**House Project Phase 1**

Student ID - **1894990**

Student Name - **Meher Vishal Malla**

Section 1 - Introduction to Business Scenarios.

Section 2 - Generate a smart discovery Report using SAP Data Analytics in Cloud

Section 3 - Generate a House Price Prediction Model.

1. What Model to use?

Decision Tree Model?

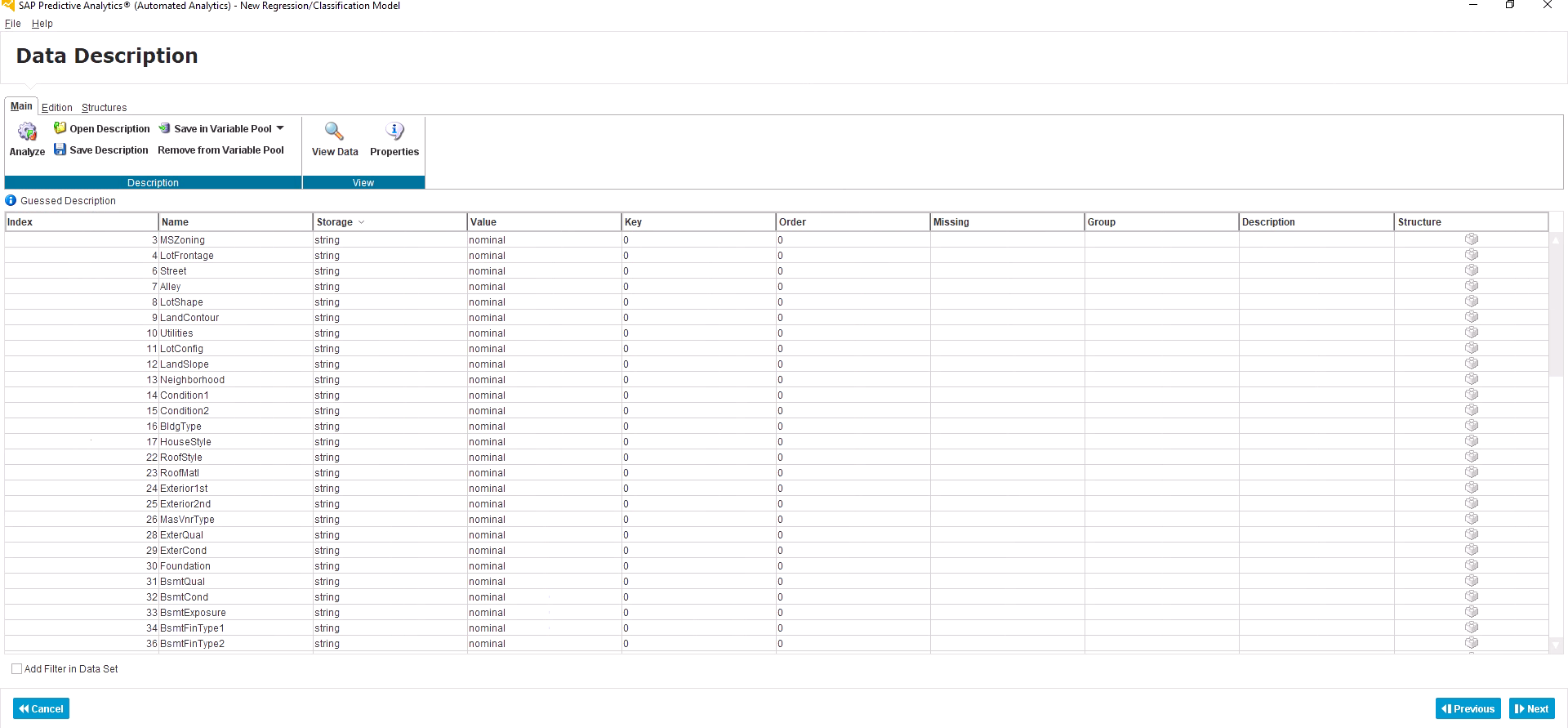
Cluster?

Regression? Linear Model, Non-Cluster Model.

