Predicting Kidney Failure Risk: Analysis of NHANES Data

ALY6040: Data Mining, Northeastern University

By: Group 2

Yulin Liang

Mohit Ravindra Kamble

Aadira Anil Ramakrishnan

Monika Gundecha

Yash Gokhale

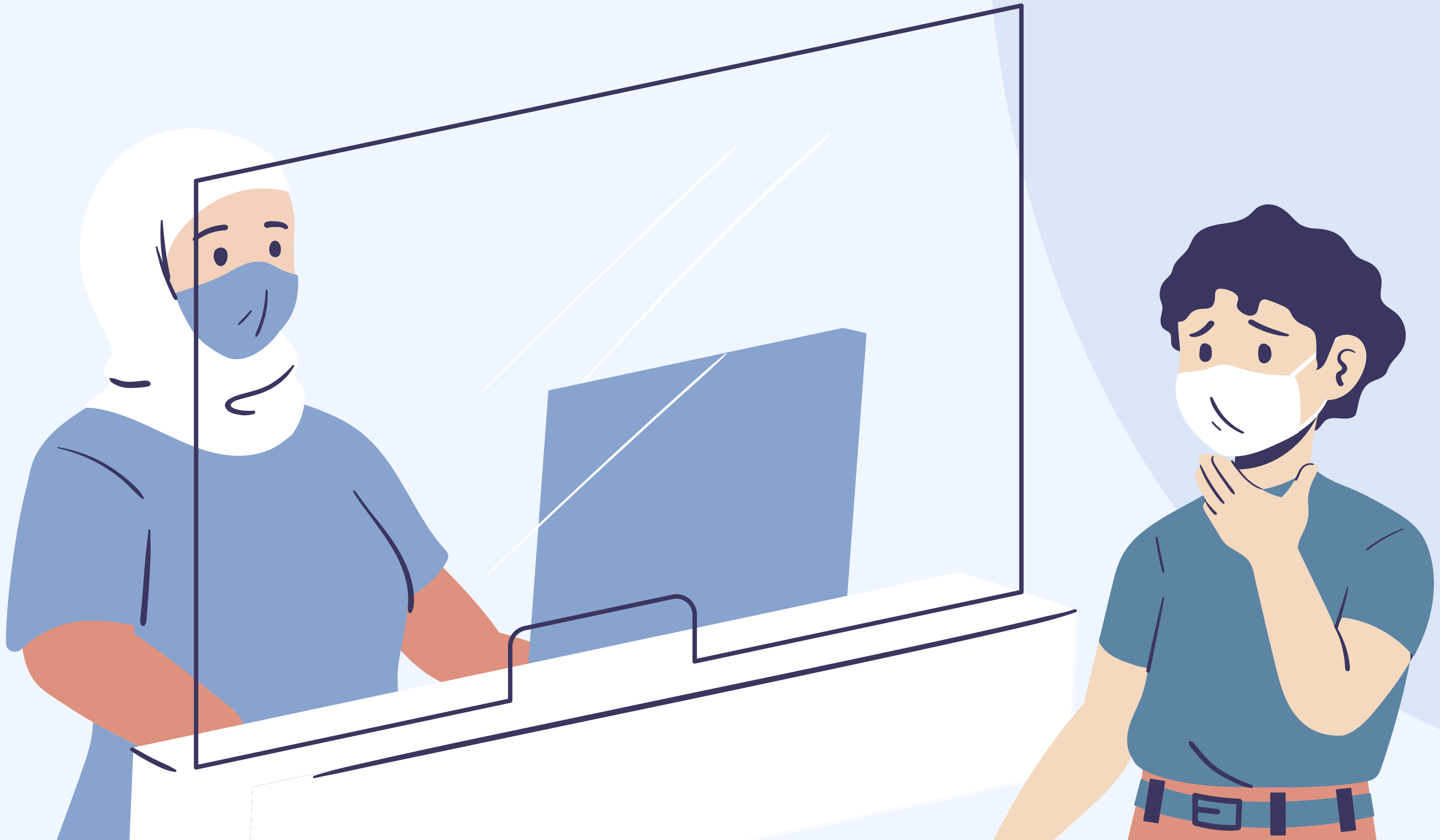
Professor Justin Grosz

June 24, 2024

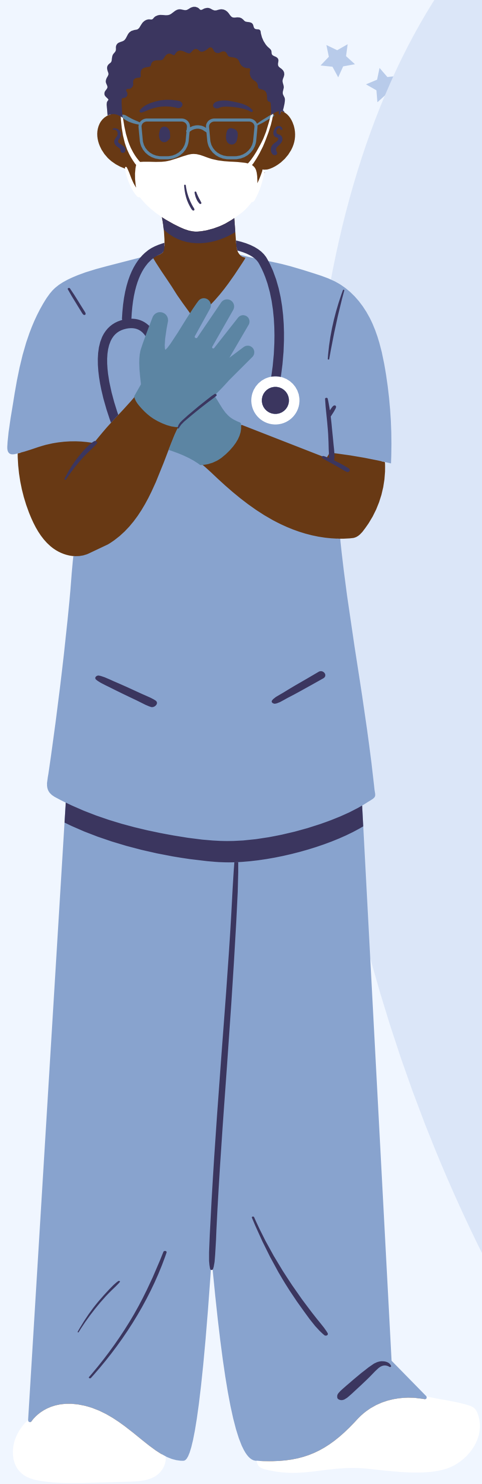
Agenda

Topics covered in this presentation

- Introduction
- Business Questions
- Methods
