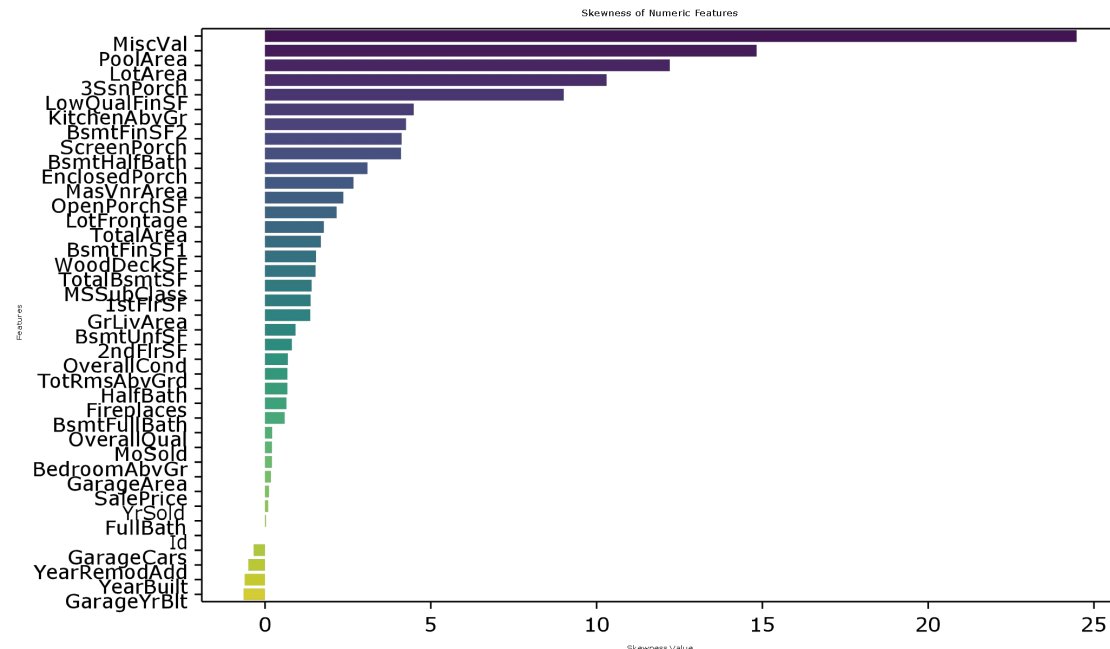


The project focuses on applying multiple linear regression models to predict house prices. Various methods are adopted in data processing. Firstly, columns with a high missing value rate (greater than 50%) are filtered out and excluded from the dataset.

In addition, to decrease white noise caused by features, we use feature engineering to combine the “TotalBsmtSF”, “1stFlrSF” and “2ndFlrSF” into a new variable “TotalArea”. The purpose is to enhance the significance of relevant columns to prevent type II error which indicates incorrectly concluding insignificance for variables that are significant.

For deeper data cleaning, we separate data into numeric and categorical. As shown in Plot 1, it can be observed that many numerical variables have high skewness. To make the variables more consistent with the assumption of the linear regression model about the normal distribution, variables with high skewness are converted to logarithmic data. After transforming the variables into a logarithmic scale, the values of outliers will be reduced, ensuring that the data better conforms to a normal distribution.