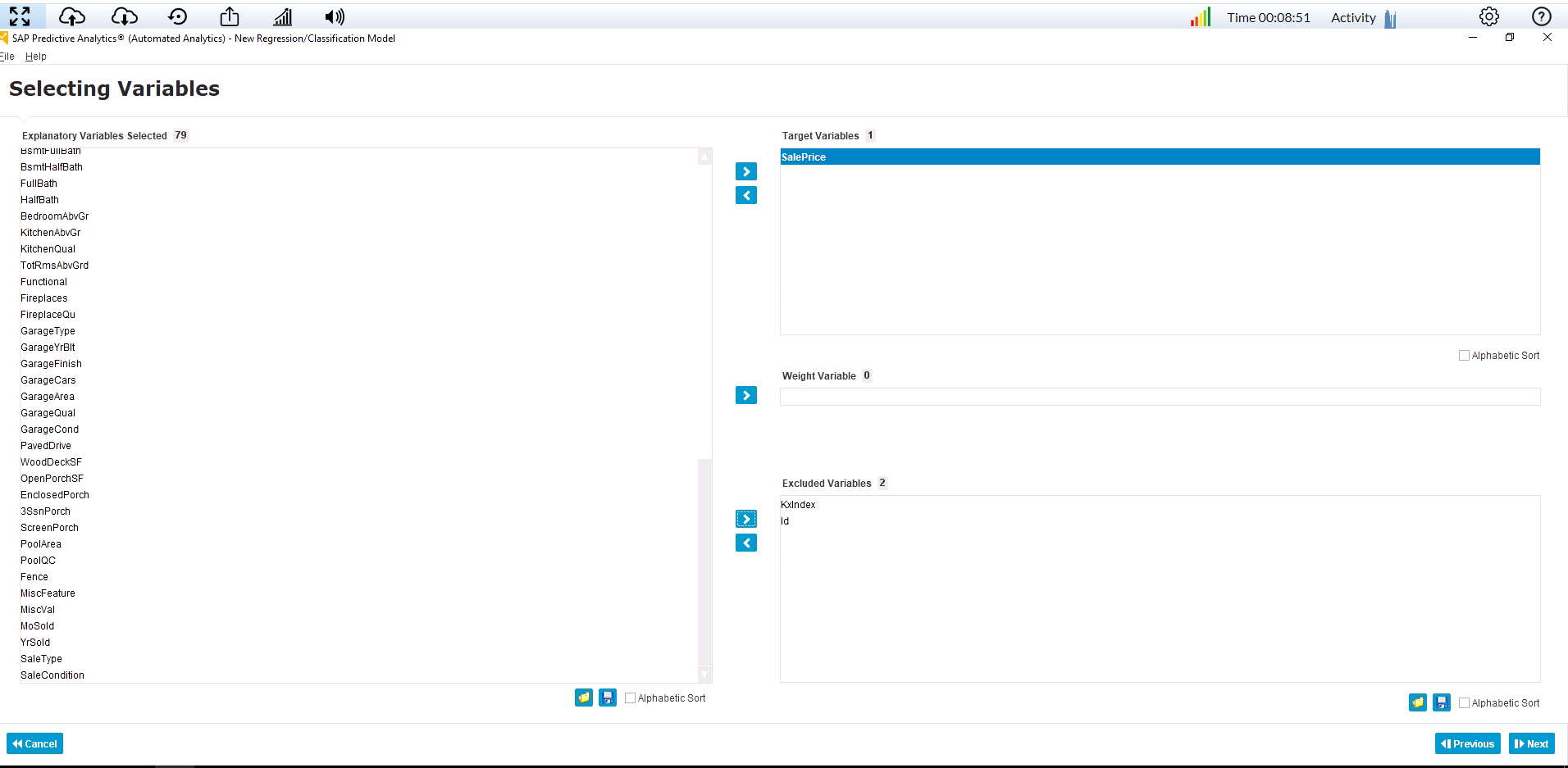
Time-Series Model?

Answer - Regression Model:

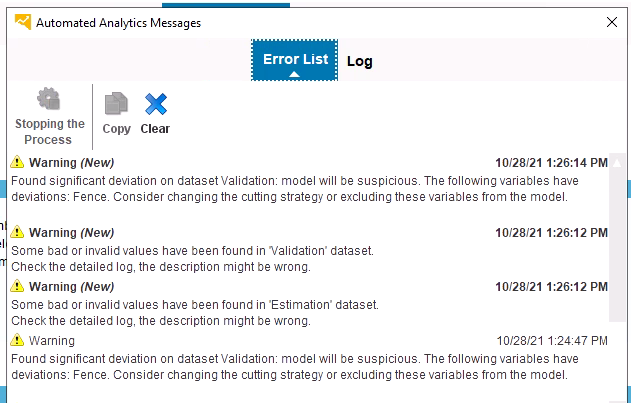
1. Automatic Regression Model.



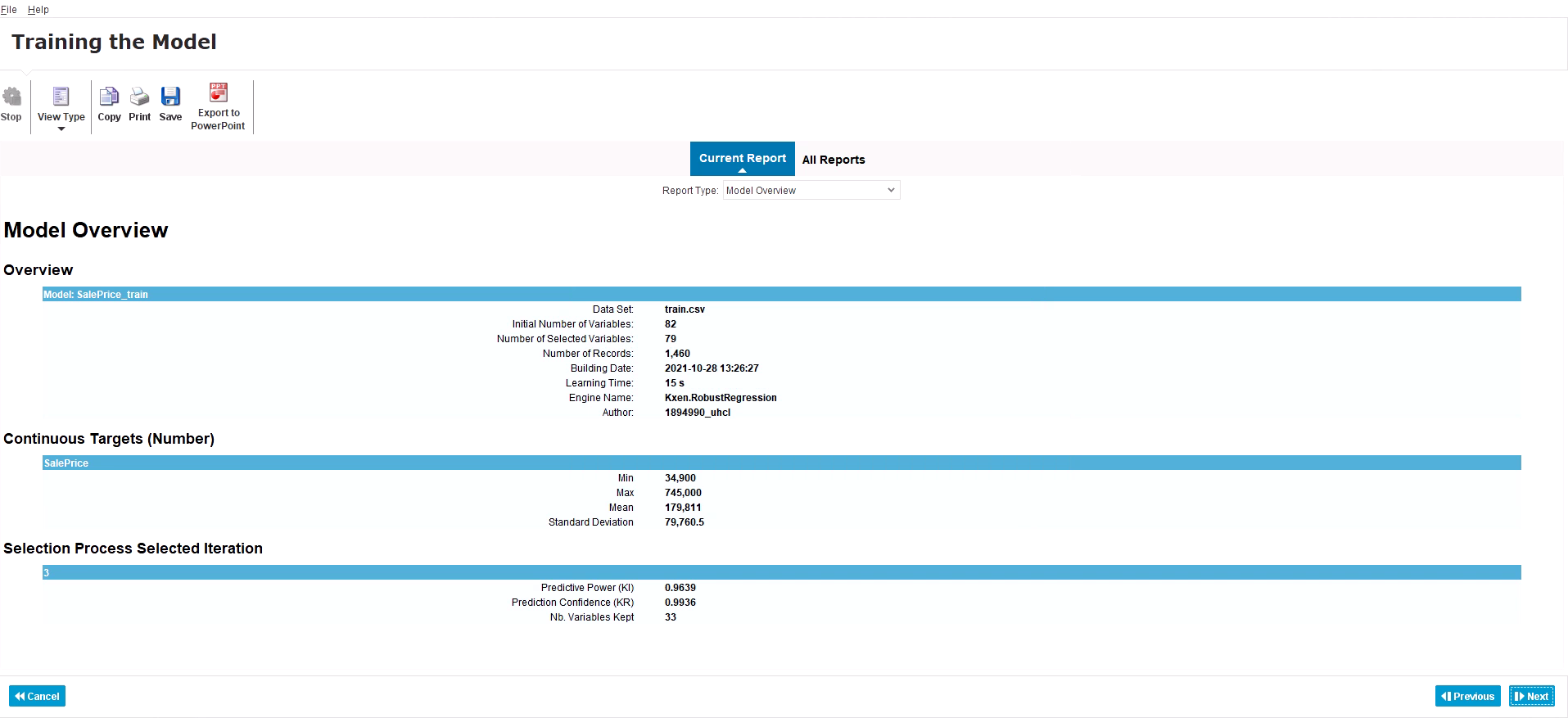
**Screenshot 1** - Load the raw training dataset without any further data cleaning.