- Recommendations
- Conclusion
- References



Introduction



Our goal is to predict kidney failure risk using the National Health and Nutrition Examination Survey (NHANES) dataset.

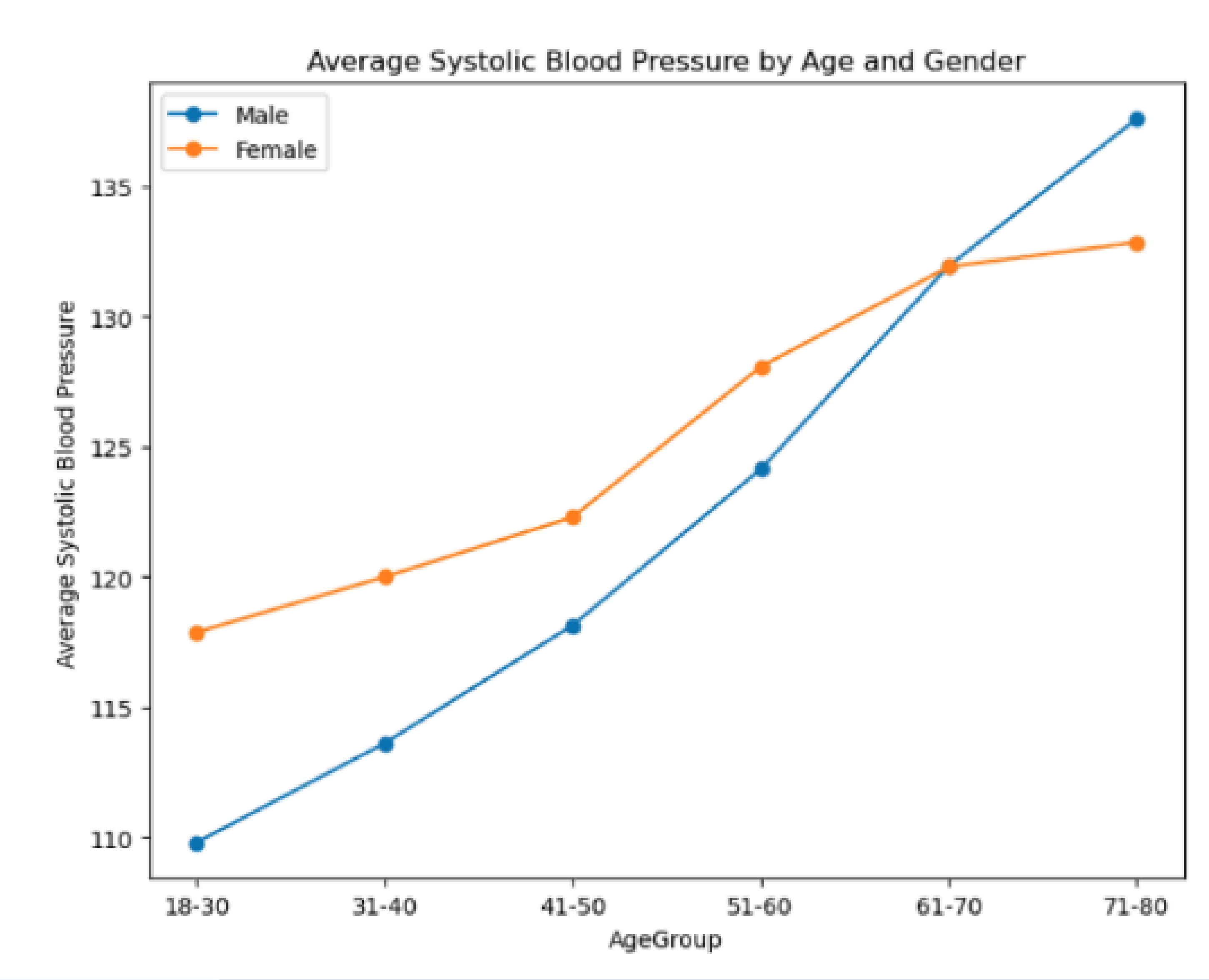
We analyzed relevant variables to develop predictive models for identifying at-risk individuals, aiding stakeholders like insurance companies in setting policies tailored to kidney failure risk.