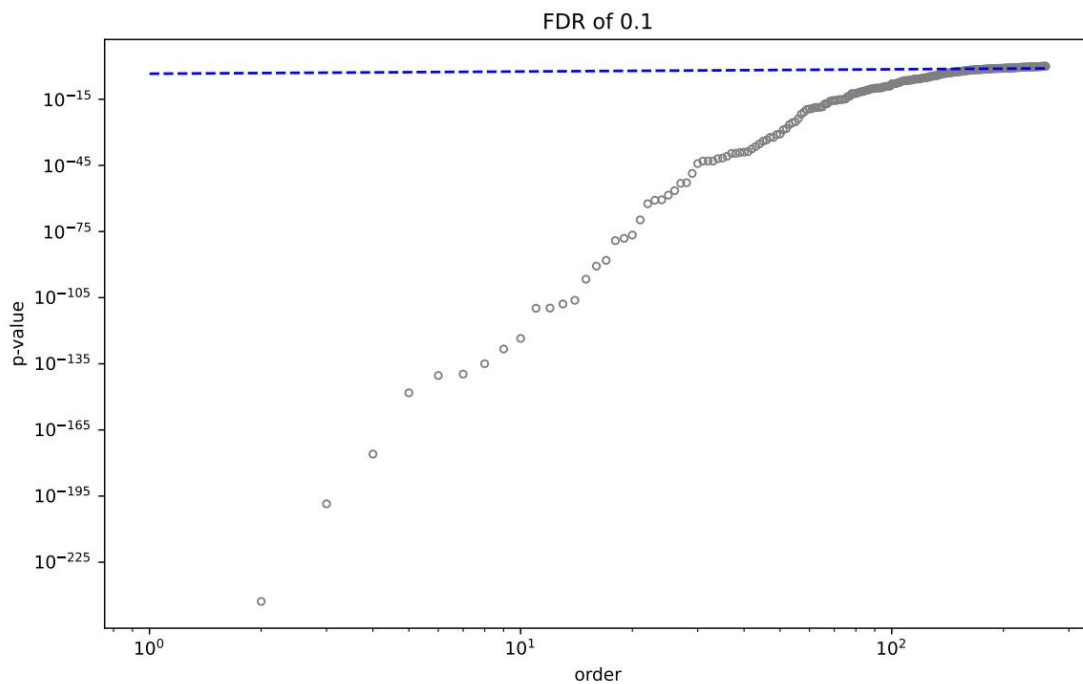


Plot 1: Skewness Values of Numerical Variables

Subsequently, null numerical variables including the “TotalArea” are filled with the median of each column. The null categorical variable values are filled with mode, which is the most frequent value in the dataset of each column. Then all categorical

variables are converted to dummy variables to make them accessible in regression.

Finally, by applying the Benjamini-Hochberg method, False Discovery Rate Curve (shown in plot 2) is used to control the proportion of type I errors that falsely reject null hypotheses during multiple hypothesis testing among all the rejections. In this project, we rejected all variables to be higher than the threshold of 0.1.



Plot 2: False Discovery Rate Curve


In conclusion, through a series of processing of the data, we have obtained an efficient dataset. Then multiple linear regression models are trained and predicted. Lastly, the regression result is posted as an appendix.

Python Code Link:

https://colab.research.google.com/drive/1rRZktfdaRls9T0t7DZJLK2RmeAlHmVcB?usp=drive_link

Kaggle Score:

Submissions

		Recent
All		
Successful		
Errors		
Submission and Description		Public Score ⓘ
	Assignment 2.csv Complete · now	0.13914

