8. Appendix

All relevant figures are included here and some are also in main text, but the ones in the appendix are higher resolution. This summarizes the all the output from the .ipynb file to high-definition images and is also grouped together by type of graph instead of by regression method.

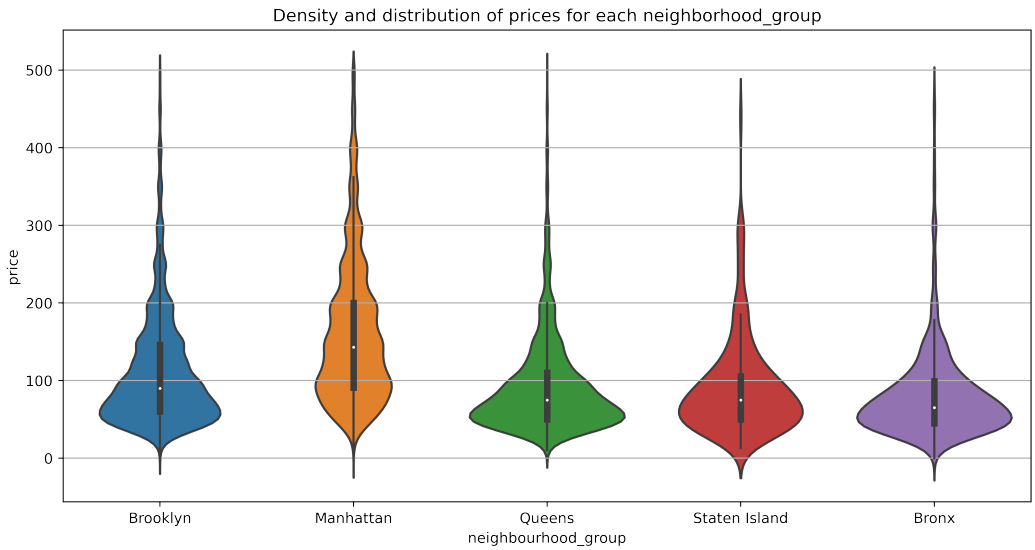


Figure 1: Density

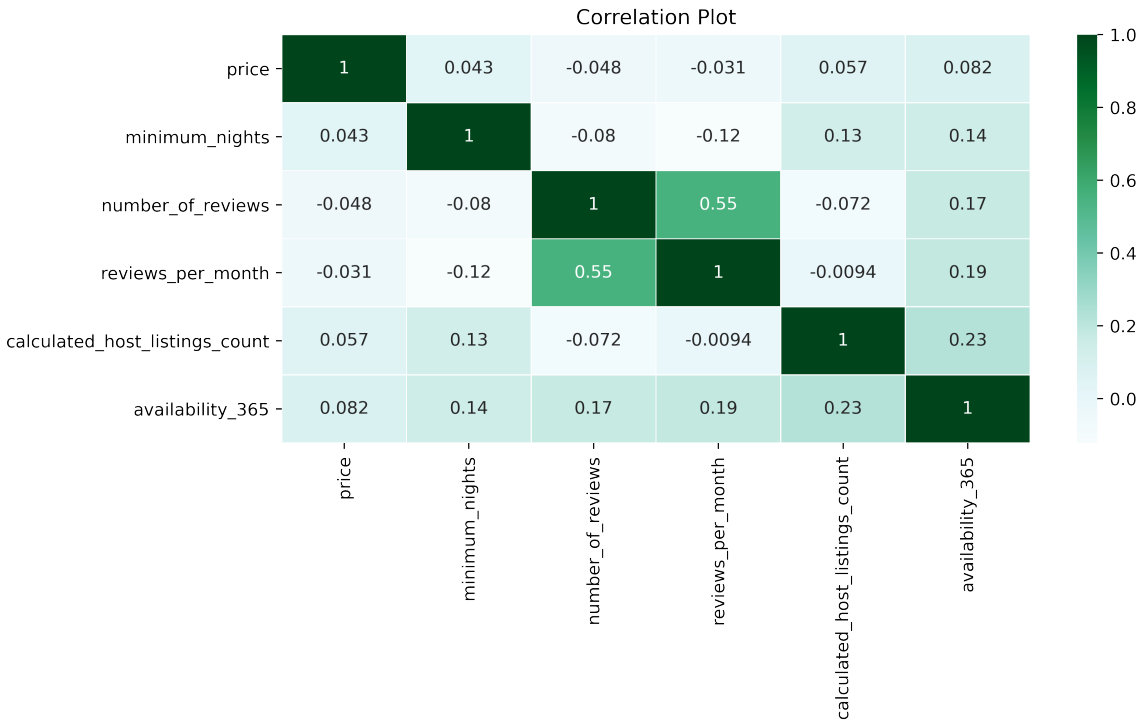


Figure 2: Correlation

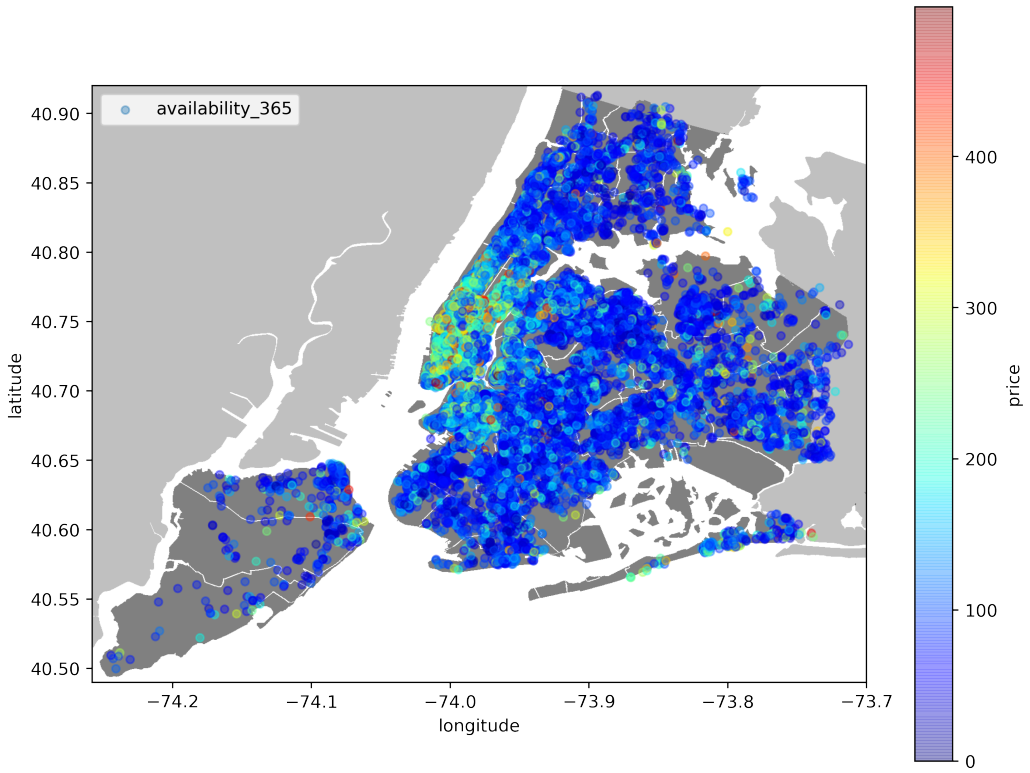


Figure 3: Map

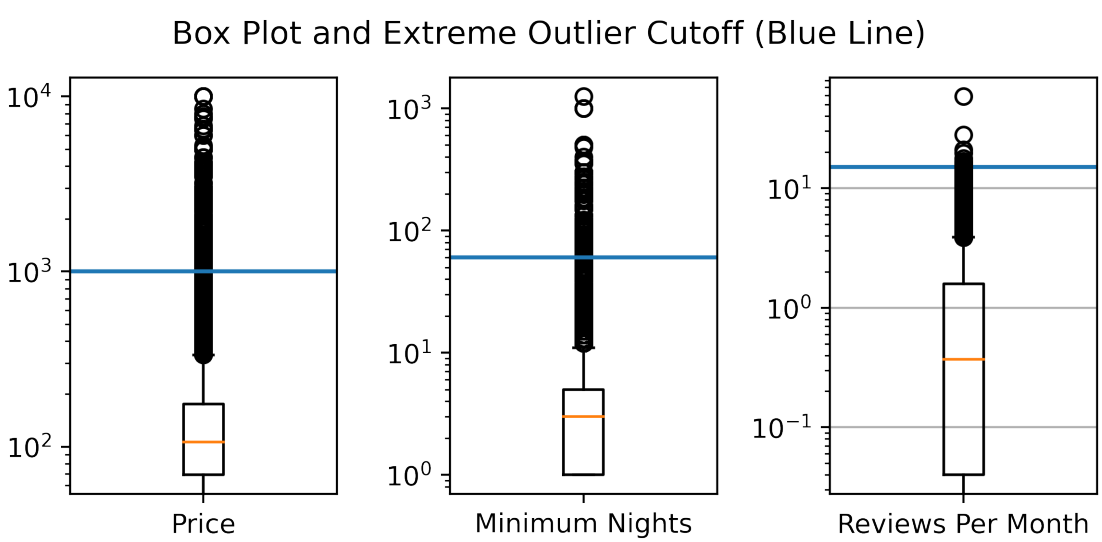


Figure 4: Outliers

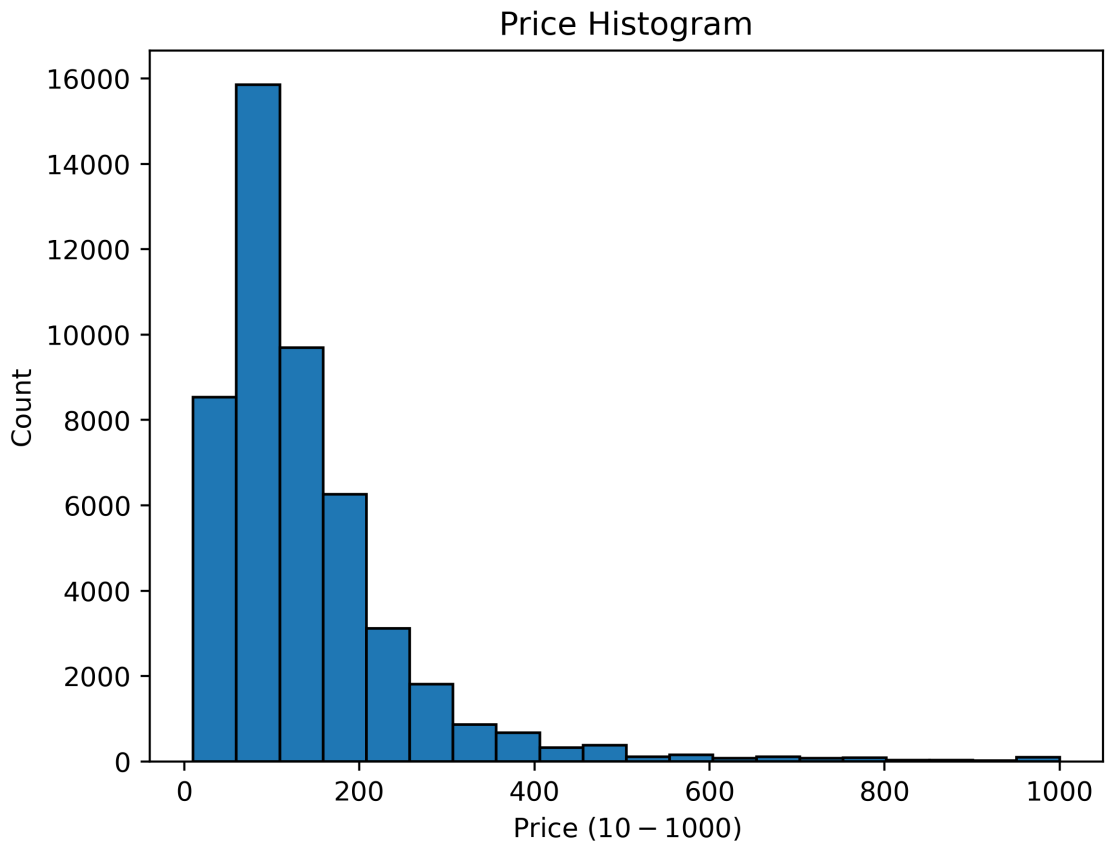


Figure 5: Histogram of Price

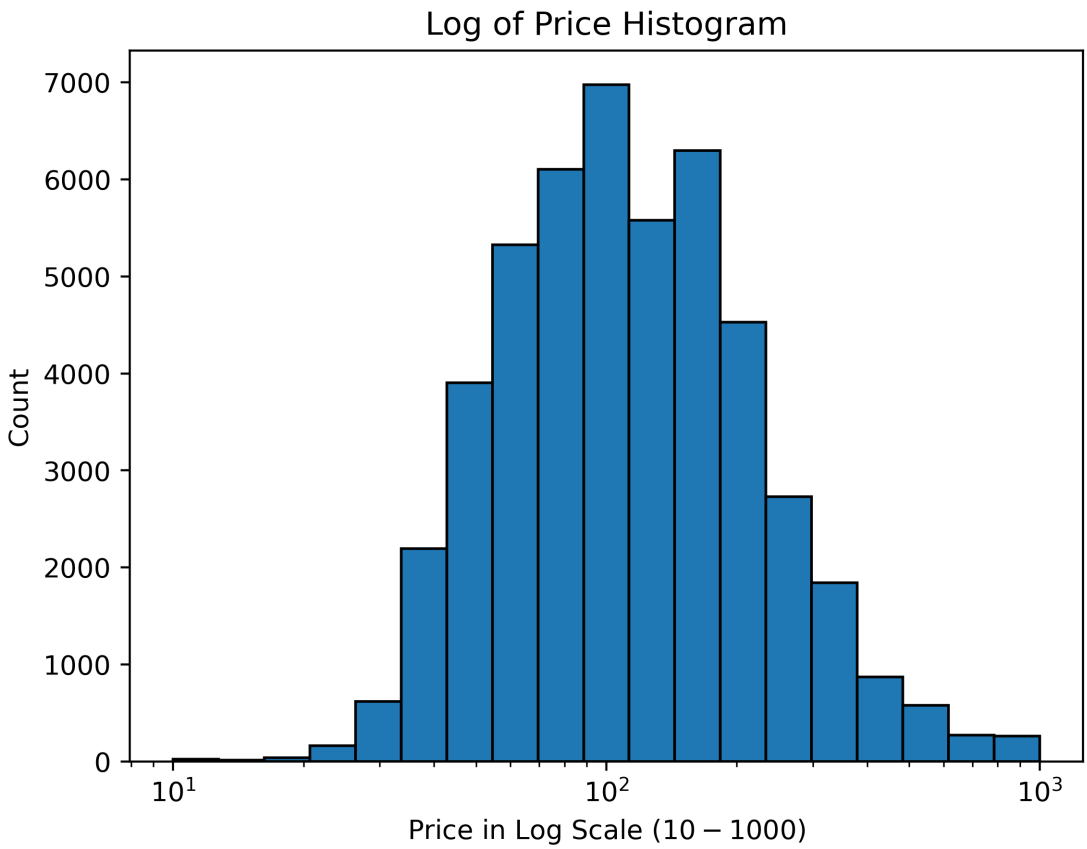


Figure 6: Histogram of Price with Logarithmic Scale

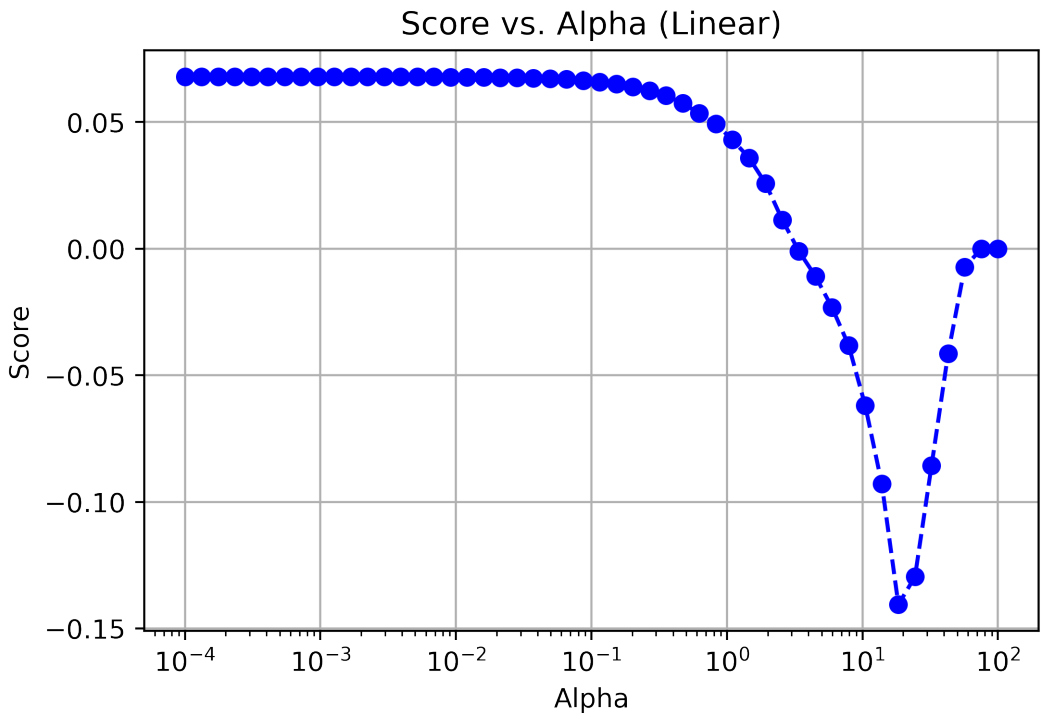


Figure 7: Linear (Non-Transformed Price) Score Alpha (for Lasso)

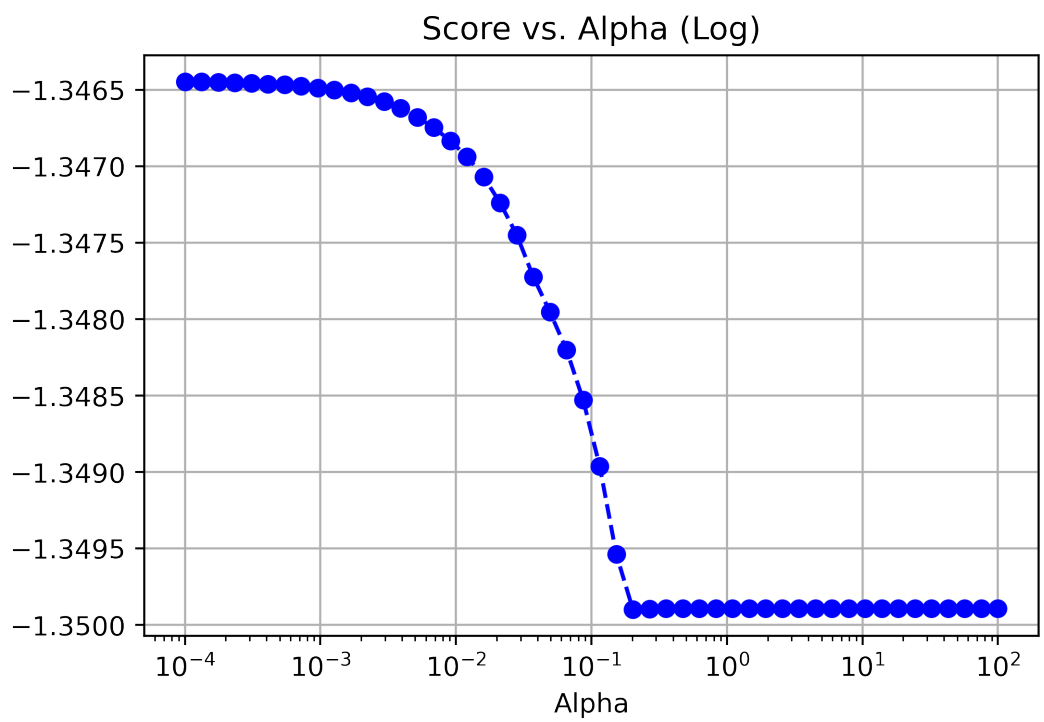


Figure 8: Log (Log Transformed Price) Score Alpha (for Lasso)

