

# RDS Project Proposal

**Ke Yang (ky630) and Zane Dennis (zdd210)**

**Data:** health insurance personal data. The attributes include sex, gender, bmi (body mass index), children, smoker, region, and charges (predicted values).

**Link of data:** <https://www.kaggle.com/mirichoi0218/insurance>

**ADS:** regression model to predict the insurance costs, including a linear regression and a random forest regression model.

**Link of ADS:** <https://www.kaggle.com/ruslankl/health-care-cost-prediction-w-linear-regression>

**Nutrition label of health insurance cost predictor:** focus on explanation of the data, importance of the attributes to the predicted values, how the correlation between attributes affect the predicted values, whether the predicted values are stable considering the input, etc.