Report of Linear Regression of Monet Dataset

* **Explanation of the meaning of each parameter in the Data, and describe the relationship between training datasets, validation datasets, and testing datasets if have. Describe the data cleaning, transformation, and normalization if have.**

Monet dataset has different columns, but in this assignment just the target(‘price’) and the independent variables (Width and Hight) are important to create a linear regression model. That because they have a good correlation compering with other columns.

First step before starting to build the linear regression model, cleaning the dataset is required. Fortunately, this data set doesn’t have any null values or errors, so I have started creating the model.

* **Explanation of the machine learning models that you implemented including the structure of the models, parameters, training process, and cost functions.**
* **Describe the procedure and results of the evaluation of the trained models.**
* First simple model:

This model was simple model. The dependent variable is the price, and height is the independent variable. I split the data as training dataset that has 80% of the data and testing dataset that has 20% of the data. The results of the cost function shows that this model is insufficient model for predicting the price. Figure 1.

Graphical user interface, text, application, email

Description automatically generated

Figure 1

* Second simple model:

Same previous steps I did with this model, but instead of hight I chose width as independent variable. Training model 80% and testing 20%.

The results shows that this model is better than the first model, but the mean error still high. Figure 3. Graphical user interface, text, application, email

Description automatically generated

Figure 3

* Third simple model:

I have added a new column to the dataset which is size = Hight \* width in order to see if this will work better than others. The training also 80% and testing is 20%. This model results is worse than the second model, but better than first model. Figure 3

Graphical user interface, text, application

Description automatically generated

Figure 4

* Final model is the multivariate linear regression model:

Before starting to create this model, I have normalized the data because the difference between the variable’s values is large. By normalization, the data range now is between 0 and 1.

This model has more than one independent variable. It has width, height, and size. The results are better than last models. The mean squared error in this model is less than all the previous observations. Figure 6.

Graphical user interface, text, application

Description automatically generated

Figure 6

Normalization:

Graphical user interface

Description automatically generated with medium confidence

Figure 7

* **Visualization of results and explain what does mean of the visualizations. The visualizations must enhance the interpretation of the results in an accurate and concise way.**
* Model 1:

We can see many points are so far of our prediction.

Graphical user interface, application

Description automatically generated

* Model 2:

It’s still many points are far than the predictuin, but it’s better than first model.

Graphical user interface, chart, application

Description automatically generated with medium confidence

Model 3:

Model is worse than the second model, but better than first model.

Chart

Description automatically generated

* Summary and conclusion:

After creating different linear regression models and observing the results of the models. I can say that last model which is multivariate model is the better one considering the results of the mean squared error and the r2 score. However, all the models can’t be considered as good models because the mean squared error is high and the r2 score is very low.