

# Package ‘ExamPData’

October 12, 2022

**Type** Package

**Title** Data Sets for Predictive Analytics Exam

**Version** 0.5.0

**Date** 2021-10-20

**Description** Contains all data sets for Exam PA: Predictive Analytics at  
<<https://exampa.net/>>.

**URL** <https://github.com/sdcastillo/ExamPData>

**BugReports** <https://github.com/sdcastillo/ExamPData/issues>

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**Depends** R (>= 3.5.0)

**RoxygenNote** 7.1.1

**NeedsCompilation** no

**Author** Guanglai Li [aut, cre],  
Sam Castillo [aut]

**Maintainer** Guanglai Li <[liguanglai@gmail.com](mailto:liguanglai@gmail.com)>

**Repository** CRAN

**Date/Publication** 2021-10-23 15:10:01 UTC

## R topics documented:

actuary_salaries . . . . .	2
apartment_apps . . . . .	3
auto_claim . . . . .	4
bank_loans . . . . .	5
bike_sharing_demand . . . . .	6
boston . . . . .	7
customer_phone_calls . . . . .	8
customer_value . . . . .	9

exam_pa_titanic . . . . .	9
health_insurance . . . . .	10
june_pa . . . . .	11
patient_length_of_stay . . . . .	12
patient_num_labs . . . . .	13
pedestrian_activity . . . . .	13
readmission . . . . .	14
student_success . . . . .	15
travel_insurance . . . . .	16
travel_spending . . . . .	17
<b>Index</b>	<b>18</b>

---

actuary_salaries	<i>DW Simpson actuarial salary data</i>
------------------	---

---

## Description

The data actuary\_salaries contains the salaries of actuaries collected from the DW Simpson survey.

## Usage

```
actuary_salaries
```

## Format

data.frame, 138 observations of 6 variables:

**industry** The industry of the actuary, having values of Casualty, Health, Pension, and Life

**exams** The number of exams passed. Values of ASA, FSA, 5,4,3,2,1

**experience** Years of experience, in the range 1 - 20

**salary** Annual salary range, in \$1,000

**salary\_low** Lower end of the annual salary range

**salary\_high** Higher end of the annual salary range

---

apartment_apps	<i>Apartment Apps</i>
----------------	-----------------------

---

### Description

Apartment applications as used in ExamPA.net's Practice Exam

### Usage

apartment\_apps

### Format

data.frame, 1430 observations of 41 variables:

**applicants** The total number of people who apply for a lease at that apartment building, including all apartment units.

**sale\_price** The sale price of each apartment unit.

**num\_units** The number of units in the apartment building.

**year\_sold** Year that apartment building was sold or remodeled.

**month\_sold** Month that apartment building was sold or remodeled.

**overall\_qual** Rates the overall material and finish of the building on a scale from 1 to 10 with 10 being the best and 1 being the worst.

**total\_sq\_feet** Total square feet.

**gr\_liv\_area** Above ground living area in square feet.

**tot\_bathrooms** The number of bathroom of each unit.

**lot\_area** Lot size in square feet.

**exter\_qual** Rates the external quality of the building on a scale from 1 to 10 with 10 being the best and 1 being the worst.

**full\_bath** The number of full-size bathroom in each unit.

**central\_air** Whether or not each unit has a central air conditioning system (1 = yes, 0 = no).

**garage\_type\_attchd** 1 = Attached garage.

**garage\_type\_basment** 1 = Basement garage.

**garage\_type\_builtIn** 1 = Build in garage.

**garage\_type\_detchd** 1 = Detached garage.

**garage\_type\_no\_garage** 1 = No garage.

**NeighborhoodBrDale** 1 = Dale

**neighborhood\_brk\_side** 1 = Brookside.

**neighborhood\_clear\_cr** 1 = Clear Circle.

**neighborhood\_collg\_cr** 1 = College Circle.

**neighborhood\_crawfor** 1 = Crawford.  
**neighborhood\_edwards** 1 = Edwards.  
**neighborhood\_gilbert** 1 = Gilbert.  
**neighborhood\_idottrr** 1 = DOTRR.  
**neighborhood\_meadowv** 1 = Meadow.  
**neighborhood\_mitchel** 1 = Mitchel.  
**neighborhood\_n\_ames** 1 = North Ames  
**neighborhood\_n\_ridge** 1 = North Ridge.  
**neighborhood\_n\_ridge\_hghts** 1 = North Ridge Heights.  
**neighborhood\_n\_w\_ames** 1 = North West Ames.  
**neighborhood\_old\_town** 1 = Old Town.  
**neighborhood\_sawyer** 1 = Sawyer.  
**neighborhood\_sawyer\_w** 1 = Sawyer West.  
**neighborhood\_somerst** 1 = Somer St.  
**neighborhood\_stone\_br** 1 = Stone Bridge.  
**neighborhood\_swisu** 1 = SWISU.  
**neighborhood\_timber** 1 = Timber.  
**neighborhood\_veenker** 1 = Veenker.  
**neighborhood\_saleprice** The mean sale price for all units in that neighborhood.

