

Can Physical Characteristics Predict Health Insurance Costs?

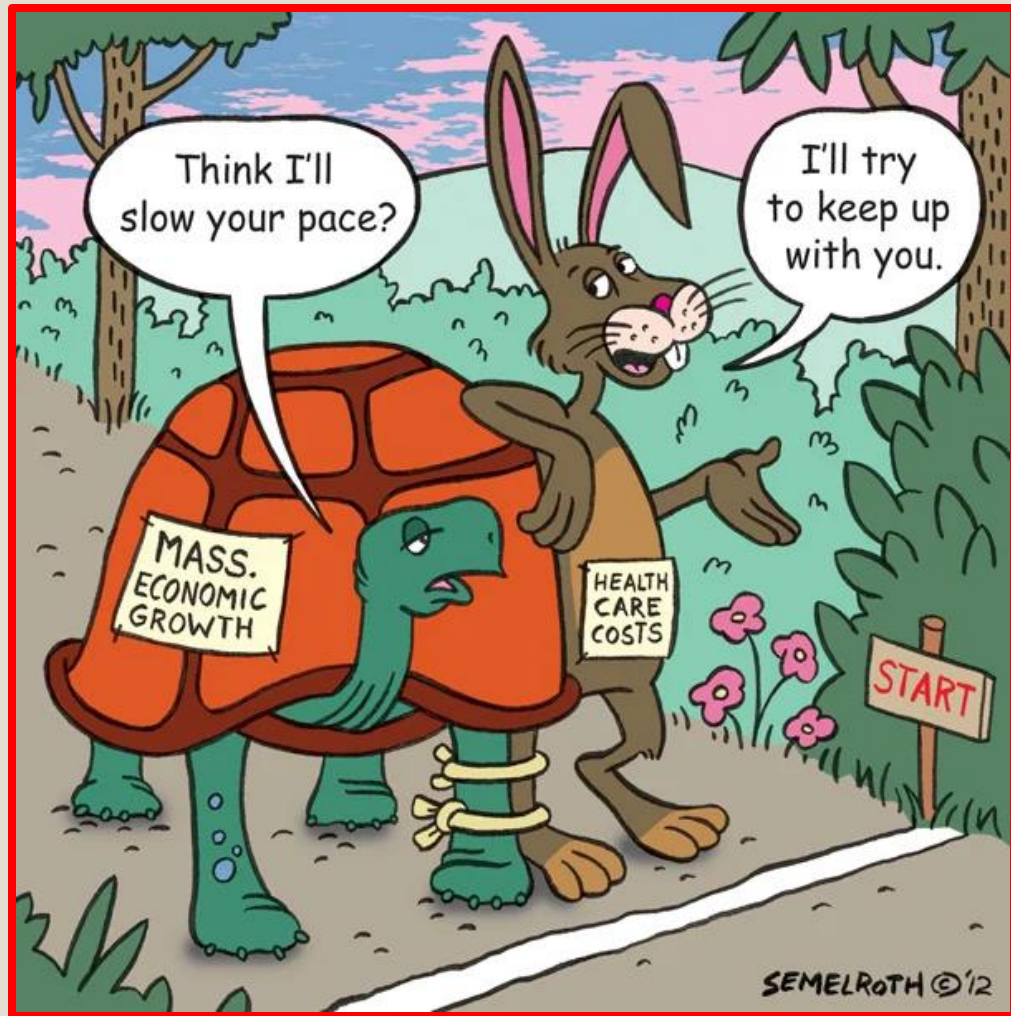
A Predictive Modeling Approach Using Machine