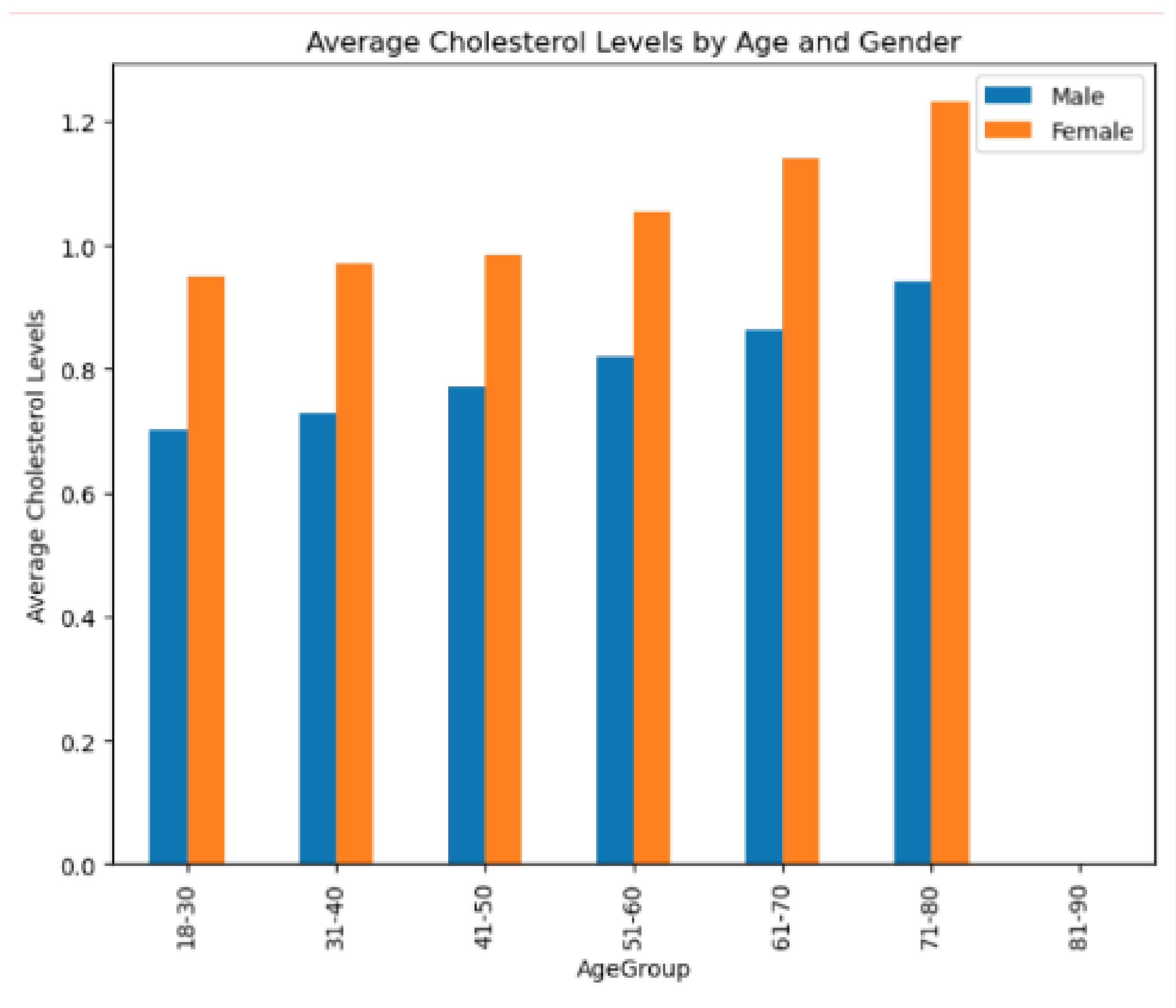
Business Question
Stakeholder: Insurance Company