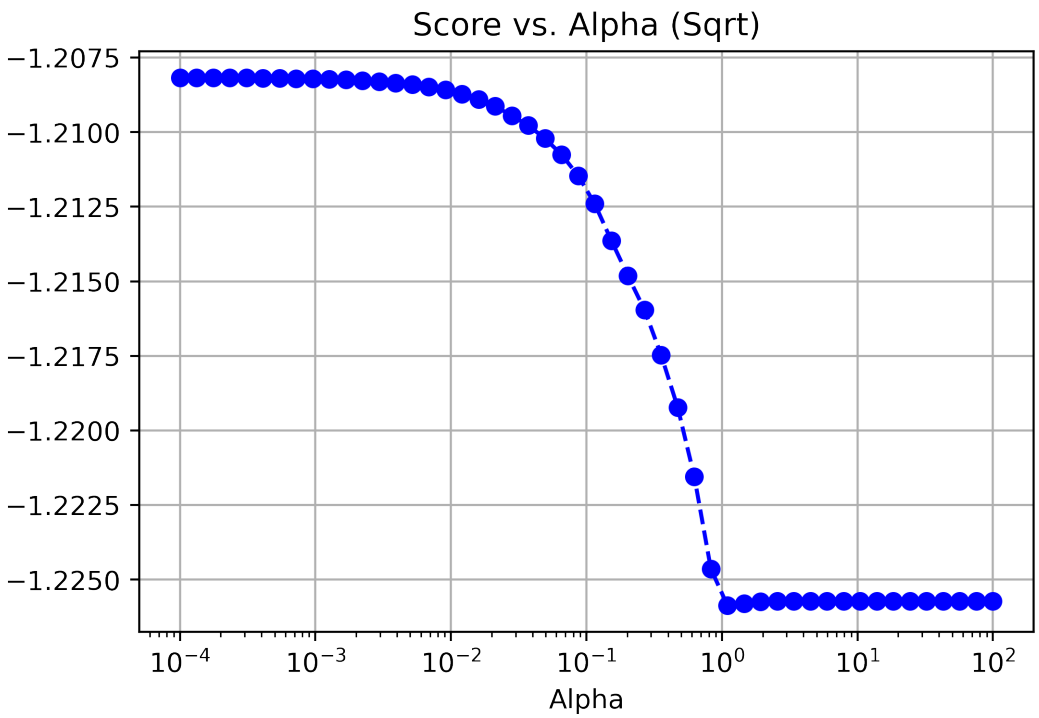


Figure 9: Sqrt (Sqrt Transformed Price) Score Alpha (for Lasso)

	MAPE	MAE	MAD_ratio	r2_score	MSLE	Median Absolute Error	MSE
Null	0.805	78.56	0.000	0.00%	0.496	61.89	13941
OLS (Linear)	0.645	72.18	0.662	6.85%	0.434	48.20	12986
OLS (Log)	0.451	61.61	0.571	11.82%	0.291	33.24	12294
OLS (Sqrt)	0.521	64.79	0.597	12.47%	0.324	38.84	12202
LASSO (Linear)	0.645	72.18	0.662	6.85%	0.434	48.20	12986
LASSO (Log)	0.451	61.60	0.572	11.85%	0.291	33.23	12289
LASSO (Sqrt)	0.521	64.79	0.597	12.48%	0.324	38.84	12201
kNN (Linear)	0.540	65.42	0.687	10.73%	0.325	37.28	12446
kNN (Log)	0.449	60.91	0.619	14.22%	0.293	31.95	11958
kNN (Sqrt)	0.489	62.59	0.648	13.50%	0.303	34.48	12059
SVM (Linear Kernel)	0.466	60.93	0.500	15.81%	0.301	35.94	11736
SVM (Linear/Gaussian)	0.449	61.10	0.514	14.21%	0.301	34.18	11959
SVM (Log/Gaussian)	0.427	61.11	0.613	12.15%	0.296	31.31	12247
SVM (Sqrt/Gaussian)	0.431	61.11	0.597	12.91%	0.297	31.92	12141
Random Forest (Linear)	0.613	72.48	0.776	-1.40%	0.391	42.47	14136
Random Forest (Log)	0.489	65.20	0.663	8.53%	0.339	35.83	12752
Random Forest (Sqrt)	0.545	68.47	0.717	4.71%	0.360	39.01	13284

Figure 10: Evaluation Metrics Table for All 17 Models

8.1 Median Absolute Error

Quantile	(9.999, 45.0]	(45.0, 58.0]	(58.0, 69.0]	(69.0, 80.0]	(80.0, 95.0]	(95.0, 105.0]	(105.0, 125.0]	(125.0, 150.0]	(150.0, 175.0]	(175.0, 205.0]	(205.0, 293.25]	(293.25, 1000.0]
Null	101.89	91.89	77.89	66.89	51.89	41.89	21.89	8.11	23.11	55.11	107.11	238.11
OLS(Linear)	49.21	42.02	39.83	38.54	45.13	45.79	43.90	37.85	33.79	39.52	71.20	213.70
OLS(Log)	31.95	22.84	15.65	16.26	24.29	29.15	31.71	27.55	31.23	50.03	88.85	225.37
OLS(Sqrt)	38.51	29.80	24.16	23.92	30.52	35.00	35.11	30.94	30.59	42.73	79.93	218.59
LASSO(Linear)	49.21	42.02	39.83	38.54	45.13	45.79	43.90	37.85	33.79	39.52	71.20	213.70
LASSO(Log)	31.98	22.94	15.76	16.22	24.22	29.16	31.68	27.59	31.12	50.02	88.95	225.42
LASSO(Sqrt)	38.51	29.81	24.16	23.92	30.54	34.99	35.11	30.95	30.59	42.73	79.89	218.63
kNN(Linear)	34.88	25.17	19.90	19.95	26.35	33.28	36.47	35.88	36.05	45.55	76.46	207.97
kNN(Log)	27.55	17.96	13.55	17.48	23.96	29.23	31.90	32.62	34.67	50.86	87.47	220.83
kNN(Sqrt)	31.44	21.64	16.10	18.04	24.78	30.82	33.52	33.51	34.11	47.08	80.57	214.06
SVM(Linear Kernel)	37.32	28.35	23.51	23.12	29.57	33.30	31.42	22.83	22.27	42.89	85.69	224.64
SVM(Linear/Gaussian)	33.52	21.83	16.63	17.03	23.93	27.59	30.72	31.10	32.98	51.18	90.80	232.38
SVM(Log/Gaussian)	25.24	16.78	12.48	16.04	23.31	28.68	33.67	34.11	35.46	51.24	88.13	229.16
SVM(Sqrt/Gaussian)	26.59	17.62	12.52	15.97	22.82	28.49	33.75	33.95	35.36	51.30	88.47	228.69
RandomForest(Linear)	39.33	29.70	22.72	23.20	27.49	35.40	40.02	44.08	43.39	55.21	84.10	210.14
RandomForest(Log)	30.81	21.89	15.83	18.42	23.79	31.42	35.90	37.41	39.54	57.32	89.40	225.04
RandomForest(Sqrt)	34.82	25.35	19.61	20.88	25.30	33.17	38.72	41.13	40.90	56.81	86.61	215.88

Figure 11: Median Absolute Error by 12-Quantile of Price for All 17 Models

Neighborhood	Bronx	Brooklyn	Manhattan	Queens	Staten Island
Null	76.89	66.89	57.11	72.89	72.39
OLS(Linear)	38.36	47.03	51.30	43.54	44.24
OLS(Log)	19.99	29.86	41.46	24.89	28.51
OLS(Sqrt)	25.38	36.49	44.79	32.10	31.65
LASSO(Linear)	38.36	47.03	51.30	43.54	44.26
LASSO(Log)	20.11	29.87	41.46	24.90	28.81
LASSO(Sqrt)	25.32	36.51	44.78	32.07	31.92
kNN(Linear)	27.00	33.74	45.62	28.39	38.99
kNN(Log)	20.61	27.96	41.64	22.98	29.44
kNN(Sqrt)	23.71	30.40	43.28	25.70	36.61
SVM(Linear Kernel)	27.25	33.16	41.88	31.54	33.63
SVM(Linear/Gaussian)	23.81	30.09	43.08	27.10	33.43
SVM(Log/Gaussian)	17.13	27.14	42.29	21.34	23.23
SVM(Sqrt/Gaussian)	16.95	27.73	42.45	21.25	24.78
RandomForest(Linear)	42.33	36.51	49.91	39.56	47.40
RandomForest(Log)	32.06	30.05	45.36	29.00	34.70
RandomForest(Sqrt)	37.47	33.09	47.54	34.46	43.76

Figure 12: Median Absolute Error by Neighborhood for All 17 Models

RoomType	Entire home/apt				Private room				Shared room			
Quartile	(9.999, 69.0]	(69.0, 105.0]	(105.0, 175.0]	(175.0, 1000.0]	(9.999, 69.0]	(69.0, 105.0]	(105.0, 175.0]	(175.0, 1000.0]	(9.999, 69.0]	(69.0, 105.0]	(105.0, 175.0]	(175.0, 1000.0]
Null	81.89	47.89	16.89	108.11	91.89	56.89	16.89	88.11	103.89	61.89	21.89	183.11
OLS(Linear)	103.54	75.25	38.03	77.66	40.65	31.76	41.69	133.39	40.07	33.67	63.16	264.36
OLS(Log)	78.17	51.68	25.47	88.71	23.08	14.82	46.51	139.96	17.04	23.84	71.06	280.65
OLS(Sqrt)	89.27	62.84	29.84	81.42	29.47	19.76	41.93	138.84	26.37	21.03	67.56	278.38
LASSO(Linear)	103.54	75.25	38.03	77.66	40.65	31.76	41.69	133.39	40.07	33.66	63.16	264.36
LASSO(Log)	78.14	51.79	25.45	88.77	23.14	14.78	46.52	139.80	17.22	23.66	70.80	280.47
LASSO(Sqrt)	89.26	62.84	29.85	81.44	29.46	19.76	41.92	138.84	26.40	20.98	67.50	278.31
kNN(Linear)	87.80	65.38	34.49	80.12	25.78	16.28	40.45	136.43	36.50	20.25	49.97	259.81
kNN(Log)	75.21	51.87	28.94	88.67	19.00	15.84	44.77	141.68	18.16	20.98	65.93	269.06
kNN(Sqrt)	81.35	58.02	31.07	83.06	22.44	15.73	42.27	138.62	27.36	20.09	52.93	266.12
SVM(Linear Kernel)	84.74	56.69	21.97	86.35	28.05	18.25	41.72	140.15	26.49	18.98	69.11	278.34
SVM(Linear/Gaussian)	79.16	49.71	27.88	93.89	22.71	16.00	44.33	138.89	36.27	21.00	47.95	262.18
SVM(Log/Gaussian)	71.68	50.99	30.31	93.21	17.48	15.50	46.47	142.05	10.73	22.76	67.78	275.55
SVM(Sqrt/Gaussian)	74.96	51.47	30.56	93.78	18.25	15.29	45.94	141.79	12.43	22.46	68.67	276.29
RandomForest(Linear)	100.85	66.19	41.96	89.89	28.90	19.21	43.01	132.49	36.60	32.33	53.53	209.96
RandomForest(Log)	87.18	53.79	35.35	94.12	21.86	16.95	46.40	141.01	19.63	24.61	45.50	233.16
RandomForest(Sqrt)	96.19	60.02	38.73	91.76	25.18	17.85	44.12	138.24	23.74	27.08	36.48	227.93

Figure 13: Median Absolute Error by Price Quartile and Room Type for All 17 Models

RoomType	Entire home/apt	Private room	Shared room
Null	47.11	72.89	96.89
OLS(Linear)	59.16	38.80	41.43
OLS(Log)	51.58	22.72	21.64
OLS(Sqrt)	54.32	27.74	27.71
LASSO(Linear)	59.16	38.80	41.43
LASSO(Log)	51.55	22.72	21.72
LASSO(Sqrt)	54.32	27.75	27.76
kNN(Linear)	56.65	24.53	34.92
kNN(Log)	51.75	20.43	21.96
kNN(Sqrt)	53.80	22.21	29.58
SVM(Linear Kernel)	49.36	26.43	28.08
SVM(Linear/Gaussian)	51.89	22.84	34.00
SVM(Log/Gaussian)	53.38	19.62	15.35
SVM(Sqrt/Gaussian)	53.55	19.89	17.81
RandomForest(Linear)	64.67	27.54	37.40
RandomForest(Log)	58.67	22.39	23.30
RandomForest(Sqrt)	61.48	24.83	28.10

Figure 14: Median Absolute Error by Room Type for All 17 Models

8.2 Feature Importance

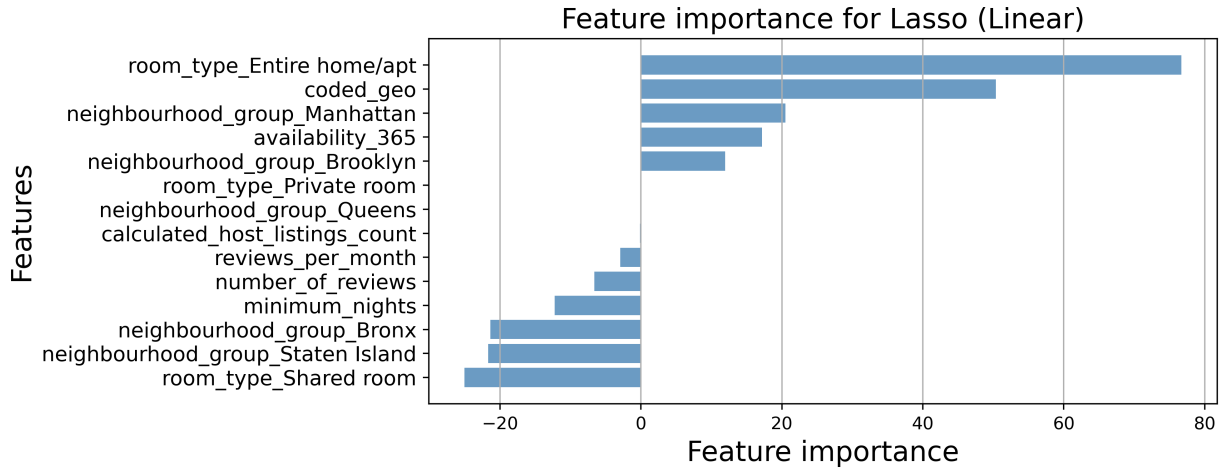


Figure 15: Feature Importance for Lasso (Linear Price)

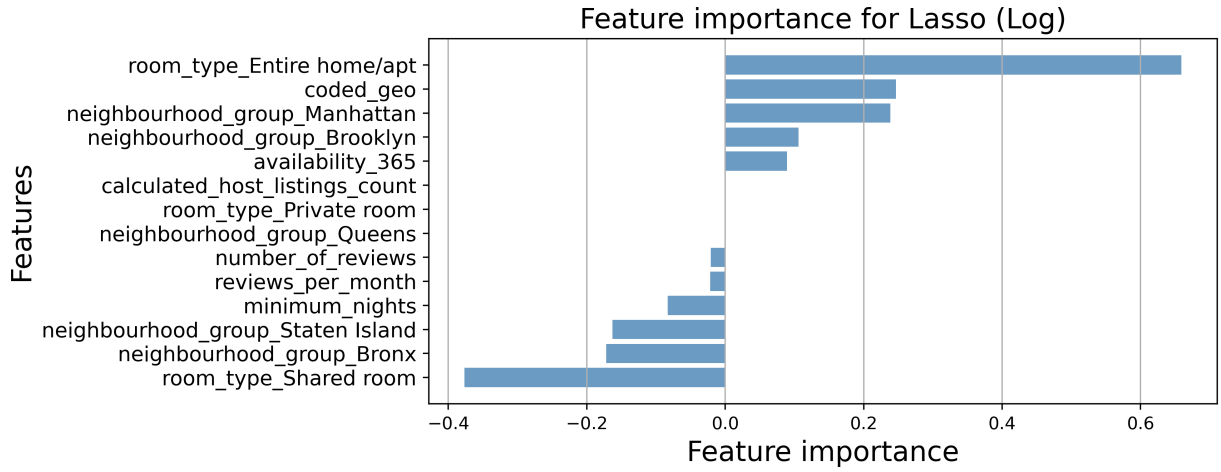


Figure 16: Feature Importance for Lasso (Log Price)

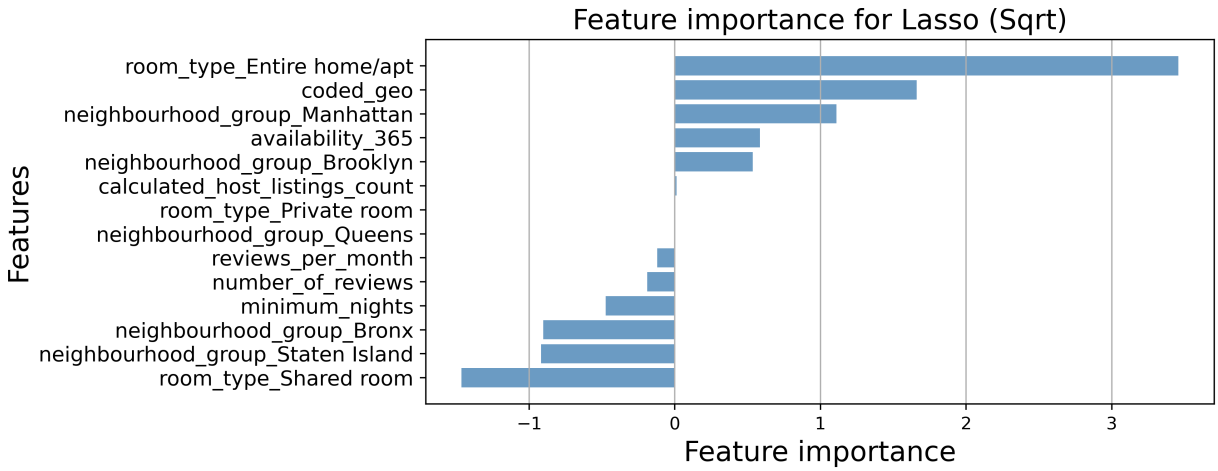


Figure 17: Feature Importance for Lasso (Sqrt Price)

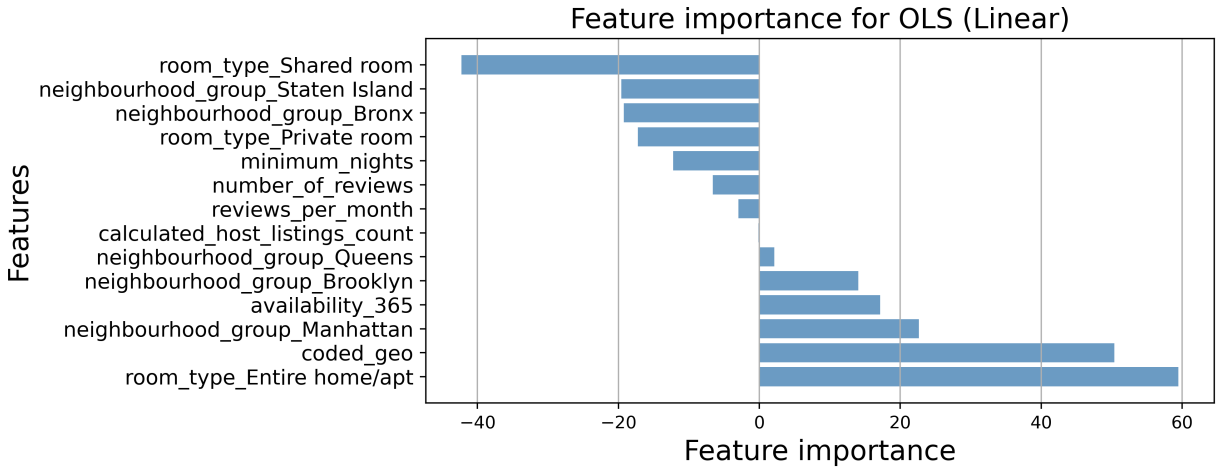


Figure 18: Feature Importance for OLS (Linear Price)

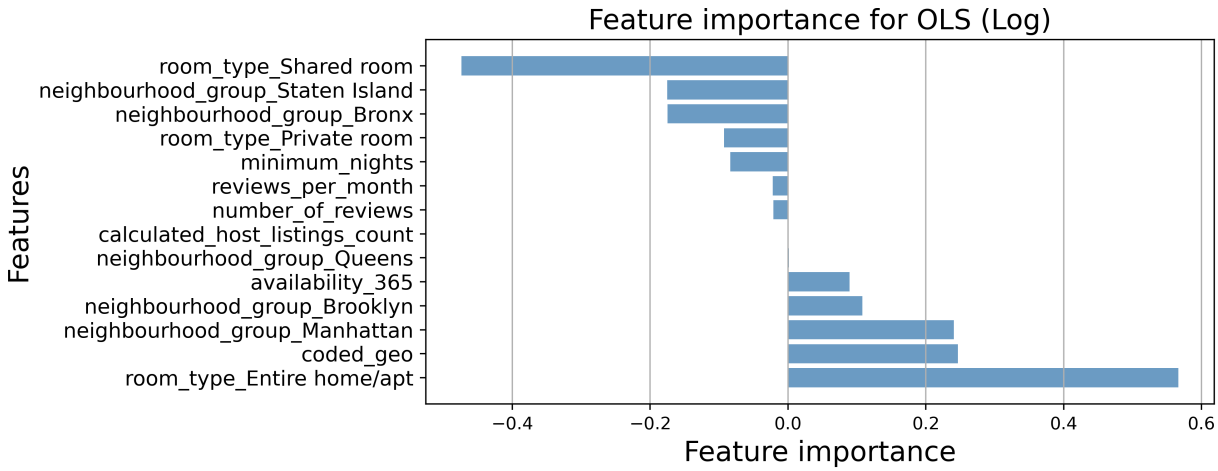


Figure 19: Feature Importance for OLS (Log Price)

