**Real Estate Regression Project**

You have been hired by a local real estate investor who is interested in understanding more about what factors are influencing the prices of homes in his area. He will use this information to help set prices for homes that he is preparing to sell (prediction), and he will also use this information to help make decisions about which homes to buy and how to renovate the homes that he currently owns (inference). (NOTE: These data include homes that were sold between 2006 and 2010. You can assume that he is using these data to make decisions about real estate investments in that same time frame (not current day).)

He has pulled information on 1460 homes in his area, and he would like for you to use these data to build a multiple linear regression model to help him make some data-driven decisions about his real estate investments.

The data set includes the following columns:

|  |  |
| --- | --- |
| **Variable Name** | **Variable Description** |
| Id | Property’s Unique Numeric Identifier |
| SalePrice | Sale Price ($) when the home was sold |
| LotArea | Area of the lot in square feet (SF) |
| TotalBsmt | Area of the basement in SF (0 if the home does not have a basement) |
| 1stFlrSF | Total area of the 1st floor of the home in SF |
| 2ndFlrSF | Total area of the 2nd floor of the home in SF (0 if the home doesn’t have a 2nd floor) |
| FullBath | Total number of full bathrooms in the house |
| HalfBath | Total number of half bathrooms in the house |
| BedroomAbvGr | Total number of bedrooms in the house (excluding the basement) |
| TotRmsAbvGr | Total number of rooms in the house (excluding the basement) |
| Fireplaces | Total number of fireplaces in the house |
| GarageCars | Total number of cars that fit into the garage (0 if no garage) |
| WoodDeckSF | Total square footage of wooden deck (0 if home doesn’t have a wooden deck) |
| OpenPorchSF | Total square footage of open porch (0 if home doesn’t have an open porch) |
| EnclosedPorch | Total square footage of enclosed porch (0 if home doesn’t have an enclosed porch) |
| 3SsnPorch | Total square footage of three season sunroom (0 if home doesn’t have a three season sunroom) |
| ScreenPorch | Total square footage of screened porch (0 if home doesn’t have a screened porch) |
| PoolArea | Total square footage of outdoor pool (0 if home doesn’t have a pool) |
| LotConfig | Location of the lot with respect to the street (Inside (between 2 homes), FR3 (facing road on 3 sides), FR2 (facing road on 2 sides), CulDSac (cul de sac lot), corner (corner lot)) |
| BldgType | Type of home (1Fam (single family detached home), 2fmCon (unit in a condominium building), Duplex (unit in a duplex), Twnhs (townhouse), TwnhsE (end unit townhouse) |
| YearBuilt | Year the home was built |
| YearRemodAdd | Year the home was remodeled (updated) |
| Foundation | Type of foundation the home is built on (BrkTil (brick), CBlock (concrete blocks), PConc (poured concrete), Slab, Stone, Wood) |
| CentralAir | Y = home has central air conditioning, N = home does not have central air conditioning |
| KitchenQual | Condition of the kitchen (Gd (good condition), Fa (fair condition), Ex (excellent condition), TA (condition unknown) |
| PavedDrive | Y = driveway is paved, N = driveway is not paved, P = driveway is partially paved |
| YrSold | Year the home was last sold |

**Rubric (total of 70 points):**

Slides are professional, clearly annotated, and do not exceed the page limit (10 points)

Slides and annotation are appropriate for a business audience (10 points)

Regression model meets the criteria for a good regression model (15 points)

Explanation of regression model is clear and accurate (15 points)

Python code is included and clearly documented (5 points)

Python code clearly demonstrates the iterative regression modelling process (5 points)

Presentation and analysis exhibit clear mastery of multiple regression (10 points) (For this item, 10 points = Exceeds Expectations, 5 points = Meets Expectations, 2 points = Below Expectations)