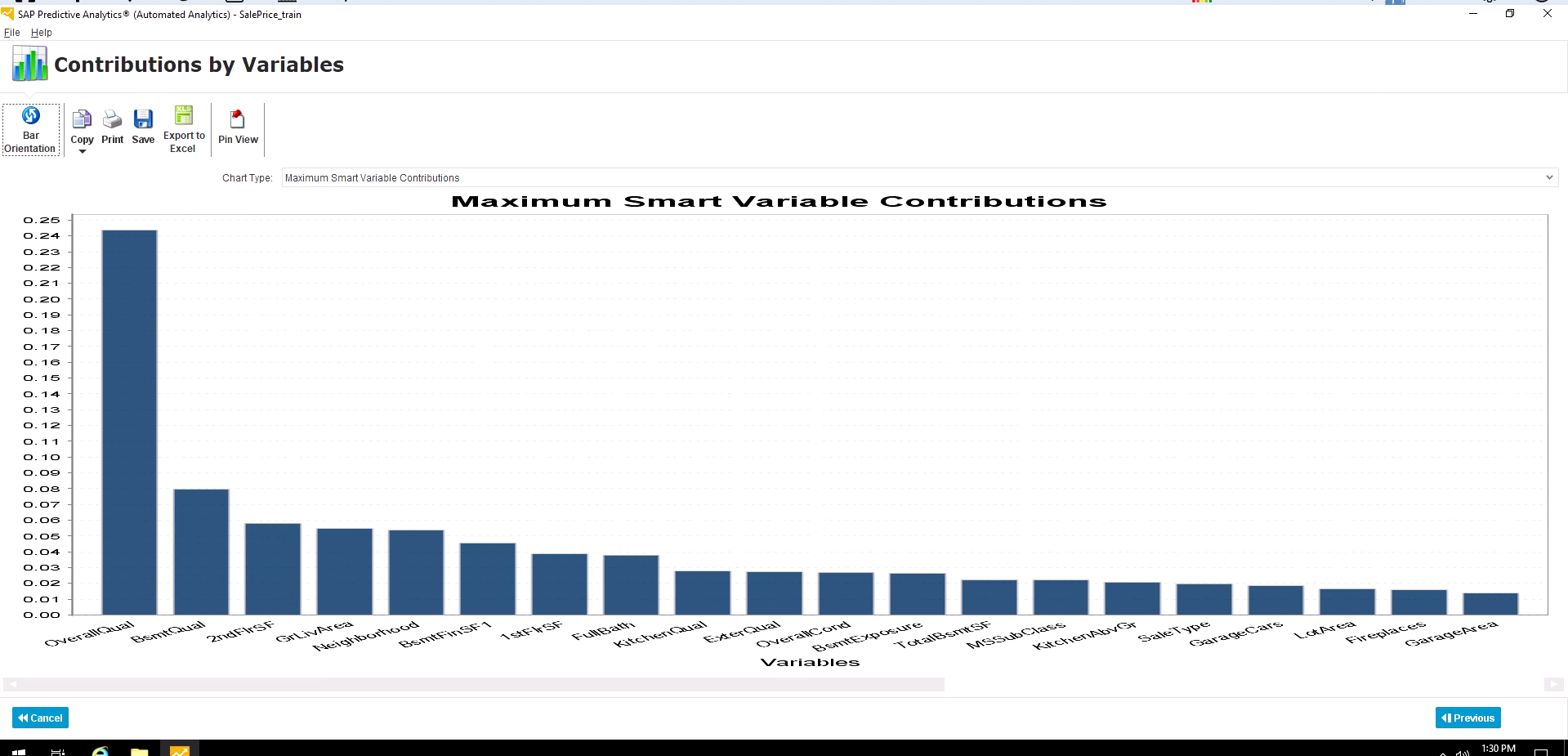
**Screenshot 2** - Selected variables for your regression model.