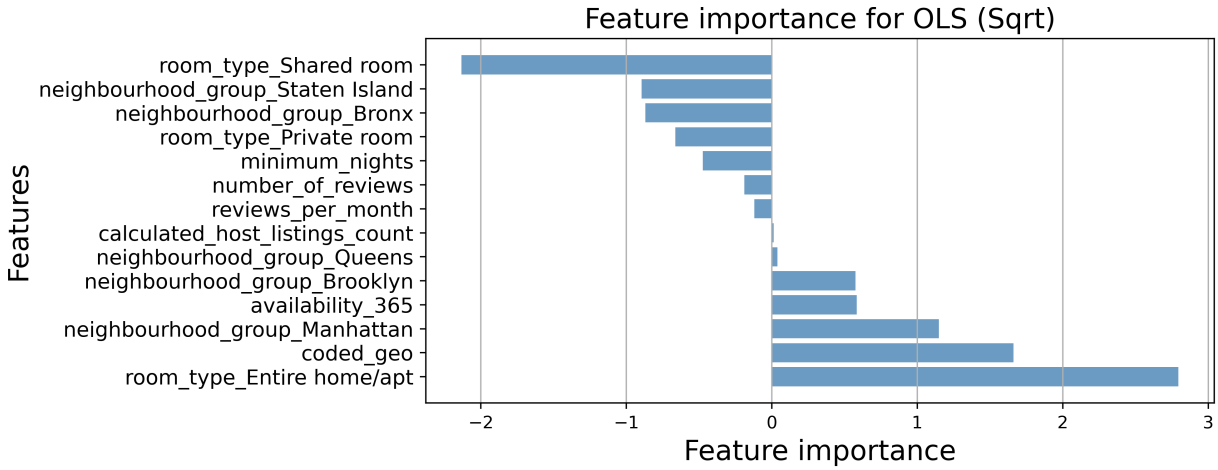


Figure 20: Feature Importance for OLS (Sqrt Price)

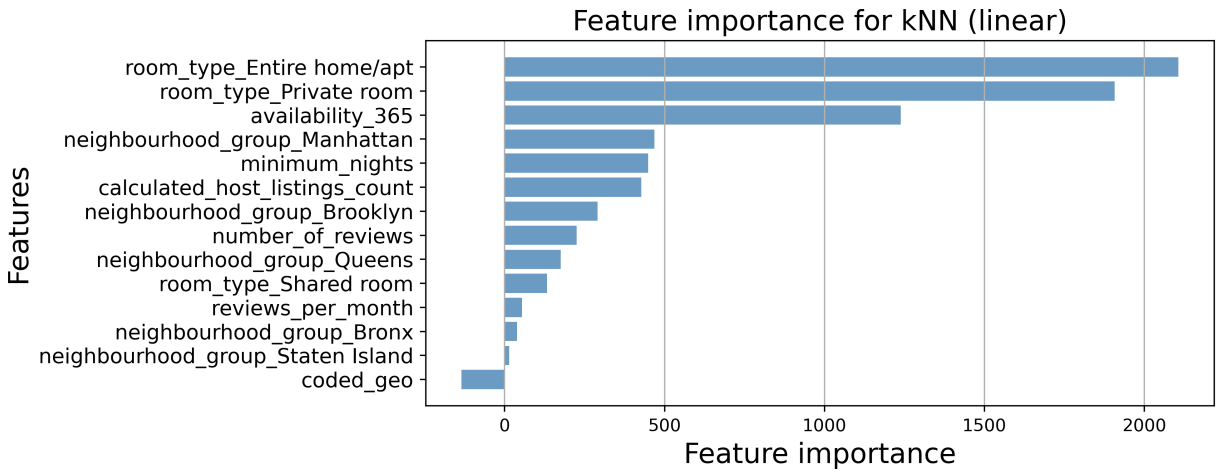


Figure 21: Feature Importance for kNN (Linear Price)

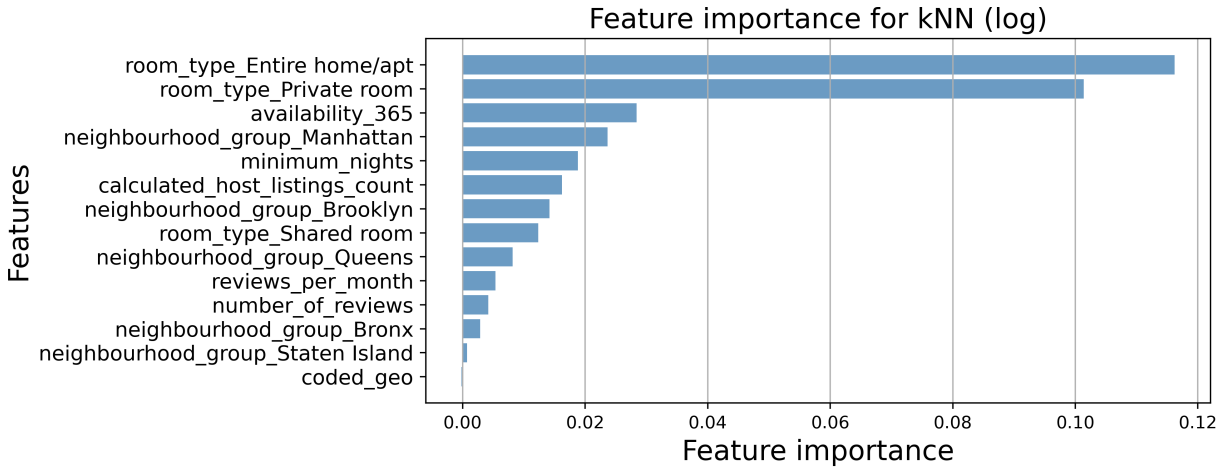


Figure 22: Feature Importance for kNN (Log Price)

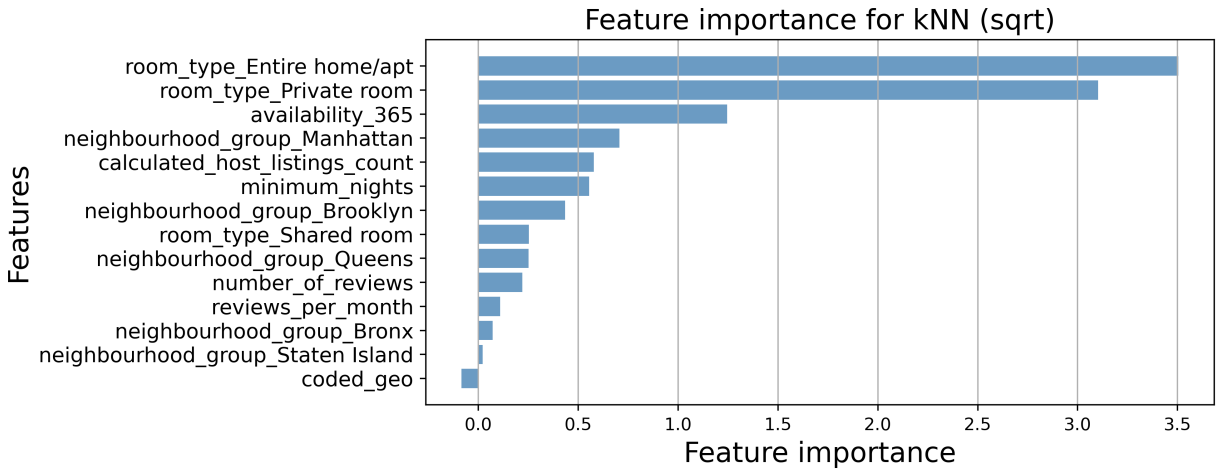


Figure 23: Feature Importance for kNN (Sqrt Price)

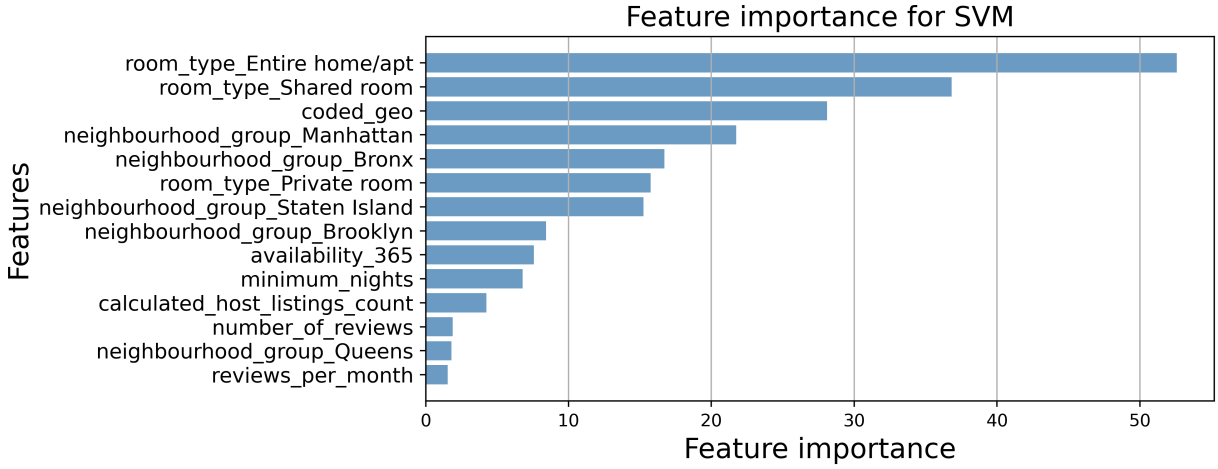


Figure 24: Feature Importance for SVM

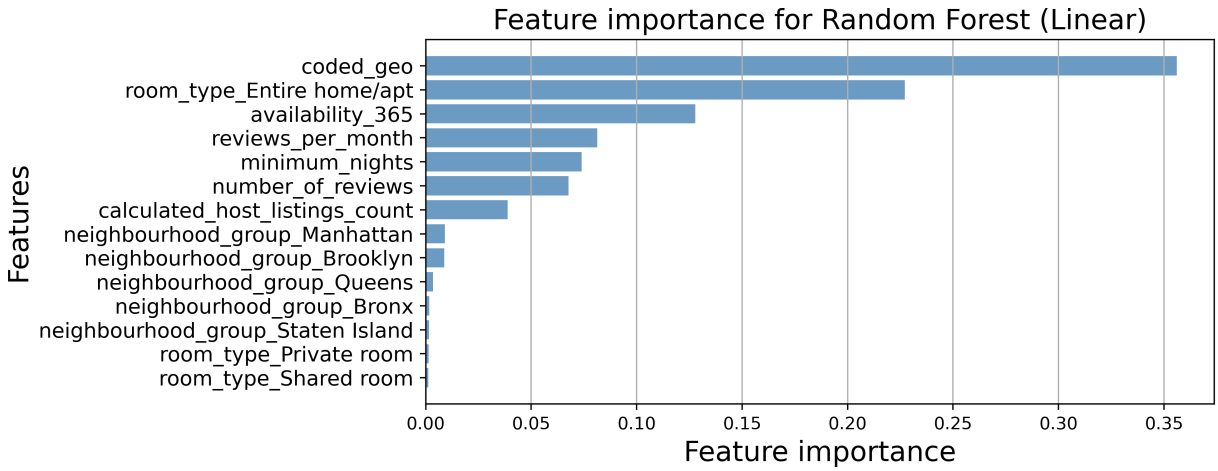


Figure 25: Feature Importance for Random Forest (Linear Price)

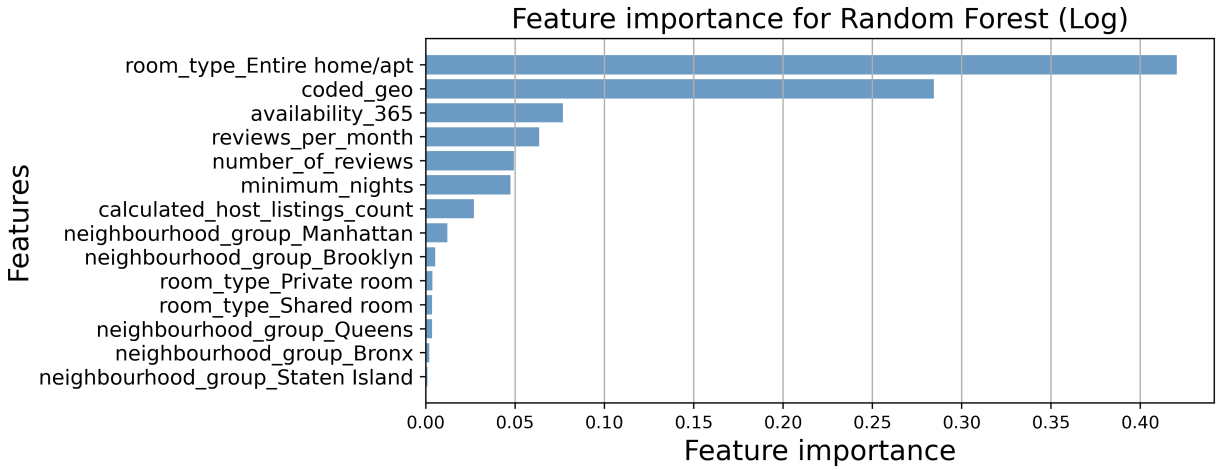


Figure 26: Feature Importance for Random Forest (Log Price)

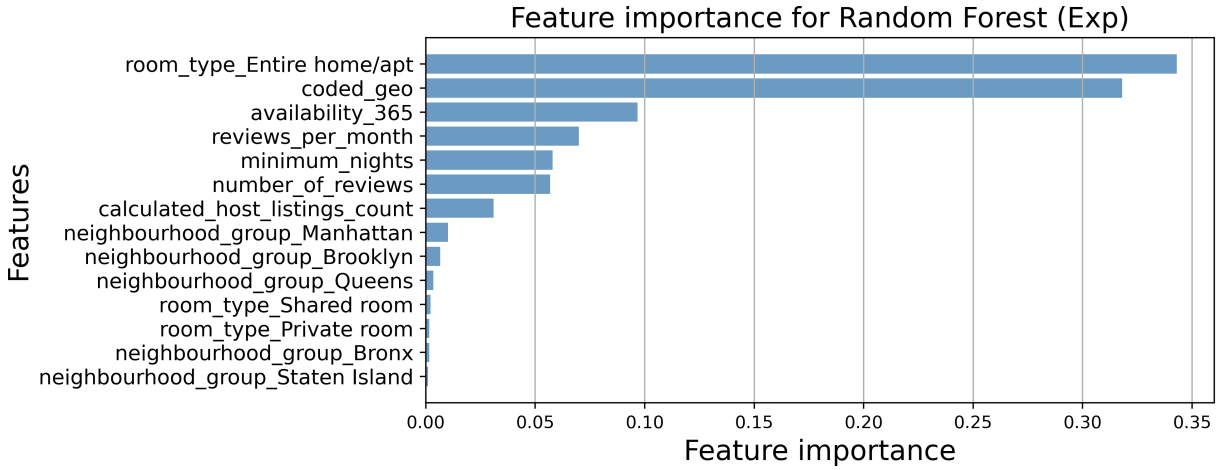
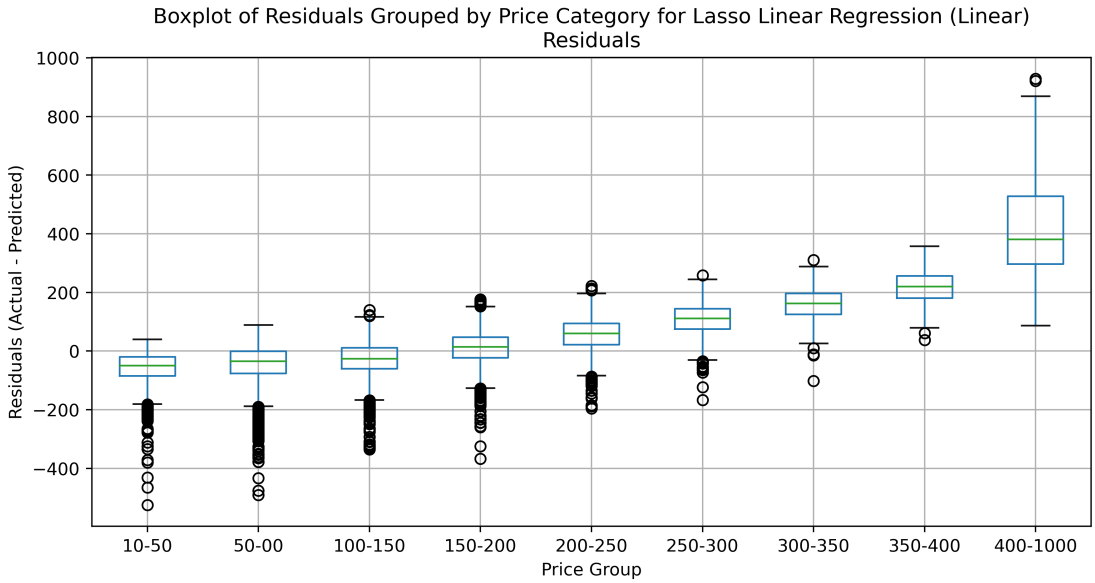
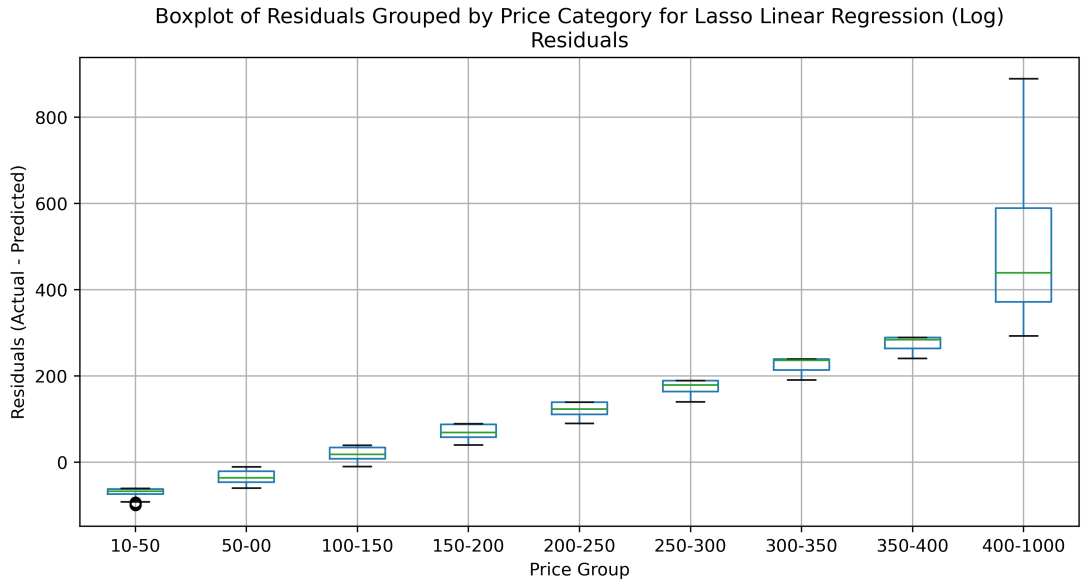


Figure 27: Feature Importance for Random Forest (Sqrt Price)

8.3 Price Category Boxplots





u

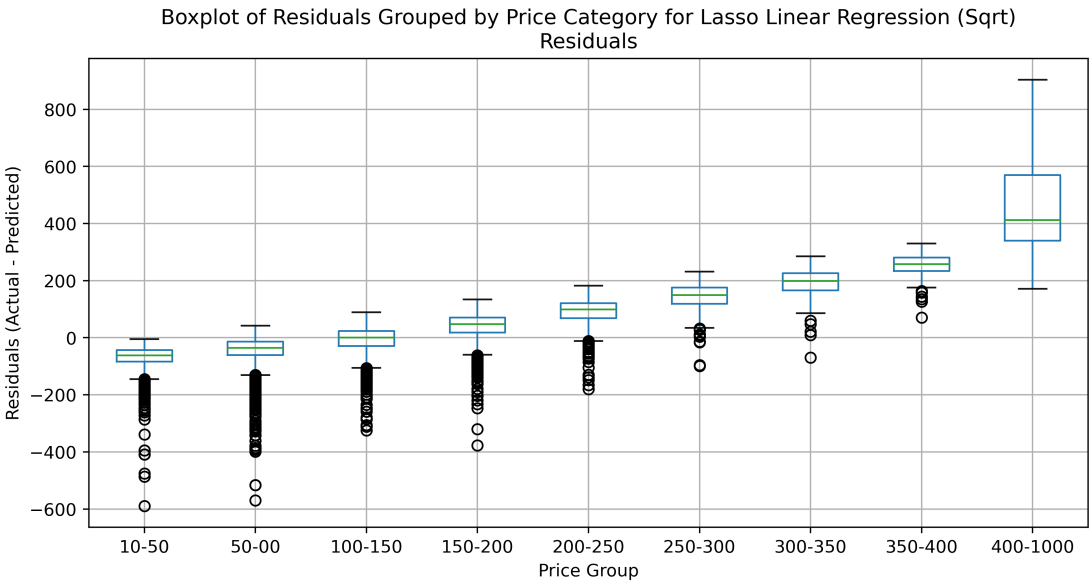


Figure 28: Price Category for Lasso Linear Regression (Sqrt)

