Predicting Sales Using Neural Networks

Martin Burger
R PROGRAMMING TUTOR

www.r-tutorials.com

Neural Networks

Popular prediction method in time series analysis

Implementation with the library ‘forecast’

Data: Restaurant sales

Data cleaning

Building neural networks with ‘nnetar’

Interactive data visualization with ‘dygraphs’

The Beach Restaurant Dataset

Sales of a restaurant on the beach

Monthly revenue in USD

- January 1998 – December 2017

Formatting issues to be addressed

- Quotes
- Inappropriate data types
- Outliers
- Missing values

Modeling the data with neural networks

- Function ‘`nnetar`’ in the library ‘`forecast`’
- Producing 3 years of forecast

Cleaning the Dataset

Data Cleaning Takes Time

Analytical tools require clean datasets

Data rarely comes with proper formatting

Data collectors often lack the right tools and training

Steps of Data Cleaning

Data import

Removing quotations

Converting into a time series

Imputation of missing data and outliers

```
library(tidyr)

revenue <- separate(revenue, col = V2,
                    sep = c(2, -3),
                    into = c("rest", "data", "rest2"))
```

Removing the Quotations

Separating a string into substrings of given lengths

- Argument '`sep`' specifies where the new substrings begin
- Argument '`into`' specifies the header for the new columns

The Function ‘separate’

```
myts <- ts(as.numeric(revenue$data),
           start = 1998, frequency = 12)
```

Converting the Data into a Time Series

Conversion to numeric and then into time series

Time stamp: Monthly data starting in 1998

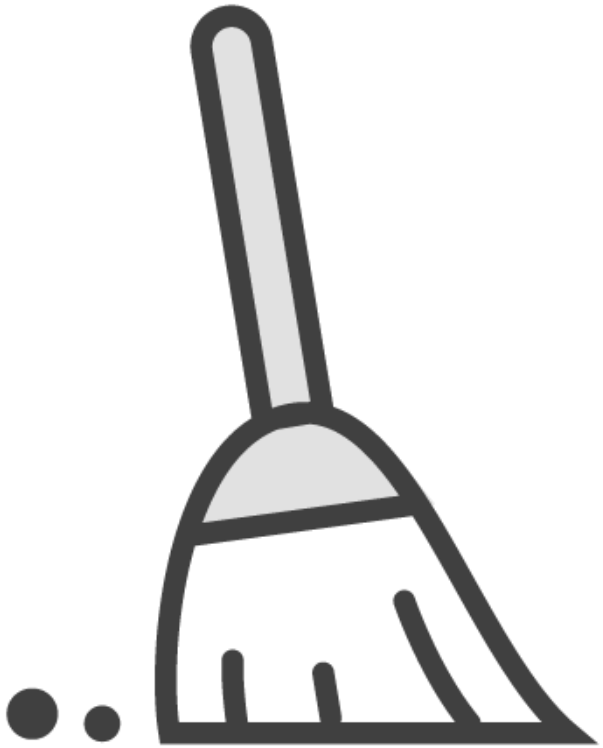
The Function ‘tsclean’ from Forecast

Missing data and outlier detection

Implementation with linear interpolation

Seasonally adjusted results

Data Cleaning Steps



Quote removal with `'separate()'`

Data type conversion

- Into numeric with `'as.numeric()'`
- Into time series with `'ts()'`

Cleaning with `'tsclean()'`

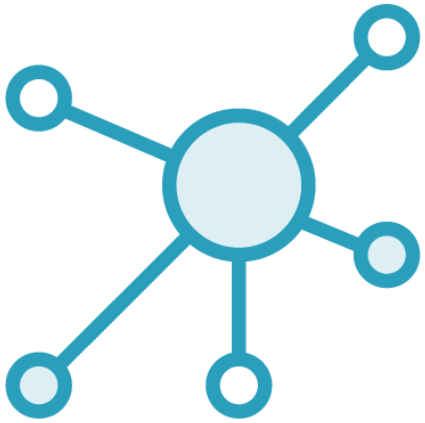
- Outliers
- Missing values
- Seasonal adjustment

