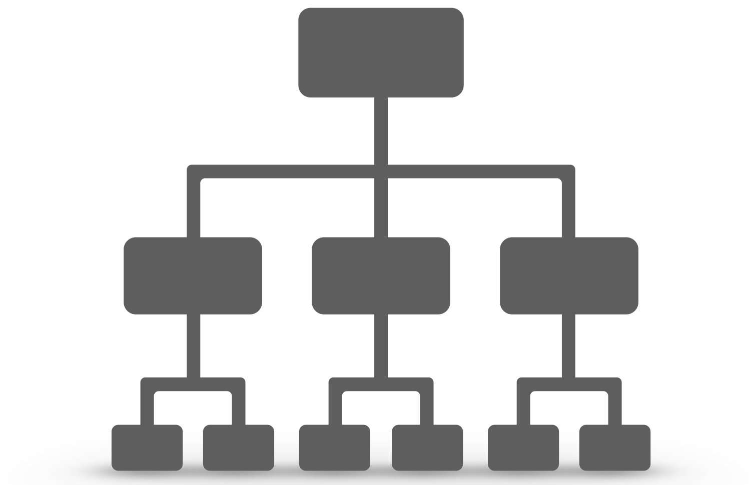
MODELING APPROACHES

TOOLS AVAILABLE



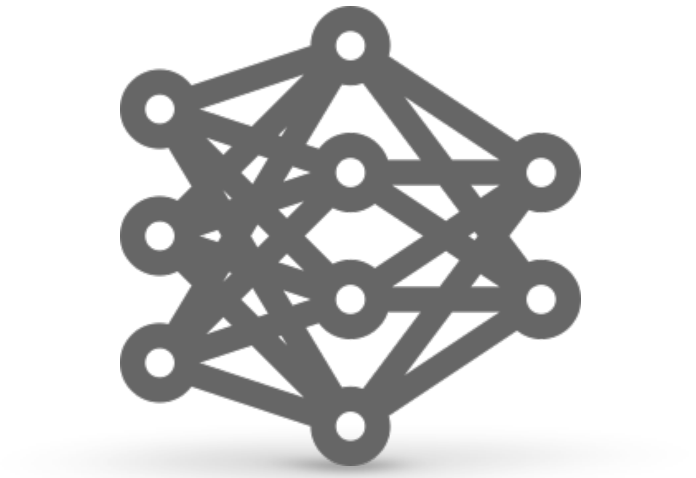
TIME SERIES MODELS

- (S)ARIMA
- (G)ARCH
- VAR
- FB Prophet



MACHINE LEARNING

- Linear Regression
- SVM
- Gaussian Process
- Tree Based Models
    - Random Forests
    - XGBoost
    - Catboost
    - LightGBM



DEEP LEARNING

- MLP
- RNN
- LSTM
- SEQ2SEQ

MODELING APPROACHES

MODEL SCORE CARD

Characteristic / Requirement	Score
Highly non-stationary	
Multiple time series	
Multi-horizon forecast	
Model interpretability	
Model Capability	
Computational Efficiency	

ARIMA

BASIC CONCEPTS

Auto-**R**egressive **I**ntegrated **M**oving **A**verage

ARIMA(p, d, q)

SARIMA(p, d, q)x(Q,D,P,m)

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

AR(p)

Past Values

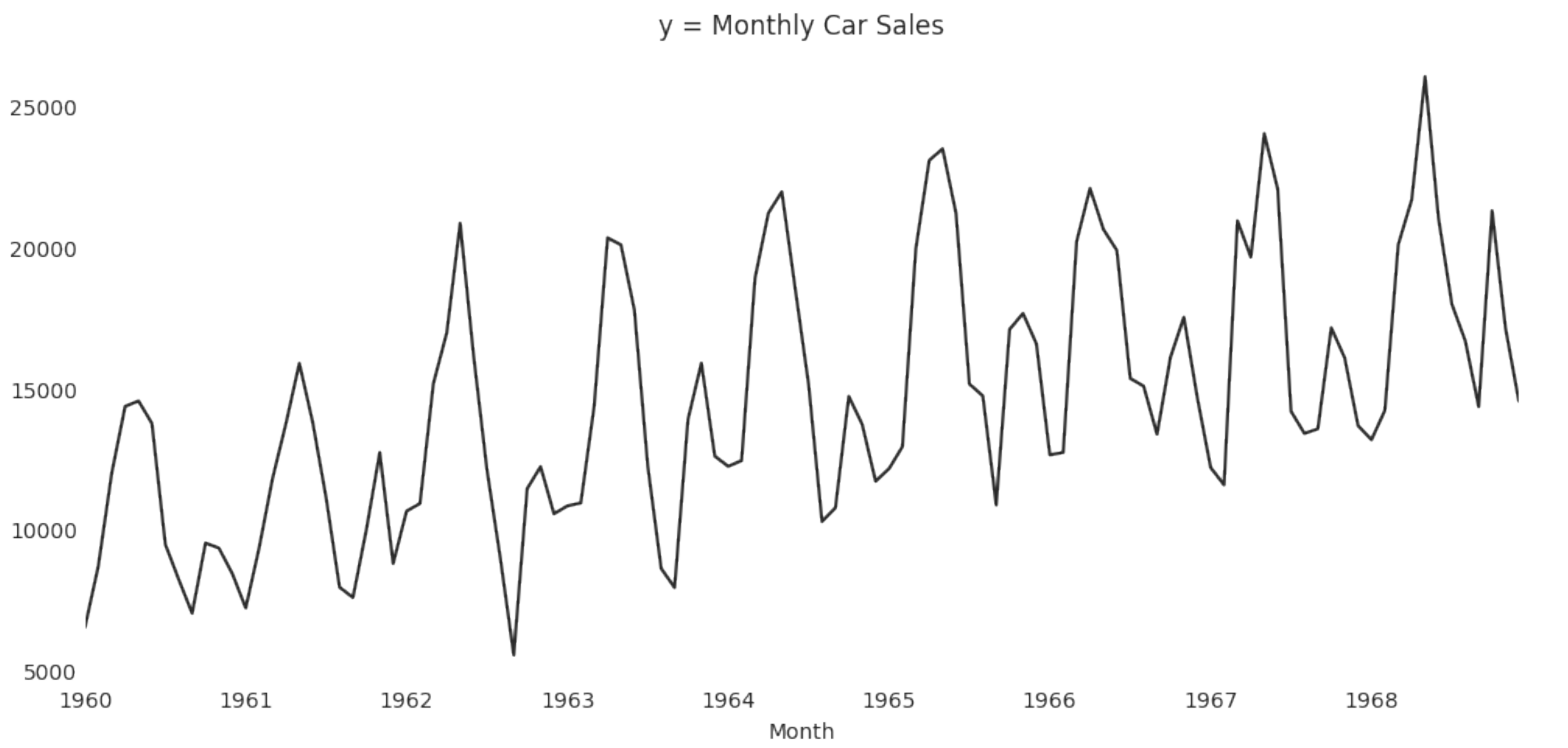
MA(q)