How can we accurately identify individuals at high risk of kidney failure using available health data?

How can we optimize our insurance packages and pricing strategies based on these risk assessments?

Hypothesis





High blood pressure is a key risk factor for kidney disease, making it crucial in predicting kidney failure.

Gender differences in systolic blood pressure by age may impact kidney failure risk due to lifestyle, diet, or physiological factors.

Method

- Models Used:
  - Logistic Regression (Backward Selection Method)
  - Gradient Boosting
- Factors included:
  - Health Indicators: Drugs, Cholesterol, Blood Pressure, Age, Diabetes, Serum Creatinine, Albumin.
  - Diet: Fat and Protein Intake
- Focusing on the likelihood of Kidney Failure

Logistic Regression Model Results

Strengths:

- The model correctly predicted the kidney failure status (97%)
- Low rate of false alarms for kidney failure

Limitations:

- The model correctly identified only 19% of the actual kidney failure cases. This means that it missed 81% of the kidney failure cases, which is a significant risk.

Key point:

- This model is more cautious, flagging fewer people for kidney failure risk but with higher confidence

Gradient Boosting Model Results

Strengths:

- The model correctly predicted the kidney failure status (96%)

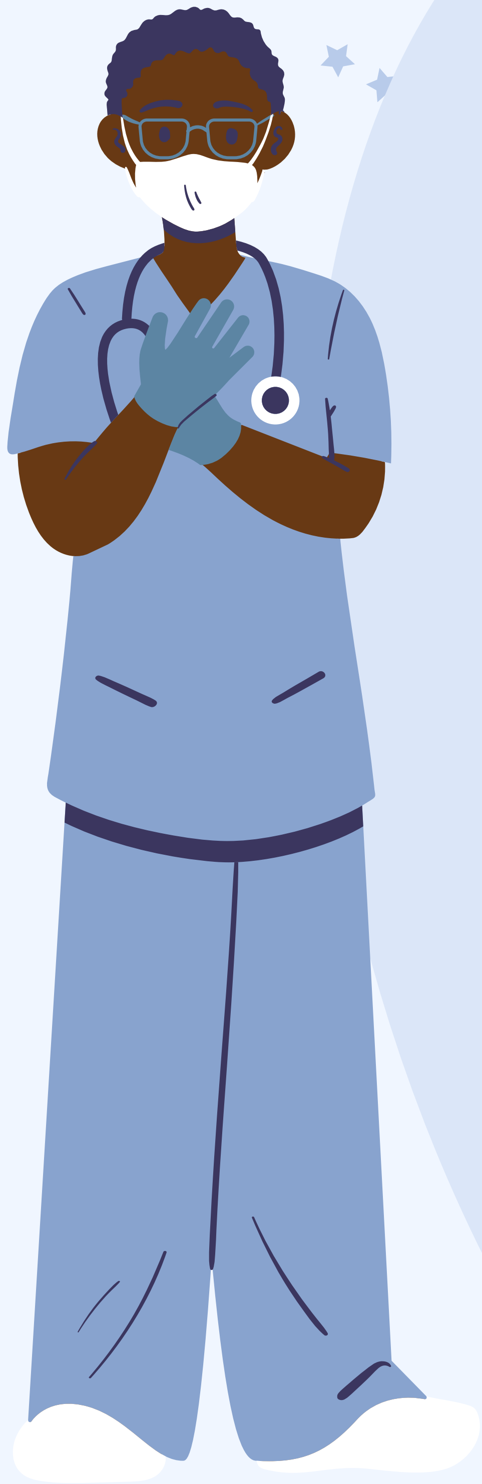
Limitations:

- The model correctly identified 21% of the actual kidney failure cases. This means that it missed 79% of the kidney failure cases, which is still a significant risk.

