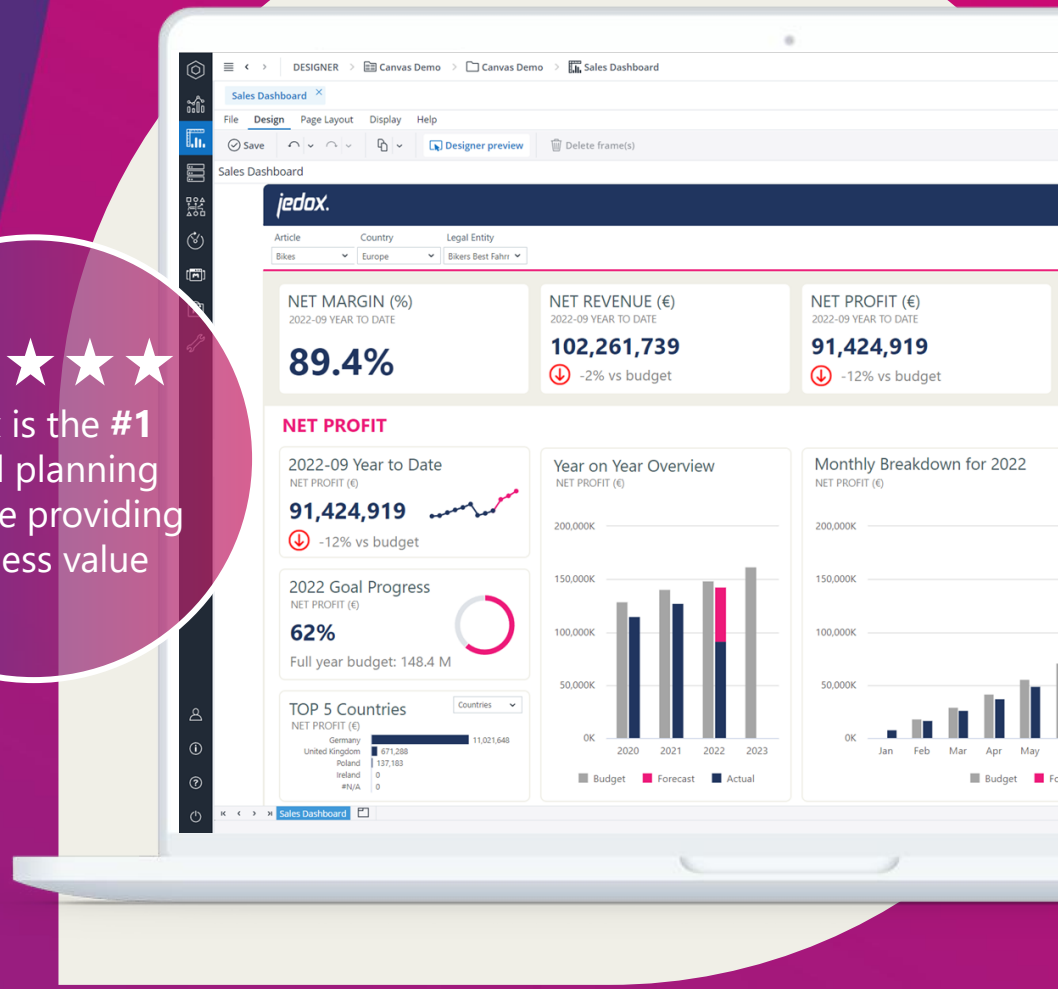
Jedox AIssisted™ Planning

ALGORITHM CHEAT SHEET
FOR AISSISTED™ PLANNING SERVICES

★★★★★
Jedox is the **#1** global planning software providing business value



Time Series Prediction

CLASSIC

Linear Model

Data Needs: Trend & Seasonality

Speed & Quality: Fast speed; High quality (Highly data dependent)

- Uses trend and seasonality to create a linear model for prediction.

Holt Winters

Data Needs: Periodicity & Seasonality required, at least 2 periods of data

Speed & Quality: Fast speed; High quality (Highly data dependent)

- Uses trend, season and level to predict time series forecast.
- Recommended for larger prediction ranges (> 1 year).

Seasonal Naïve

Data Needs: Seasonality required

Speed & Quality: Fastest speed; Intermediate quality (Heavily data dependent)

- Basic ARIMA modelling assuming seasonality.

