Package: peopleanalytics (via r-universe)

August 30, 2024

Title Data Sets for Craig Starbuck's Book, ``The Fundamentals of People Analytics: With Applications in R''

Version 0.1.0

Description Data sets associated with modeling examples in Craig Starbuck's book, ``The Fundamentals of People Analytics: With Applications in R''.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.1

Depends R (>= 3.5.0)

LazyData true

NeedsCompilation no

Author Craig Starbuck [aut, cre]
(<https://orcid.org/0000-0001-5089-2850>)

Maintainer Craig Starbuck <cstarbuck@orgacuity.com>

Date/Publication 2022-10-25 17:22:35 UTC

Repository https://cstarbuck.r-universe.dev

RemoteUrl https://github.com/cran/peopleanalytics

RemoteRef HEAD

RemoteSha c43a4544c0c3a2b85faa60dbf6641171ada4e16c

Contents

---

benefits   *benefits*

Description

Fictitious benefits data for employees in a mid-size company

Usage

```
data("benefits")
```

Format

A data frame with 1471 observations on the following 3 variables.

employee_id Unique identifier for each employee

stock_opt_lvl Job level, where 1 = 'Junior' and 5 = 'Senior'

trainings Number of trainings completed within the past year

Examples

```
data(benefits)
```

---

demographics   *demographics*

Description

Fictitious demographics data for employees in a mid-size company

Usage

```
data("demographics")
```

Format

A data frame with 1470 observations on the following 7 variables.

employee_id Unique identifier for each employee

age Employee age in years

commute_dist Commute distance in miles

ed_lvl Education level, where 1 = 'High School', 2 = 'Associate Degree', 3 = 'Bachelor's Degree', 4 = 'Master's Degree', and 5 = 'Doctoral Degree'

ed_field Education field associated with most recent degree

gender Gender self-identification

marital_sts Marital status

Examples

```
data(demographics)
```

---

employees   *employees*

Description

Fictitious data on employees in a mid-size company

Usage

```
data("employees")
```

Format

A data frame with 1470 observations on the following 36 variables.

employee_id Unique identifier for each employee

active Flag set to 'Yes' for active employees and 'No' for inactive employees

stock_opt_lvl Stock option level

trainings Number of trainings completed within the past year

age Employee age in years

commute_dist Commute distance in miles

ed_lvl Education level, where 1 = 'High School', 2 = 'Associate Degree', 3 = 'Bachelor's Degree', 4 = 'Master's Degree', and 5 = 'Doctoral Degree'

ed_field Education field associated with most recent degree

gender Gender self-identification

marital_sts Marital status

dept Department of which an employee is a member

engagement Employee engagement score measured on a 4-point Likert scale, where 1 = 'Highly Disengaged' and 4 = 'Highly Engaged'

job_lvl Job level, where 1 = 'Junior' and 5 = 'Senior'

job_title Job title

overtime Flag set to 'Yes' if the employee is nonexempt and works overtime and 'No' if the employee does not work overtime

business_travel Business travel frequency

hourly_rate Hourly rate calculated irrespective of hourly/salaried employees

daily_comp Hourly rate * 8

monthly_comp Hourly rate * 2080 / 12

annual_comp Hourly rate * 2080

ytd_leads Year-to-date (YTD) number of leads generated for employees in Sales Executive and Sales Representative positions

ytd_sales Year-to-date (YTD) sales measured in USD for employees in Sales Executive and Sales Representative positions

standard_hrs Expected working hours over a two-week payroll cycle

salary_hike_pct The percent increase in salary for the employee's most recent compensation adjustment (whether due to a standard merit increase, off-cycle adjustment, or promotion)

perf_rating Most recent performance rating, where 1 = 'Needs Improvement', 2 = 'Core Contributor', 3 = 'Noteworthy', and 4 = 'Exceptional'

prior_emplr_cnt Number of prior employers

env_sat Environment satisfaction score measured on a 4-point Likert scale, where 1 = 'Highly Dissatisfied' and 4 = 'Highly Satisfied'

job_sat Job satisfaction score measured on a 4-point Likert scale, where 1 = 'Highly Dissatisfied' and 4 = 'Highly Satisfied'

rel_sat Collegue relationship satisfaction score measured on a 4-point Likert scale, where 1 = 'Highly Dissatisfied' and 4 = 'Highly Satisfied'

wl_balance Work-life balance score measured on a 4-point Likert scale, where 1 = 'Poor Balance' and 4 = 'Excellent Balance'

work_exp Total years of work experience

org_tenure Years at current company

job_tenure Years in current job

last_promo Years since last promotion

mgr_tenure Years under current manager

interview_rating Average rating across the interview loop for the onsite stage of the employee's recruiting process, where 1 = 'Definitely Not' and 5 = 'Definitely Yes'

Examples

```
data(employees)
```

---

job   *job*

Description

Fictitious job data for employees in a mid-size company

Usage

```
data("job")
```

Format

A data frame with 1470 observations on the following 6 variables.

employee_id Unique identifier for each employee

dept Department of which an employee is a member

job_lvl Job level, where 1 = 'Junior' and 5 = 'Senior'

job_title Job title

overtime Flag set to 'Yes' if the employee is nonexempt and works overtime and 'No' if the employee does not work overtime

business_travel Business travel frequency

Examples

```
data(job)
```

---

payroll   *payroll*

Description

Fictitious payroll data for employees in a mid-size company

Usage

```
data("payroll")
```

Format

A data frame with 1470 observations on the following 6 variables.

