

## Property Price Prediction

### Problem Statement

There are a number of factors which determine property prices, some are logical, based on economic theories and population density and some are based on more intangible factors, like availability of amenities & necessities, neighborhood, etc.

Build a linear regression model with stochastic gradient descent to predict the price of the property from the dataset having attributes such as sale type, sale condition etc.

### Data Description

- **Zoning\_Class:** Identifies the general zoning classification of the sale
- **Building\_Class:** Identifies the type of dwelling involved in the sale
- **Lot\_Extent:** Linear feet of street connected to property
- **Lot\_Size:** Lot size in square feet
- **Road\_Type:** Type of road access to property
- **Lane\_Type:** Type of alley access to property
- **Property\_Shape:** General shape of property
- **Land\_Outline:** Flatness of the property
- **Utility\_Type:** Type of utilities available
- **Lot configuration:** Lot configuration
- **Property\_Slope:** Slope of property
- **Neighborhood:** Physical locations within Ames city limits
- **Condition1:** Proximity to various conditions
- **Condition2:** Proximity to various conditions (if more than one is present)
- **House\_Type:** Type of dwelling
- **House\_Design:** Style of dwelling
- **Overall\_Material:** Rates the overall material and finish of the house
- **House\_Condition:** Rates the overall condition of the house
- **Construction\_Year:** Original construction date
- **Remodel\_Year:** Remodel date (same as construction date if no remodeling or additions)
- **Roof\_Design:** Type of roof
- **Roof\_Quality:** Roof material
- **Exterior1st:** Exterior covering on house
- **Exterior2nd:** Exterior covering on house (if more than one material)
- **Brick\_Veneer\_Type:** Masonry veneer type
- **Brick\_Veneer\_Area:** Masonry veneer area in square feet
- **Exterior\_Material:** Evaluates the quality of the material on the exterior
- **Exterior\_Condition:** Evaluates the present condition of the material on the exterior
- **Foundation\_Type:** Type of foundation
- **Basement\_Height:** Evaluates the height of the basement

# KPMG Data Science Prodegree

## Linear Regression: Problem Statement

---



- **Basement\_Condition:** Evaluates the general condition of the basement
- **Exposure\_Level:** Refers to walkout or garden level walls
- **BsmtFinType1:** Rating of basement finished area
- **BsmtFinSF1:** Type 1 finished square feet
- **BsmtFinType2:** Rating of basement finished area (if multiple types)
- **BsmtFinSF2:** Type 2 finished square feet
- **BsmtUnfSF:** Unfinished square feet of basement area
- **Total\_Basement\_Area:** Total square feet of basement area
- **Heating\_Type:** Type of heating
- **Heating\_Quality:** Heating quality and condition
- **Air\_Conditioning:** Central air conditioning
- **Electrical\_System:** Electrical system
- **First\_Floor\_Area:** First Floor square feet
- **Second\_Floor\_Area:** Second floor square feet
- **LowQualFinSF:** Low quality finished square feet (all floors)
- **Grade\_Living\_Area:** Above grade (ground) living area square feet
- **Underground\_Full\_Bathroom:** Basement full bathrooms
- **Underground\_Half\_Bathroom:** Basement half bathrooms
- **Full\_Bathroom\_Above\_Grade:** Full bathrooms above grade
- **Half\_Bathroom\_Above\_Grade:** Half baths above grade
- **Bedroom:** Bedrooms above grade (does NOT include basement bedrooms)
- **Kitchen:** Kitchens above grade
- **Kitchen\_Quality:** Kitchen quality
- **Rooms\_Above\_Grade:** Total rooms above grade (does not include bathrooms)
- **Functional\_Rate:** Home functionality (Assume typical unless deductions are warranted)
- **Fireplaces:** Number of fireplaces
- **Fireplace\_Quality:** quality of fireplaces
- **Garage:** Garage location
- **Garage\_Built\_Year:** Year garage was built
- **Garage\_Finish\_Year:** Interior finish of the garage
- **Garage\_Size:** Size of garage in car capacity
- **Garage\_Area:** Size of garage in square feet
- **Garage\_Quality:** Garage quality
- **Garage\_Condition:** Garage condition
- **Pavedd\_Drive:** Paved driveway
- **W\_Deck\_Area:** Wood deck area in square feet
- **Open\_Lobby\_Area:** Open porch area in square feet
- **Enclosed\_Lobby\_Area:** Enclosed porch area in square feet
- **Three\_Season\_Lobby\_Area:** Three season porch area in square feet
- **Screen\_Lobby\_Area:** Screen porch area in square feet
- **Pool\_Area:** Pool area in square feet
- **Pool\_Quality:** Pool quality

# KPMG Data Science Prodegree

## Linear Regression: Problem Statement

---



- **Fence\_Quality:** quality of fence
- **Miscellaneous\_Feature:** Miscellaneous feature not covered in other categories
- **Miscellaneous\_Value:** \$Value of miscellaneous feature
- **Month\_Sold:** Month Sold (MM)
- **Year\_Sold:** Year Sold (YYYY)
- **Sale\_Type:** Type of sale
- **Sale\_Condition:** Condition of sale

### Evaluation Parameters

Evaluation will be based on:

- Data Preprocessing
- Model Comparison
- Model Selection

### Data Preprocessing

Check the data distribution of variables and perform transformation if a variable's distribution is skewed. Perform label encoding on categorical variables.

### Model Comparison

Build linear regression models and compare results.

### Model Selection

Select the best model. Model selection to be based on model accuracy.

### Expected Outcome

Low RMSE and high coefficient of determination( $R^2$ ) is expected while predicting the outcome using test data.