Time Series Prediction

CLASSIC

Exponential Smoothing

Data Needs: Periodicity, at least 2 periods of data

Speed & Quality: Intermediate speed; High quality (Highly data dependency)

- Uses trend and season to predict time series forecast.
- Recommended for larger prediction ranges (> 1 year).

ARMA

Data Needs: No specific data characteristics needed

Speed & Quality: Intermediate speed; Highest quality (Lowest data dependency)

- Uses trend, periodicity, seasonality, cyclic nature of time to create optimized ARIMA model series.
- Recommended for larger prediction ranges (> 1 year).

Random Walk with Drift

Data Needs: No specific data characteristics needed

Speed & Result Quality: Fastest speed; Intermediate quality (Heavily data dependent)

- Uses automated drift (error) calculation

Time Series Prediction

INNOVATIVE

STL Model	**Data Needs:** Seasonality required **Speed & Quality:** Intermediate speed; High quality (Highly data dependent)	▪ Decompose non-linear relationships in season and trend, bias adjustment (Box Cox transformation). ▪ Recommended for larger prediction ranges (> 1 year).
STLF	**Data Needs:** Seasonality required **Speed & Quality:** Intermediate speed; High quality (Highly data dependent)	▪ Decompose non-linear relationships in season (automatically optimized) and trend , bias adjustment (Box Cox transformation). ▪ Recommended for larger prediction ranges (> 1 year).
Neural Network (TAR)	**Data Needs:** No specific data characteristics needed **Speed & Quality:** Slow speed; Highest quality (Lowest data dependency)	▪ Univariate Neural Network model (layers, weights), bias adjustment.
TBATS	**Data Needs:** Seasonality required **Speed & Quality:** Slow speed; High quality (Highly data dependent)	▪ Uses multiple, shifted seasonalities, bias adjustment (Box Cox transformation). ▪ Recommended for larger prediction ranges (> 1 year).

Time Series Prediction

INNOVATIVE

Croston's Method	**Data Needs:** Periodicity, at least 2 time series, for discrete time series (i.e. no decimals), no negative values	▪ Combines exponential smoothing estimates with average demand intervals to create a model for prediction
	Speed & Quality: Fast speed; Intermediate quality (Highly data dependent)	
LGBM	**Data Needs:** No specific data characteristics needed.	▪ Good at covering complex patterns in data.
	Speed & Quality: Extremely fast with high quality forecasts when data is clean. Risk of overfitting when data is unprepared.	
SARIMAX	**Data Needs:** No specific data characteristics needed.	▪ Uses the same model type as ARIMA for forecasting but with more customization options, e.g. seasonality.
	Speed & Quality: Intermediate speed and high quality.	

Driver Based Forecasting

DRIVER BASED

Linear Regression	**Data Notes:** Suitable for simpler problems with fewer features **Speed & Quality:** Fastest speed; best results with small number of features	▪ The regression model allows for a linear relationship between the forecast of dependent variable y and one or more predictor variable(s).
Support Vector Machine	**Data Notes:** Works well with large number of features, requires extensive hyper-parameter search for best results **Speed & Quality:** Slower to get highest quality results with hyper-parameter search	▪ Can model more complex non-linear relationships ▪ Works well with a large number of features ▪ Training time can increase heavily with bigger database
Extremely Random Trees	**Data Notes:** Handles cases with large number of features but smaller number of samples **Speed & Quality:** Intermediate speed, good results possible out of the box, start here	▪ Similar to Random Forest but with more randomization ▪ Faster to compute than Random Forest ▪ Performs better than Random Forest with more features
Random Forest	**Data Notes:** Handles cases with large number of features but smaller number of samples **Speed & Quality:** Slow speed, good results possible out of the box, start here	▪ Each estimator trained on random sample from data

Driver Based Forecasting

DRIVER BASED

Gradient Boost

Data Notes: Sensitive to data with outliers

Speed & Quality: Fast speed; Intermediate quality (overfitting can be a problem)

- Continuously improve estimator by learning from errors of previous one.
- Predict based on weighted average.

K Nearest Neighbors

Data Notes: Handles more complex non-linear relationships and large amount of features

Speed & Result Quality: Speed increases heavily with larger datasets, good results possible with little tuning

- Uses proximity to make predictions about the grouping of an individual data point.

Polynomial Regression

Data Notes: Not suitable for high number of features

Speed & Quality: Runs quickly, good results possible with smaller number of features

- Similar to linear regression but can handle non-linear relationships between target feature and input feature.
- Able to capture more complex relationships than linear regression.

