Holmusk Assignment Report Draft 1

# Introduction

# Data –

## Preprocessing

Repeated Patient IDs – Handling strategy? 🡪 Remove all?

## Exploratory Data Analysis (Univariate & Multivariate)

1. Distribution of cost
2. Distribution between gender
3. Distribution between races
4. Distribution between age
5. Distribution between BMI
6. Distribution of lab tests
7. Distribution of medical history availability

# Method

GLM

Regularization method? Lasso

Cross-validation strategy

Variable Importance

# Results

# Discussion

Statistical Significance vs Practical Significance

Existing Literature

# Conclusion

# Appendix