

MS Azure ML Designer를 활용한 개인 실습

Lab : 의료비 데이터

강명호

배경 및 모델링 목표

배경 및 목표

Medical Cost Personal Datasets 를 분석하여 개인 의료비를 예측할 수 있는 회귀 모델 구현

- 의료 보험 상품을 개발하기 위해 보험사는 과거의 데이터를 이용하며, 이를 통해 정확한 보험 모델의 개발이 가능함



배경 및 목표

Medical Cost Personal Datasets 를 분석하여 개인 의료비를 예측할 수 있는 회귀 모델 구현



- Brett Lantz가 쓴 Machine Learning with R에서 데이터 발췌하여 가공한 데이터세트 사용
- 특성 :
 - age, sex, bmi,
 - children, smoker,
 - region, charges
- age, bmi, children 수 등의 특성에 따른 진료비(charges)를 예측하는 회귀 문제

데이터 수집

데이터 수집

Kaggle에 접속하여 Medical Cost Personal 데이터 세트 검색

The screenshot shows the Kaggle homepage. On the left, there is a sidebar with various navigation options: Create, Home, Competitions, Datasets, Models, Code, Discussions, Learn, More, Your Work, and VIEWED. The VIEWED section contains two entries: 'Medical Cost Perso...' and 'Medical Cost Perso...'. At the top center, there is a search bar with the placeholder 'Search'. To the right of the search bar, it says 'LOGIN STREAK' with a value of '4 days a new record!'. Below the search bar, the main content area features a large 'Welcome, Myeongho Kang!' message. Underneath it, a sub-message says 'You're on a roll! Jump back in, or start something new.' There are four summary cards: 'Datasets' (0 total created), 'Notebooks' (0 total created), 'Competitions' (0 total joined), and 'Discussions' (0 total posted). At the bottom, there is a section titled 'How to start: Choose a focus for today' with the sub-instruction 'Help us make relevant suggestions for you'.

데이터 수집

Kaggle에 접속하여 Medical Cost Personal 데이터 세트 검색

← medical cost personal

Notebooks 444 Comments 285 Topics 27 Datasets 14

Filter by 770 Results

DATE

Last 90 days 37

VIEWED BY YOU

Viewed 3

Not Viewed 767

CREATOR

You 0

Medical Cost Personal Datasets
Notebook · 3y ago · by PAVAN KUMAR D
Data source: [Medical Cost Personal Datasets](#)

Medical Cost Personal Cost Prediction
Notebook · 4y ago · by Mohaiminul Islam
Data source: [Medical Cost Personal Datasets](#)

Medical Cost Personal
Notebook · 3y ago · by mohamed_zaytoon
Data source: [Medical Cost Personal Datasets](#)

데이터 수집

Kaggle에 접속하여 Medical Cost Personal 데이터 세트 검색

← medical cost personal

Notebooks 444

Comments 285

Topics 27

Datasets 14 X

Filter by

DATE

Last 90 days

14 Results

Relevance ▾

VIEWED BY YOU

Viewed

1



Medical Cost Personal Datasets

Dataset · 6y ago · by Miri Choi

Insurance Forecast by using Linear Regression

2651

254,391 downloads

Not Viewed

13



Medical Cost Personal Datasets

Dataset · 3y ago · by PAVAN KUMAR D

Medical Cost Personal Datasets

24

893 downloads

CREATOR

You

0



Medical Cost Personal Dataset

Dataset · 4y ago · by Abdel Homi

This dataset is from Machine Learning with R by Brett Lantz.

8

359 downloads

데이터 수집

Kaggle에 접속하여 Medical Cost Personal 데이터 세트 검색

MIRI CHOI · UPDATED 6 YEARS AGO

▲ 2651 New Notebook

Download (16 kB)

Medical Cost Personal Datasets

Insurance Forecast by using Linear Regression

Data Card Code (1489) Discussion (16) Suggestions (0)

About Dataset

Context

Machine Learning with R by Brett Lantz is a book that provides an introduction to machine learning using R. As far as I can tell, Packt Publishing does not make its datasets available online unless you buy the book and create a user account which can be a problem if you are checking the book out from the library or borrowing the book from a friend. All of these datasets are in the public domain but simply needed some cleaning up and recoding to match the format in the book.

Usability

8.82

License

Database: Open Database, Cont.

Expected update frequency

Not specified

Tags

데이터 수집

Kaggle - Medical Cost Personal

About Dataset

Context

Machine Learning with R by Brett Lantz is a book that provides an introduction to machine learning using R. As far as I can tell, Packt Publishing does not make its datasets available online unless you buy the book and create a user account which can be a problem if you are checking the book out from the library or borrowing the book from a friend. All of these datasets are in the public domain but simply needed some cleaning up and recoding to match the format in the book.

Content

Columns

- age: age of primary beneficiary
- sex: insurance contractor gender, female, male

Usability ⓘ

8.82

License

Database: Open Database, Cont...

Expected update frequency

Not specified

Tags

Education

Health

Finance

Insurance

Healthcare

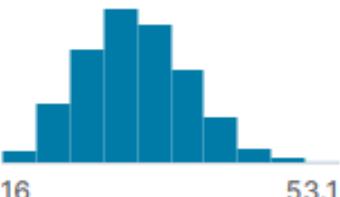
Content

Columns

- age: age of primary beneficiary
- sex: insurance contractor gender, female, male
- bmi: Body mass index, providing an understanding of body, weights that are relatively high or low relative to height, objective index of body weight (kg / m^2) using the ratio of height to weight, ideally 18.5 to 24.9
- children: Number of children covered by health insurance / Number of dependents
- smoker: Smoking
- region: the beneficiary's residential area in the US, northeast, southeast, southwest, northwest.
- charges: Individual medical costs billed by health insurance

데이터 수집

Kaggle - Medical Cost Personal

About this file						 Add Suggestion
This dataset consists of 1338 rows.						
# age	sex	# bmi	# children	smoker	region	
Edad del asegurado	Género	Índice de masa corporal	Número de hijos	Indicador si fuma	Región del asegurado	
	male female	51% 49%			true false	southeast southwest Other (6)
18	female	27.9	0	yes	southwest	
18	male	33.77	1	no	southeast	
28	male	33	3	no	southeast	
33	male	22.705	0	no	northwest	
32	male	28.88	0	no	northwest	
31	female	25.74	0	no	southeast	

데이터 수집

Kaggle - Medical Cost Personal