---

 auto\_claim

*Automotive claims*


---

### Description

Automotive claims

### Usage

auto\_claim

### Format

data.frame, 10296 observations of 29 variables:

**POLICYNO** Policy number.

**PLCYDATE** Date that policy was signed.

**CLM\_FREQ5** Number of claims.

**CLM\_AMT5** Aggregate claim loss of policy (in thousands).

**CLM\_AMT**

**KIDSDRIV** Number of child passengers.

**TRAVTIME** Time to commute.

**CAR\_USE** (1) Private or (2) commercial use.

**BLUEBOOK** (log) car value.

**RETAINED** Whether the policy was retained or not.

**NPOLICY** Number of policies.

**CAR\_TYPE** (0-1 dummy variables) Type of car : (base) Panel Truck, (2) Pickup,(3) Sedan, (4) Sports Car, (5) SUV, (6) Van

**RED\_CAR** Whether the color of the car is (2) car or (1) not.

**REVOLVED** Whether the policyholder's license was (2) revoked in the past or (1) not.

**MVR\_PTS** Number of motor vehicle record points.

**CLM\_FLAG** Whether there was a claim or not.

**AGE** Age.

**HOMEKIDS** Number of children at home.

**YOJ** Year of job.

**INCOME** Annual income.

**GENDER** Gender of policyholder : (1) female or (2) male.

**MARRIED** Whether the policyholder is (2) married or (1) not.

**PARENT1** Whether (2) the policyholder grew up in a single-parent family or (1) not.

**JOBCLASS** (0-1 dummy variables) Job class of policyholder: (base) Unknown, (2) Blue Collar, (3) Clerical, (4) Doctor, (5) Home Maker, (6) Lawyer, (7) Manager, (8) Professional, (9) Student

**MAX\_EDUC** (0-1 dummy variables) Maximal level of education of policyholder: (base) less than High School, (2) Bachelors, (3) High School, (4) Masters, (5) PhD.

**HOME\_VAL** Value of home.

**SAMEHOME** Whether they grew up in the same home as their current home.

**AREA** (1) Rural or (2) urban area.

**IN\_YY** Year.

---

 bank\_loans

*Bank Loans*


---

### Description

Credit data from UCI Machine Learning Repository.

### Usage

bank\_loans

**Format**

data.frame, 41188 observations of 21 variables:

**age** age (numeric).

**job** type of job (categorical).

**marital** marital status (categorical).

**education** 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')

**default** has credit in default? (categorical).

**housing** has housing loan? (categorical).

**loan** has personal loan? (categorical).

**contact** contact communication type (categorical).

**month** last contact month of year (categorical).

**day\_of\_week** last contact day of the week (categorical).

**duration** last contact duration, in seconds (numeric). Important note - this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

**campaign** number of contacts performed during this campaign and for this client (numeric, includes last contact)

**pdays** number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted).

**previous** number of contacts performed before this campaign and for this client (numeric).

**poutcome** outcome of the previous marketing campaign (categorical).

**emp.var.rate** employment variation rate.

**cons.price.idx** consumer price index.

**cons.conf.idx** consumer confidence index.

**euribor3m** euribor 3 month rate.

**nr.employed** number of employees.

**y** has the client subscribed a term deposit?

---

bike_sharing_demand	<i>Bike sharing demand</i>
---------------------	----------------------------

---

**Description**

bike sharing demand dataset

**Usage**

bike\_sharing\_demand

**Format**

data.frame, 17376 observations of 10 variables:

**season** Season. 1 - winter, 2 - spring, 3 - summer, 4 - fall.

**year** Year. 0 - 2011, 1 - 2012

**hour** Hour.

**holiday** Whether the day is a holiday.

**weekday** Day of the week.

**weathersit** Weather situation. 1 - clear of partly cloudy, 2 - mist, 3 - rain or snow.

**temp** Normalized temperature in Celsius. The values are derived via  $(t - t_{\min}) / (t_{\max} - t_{\min})$ ,  
 $t_{\min} = -9$ ,  $t_{\max} = +39$ .