Classification Wizard

Logistic Regression

Data Notes: Suitable for simpler problems with fewer features

Speed & Quality: Fastest speed; best results with small number of features

- Good to start here if a linear relationship between the target and input features is expected and if number of features is small relative to number of samples.

Support Vector Machines

Data Notes: Works well with large number of features, requires extensive hyper-parameter search for best results

Speed & Quality: Slower to get highest quality results with hyper-parameter search

- Can model more complex non-linear relationships. Training time can increase heavily with bigger database.

Extremely Random Trees

Data Notes: Handles cases with large number of features but smaller number of samples

Speed & Result Quality: Intermediate speed, good results possible out of the box, start here

- Similar to Random Forest but with more randomization.
- Faster to compute than Random Forest.
- Performs better than Random Forest with more features.

Classification Wizard

Random Forest

Data Notes: Handles cases with large number of features but smaller number of samples

Speed & Quality: Slow speed, good results possible out of the box, start here

- Each estimator trained on random sample from data.

Gradient Boost

Data Notes: Sensitive to data with outliers

Speed & Quality: Fast speed; Intermediate quality (overfitting can be a problem)

- Continuously improve estimator by learning from errors of previous one. Predict based on weighted average.

K Nearest Neighbors

Data Notes: Handles more complex non-linear relationships and large amount of features

Speed & Result Quality: Speed increases heavily with larger datasets, good results possible with little tuning

- Uses proximity to make classifications about the grouping of an individual data point.

Classification Wizard

GaussianNB

Data Notes: Suitable for all data sets, including smaller sample sizes, usually used for text classification

Speed & Quality: Speed increases heavily with larger datasets, good results possible with little tuning

- Estimate the probability of a feature belong to a certain target class using Bayes' theorem.
- All naive bayes algorithms train fast and can work really well even with a small sample size (not many rows).

MultinomialNB

Data Notes: Suitable for all data sets, including smaller sample sizes, usually used for text classification

Speed & Quality: Speed increases heavily with larger datasets, good results possible with little tuning

- Estimate the probability of a feature belong to a certain target class using Bayes' theorem.
- All naive bayes algorithms train fast and can work really well even with a small sample size (not many rows).

ComplementNB

Data Notes: Suitable for all data sets, including smaller sample sizes, usually used for text classification

Speed & Result Quality: Speed increases heavily with larger datasets, good results possible with little tuning.

- Estimate the probability of a feature belong to a certain target class using Bayes' theorem.
- All naive bayes algorithms train fast and can work really well even with a small sample size (not many rows).

Data Preparation: Outlier Detection

SIMPLE | MEDIUM | ADVANCED (SEASONAL)

Local Outlier Factor

- **Data Notes:** Works well with most types of time series as it makes no strong assumptions about input time series
- **Speed & Quality:** Runs quickly with good quality (good starting point)

Isolation Forest

- **Data Notes:** Works well with numerical time series; it makes no assumptions about seasonality
- **Speed & Quality:** Quality is generally good, though it may underperform on very short or highly seasonal time series

Interquartile Range Based

- **Data Notes:** Best for stable time series without strong trends or seasonality
- **Speed & Quality:** Extremely fast and high quality when data is clean and consistent but may miss context-aware anomalies

Z-Score Statistic Based

- **Data Notes:** Best for stationary time series and normally distributed data. May require normalization, is sensitive to extreme values
- **Speed & Quality:** Very fast good quality for clean, symmetric data but may struggle with skewed or seasonal time series

Change Point

- **Data Notes:** Works well with time series that contain abrupt changes or transitions. Assumes changes are meaningful and not just noise
- **Speed & Quality:** Moderate speed and high quality for detecting structural shifts but not designed for spotting isolated anomalies.

Data Preparation: Interpolation

SIMPLE | MEDIUM |ADVANCED (SEASONAL)

Gaussian Regression

Data Notes: Very flexible and powerful. Handles nonlinear patterns well but requires sufficient data.

Speed & Quality: Slower but delivers excellent quality. Ideal for high-stakes or precision-critical applications

Linear Regression

Data Notes: Works well with linear trends. Less effective for nonlinear or noisy data

Speed & Quality: Slower but offers better quality for trend-driven data

Linear Approx.

Data Notes: Best for gradually changing data without complex patterns or sudden shifts. No assumptions about seasonality or stationarity.