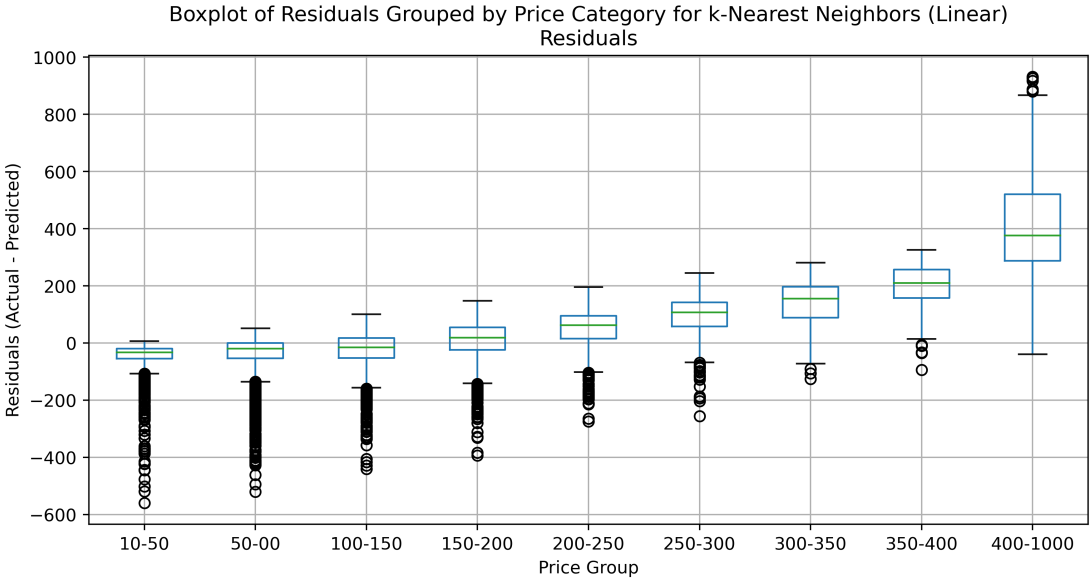


Figure 29: Price Category for k-Nearest Neighbors (Linear)

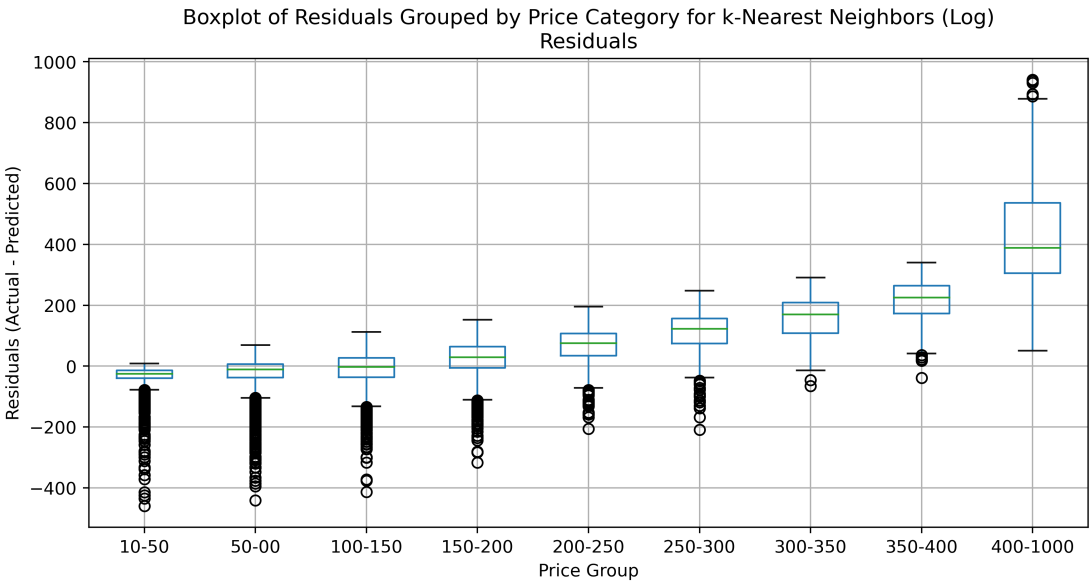


Figure 30: Price Category for k-Nearest Neighbors (Log)

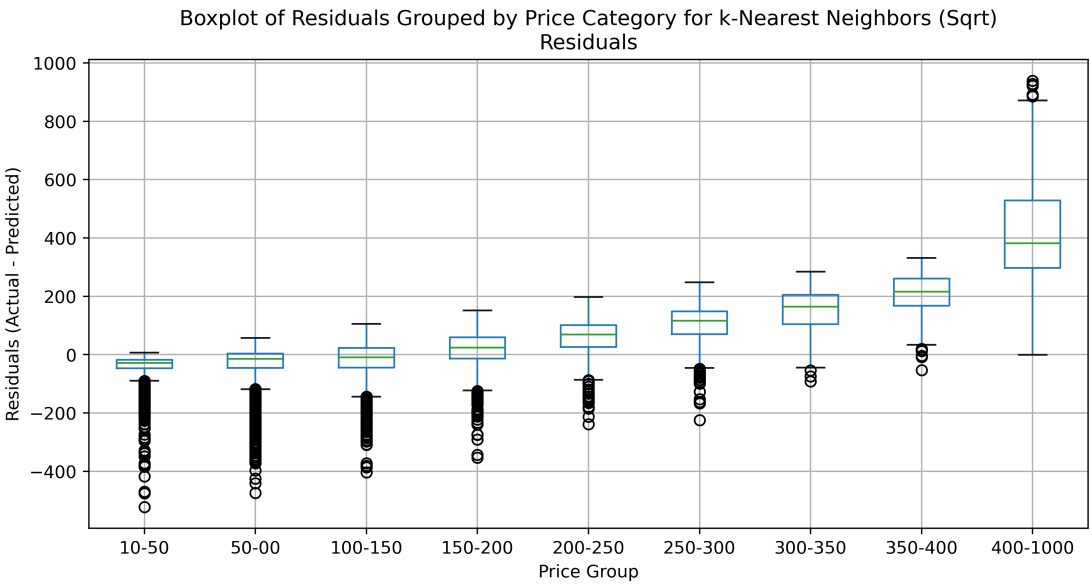


Figure 31: Price Category for k-Nearest Neighbors (Sqrt)

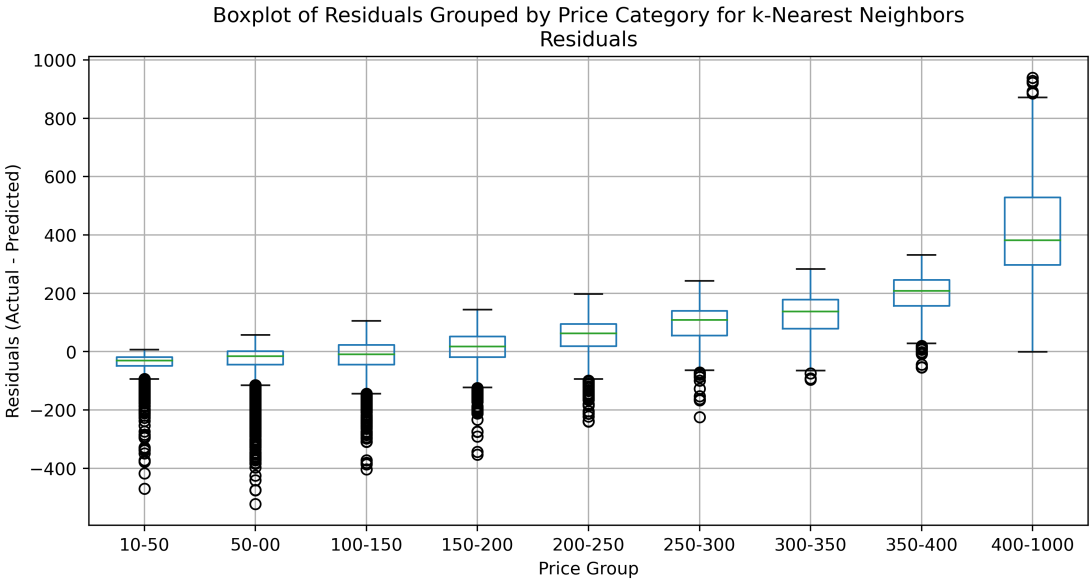


Figure 32: Price Category for k-Nearest Neighbors

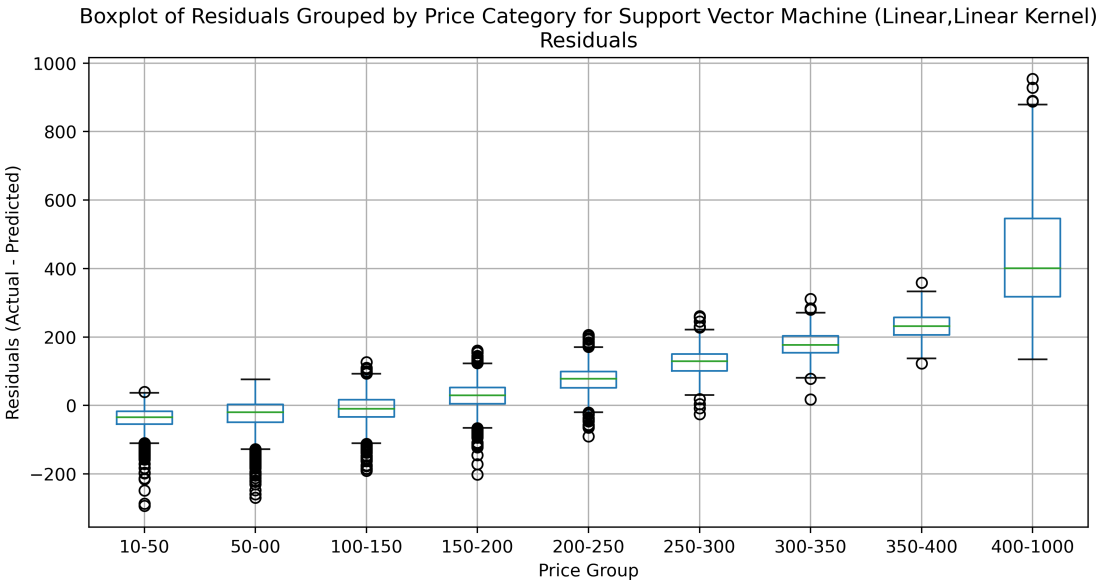


Figure 33: Price Category for Support Vector Machine (Linear,Linear Kernel)

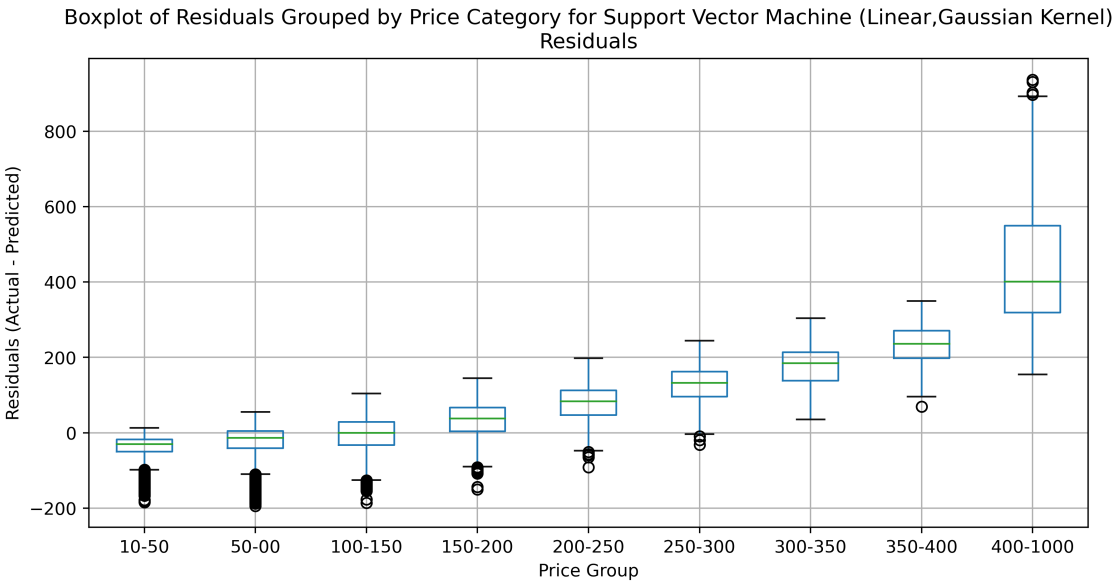


Figure 34: Price Category for Support Vector Machine (Linear,Gaussian Kernel)

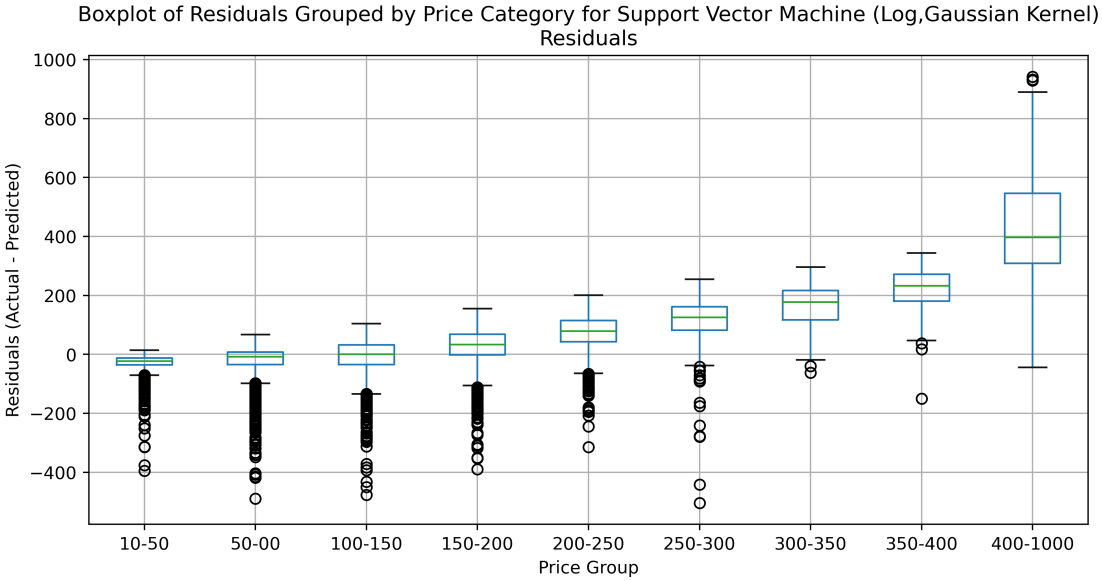


Figure 35: Price Category for Support Vector Machine (Log,Gaussian Kernel)

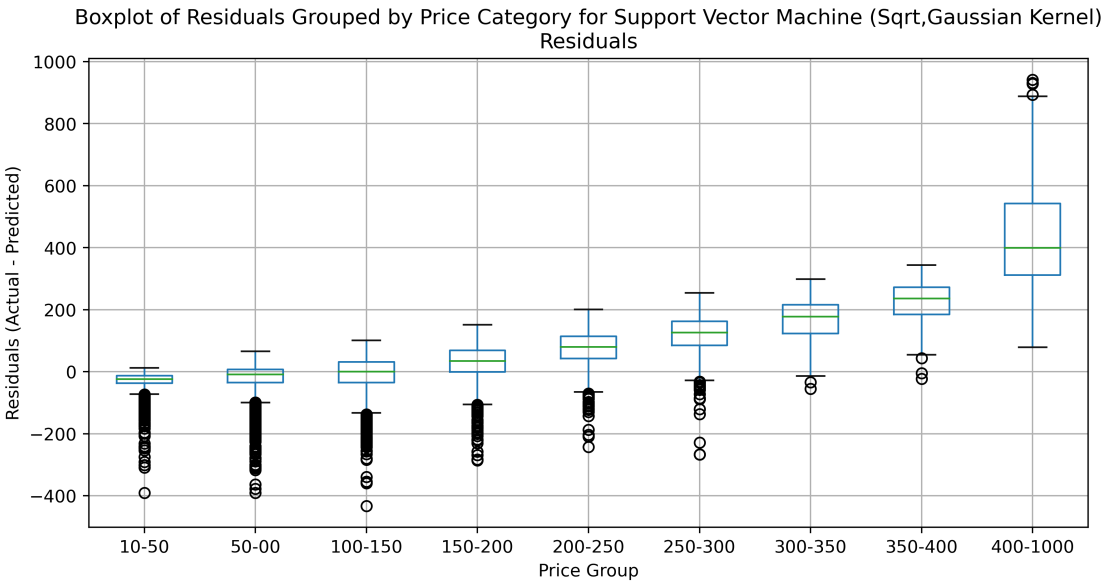


Figure 36: Price Category for Support Vector Machine (Sqrt,Gaussian Kernel)

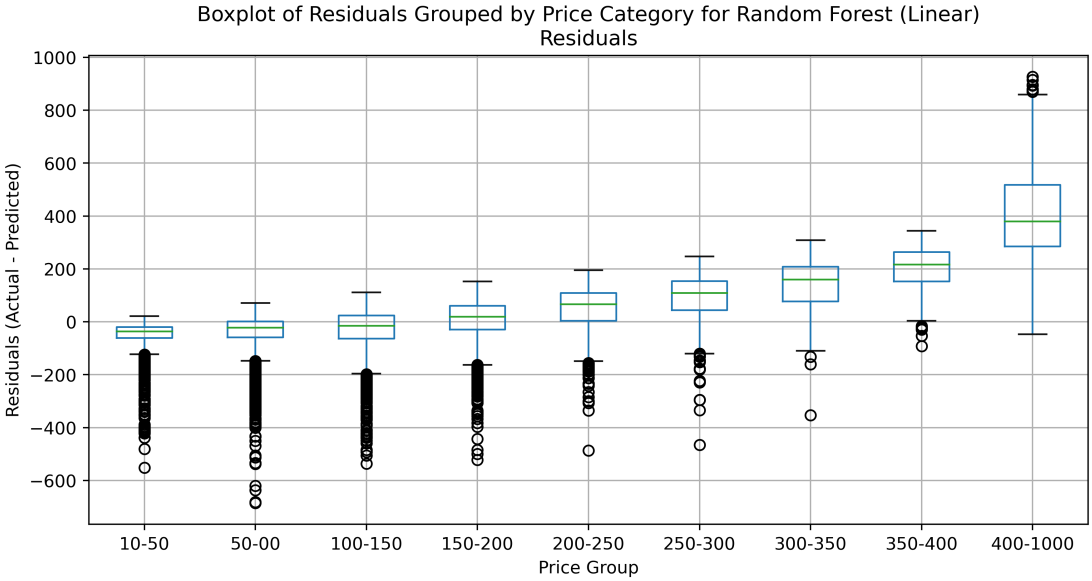


Figure 37: Price Category for Random Forest (Linear)

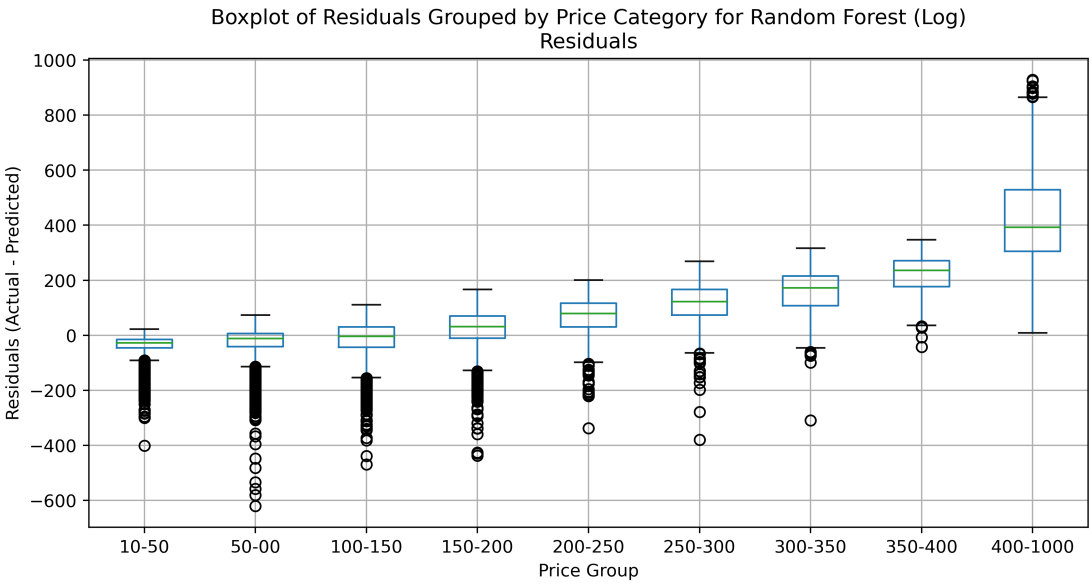


Figure 38: Price Category for Random Forest (Log)