**humidity** Normalized humidity. The values are divided by 100 (max).

**windspeed** Normalized windspeed. The values are divided by 67 (max).

**bikes\_per\_hour** Count of rental bikes in each hour.

---

boston

*Boston*

---

**Description**

Boston housing data set

**Usage**

boston

**Format**

data.frame, 506 observations of 14 variables:

**crim** per capita crime rate by town.

**zn** proportion of residential land zoned for lots over 25,000 sq.ft.

**indus** proportion of non-retail business acres per town.

**chas** Charles River dummy variable (= 1 if tract bounds river; 0 otherwise).

**nox** nitrogen oxides concentration (parts per 10 million).

**rm** average number of rooms per dwelling.

**age** proportion of owner-occupied units built prior to 1940.

**dis** weighted mean of distances to five Boston employment centers.

**rad** index of accessibility to radial highways.  
**tax** full-value property-tax rate per \$10,000.  
**ptratio** pupil-teacher ratio by town.  
**black**  $1000(Bk - 0.63)^2$  where Bk is the proportion of blacks by town.  
**lstat** lower status of the population (percent).  
**medv** median value of owner-occupied homes in \$1000s.

---

customer\_phone\_calls *Customer Phone Calls*

---

### Description

Data used on June 18, 2020 Exam PA

### Usage

customer\_phone\_calls

### Format

data.frame, 10000 observations of 14 variables:

**age** Age of the prospective customer. Integer from 17 to 98.  
**job** Occupation category. Factor with 11 levels.  
**marital** Marital status. Factor with levels divorced, married, and single  
**housing** Indicates whether the prospect has a housing loan. Factor with levels yes and no.  
**loan** Indicates whether the prospect has a consumer loan. Factor with levels yes and no.  
**phone** The type of phone the prospect uses. Factors with levels cellular and landline.  
**month** The month of the marketing call. Factor with 12 levels.  
**weekday** The day of the week of the marketing call. Factor with five levels.  
**CPI** Consumer price index at the time of the call. Numeric from 92.20 to 94.77.  
**CCI** Consumer confidence index at the time of the call. Numeric from -50.8 to -26.9.  
**irate** Short term interest rate at the tie of the call. Numeric from 0.634 to 5.045.  
**employment** Number of employees of ABC Insurance at the time of the call. Numeric from 4964 to 5228.  
**purchase** Indicator of purchase. Integer (1 for purchase, or 0 for no purchase.)  
**edu\_years** Years of education. Integer from 1 to 16.



---

customer_value	<i>Customer Value</i>
----------------	-----------------------

---

**Description**

Customer value data set from December 2019 PA

**Usage**

customer\_value

**Format**

data.frame, 48842 observations of 8 variables:

**age** Age of the prospective policyholder. Integer from 17 - 90

**education\_num** Indicator of the amount of education - it is not the number of years of education, but a higher number does mean more years. Integer from 1 to 16.

**marital\_status** For married, AF means alternative form while civ means civil. Factor with seven levels.

**occupation** Occupations have been grouped into five categories. There is no indication regarding what they mean. A sixth group represents cases where the occupation is unknown. Factor with six levels.

**cap\_gain** Capital gains recorded on investments. Integer from 0 to 99,999.

**hours\_per\_week** Number of hours worked per week. Integer from 1 to 99

**score** A proprietary "insurance score" developed by MEB. Real number with two decimal places.

**value\_flag** Indicator a policy holder being High or Low value. Factor with two levels.

---

exam_pa_titanic	<i>Exam PA Titanic</i>
-----------------	------------------------

---

**Description**

Titanic passengers as used in ExamPA.net's practice exam

**Usage**

exam\_pa\_titanic

**Format**

data.frame, 906 observations of 11 variables:

**passengerid** Passenger id

**survived** Survived Y/N

**pclass** Ticket class

**name** Name

**sex** male, female

**age** Age

**sibsp** # of siblings

**parch** # of parents or children aboard the Titanic.

**ticket** Number fare

**fare** Cost of ticket.

**embarked** Port of Embarkation. C = Cherbourg, Q = Queenstown, S = Southampton.

---

health\_insurance

*Health insurance*

---

**Description**

Health insurance claims as used in ExamPA.net's Practice Exam. The data set consists of prior year's health insurance claims, along with patient demographic information, from Freedom Health.

