**BAIS 329 FINAL PROJECT**

**SNEAKERS**

**Problem Description**

You are getting ready to graduate with your undergraduate degree in Business Analytics from Xavier University and currently on the job market. You and are being interviewed by an online footwear company that sells everything from sneakers, to sling back pumps, to hiking books. They do not have any brick and mortar stores, but they are known for their certified “Sneakerheads” that can be reached via phone or Zoom. This has been a competitive advantage for them in this market as they often sell high-end products from brands such as Balenciaga, Berluit, and Gucci, and they have found that customers sometimes like to ask questions and complete their purchases with one of the company’s certified sneakerheads when spending upwards of $1,000 on a pair of shoes. The have a data set they are providing to you that has information about a recent marketing campaign they executed. They ran 15 targeted online ad campaigns at places such as Facebook and Instagram, and even on a podcast. They want you to build a predictive model to help them predict purchasing behavior (to buy or not buy) to help their marketing team with their eCommerce strategy.

You use the following files to conduct your analysis.

1. “ShoesStudents.xlsx” This is the dataset with 1,600 observations you will use to build your model. You will want to divide this data into a training and validation data set to test your models.
2. “Data Dictionary.docx”
3. “TestData.xlsx”: This file contains an additional 400 lines of data that you will input into your model for me to evaluate your model. These 400 observations CANNOT be used in your model building. This is the very last thing that you will do for this project: drop this data into JMP (anyway that you want) and generate predictions for these 400 customers. This is where you will submit your final predictions for the project. (For grading, I must have the columns “Prob 1”, “Prob 0”, and “Most Likely Purchase”)

**Grading Rubric**

Quality of the Analysis 50 pts

* Thoroughness and completeness of the analysis
* Appropriate and accurate application of methods
* Accuracy of analysis and interpretations
* Make sure to include your final script in your model submitted

Interpretation and Implications of the analysis (Management Report) 20 pts

* How do you improve business operations?
* What future variables should they focus on?

Quality of Visualization, Presentation, and Format 25 pts

* Are your visual aids novel and easy to interpret?
* Does your presentation flow well from initial analysis of the

data to the methods used and conclusions?

Accuracy of Prediction

* 1st Place +10 pts
* 2nd Place + 8 pts
* 3rd Place + 6 pts
* 4th - 6th Place + 0-5 pts

Group work (scaled portion of the final score)

**Project Steps**

**Stage 1: Prepare for Modeling (Initial Data Analysis)**

* *Examine and understand the data*: Look at the individual variables and explore the relationships between the independent and dependent variables. Any method of exploration is acceptable (histograms, scatter plot matrices, graph builder, export to Excel, etc…) and use of multiple methods is recommended.
* *Clean, organize and transform data*: Assess the cleanliness of the data (missing values, erroneous data, outliers, etc…) and determine if data needs to be edited and/or transformed. (CLEARLY DOCUMENT ALL CHANGES)
* *Define relevant features of the data*: Determine which variables will be included in your analysis and what type of analysis you will conduct.

**Stage 2: Modeling**

* Use multiple models with a variety of inputs and options to not only analyze your data but to build your model. The following types of models/analysis could be considered:
  + Logistic Regression- Michael, Abigail
  + Decision Tree- Michael, Connor
  + Boosted Tree - Connor, Abigail
  + Random Forest - Connor, Abigail
* Compare the models to select the best model for your data.

**Stage 3: Developing a Management Report and Powerpoint Presentation**

Key elements of a management report:

* *Executive Summary: The business problem and the proposed solution*.

This should include a short (1-2 sentences) summary of the business problem and a brief overview of the report. The purpose is to provide the manager with enough details that they understand the business problem and proposed solutions.

* *Introduction: The problem, why it is important, and the role of analytics in solving the problem.*

Here, the team provides additional background information on why this is a problem that the company wants to solve and why it is important. Information on how analytics and modeling tools can help solve the problem should be provided.

* *Methods: The data, exploratory analysis, and modeling approach used.*

This should include a high-level summary of the data for key variables (work from Step 1 above), and a summary of the final model in non-technical terms. The core objective here is to provide some level of understanding of the data used in the model and how the problem was solved without being overly detailed or technical.

* *Results: The selected model and model performance.*

This is where you report on what was learned from the final model and how well the model performs. For example, this section should include a summary of the most important variables, misclassification rate, and statements such as, “increases the value of this variable increases the response rate by this amount on average.” Graphics should be used to provide insights, communicate results, and provide a better understanding of the model and modeling results. This is also a place to report any concerns or risks related to the model and the findings. For example, the model may produce a high false positive rate or there may be issues with data quality.

* *Conclusion: Link to business goal fit and key management insights.*

In this section, you should connect back to the business problem and provide a summary of what was covered and final results, long term concerns, and potential next steps.

* *Appendix: Background information if needed.*

You must provide a brief summary of the model(s) you did NOT select. Additional technical details, including JMP output and other details and findings that should be brought to my attention. This content should NOT be in your presentation but should enable you to answer deeper technical questions should they arise and support your analysis. This needs to be very neatly organized and should only be turned into me.