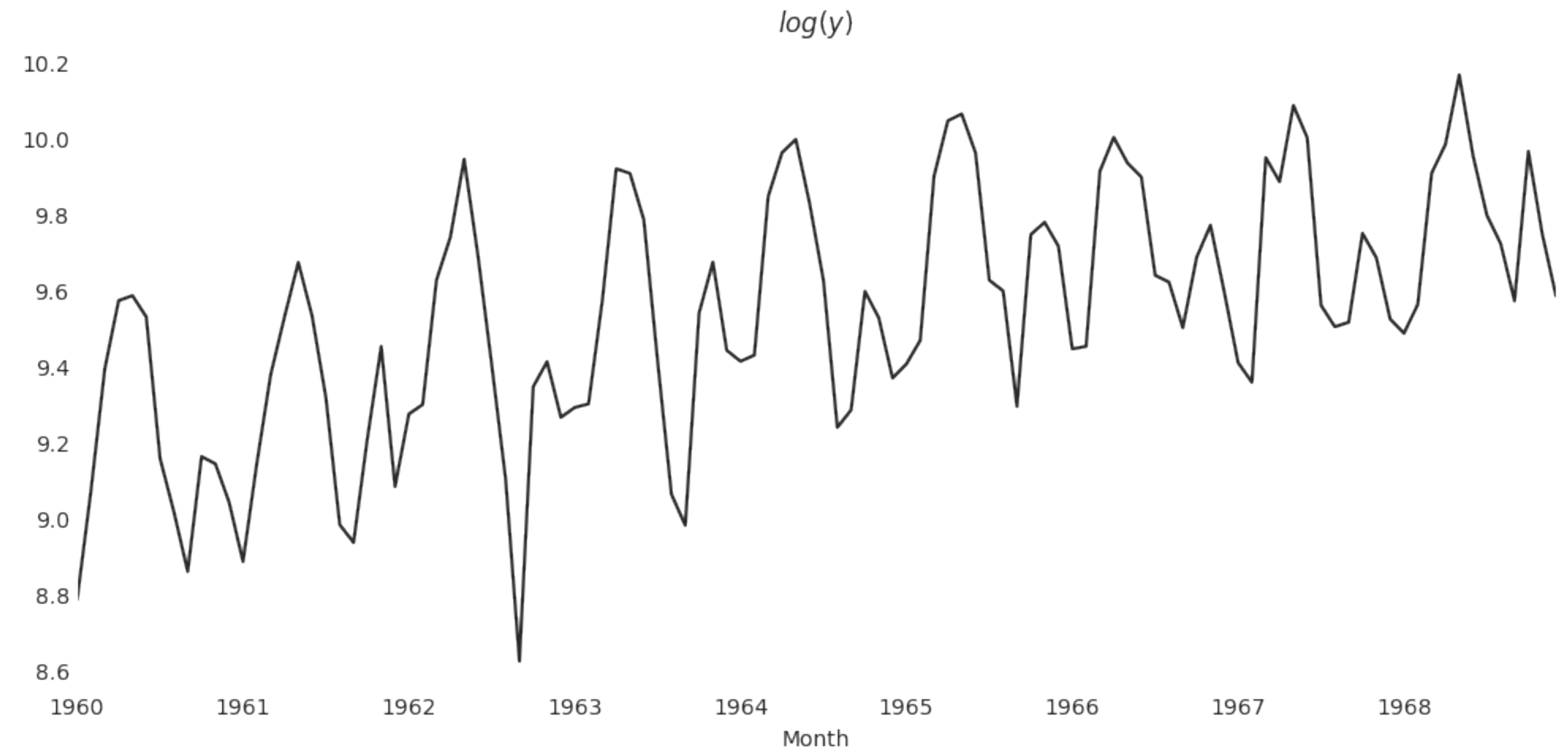
Past Errors

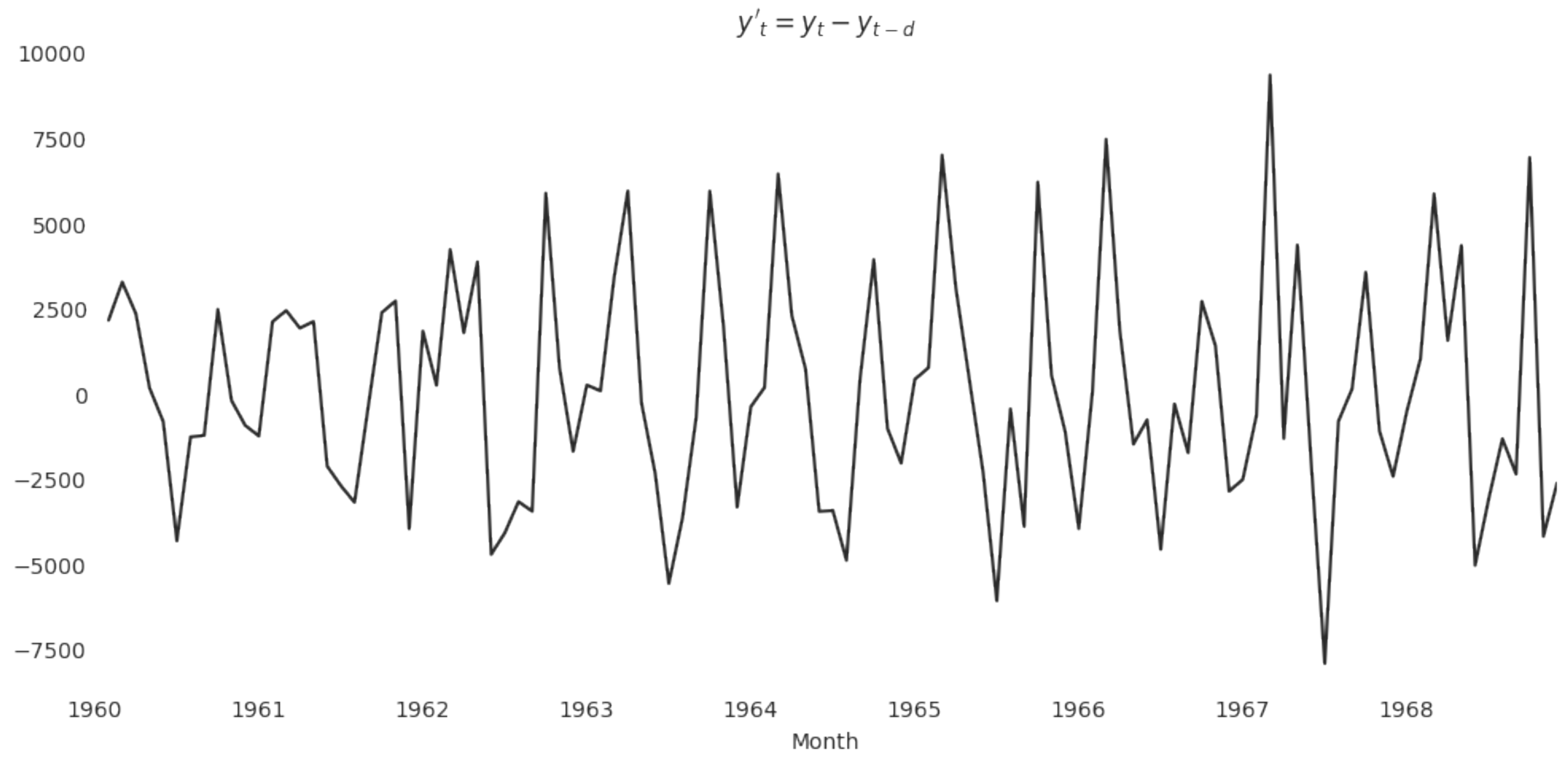
$$y'_t = y_t - y_{t-d}$$

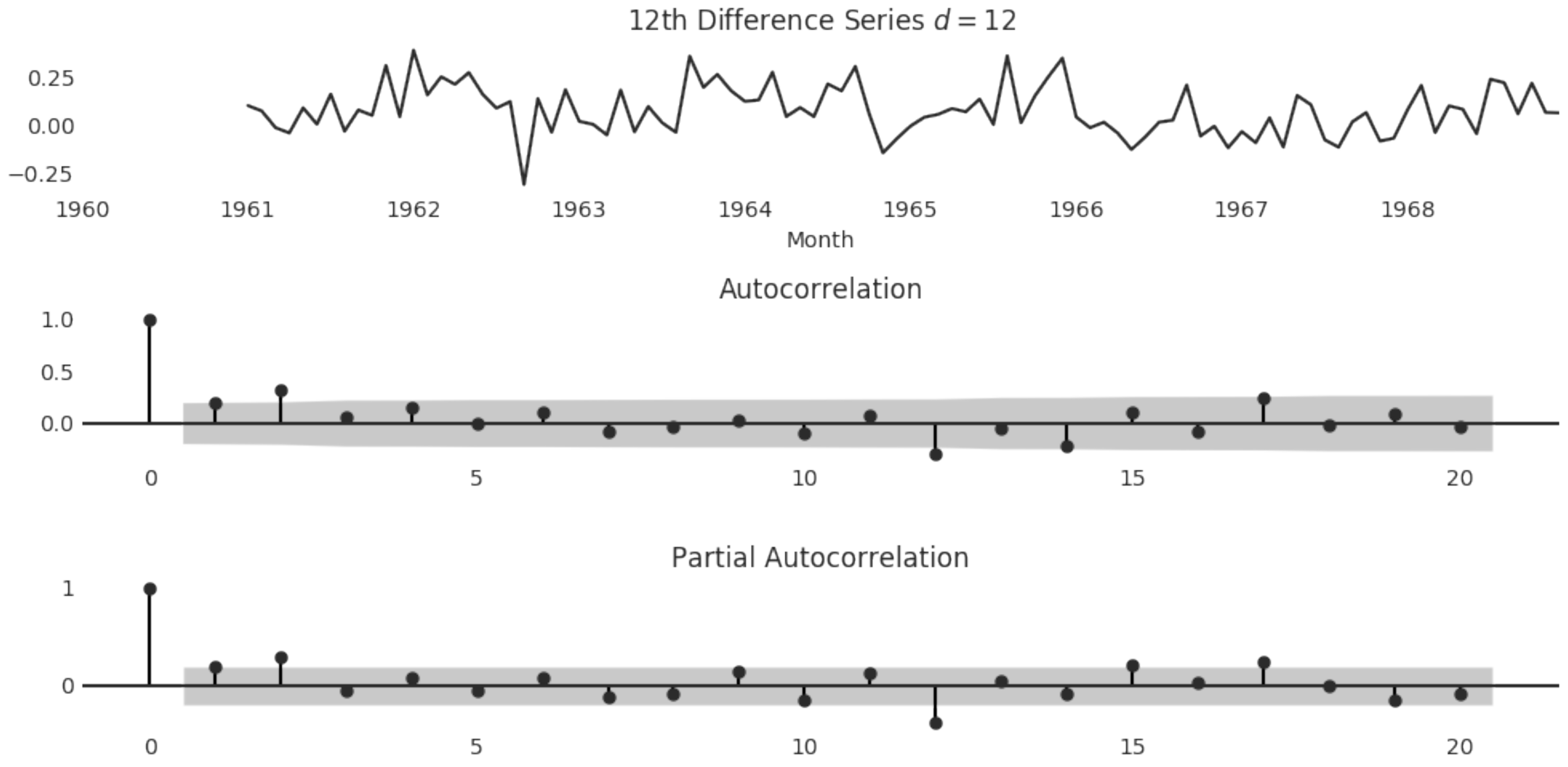
ARIMA

EXAMPLE





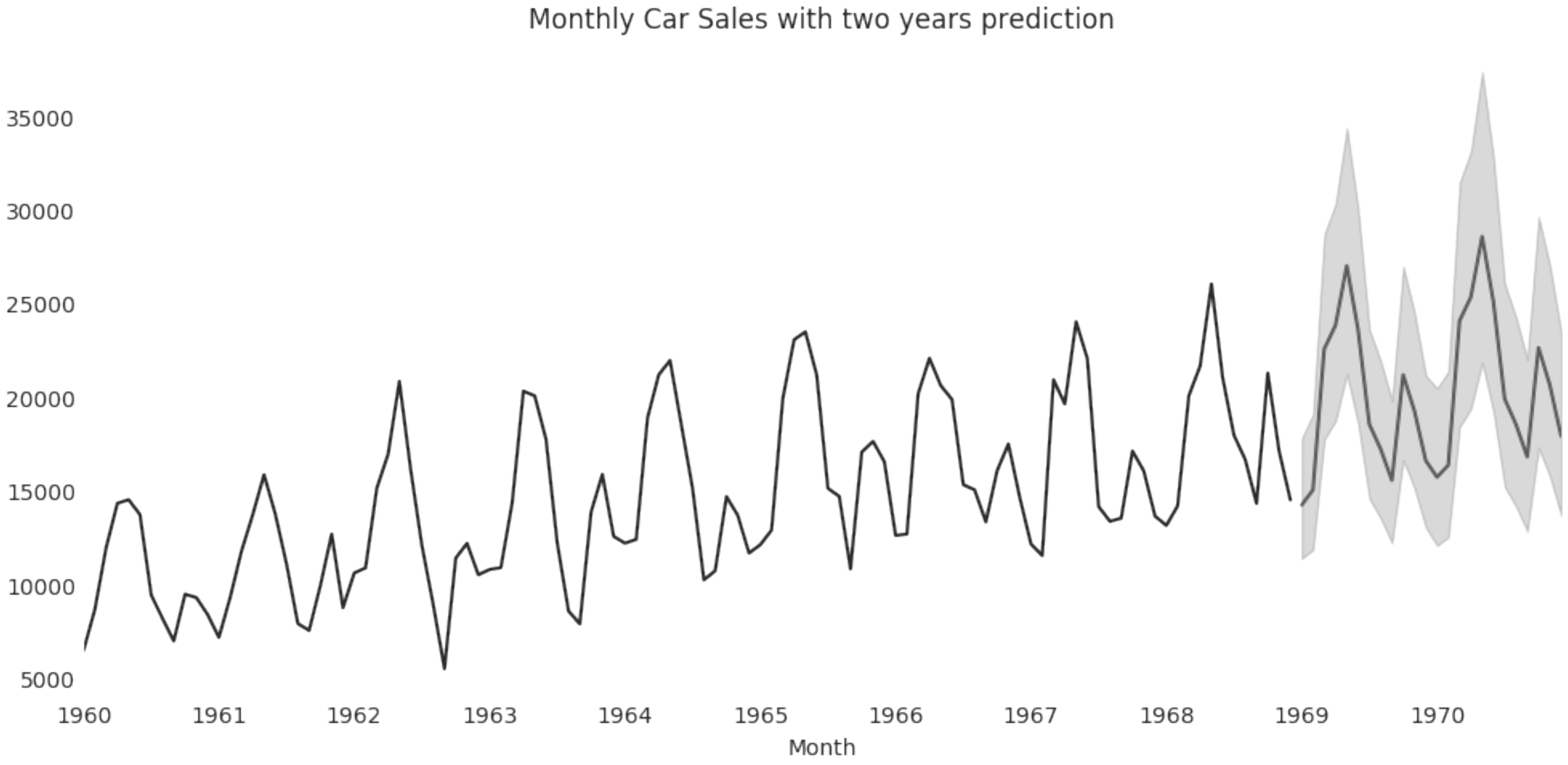




ARIMA

EXAMPLE

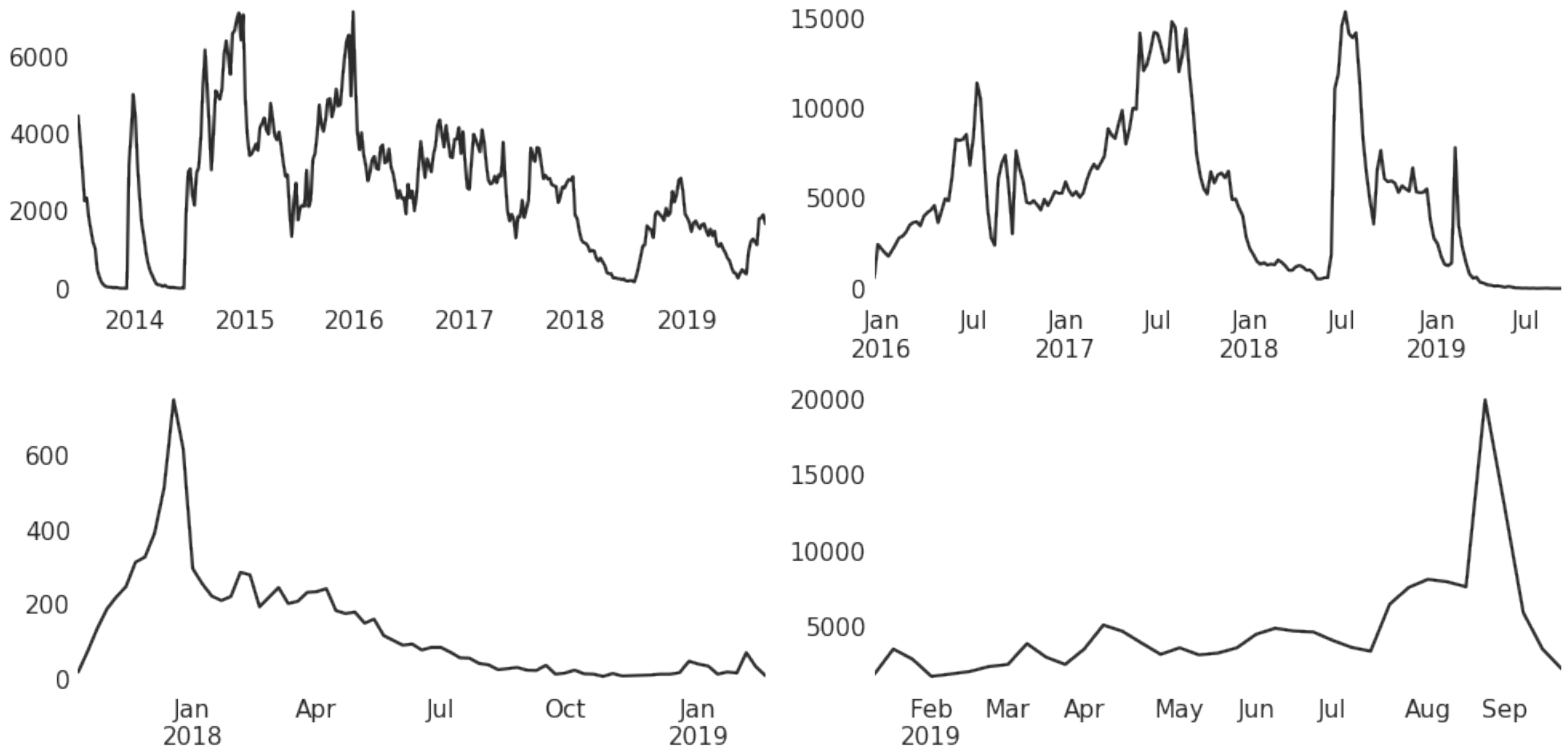
- Study ACF/PACF charts and determine the parameter or use an automated algorithm.
- Seasonal pattern (Strong correlation between y'_t and y'_{t-12})
- Algorithm found: SARIMAX(1, 1, 1)x(0, 1, 1)^12



TIME SERIES MODELS

LIMITATIONS

Characteristic / Requirement	Score
Highly non-stationary	Limited
Multiple time series	Limited
Multi-horizon forecast	Yes
Model interpretability	High
Model Capability	Low
Computational Efficiency	High

