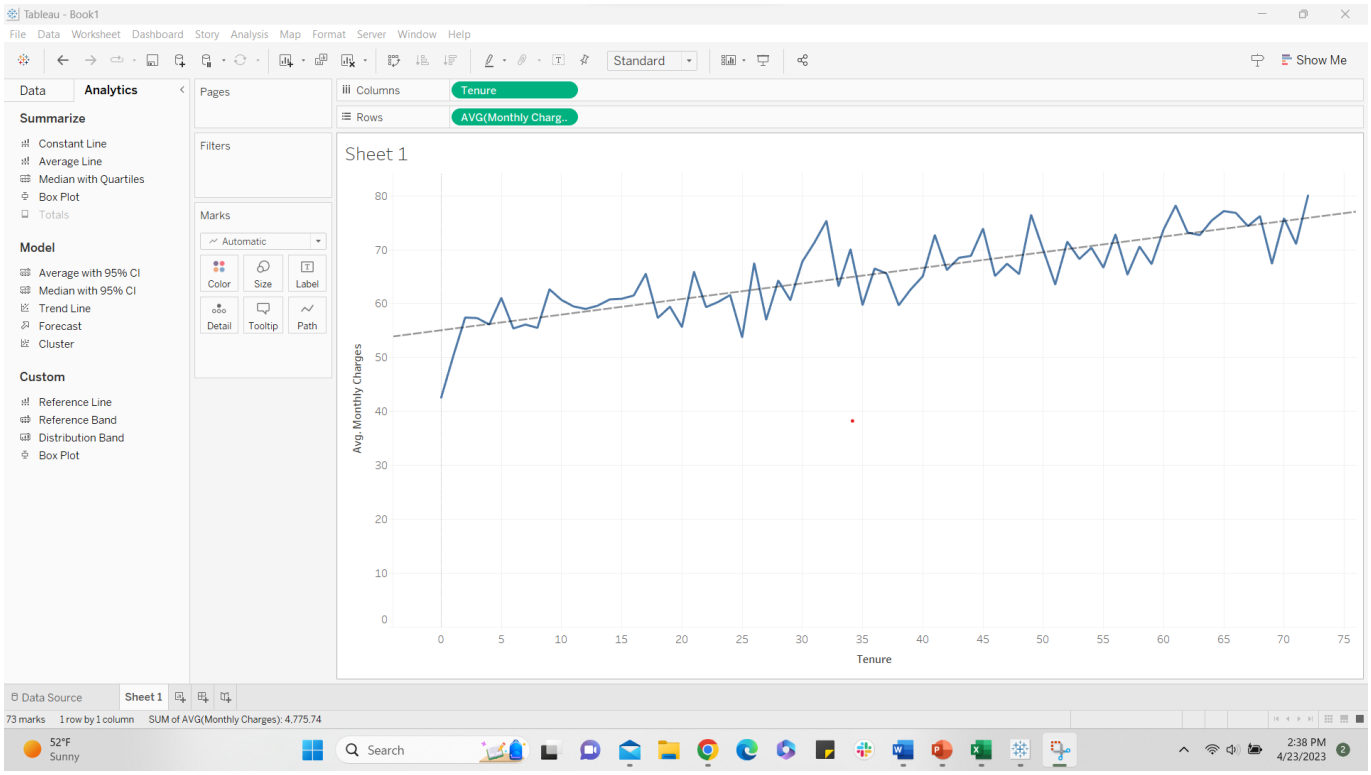
1. Step 2.4 Screenshot 1: A screenshot of your line chart with the trend line with date and time (5 points). Your screenshot must correctly and clearly show line chart with the trend line; it must include the line chart, trendline, x axis with a meaningful label, and y-axis with a meaningful label in Tableau, and display appropriate date and time.