**Usage**

health\_insurance

**Format**

data.frame, 1338 observations of 7 variables:

**age** Age of policy holder.

**sex** M or F.

**bmi** Body Mass Index: weight divided by height.

**children** Number of children.

**smoker** Smoker status. Yes or No.

**region** Geographic region.

**charges** Annual medical claims for this policy.

june\_pa

*June\_pa***Description**

Auto crash data set from SOA June 2019 PA

**Usage**

june\_pa

**Format**

data.frame, 23137 observations of 14 variables:

**Crash\_Score** Measure the extent of the crash using factors such as number of injuries and fatalities, the number of vehicles involved, and other factors. A positive number with two decimal places.

**year** Calendar year of the crash. Integer 2014 - 2019.

**Month** Calendar month of the crash. Integer 1 - 12 (1 = January, 12 = December.)

**Time\_of\_Day** Time of day, on 4-hour blocks. Integer 1 - 6 (1 = midnight to 4am, 6 = 8pm to midnight.)

**Rd\_Feature** Special features of the road where the crash occurred. NONE = no special feature, INTERSECTION = the meeting of at least two roads, RAMP = exit or entrance ramp to a controlled access road, DRIVEWAY = entrance to home of business, OTHER.

**Rd\_Character** Description of the road where the crash occurred. STRAIGHT-LEVEL = no curves or hills, STRAIGHT-GRADE = no curves, but on a hill (up or down), STRAIGHT-OTHER, CURVE-LEVEL = on a curve but no hill, CURVE-GRADE = on a curve and on a hill, CURVE-OTHER, OTHER.

**Rd\_Class** Classification of the road type. STATE HWY = maintained by the state government, US HWY = maintained by the federal government.

**Rd\_Configuration** Design of the road. TWO-WAY-PROTECTED-MEDIAN = traffic in both directions, separated with a barrier, TWO-WAY-UNPROTECTED-MEDIAN = separated but with no barrier, TWO-WAY-NO-MEDIAN = no separation, ONE-WAY, UNKNOWN.

**Rd\_Surface** Material used for the road surface. SMOOTH ASPHALT, COARSE ASPHALT, CONCRETE, GROOVED CONCRETE, OTHER.

**Rd\_Conditions** Condition of the road. DRY, WET, ICE-SNOW-SLUSH, OTHER.

**Light** Lighting. DAYLIGHT, DARK-NOT-LIT = no street lamps in area, DARK-LIT, DUSK, DAWN, OTHER.

**Weather** Weather conditions. CLEAR, RAIN, CLOUDY, SNOW, OTHER.

**Traffic\_Control** Any items that control traffic flow. SIGNAL = lighted stop/go signal, STOP-SIGN, YIELD, NONE, OTHER.

**Work\_Area** Whether the crash in a work area? YES/NO

---

patient\_length\_of\_stay

*Patient Length of Stay*

---

### Description

Data used on June 16, 2020 Exam PA

### Usage

patient\_length\_of\_stay

### Format

data.frame, 10000 observations of 13 variables:

**days** Number of days between admission into and discharge from hospital. Integer 1 - 14.

**gender** Patient gender. Male or Female.

**age** Patient age (in 10-year age bands). [0, 10), [10, 20), ..., [90, 100)

**race** patient race. AfricanAmerican, Asian, Caucasian, Hispanic, Other.

**weight** Patient weight (in 25-pound weight bands). [0, 25), [25, 50), [175, 200)

**admit\_type\_id** Identifier corresponding to the type of hospital admission. 1 = Emergency, 2 = Urgent, 3 = Elective, 4 = Not Available.

**metformin** Indicates whether upon admission, metformin was prescribed or there was a change in the dosage. Up = dosage was increased, Down = dosage was decreased, Steady = dosage did not change, No = drug was not prescribed.

**insulin** Indicates whether upon admission, insulin was prescribed or there was a change in the dosage. Up = dosage was increased, Down = dosage was decreased, Steady = dosage did not change, No = drug was not prescribed.

**readmitted** Indicates whether patient had been readmitted after an inpatient stay in the twelve months preceding the encounter. <30 = patient was readmitted in less than 30 days, >30 = patient was readmitted in more than 30 days, No = no record of readmission.