Appendix

	coef	std err	t	P> t	[0.025	0.975]
Exterior2nd_CmentBd	0.0141	0.074	0.191	0.848	-0.13	0.158
ExterCond_Fa	-0.0475	0.031	-1.557	0.12	-0.107	0.012
PoolArea	-0.0032	0.008	-0.405	0.685	-0.019	0.012
Condition1_Feodr	-0.0087	0.025	-0.351	0.725	-0.057	0.04
Condition1_PosA	0.0449	0.048	0.927	0.354	-0.05	0.14
PavedDrive_P	0.4715	0.11	4.287	0	0.256	0.687
PavedDrive_Y	0.4979	0.11	4.516	0	0.282	0.714
MasVnrType_Stone	0.0567	0.035	1.628	0.104	-0.012	0.125
Neighborhood_StoneBr	0.1619	0.028	5.695	0	0.106	0.218
BsmtExposure_Gd	0.3995	0.083	4.816	0	0.237	0.562
BsmtQual_TA	0.3461	0.081	4.251	0	0.186	0.506
ExterQual_TA	0.3494	0.082	4.26	0	0.188	0.51
Electrical_FuseP	-0.0294	0.217	-0.135	0.892	-0.455	0.397
GarageType_Detchd	0.0401	0.028	1.454	0.146	-0.014	0.094
1stFlrSF	0.0472	0.049	0.962	0.336	-0.049	0.144
GarageArea	-6.14E-06	3.73E-05	-0.164	0.869	-7.93E-05	6.71E-05
LotFrontage	0.0009	0.017	0.053	0.958	-0.032	0.034
Exterior2nd_AsbShng	-0.0265	0.057	-0.463	0.643	-0.139	0.086
Neighborhood_Sawyer	-0.0032	0.019	-0.163	0.871	-0.041	0.035
MasVnrArea	-0.0068	0.006	-1.204	0.229	-0.018	0.004
BldgType_1Fam	0.039	0.019	2.105	0.036	0.003	0.075
Neighborhood_NWAmes	-0.0037	0.02	-0.19	0.85	-0.042	0.035
ExterQual_Ex	0.3573	0.084	4.257	0	0.193	0.522
GarageFinish_RFn	0.4845	0.109	4.447	0	0.271	0.698
LotArea	0.0569	0.012	4.617	0	0.033	0.081
MSZoning_C (all)	-0.0516	0.077	-0.671	0.503	-0.203	0.099
LowQualFinSF	-0.0013	0.005	-0.252	0.801	-0.011	0.009
YearBuilt	0.0006	0	1.574	0.116	0	0.001
TotRmsAbvGrd	0.0053	0.005	1.183	0.237	-0.004	0.014
Exterior1st_BrkComm	-0.2172	0.109	-1.986	0.047	-0.432	-0.003
Heating_Grav	-0.2476	0.062	-4.021	0	-0.368	-0.127
GarageType_BuiltIn	0.021	0.031	0.676	0.499	-0.04	0.082
KitchenAbvGr	-0.2031	0.067	-3.035	0.002	-0.334	-0.072
Exterior2nd_MetalSd	-0.0054	0.042	-0.127	0.899	-0.088	0.078
Exterior1st_CemntBd	-0.0302	0.073	-0.416	0.678	-0.173	0.112
Electrical_SBrkr	-0.0633	0.214	-0.296	0.767	-0.483	0.356
FireplaceQu_Gd	-0.0011	0.022	-0.049	0.961	-0.045	0.043
GarageCond_Po	0.04	0.08	0.499	0.618	-0.117	0.197
Functional_Min1	0.0298	0.033	0.911	0.362	-0.034	0.094
RoofMatl_CompShg	0.2151	0.047	4.58	0	0.123	0.307
GarageQual_Po	-0.1157	0.124	-0.933	0.351	-0.359	0.128
PavedDrive_N	0.4829	0.109	4.425	0	0.269	0.697
Condition1_Norm	0.0506	0.021	2.441	0.015	0.01	0.091
Electrical_FuseA	-0.0455	0.213	-0.213	0.831	-0.464	0.373
Neighborhood_NAmes	0.002	0.017	0.12	0.904	-0.031	0.035
Neighborhood_Timber	0.0264	0.024	1.091	0.275	-0.021	0.074