Speed & Quality: Very fast. Good for smooth data or as a baseline but may struggle with highly variable or nonlinear series.

Linear Approx + Seasonality

Data Notes: Requires sufficient data to detect seasonality. Works well with periodic or seasonal time series

Speed & Quality: Moderately fast. Produces high-quality results for seasonal data, but less effective for non-seasonal or irregular

Spline Interpolation

Data Notes: Captures complex patterns better than linear methods. Suited for gradual changes or curve, but less robust to noise.

Speed & Quality: Slightly slower than linear interpolation but more accurate for smooth, nonlinear data. Balances flexibility and stability

Polynomial Interpolation

Data Notes: Can be sensitive to noise and may overfit with too many data points. Not ideal for long or uneven time series

Speed & Quality: Performs well on small, clean data but quality drops quickly otherwise. Slow and unstable on large datasets

LOCF

Data Notes: Best for stable or slowly changing time series. Can misrepresent data if used during rapid changes or long gaps.

Speed & Quality: Extremely fast and simple. Quality depends on data stability; it can be misleading in volatile series

Nearest Point Replacement

Data Notes: Best for discrete or slowly changing time series. Can introduce bias or abrupt jumps in highly variable or continuous data

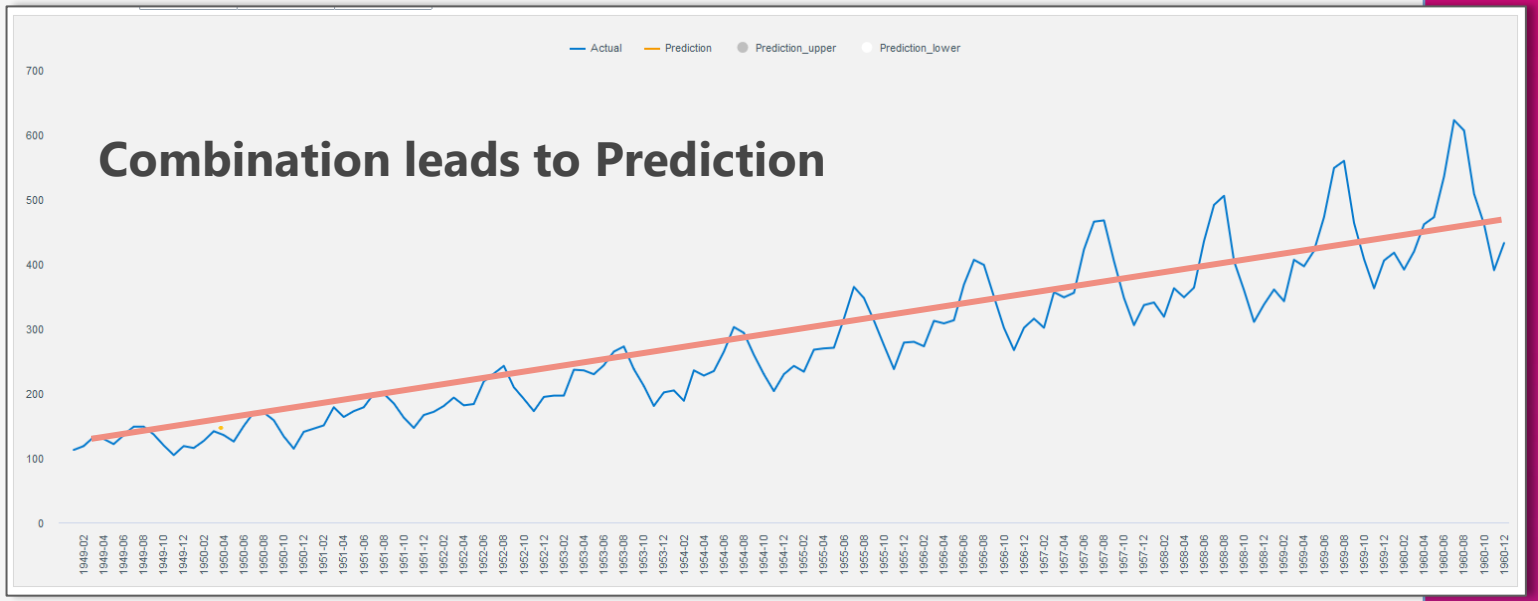
Speed & Quality: Extremely fast and easy to implement. Quality is low for dynamic data but acceptable for stable or categorical series