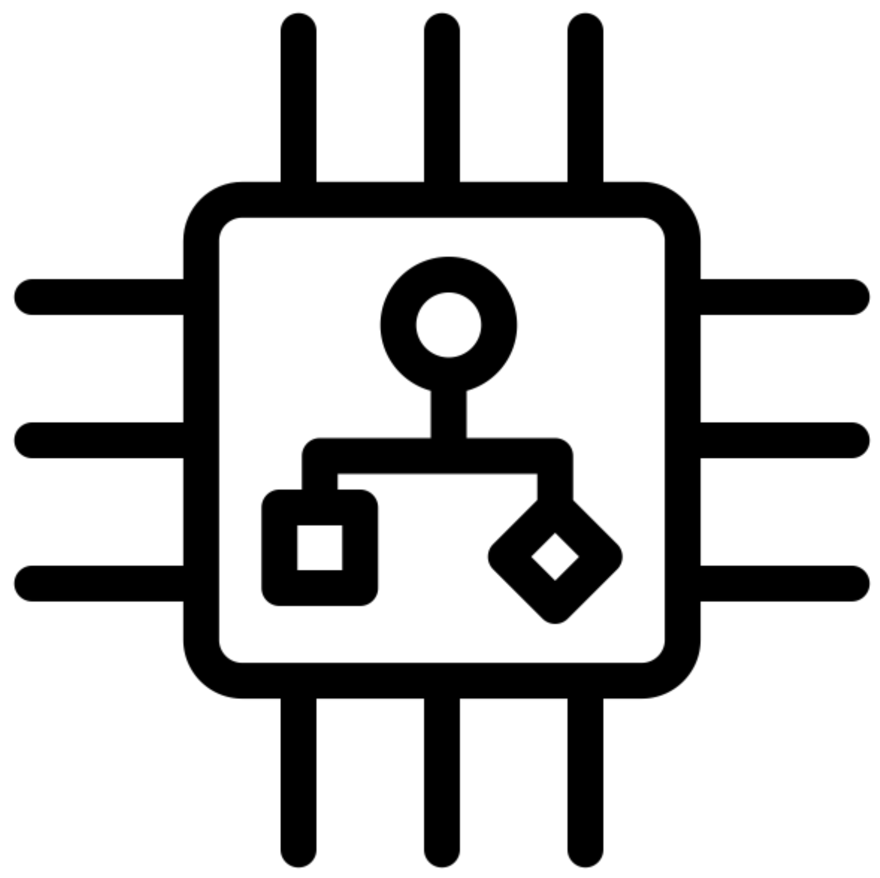


Sample plots of fashion product sales

MACHINE LEARNING

MACHINE LEARNING MODELS ARE MORE FLEXIBLE

- Additional features in the model.
- No assumption about the demand distribution.
- One single model can handle many or all products.
- Feature Engineering is very important.



MACHINE LEARNING - FEATURES

FEATURE ENGINEERING IS AN IMPORTANT STEP IN THE MACHINE LEARNING APPROACH

SOURCE	EXTRACTION	ENCODING
Sales Data	Time Series, moving averages, statistics, lagged features, stock	Numerical
Product Attributes	category, brand, color, size, style, identifier	One Hot Encoding
Time	Day of week, month of year, number of week, season	Feature Hashing
Location	Holiday, weather, macroeconomic information	Embeddings

→ FEATURES

MACHINE LEARNING - FEATURES

FEATURE ENGINEERING IS AN IMPORTANT STEP IN THE MACHINE LEARNING APPROACH

SOURCE

	date	product_id	color	price
0	2018-08-27	4428029	BLACK	4.95
1	2018-09-03	4428029	BLACK	4.95
2	2018-09-10	4428029	BLACK	4.95
3	2018-09-17	4428029	BLACK	4.95
4	2018-09-24	4428029	BLACK	4.95
5	2018-10-01	4428029	BLACK	4.95
6	2018-10-08	4428029	BLACK	4.95
7	2018-10-15	4428029	BLACK	4.95
8	2018-10-22	4428029	BLACK	4.95
9	2018-10-29	4428029	BLACK	4.95

ENCODING

	month	week_of_year	price_bin	color_black	color_red	color_blue	color_yellow	color_white	color_navy	color_pink
0	8	35	0	1	0	0	0	0	0	0
1	9	36	0	1	0	0	0	0	0	0
2	9	37	0	1	0	0	0	0	0	0
3	9	38	0	1	0	0	0	0	0	0
4	9	39	0	1	0	0	0	0	0	0
5	10	40	0	1	0	0	0	0	0	0
6	10	41	0	1	0	0	0	0	0	0
7	10	42	0	1	0	0	0	0	0	0
8	10	43	0	1	0	0	0	0	0	0
9	10	44	0	1	0	0	0	0	0	0

MACHINE LEARNING - MODELS

SOME OF MODELS IN THE ZOO

LINEAR REGRESSION

Estimate the independent variable as the linear expression of the features.

- Least Squares
- Ridge / Lasso
- Elastic Net
- ARIMA + X

TREE BASED

Use decision trees to learn the characteristics of the data to make predictions

- Regression Tree
- Random Forest
- Gradient Boosting
  - Catboost
  - LightGBM
  - XGBoost

SUPPORT VECTOR REGRESSION

Minimise the error within the support vector threshold using a non-Linear kernel to model non-linear relationships.

- NuSVR
- LibLinear
- LibSVM
- SKLearn

MACHINE LEARNING

LIMITATIONS

Characteristic / Requirement	Score
Highly non-stationary	Yes
Multiple time series	Yes
Multi-horizon forecast	Yes
Model interpretability	Medium
Model Capability	Medium
Computational Efficiency	Medium

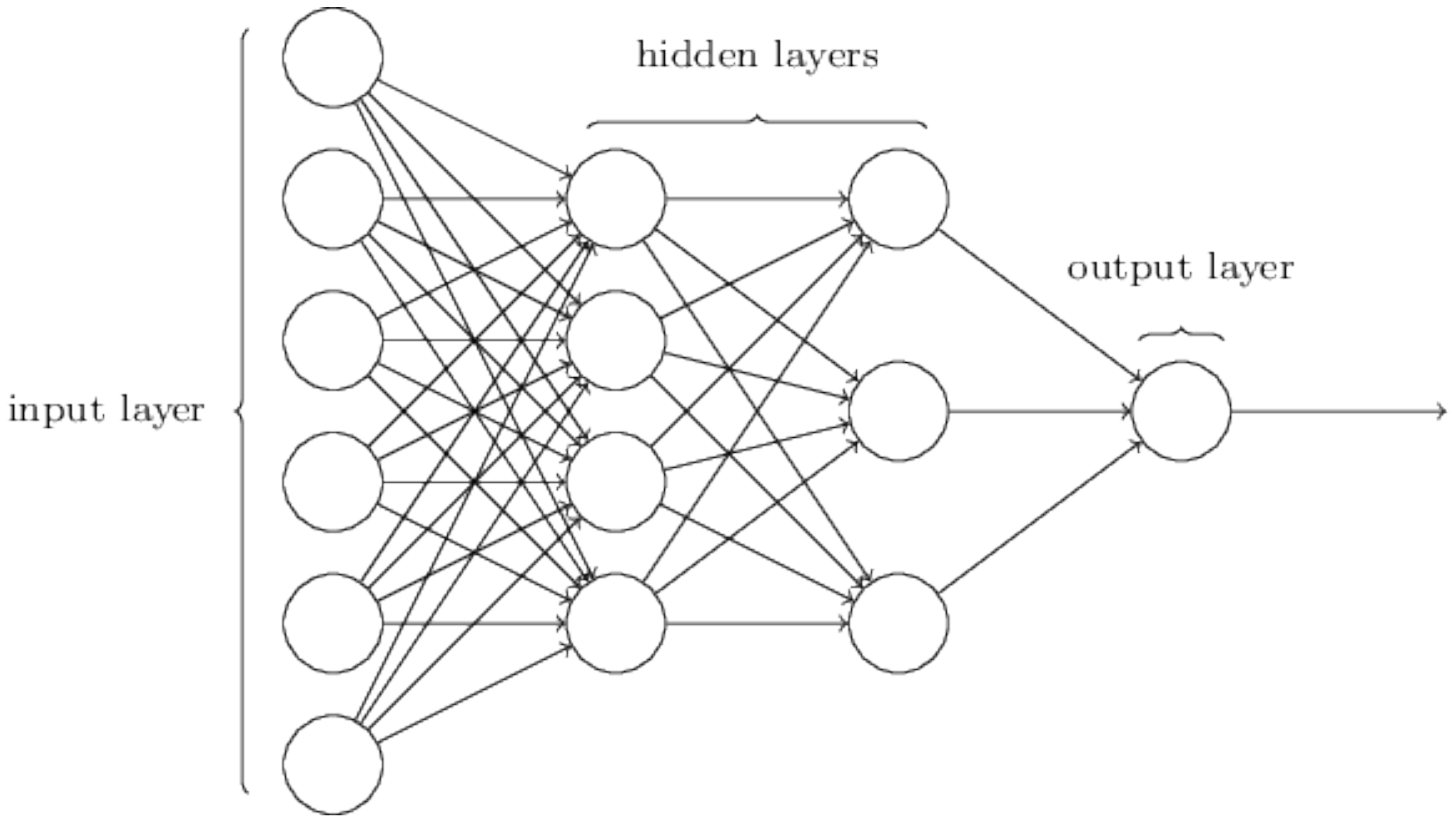
- Requires expert knowledge
- Time consuming feature engineering required
- Some features are difficult to capture