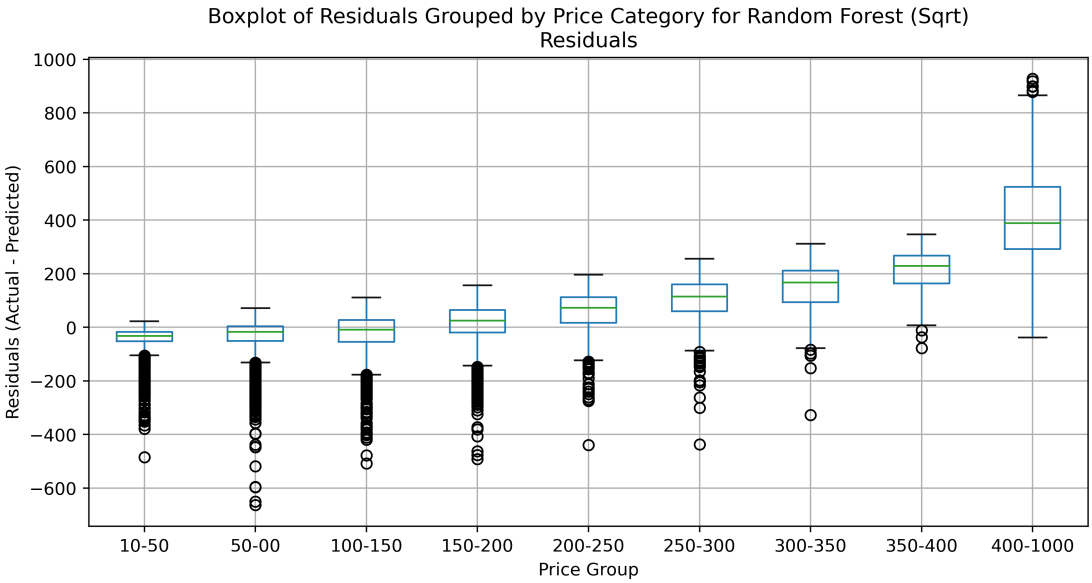


Figure 39: Price Category for Random Forest (Sqrt)

8.4 Actual Price vs Residual Price

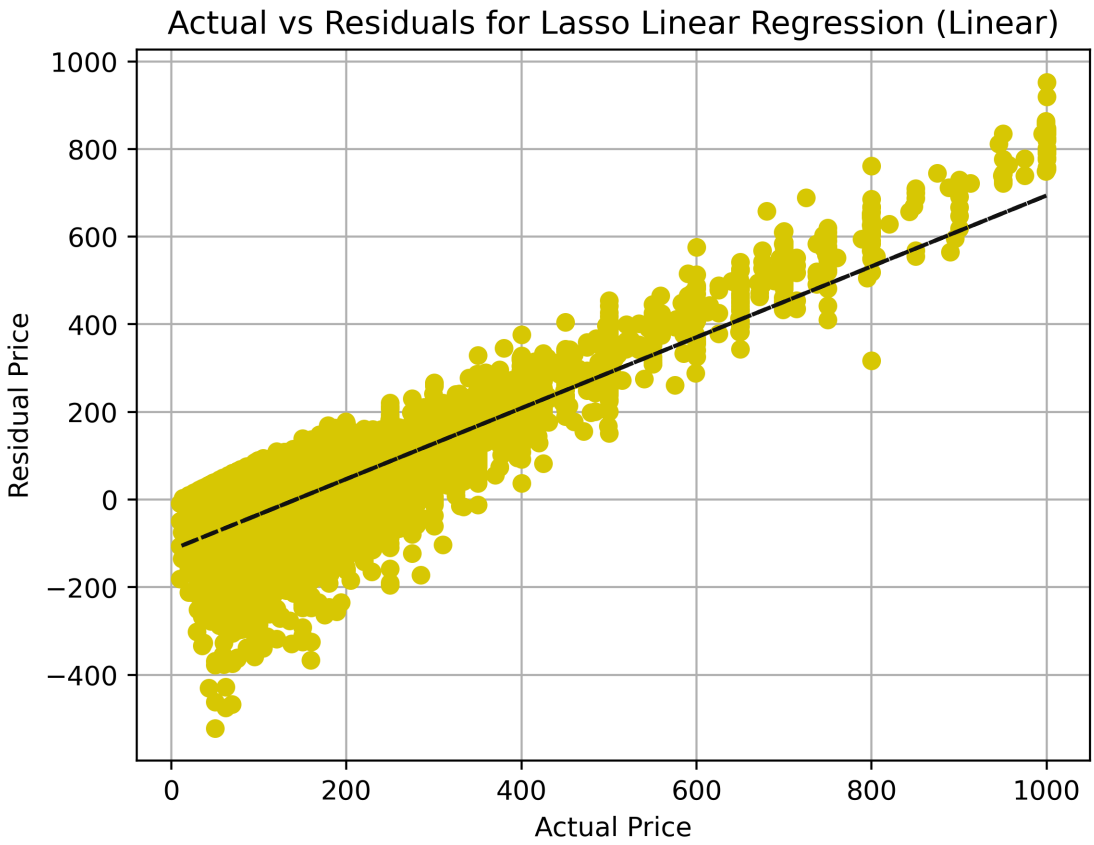


Figure 40: Actual vs Residuals for Lasso Linear Regression (Linear)

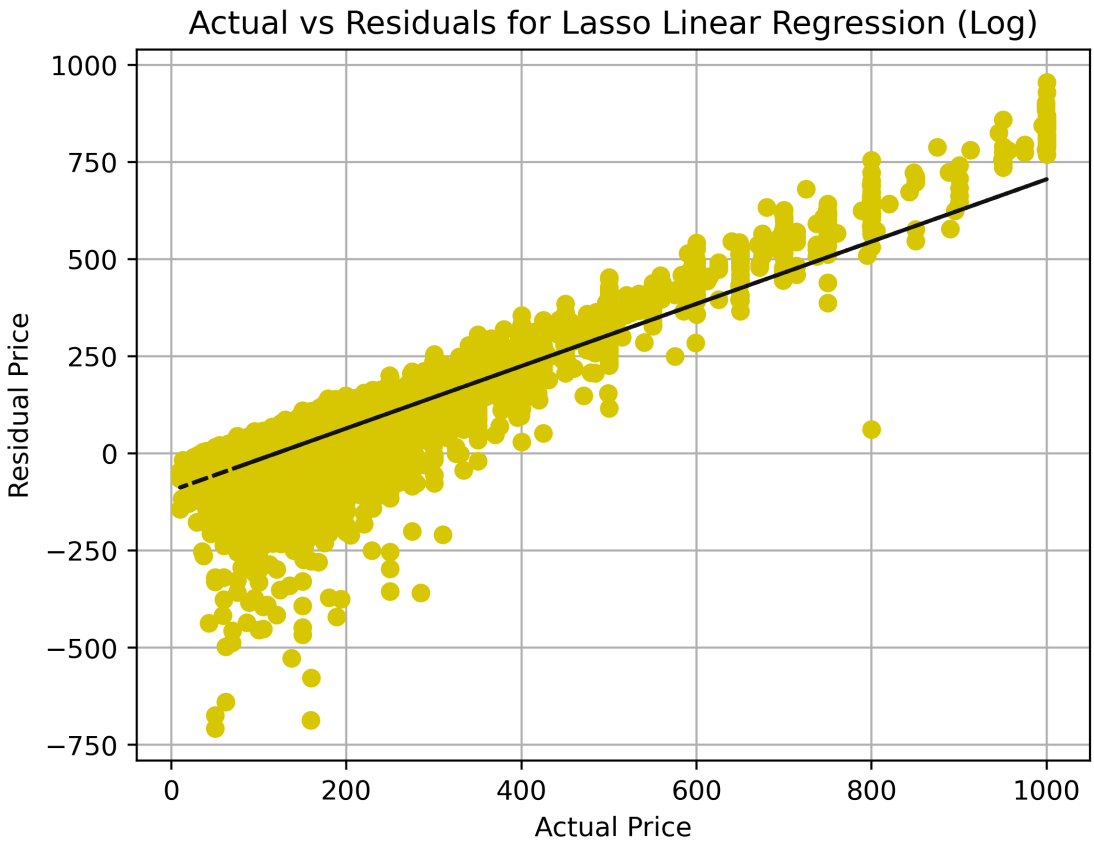


Figure 41: Actual vs Residuals for Lasso Linear Regression (Log)

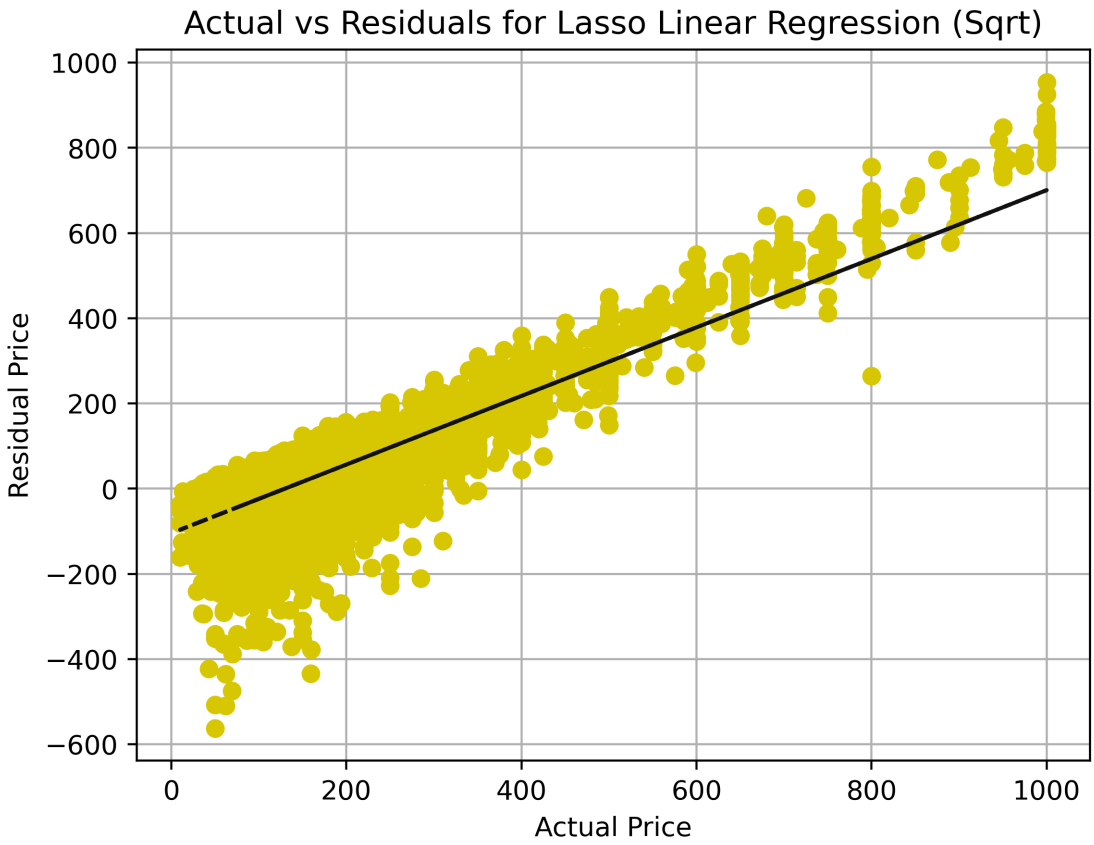


Figure 42: Actual vs Residuals for Lasso Linear Regression (Sqrt)

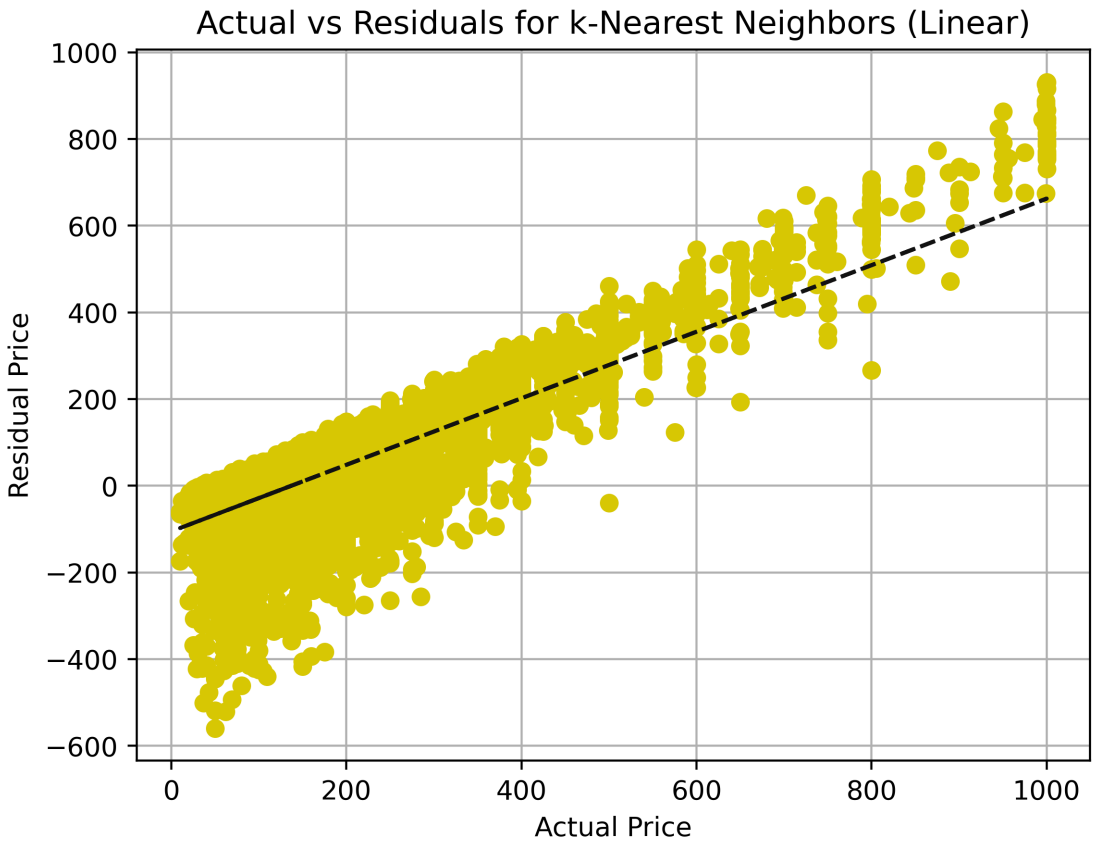


Figure 43: Actual vs Residuals for k-Nearest Neighbors (Linear)

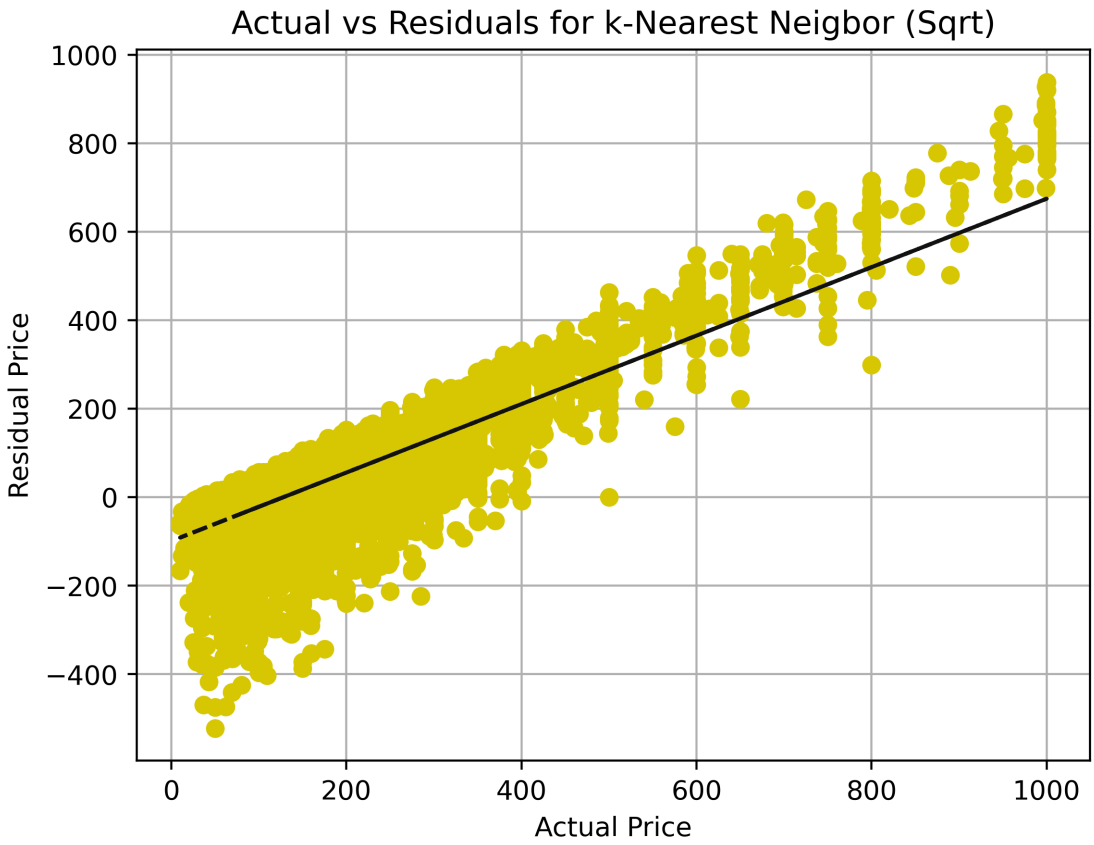


Figure 44: Actual vs Residuals for k-Nearest Neigbor (Sqrt)

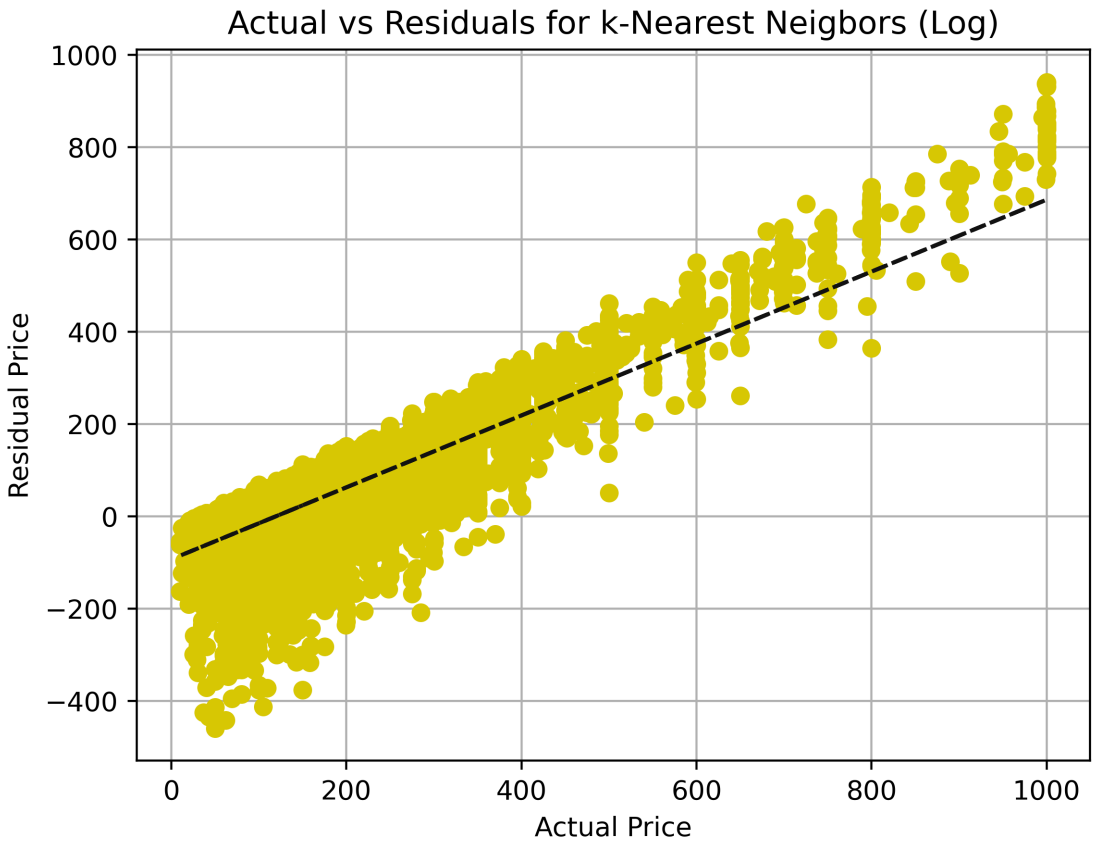


Figure 45: Actual vs Residuals for k-Nearest Neigbors (Log)