Jedox AIssisted™ Planning – THE AI

What the Algorithms Do

Trend

Decrease, Increase, Static…



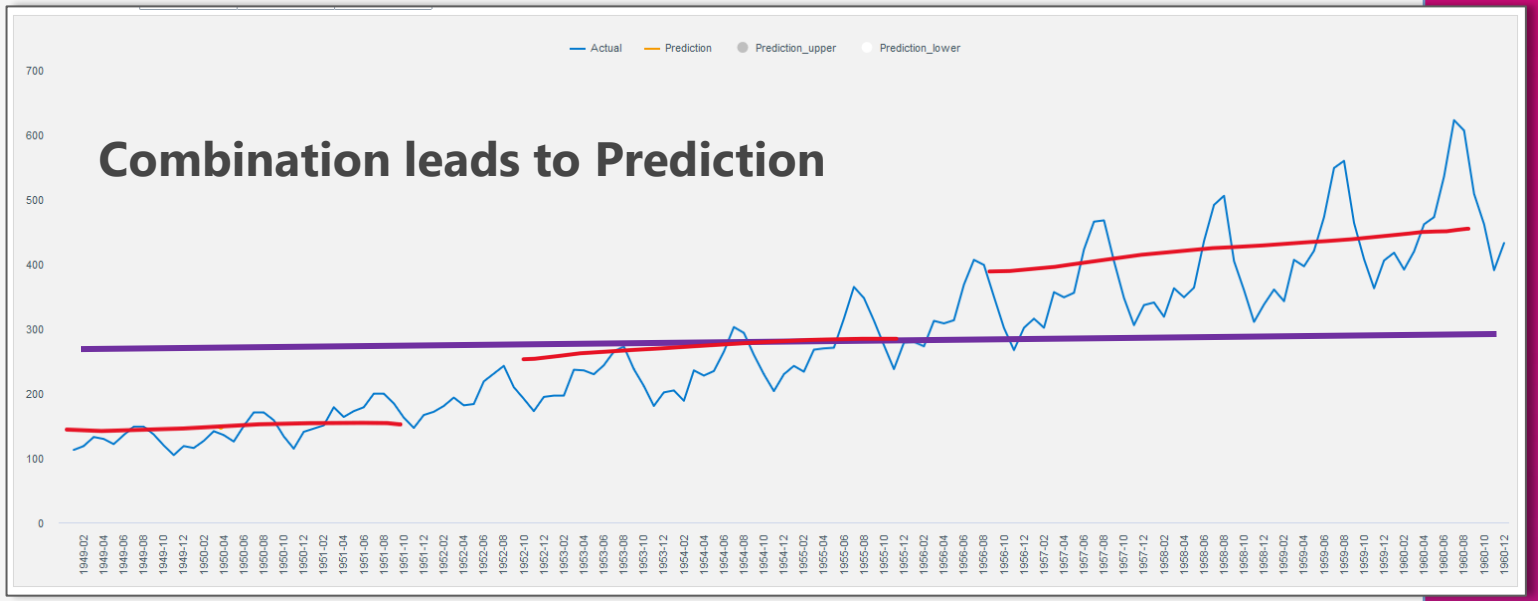
Always Look at Data Before the Project

Saves time, gives you better idea of what level to predict

Jedox AIssisted™ Planning – THE AI

What the Algorithms Do

