Medical Student Health

LHL-Final-Project

Shaojun Zhang

2023-08-17

CONTENTS

01

Goals:

Goals:

- 1. To **predict** the existing data in training a predictive model, with an aim to supplement the diabetes section, which lacked around 20,000 entries.
- 2. To execute regression analysis on the supplemented data to identify **correlation** between health features and diabetes.

02

Steps Adopted:

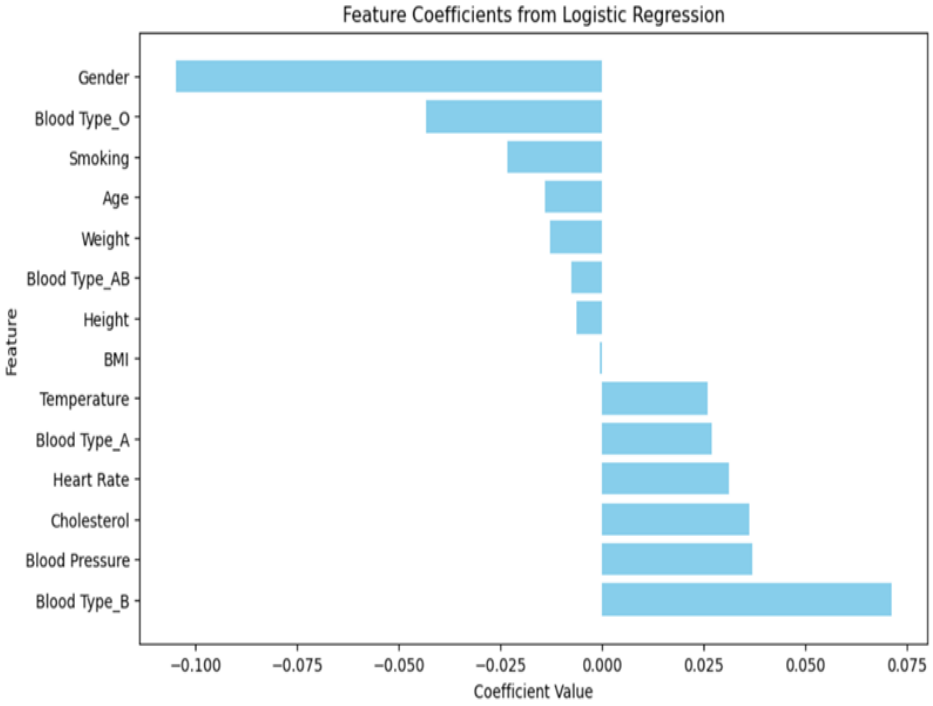
Steps Adopted:

1. **Data cleaning and EDA**
2. **Model training.**
3. **Regression Analysis.**
4. **Visualization**

03

Results:

Results:

