# Introduction to Data Modeling

---

## Example

**Dataset** للشقق
المساحة، عدد الغرف، الاتجاهات، المنطقة، عمر البناء، السعر

**Training set:**
140, 3, W, Khalda, 12, **180000**
120, 3, E, AA, 20, **125000**
200, 5, E, AA, 5, **240000**
170, 4, W, BB, 15, **190000**

**Test set:**
150, 4, W, Khalda, 18, ?

**Some problems related to datasets:**
Underfitting**:** cannot predict any class, because the training set is few
Overfitting: wrongly predict, because the training set is very huge

---

## Phases of Create DS Model

1. Data collection
2. Data understanding
3. Data preprocessing
4. Data modeling:
    a) Split dataset into training set (60-80%) and test set (20-40%)
    b) Choose task
        i. Predictive - Supervised Learning (mission: predict, base: target)
        ii. Descriptive - Unsupervised Learning
    c) Choose model/algorithms [the algorithms are various according to (1) how to work and (2) accuracy result]
    d) Training
    e) Testing
    f) Evaluation

---

## Predictive Modeling

Some of its:
1. Classification (predict category/class)
    a) Decision Tree

b) Logistic Regression

2. Regression (predict continues numeric value)
   a) Linear Regression