Fitting the Neural Network

Predicting Sales Using Neural Networks

Data Cleaning

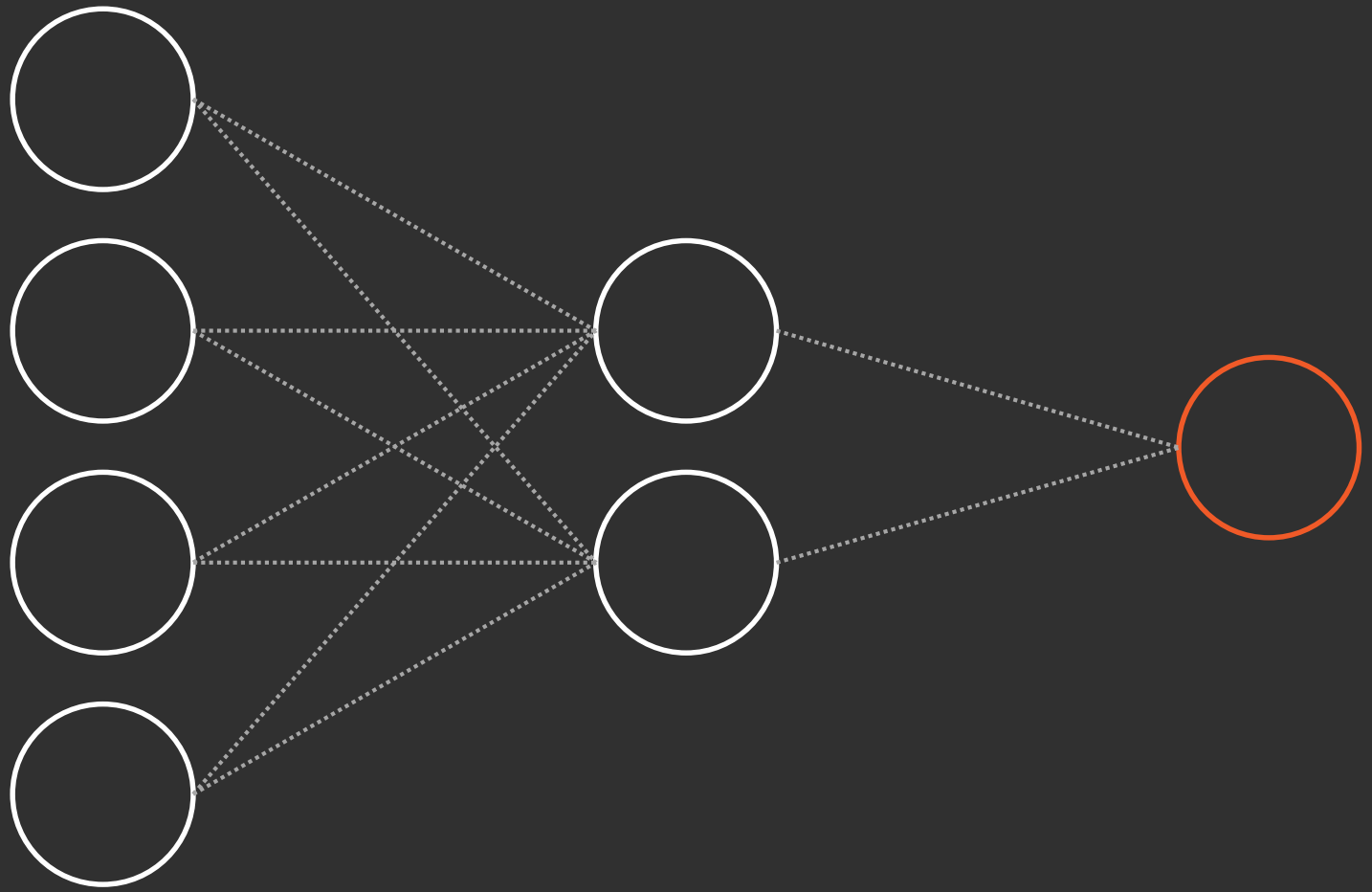
Preparing the dataset to be analyzed