DEEP LEARNING - MODELS

SOME OF MODELS IN THE ZOO

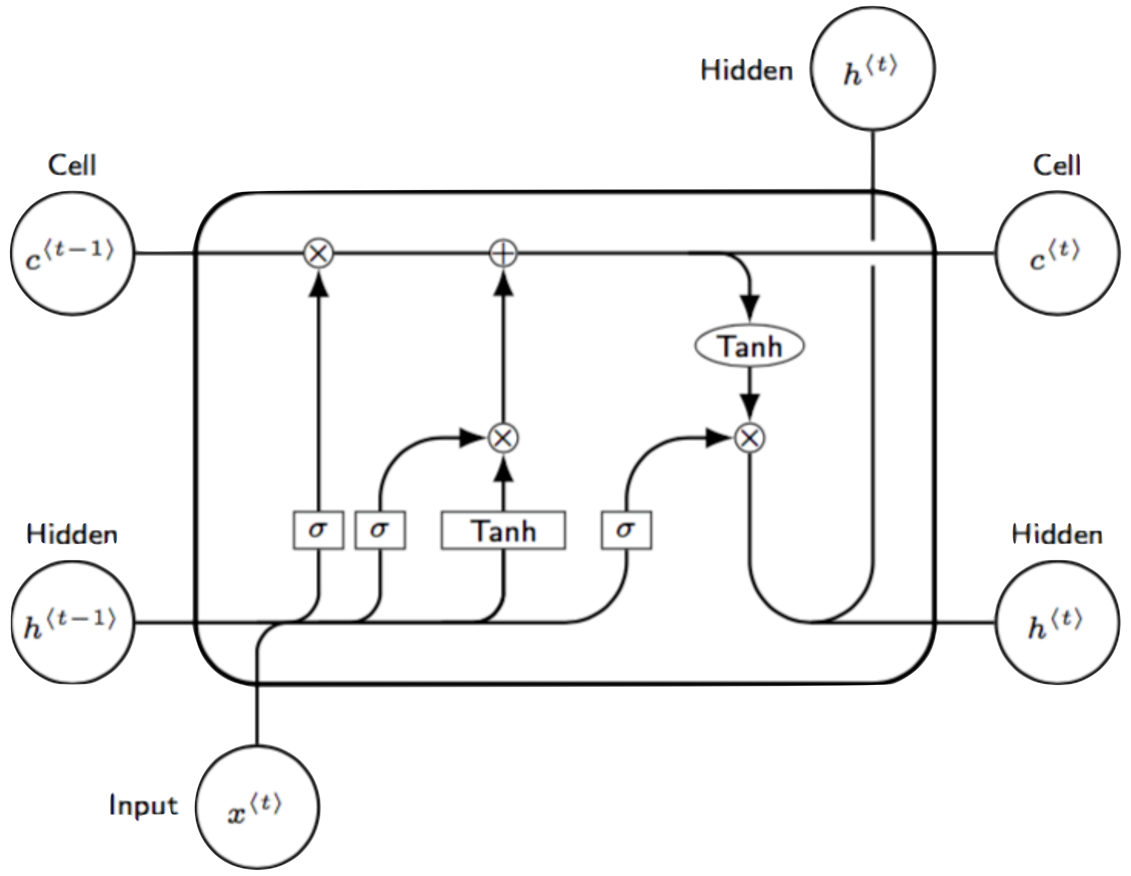
MULTILAYER PERCEPTRON

Fully connected multilayer artificial neural network.



LONG SHORT TERM MEMORY

A type of recurrent neural network used for sequence learning.
Cell states updated by gates.
Used for speech recognition, language models, translation, etc.



SEQ2SEQ

Encoder decoder architecture.

It uses two RNN that will work together trying to predict the next state sequence from the previous sequence.

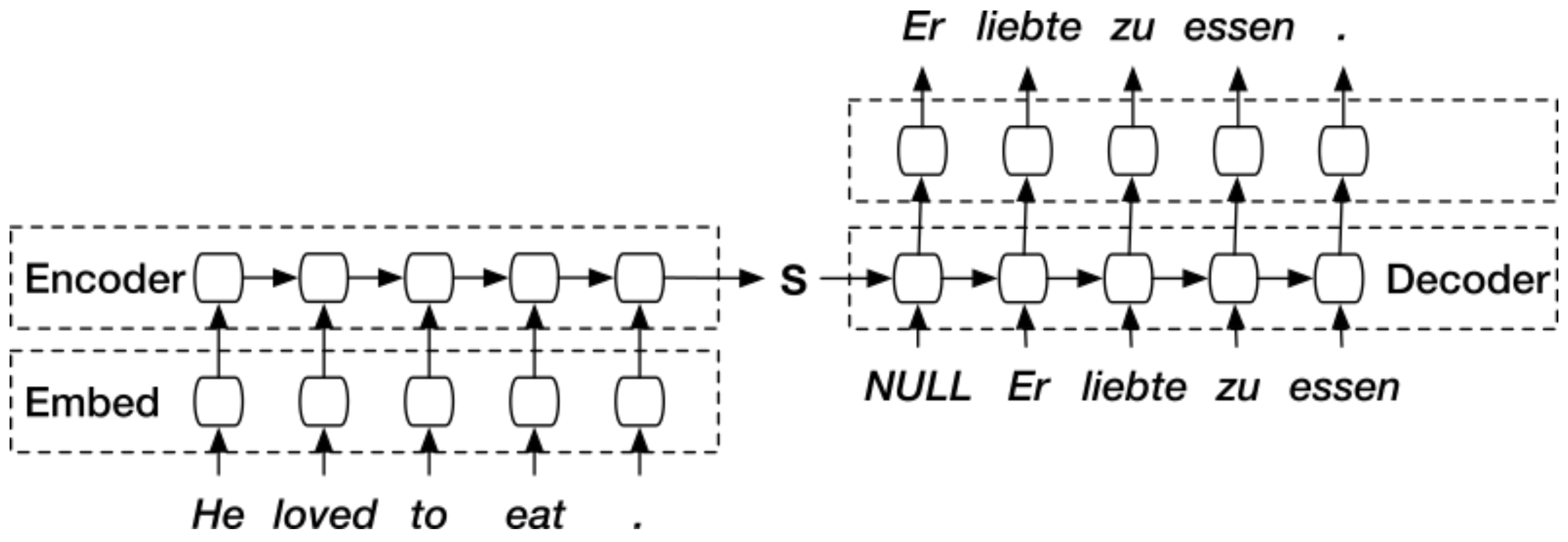
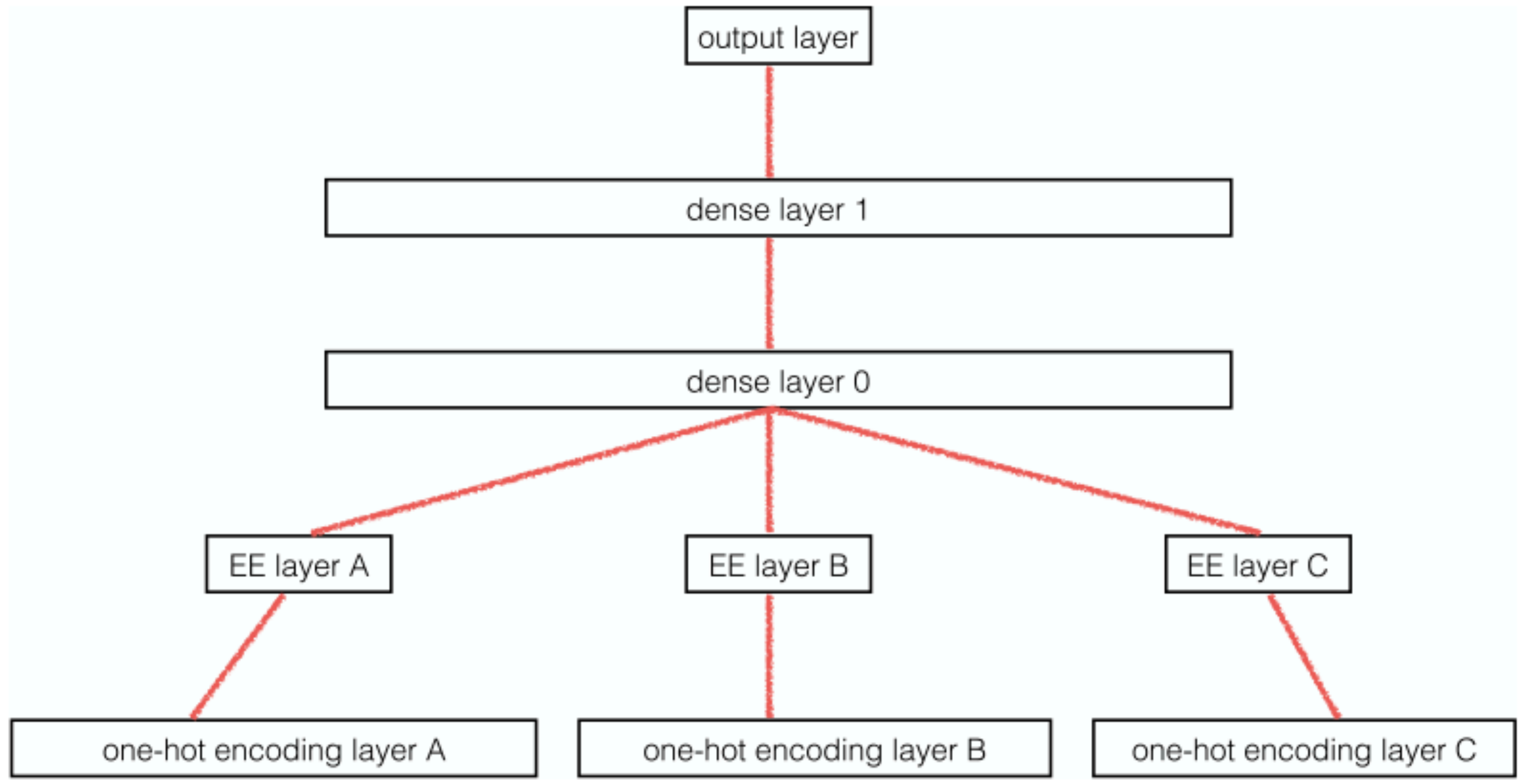
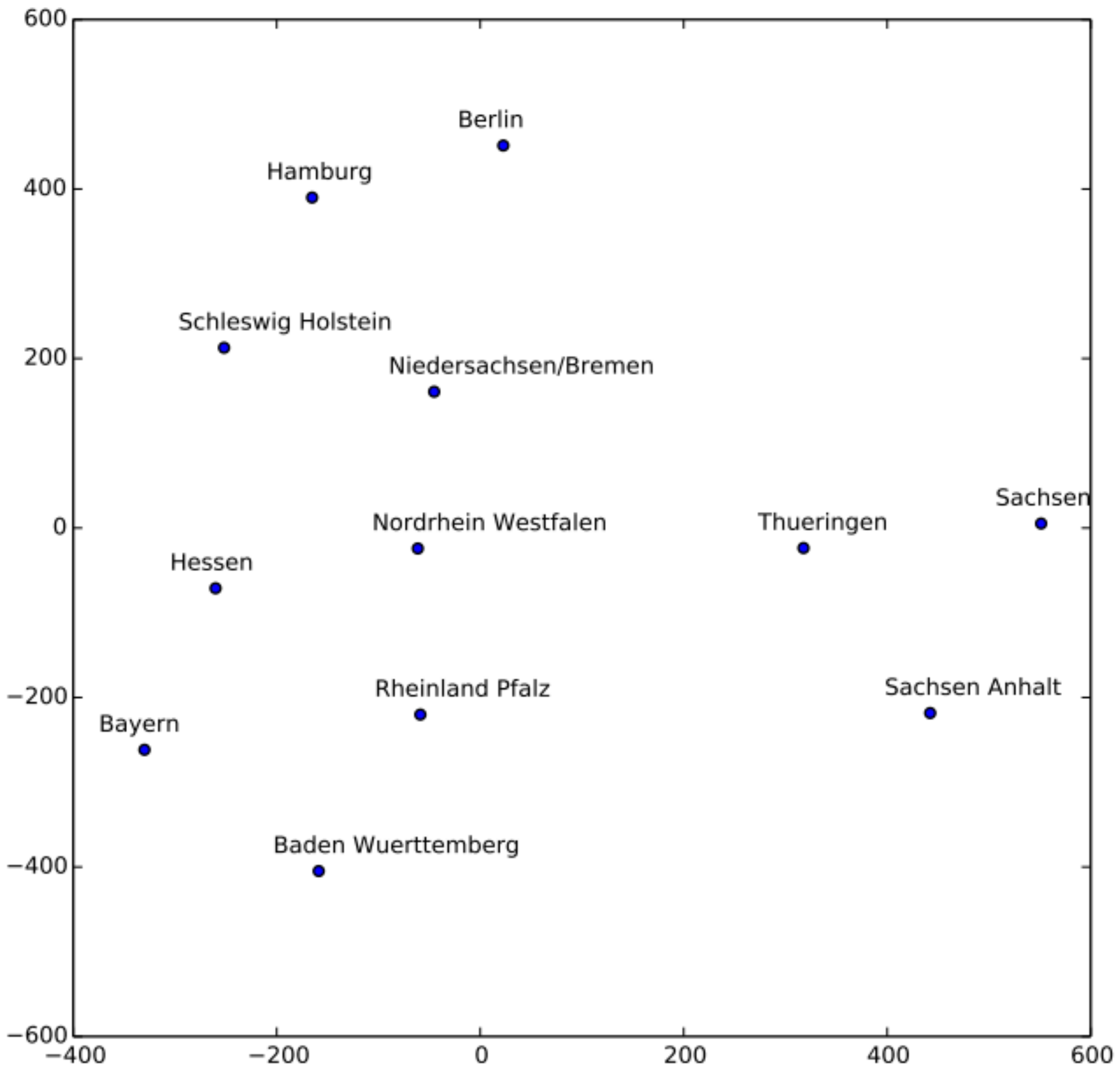