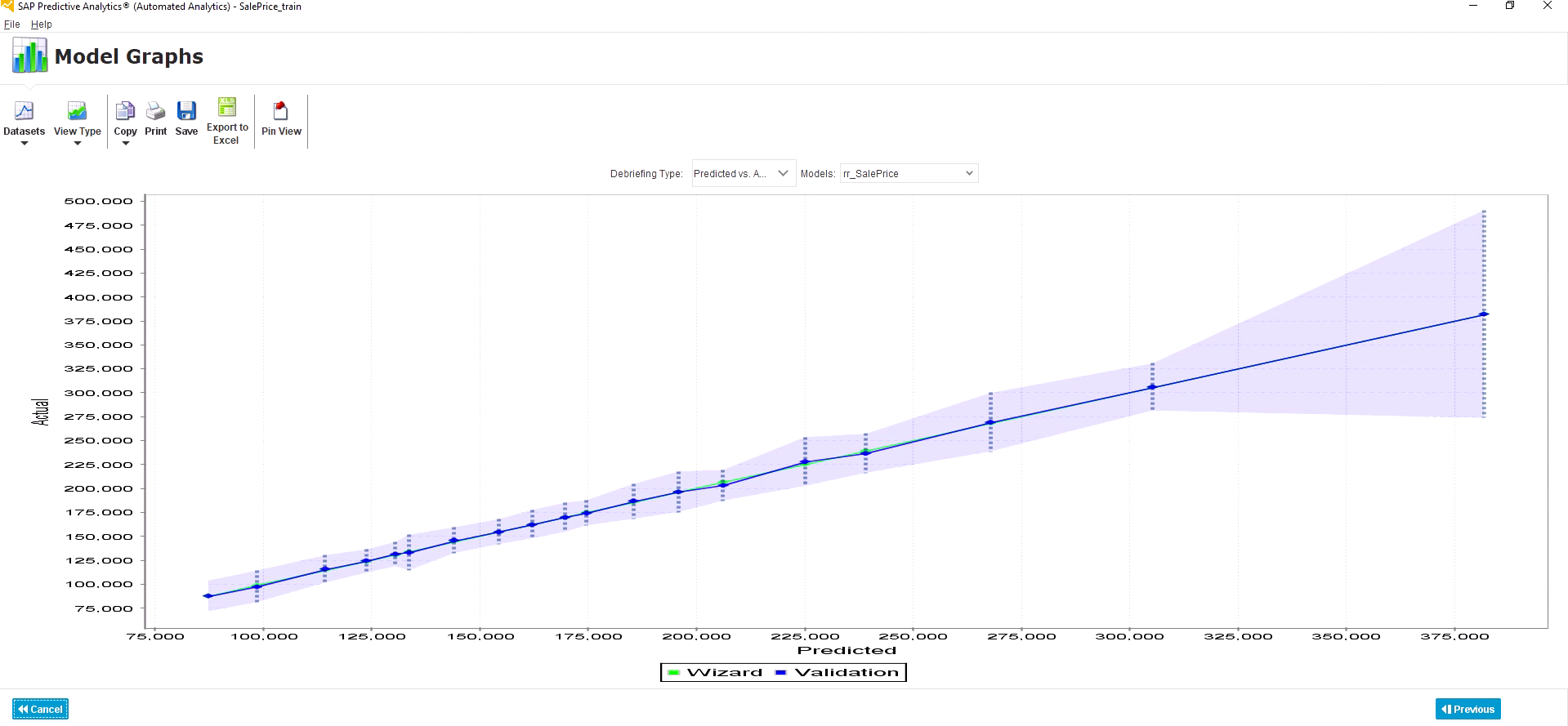
**Screenshot 3** - Warning Message of the model says, Fence has found significant deviation on data validation.



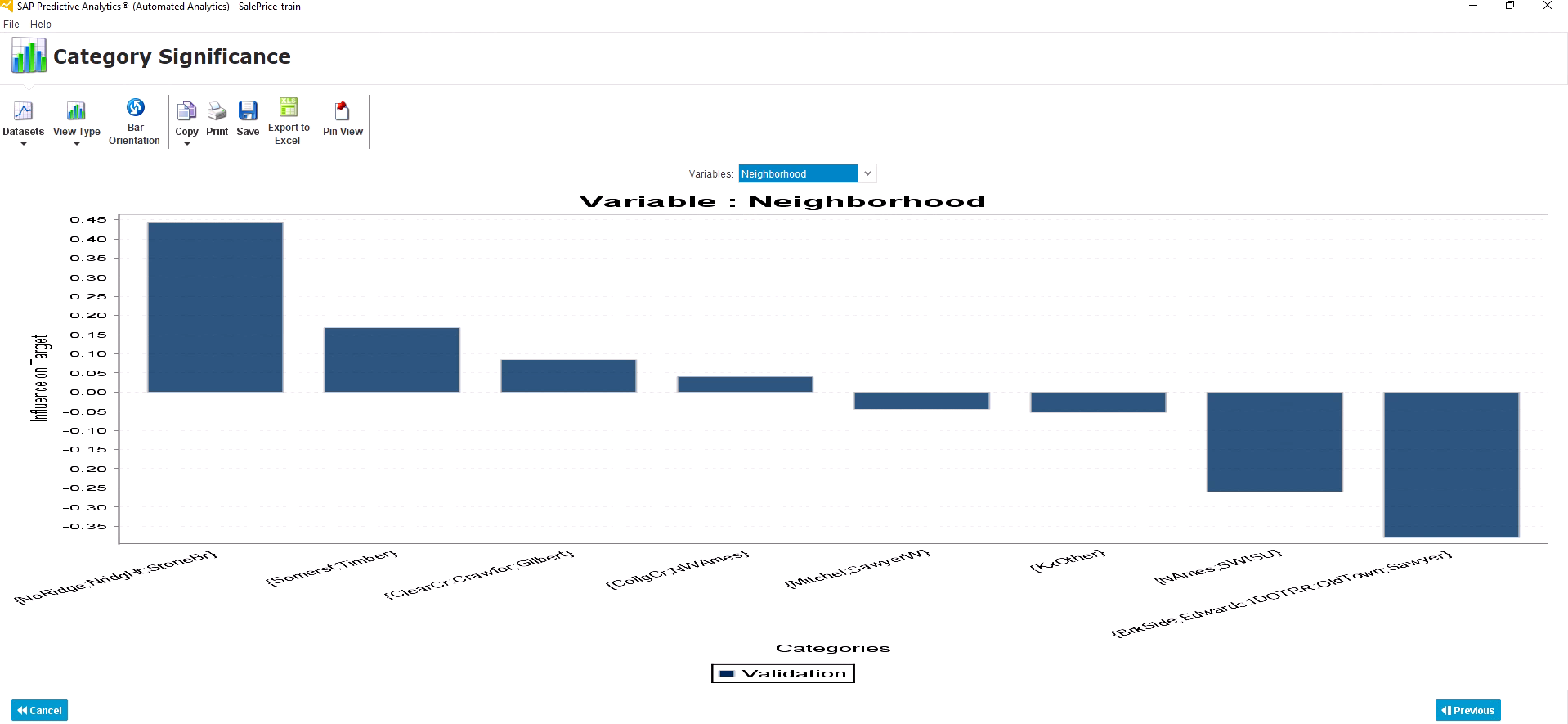
**Screenshot 4** - Model Overview.



