



## Introduction to Data Science

### Supervised learning (Model evaluation)

Suppose you are working on a project to predict the price of a house based on its size in square feet. Here are the variables in the dataset:

- Size (numeric): The size of the house in square feet.
- Price (numeric): The price of the house in dollars.

The following dataset, shows a regression model prediction for the price of a house based on its size.

Size (sq ft)	Price (OMR)	
1000	15000	14500
1500	25000	26500
2000	35000	33000
2500	45000	44500
3000	55000	60000

#### Tasks:

- **(10pt)** Calculate the mean squared error (MSE) of the model
- **(10pt)** Calculate the accuracy and error rate of the model (assuming 10% acceptable error margin). That means, if the predicted price is within 10% of the actual price, it is considered a correct prediction.
- **(10pt)** Comment on the accuracy of this model





Suppose you are working on a project to predict whether a customer will buy a product based on their demographic information. Here are the variables in the dataset:

- Age (numeric): The age of the customer.
- Income (numeric): The annual income of the customer, in thousands of dollars.
- Gender (binary): A binary variable representing the gender of the customer, either "male" or "female".
- Education level (categorical): A categorical variable representing the education level of the customer, either "high school", "college", or "graduate school".
- Bought product (binary): A binary variable representing whether or not the customer bought the product, either "yes" or "no".

Using this dataset, a build a binary classification model to predict whether a customer will buy the product based on their demographic information was built. The following is the confusion matrix for the model

performance of your model:

	Predicted Yes	Predicted No
Actual Yes	20	10
Actual No	30	30

- **(10 pt)** Using this confusion matrix, you can calculate the following evaluation criteria:
  - Accuracy
  - Precision
  - Recall
  - F1 Score
- **(10pt)** Explain your results