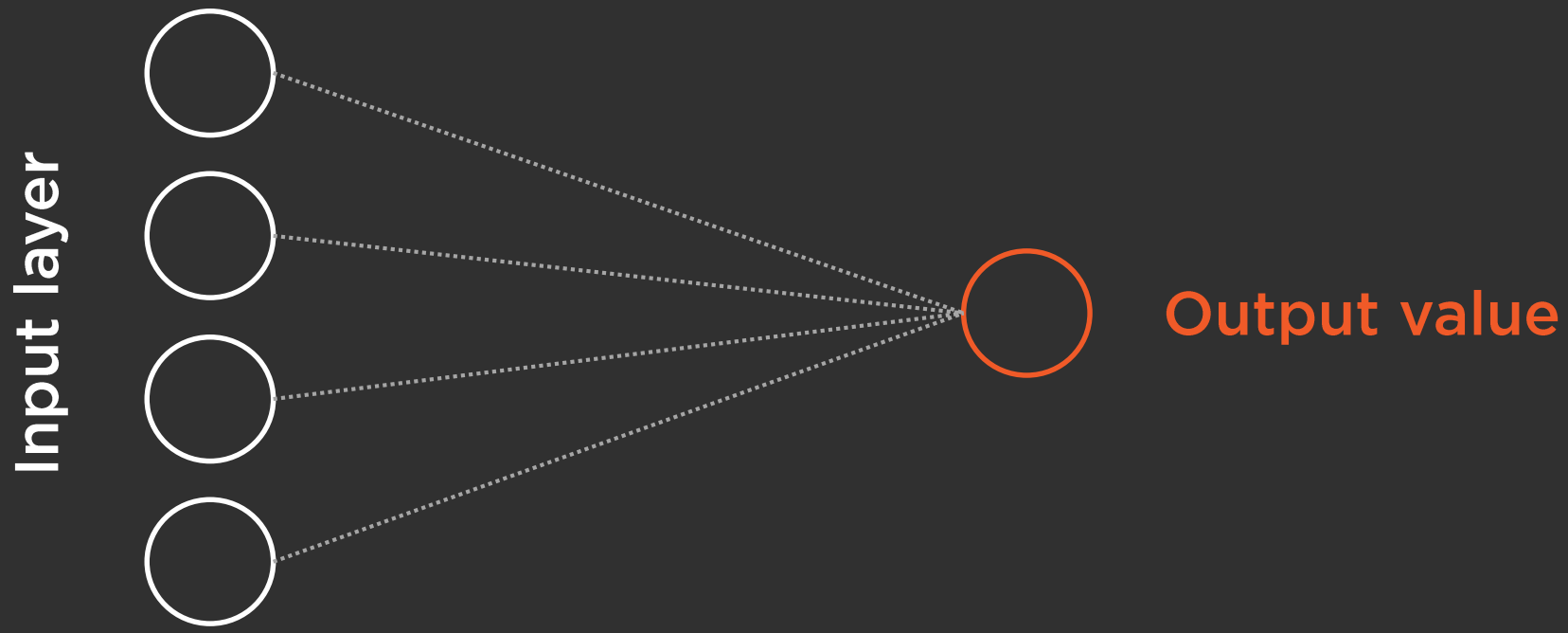
Fitting the Neural Network

Prediction using a machine learning technique

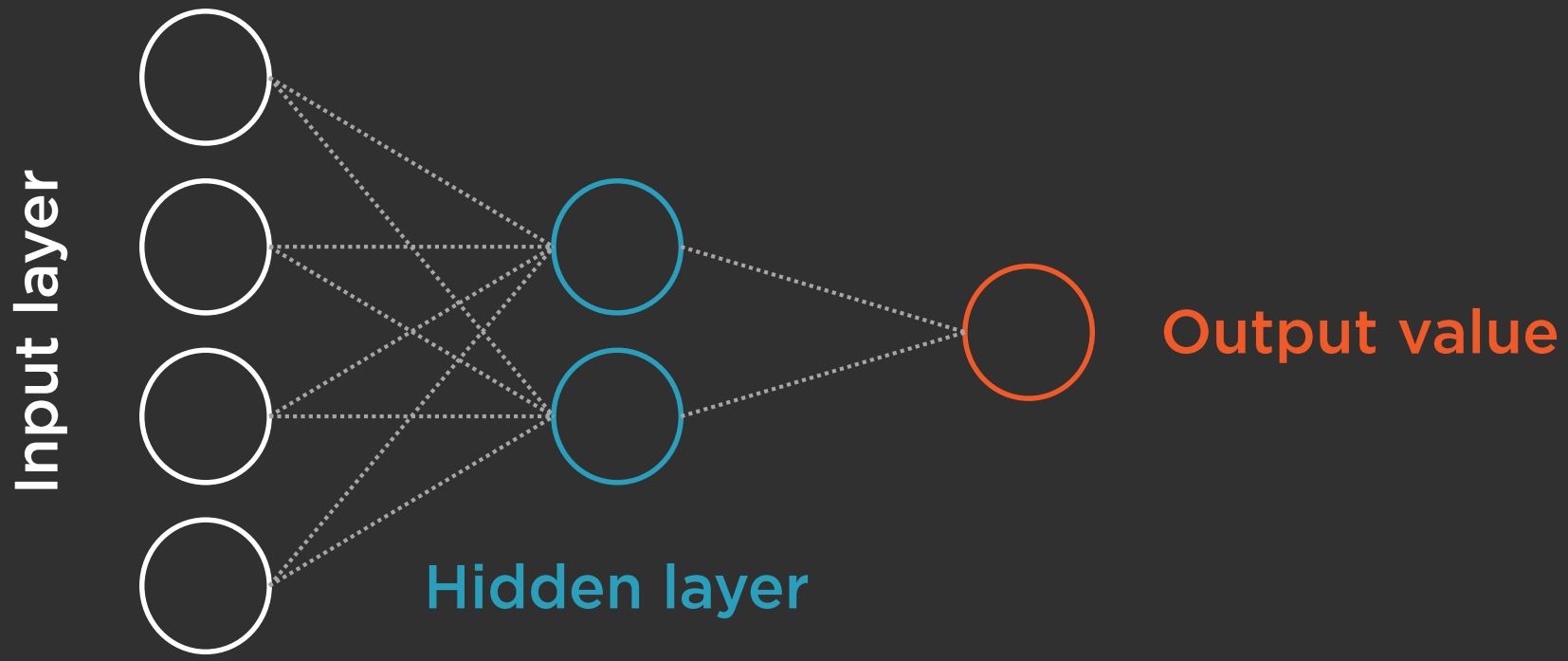
