Prediction of sale prices of house

Problem Statement

To predict the sale prices of houses

Project Team

1. Tejaswini Nutalapati
2. Aditi Bhargava

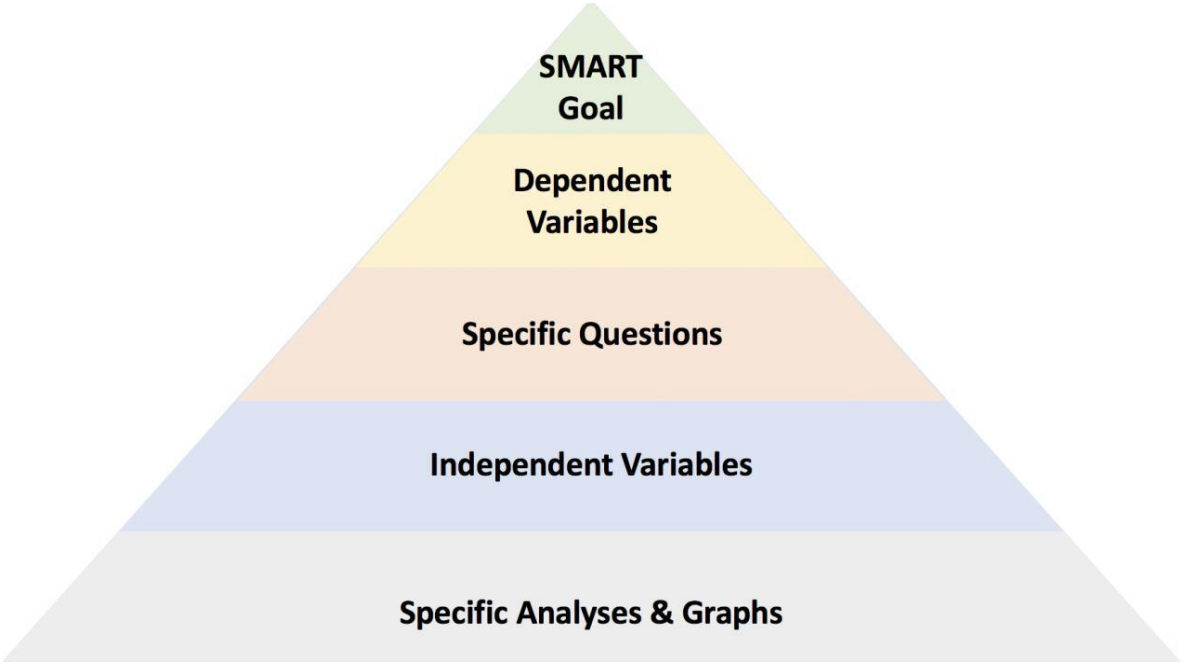
About the Data

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence. With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this dataset allows us to predict the final price of each home.

The Ames Housing dataset was compiled by Dean De Cock for use in data science education. It's an incredible alternative for data scientists looking for a modernized and expanded version of the often-cited Boston Housing dataset.

Data Souce: https://www.kaggle.com/c/house-prices-advanced-regression-techniques

Structured Pyramid Analysis Plan



Smart Goal

A goal that is Specific, Measurable, Attainable, Realistic, and Time-bound.
In this dataset, the smart goal is to predict the sale prices of houses.

Dependent Variables

Variables that directly pertain to the accomplishment of the SMART goal.
In this dataset, there is only one dependent variable;

- SalePrice

Specific Questions

Ideas and intuitions about what parameters may impact the independent variables.
In this dataset, the specific questions that can be asked are ;

- What is the effect of various variables on SalePrice
- What are the variables that have least/highest effect on SalePrice.
- What variables can be clustered into one variable
- What variables can be ignored while prediction

Independent Variables

Variables that may help explain the possible connections in the questions, above, to the dependent variables.
In this dataset, all the variables that help in predicting the sale price here are independent variables. i.e all variables other than SalePrice are independent.

- MSSubClass
- MSZoning
- LotFrontage
- LotArea
- Street
- Alley
- LotShape
- LandContour
- Utilities
- LotConfig
- LandSlope
- Neighborhood
- Condition1
- Condition2
- BldgType
- HouseStyle
- OverallQual
- OverallCond
- YearBuilt

- YearRemodAdd
- RoofStyle
- RoofMatl
- Exterior1st
- Exterior2nd
- MasVnrType
- MasVnrArea
- ExterQual
- ExterCond
- Foundation
- BsmtQual
- BsmtCond
- BsmtExposure
- BsmtFinType1
- BsmtFinSF1
- BsmtFinType2
- BsmtFinSF2
- BsmtUnfSF
- TotalBsmtSF
- Heating
- HeatingQC
- CentralAir
- Electrical
- 1stFlrSF
- 2ndFlrSF
- LowQualFinSF
- GrLivArea
- BsmtFullBath
- BsmtHalfBath
- FullBath
- HalfBath
- Bedroom
- Kitchen
- KitchenQual
- TotRmsAbvGrd
- Functional
- Fireplaces
- FireplaceQu
- GarageType
- GarageYrBlt

- GarageFinish
- GarageCars
- GarageArea
- GarageQual
- GarageCond
- PavedDrive
- WoodDeckSF
- OpenPorchSF
- EnclosedPorch
- 3SsnPorch
- ScreenPorch
- PoolArea
- PoolQC
- Fence
- MiscFeature
- MiscVal
- MoSold
- YrSold
- SaleType
- SaleCondition

Specific Analysis and Graphs

Statistical analyses and plots of independent versus dependent variables.

In this dataset, the analysis that can be done are,

- Univariate Analysis of independent variables vs SalePrice
- Bivariate Analysis on continuous variables vs SalePrice
- Bivariate Analysis on continuous variables vs SalePrice
- Multivariate Analysis to create the correlation plot and find significance of each feature