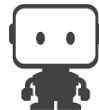


Automated Time Series modelling with DataRobot

Jarlath Quinn

www.sv-europe.com

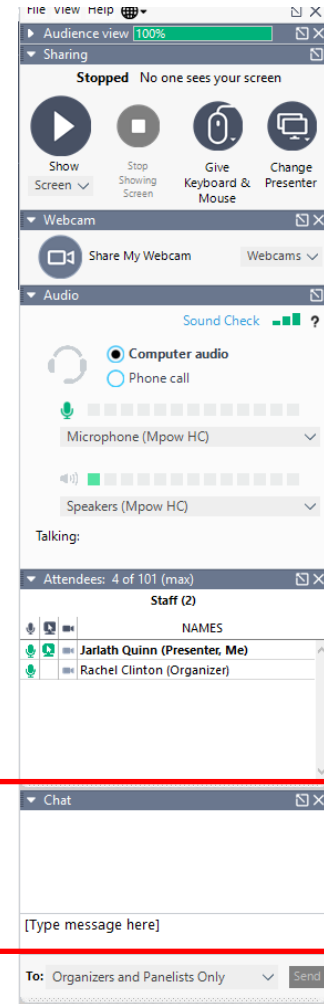


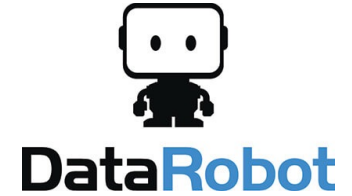
DataRobot

A SELECT INTERNATIONAL COMPANY

FAQ's

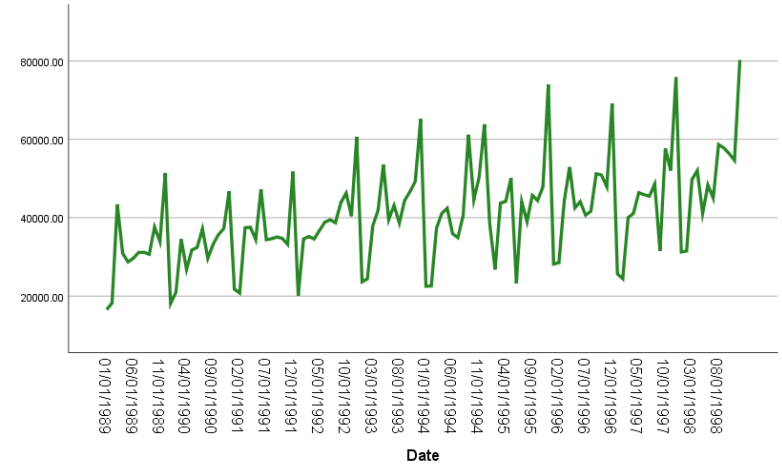
- Is this session being recorded? Yes
- Can I get a copy of the slides? Yes, we'll email a PDF copy to you after the session has ended.
- Can we arrange a re-run for colleagues? Yes, just ask us.
- How can I ask questions? All lines are muted so please use the chat facility – if we run out of time we will follow up with you.





- Premier accredited partner to IBM, Predictive Solutions and DataRobot specialising in advanced analytics & big data technologies
- Work with open source technologies (R, Python, Spark etc.)
- Team each has 15 to 30 years of experience working in the advanced and predictive analytics industry
- Deep experience of applied advanced analytics applications across sectors
 - Retail
 - Gaming
 - Utilities
 - Insurance
 - Telecommunications
 - Media
 - FMCG





