The background features a collection of abstract geometric shapes. In the top left, there is a small white square with a black outline, partially overlapping a larger solid orange square. On the right side, a tall, thin black vertical bar is positioned next to a wider orange vertical bar. A thin orange line curves from the top right towards the center. In the bottom left, there are several horizontal orange lines of varying lengths, a black circle partially enclosed by a teal circle, and a solid orange rectangle. A thick teal arc curves from the bottom right towards the center. Another orange square with a white outline is located on the right side, partially overlapping the teal arc.

# Healthcare Cost Calculator

*Say Goodbye to Insurance  
Rip-offs*

# Motivation

- Insurers invest time and money to develop models that forecast medical expenses.
- Health insurance buyers should do the same before picking an insurance





# Data Overview

01

## Overview

Insurance.csv (54KB)  
Kaggle open dataset  
1338 observations

02

## Target

Annual medical  
expenses charged  
in USD (*numerical*)

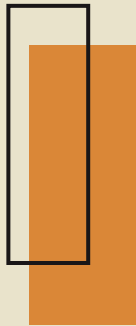
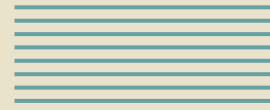
## Predictors

- Age (*numerical*)
- Sex (*categorical*)
- BMI (*numerical*)
- Number of Children (*numerical*)
- Smoker or Not (*categorical*)
- Region (*categorical*)

03

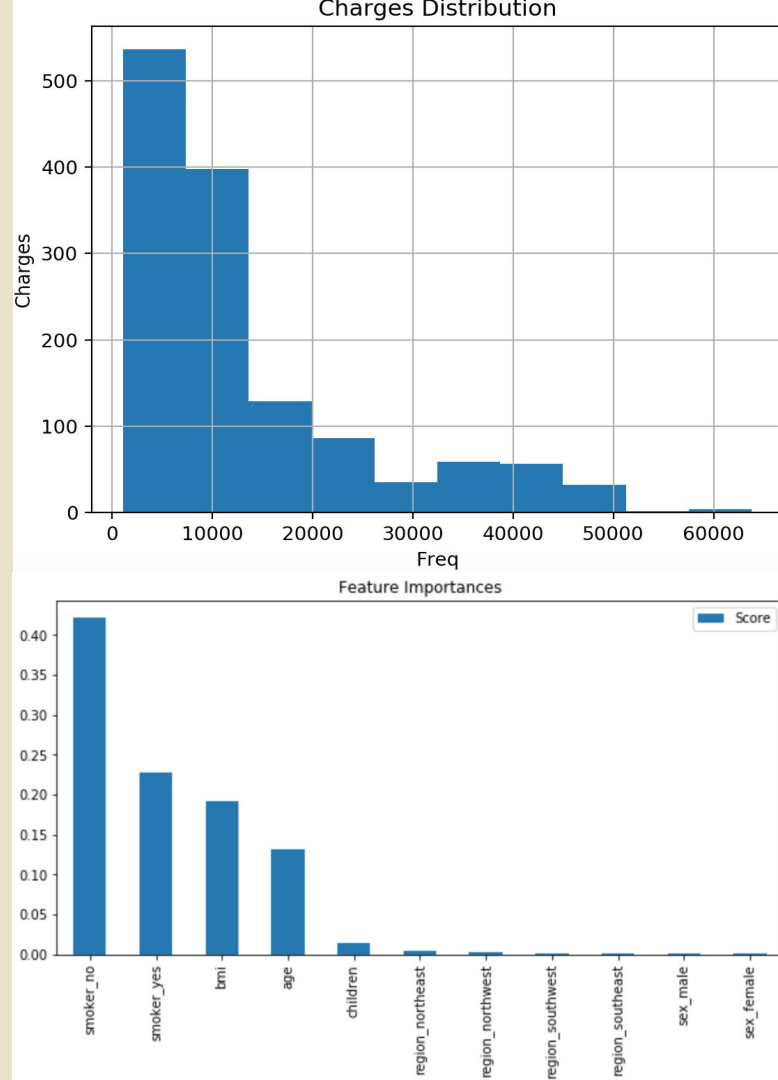


## Success Criteria

- Random Forest Regressor
  - *Success Criteria:  $R^2 > 0.85$*
  - Test  $R^2 = 0.87$
  - Split the data into training and test set. Evaluate model on the test set.
- 
- 

# Insights

1. Most of the time medical costs are below \$10,000 per year
2. Smoker or not is the most important indicator of medical costs. Followed by BMI



# THANKS!

Do you have any questions?

zhangyiye2020@u.northwestern.edu

<https://github.com/rocky-ye>