**Screenshot 5** - Model Contribution by Variables.

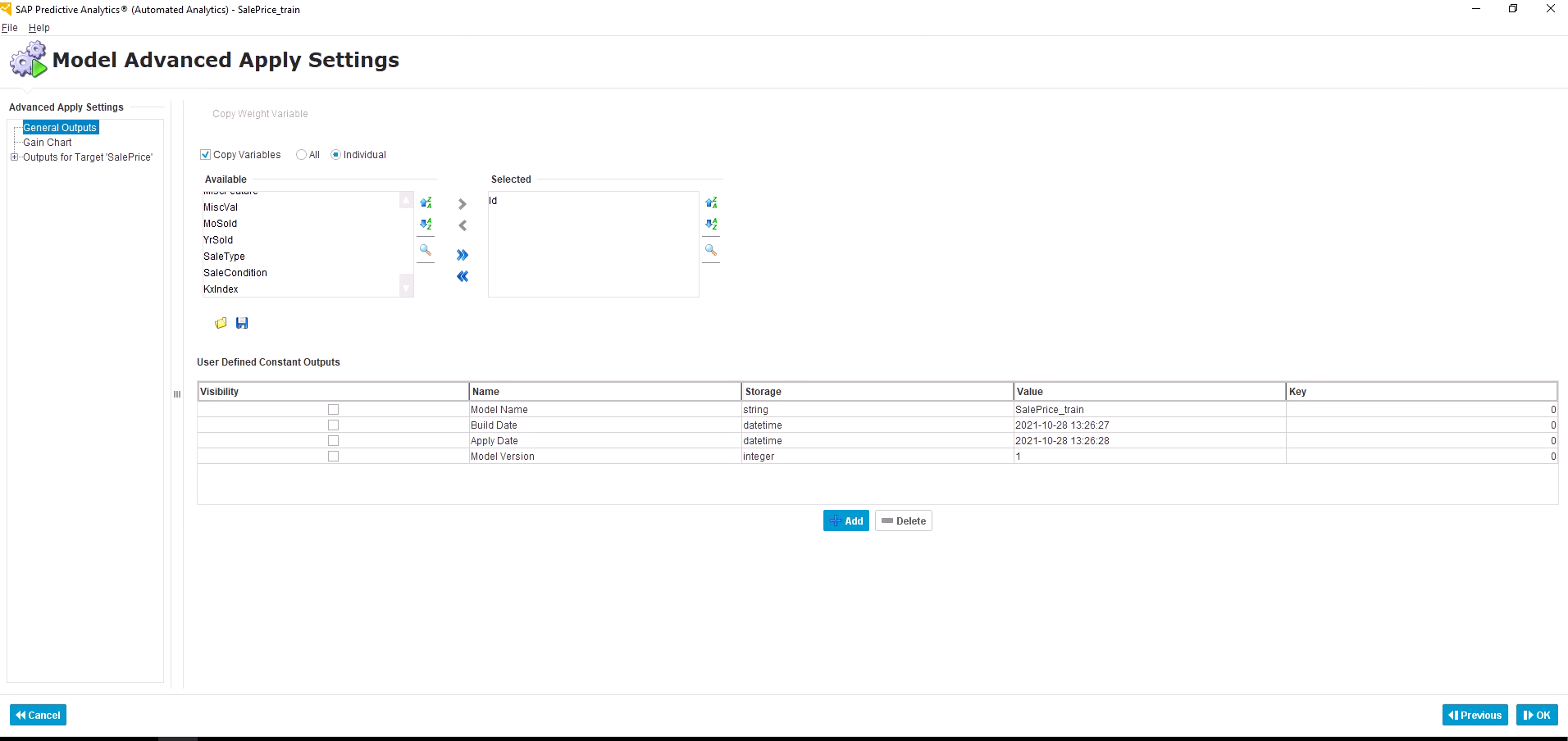


**Screenshot - 6** Model Graph shows that your model has better performance when house price < 200k and the performance decreases while the house price goes up. Our explanations that the house price goes up exponentially, which does not fit into the linear regression model.