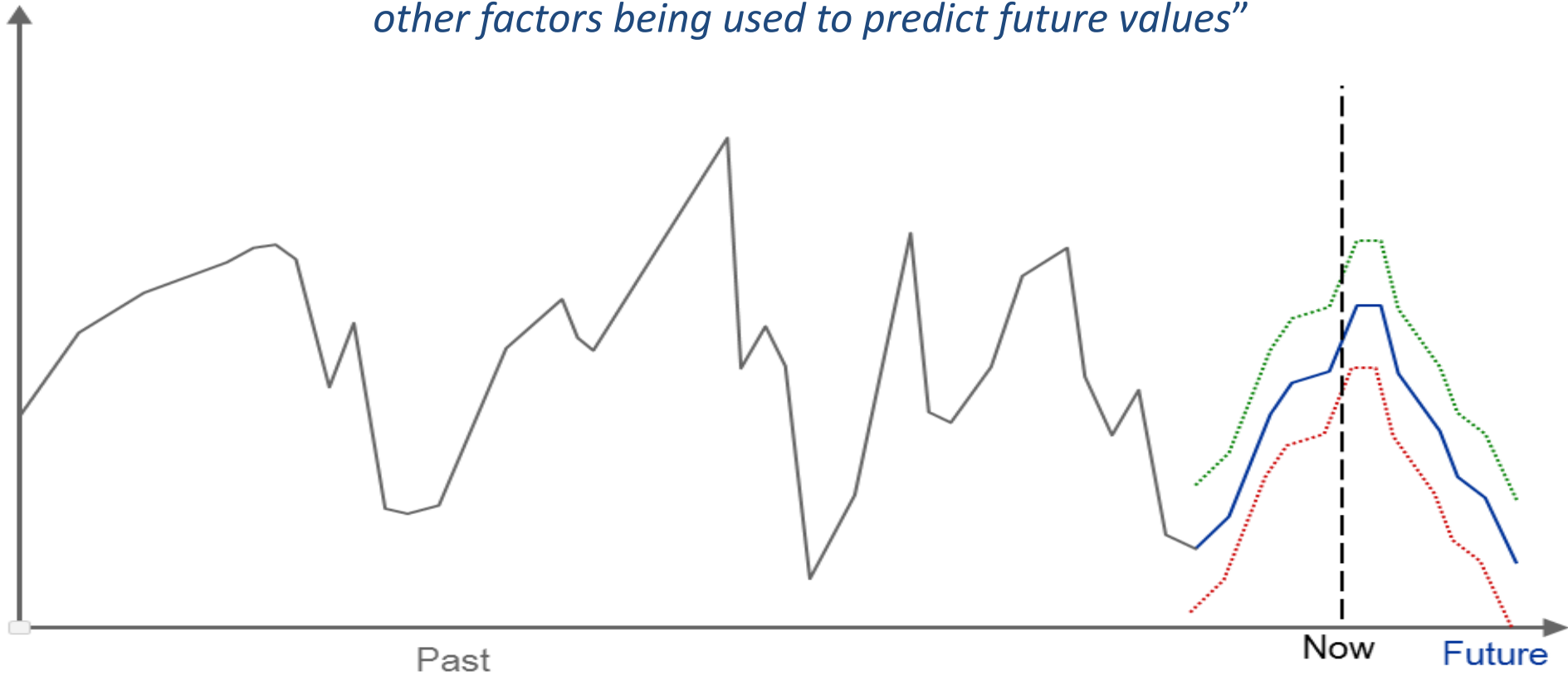
The Principles of Time Series Forecasting

What is Time Series?

- A 'Time Series' is simply a series of values of a quantity collected over a specific time period, often with equal intervals between them
- Examples of time series include:
 - Airline passenger numbers for a particular country over the last 40 years
 - Daily website hits during a three-month period
 - Hourly traffic volumes over the course of a week

Time Series Forecasting

“ ‘Things’ that are observed repeatedly over time, with past values and other factors being used to predict future values”

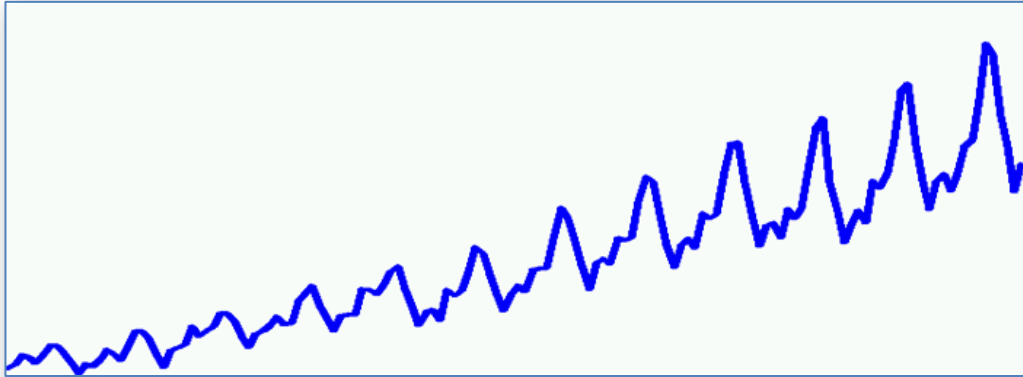


What is Time Series?

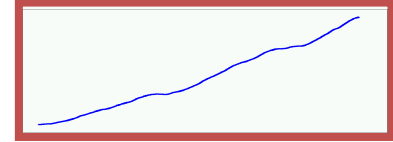
- Time series analysis is based on the principle that the past provides a model for the future
- Time series forecasting models often don't require predictor/independent variables
- The goal of time series analysis is to separate the random variability ('noise') from the variability that can be explained
- A single time series may have several elements that enable effective forecasting

What's *in* a 'Time Series'?