2. Step 2.4 Questions:

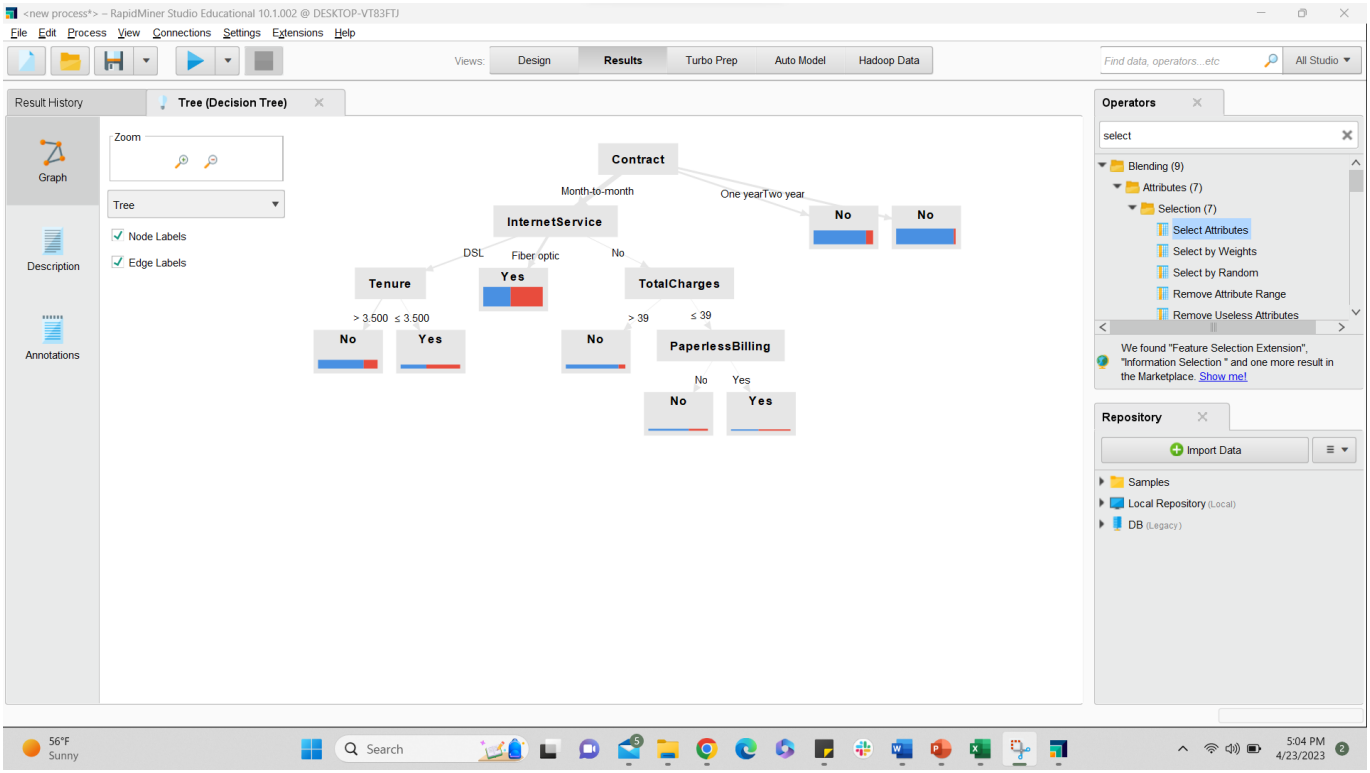
- What is this chart about? (2 points)
- What meaningful pattern or rule can you observe from the chart? (2 points)
- Based on what you observe, please present why it is important to retain customers. (3 points).

Requirements: Your responses must be clear, relevant, and concise (No more than 150 words in total; please indicate the word count at the end); if you fail to provide the word count, a penalty of 1 point will be applied; if your response is more than 150 words, a penalty of 1-3 points will be applied.

The chart shows customers at each tenure and what average monthly charge they pay. It was used to analyze whether customers pay more or less the longer the tenure. There appears to be an increasing pattern in this chart. It is important to retain customers for overall lower prices.

WC: 49

3. Step 4.7 Screenshot 2: A screenshot of your decision tree model (the one with a tree graph) with date and time after running your decision tree model (5 points). Your screenshot must correctly and clearly show the decision tree model; it must include all the nodes and branches in RapidMiner and display appropriate date and time.



4. Step 4.7 Question: If you are asked to present this decision tree to someone who have no background about data mining, what are you going to say. (9 points).

Requirements: Your response must

- Describe a decision tree model generally, including what a decision tree is and why it is used (3 points).
- Describe the decision tree graph in your screenshot specifically (specific to this case), including what it is for (i.e., purpose) and what it consist of (i.e., elements). Your response must be (3 points).
- Illustrate how the decision tree is used using 1-2 examples (3 points).
- Be no more than 200 words and indicate the word count at the end.

