**Members: BSIT-3A**Mj Enriquez

Desiree Domingo  
Ronel Venturina

Kent Lee Superalez

Jonnerey Simbanagan

Rosalie Sumaligpon **Use any of the available public datasets on the internet ( e.g. Kaggle, UCI) and build your classification/prediction model. Suggested datasets: loan prediction dataset, supermarket dataset, iris dataset, wine quality dataset, etc.**

**Submit a document with the following contents:**

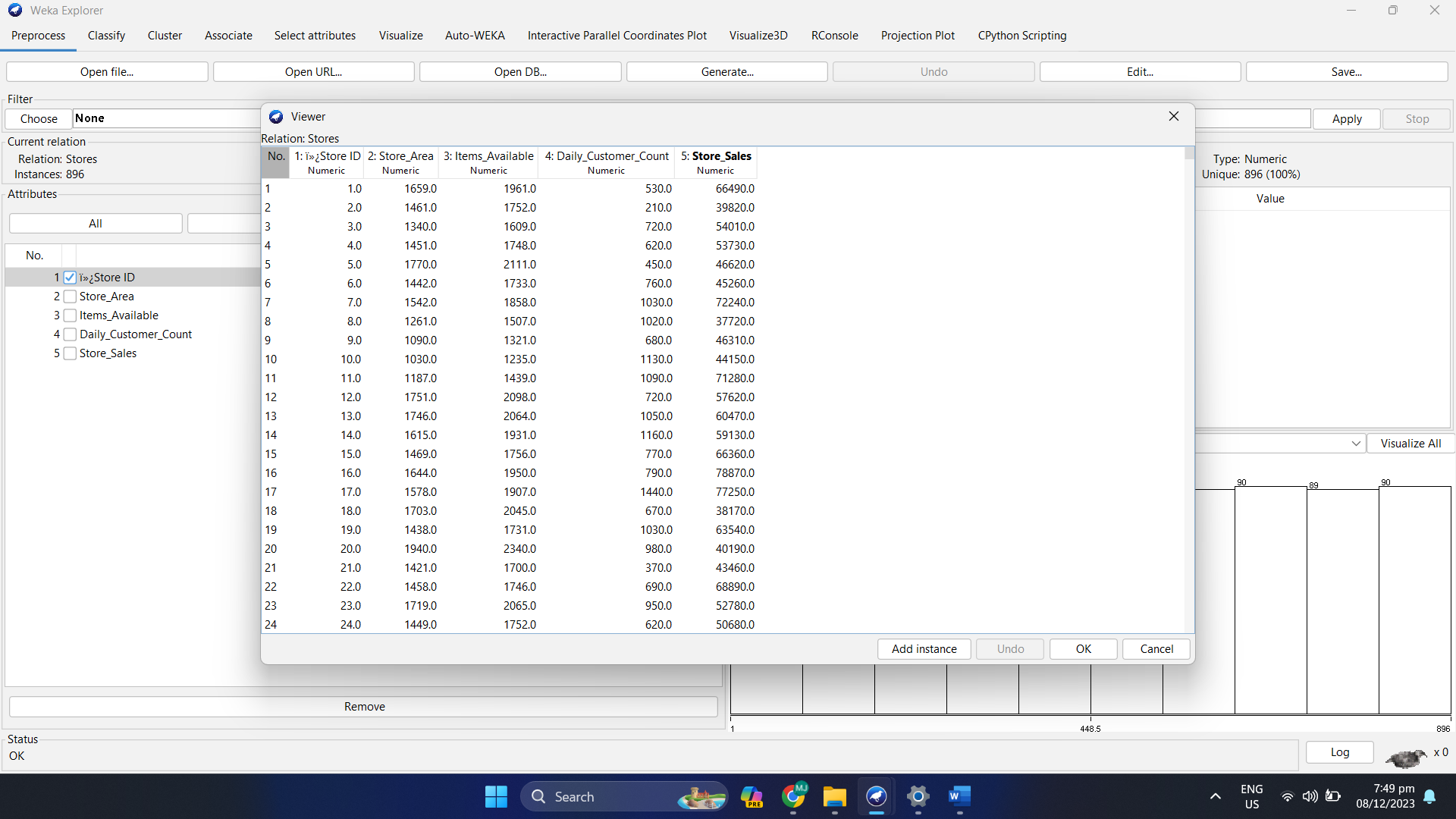
1. Dataset Description

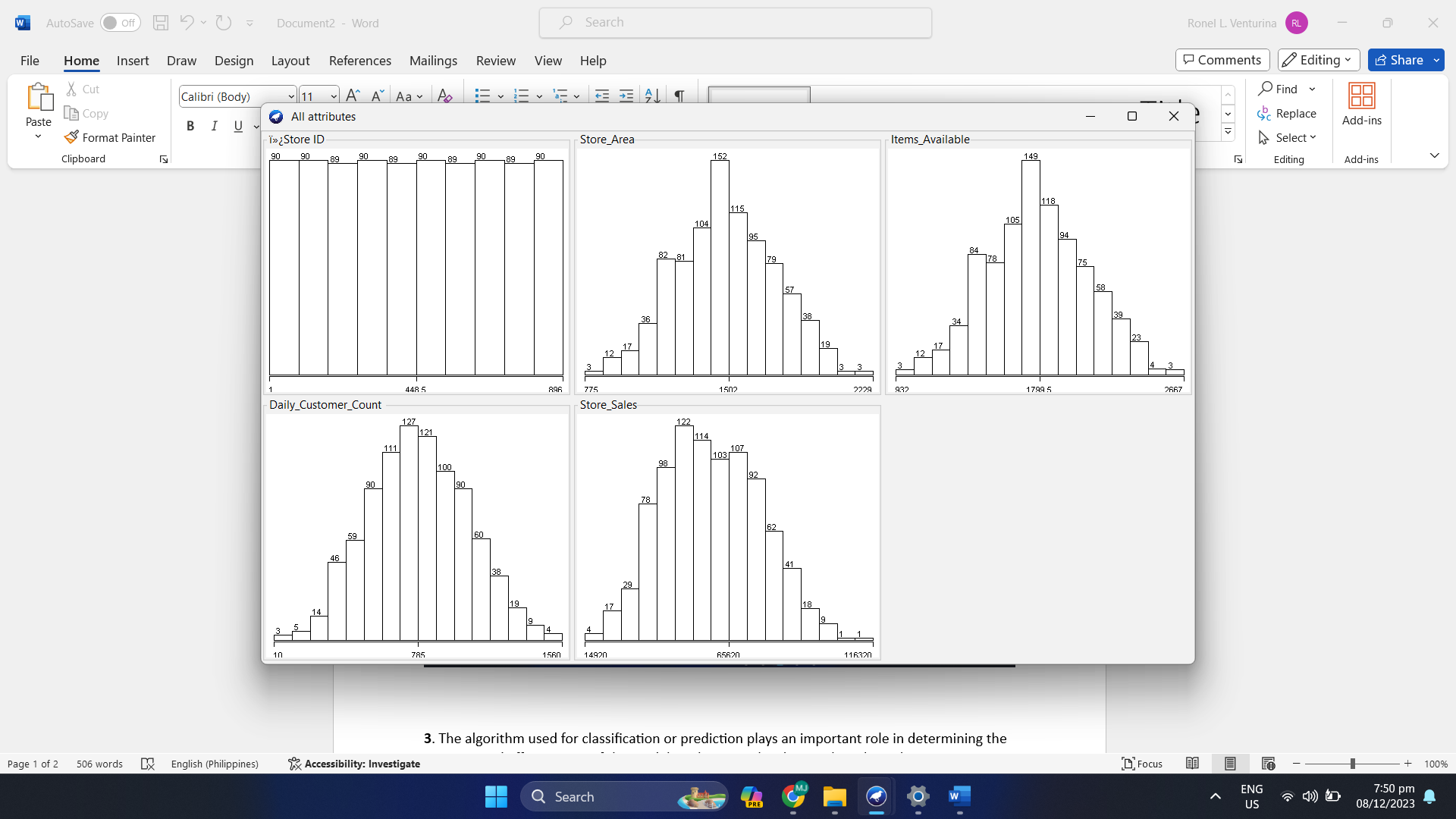
2. Screenshot of the Dataset Sample

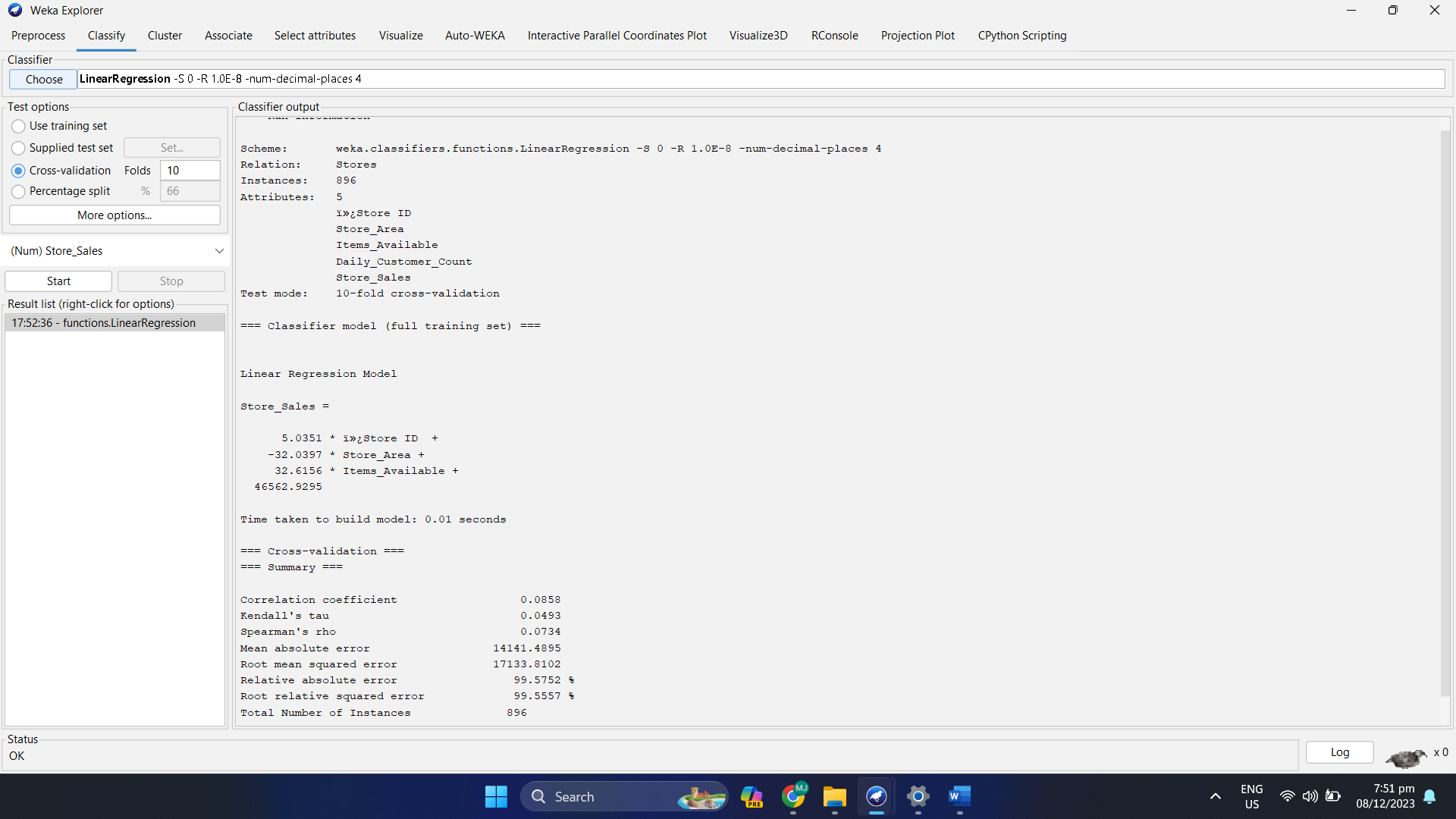
3. Description of the Algorithm Used for classification/prediction

4. Classification/Prediction Model screenshot (using Weka or any data mining tool)

5. Interpretation of the output

**1.** A supermarket is a self-service shop offering a wide variety of food, beverages and household products, organized into sections. This kind of store is larger and has a wider selection than earlier grocery stores, but is smaller and more limited in the range of merchandise than a hypermarket or big-box market. In everyday U.S. usage, however, "grocery store" is synonymous with supermarket, and is not used to refer to other types of stores that sell groceries.  
  