Level
Average of Value

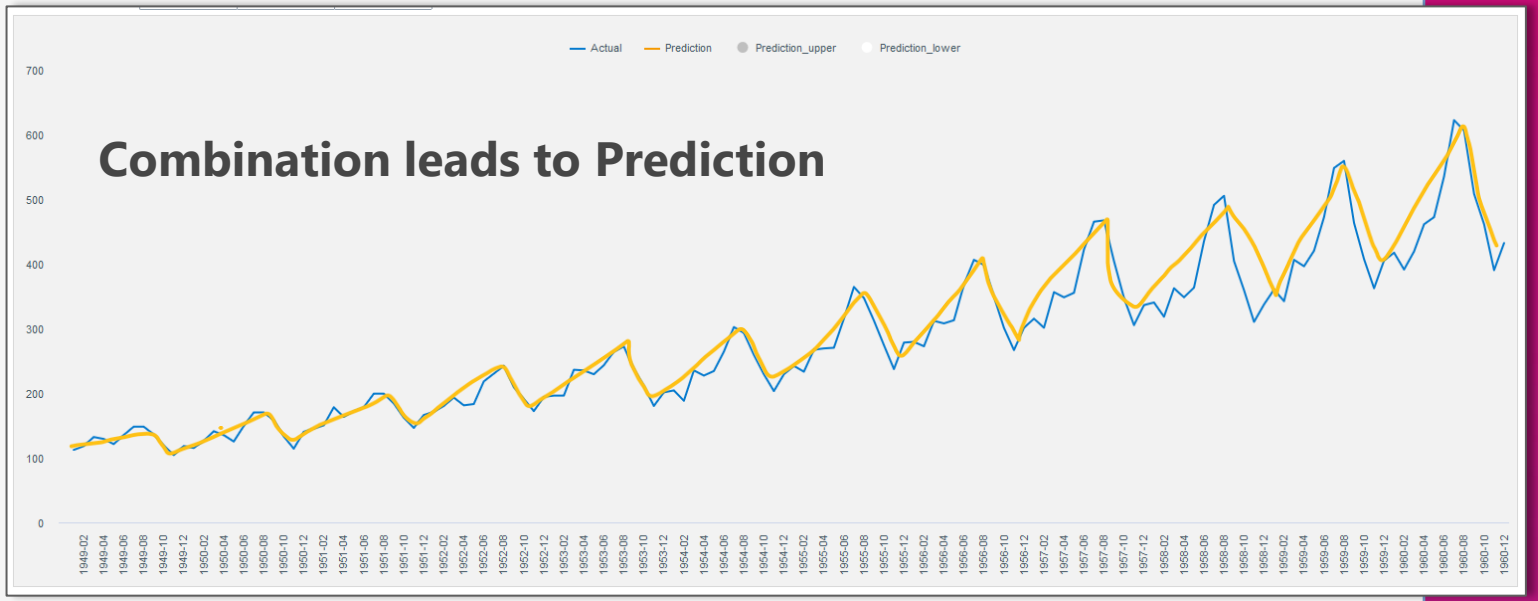


Always Look at Data Before the Project
Saves time, gives you better idea of what level to predict

Jedox AIssisted™ Planning – THE AI

What the Algorithms Do

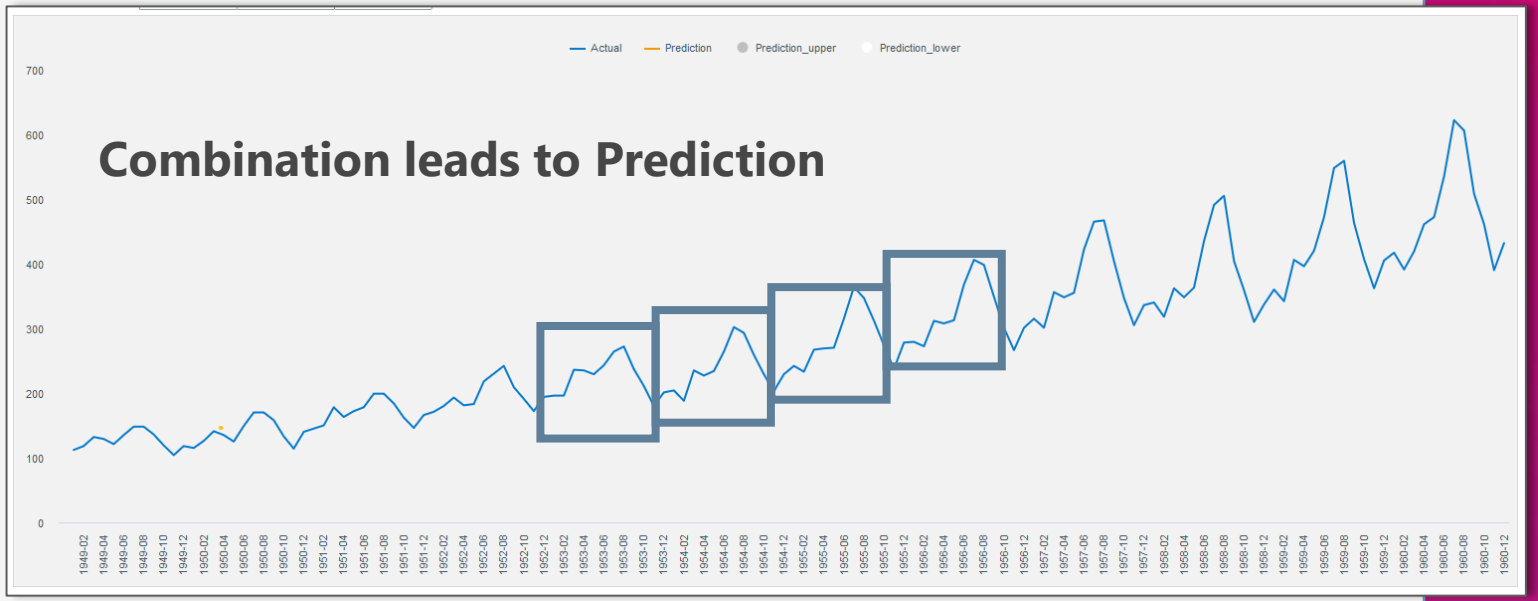
Seasonality
Peaks, valley, etc.



