# Predicting future sale prices

With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, predict the final price of each home. In addition, find out what correlates strongly to housing price.

**Summary**: We have secured a contract with a start-up company ("E-Z-Houses", and subsequently referred to in this document as "the client"), which is interested in business insights leveraging machine learning. This project will focus on the clients housing data to find insights and predict future housing prices.

1. **Context**: The client is interested in the following:
    a. Leveraging machine learning to see what factors affect housing prices
        i. Specific variables you want to analyse / visualise
            1. YearBuilt,Saleprice,OverallQual,TotalBsmtSF,GrLivArea,LotArea, OverallCond
    b. Perform an EDA to get a visual understanding of the housing market and how it made up.
    c. A Machine Learning system that can predict housing prices with >80% accuracy.

2. **Criteria for Success**: The approach to be developed will be deemed successful if :
    a. Business insights are found
    b. We find what factors are most important to the sale price
    c. Create a model that will predict future housing prices with accuracy > 80%

3. **Scope of the Solution Space**: This project will be restricted to the following:
    a. This project will be catered to our client but in reality could be sold to any competing firm in Ames,Iowa.

4. **Constraints within the Solution Space**:
    a. Only applies to Ames,Iowa.
    b. What company's data allows us to see.

5. **Stakeholders to Provide Key Insights**: Mr. Smith, Managing Partner. His email is Smith@gmail.com

6. **Required Data Sources**:
    a. We will be acquiring the data through UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/Bank+Marketing
    b. Data from the client

7. **Anticipated Data Science Approaches to be Used**:
    a. Regression focused approach
    b. Feature Creation
    c. Creating regression models
    d. EDA
    e. Scikit learn library