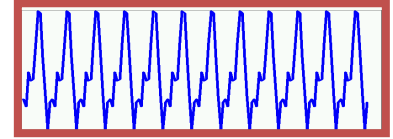
Time Series



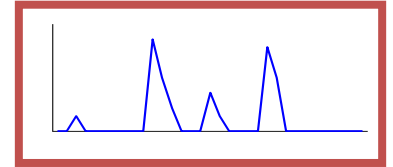
Trend



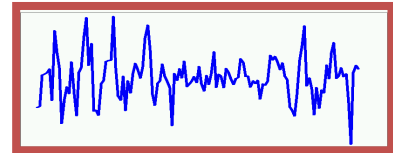
Seasonality

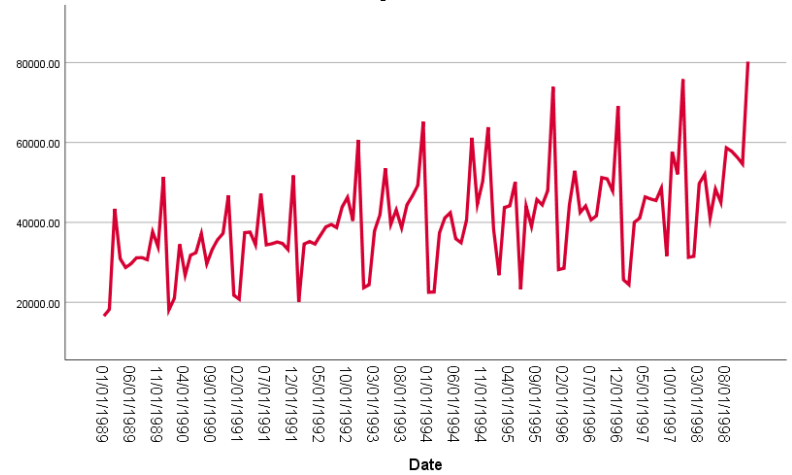


Events



Noise



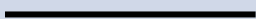

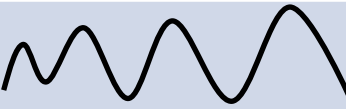


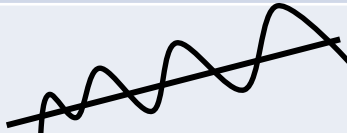








Types of Time Series

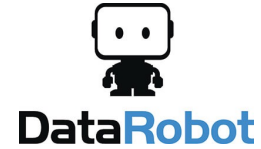
Two classic approaches to Time Series forecasting

- **ETS** (Error, Trend and Seasonality) / Exponential Smoothing models
 - Often used when there are strong seasonal trends (e.g. you see the same patterns from one year to the next)
- **ARIMA** (Autoregression, Integrated, Moving Average)
 - Often used when the only most recent values are needed to forecast what happens next
 - Can incorporate independent ‘predictor’ variables

Example: Basic Exponential Smoothing Methods

| | Non Seasonal | Additive Seasonal | Multiplicative Seasonal |
|-------------------|---|---|---|
| Constant Level |  |  |  |
| Linear Trend |  |  |  |
| Damped Trend |  |  |  |
| Exponential Trend |  |  |  |

DataRobot also includes a number of advanced machine learning methods



Machine learning and linear models

- RNN, XGBoost, elastic-net, etc.
- Linear trends, Ridge regression, Fourier models

Open Source models

- Facebook Prophet
- Spark
- Eureqa

The screenshot displays a list of machine learning models in a dark-themed interface. Each model entry includes a name, a brief description of its features, a performance score (M and BP values), and a baseline configuration. The models are:

- Error-Trend-Seasonal exponential smoothing model**: Error-Trend-Seasonal exponential smoothing model. Baseline Only (latest). 1 year • 4 months • 6 days. Performance: M8 BP69.
- Prophet**: Prophet. Date Only. 1 year • 4 months • 6 days. Performance: M10 BP61.
- Ridge Regressor with Forecast Distance Modeling**: One-Hot Encoding | Numeric Data Cleansing | Standardize | Naive Predictions as Feature | Ridge Regressor with Forecast Distance Modeling. No Differencing. 1 year • 4 months • 6 days. Performance: M11 BP70.
- eXtreme Gradient Boosted Trees Regressor with Early Stopping (learning rate =0.3)**: Ordinal encoding of categorical variables | Missing Values Imputed | Extract Forecast Distance Feature | Naive Predictions as Feature and Offset | eXtreme Gradient Boosted Trees Regressor with Early Stopping (learning rate =0.3) | Text fit on Residuals (L2 / Least-Squares Loss). With Differencing (average baseline). 1 year • 4 months • 6 days. Performance: M16 BP74 MONO.