Heating_GasA	-0.0733	0.031	-2.338	0.02	-0.135	-0.012
Foundation_Slab	-0.0187	0.062	-0.299	0.765	-0.141	0.104
BsmtFinType1_Rec	0.2357	0.055	4.249	0	0.127	0.344
Functional_Min2	0.0352	0.033	1.084	0.279	-0.029	0.099
BsmtFinType1_ALQ	0.2556	0.056	4.593	0	0.146	0.365
GarageType_Attchd	0.0335	0.027	1.242	0.214	-0.019	0.086
Neighborhood_CollgCr	0.0231	0.016	1.43	0.153	-0.009	0.055
Neighborhood_Veenker	0.0662	0.039	1.701	0.089	-0.01	0.143
WoodDeckSF	0.004	0.001	2.817	0.005	0.001	0.007
3SsnPorch	0.007	0.005	1.46	0.145	-0.002	0.016
MiscVal	-0.0007	0.003	-0.252	0.801	-0.006	0.005
HouseStyle_1.5Unf	0.0272	0.041	0.669	0.503	-0.052	0.107
SaleType_WD	-0.0519	0.028	-1.831	0.067	-0.107	0.004
HouseStyle_SFoyer	0.0233	0.027	0.858	0.391	-0.03	0.077
GarageQual_Fa	-0.1609	0.042	-3.872	0	-0.242	-0.079
YearRemodAdd	0.0015	0	5.971	0	0.001	0.002
LandContour_Bnk	-0.0465	0.017	-2.684	0.007	-0.081	-0.013
KitchenQual_Gd	0.3493	0.083	4.197	0	0.186	0.513
HeatingQC_Fa	0.1312	0.13	1.01	0.313	-0.124	0.386
Condition1_PosN	-0.0093	0.035	-0.267	0.79	-0.077	0.059
FullBath	0.0209	0.011	1.973	0.049	0	0.042
HeatingQC_TA	0.1212	0.128	0.944	0.345	-0.131	0.373
SaleCondition_Partial	-0.0583	0.075	-0.779	0.436	-0.205	0.089
Foundation_CBBlock	-0.011	0.044	-0.251	0.801	-0.097	0.075
Exterior2nd_Brk Cmn	0.0306	0.059	0.516	0.606	-0.086	0.147
HeatingQC_Gd	0.1388	0.129	1.079	0.281	-0.113	0.391
LotShape_IR2	0.1176	0.044	2.663	0.008	0.031	0.204
Exterior2nd_Wd Shng	-0.0124	0.028	-0.445	0.656	-0.067	0.042
OverallQual	0.0443	0.005	9.052	0	0.035	0.054
KitchenQual_TA	0.3341	0.081	4.101	0	0.174	0.494
SaleType_COD	-0.0427	0.034	-1.243	0.214	-0.11	0.025
Neighborhood_BrkSide	0.0529	0.028	1.892	0.059	-0.002	0.108
BsmtFinSF1	0.0086	0.007	1.292	0.197	-0.004	0.022
TotalBsmtSF	-0.0008	0.009	-0.086	0.932	-0.018	0.017
Neighborhood_BrDale	0.0479	0.044	1.097	0.273	-0.038	0.134
BsmtFinType1_GLQ	0.248	0.056	4.398	0	0.137	0.359
HouseStyle_1Story	0.005	0.018	0.275	0.783	-0.031	0.041
RoofStyle_Gambrel	-0.1464	0.055	-2.67	0.008	-0.254	-0.039
GarageFinish_Fin	0.4881	0.11	4.448	0	0.273	0.703
FireplaceQu_TA	0.001	0.022	0.043	0.966	-0.043	0.045
SaleCondition_Normal	0.0416	0.023	1.837	0.066	-0.003	0.086
SaleType_ConLD	0.0577	0.052	1.118	0.264	-0.044	0.159
Condition2_Feodr	0.0557	0.056	0.992	0.321	-0.054	0.166
Exterior2nd_Stucco	-0.0493	0.029	-1.702	0.089	-0.106	0.008
Exterior1st_Wd Sdng	-0.0835	0.021	-3.996	0	-0.124	-0.042
BsmtFinType1_BLQ	0.2358	0.055	4.26	0	0.127	0.344
GarageQual_TA	-0.0955	0.034	-2.811	0.005	-0.162	-0.029
BsmtFinType2_BLQ	-0.0461	0.022	-2.097	0.036	-0.089	-0.003
BsmtFullBath	0.0198	0.009	2.193	0.029	0.002	0.037