Always Look at Data Before the Project
Saves time, gives you better idea of what level to predict

Jedox AIssisted™ Planning – THE AI

What the Algorithms Do



Period
e.g. 1 year

Always Look at Data Before the Project
Saves time, gives you better idea of what level to predict

Jedox AIssisted™ Planning – THE AI

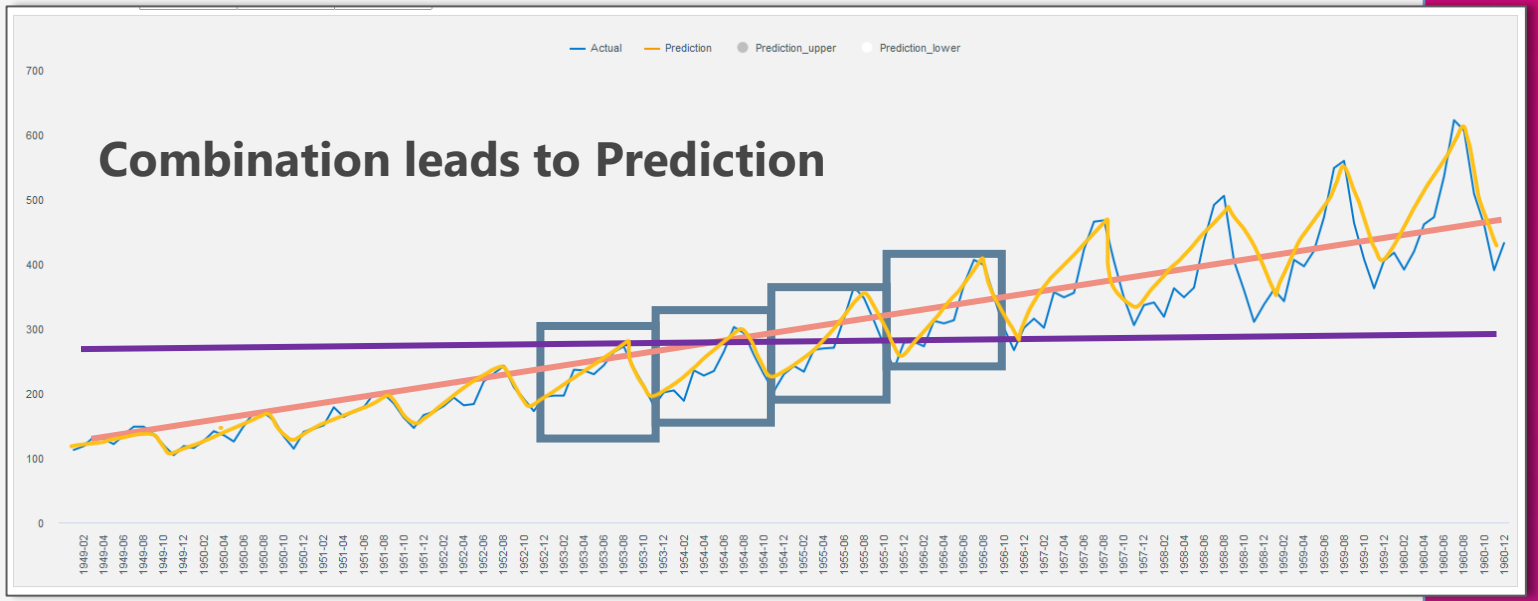
Algorithms take all of this into account to calculate Prediction

Trend
Decrease, Increase, Static…

Level
Average of Value

Seasonality
Peaks, valley, etc.

Period
e.g. 1 year



Always Look at Data Before the Project
Saves time, gives you better idea of what level to predict

jedox.

Thank You