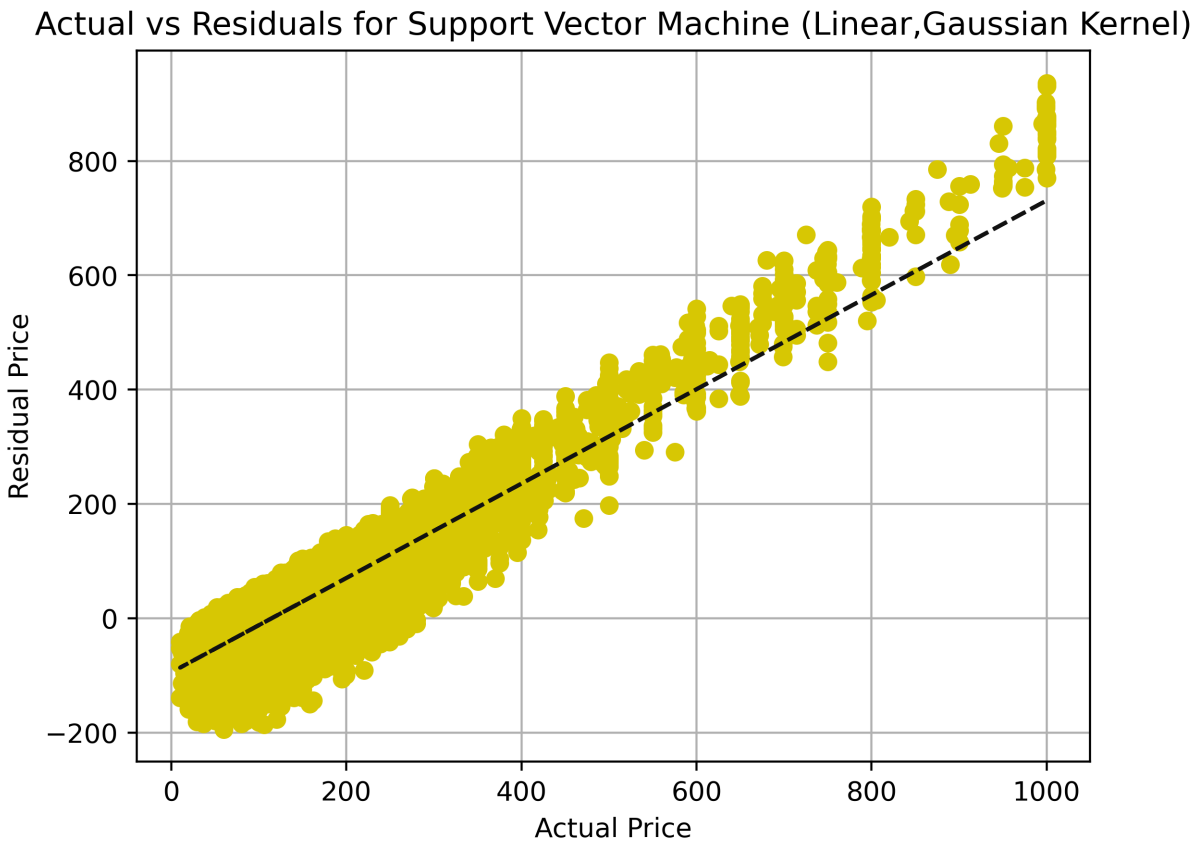


Figure 46: Actual vs Residuals for Support Vector Machine (Linear,Gaussian Kernel)

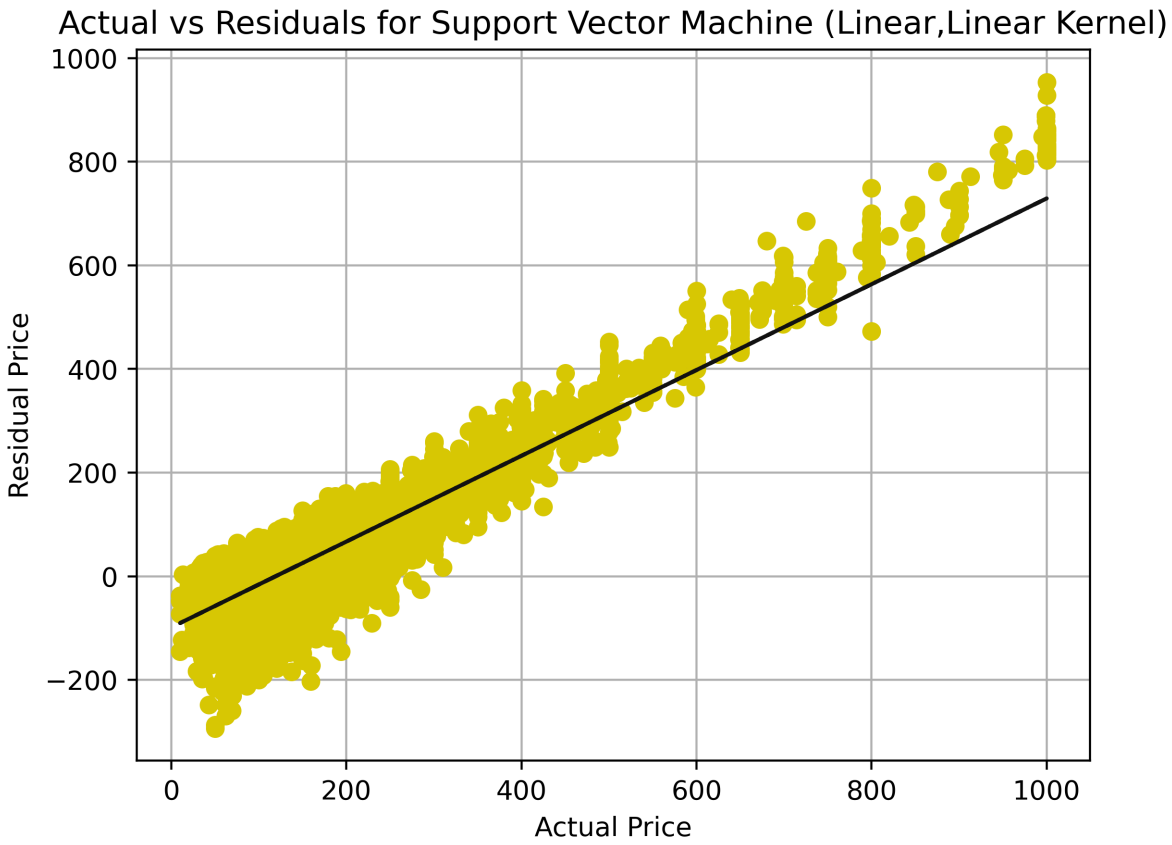


Figure 47: Actual vs Residuals for Support Vector Machine (Linear,Linear Kernel)

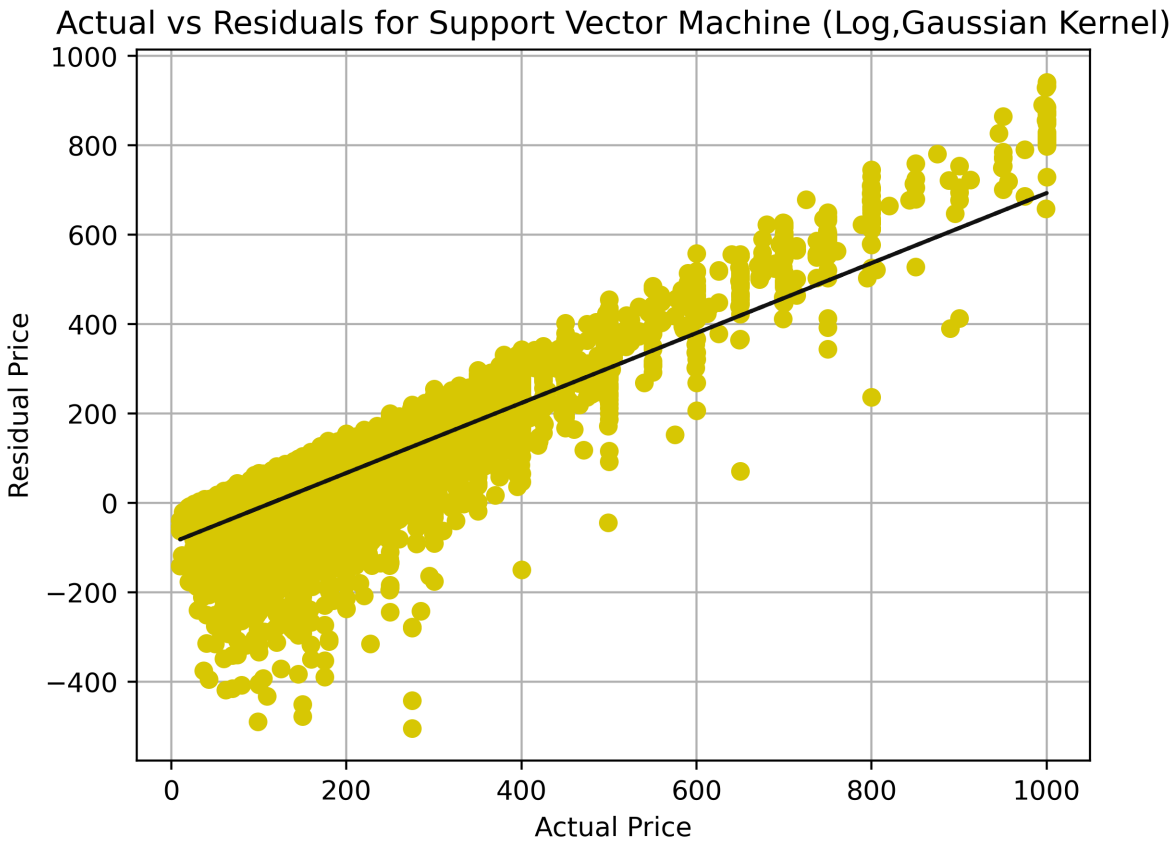


Figure 48: Actual vs Residuals for Support Vector Machine (Log,Gaussian Kernel)

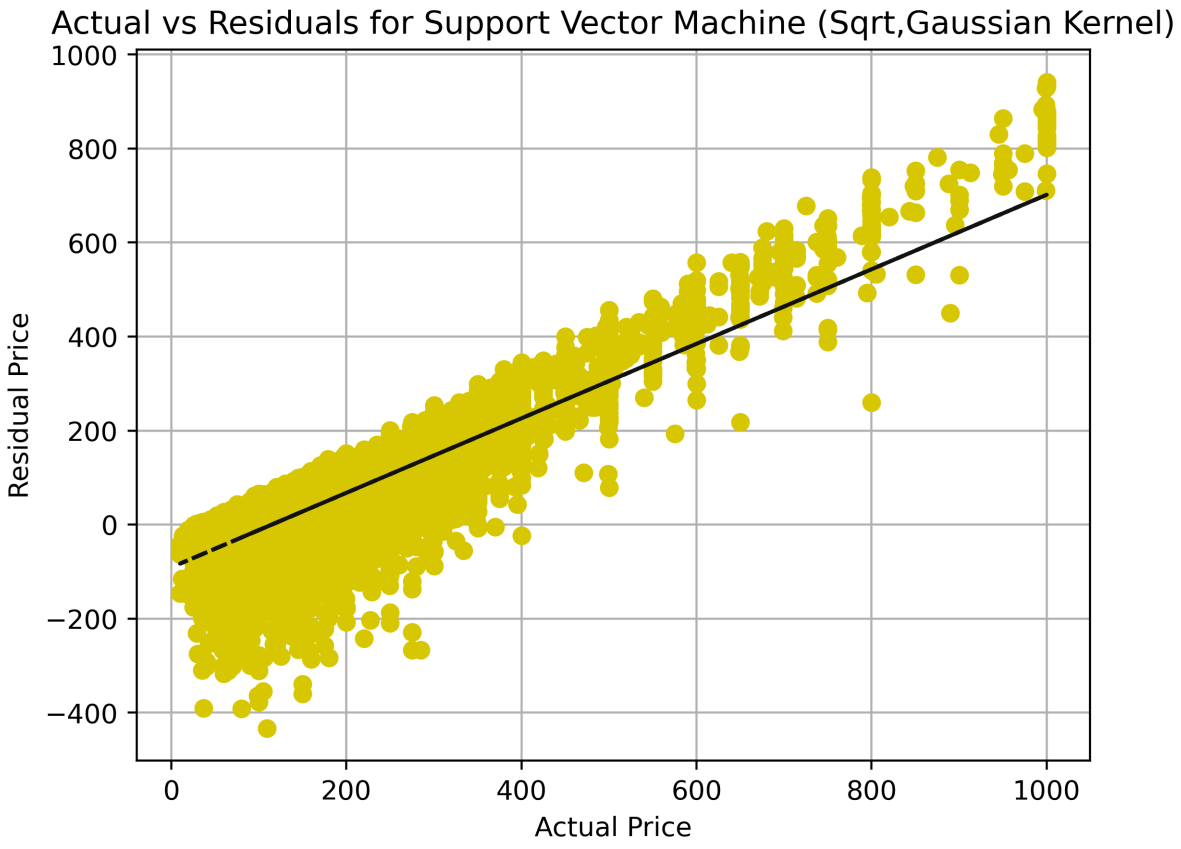


Figure 49: Actual vs Residuals for Support Vector Machine (Sqrt,Gaussian Kernel)

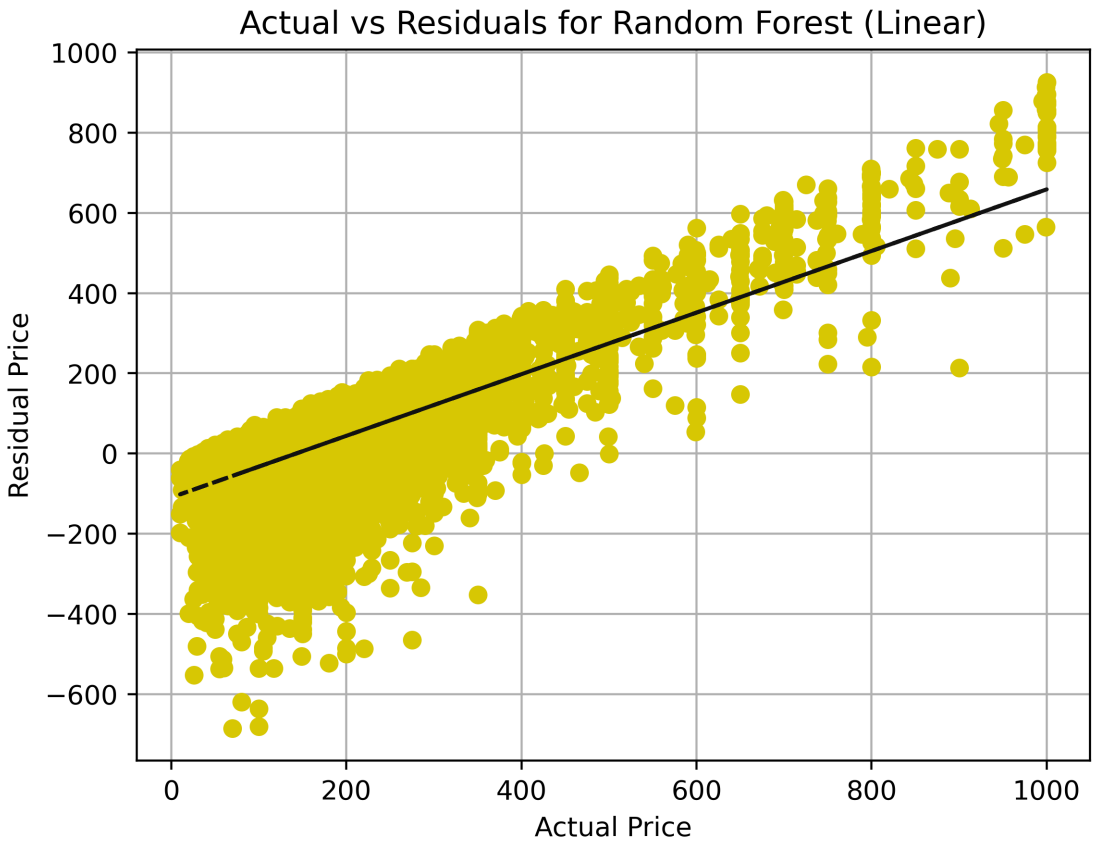


Figure 50: Actual vs Residuals for Random Forest (Linear)

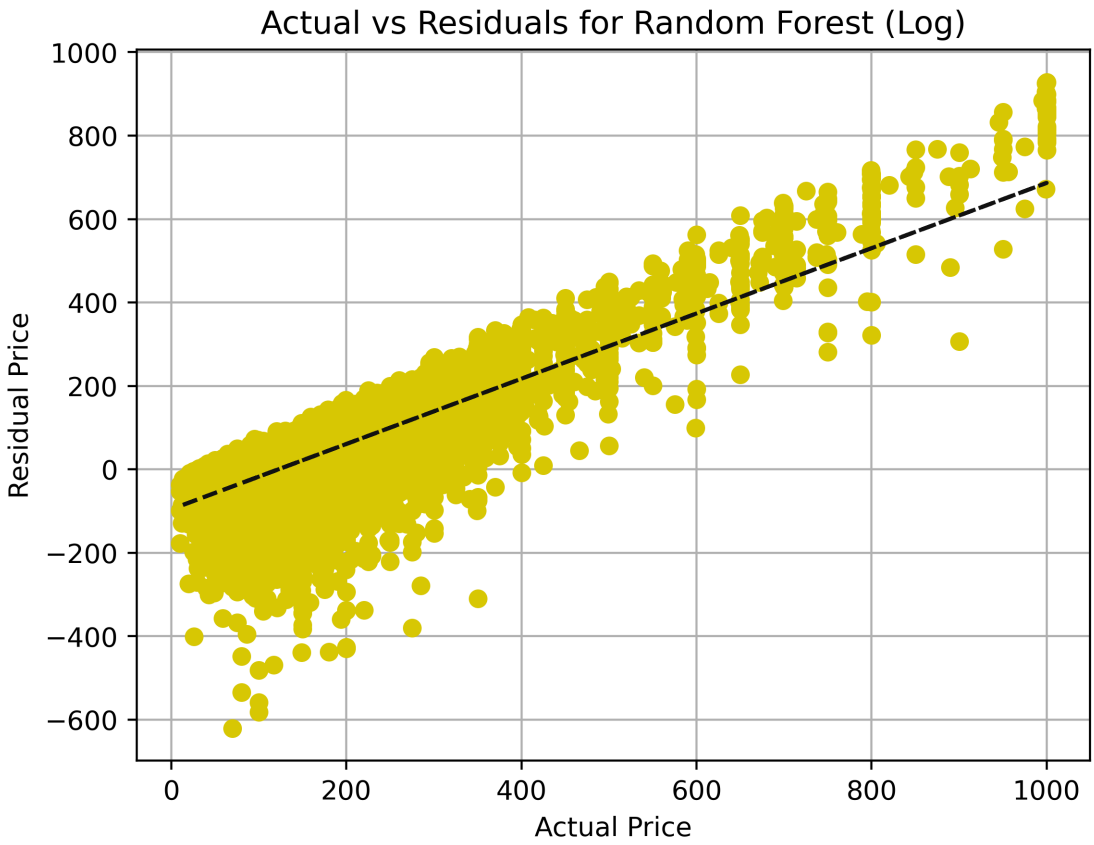


Figure 51: Actual vs Residuals for Random Forest (Log)

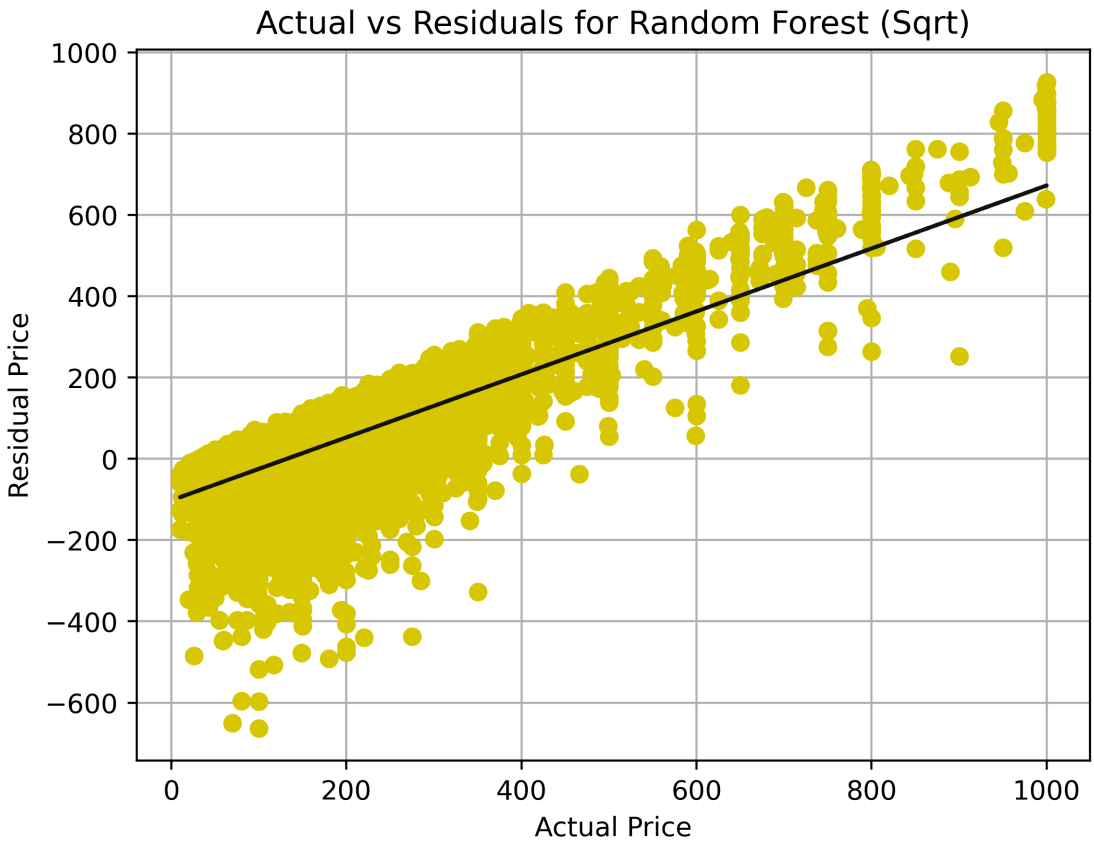


Figure 52: Actual vs Residuals for Random Forest (Sqrt)

