

## DATA











Lot Area

Garage

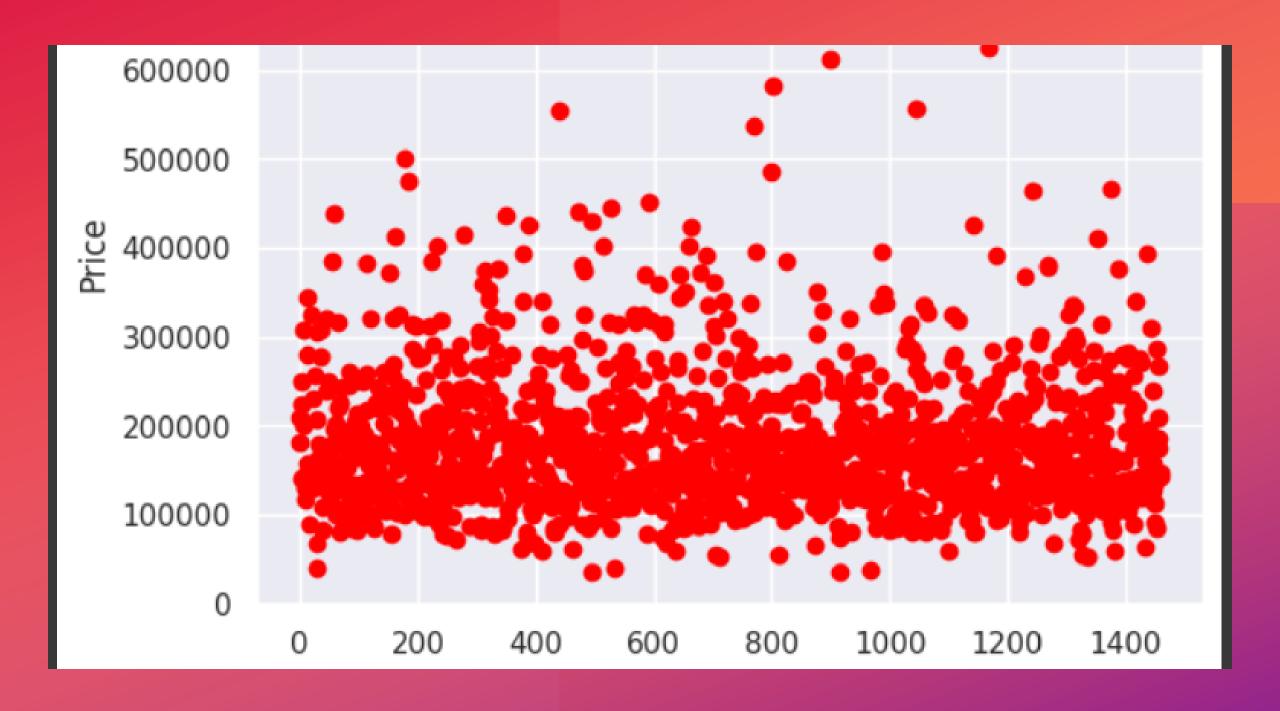
Year built

Materials

Alley

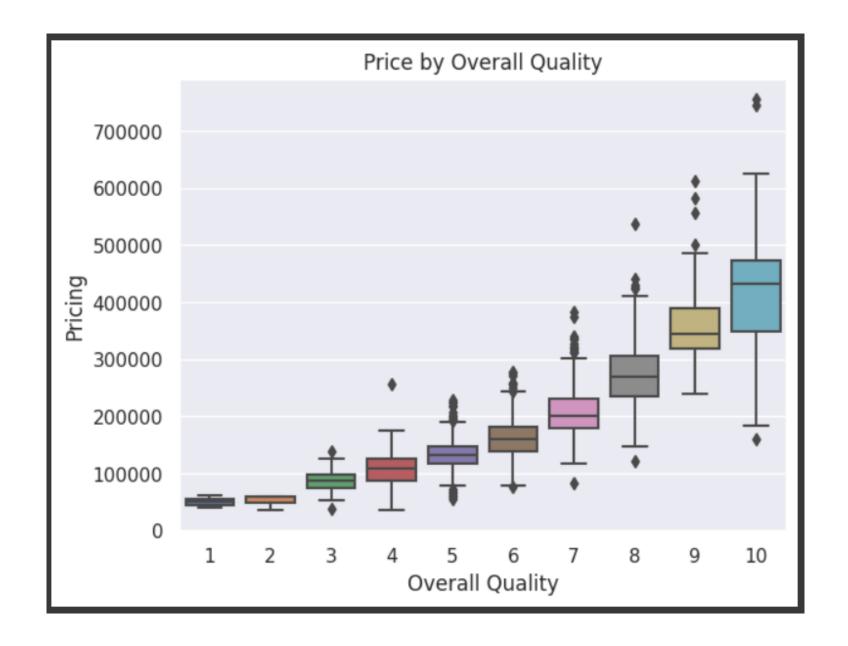


And so on

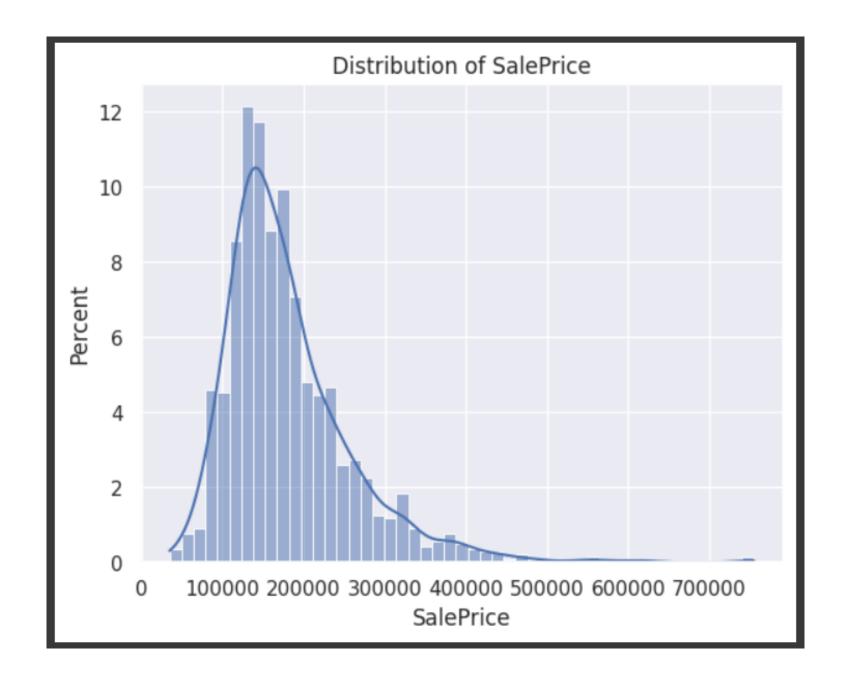


# PRICE BY QUALITY

### AS EXPECTED



# PRICES DISTRIBUTION





#### CLEANING

- IQR top 7 features with most outliers
- i. 798/1460 entries left
- Correlation (62 out of 80 features left)
- i. Higher Bound: +0.6
- ii. Lower Bound: -0.36
- One-hot encoding
- i. Heavy interference with normalization
- Removal of non-numerical features
- . Heavily decreases performance