**num\_procs** Number of procedures performed in the twelve months preceding the encounter. Integer 0 - 6.

**num\_meds** Number of distinct medications administered in the twelve months preceding the encounter. Integer 1 - 67.

**num\_ip** Number of the inpatient visits of the patient in the twelve months preceding the encounter. Integer 0 -21.

**num\_diags** Number of diagnoses entered to the system in the twelve months preceding the encounter. Integer 1 - 16.

---

patient_num_labs	<i>Patient Number of Labs</i>
------------------	-------------------------------

---

**Description**

Data used on June 19, 2020 Exam PA

**Usage**

patient\_num\_labs

**Format**

data.frame, 10000 observations of 14 variables:

**age** Age of prospective customer. Integer from 17 to 98.

**job** Occupation category. Factor with 11 levels.

**marital** Marital status. Factor with levels divorced, married, and single

**housing** Indicates whether the prospect has a housing loan. Factor with levels no, yes.

**loan** Indicates whether the prospect has a consumer loan. Factor with levels no, yes.

**phone** The type of phone the prospect uses. Factor with levels cellular, landline.

**month** The month of the marketing call. Factor with 12 levels.

**weekday** The day of the week of the marketing call. Factor with five levels.

**CPI** Consumer price index at the time of the call. Numeric from 92.20 to 94.77.

**CCI** Consumer confidence index at the time of the call. Numeric from -50.8 to -26.9.

**irate** Short term interest rate at the tie of the call. Numeric from 0.634 to 5.045.

**employment** Number of employees of ABC Insurance at the time of the call. Numeric from 4964 to 5228.

**purchase** Indicator of purchase. Integer (1 for purchase, or 0 for no purchase.)

**edu\_years** Years of education. Integer from 1 to 16.

---

pedestrian_activity	<i>Pedestrian activity</i>
---------------------	----------------------------

---

**Description**

pedestrian activity dataset

**Usage**

pedestrian\_activity

**Format**

data.frame, 11373 observations of 7 variables:

**pedestrians** The count of pedestrians during one hour starting at the indicated time.

**weather** Hourly weather condition, eleven categories.

**temperature** Hourly temperature in degrees Fahrenheit.

**precipitation** Hourly precipitation in inches.

**hour** Time at beginning of the measuring hour.

**weekday** Day of the week.

**temp\_forecast** Predicted daily average temperature in degrees Fahrenheit.

---

readmission

*Readmission*

---

**Description**

SOA Hospital Readmissions Sample Exam, 2019.

**Usage**

readmission

**Format**

data.frame, 66782 observations of 9 variables:

**Readmission.Status** The target variable, it is 1 for patients that were readmitted, 0 otherwise.

**Gender** M indicates male, F indicates female.

**Race** There are four categories: Black, Hispanic, Others, White.

**ER** The number of emergency room visits prior to the hospital stay associated with the readmission, an integer.

**DRG.Class** Diagnostic Related Group classification. There are three categories: MED (for medical), SURG (for surgical), UNGROUP.

**LOS** Length of hospital stay in days, an integer.

**Age** The patient's age in years, an integer. (Note that while most Medicare recipients are age 65 or older there are circumstances in which those under 65 can receive benefits.)

**HCC.Riskscore** Hierarchical Condition Category risk score. It is designed to be an estimate of a patient's condition and prospective costs. It is a continuous variable, rounded to three decimal places. Higher numbers indicate greater risk.

**DRG.Complication** Complications, with five levels: MedicalMCC.CC, MecalNoc, Other, SurgMCC.CC, SurgNoC, MCC.CC complications or comorbidities that may be major. NoC means no complications or comorbidities.

---

student_success	<i>Student Success</i>
-----------------	------------------------

---

### Description

SOA Student Success PA Sample Project, 2019.

### Usage

student\_success

### Format

data.frame, 585 observations of 33 variables:

**school** student's school (binary: GP (Grand Pines) or MHS (Marble Hill School)).

**sex** student's sex (binary: female or male).

**age** student's age (numeric: from 15 to 22).

**address** student's home address type (binary: U (Urban) or R (Rural)).

**famsize** family size (binary: GT3 (>3) or LE3 (<3)).

**Pstatus** parent's status (binary: A (Apart) or T (Together)).

**Medu** mother's education (numeric from 0 - 4. 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education (high school), or 4 - higher education (college)).

**Fedu** father's education. (numeric from 0 - 4. 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education (high school), or 4 - higher education (college)).