How Does a Neural Network Work?



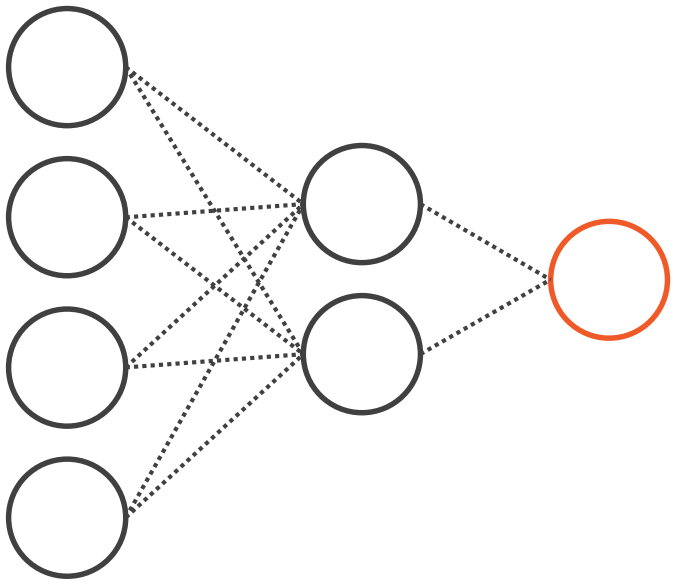
Simple Neural Net



Multilayer Feed Forward Network



Neural Network Auto Regression Model (NNAR)