Image credits: https://github.com/ledell/sldm4-h2o/ https://smerity.com/articles/2016/google_nmt_arch.html

DEEP LEARNING

FEATURE ENGINEERING: ENTITY EMBEDDING FOR CATEGORICAL VARIABLES





The learned German state embedding mapped to a 2D space with t-SNE.

Source: Cheng Guo and Felix Berkhahn. 2016. Entity embeddings of categorical variables. arXiv preprint arXiv:1604.06737 (2016).

DEEP LEARNING

LIMITATIONS

Characteristic / Requirement	Score
Highly non-stationary	Yes
Multiple time series	Yes
Multi-horizon forecast	Yes
Model interpretability	Low
Model Capability	High
Computational Efficiency	Low

- Very flexible approach
- Automated feature learning is more limited due to the lack of unlabelled data.
- Some feature engineering is still necessary.
- Poor model interpretability

MODELS - SUMMARY

TIME SERIES MODELS

- Good model interpretability
- Limited model complexity to handle non-linear data
- Difficult to model multiple time series
- Difficult to integrate shared features across different time series

MACHINE LEARNING

- Flexible
- Can incorporate many features across the time series
- A lot of feature engineering required

DEEP LEARNING

- Very flexible
- Automated feature learning via embeddings
- Still some degree of feature engineering necessary
- Poor model interpretability

