**CIDM/ECON 6308 Exam 2 Screenshots Submission**

**(50 points; Due 11:59 CT April 23, 2023)**

**Requirements: This exam is open book, open slides, and open notes. Because this is an individual exam, you are not allowed to collaborate nor discuss with anyone else during the exam period. Sharing anything related to the exam with any person or party violates the University's Academic Integrity Code, as well as the PEV COB Student Code of Ethics listed in our syllabus, and will be reported to the Dean Office of PEV COB. Any question about the exam should be directed to the instructor.**

**Please follow the instruction and requirements to answer each question. After completing the exam, please compile your screenshots and responses to open-ended questions here and then submit it via Exam 2 Screenshot Submission while submit your answers to other questions via Exam 2 Objective Questions Submission on WT Class. It is your responsibility to make your screenshots and answers meet the required format; otherwise, you would lose points because of wrong format. Specifically,** screenshots without appropriate date and time will receive a penalty of at least 50% of your grade. **If you do not use this document as a template, your submission WILL NOT be graded and a zero point will be assigned.**

When answering the open-ended questions, you are allowed cite any reliable reference, but you have to rephrase it in your own words and use the APA format to list your reference at the end, which will not be counted to the word limit. If you fail to cite or rephrase anything you actually cite, you are committing academic violation, which will be reported to the Dean Office of PEV COB. A possible penalty may include a zero point in this exam, a failing grade in this course, or suspension from the College.

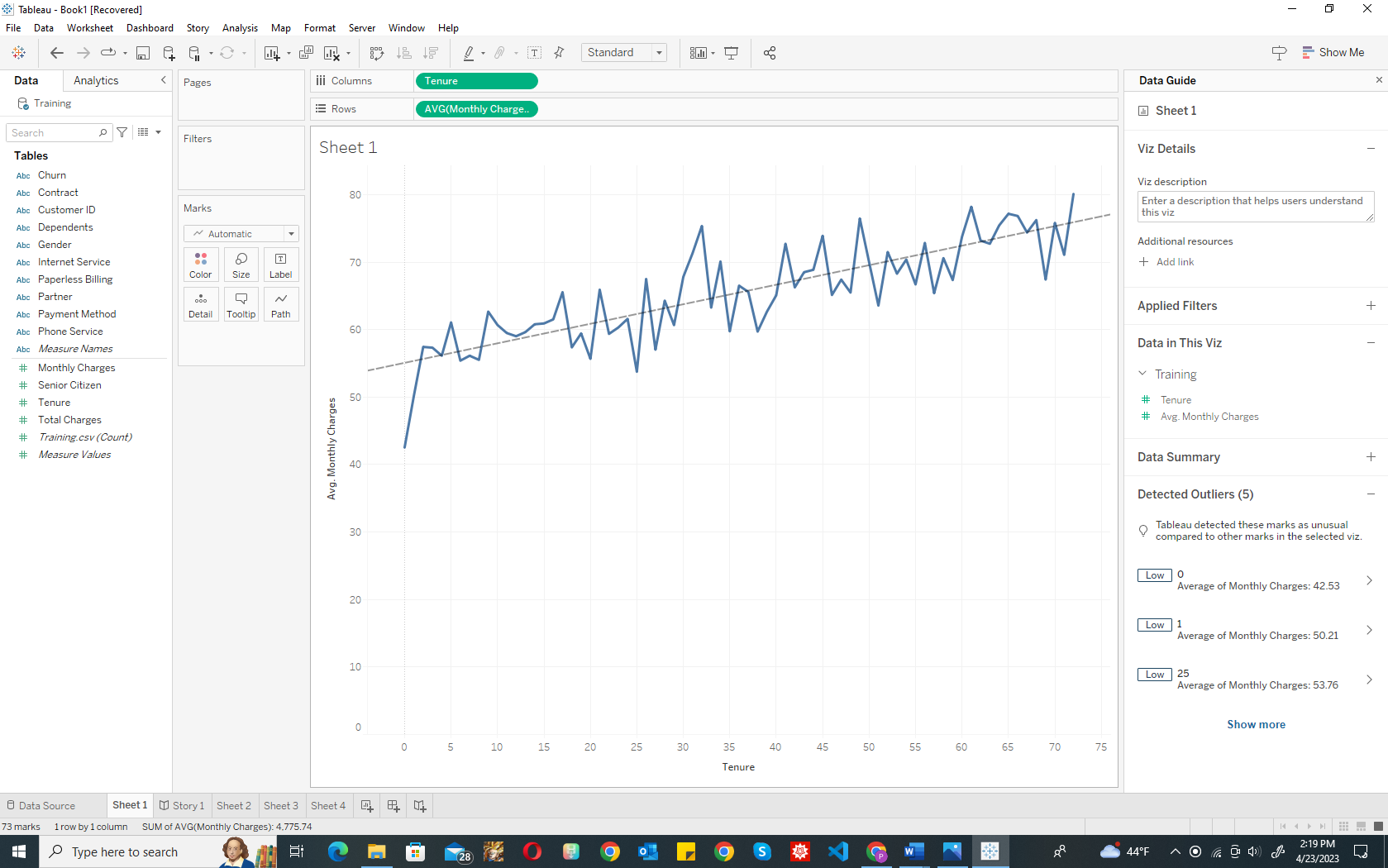
Please read, understand, and comply with these requirements carefully in this exam by typing your name below.

