**Proposal** 1-2 pages (due Class 6) (5 pts)

a. What will be your approach to solve the given problem?

b. What information/data will you use? (Besides the provided data)

c. What will be your performance measure and success criteria?

1. Calculate correlation and information gain on train.csv for each of the target variable. Since there are four policy options available to the customer, we can train our model by applying classification technique for target variable ‘policy’. As the price of the policy depends on the customer characteristics and the policy selected, we can apply regression techniques to estimate what the policy ‘price’ to quote.
2. We will use train.csv to train our model. The trained model will be used on test\_session\_history.csv to test the accuracy. Cross-validation?
3. Performance measure: We will perform sensitivity analysis and calculate accuracy for model. Success criteria: we would calculate the absolute errors between predicted price and actual price on test\_session\_history.csv. We will define the tolerance value, any predicted value for ‘price’ which lies within the tolerance range is considered to be successful prediction.