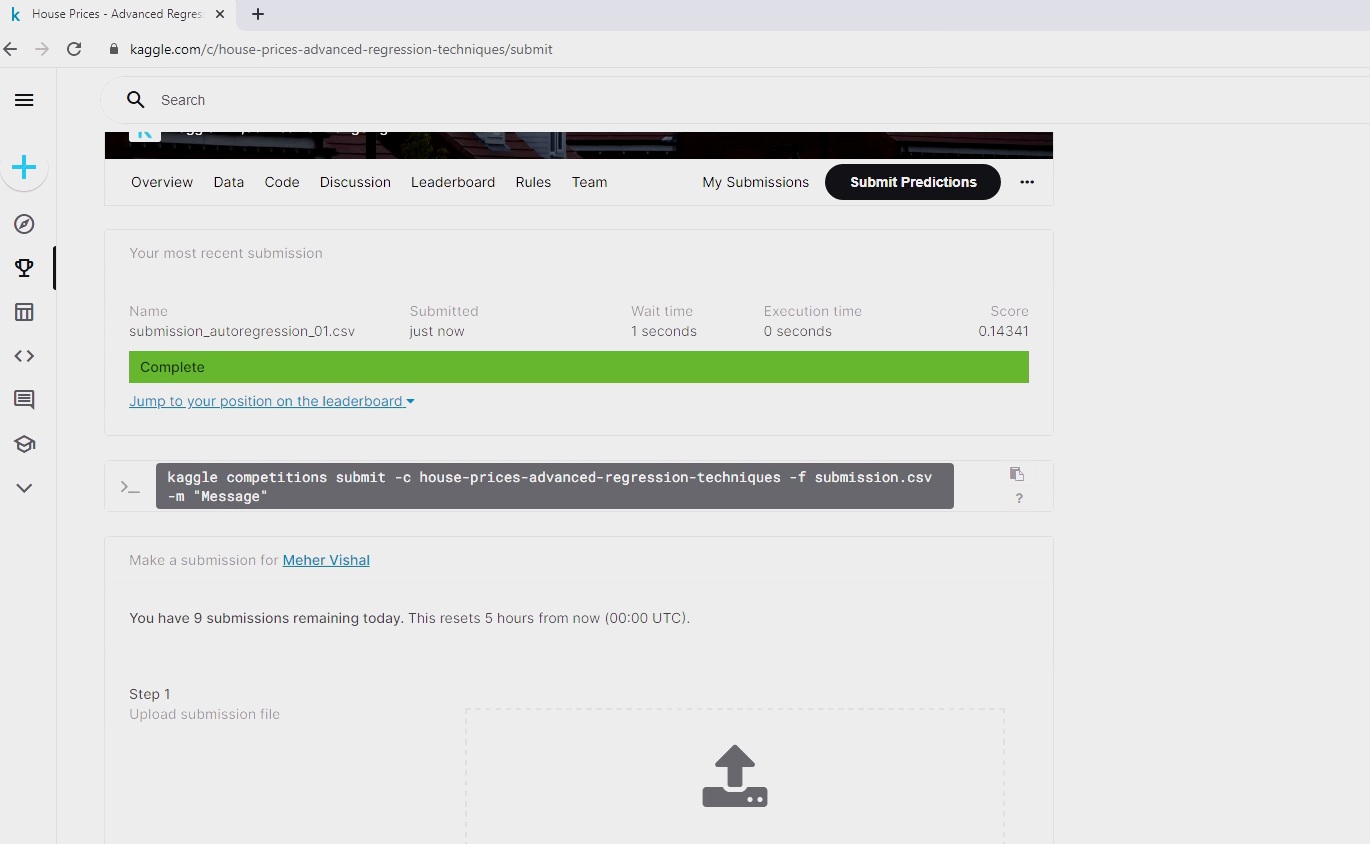
**Screenshot 7** - Neighbourhood category significance

After we build the regression model, we can apply our model to a testing dataset to predict our house prices.

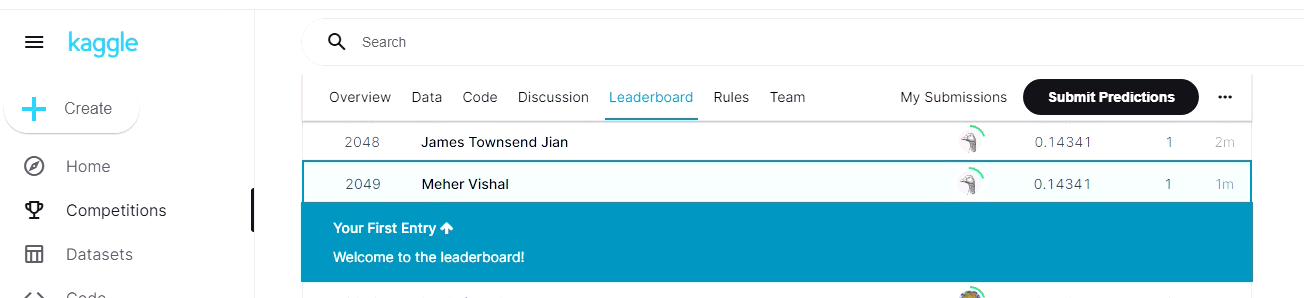


**Screenshot 8** - Configure your model’s output for testing the dataset.

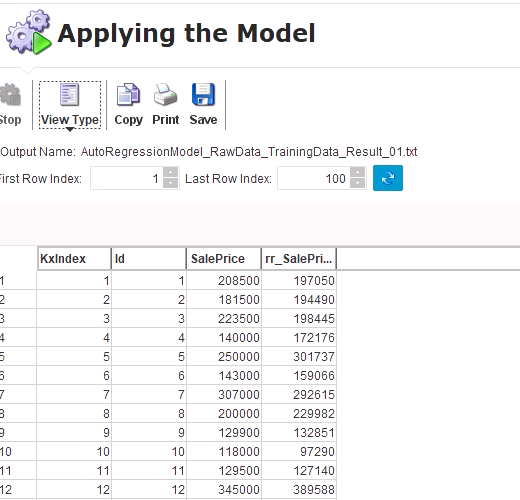
2. Prediction results, And your ranking screenshot from Kraggle.com.



