**Insurance premium prediction**

Insurance Dataset for Predicting Health Insurance Premiums in the US is a collection of data on various factors that can influence medical costs and premiums for health insurance in the United States. The dataset includes information on 10 variables, including age, gender, body mass index (BMI), number of children, smoking status, region, occupation, type of insurance plan etc

There are 1 million records in this dataset.

You can use the first 700k records for building the model, 200k records for evaluating the model and the rest of the records as live records on which model would be evaluated in production.

Please create separate .py files for model training and prediction.

Make sure that your dataset during the model building and prediction phase goes through the same data processing and feature creation steps.

Your final deliverable can either be a python script which can be scheduled as a batch script to process premium calculations in batch or a flask api endpoint which can be used to calculate the premium in real time.

Please submit the final report which give clear idea on below points

1. ML model selected.
2. Features selected.
3. Feature importance
4. Evaluation of model
5. Original code

Use the “Insurance\_Prediction’’ table from the Database.db file for the project on predicting insurance.

Please feel free to reach out to your mentor in case of any doubts.