PRACTICE

EVALUATION AND METRICS

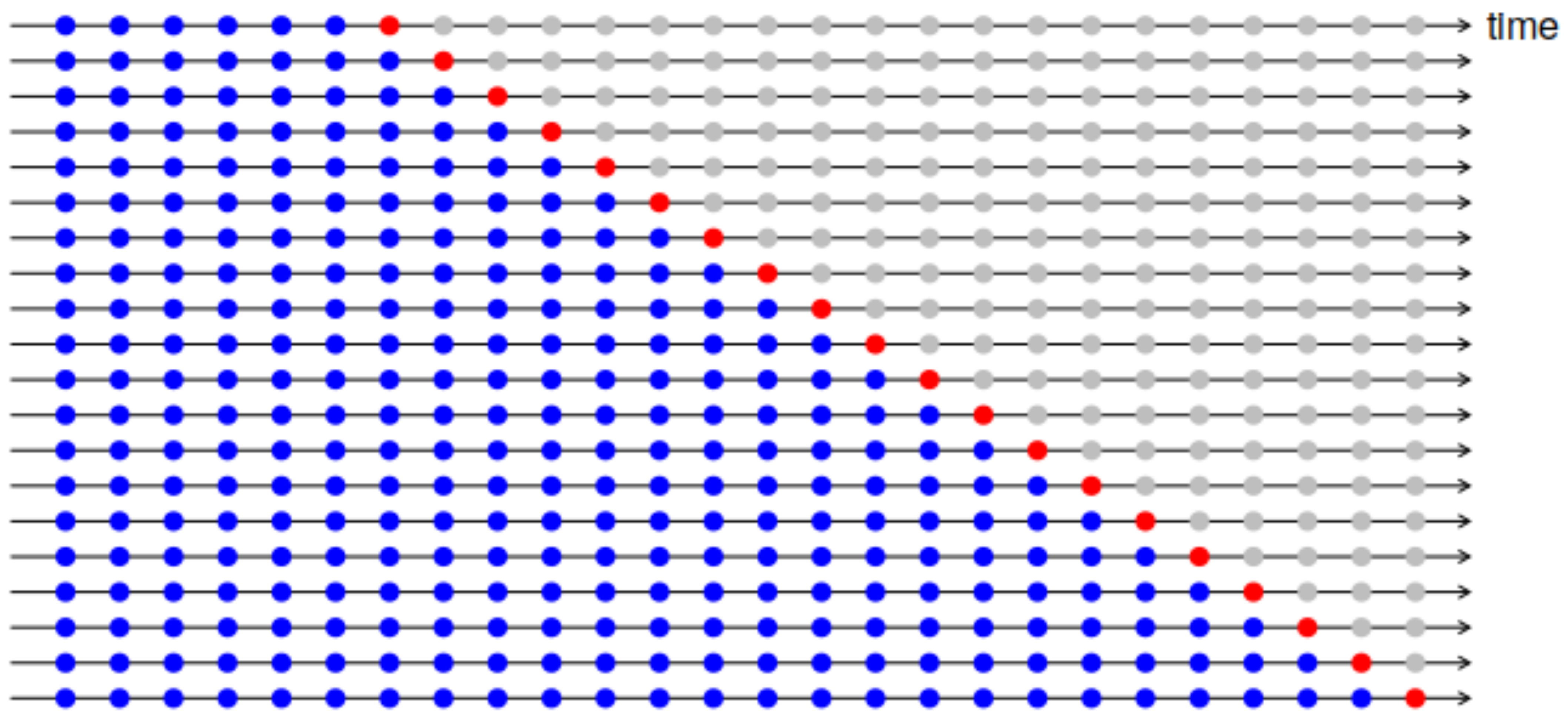
KPI vs LOSS FUNCTIONS

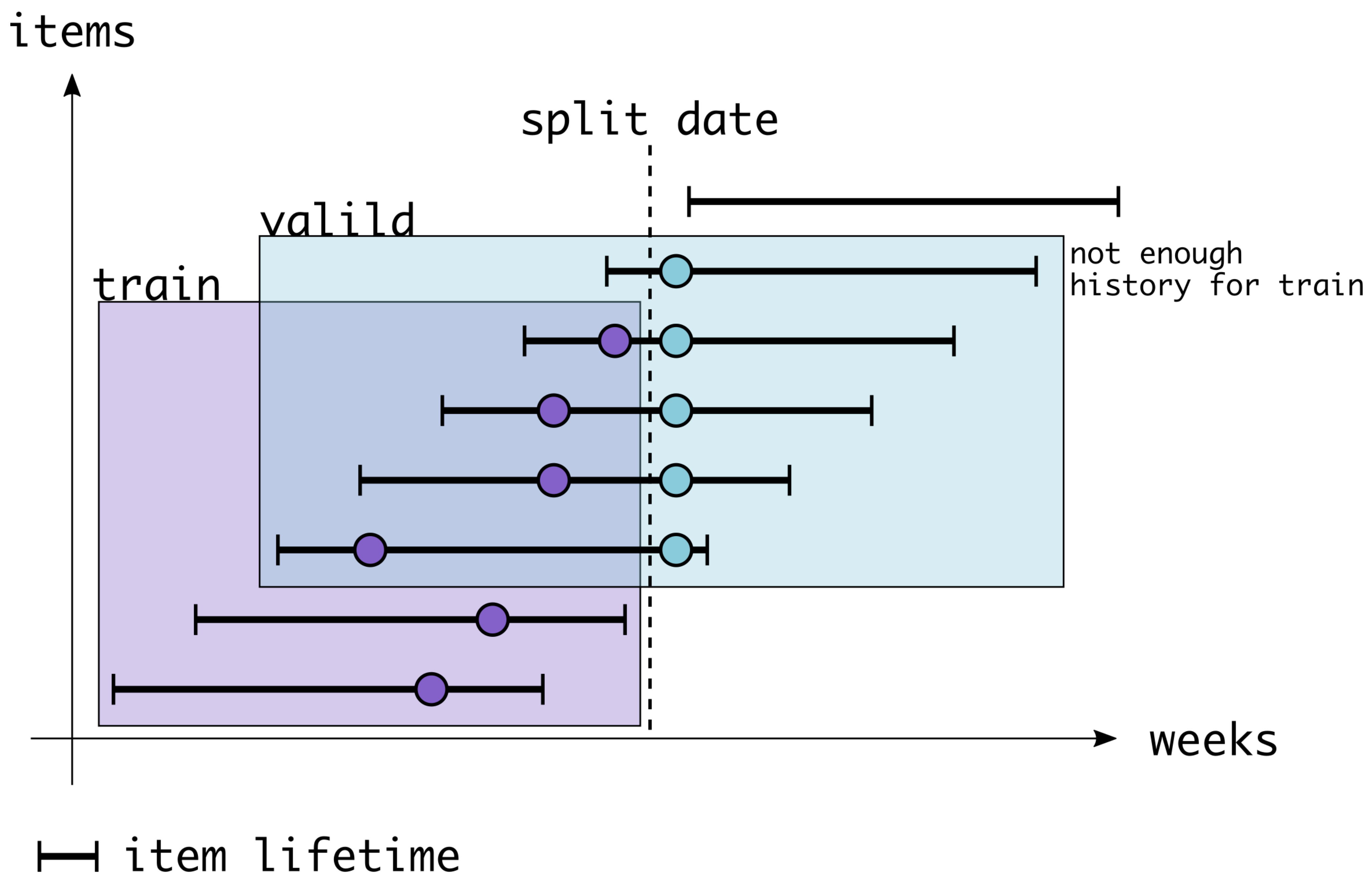
Metric	Formula	Notes
MAE (mean absolute error)	$\frac{1}{N}\sum_i \left\|\hat{y}_i - y_i\right\|$	Intuitive
MAPE (mean absolute percentage error)	$\frac{1}{N}\sum_i \left\|\frac{\hat{y}_i - y_i}{y_i}\right\|$	Independent of the scale of measurement
SMAPE (symmetric mean absolute percentage errror)	$\frac{1}{N}\sum_i \frac{2\left\|\hat{y}_i - y_i\right\|}{\left\|y_i\right\| + \left\|\hat{y}_i\right\|}$	Avoid Asymmetry of MAPE
MSE (Mean squared error)	$\frac{1}{N}\sum_i (\hat{y}_i - y_i)^2$	Penalize extreme errors
MSLE (Mean Squared Logarithmic loss)	$\frac{1}{N}\sum_i log(y_i + 1) - log(\hat{y}_i + 1)$	Large errors are not more significantly penalised than small ones
Quantile Loss	$\frac{1}{N}\sum_i q(\hat{y}_i - y_i)^+ + (1-q)(\hat{y}_i - y_i)^+$	Measure distribution
RMSPE (Root Mean Squared Percentage Error)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2}$	Independent of the scale of measurement

DSPT DAY 2019

CROSSVALIDATION

TIME SERIES BRINGS SOME RESTRICTIONS



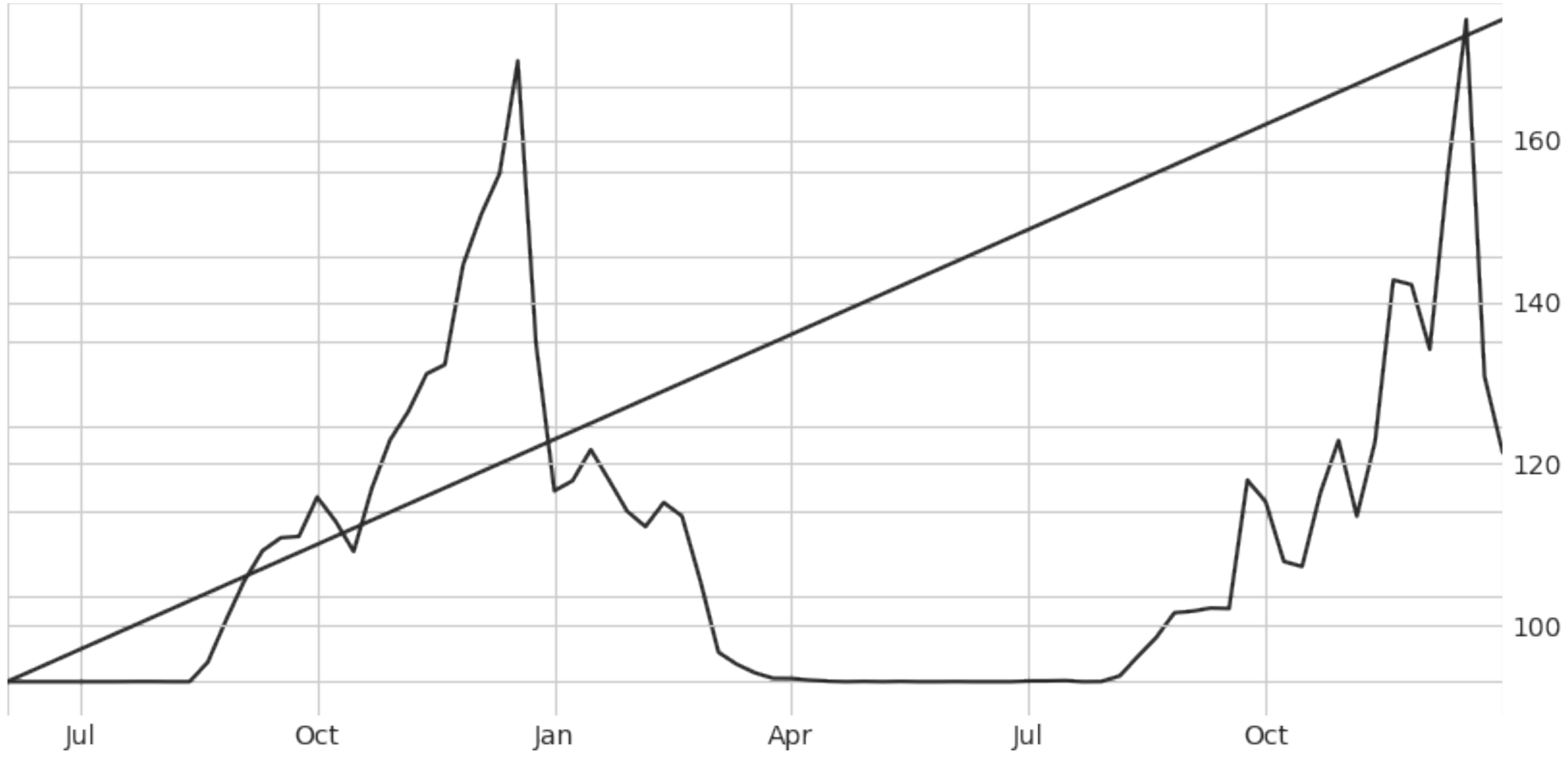


Left image source: Hyndman, R.J., & Athanasopoulos, G. (2019) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3.