Your Name (First Last): Mehnaz Afrose

If you still do not know how to take a screenshot with date and time, please check the website <https://www.take-a-screenshot.org/> for more instructions.

**If you do not know how to show the date and time on your MAC Book, Google your question or try** [**this site**](https://osxdaily.com/2014/06/23/show-date-menu-bar-mac-os-x/)**.**

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.

|  |
| --- |
| * This chart is about the relation between how long the customer is staying with the company and the average monthly charge. * This chart has an upper trend, the line chart has an increasing pattern. * This chart is important because we can see that the average monthly charge is increasing along with tenure. From the very first month, a customer is paying on average $42.53, and after one month a customer is paying on average $50.21. A customer with 72 months staying (which is the highest) is paid on average $80.10 per month. That is why it is getting tough to retain customers.   WC: 102 |

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.

Graphical user interface, text, application

Description automatically generated

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 is a tree-like machine learning model that makes judgments depending on input features and it can handle both numerical and category data and is interpretable. * In this case, this decision tree graph is used to observe which customers are more likely to churn. It includes the types of contracts the company has with its customers (i.e.: month-to-month, one year, two year), the types of internet services the company provides (e.g.: DSL, fiber optic), total charge, and the types of billing system. * This DT model is important because, from this we can observe that customers who have long term contract (one or two years) are more likely to stay. But the customers who have month-to-month contracts are more likely to churn. Fiber optic users are more likely to churn. DSL users who have tenure of more than 3.5 months are more likely to stay, but customers who has less than or equal 3.5 months tenure are more likely to churn. Among the non-internet service categorized customers, who use paperless billing system, they are more likely to churn. With this DT model we were able to find out all these important facts about customer churn.   WC: 196 |

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.

Diagram

Description automatically generated

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.

|  |
| --- |
| * Step of CRISP-DM:   + Business Understanding: Define the problem and project objectives.   + Data Understanding: Explore and understand the data.   + Data Preparation: Clean, transform, and format the data.   + Modeling: Select and apply a suitable modeling technique.   + Evaluation: Assess the model performance and check if it meets project objectives.   + Deployment: Deploy the model in a production environment. * First CRISP-DM step: The problem was having trouble retaining the customers, and I need to analysis whether 100 new customers (stored in Prediciton.csv) will churn. Second CRISP-DM step, I reviewed the dataset, how many attributes and which attributes it has (e.g., Gender, tenure, internet service, monthly charges, etc.). Third CRISP-DM step, I used “select attribute” operator for cleaning the data and removing an unnecessary column (e.g., customer ID). I also included special attributes. Fourth CRISP-DM step, I used Decision Tree model to identify which type of customers are more likely to churn, logistic regression model to compare both models. After that I used a predictive analysis to predict whether the new customers are likely to churn or not. Fifth CRISP-DM step: After comparing the models I was able to evaluate which customers are more likely to churn, which sectors need to be improved, how many predictions from the decision tree model will be dropped because of high uncertainty, how many predictions from the logistic regression model will be dropped because of high uncertainty. Sixth CRISP-DM step: The performance of this model, including these operators, is accurate and reliable. Based on this evaluation I propose to deploy this model into the main operating system to identify the areas for improvements to manage the retention of the customers.   WC: 276 |