DataRobot automates most steps that are traditionally manually configured

- Uploads, reads and explores the data
- Identifies extremes, missing values and data issues
- Detects the time series periodicity (minutes, days, weeks, months, years etc)
- Detects multiple series and hierarchies
- Creates a portfolio of potential predictor fields in the data
- Identifies how best to test the model using back-steps
- Carries out a range of data preparation steps such as missing value imputations
- Creates a series of Blueprints showing the data preparation steps and configuration of each model
- Generates a leader board of the best performing models
- Creates a series of evaluation plots to test and understand model performance

DataRobot uses back tests to test and rank the models

Configure Backtesting

Number of Backtests

Validation Length

Gap Length

3 + Holdout

Y:0 M:0 D:92 H:0 m:0 S:0

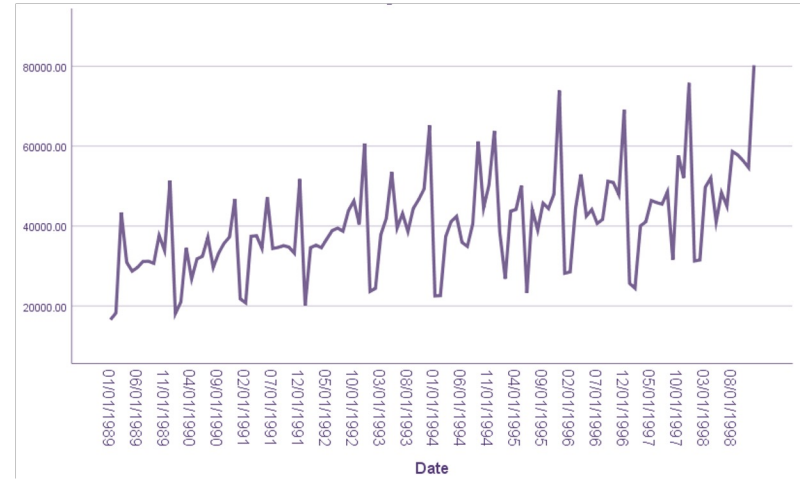
Y:0 M:0 D:0 H:22 m:0 S:0

These are the default settings for all backtests. However, modifying individual backtests will override these settings.



Each overlapping back test has a validation time zone to test a given model's accuracy

Equal rows per backtest
Sets the backtest configuration to use the same number of rows instead of the same length of time (duration).