FireplaceQu_Po	-0.0005	0.035	-0.016	0.988	-0.068	0.067
GarageType_CarPort	0.0318	0.051	0.622	0.534	-0.069	0.132
BsmtFinType1_Unf	0.2582	0.062	4.182	0	0.137	0.379
Neighborhood_MeadowV	-0.0562	0.044	-1.282	0.2	-0.142	0.03
CentralAir_N	0.6886	0.163	4.234	0	0.37	1.008
Heating_Wall	-0.0309	0.075	-0.414	0.679	-0.177	0.115
Neighborhood_SWISU	0.0445	0.033	1.336	0.182	-0.021	0.11
Neighborhood_Gilbert	0.0156	0.02	0.77	0.442	-0.024	0.055
Fireplaces	0.0165	0.007	2.199	0.028	0.002	0.031
OpenPorchSF	0.0027	0.002	1.448	0.148	-0.001	0.006
BsmtQual_Fa	0.3675	0.081	4.546	0	0.209	0.526
HalfBath	0.0099	0.01	0.966	0.334	-0.01	0.03
MSZoning_RL	0.3689	0.066	5.567	0	0.239	0.499
BedroomAbvGr	-0.0005	0.007	-0.068	0.946	-0.013	0.013
Condition1_Artery	-0.029	0.028	-1.019	0.309	-0.085	0.027
BsmtCond_Po	-0.1639	0.135	-1.211	0.226	-0.429	0.102
MasVnrType_None	0.0161	0.044	0.369	0.712	-0.07	0.102
Electrical_FuseF	-0.0369	0.217	-0.17	0.865	-0.462	0.388
SaleCondition_Abnorml	-0.0333	0.026	-1.286	0.199	-0.084	0.018
LotConfig_Inside	-0.0077	0.008	-0.946	0.344	-0.024	0.008
BsmtUnfSF	-2.33E-05	1.86E-05	-1.251	0.211	-5.99E-05	1.33E-05
HouseStyle_1.5Fin	-0.0225	0.02	-1.149	0.251	-0.061	0.016
Exterior1st_AsbShng	-0.0035	0.059	-0.06	0.952	-0.12	0.112
Foundation_BrkTil	-0.0472	0.044	-1.066	0.287	-0.134	0.04
Neighborhood_ClearCr	0.06	0.028	2.117	0.034	0.004	0.116
Street_Grvl	0.6973	0.166	4.19	0	0.371	1.024
LandContour_HLS	0.0178	0.019	0.931	0.352	-0.02	0.055
SaleType_New	0.081	0.077	1.051	0.294	-0.07	0.232
GarageCars	0.0524	0.011	4.975	0	0.032	0.073
CentralAir_Y	0.7638	0.165	4.633	0	0.44	1.087
Neighborhood_OldTown	-0.0201	0.028	-0.715	0.474	-0.075	0.035
BsmtExposure_Av	0.3551	0.082	4.326	0	0.194	0.516
KitchenQual_Fa	0.3451	0.082	4.191	0	0.184	0.507
BsmtCond_Fa	-0.0706	0.021	-3.42	0.001	-0.111	-0.03
LotShape_Reg	0.1006	0.042	2.42	0.016	0.019	0.182
Exterior2nd_Wd Sdng	0.0573	0.021	2.696	0.007	0.016	0.099
BsmtExposure_No	0.3383	0.082	4.132	0	0.178	0.499
HeatingQC_Ex	0.1567	0.128	1.22	0.223	-0.095	0.409
EnclosedPorch	0.0013	0.002	0.615	0.538	-0.003	0.006
FireplaceQu_Ex	0.0129	0.034	0.384	0.701	-0.053	0.079
Neighborhood_NridgHt	0.1282	0.022	5.824	0	0.085	0.171
RoofMatl_WdShake	0.1803	0.066	2.731	0.006	0.051	0.31
MasVnrType_BrkFace	0.0479	0.033	1.462	0.144	-0.016	0.112
Exterior2nd_VinylSd	0.027	0.035	0.772	0.44	-0.042	0.096
Foundation_PConc	-0.0172	0.044	-0.393	0.695	-0.103	0.069
2ndFlrSF	6.58E-05	3.30E-05	1.995	0.046	1.09E-06	0
Functional_Maj2	-0.2487	0.066	-3.751	0	-0.379	-0.119
Exterior1st_VinylSd	-0.0338	0.035	-0.959	0.338	-0.103	0.035
BldgType_Duplex	0.0095	0.035	0.269	0.788	-0.06	0.079