#### CORRELATION

• YEAR BUILT-GARAGE
BUILT
QUALITY-SIZE OF GARAGE
• BASEMENT (FINISHED OR
NOT) - BASEMENT
QUALITY

MSSubClass 0.03 10.2-0.2.1-0.14 0208.5-050450.09416.10280 070806206.19.107.065207.103.102.0308900206.456-0.0420.048002402014 LotFrontage -0.04.2710.20.240.101.06099927053 0.16.20.240.109.0536692.06320.140.309054307.1 D.D.29.302.08084 0.1 0.04070806680.3026 LotArea-0.020.0.2 74 0970560.702007.16-0.0008960.0.03800053300004407.D9.895009.03.24.08.21.08.20890.0260.043014421047 OverallQual0,002.10,2709 140,286 0, 6, 30, 17 0, 30, 55, 46, 27, 0659, 10, 04, 20, 20, 20, 40, 40, 59, 60, 58, 30, 35 0, 022-0, 90, 908 985, 01 - 0.8 OverallCond-0.0-D41Ql-D1056281-0.440.-D.0-D3012-0.107.20.-D10069702-D1001G57-0.706.D4071G.-D60707402-D.0-D300112-0.0420.08011Ql-004965 YearBuilt0.00482.0601766.4710.70.25.13 0.20.40.26.1-0.10929.15.04.55-2.0-251810.10.80.58.50.24.22-0.00250.0299.91.817 YearRemodAdd0.006078599.920.60.0.7610.107068 0.24.35.26.140.10430.108.020750.108.05820120.1070.48.443.20.25 0.03-0.04030.6001.0014 MasVnrArea-0.002650520.007.340, 10326.1 110.21 0.08734, 350-08206.284, D2030626, 04.0.2004.288, 230, 20.34, 35, 16, 13 0.053-0.008.50484.0006 -0.6D.60.40.30.-D7007075 10.0-D2007.B-D8D599.906609070514.20.2050870.026-0.02.64.7001.943 BsmtFinSF1-0.903.2046503165.107.00.20306082111 BsmtFinSF2 BsmtUnfSF 0.0250940300890.10722.2040807.6 **10.48.37.0561**32.2.40.06.32.09**01**.600.732202424.220.20.00.11.10.00087.00260606.05 TotalBsmtSF0.096.70620.16.56.20140.36.340.4 0.48 10.75 -0.00 B0-270 0839 . D 30 101 101 108 24.36.40 . 46.26 . 22 0.029 -0.020 60 224 90 11 1stFirSF-0.0-D1102240.20.40.20126.25.350.3 0.3 0.75110.-0.004542.107.00.840.203082120.30.29.20.40.406.20.17 0.0390.042909076.911 2ndFirSF 0.0 10428, 1000 8620, 0691 0. 1040 812 17-0.05 625.3 41 00770. 1030 105 60, 70050 00 660.2 20. 10.18. 1030 962 5-0.00 90.0 912 904 918 10 405 LowQualFinSF-0,026:00705080508050869702 199.104906.70 7 70.0003907.10450 100042931.600.8007.702805502046 193.604.505.301.60060 1 90.00646.0047055901.40 3 5 GrLivArea0.000.748.36.26.59.20129.30.334051 0.20.30.49.01.0026026520.50.53409.80.46.20.40.46.20.37 0.0210.0250.080097 0.40.20.1-0.-030460 10.-0308099.40892.06.710.10.10.10.16.20.08 -0.010.001090802084 BsmtFullBath0.000206620201.D.103.00.316.103.10.74 BsmtHalfBath 0.0 105906 0.00504.0042.07604.002.005.0652-0.-0507008501.790.5010.8-1281 11.00802.8-07906.700442 10049-09902.0-00375 0.04-0.-00090204-03019 -0.2FullBath0.003.29.22070157.50.55.50.50.2060730.32.29.34.33807.707.93.38110.13.340.D.53.22.20.50.407.39.14.29 0.0260.096.8030.6018 BedroomAbvGr0.001.065x809507.04.703.068.20.15 0.10601.482.5.02554.083734.2810.19.60.07.00504.50490.403490.4035 KitchenAbvGr 0.0 1927/0-B40-29202 103.10802 0.004 8 9 9 .00 0.007 1810. 20 0.05 812 0.09 70 02. D 607 D .00 9 1 103. 20 1. 108. D 501. 0.4 0.5 D 501. 0.4 0.5 D 501. 0.4 0.5 D 501. 0.0 1 103. D 501. -0.0TotRmsAbvGrd0.004.303.307.103.400.106107.20.208.0660.202.180.30.000109.000109.202.008.00450.402.002.202.008.00450.402.002.000109.202.008.00450.402.002.000109.202.008.00450.402.002.008.00450.402.002.008.00450.402.002.008.00450.402.008.00450.00 Fireplaces-0.00.710.10.240-0.00316.16.20.19 0.02424.29.22.06.640. D1020120.20.00.162.910. D.26.20.28.180.00410.08300.909067 GarageYrBit0.006.61.30-0.06.220.4280.750.02.075 0.24.36.230-0.04.507.10.04.50.4080-501105170.1 10.530.50.26.24-0.00590.93.2026040518 GarageCars 0.06.0 8929.10.6 0.20.58.48.34.14 0.20.40.10.0 05.347.0 20.0040.0 070450 0.807.2 0.55 10.80.20.22 0.023-0.00.0 05.55 GarageArea 0.02/20/22/2.18,5-8, 2055/2.40, 35, 21 - 0.2).46, 46, 103 0 0 346, 106 0 22/9, 102 0 49 0 45/4.2 3).5).88 1 1.26, 23 0.02 5-0.00 30 46 9 10 39 <del>-</del> -0.2 WoodDeckSF0.00890408.22.9.9.06124.20.105.25 -0.001205.2409.6105207.201010314-03094.603105.28.20.20.201013-0.0560.1030020801078 OpenPorchSF-0.0BQ5.08.4893-5.10220.25.D3087 0.10.20.10.20.10.20.090.0B005.9.D904508.20.18.24.20.20.1 EnclosedPorch 3SsnPorch-0,97,79420.00,602.204.2002.60305.802.60.000.802.90890.000066520.001.00402050-05564902.902005393.902595.602.6 1 -0.006684.902.8048 ScreenPorch PoolArea 0.06.503435075643000981-0294.333.5026.0902.602602903340.4002501250.92006388.644001.001.8563333.202.002.00330680.0065 MiscVal-0, 96.40240631.4089104 0.9900694.8047-0.96.-D299.704955.03D0000440393.705.904.00801.902.605.904.802.70680.0490.00 MoSold 0.0**6.40.24.82.D8594.90.00.704**7019 0.0060**4.906.9-8201481.02**048906.02**207.86.506.207.90.438.506.D8.**D6.40.028 0.0020 **810**.16 

