



Health Insurance cost

predict insurance costs?

METHODOLOGE



INTRODUCTION



EDA



Feature engineering



**Linear regression
models**



CONCLUSION





01

INTRODUCTION

INTRODUCTION

Today we'll explore a Data Set dedicated to the price of treatment of various patients. the price of treatment depends on several factors: designation, form of clinic, town of residence, age then on. we've got no data on the diagnosis of patients.

However, we have alternative information which will facilitate U.S.A. to create a conclusion regarding the health of patients and observe regression analysis.



FEATURES



SEX

insurance contractor
gender, female, male



AGE

age of primary beneficiary



BMI

Body mass index , providing an
understanding of body.



CHILDREN

Number of children covered by
health insurance




SMOKER

Smoking , yes or No




CHARGES

Individual medical costs billed by
health insurance



REGION

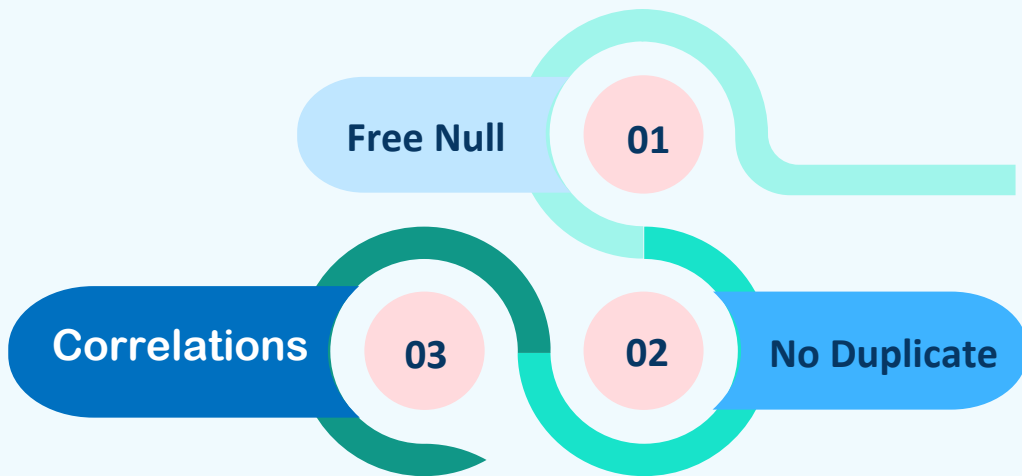
the beneficiary's residential area in the
US, south , north



