Name \_\_\_\_ Zhang Xinge,QianZhang\_\_ Date: \_\_2023.12.6\_\_\_

NYU ID: \_\_\_\_N10837293, N19756113\_\_\_\_\_\_

Net ID: XZ4513, QZ2570

Course Section: \_\_\_\_\_ CSCI-GA.2433-001\_\_\_\_\_\_\_\_\_

**Project #3**

Total in points (100 points total): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Professor’s Comments:

Affirmation of my Independent Effort:

\_\_\_\_\_ZHANG XINGE\_Zhang Qian\_\_\_\_\_\_\_\_\_

(Sign here)

README:

This assignment was completed by a group consisting of two students Xinge Zhang and Qian Zhang. Each member is considered to contribute equal effort to this solution.

In this section, we plan to use the datasets [Medical Insurance Premium Prediction](https://www.kaggle.com/datasets/tejashvi14/medical-insurance-premium-prediction) as our data lake.

ossible to handle.

Heatmap for correlation between columns:

图表, 树状图

描述已自动生成

Distribution of Age:

图表, 直方图

描述已自动生成

Distribution of height:

图表, 直方图

描述已自动生成

Dependent and independent feature split

Daata normalization

Train test split

图形用户界面, 文本, 应用程序

描述已自动生成

Find important feature through linear regression:

图表, 条形图

描述已自动生成

Visualize important features through random forest regressor:

图表, 条形图

描述已自动生成

Visualize important features through xgboost:

图表, 条形图

描述已自动生成