Simple Time Series

Create your own time series

- <https://trends.google.com>

● BBQ + Barbecue + Barbeque
Search term




+ Compare

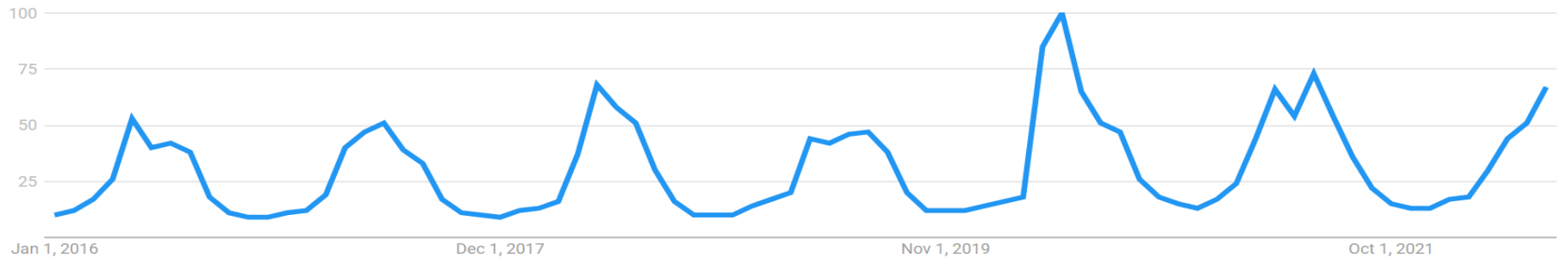
United Kingdom ▾

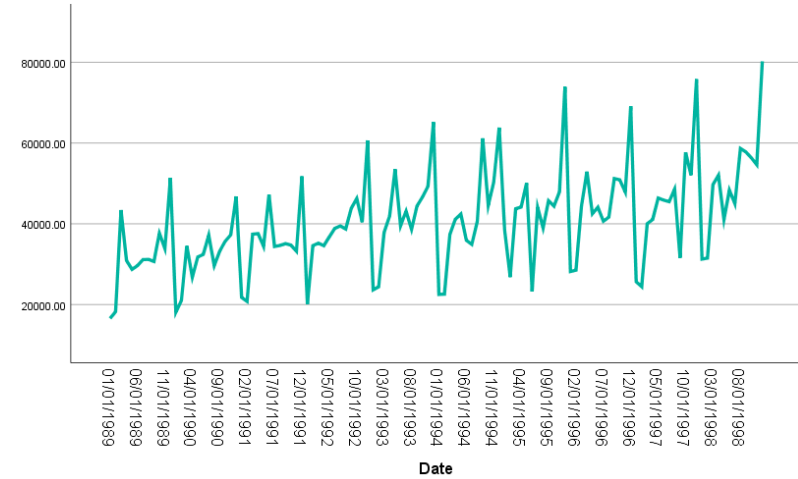
1/1/16 - 6/6/22 ▾

All categories ▾

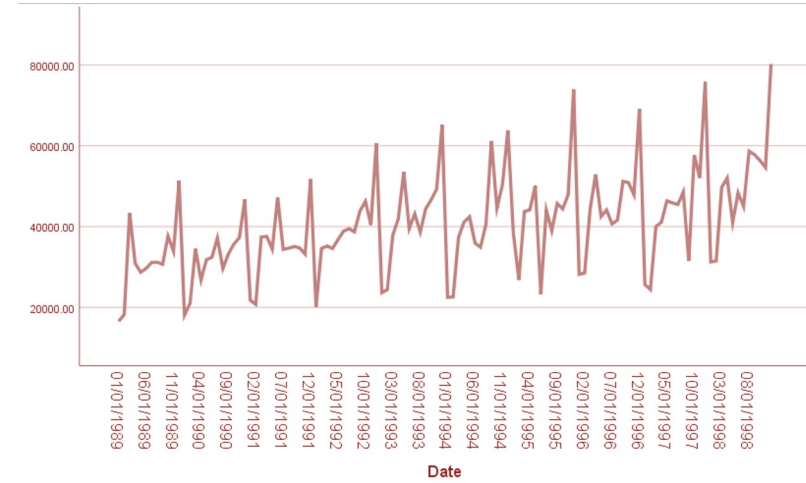
Web Search ▾

Interest over time 





Time Series with Predictors



Multiple Time Series

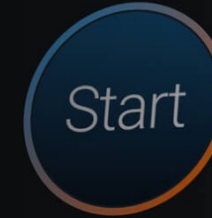
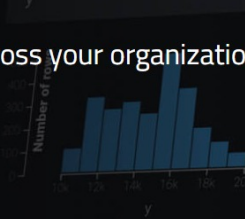
Automated Time Series

Effortlessly create, deploy, and maintain high-impact forecasts across your organization.

DataRobot Data Models Deployments Insights Jupyter Repository AI Catalog

What would you like to predict?

y REGRESSION



Modeling Mode: Autopilot

Feature list: Informative Features

Optimization Metric: RMSE

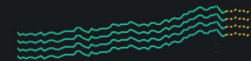
← Turn off time-aware modeling
What is the primary date/time fea

msi_date MILLISECONDS



- Fully Automated
- AI-Driven innovative models
- Granular forecasting at scale
- Real-time Anomaly detection
- Easy Integration
- On-Cloud and On-Prem

Time-Aware Modeling



The selected date/time feature "Date" has multiple rows with the same timestamp. If this is due to the data including multiple series or groups, select a series identifier.

What is the series identifier?

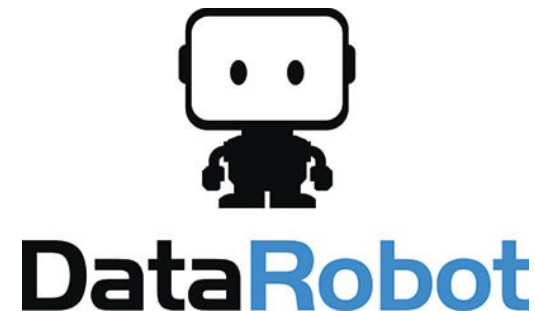
Store_Size

10 SERIES

2 potential identifiers: Store_Size, Store

← Go back

Set series ID



Please contact Rachel Clinton

rclinton@sv-europe.com

To find out more about how Smart Vision and

DataRobot can help accelerate your

organisation's efforts in Cloud AI



A SELECT INTERNATIONAL COMPANY



Contact us:

+44 (0)207 786 3568

info@sv-europe.com

Twitter: @sveurope



[Follow us on Linked In](#)



[Sign up for our Newsletter](#)

Thank you