**Screenshot 9** - Error Score

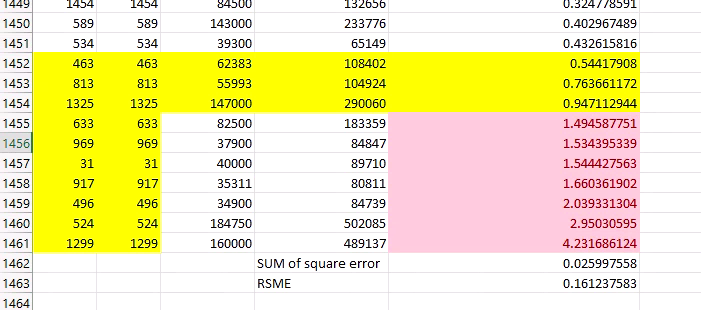


**Screenshot 10** - Kraggle Ranking - 2049



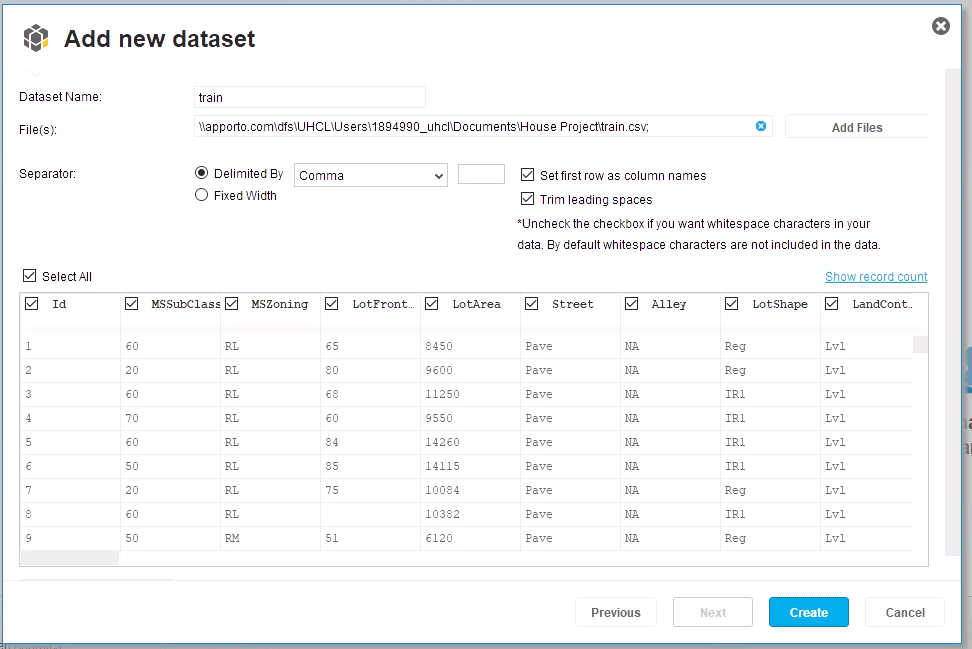
**Screenshot 11** - Model Prediction Sale Price for training dataset. We will use this output to calculate the model score (error rate).

1. Using Excel to calculate your Model’s error number.



**Screenshot 12** - RMSE Error Score with house outliers

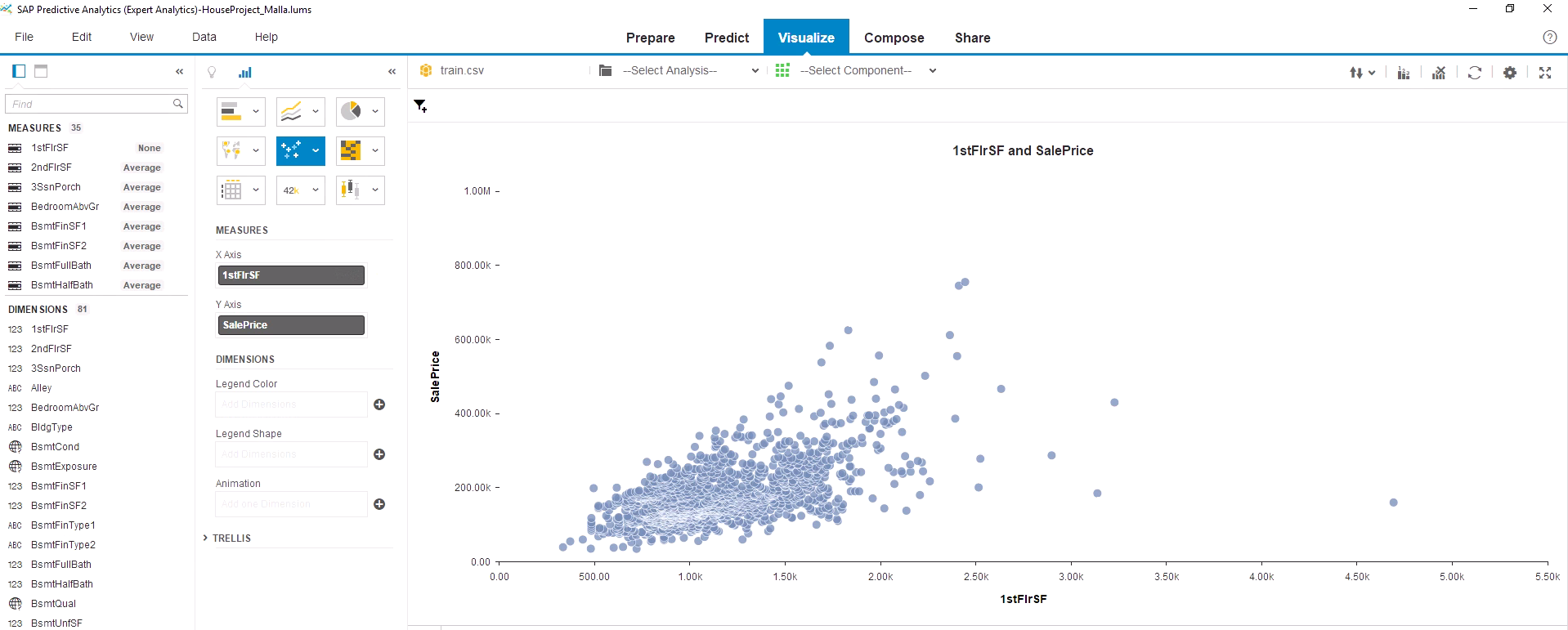
1. **Expert Analytic Regression Model.**



