# Housing Market Price Analysis

Team Members: Scott Breitbach, Pushkar Chougule, and Rachel Nelson

# Introduction

### Background

After the impact of COVID-19 pandemic, the economy is looking get back on track. One of the indicators of recovery is the housing market demand / supply and prices. So, we were curious about performing the detailed analysis of the housing market and determine what all different factors contribute to the housing prices. Whether looking to buy your first home, sell a home or invest in a rental property, today’s housing market moves fast. So, this will be an interesting work ahead of us.

### Problem Statement

Is this house priced right? Due to market conditions, it’s hard to know what is considered a “good deal”. By predicting house prices based on factors, we can compare our predicted price to the asking price of a home for sale to help us place a proper bid on a house.

### Scope

Looking at Ames, Iowa, we will download market data and leverage that to build out a predictive model. We want to be able to change the city analytics based on user preference and output predictive housing prices to determine if the expected sale price in the specified market.

### Document Overview

In this project proposal document, we will be briefly touching upon below aspects of the project planning and execution for the DSC 630 course.

* Preliminary Requirement
  + Technical Approach
  + Data Sources / Plan
  + Model Deployment
  + Testing and Evaluation
* Expected Results
* Execution and Management of Project
  + Project Plan
  + Project Risk

# Preliminary Requirement

### Technical Approach

Our technical approach will be to leverage housing data and test several models including clustering, regression, and machine learning to find a model that best fits the data that we can use to determine housing cost.

### Data Sources or Plan for Data

We plan to download a Kaggle data set with common housing metrics including location, square footage of the house and lot size, number of bedrooms and bathrooms, year built, garage area, basement area etc. with the sale price as the target variable. Overall, there are 80 predictor variables available for analysis to estimate the value of the house.

Source: <https://www.kaggle.com/c/house-prices-advanced-regression-techniques/overview>

### Analysis

Our preliminary analysis will include the following summary statistics:

* Mean
* Median
* Mode
* Range
* Quartiles
* Skewed
* Kurtosis

We will be leveraging different charts to tell these statistics which may include histograms to help us determine distribution, box plots, bar charts, and scatter plots. We will also be looking at correlations in the data.

### Requirement for Development

We will be using the CRISP-DM framework to develop our requirements. To complete the analysis, we need to solidify our problem statement/hypothesis. We also need to determine what is the right data to gather that we wish to include into the analysis. gather the data, prepare and cleanse the data into a data frame which contains decimal data types.

1. Business understanding – What does the business need?
2. Data understanding – What data do we have / need? Is it clean?
3. Data preparation – How do we organize the data for modeling?
4. Modeling – What modeling techniques should we apply?
5. Evaluation – Which model best meets the business objectives?
6. Deployment – How do stakeholders access the results?

### Model Deployment

Our model deployment will include the following four tasks:

1. Plan deployment – Define our plan for deploying the model
2. Plan monitoring and maintenance - Develop a thorough monitoring and maintenance plan to update the model ongoing for different cities/areas
3. Product Final Report – As a team, create our summary of the project which includes the final presentation of results
4. Review Project - Conduct a project retrospective lesson learned about what went well, what could have been better, and how to improve in the future and what we could have changed given more time.

### Testing and Evaluation

We will compare accuracy and precision of models to determine which model performs best.

We are planning to use R squared method, Mean / Median Absolute Error, Mean Square Error and any other applicable methods we come across as an indication of variance between actuals and predicted values.

# Expected Results

We expect to be able to predict prices of houses based on factors available within the data set with at least 80% accuracy. We expect that the predictive model for the house prices predictions would be helpful to the end users in the market, who are searching for the house and would serve as a handy guide for them.

# Execution and Management of Project

### Project Plan

Communicate with each other via teams / emails. Keep shared files in our Teams chat.

* Week 1: Milestone 1 Due (Team Information/Communication Plan)
  + Start dedicated teams chat with team - Internal Teams channel created
  + Start talking about possible topics (finance, healthcare, weather, or other areas of interest) and suitable data sources (kaggle, google cloud public datasets, Data.gov, National Center for Environment Information etc.)
* Week 2: Milestone 2 Due (Data Selection and Project Proposal) & Peer Review
  + Finalize the project topic
  + Finalize data selection and share data on Teams
  + Draft project proposal
  + Determine the potential outcome / target variables from the project
  + Discuss about the potential challenges / risks
* Week 3 & 4: Project prep work and Peer Review
  + Start discussions around approach for preliminary analysis and set up time to discuss as a team on findings/next steps
  + Divide the work amongst group members
  + Decide on the approach for the usage of programming languages R and Python
* Week 5: Milestone 3 Due (Preliminary Analysis)
  + Deliver preliminary analysis which should be a combination of R and Python scripts.
  + Prepare suitable visualizations
  + Summarize the data characteristics and structures
  + Data clean up processing using python
* Week 6: Peer Review
  + Start project presentation draft in PowerPoint
* Week 7, 8 and 9: Milestone 4 Due (Project Presentation & Status) & Peer Review
  + Prepare the draft version of the predictive models in R
  + Coordinate on the code changes / seek each other’s help and feedback
  + Evaluate the model performance and identify improvements needed
  + Seek feedback from the other students in the class / exchange views with each other
  + Submit milestone 4 project documentation and status
* Week 10: Peer Review
  + Continue to refine draft documents and make any changes on shared documentation
* Week 11 and 12: Milestone 5 Due (Final project paper and presentation) & Peer Review (Due Saturday!)
  + Finalize / fine tune the predictive model, considering the feedback from previous weeks
  + Document the results and prepare the presentation
  + Coordinate for the preparation of final project presentation
  + Submit final paperwork

### Project Risk

The following are the risks to our project:

* Finding the correct data
* All members of the team are in different time-zones so coordinating efforts will be difficult
* Ensuring we meet all of the project requirements and use both Python and R for our final analysis results
* Dividing the work up properly so each team-member contributes the same amount of time and effort
* Learning from one another while remote
* Communicating project status and not duplicating efforts