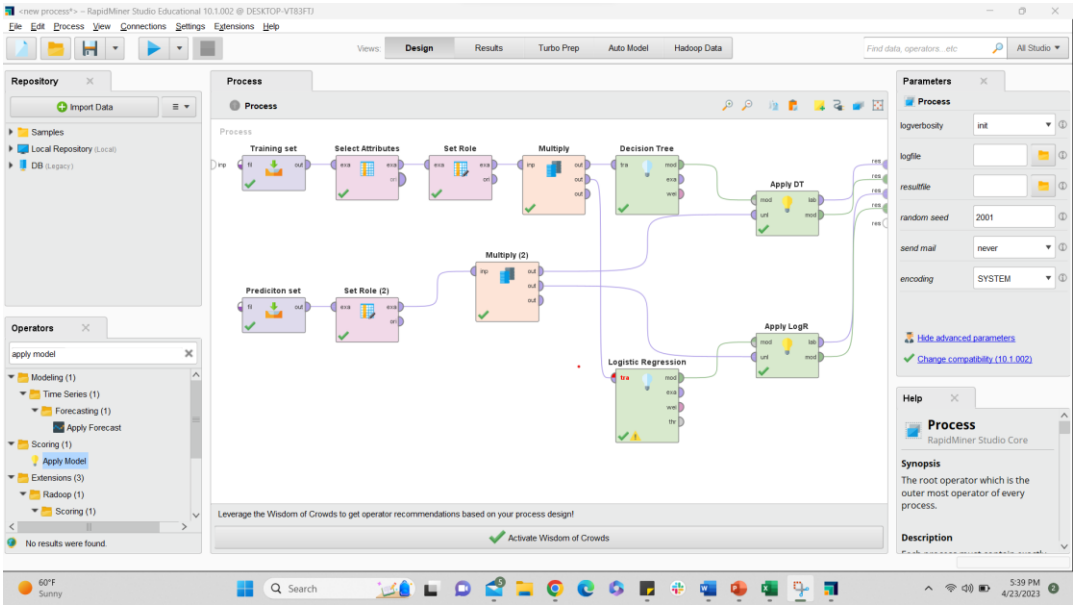
If you fail to provide the word count, a penalty of 1.5 points will be applied; if your response is more than 200 words, a penalty of 1-3 points will be applied.

A decision tree model is generally used to provide a visual on predictions based on given attributes. The decision tree model above provides a visual on what attributes indicate whether a customer is likely to churn. Given parameters, if an attribute leads a customer to not churning, we move down to the next attribute and so on. For this specific example, if a customer is signed up for a month-to-month contract, they are more likely to churn than a customer who has signed a one- or two-year contract. Decision tree models can be used on situations like determining turnover in jobs and what kind of factors affect that.

WC: 108

5. Step 5.5 Screenshot 3: A screenshot of your RM process with date and time at Step 5.5 (6 points).

Requirements: your screenshot must include all the operators (both training and prediction steps, both decision tree and logistic regression models) in a correct sequence in ONE process.



6. Step 5.5 Question: Please present the data mining process (CRISP-DM) using your RM process to someone who has no data mining background (18 points).

Your presentation must:

- Briefly describe each step of CRISP-DM using your own words in 1 or 2 sentences
- Clearly describe how each step is applied to this case; if any operator(s) in your RM process is (are) involved in this step, please incorporate the description of this operator in your presentation (see an example below);
- Be no more than 400 words in total and indicate the word count at the end.

An example in Step 3.4 in HW4: In the modeling step, this case uses the linear regression operator to create a linear regression model with quality as the target attribute and other variables as predictors selected via the M5 prime method.

Recommendation: you may come back to answer this question after completing Step 5.6-5.8.

3 points for each step: 1 point for the general description and 2 points for the specific description in this case. If you fail to provide the word count, a penalty of 2 points will be applied; if your response is more than 400 words, a penalty of 2-5 points will be applied.

The first part of the data mining process is business understanding and we did this by reading the description about Mark, his organization, and what the current business needs were. The next step is data understanding by analyzing the data provided and where it's coming from. We received the data in CSV files with selected attributes. After data understanding, the next step is data preparation. I prepared the data by using the operators above like select attributes to select the relevant attributes in the current case. I also used the decision tree and logistic regression operators to reformat the data. Modeling is the next step in data mining, and I was able to use the apply model operator, labeled as Apply DT and Apply LogR in the above visualization. For the next step evaluation, there is going to be an evaluation of how helpful the model will be for business needs. I was able to implement the evaluation stage by analyzing the data in visuals and charts, while also answering business need questions and implementing calculations. The last step is deployment, where the model is used by the business and can help the business understand the output. I was able to deploy the model by describing the screenshots in this assignment.

WC: 211