



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

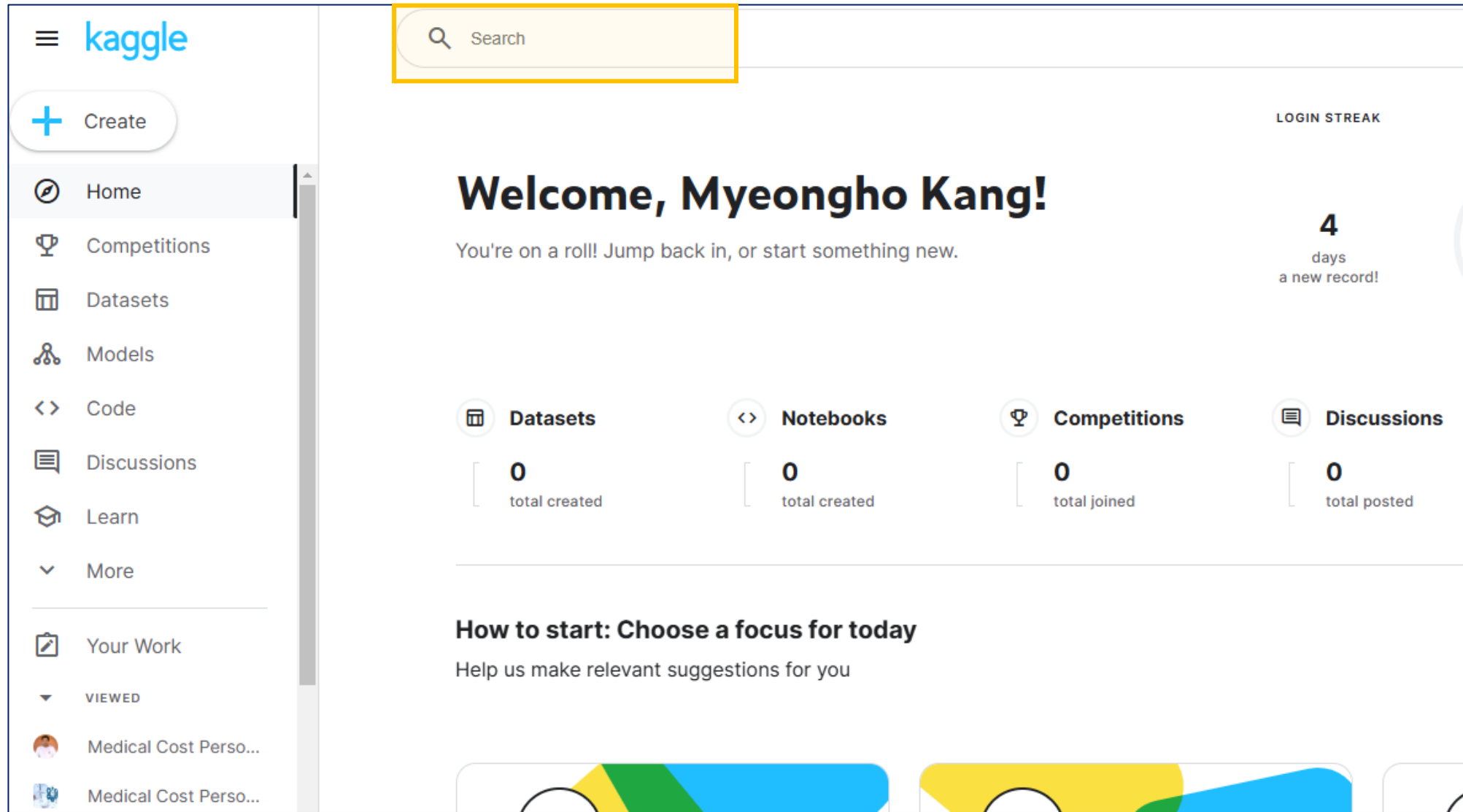


- Brett Lantz가 쓴 **Machine Learning with R**에서 데이터 발췌하여 가공한 데이터세트 사용
- 특성 :
 - age, sex, bmi,
 - children, smoker,
 - region, charges
- age, bmi, children 수 등의 특성에 따른 진료비(charges)를 예측하는 회귀 문제

데이터 수집

데이터 수집

Kaggle에 접속하여 Medical Cost Personal 데이터 세트 검색



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

✕

Filter by

14 Results

Relevance ▾

DATE

☐ Last 90 days

VIEWED BY YOU


☐ Viewed

☐ Not Viewed

CREATOR

☐ You

1



Medical Cost Personal Datasets


Dataset · 6y ago · by [Miri Choi](#)

Insurance Forecast by using Linear Regression

2651

254,391 downloads

1



Medical Cost Personal Datasets


Dataset · 3y ago · by [PAVAN KUMAR D](#)

Medical Cost Personal Datasets

24

893 downloads

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

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



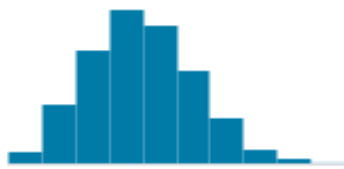

Content

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 Edad del asegurado	# sex Género	# bmi Indice de masa corporal	# children Número de hijos	✓ smoker Indicador si fuma	A region Región de residencia del asegurado	
 1864	male 51% female 49%	 1653.1	 05	true 0 0% false 0 0%	southeast southwest Other (6)	
19	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

Add Suggestion

	✓ smoker Indicador si fuma	Δ region Región donde vive el asegurado	# charges Prima del seguro
5	<div><div>true</div><div>0 0%</div><div>false</div><div>0 0%</div></div>	<div><div>southeast</div><div>27%</div><div>southwest</div><div>24%</div><div>Other (649)</div><div>49%</div></div>	<div><div>1.12k63.8k</div></div>
	yes	southwest	16884.924
	no	southeast	1725.5523
	no	southeast	4449.462
	no	northwest	21984.47061
	no	northwest	3866.8552
	no	southeast	3756.6216