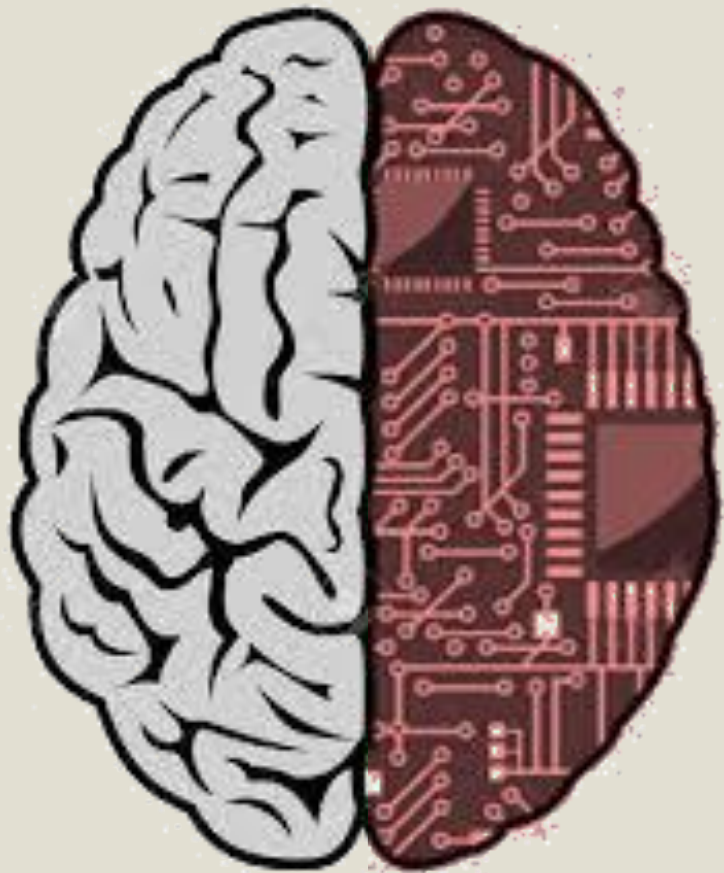


The image shows a medical invoice with a table of services and costs. The invoice is titled 'MEDICAL INVOICE' and includes a section for 'ACCOUNT SUMMARY' and 'PATIENT SERVICES PROVIDED'. The table lists various medical services and their corresponding costs. A calculator is visible in the background, suggesting the financial nature of the data.

DESCRIPTION	AMOUNT
Office Visit	125.00
Lab Work	225.00
X-Rays / Abdominal	350.00
Surgery	7,500.00
Anesthesia	1,000.00
Pathology	531.00
Medical/Surgical Supplies	357.00
Post-Op Care	482.00
TOTAL	\$10,570.00
CLAIM	0.00
NET PAYABLE	\$10,570.00

Spending in
the U.S. was
\$4.8
trillion..





**Our strategy
depends on
accurately
estimating
cost risk**

Who would Benefit



Insurance Analysts & Actuaries



Healthcare Policy Designers



Patients & Policyholders

The Dataset

kaggle

Medical Cost Personal Datasets

Insurance Forecast by using Linear Regression

Data Card

Code (1705)

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.