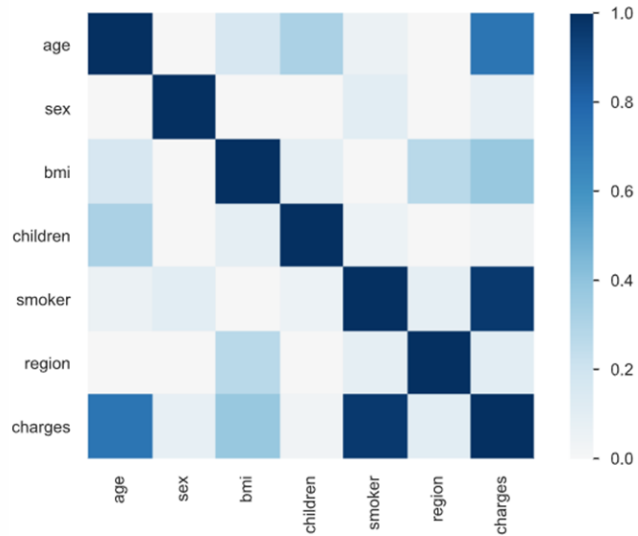


02

EDA



Correlations

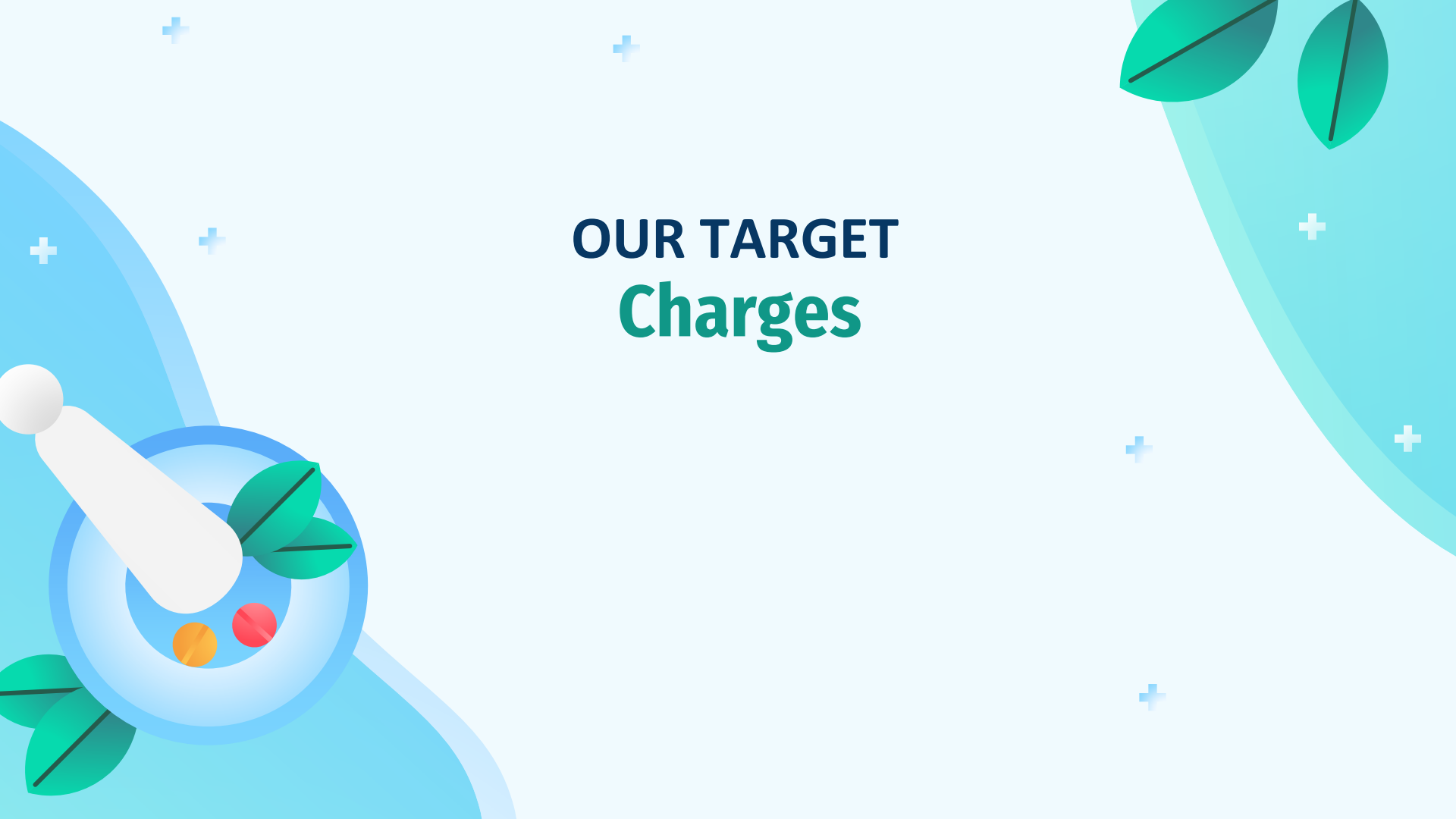


The background is a light blue gradient with various medical-themed icons. In the top center, a blue syringe is shown with a single drop of dark blue liquid falling into a light blue, irregularly shaped blob. The number '03' is printed in dark blue inside this blob. To the right, there are several pills: two red and blue capsules, one orange round pill, and one blue round pill. On the left side, there are two green leaves. Scattered throughout the background are several small, light blue plus signs. The overall design is clean and modern, with a focus on medical or scientific themes.

03

Feature engineering

OUR TARGET Charges

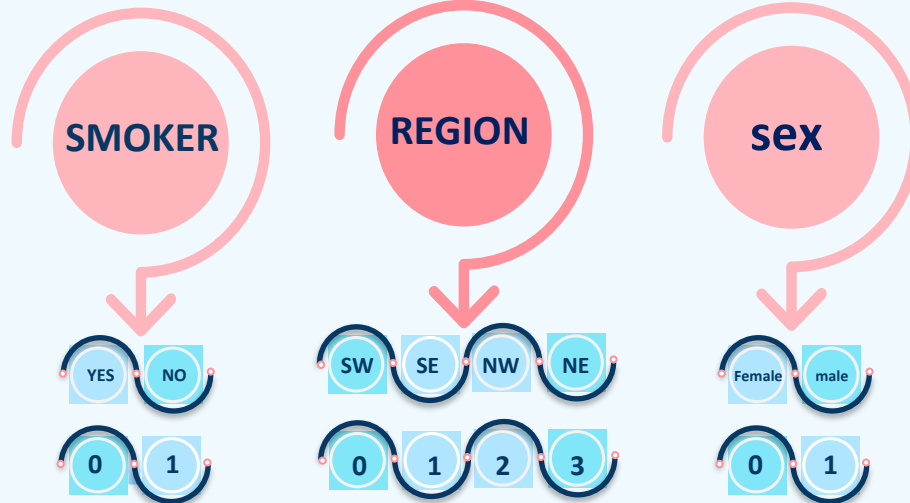


WHY Charges !!

Columns	Correlation
Region	- 0.0062
SEX	0.0572
Children	0.0679
BMI	0.1983
AGE	0.2990
Smoker	0.7872
Charges	1.0000



In our project we used Label Encoder which converts from non-numerical labels to numerical.