8.5 Actual Price vs Predicted Price

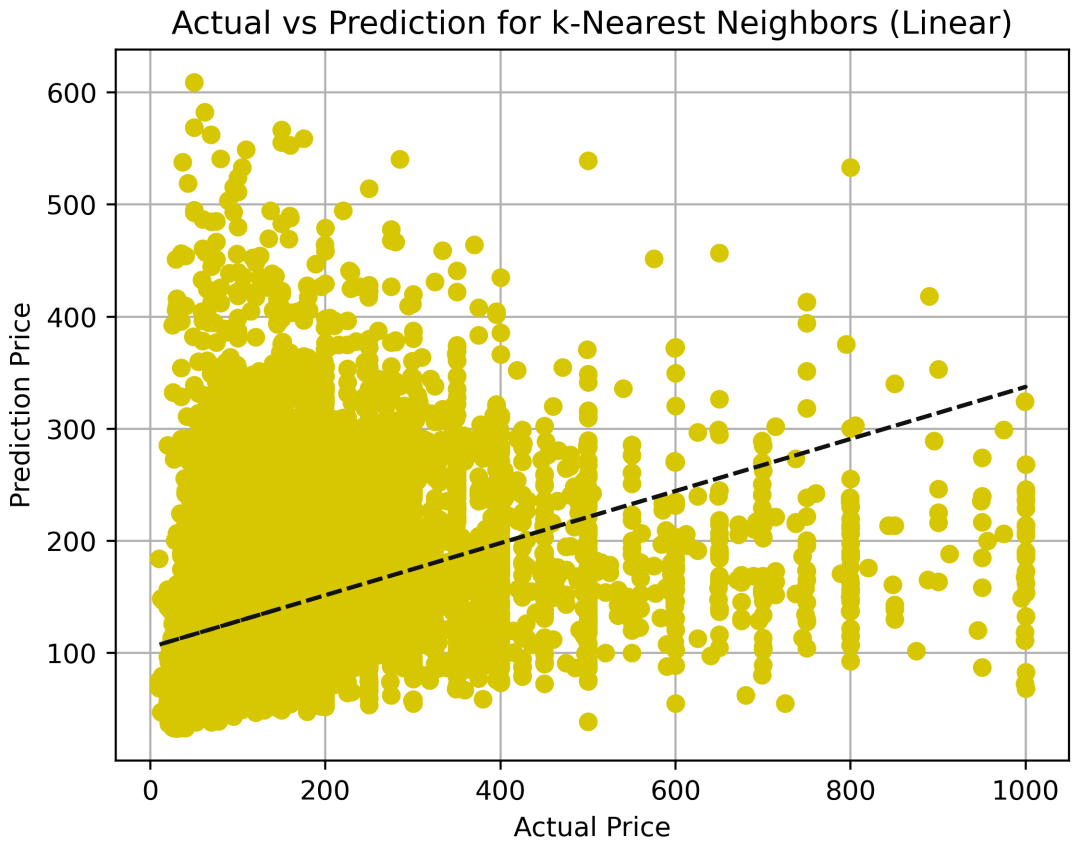


Figure 53: Actual vs Prediction for k-Nearest Neighbors (Linear)

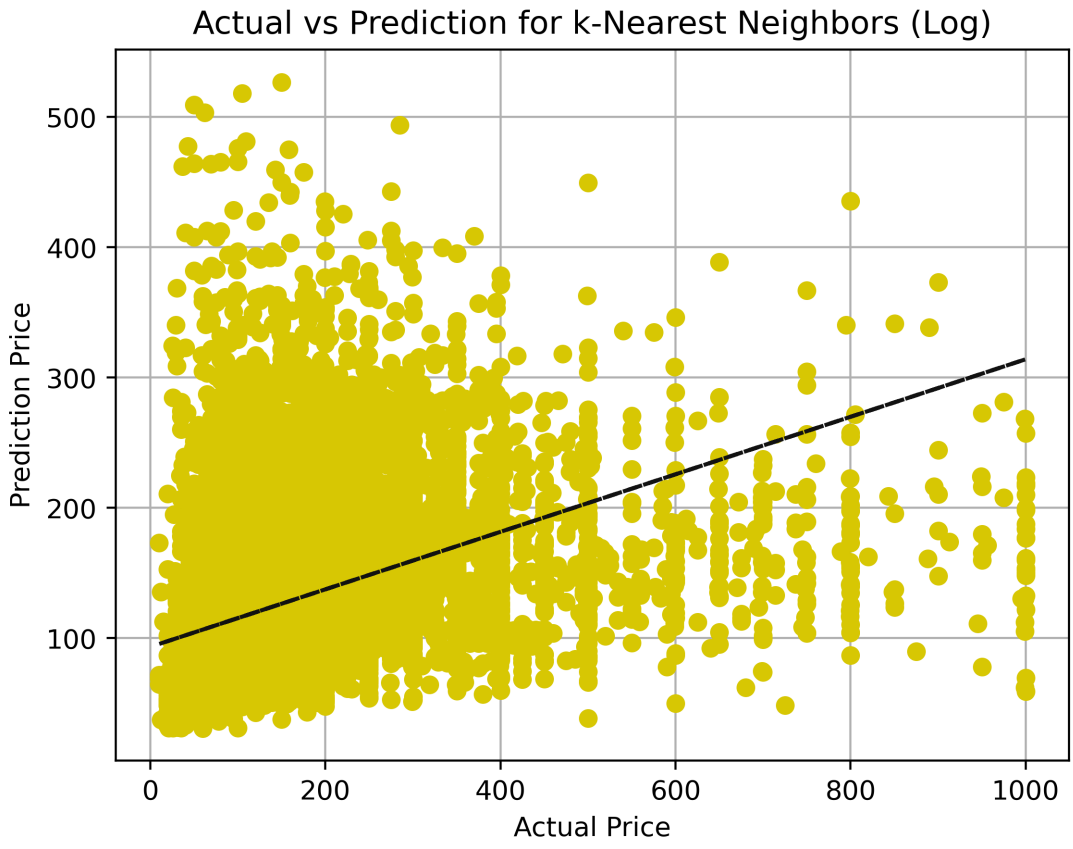


Figure 54: Actual vs Prediction for k-Nearest Neighbors (Log)

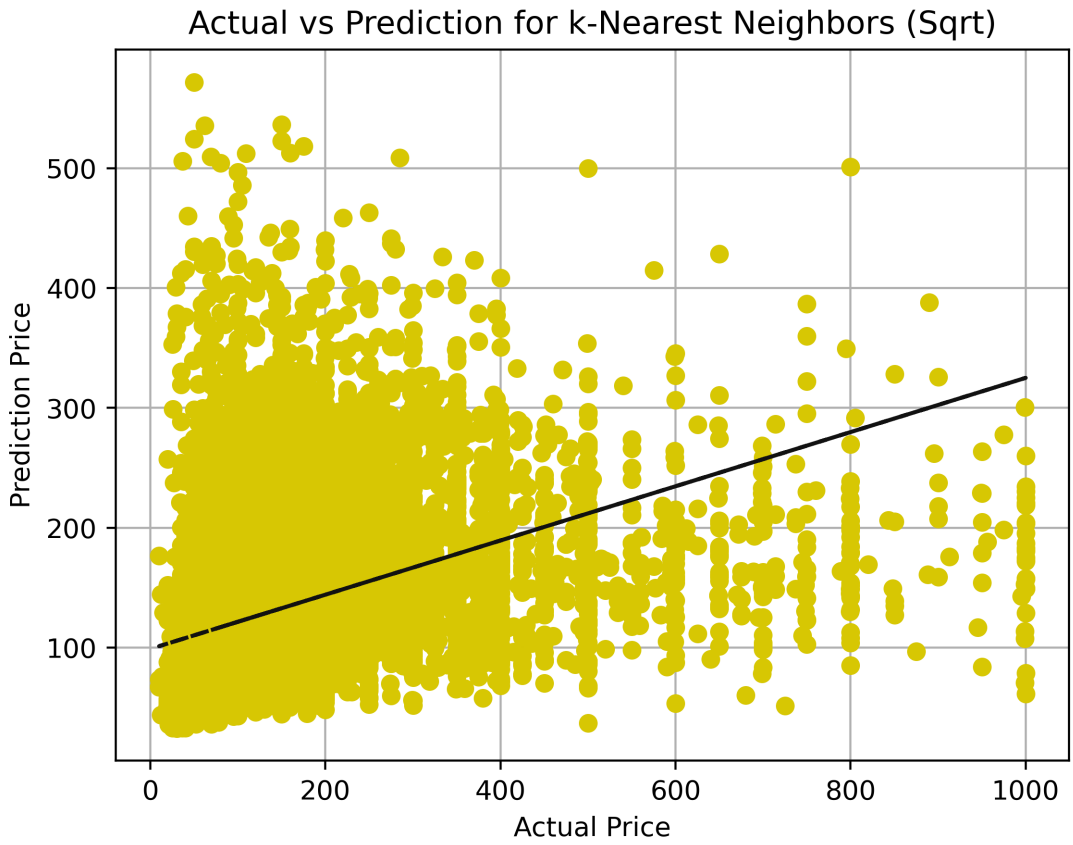


Figure 55: Actual vs Prediction for k-Nearest Neighbors (Sqrt)

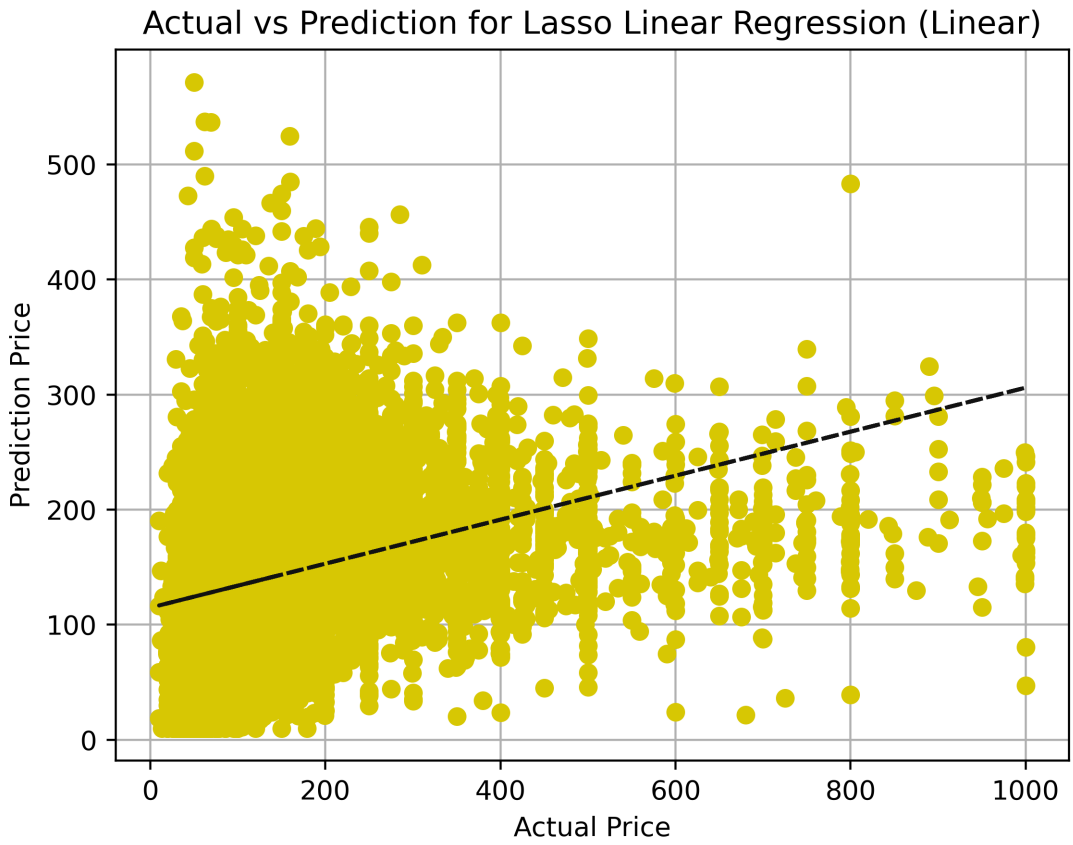


Figure 56: Actual vs Prediction for Lasso Linear Regression (Linear)

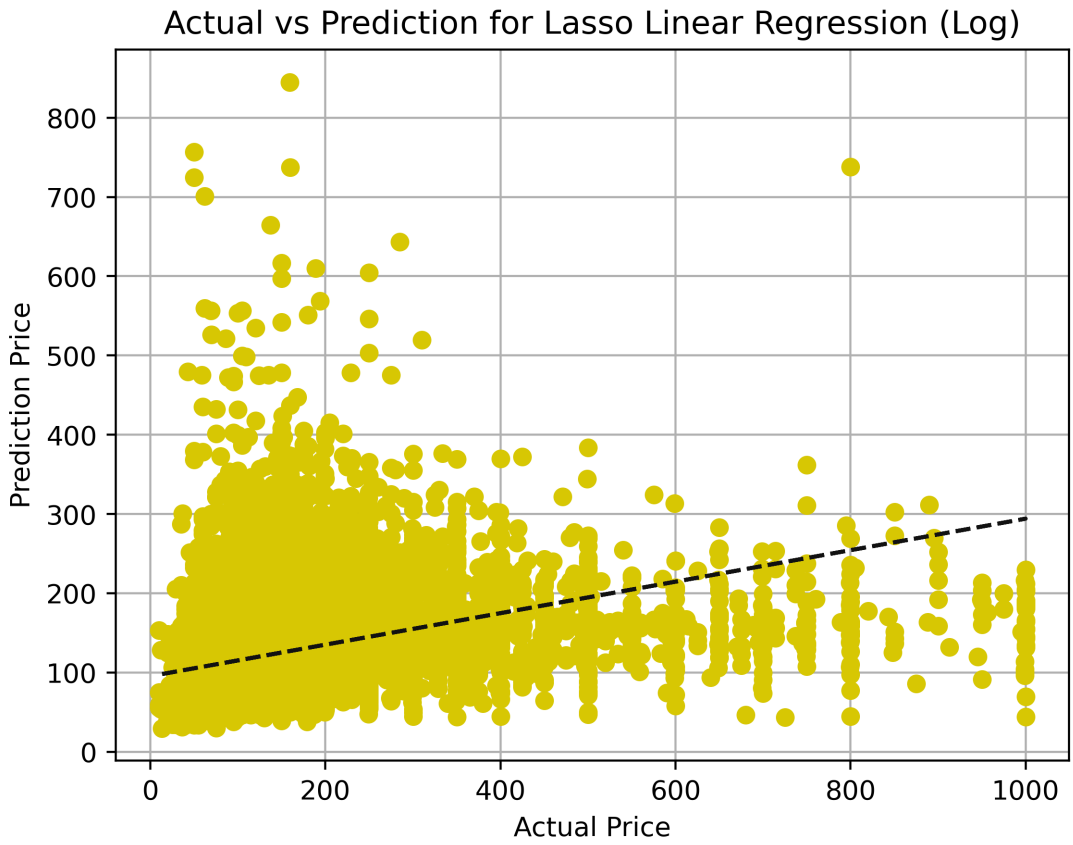


Figure 57: Actual vs Prediction for Lasso Linear Regression (Log)

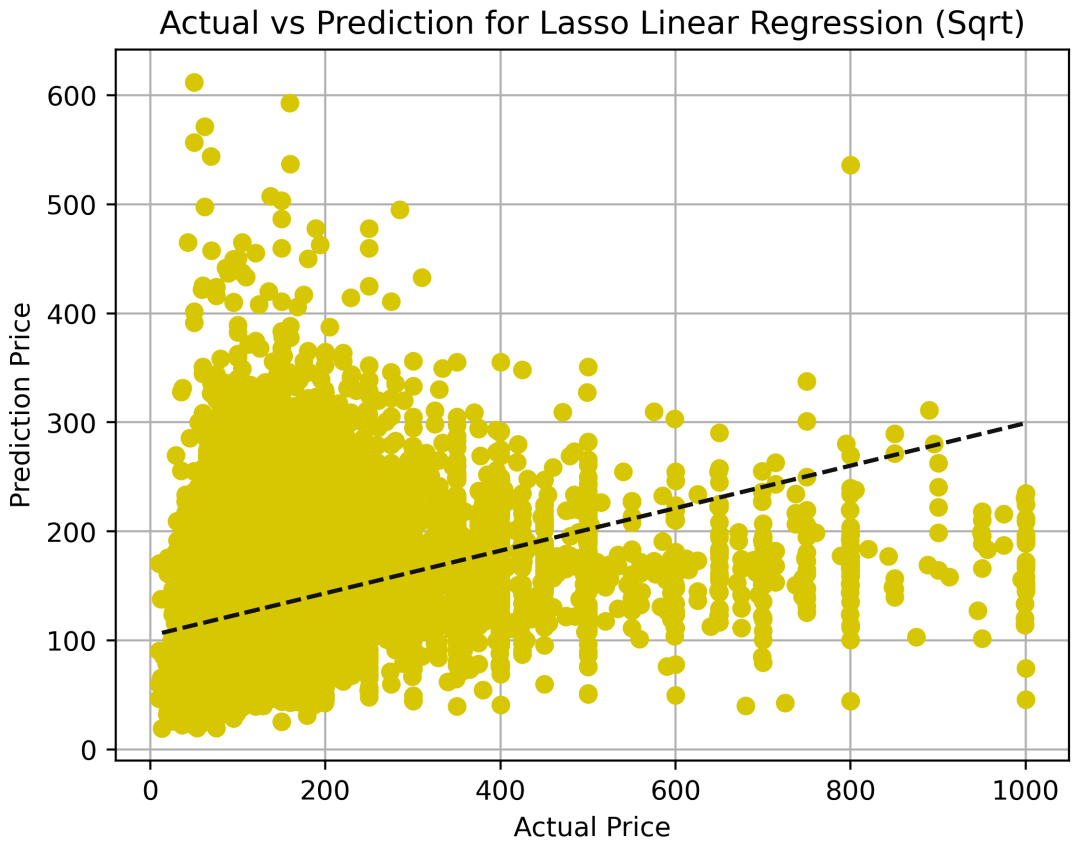


Figure 58: Actual vs Prediction for Lasso Linear Regression (Sqrt)