RoofStyle_Gable	-0.1265	0.039	-3.227	0.001	-0.203	-0.05
HouseStyle_2Story	-0.0465	0.02	-2.319	0.021	-0.086	-0.007
Neighborhood_NoRidge	0.1344	0.025	5.424	0	0.086	0.183
GrLivArea	0.1985	0.075	2.651	0.008	0.052	0.345
Neighborhood_IDOTRR	-0.0002	0.036	-0.005	0.996	-0.072	0.071
ExterCond_TA	-0.0015	0.011	-0.133	0.894	-0.024	0.021
LotShape_IR1	0.0952	0.041	2.311	0.021	0.014	0.176
BsmtQual_Ex	0.3869	0.086	4.522	0	0.219	0.555
Functional_Typ	0.0966	0.026	3.785	0	0.047	0.147
Neighborhood_Edwards	-0.0661	0.019	-3.446	0.001	-0.104	-0.028
ScreenPorch	0.0098	0.002	4.157	0	0.005	0.014
Neighborhood_Somerst	0.0813	0.031	2.606	0.009	0.02	0.142
GarageCond_TA	-0.0242	0.041	-0.582	0.56	-0.106	0.057
MSZoning_RH	0.4029	0.072	5.632	0	0.263	0.543
BldgType_Twnhs	-0.0237	0.025	-0.945	0.345	-0.073	0.026
MSZoning_RM	0.3416	0.067	5.061	0	0.209	0.474
ExterQual_Gd	0.3589	0.083	4.338	0	0.197	0.521
Exterior1st_WdShing	-0.0309	0.031	-0.981	0.327	-0.093	0.031
ExterCond_Po	-0.2449	0.126	-1.943	0.052	-0.492	0.002
GarageFinish_Unf	0.4798	0.109	4.409	0	0.266	0.693
Exterior1st_MetalSd	0.0026	0.042	0.062	0.951	-0.079	0.085
Neighborhood_Crawfor	0.1629	0.025	6.636	0	0.115	0.211
Exterior2nd_HdBoard	0.0086	0.02	0.422	0.673	-0.031	0.049
BsmtExposure_Mn	0.3595	0.082	4.392	0	0.199	0.52
LotConfig_CulDSac	0.0374	0.016	2.365	0.018	0.006	0.068
Exterior1st_HdBoard	-0.0425	0.022	-1.959	0.05	-0.085	6.75E-05
MSZoning_FV	0.3906	0.071	5.492	0	0.251	0.53
BsmtFinType1_LwQ	0.219	0.055	3.958	0	0.11	0.328
Street_Pave	0.7551	0.166	4.562	0	0.43	1.08
ExterQual_Fa	0.3868	0.089	4.324	0	0.211	0.562
SaleCondition_AdjLand	0.0584	0.075	0.783	0.434	-0.088	0.205
KitchenQual_Ex	0.4239	0.084	5.028	0	0.259	0.589
RoofStyle_Hip	-0.1141	0.04	-2.854	0.004	-0.193	-0.036
MoSold	-0.0011	0.001	-0.938	0.349	-0.003	0.001
RoofMatl_WdShngl	0.3655	0.069	5.304	0	0.23	0.501
GarageYrBlt	-0.0005	0	-1.994	0.046	-0.001	-8.71E-06
TotalArea	0.2042	0.076	2.679	0.007	0.055	0.354
GarageCond_Fa	-0.0643	0.048	-1.334	0.182	-0.159	0.03
BsmtQual_Gd	0.3519	0.084	4.202	0	0.188	0.516
BsmtCond_Gd	0.0299	0.016	1.903	0.057	-0.001	0.061
BldgType_2fmCon	0.0261	0.035	0.741	0.459	-0.043	0.095

Table1: OLS Regression Results