04

Linear regression models

MODELS SCORES

Linear Regression

Train : 0.749008
validation : 0.742403

Random Forest

Train : 0.974000
validation : 0.869000

Polynomial Features

Train : 0.843331
validation : 0.822541

Lasso

Train : 0.749008
validation : 0.742103



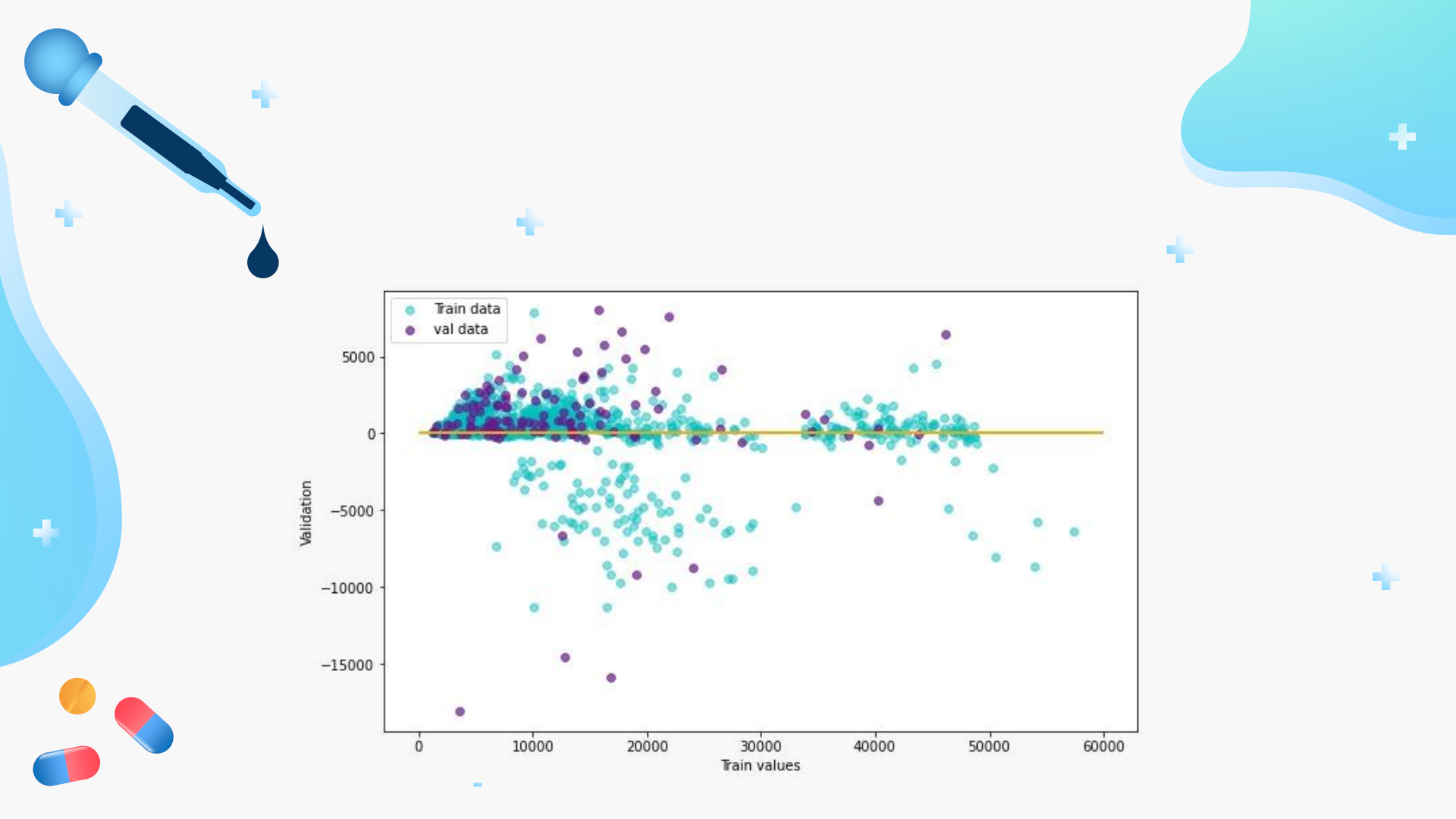
The background is a light blue gradient with various medical-themed icons. In the top center, a blue syringe is shown dripping a dark blue drop into a light blue, irregularly shaped blob. Inside this blob is the number '05'. To the right, there are several pills: two red and blue capsules, one orange round pill, and one blue round pill. On the left side, there are two green leaves. Scattered throughout the background are small white plus signs and light blue curved shapes.

05

CONCLUSION

CONCLUSION

- In the end, the Polynomial has been selected because it's the highest ratio experiment .
- i



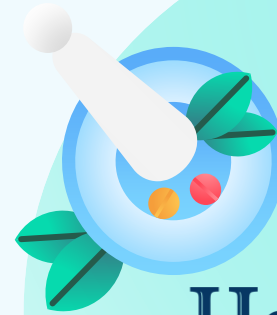


THANKS

For watching..

Presented by

Rawan Alshehri
Khalid Alharbi
Abdulrahman Qureshi



Health Insurance
+ cost