**Mjob** mother's job (nominal, teacher, health (health care related), services (civil services, administrative or police), at\_home, or other)

**Fjob** father's job (nominal, teacher, health (health care related), services (civil services, administrative or police), at\_home, or other)

**reason** reason to choose school (nominal: home (close to home), reputation (school reputation), course (course preference), other).

**guardian** student's guardian (nominal: mother, father, or other).

**traveltime** home to school travel time (numeric: 1 - < 15 min, 2 - 15 to 30 min, 3 - 30 min to 1 hour, or 4 - > 1 hour).

**studytime** weekly study time (numeric: 1 - < 2 hour, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - > 10 hours).

**failures** number of past class failures (numeric: n if  $0 \leq n < 3$ , else 3).

**schoolsup** extra educational support (binary: yes or no).

**famsup** extra family supplement (binary: yes or no).

**paid** extra paid classes (binary: yes or no).

**activities** extra-curricular activities (binary: yes or no).

**nursery** attended nursery school (binary: yes or no).  
**higher** wants to take higher education (binary: yes or no).  
**internet** internet access at home (binary: yes or no).  
**romantic** has a romantic relationship (binary: yes or no).  
**famrel** quality of family relationships (numeric: from 1 - very bad to 5 - very excellent).  
**freetime** free time after school (numeric: from 1 - very low to 5 - very high).  
**goout** going out with friends (numeric: from 1 - very low to 5 - very high)  
**Dalc** weekday alcohol consumption (numeric: from 1 - very low to 5 - very high).  
**Walc** weekend alcohol consumption (numeric: from 1 - very low to 5 - very high).  
**health** current health status (numeric: from 1 - very bad to 5 - very good).  
**absences** number of school absences (numeric: from 0 to 75).  
**G1** first trimester grade (numeric: from 0 to 20).  
**G2** second trimester grade (numeric: from 0 to 20).  
**G3** third trimester grade (numeric: from 0 to 20).

---

travel_insurance	<i>Travel insurance data</i>
------------------	------------------------------

---

### Description

The travel insurance dataset.

### Usage

travel\_insurance

### Format

data.frame, 10000 observations of 7 variables:

**Distance** Distance traveled in trip, in km

**Duration** Number of nights spent on trip

**Reason** Main reason for the trip. Vacation includes holiday, leisure, or recreation. Visit includes visiting friends or relatives.

**Age** Age of adult survey respondent in six age bins. 1: 19-24, 2: 25-34, 3: 35:44, 4: 45:54, 5: 55-64, 6: 65+

**Others** Number of other persons that accompanied the respondent on the trip

**Mode** Main mode of transportation, car or plane

**Cost** Total spending on trip, in Canadian \$



---

travel_spending	<i>Travel spending data</i>
-----------------	-----------------------------

---

**Description**

The travel spending dataset.

**Usage**

travel\_spending

**Format**

data.frame, 4884 observations of 11 variables:

**Q** Calendar quarter of trip

**ProvO** Trip province of origin

**Distance** Distance traveled in trip, in km

**Duration** Number of nights spent on trip

**Reason** Main reason for the trip. Vacation includes holiday, leisure, or recreation. Visit includes visiting friends or relatives.

**Age** Age of adult survey respondent in six age bins

**Gender** Gender of adult survey respondent

**HHI** Household income, in Canadian \$

**Others** Number of other persons that accompanied the respondent on the trip

**Mode** Main mode of transportation, car or plane

**Cost** Total spending on trip, in Canadian \$

# Index

## \* datasets

- actuary\_salaries, 2
  - apartment\_apps, 3
  - auto\_claim, 4
  - bank\_loans, 5
  - bike\_sharing\_demand, 6
  - boston, 7
  - customer\_phone\_calls, 8
  - customer\_value, 9
  - exam\_pa\_titanic, 9
  - health\_insurance, 10
  - june\_pa, 11
  - patient\_length\_of\_stay, 12
  - patient\_num\_labs, 13
  - pedestrian\_activity, 13
  - readmission, 14
  - student\_success, 15
  - travel\_insurance, 16
  - travel\_spending, 17
- 
- actuary\_salaries, 2
  - apartment\_apps, 3
  - auto\_claim, 4
- 
- bank\_loans, 5
  - bike\_sharing\_demand, 6
  - boston, 7
- 
- customer\_phone\_calls, 8
  - customer\_value, 9
- 
- exam\_pa\_titanic, 9
- 
- health\_insurance, 10
- 
- june\_pa, 11
- 
- patient\_length\_of\_stay, 12
  - patient\_num\_labs, 13
  - pedestrian\_activity, 13