Fitting the neural network on time series

NNAR(p, k)

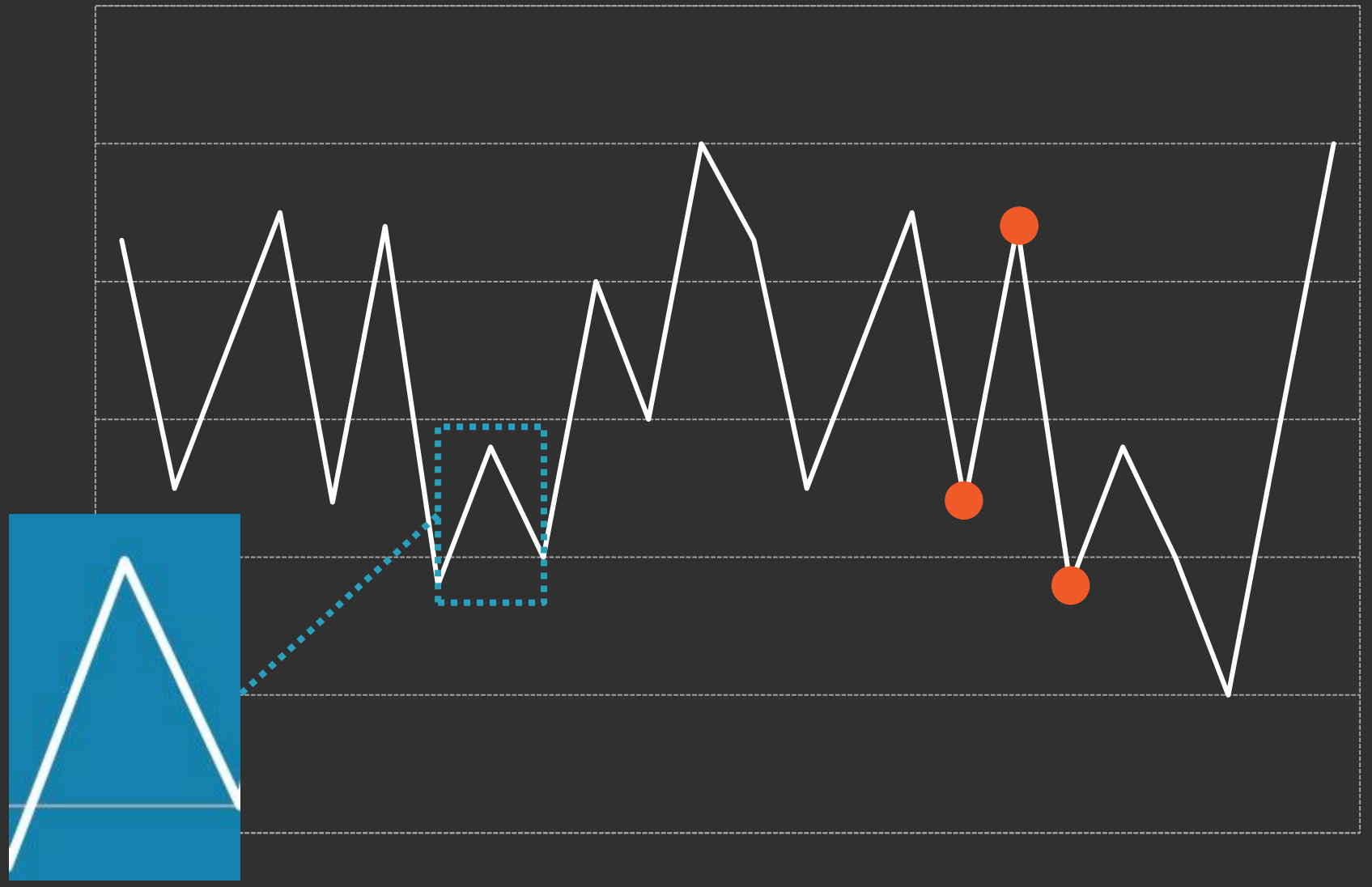
- Lagged values (p) used as inputs
- Number of nodes (k) in the hidden layer