**2**. 



**3**. The algorithm used for classification or prediction plays an important role in determining the accuracy and effectiveness of the model. In this case, the chosen algorithm is linear regression. Linear regression is a supervised machine learning algorithm used to predict continuous outcomes based on one or more predictor variables. The model assumes a linear relationship between the predictors and the target variables, and aims to find the best fitting formula that minimizes the squared difference between the predicted value and the associated actual value the overall The coefficients quantify their impact on forecasted sales. The model is trained with the given data, and its performance is evaluated using cross-validation metrics, including correlation coefficients, mean absolute error, and root mean squared error, to provide insight into the accuracy and predictive ability of the model around  
  
**4**. 

**5**. The linear regression model aims to predict Store\_Sales based on four predictor variables: ï»¿Store ID, Store\_Area, Items\_Available, and Block The coefficients of the model provide insight into the relationship between these variables and sales. ï»¿A unit increase in Store ID is associated with a predicted increase in sales of 5.0351, while a unit increase in Store\_Area is associated with a predicted increase in sales of 32.0397 Conversely , the unit increase in Items\_Available corresponds to the predicted increase in sales of 32.6156. The intercept, 46562.9295, represents predicted sales when all predictor variables are zero, but its interpretation may be restricted depending on the context of the data In sample analysis, a correlation coefficient of 0.0858 indicates a linear relationship soft between forecasted and actual sales. An average absolute error of 14141.4895 means that the average prediction range would be $14,141.49. The root squared error is 17133.8102, which provides another measure of prediction error. The absolute error and the root of the square error give percentages of the magnitude of the error with respect to the true value of 99.5752% and 99.5557%, respectively Interpretations should be made considering specific contextual information and insights regarding the predictive performance of the model. If the sample does not meet the desired criteria, further analysis of the components or alternative sampling methods may be required.