`employee_id` Unique identifier for each employee

`hourly_rate` Hourly rate calculated irrespective of hourly/salaried employees

`daily_comp` Hourly rate * 8

`monthly_comp` Hourly rate * 2080 / 12

`annual_comp` Hourly rate * 2080

`standard_hrs` Expected working hours over a two-week payroll cycle

Examples

```
data(payroll)
```

---

performance	*performance*

Description

Fictitious performance data for employees in a mid-size company

Usage

```
data("performance")
```

Format

A data frame with 1470 observations on the following 3 variables.

`employee_id` Unique identifier for each employee

`salary_hike_pct` The percent increase in salary for the employee's most recent compensation adjustment (whether due to a standard merit increase, off-cycle adjustment, or promotion)

`perf_rating` Most recent performance rating, where 1 = 'Needs Improvement', 2 = 'Core Contributor', 3 = 'Noteworthy', and 4 = 'Exceptional'

Examples

```
data(performance)
```

---

prior_employment *prior_employment*

Description

Fictitious prior employment data for employees in a mid-size company

Usage

```
data("prior_employment")
```

Format

A data frame with 1470 observations on the following 2 variables.

employee_id Unique identifier for each employee

prior_emplr_cnt Number of prior employers

Examples

```
data(prior_employment)
```

---

sentiment *sentiment*

Description

Fictitious sentiment data for employees in a mid-size company

Usage

```
data("sentiment")
```

Format

A data frame with 1470 observations on the following 6 variables.

employee_id Unique identifier for each employee

env_sat Environment satisfaction score measured on a 4-point Likert scale, where 1 = 'Highly Dissatisfied' and 4 = 'Highly Satisfied'

engagement Employee engagement score measured on a 4-point Likert scale, where 1 = 'Highly Disengaged' and 4 = 'Highly Engaged'

job_sat Job satisfaction score measured on a 4-point Likert scale, where 1 = 'Highly Dissatisfied' and 4 = 'Highly Satisfied'

rel_sat Colleague relationship satisfaction score measured on a 4-point Likert scale, where 1 = 'Highly Dissatisfied' and 4 = 'Highly Satisfied'

wl_balance Work-life balance score measured on a 4-point Likert scale, where 1 = 'Poor Balance' and 4 = 'Excellent Balance'

Examples

```
data(sentiment)
```

---

status | *status*

Description

Fictitious data on the active status of employees in a mid-size company

Usage

```
data("status")
```

Format

A data frame with 1470 observations on the following 2 variables.

employee_id Unique identifier for each employee

active Flag set to 'Yes' for active employees and 'No' for inactive employees

Examples

```
data(status)
```

---

survey_responses | *survey_responses*

Description

Fictitious survey responses for anonymized employees in a mid-size company

Usage

```
data("survey_responses")
```

Format

A data frame with 400 observations on the following 12 variables.

belong Belonging score measured on a 5-point Likert scale, where 1 = 'Highly Unfavorable' and 5 = 'Highly Favorable'

effort Discretionary Effort score measured on a 5-point Likert scale, where 1 = 'Highly Unfavorable' and 5 = 'Highly Favorable'

incl Inclusion score measured on a 5-point Likert scale, where 1 = 'Highly Unfavorable' and 5 = 'Highly Favorable'

eng_1 Engagement score on item 1 of 3 measured on a 5-point Likert scale, where 1 = 'Highly Disengaged' and 5 = 'Highly Engaged'

eng_2 Engagement score on item 2 of 3 measured on a 5-point Likert scale, where 1 = 'Highly Disengaged' and 5 = 'Highly Engaged'

eng_3 Engagement score on item 3 of 3 measured on a 5-point Likert scale, where 1 = 'Highly Disengaged' and 5 = 'Highly Engaged'

happ Happiness score measured on a 5-point Likert scale, where 1 = 'Highly Unfavorable' and 5 = 'Highly Favorable'

psafety Psychological Safety score measured on a 7-point Likert scale, where 1 = 'Highly Unfavorable' and 7 = 'Highly Favorable'

ret_1 Retention score on item 1 of 3 measured on a 5-point Likert scale, where 1 = 'Highly Unfavorable' and 5 = 'Highly Favorable'

ret_2 Retention score on item 2 of 3 measured on a 5-point Likert scale, where 1 = 'Highly Unfavorable' and 5 = 'Highly Favorable'

ret_3 Retention score on item 3 of 3 measured on a 5-point Likert scale, where 1 = 'Highly Unfavorable' and 5 = 'Highly Favorable'

ldrshp Senior Leadership score measured on a 5-point Likert scale, where 1 = 'Highly Unfavorable' and 5 = 'Highly Favorable'

Examples

```
data(survey_responses)
```

---

tenure *tenure*

Description

Fictitious tenure data for employees in a mid-size company

Usage

```
data("tenure")
```

Format

A data frame with 1470 observations on the following 6 variables.

employee_id Unique identifier for each employee

work_exp Flag set to 'Yes' for active employees and 'No' for inactive employees

org_tenure Years at current company

job_tenure Years in current job

last_promo Years since last promotion

mgr_tenure Years under current manager

Examples

```
data(tenure)
```

---

turnover_trends *turnover_trends*

Description

Fictitious monthly employee turnover rates by several dimensions

Usage

```
data("turnover_trends")
```

Format

A data frame with 3000 observations on the following 6 variables.

year Integer representing the year, which ranges from 1 (earliest) to 5 (most recent)

month Integer representing the month, which ranges from 1 (January) to 12 (December)

job Job title

level Job level, where 1 = 'Junior' and 5 = 'Senior'

remote Flag set to 'Yes' for a remote worker and 'No' for a non-remote worker

turnover_rate Monthly turnover rate, calculated by dividing the termination count into the average headcount (beginning headcount + ending headcount / 2) for the respective month

Examples

```
data(turnover_trends)
```

Index