Content

Columns

age

sex

bmi

children

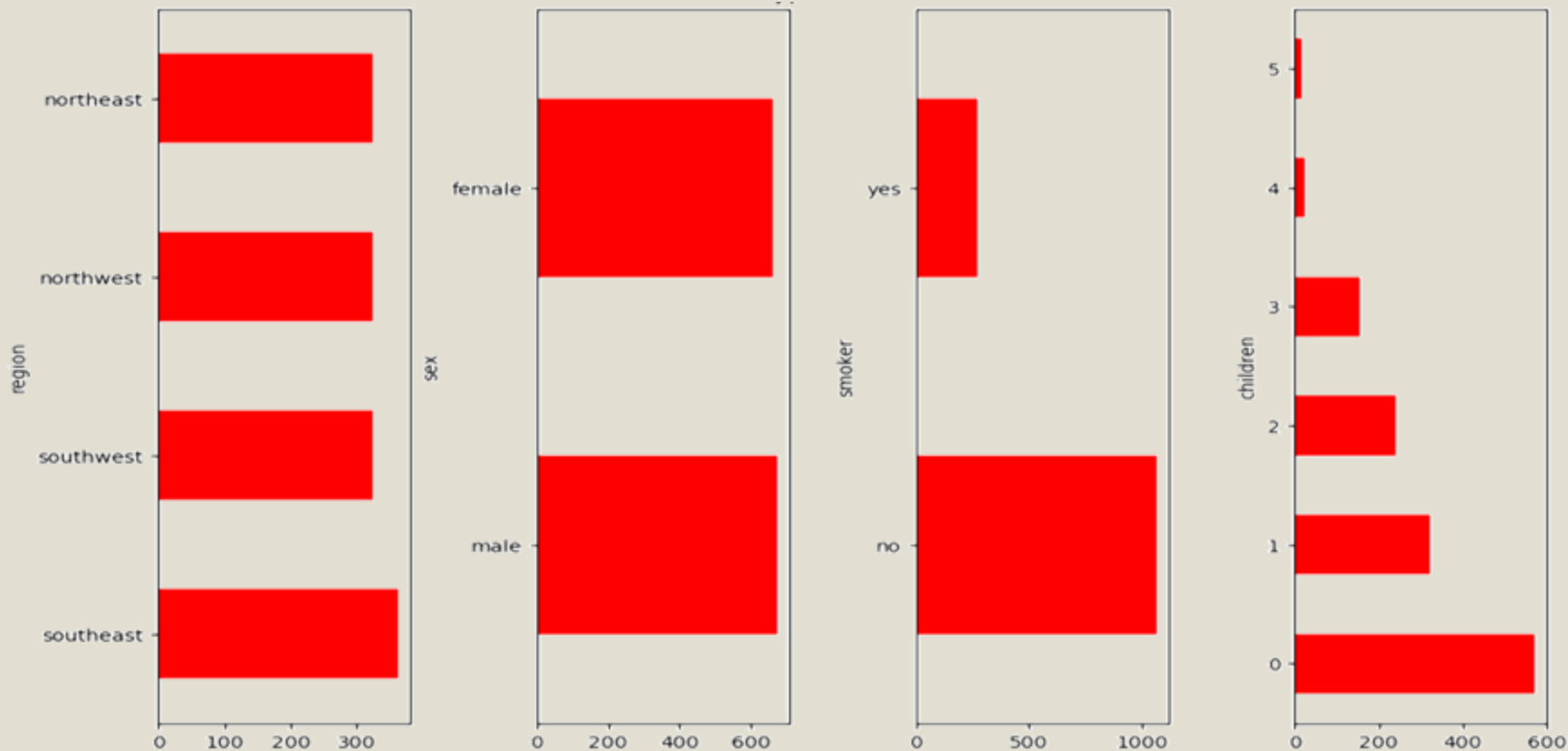
smoker

region

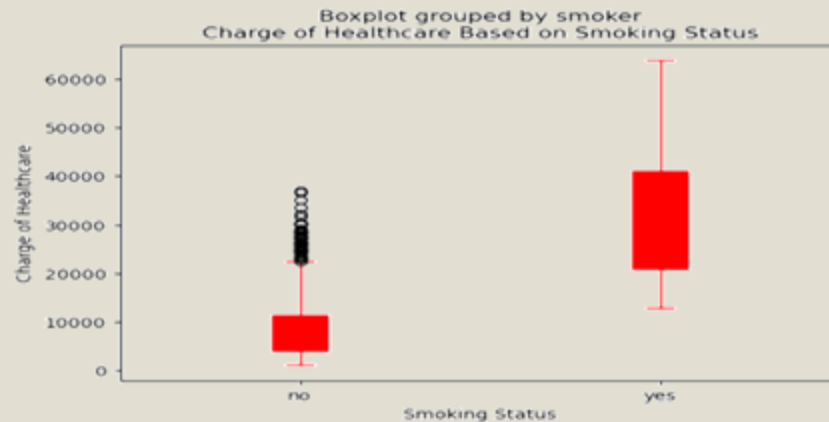