Heatmap of Numerical Features

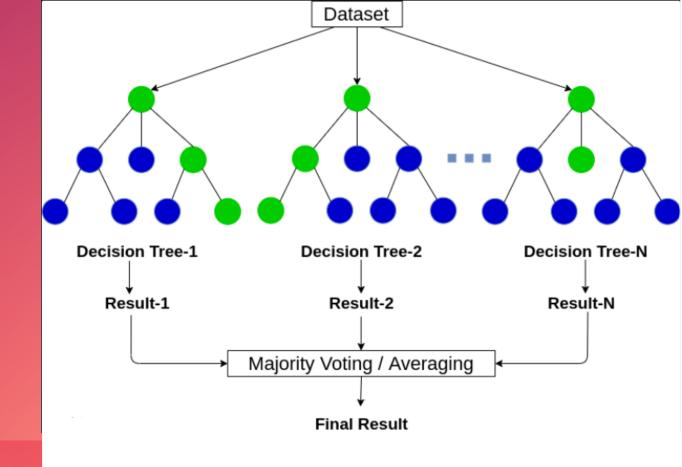
Shape of original dataset: (1460, 81)
Shape of x\_75 before correlation removal: (798, 80)
Shape of x\_75\_final (after correlation removal): (798, 62)

MSSubClass
LotFrontage
LotArea
OverallQual
OverallQual
OverallQual
OverallQual
OverallQual
WasVnrArea
BsmtFinSF1
BsmtFinSF2
BsmtFinSF2
BsmtFinSF2
CarlivArea
BsmtFullBath
FullBath
FullBath
HalfBath
FullBath
FullBath
GarageVrBlt
GarageVrBlt
GarageArea
WoodDeckSF
OpenPorchSF
EnclosedPorch
3SsnPorch
ScreenPorch
PoolArea
MiscVal
MiscVal
MiscVal