USEFUL PREDICTORS

SOME TOOLS FROM TOOLBOX

- **Trend or Sequence**
- Seasonal Variables
- Intervention Variables



$$x_{1,t} = t$$

USEFUL PREDICTORS

SOME TOOLS FROM TOOLBOX

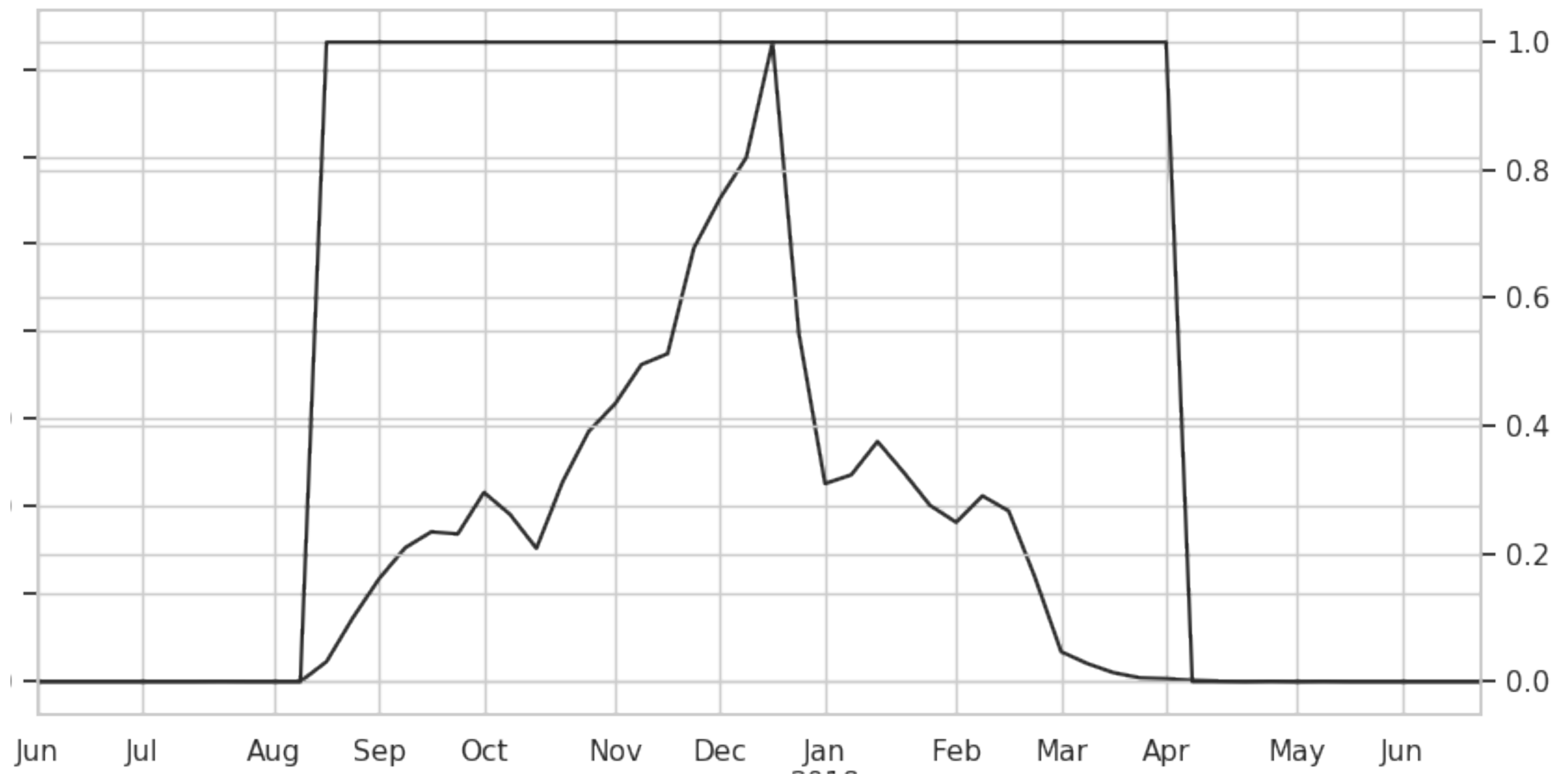
- Trend or Sequence
- **Seasonal Variables**
- Intervention Variables

	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
2018-12-31	0	1	0	0	0	0	0
2019-01-01	0	0	0	0	0	1	0
2019-01-02	0	0	0	0	0	0	1
2019-01-03	0	0	0	0	1	0	0
2019-01-04	1	0	0	0	0	0	0
2019-01-05	0	0	1	0	0	0	0
2019-01-06	0	0	0	1	0	0	0

USEFUL PREDICTORS

SOME TOOLS FROM TOOLBOX

- Trend or Sequence
- Seasonal Variables
- **Intervention Variables**



SUMMARY

- Demand Forecasting in fashion retail is challenging as the forecasting system need to deal with the certain specific characteristics: fashion trends, seasonality, influence of many external variables.
- Machine Learning, in particular Gradient Boosting seem to offer a good compromise between model capacity and interpretability.
- Feature Engineering is key, and it is still necessary when using Deep Learning.
- Avoid feature leaking by using a robust time series cross-validation approach.
- Try to match your metric to the business requirements. Business understandable metrics are necessary to explain the quality of the forecasts to the stakeholders.

REFERENCES

- [1] Choi, T. M., Hui, C. L., & Yu, Y. (2014). *Intelligent fashion forecasting systems: Models and applications*. *Intelligent Fashion Forecasting Systems: Models and Applications* (pp. 1–194). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39869-8
- [2] Hyndman, R.J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 01.10.2019
- [3] H&M, a Fashion Giant, Has a Problem: $4.3 Billion in Unsold Clothes. https://www.nytimes.com/2018/03/27/business/hm-clothes-stock-sales.html
- [4] Thomassey, S. (2014). Sales Forecasting in Apparel and Fashion Industry: A Review. In *Intelligent Fashion Forecasting Systems: Models and Applications* (pp. 9–27). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39869-8_2
- [5] Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis: Forecasting and control. San Francisco: Holden-Day
- [6] Autoregressive integrated moving average (ARIMA). https://en.wikipedia. org/wiki/Autoregressive_integrated_moving_average. Accessed: 2019-05-02
- [7] Cheng Guo and Felix Berkhahn. 2016. Entity embeddings of categorical variables. arXiv preprint arXiv:1604.06737 (2016).
- [8] Shen, Yuan, Wu and Pei - Data Science in Retail-as-a-Service Worshop. KDD 2019. London.

IMAGE CREDITS

- Photo Credit: https://www.flickr.com/photos/157635012@N07/48105526576 Artem Beliaikin on via Compfight CC 2.0
- Ship Freight by ProSymbols from the Noun Project
- warehouse by ProSymbols from the Noun Project
- Store by AomAm from the Noun Project
- Neural Network by Knut M. Synstad from the Noun Project
- Tunic Dress by Vectors Market from the Noun Project
- sales by Kantor Tegalsari from the Noun Project
- time series by tom from the Noun Project
- fashion by Smalllike from the Noun Project
- Time by Anna Sophie from the Noun Project
- linear regression by Becris from the Noun Project
- Random Forest by Becris from the Noun Project
- SVM by sachin modgekar from the Noun Project
- production by Orin zuu from the Noun Project
- Auto by Graphic Tigers from the Noun Project
- Factory by Graphic Tigers from the Noun Project
- Express Delivery by Vectors Market from the Noun Project
- Stand Out by BomSymbols from the Noun Project
- Photo Credit: https://www.flickr.com/photos/157635012@N07/47981346167/ by Artem Beliaikin on Flickr via Compfight CC 2.0
- Photo Credit: „https://www.flickr.com/photos/157635012@N07/48014587002/ Artem Beliaikin Flickr via Compfight CC 2.0
- regression analysis by Vectors Market from the Noun Project
- Research Experiment by Vectors Market from the Noun Project
- weather by Alice Design from the Noun Project
- Shirt by Ben Davis from the Noun Project
- fashion by Eat Bread Studio from the Noun Project
- renew by david from the Noun Project
- price by Adrien Coquet from the Noun Project
- requirements by ProSymbols from the Noun Project
- marketing by Gregor Cresnar from the Noun Project
- macroeconomic by priyanka from the Noun Project
- competition by Gregor Cresnar from the Noun Project

THANK YOU

QUESTIONS?

DSPT DAY 2019