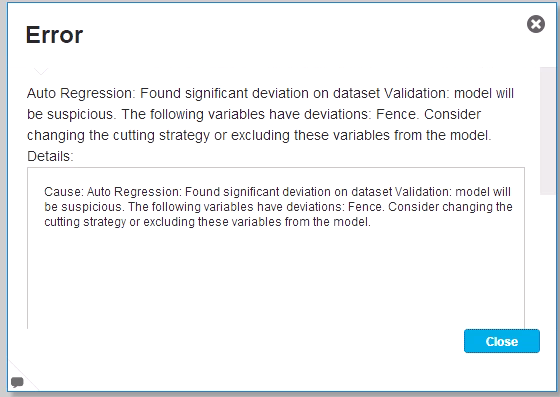
**Screenshot 13** Load Training Data Set



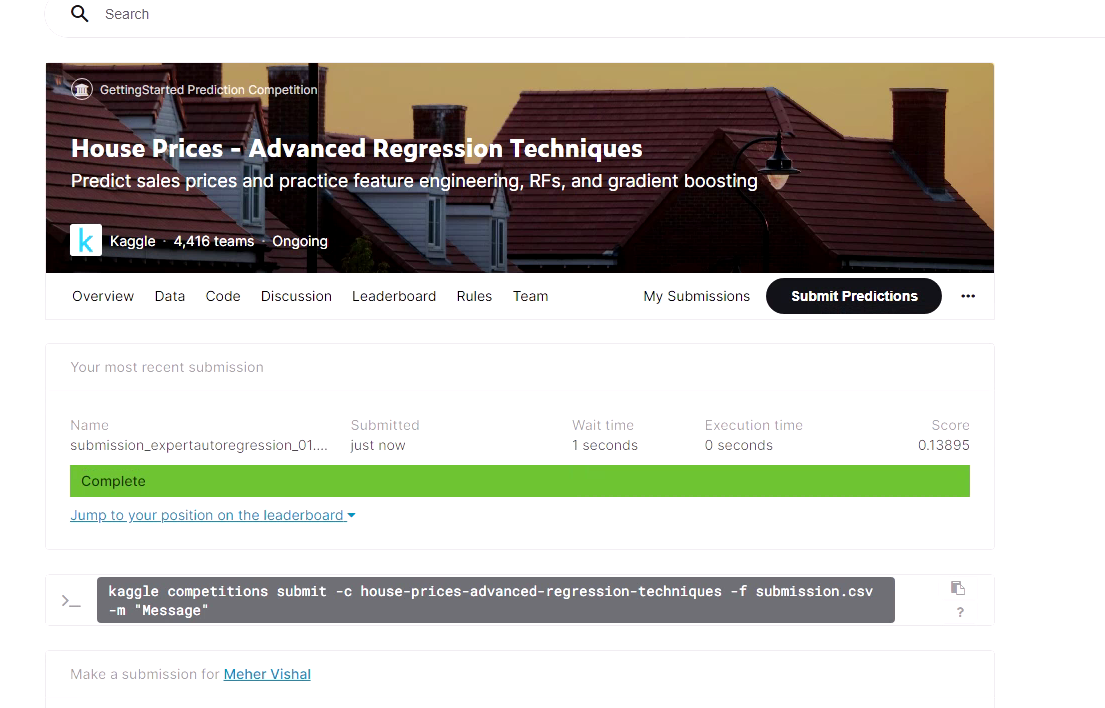
**Screenshot 14**- Correlation between Quality and House Price.



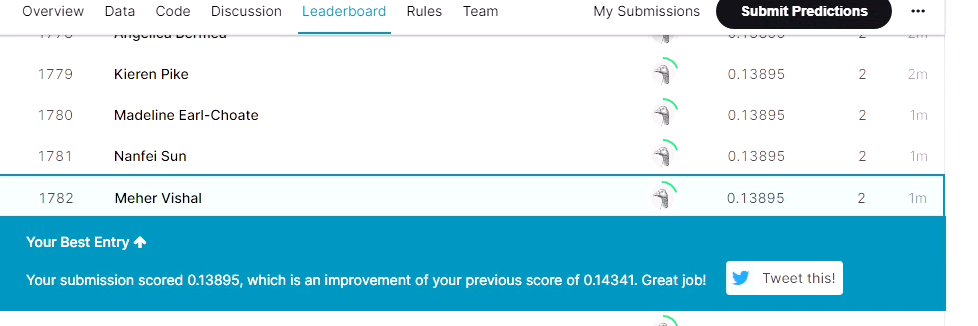
**Screenshot 15**- Correlation between 2ndflrSqt and House Price



**Screenshot 16** - Error in Fence Column.



**Screenshot 17** - Testing Result Submission in Kaggle.



**Screenshot 18** - Improved Kaggle Ranking-1782