

# Package ‘MSBStatsData’

January 20, 2025

**Title** Data Sets for Courses at the Münster School of Business

**Version** 0.0.2

**Description** Provides sample data sets that are used in statistics and data science courses at the Münster School of Business. The datasets refer to different business topics but also other domains, e.g. sports, traffic, etc.

**License** GPL (>= 3)

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.2.3

**Depends** R (>= 2.10)

**Imports** tibble

**NeedsCompilation** no

**Author** Michael Bucker [aut, cre] (<<https://orcid.org/0000-0003-0045-8460>>),  
Niels Schlüsener [aut]

**Maintainer** Michael Bucker <michael.buecker@fh-muenster.de>

**Repository** CRAN

**Date/Publication** 2024-03-09 10:50:02 UTC

## Contents

decathlon . . . . .	2
hr_data . . . . .	2
marketing_expenses . . . . .	3
salary_trends . . . . .	4
shipping_errors . . . . .	4
sorting_times . . . . .	5
<b>Index</b>	<b>6</b>

---

decathlon	<i>Data of competition results of decathlon for men</i>
-----------	---

---

**Description**

The decathlon for men is a combined event in athletics consisting of 10 track and field events: 100 metres, 400 metres, 1500 metres, 110 metre hurdles, long jump, high jump, pole vault, discus throw, javelin throw, and shotput.

**Usage**

decathlon

**Format**

A tibble with 7,958 rows and 10 variables:

**race100m** dbl result of 100m race in seconds

**longjump** dbl result of long jump in meters

**shotput** dbl result of shot put in meters

**highjump** dbl result of high jump in meters

**race400m** dbl result of 400m race in seconds

**race110mhurdles** dbl result of 110m hurdles race in seconds

**discus** dbl result of discus throw in meters

**polevault** dbl result of pole vault in meters

**javelinthrow** dbl result of javelin throw in meters

**race1500m** dbl result of 1500m race in seconds

---

hr_data	<i>HR data</i>
---------	----------------

---

**Description**

The dataset contains human resources data of a footwear company. Each entity of the dataset represents one employee with a total of six attributes.

**Usage**

hr\_data

**Format**

A tibble with 1,200 rows and 6 variables:

**position** [fct ] Position of the employee in the company.

**working\_hours** [dbl ] Contracted working hours of the employee per week.

**salary** [dbl ] Monthly salary of the employee.

**hourly\_wage** [dbl ] Hourly compensation of the employee.

**department** [fct ] Department in which the employee is employed.

**sick\_days** [dbl ] Sick days of the employee in the period observed.

---

marketing\_expenses      *Data of Marketing expenses*

---

**Description**

The dataset contains product-, marketing- and sales data of 235 shoes of a footwear company. Each entity represents one shoe, listed with its product data, marketing data and sales data in a total of 14 variables.

**Usage**

marketing\_expenses

**Format**

A tibble with 235 rows and 14 variables:

**marketing\_expenses** [dbl ] Expenses for marketing activities for the shoe.

**customers\_reached** [dbl ] Estimated number of customers reached by the footwear marketing activity of the shoe.

**negative\_reactions** [dbl ] Number of negative reactions to the marketing activities of the shoe.

**price** [dbl ] Retail price of the shoe.

**price\_segment** [fct ] Price segment of the shoe.

**number\_of\_sizes** [dbl ] Number of sizes in which the shoe is available.

**target\_customer** [fct ] Gender the shoe is intended for.

**rating\_testers** [dbl ] Average product rating of the test customers for the shoe.

**rating\_customers** [dbl ] Average product rating of the real customers for the shoe.

**color\_most\_sold** [fct ] Color in which the shoe is sold the most often.

**return\_rate** [dbl ] Rate how often the shoe is returned by the customer.

**sales\_volume** [dbl ] Number of sales for the shoe.

**rank\_rating\_customers** [dbl ] Attribute 'rating\_customers', divided into ranks for the calculation of the correlation coefficient according to Spearman.

**rank\_price\_segment** [dbl ] Attribute 'price\_segment', divided into ranks for the calculation of the correlation coefficient according to Spearman.

---

salary_trends	<i>Salary trend data</i>
---------------	--------------------------

---

**Description**

Dataset containing the average salary of a footwear company's employees over 10 years.

**Usage**

salary\_trends

**Format**

A tibble with 10 rows and 2 variables:

**year** [dbl ] Year of record.

**avg\_salary** [dbl ] Average salary in the corresponding year.

---

shipping_errors	<i>Shipping error data</i>
-----------------	----------------------------

---

**Description**

Dataset recording errors in the shipment and the type of delivery requested by the customer in the process.

**Usage**

shipping\_errors

**Format**

A tibble with 576 rows and 2 variables:

**error** [fct ] Indicator whether or not an error appeared during shipment.

**shipping** [fct ] Shipment method requested by the customer.

---

sorting_times	<i>Sorting times</i>
---------------	----------------------

---

**Description**

The dataset contains throughput times of a machine recorded in order to compare them to the manufacturer's specifications.

**Usage**

```
sorting_times
```

**Format**

A tibble with 60 rows and 1 variable:

**sorting\_time** [dbl ] Recorded time (in seconds) required by the machine for sorting individual parts.

# Index

## \* datasets

decathlon, 2

hr\_data, 2

marketing\_expenses, 3

salary\_trends, 4

shipping\_errors, 4

sorting\_times, 5

decathlon, 2

hr\_data, 2

marketing\_expenses, 3

salary\_trends, 4

shipping\_errors, 4

sorting\_times, 5