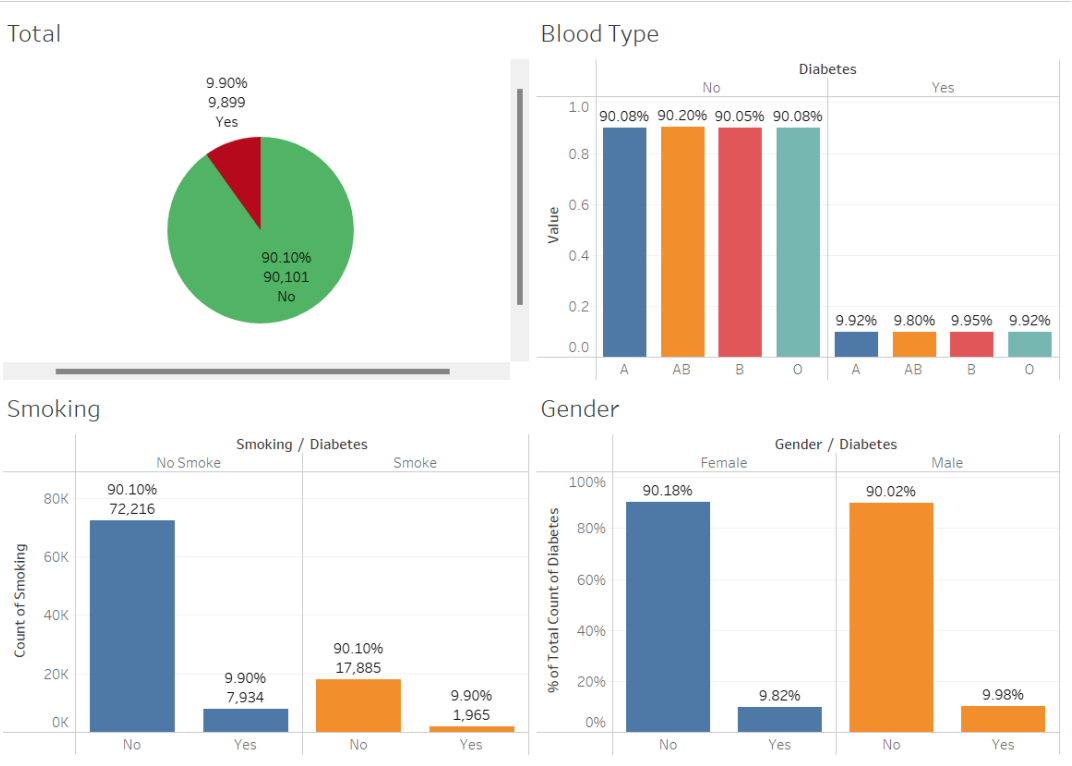


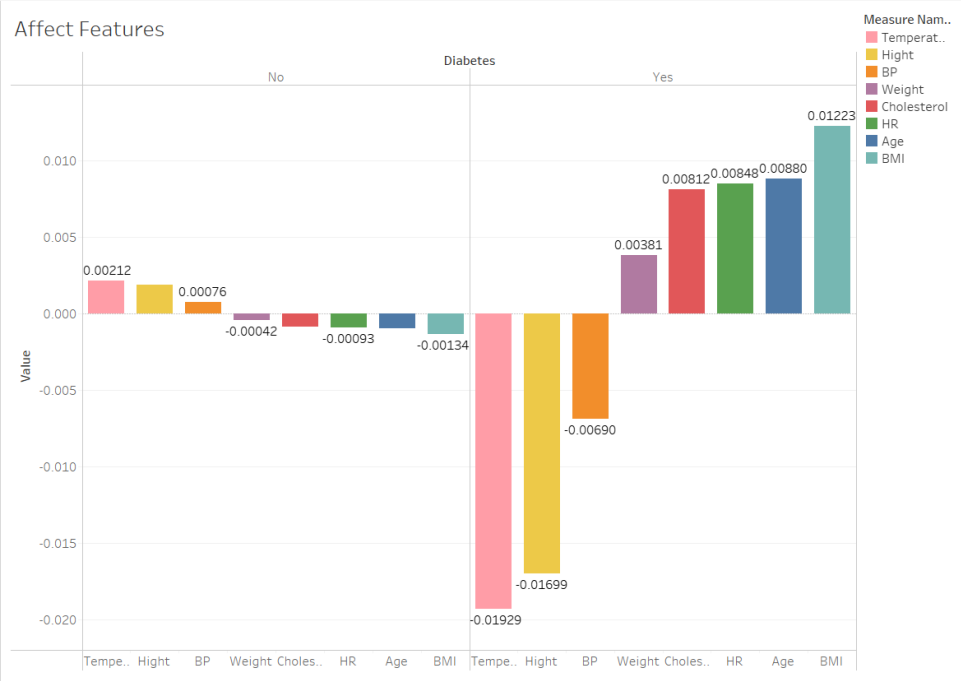
some notable observations were:

1. females are less prone to diabetes.
2. Blood type O showed a negative correlation with diabetes.
3. Non-smokers were relatively less prone to the disease.
4. Features with a positive correlation included Blood type B, Blood pressure, Cholesterol, Heart rate, Blood type A, and Body temperature.

The visualization from Tableau further depicted:

Smoking, gender, and blood type had a **minimal influence** on diabetes susceptibility.





- The data suggested individuals with **lower body temperature, height, and blood pressure** were more prone to diabetes.

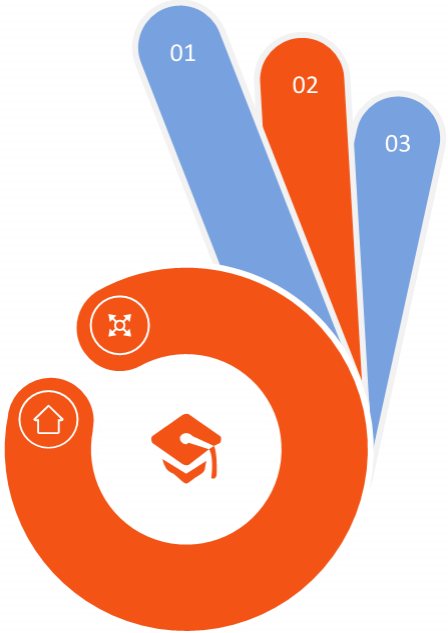
- Conversely, those with **higher weight, cholesterol, heart rate, age, and BMI** faced increased risks.

04

Challenges:

Challenges:

1. A discrepancy between the database's description and its actual content.



2. Difficult-to-detect duplicates in the data.

3. A meager correlation between data features and the predictive target.

4. Abundant missing data.

05

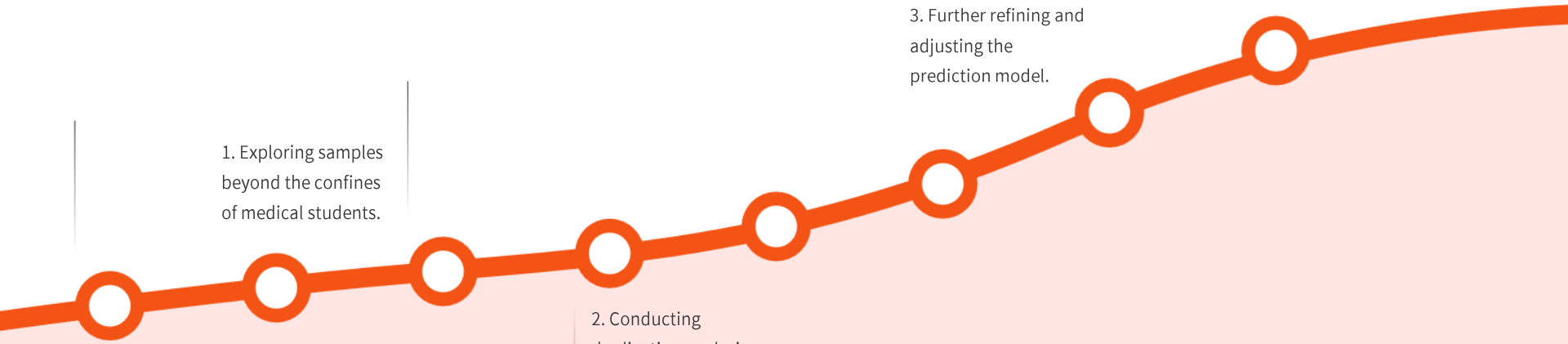
Looking Ahead:

Looking Ahead:

1. Exploring samples beyond the confines of medical students.

2. Conducting duplication analysis on columns that might serve as primary keys.

3. Further refining and adjusting the prediction model.



THE END

THANKS