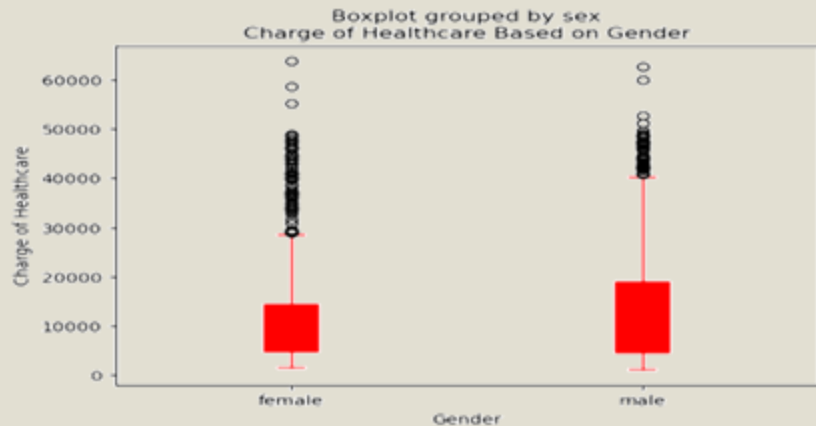
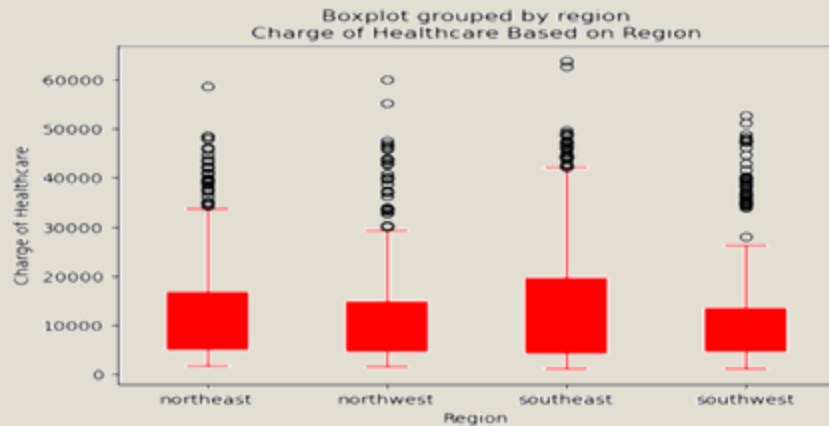
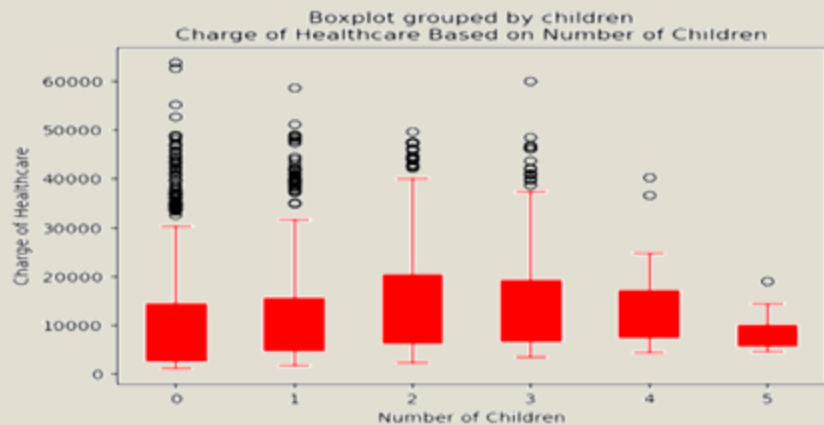
charges