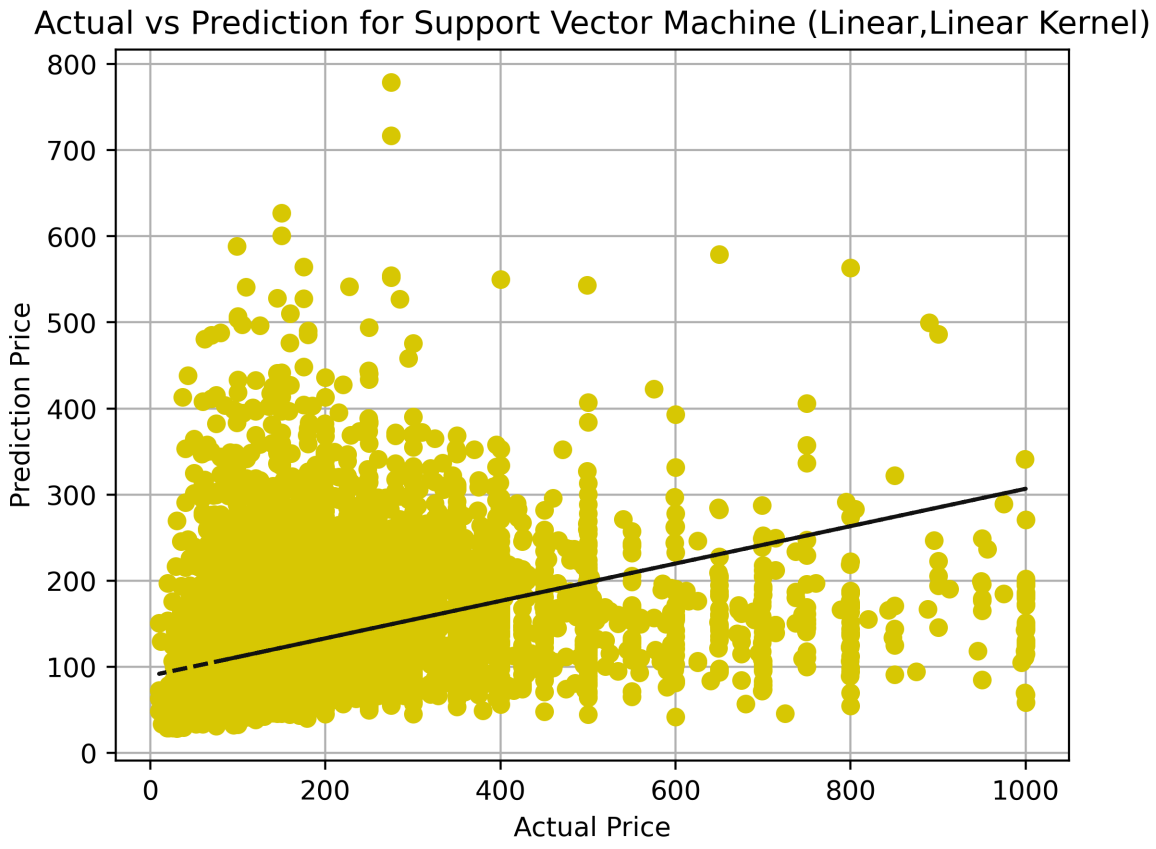


Figure 59: Actual vs Prediction for Support Vector Machine (Linear,Linear Kernel)

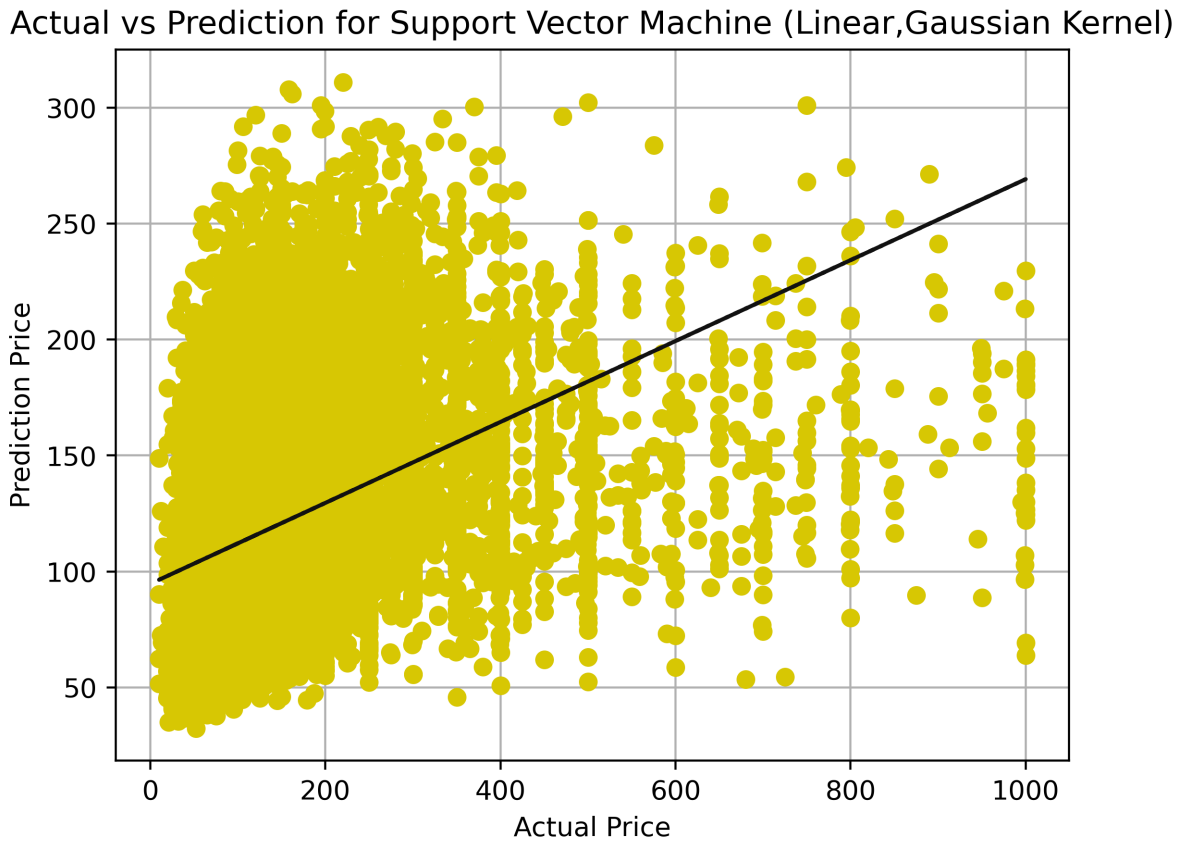


Figure 60: Actual vs Prediction for Support Vector Machine (Linear,Gaussian Kernel)

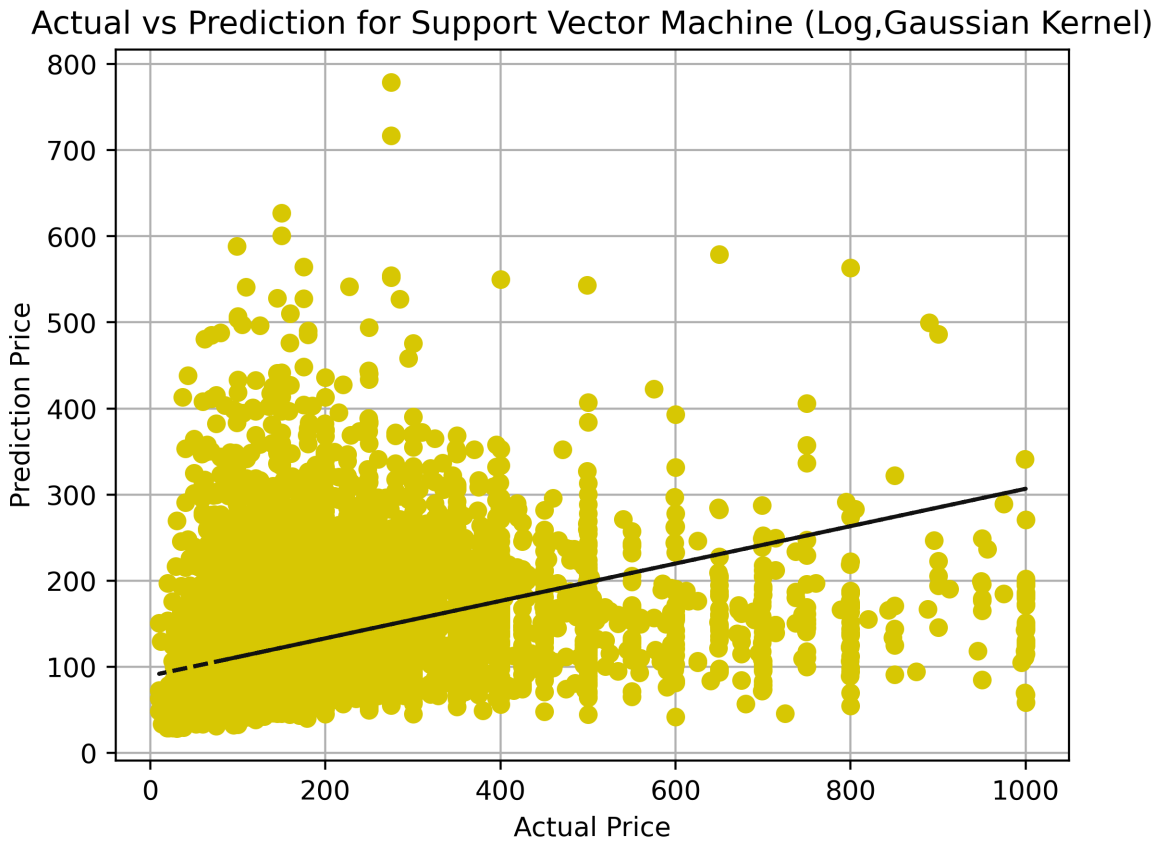


Figure 61: Actual vs Prediction for Support Vector Machine (Log,Gaussian Kernel)

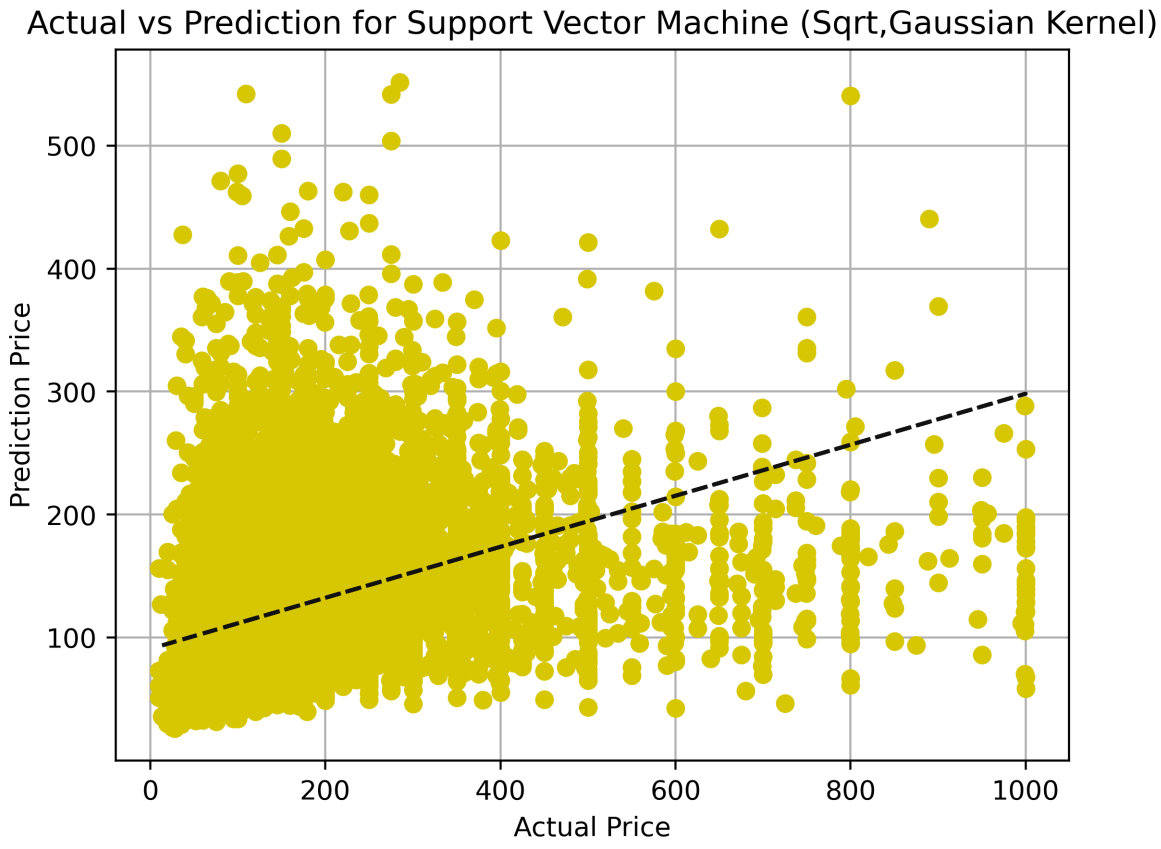


Figure 62: Actual vs Prediction for Support Vector Machine (Sqrt,Gaussian Kernel)

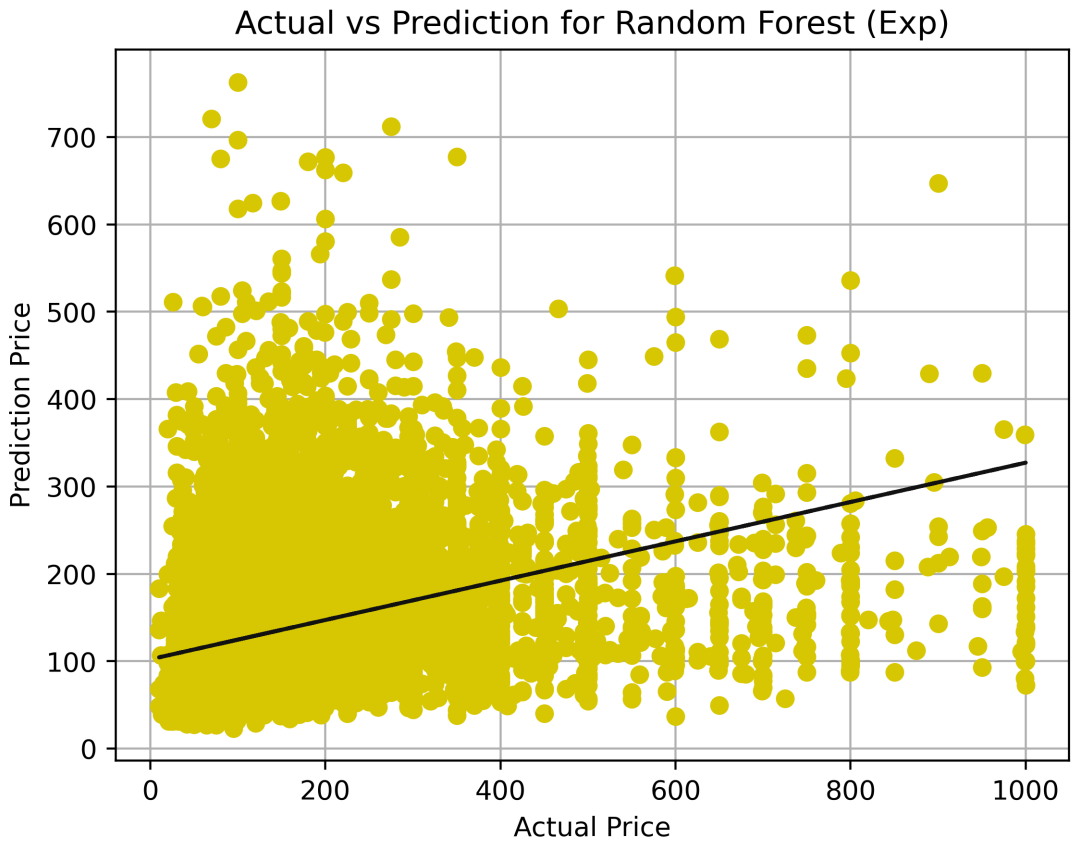


Figure 63: Actual vs Prediction for Random Forest (Exp)

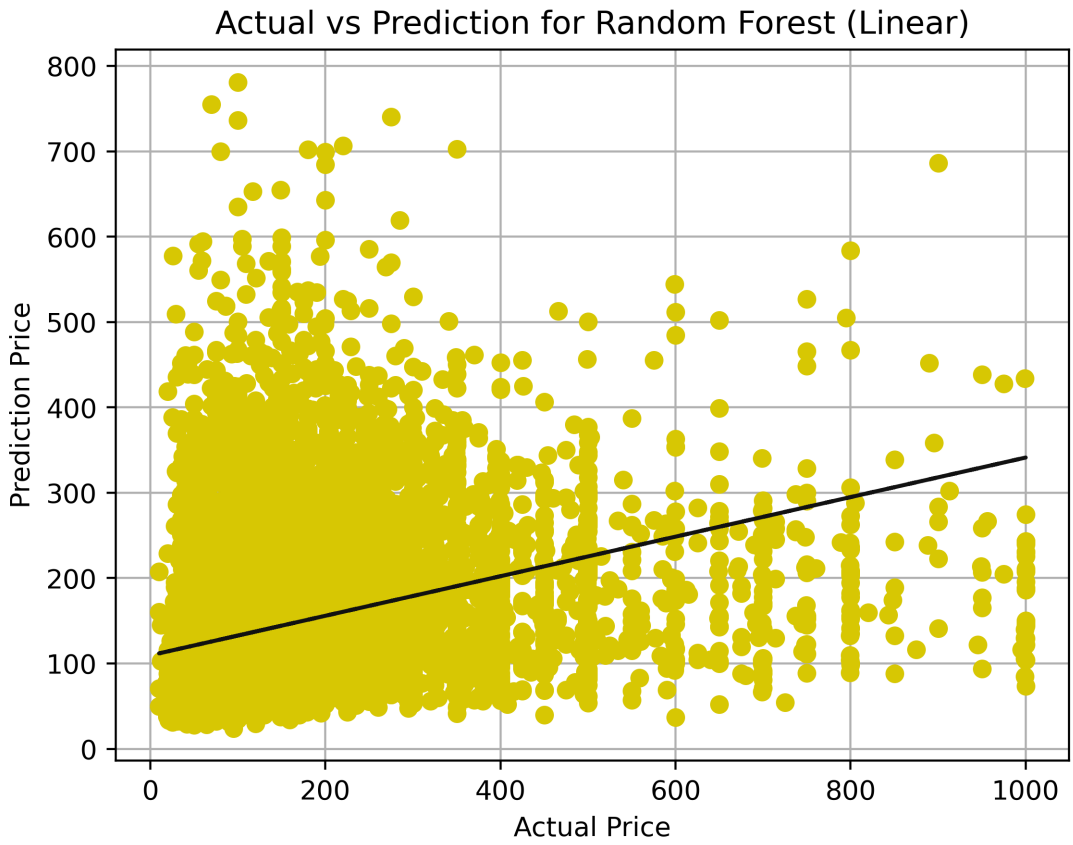


Figure 64: Actual vs Prediction for Random Forest (Linear)

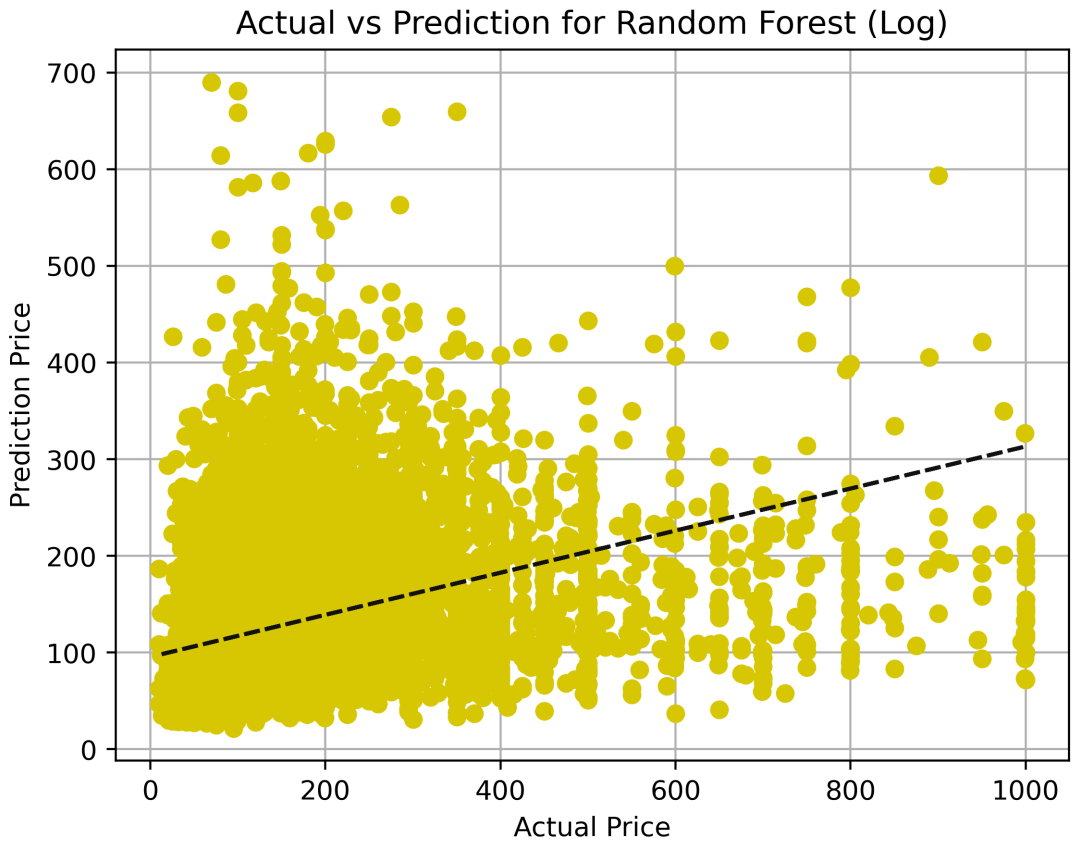


Figure 65: Actual vs Prediction for Random Forest (Log)

9. Supplementary Material

Airbnb_Project.ipynb file for all the analysis done in this project. midterm_proj.Rmd file for data Airbnb_Project.py file for the code in the 'Airbnb_Project.ipynb' file