NNAR(p, P, k)

- Seasonal lag (P): observation from the previous season(s)

Interactive Graph with Library ‘dygraphs’

Interactive Data Visualizations



Interactive Graphs with ‘dygraphs’

R library ‘dygraphs’ is an implementation of JavaScript library ‘dygraphs’

- Good quality functions
- Decent documentation
- Coding with the pipe operator %>%

Coding the Interactive Graph

```
dygraph(mydata, main = “Beach Restaurant") %>%

    dyRangeSelector() %>%

    dySeries(name = "data", label = "Revenue Data") %>%

    dySeries(c("lower","pforecast","upper"),
    label = "Revenue Forecast") %>%

    dyLegend(show = "always", hideOnMouseOut = FALSE) %>%

    dyAxis("y", label = "Monthly Revenue USD") %>%

    dyHighlight(highlightCircleSize = 5,
    highlightSeriesOpts = list(strokeWidth = 2)) %>%

    dyOptions(axisLineColor = "navy", gridLineColor = "grey") %>%
```

Coding the Interactive Graph

```
dygraph(mydata, main = “Beach Restaurant") %>%

    dyRangeSelector() %>%

    dySeries(name = "data", label = "Revenue Data") %>%

    dySeries(c("lower","pforecast","upper"),
    label = "Revenue Forecast") %>%

    dyLegend(show = "always", hideOnMouseOut = FALSE) %>%

    dyAxis("y", label = "Monthly Revenue USD") %>%

    dyHighlight(highlightCircleSize = 5,
    highlightSeriesOpts = list(strokeWidth = 2)) %>%

    dyOptions(axisLineColor = "navy", gridLineColor = "grey") %>%
```

Coding the Interactive Graph

```
dygraph(mydata, main = "Beach Restaurant") %>%

    dyRangeSelector() %>%

    dySeries(name = "data", label = "Revenue Data") %>%

    dySeries(c("lower","pforecast","upper"),
    label = "Revenue Forecast") %>%

    dyLegend(show = "always", hideOnMouseOut = FALSE) %>%

    dyAxis("y", label = "Monthly Revenue USD") %>%

    dyHighlight(highlightCircleSize = 5,
    highlightSeriesOpts = list(strokeWidth = 2)) %>%

    dyOptions(axisLineColor = "navy", gridLineColor = "grey") %>%
```

Coding the Interactive Graph

```
dygraph(mydata, main = "Beach Restaurant") %>%

    dyRangeSelector() %>%

    dySeries(name = "data", label = "Revenue Data") %>%

    dySeries(c("lower","pforecast","upper"),
    label = "Revenue Forecast") %>%

    dyLegend(show = "always", hideOnMouseOut = FALSE) %>%

    dyAxis("y", label = "Monthly Revenue USD") %>%

    dyHighlight(highlightCircleSize = 5,
    highlightSeriesOpts = list(strokeWidth = 2)) %>%

    dyOptions(axisLineColor = "navy", gridLineColor = "grey") %>%
```