Initial Observations

The Number of Individuals in Different Groups



Group Distribution & Cost Spread



Models Tested

Linear Regression

Ridge Regression

Lasso Regression

K-nearest
Neighbors

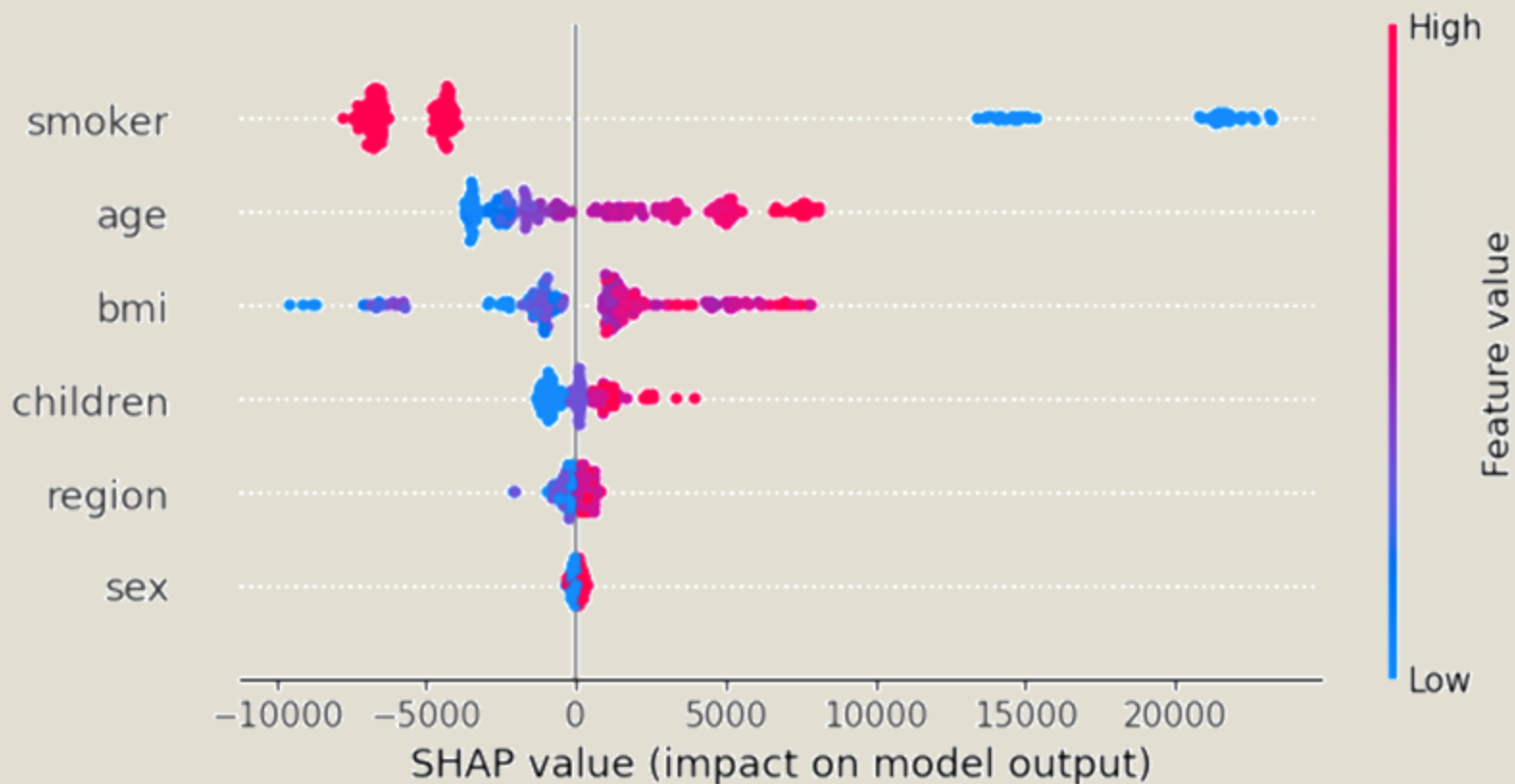
Random Forest
Regression

Gradient Boosting

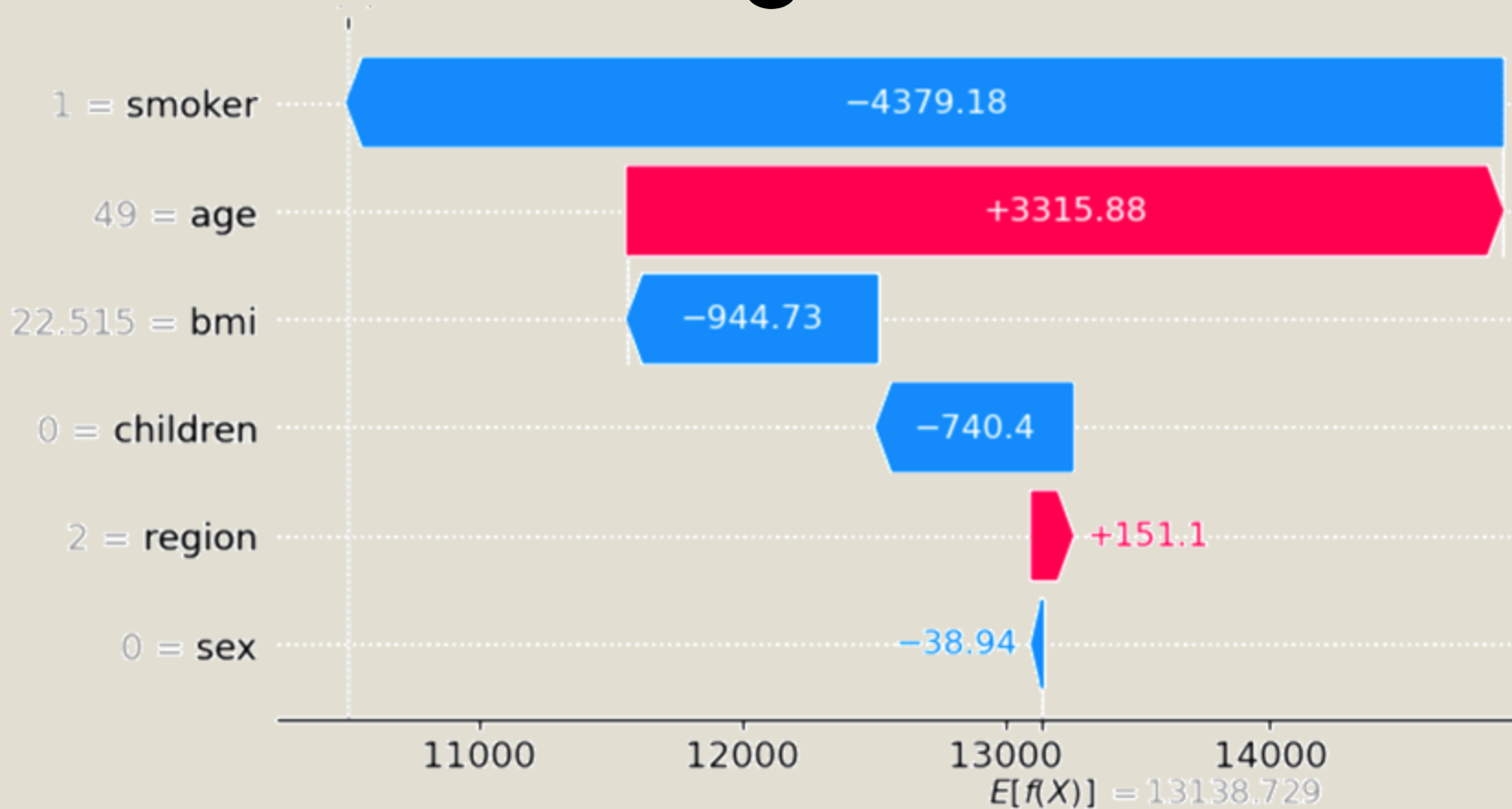
Top Performer: Gradient Boosting
explains roughly
89.4% of the
variability in
insurance charges



Which Drives Insurance Cost?



Predicting the Cost



Recommendations based on the Model

**Now We can Predict
Premiums to allow
more accurately
estimate
insurance
costs.**

Questions?