Key point:

- This model casts a wider net, potentially catching more kidney failure cases but with more false alarms

Key Metrics



- Serum creatinine, serum albumin, blood urea nitrogen, and creatinine levels are significant predictors.
- Diabetes diagnosis and long-term use of medications like prednisone or cortisone significantly increase kidney failure risk.
- Both diastolic and systolic blood pressure levels are crucial predictors.
- Dietary habits, particularly saturated fat intake, and medication usage were influential in kidney failure risks.

Recommendation

Incoporate tests like HB1AC, Urine Test and Blood Pressure

Make these tests mandatory for all policyholders to accurately assess kidney health risks and inform premium adjustments.

Record data in a better form and quantity

Better data gathering aids in analyzing trends, identifying risks, and making informed policy adjustments for improved risk management.

Prioritizing Gender, Age, and Test Results

Prioritizing gender, age, and test results helps to accurately assess kidney health risks, leading to tailored policy adjustments and improved risk management.

Conclusion

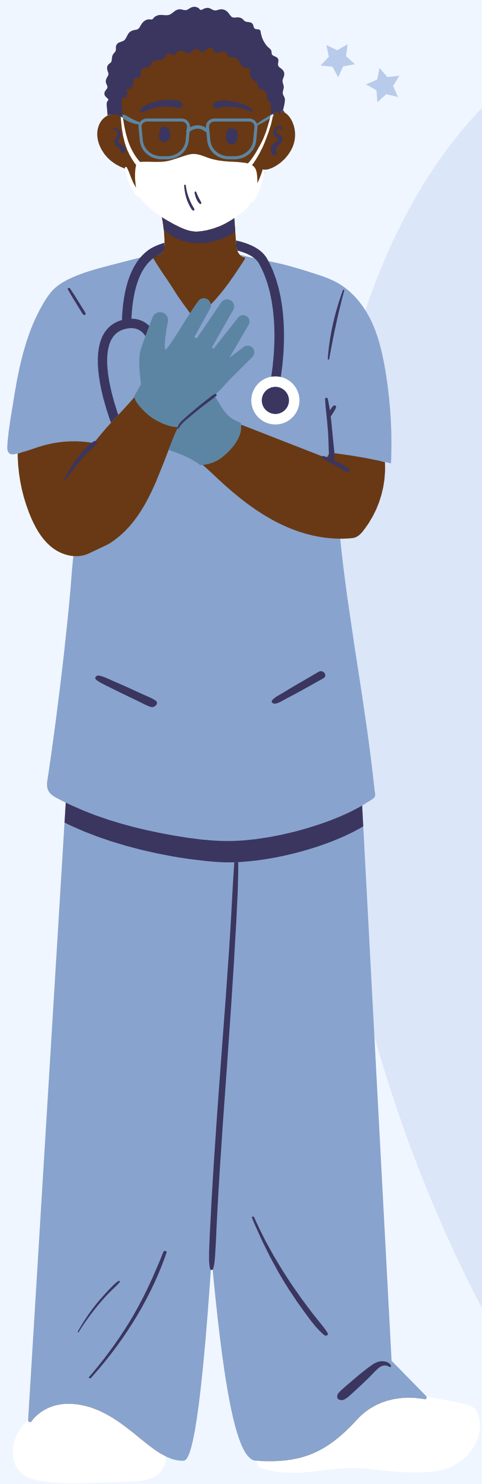
Outcome of our work

Our analysis identified key risk factors including age, dietary habits, biomarkers, blood pressure, medication usage, and comorbidities like diabetes.

Impact of our work

By incorporating these insights into screening protocols, we can customize policy premiums for higher revenue for the insurance company.

References:



Here are the links of the papers we referred:

- Whittaker, J., & Wu, K. (2022). Low-fat diets and testosterone in men: Systematic review and meta-analysis of intervention studies. arXiv:2204.00007 [q-bio.QM]. https://doi.org/10.48550/arXiv.2204.00007
- Langner, T., Östling, A., Maldonis, L., Karlsson, A., Olmo, D., Lindgren, D., Wallin, A., Lundin, L., Strand, R., Ahlström, H., & Kullberg, J. (2020). Kidney segmentation in neck-to-knee body MRI of 40,000 UK Biobank participants. arXiv:2006.06996 [q-bio.QM]. https://arxiv.org/abs/2006.06996v1
- Akinleye, A., Oremade, O., & Xu, X. (2024). Exposure to low levels of heavy metals and chronic kidney disease in the US population: A cross-sectional study. PLOS ONE, 19(4), e0288190. https://doi.org/10.1371/journal.pone.0288190

Thank you for your attention