### LINEAR REGRESSION

# RANDOM FORESTS



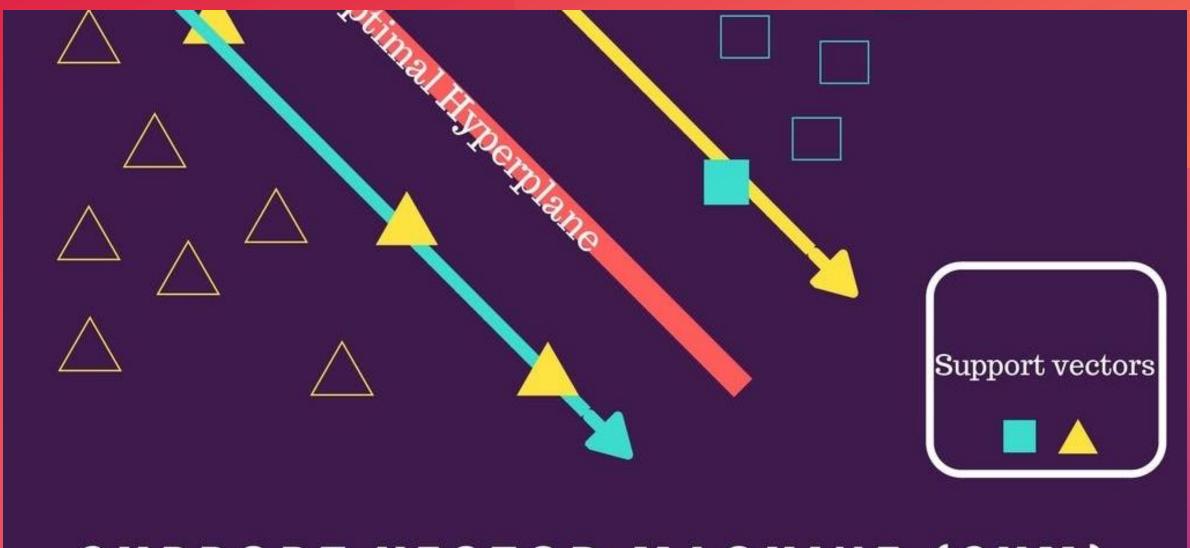
MSE Random Forests: 2410110614.83

MAE Random Forests: 34009.12

RMSE Random Forests: 49092.87743481736

Mean Random Forests: 182755.265

Normalized RMSE Random Forests: 13.655876894246832 %



SUPPORT VECTOR MACHINE (SVM)

# TESTS & METRICS

	Predicted	values (SVM)	Actual/real	values
0		189407		260000
1		197265		143000
2		224325		180000
3		154804		255900
4		239544		128000
195		221820		269500
196		138816		176000
197		167817		202500
198		230674		310000
199		160834		194000

```
# Change in C (regularization parameter) - RESULTS
# NOTE: R SQUARED MUST BE POSITIVE, otherwise we have a problem
# EPSILON = 0.1
# 0.1: r squared is -0.0222
# 1 / 2: r squared is -0.0217
# 10: r squared is -0.0195
# 50: r squared is -0.0079
# 100: r squared is 0.008969
# 500: r is 0.1
# 1.000: r is 0.1863
# 5.000: r is 0.33966
# 10.000: r is 0.37119
# 50.000: r is 0.38363
# 100.000: r is 0.3857
# Interpretation: An R^2 of 0.3857 means that about 38.57% of the total
```

# The rest (very large - 61.43%) cannot be explained using our model

# TESTS & METRICS

	–	Gap (Random Forests)	Gap (SVM)
0	-55779	-97678	-70593
1	69664	46626	54265
2	65186	28815	44325
3	-90098	-66820	-101096
4	104378	64990	111544
• •	• • •	• • •	• • •
195	-41696	-20296	-47680
196	-39303	-12423	-37184
197	-31486	-10415	-34683
198	-70495	-48505	-79326
199	-35120	-5001	-33166



Good	deals / Steals:			
(	Gap (Regression)	Gap	(Random Forests)	Gap (SVM)
103	83979		120357	87172
98	71888		110836	68684
18	56382		93166	55626
163	61363		83022	56396
87	63330		80764	47335
19	65849		78289	59631
188	27402		75069	47973
142	95203		73093	88745
151	28795		71203	22266
164	75603		66426	89225

Bad	deals / Scams:			
	Gap (Regression)	Gap	(Random Forests)	Gap (SVM)
38	-85067		-99937	-93085
186	-83412		-100009	-90934
180	-108675		-107187	-116784
143	-113067		-113345	-118025
67	-124503		-115925	-152268
105	-130736		-123457	-134925
133	-170143		-128091	-178954
42	-144406		-133586	-152032
32	-124225		-133918	-125245
25	-78026		-161012	-100378
24	-237245		-258982	-241774



Thank you