Coding the Interactive Graph

```
dygraph(mydata, main = "Beach Restaurant") %>%

    dyRangeSelector() %>%

    dySeries(name = "data", label = "Revenue Data") %>%

    dySeries(c("lower","pforecast","upper"),
    label = "Revenue Forecast") %>%

    dyLegend(show = "always", hideOnMouseOut = FALSE) %>%

    dyAxis("y", label = "Monthly Revenue USD") %>%

    dyHighlight(highlightCircleSize = 5,
    highlightSeriesOpts = list(strokeWidth = 2)) %>%

    dyOptions(axisLineColor = "navy", gridLineColor = "grey") %>%
```

Coding the Interactive Graph

```
dygraph(mydata, main = "Beach Restaurant") %>%

    dyRangeSelector() %>%

    dySeries(name = "data", label = "Revenue Data") %>%

    dySeries(c("lower","pforecast","upper"),
    label = "Revenue Forecast") %>%

    dyLegend(show = "always", hideOnMouseOut = FALSE) %>%

    dyAxis("y", label = "Monthly Revenue USD") %>%

    dyHighlight(highlightCircleSize = 5,
    highlightSeriesOpts = list(strokeWidth = 2)) %>%

    dyOptions(axisLineColor = "navy", gridLineColor = "grey") %>%
```

Coding the Interactive Graph

```
dygraph(mydata, main = "Beach Restaurant") %>%

    dyRangeSelector() %>%

    dySeries(name = "data", label = "Revenue Data") %>%

    dySeries(c("lower","pforecast","upper"),
    label = "Revenue Forecast") %>%

    dyLegend(show = "always", hideOnMouseOut = FALSE) %>%

    dyAxis("y", label = "Monthly Revenue USD") %>%

    dyHighlight(highlightCircleSize = 5,
    highlightSeriesOpts = list(strokeWidth = 2)) %>%

    dyOptions(axisLineColor = "navy", gridLineColor = "grey") %>%
```

Coding the Interactive Graph

```
dygraph(mydata, main = "Beach Restaurant") %>%

    dyRangeSelector() %>%

    dySeries(name = "data", label = "Revenue Data") %>%

    dySeries(c("lower","pforecast","upper"),
    label = "Revenue Forecast") %>%

    dyLegend(show = "always", hideOnMouseOut = FALSE) %>%

    dyAxis("y", label = "Monthly Revenue USD") %>%

    dyHighlight(highlightCircleSize = 5,
    highlightSeriesOpts = list(strokeWidth = 2)) %>%

    dyOptions(axisLineColor = "navy", gridLineColor = "grey") %>%
```

Coding the Interactive Graph

```
dygraph(mydata, main = "Beach Restaurant") %>%

    dyRangeSelector() %>%

    dySeries(name = "data", label = "Revenue Data") %>%

    dySeries(c("lower","pforecast","upper"),
    label = "Revenue Forecast") %>%

    dyLegend(show = "always", hideOnMouseOut = FALSE) %>%

    dyAxis("y", label = "Monthly Revenue USD") %>%

    dyHighlight(highlightCircleSize = 5,
    highlightSeriesOpts = list(strokeWidth = 2)) %>%

    dyOptions(axisLineColor = "navy", gridLineColor = "grey") %>%
```

Coding the Interactive Graph

```
dygraph(mydata, main = "Beach Restaurant") %>%

    dyRangeSelector() %>%

    dySeries(name = "data", label = "Revenue Data") %>%

    dySeries(c("lower","pforecast","upper"),
    label = "Revenue Forecast") %>%

    dyLegend(show = "always", hideOnMouseOut = FALSE) %>%

    dyAxis("y", label = "Monthly Revenue USD") %>%

    dyHighlight(highlightCircleSize = 5,
    highlightSeriesOpts = list(strokeWidth = 2)) %>%

    dyOptions(axisLineColor = "navy", gridLineColor = "grey") %>%
```

Neural Networks

Modeling sales data of a restaurant with neural networks

Data cleaning:

- Symbol removal with ‘`tidyr`’
- Cleaning with the ‘`tsclean`’ function
- Missing data imputation
- Outlier detection and replacement

Modeling and forecasting with ‘`nnetar`’

Interactive data visualization with ‘`dygraphs`’