ISAM 4391

Thanh Quang Phan – 2002680

**Final Project – House Prices – Final Submission**

Summary

The Phase II returned the result with RMSE of 0.13104. Five more models were run using the clean data and different combinations of Score Bin Count and Correlation Value.

The data cleaning process included:

* Deleting columns with more than 1,000 NA values including Alley, PoolQC, Fence, and MiscFeature.
* Splitting LotShape column into four separate columns Regular, IR1, IR2, and IR3 with value of 1 (means Yes) and 0 (means No).
* Deleting LotShape column.

The model runs are summery as below table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Polynomial | Score Bin Count | Correlation | Top 3 Contributors | RMSE |
| Phase II Model | **2** | **20** | **0.50** | **- OverallQual - Fireplaces - Neighborhood** | **0.13104** |
| Model 1 | 2 | 30 | 0.50 | - Neighborhood - 1stFlrSF - TotalBsmtSF | 0.13225 |
| Model 2 | 2 | 45 | 0.60 | - Neighborhood - GrLivArea - OpenPorchSF | 0.13166 |
| Model 3 | 2 | 30 | 0.60 | - OverallQual - Fireplaces - Neighborhood | 0.12884 |
| Model 4 | 2 | 50 | 0.60 | - OverallQual - Fireplaces - Neighborhood | 0.12677 |
| Model 5 | **2** | **40** | **0.60** | **- Neighborhood - GrLivingArea - 2ndFlrSF** | **0.12457** |

The above runs showed that the larger Score Bin Count and larger Correlation do not always give a better result.

1. Model 01

Graphical user interface, text, application

Description automatically generated

Screenshot . Advanced Model Parameters: Polynomial Degree: 2, Score Bin Count: 30, Correlation Settings: 0.50

Table

Description automatically generated with low confidence

Screenshot . Model Overview

Chart, line chart

Description automatically generated

Screenshot . Model Graph

Chart, bar chart

Description automatically generated

Screenshot . Top 3 Contributors: Neighborhood, 1stFlrSF, TotalBsmtSF

Graphical user interface, text, application, email

Description automatically generated

Screenshot . RMSE is 0.13225, which is not an improvement from the Phase II Model.

1. Model 02

Graphical user interface, text, application, email

Description automatically generated

Screenshot . Advanced Model Parameters: Polynomial Degree: 2, Score Bin Count: 45, Correlation Settings: 0.60

A picture containing table

Description automatically generated

Screenshot . Model Overview

Chart, line chart

Description automatically generated

Screenshot . Model Graph

Chart, bar chart

Description automatically generated

Screenshot . Top 3 contributors are Neighborhood, GrLivArea, OpenPorchSF

A picture containing logo

Description automatically generated

Screenshot . RMSE is 0.13166. This is pretty close to the Phase II Model, but not an improvement.

1. Model 03

Graphical user interface, text, application

Description automatically generated

Screenshot . Advanced Model Parameters: Polynomial Degree: 2, Score Bin Count: 30, Correlation Settings: 0.60

A picture containing table

Description automatically generated

Screenshot . Model Overview

Chart, line chart

Description automatically generated

Screenshot . Model Graph

Chart, bar chart

Description automatically generated

Screenshot . Top 3 contributors are OverallQual, Fireplaces, Neighborhood

Graphical user interface, text, application

Description automatically generated

Graphical user interface

Description automatically generated with low confidence

Screenshot . RMSE is 0.12884, which is an improvement from the Phase II Model

1. Model 04

Graphical user interface, text, application, email

Description automatically generated

Screenshot . Advanced Model Parameters: Polynomial Degree: 2, Score Bin Count: 50, Correlation Settings: 0.6

Graphical user interface, application

Description automatically generated

Screenshot . Model Overview

Chart, line chart

Description automatically generated

Screenshot . Model Graph

Chart, bar chart

Description automatically generated

Screenshot . Top 3 contributors are OverallQual, Fireplaces, and Neighborhood

Graphical user interface, text, application, email

Description automatically generated

Screenshot . RMSE is 0.12677. This is better than Model 04.

1. Model 05 – Best Result

Graphical user interface, text, application, email

Description automatically generated

Screenshot . Advanced Model Parameters: Polynomial Degree: 2, Score Bin Count: 40, Correlation Settings: 0.6

A picture containing graphical user interface

Description automatically generated

Screenshot . Model Overview

Chart, line chart

Description automatically generated

Screenshot . Model Graph

Chart, bar chart

Description automatically generated

Screenshot . Top 3 contributors are Neighborhood, GrLivArea, 2ndFlrSF

Graphical user interface, text, application, email

Description automatically generated

Screenshot . RMSE is 0.12457. This is the best result I got.

Graphical user interface, text, application, email

Description automatically generated

Screenshot : My highest ranking.