Unit 5 Assignment: Should I Sell You Car Insurance? PLEASE FOLLOW INSTRUCTION AND LOAD WORKED SCREEN SHOTS TO DOCUMENT Outcomes addressed in this activity: Unit Outcomes: (Quantitative Risk Analysis). Predict risk outcomes using historical data. Create a data analysis model appropriate to a given risk scenario. Recommend a course of action based on a model’s output and acceptable business norms. Course Outcome: Recommend proactive measures to address ethical pitfalls to risk analytics activities. Purpose The purpose of this Assignment is to give you the opportunity to build a decision tree in R Studio, interpret the tree, and apply it to an uncategorized data set. In completing the Assignment, you will see a real decision tree in action in a context that is used every day in the real world of selling insurance policies. Buying and selling insurance is all about transferring risk. Companies that sell insurance assume some of the risk of their policy holders in exchange for a paid monthly premium. If claims paid out exceed premiums paid in, the insurance company risks bankruptcy. Decision trees are a useful tool to decide who can buy insurance and on what terms. Assignment Instructions Scenario: You work for an insurance company that has many policy holders, and many agents who sell insurance to new customers every day. You have been asked to use historical data about past and current policy holders to build a decision tree that will be used by sales agents to determine the insurability of potential new clients. You will use two data sets to do this. The Policy Holders data set contains information about current and past auto insurance customers, such as whether or not they have a claim or ticket in the past 12 months, an accident in the past 36 months, how they pay for their policy, their gender and marital status, and the level of activity associated with their insurance account (this is Low, Moderate or High based on frequency of changes to the policy, frequency of late or partial payments, and other similar account activity). Note that the only variable in the Policy Holders data set that is not also in the Policy Buyers data set is Insurance Category variable. This is the dependent variable that you will predict using a decision tree model. For the Policy Holders, you have the benefit of hindsight since your company did sell auto insurance policies to all of the people in this data set, and looking back on their activity as policy holders, they have each been assigned one of Insurance Category values: Insure-Best Terms, Insure-Risk Terms, Insure-High Premium, or Do Not Insure. The “Best Terms” customers are those who have paid their premiums and had no or few claims that have cost your company money. They are the lowest risk customers. The “Risk Terms” customers have been good for your company, but have had a few claims or incidents that have cost the company money. They are still a good risk for the company, but may have slightly higher premiums or lower coverage amounts in order to account for the higher risk to the company. The “High Premium” customers are those who have had a number of claims or other problems that have cost the company money (e.g., maybe they have not always paid their premiums on time or in full), but still have been worth insuring as long as they paid higher premiums than most of the other customers. They represent a higher risk for the company, and therefore must be sold policies at higher premiums and lower coverage. The “Do Not Insure” customers are those who have filed too many claims and/or claims that have cost more than what they have paid in premiums; or who have been unreliable in paying their premiums to the point where they cost the company more money than they pay in, and are therefore not a good risk for the company. They may have had their policies cancelled by the company due to excessive risk that the company cannot bear. Assignment Requirements Prepare your Assignment submission in Microsoft Word following standard APA formatting guidelines: Double spaced, Times New Roman 12-point font, one inch margins on all sides. Include a title page and references page. You do not need to write an abstract. Label all tables and figures. Cite sources appropriately both in the text of your writing (parenthetical citations) and on your references page (full APA citation format).