

7126 Assignment 1 – Insurance Claims Prediction Report

February 28, 2023 MAI Miaoling UID: 3036043761

In this project, I executed the following steps:

- download the data
- import the Python packages
- have a quick look at the data structure (using histograms)
- create a test set (randomly pick 20%)
- look for correlations (found that age and BMI are highly correlated, and age is the most promising attribute)
- clean the data (replace the missing numeric data with median, transform the text of categorical data into numeric data)
- train the linear regression model
- Evaluate the model

Conclusion and observation

Age is the most promising attribute to predict the insurance claim and linear regression was trained in this project. The RMSE of this model is 79828.

For the company's management, they should pay much more attention to the age of their clients. To get a more accurate model, they can try and compare more analysis models and do further research on BMI or other combination factors.