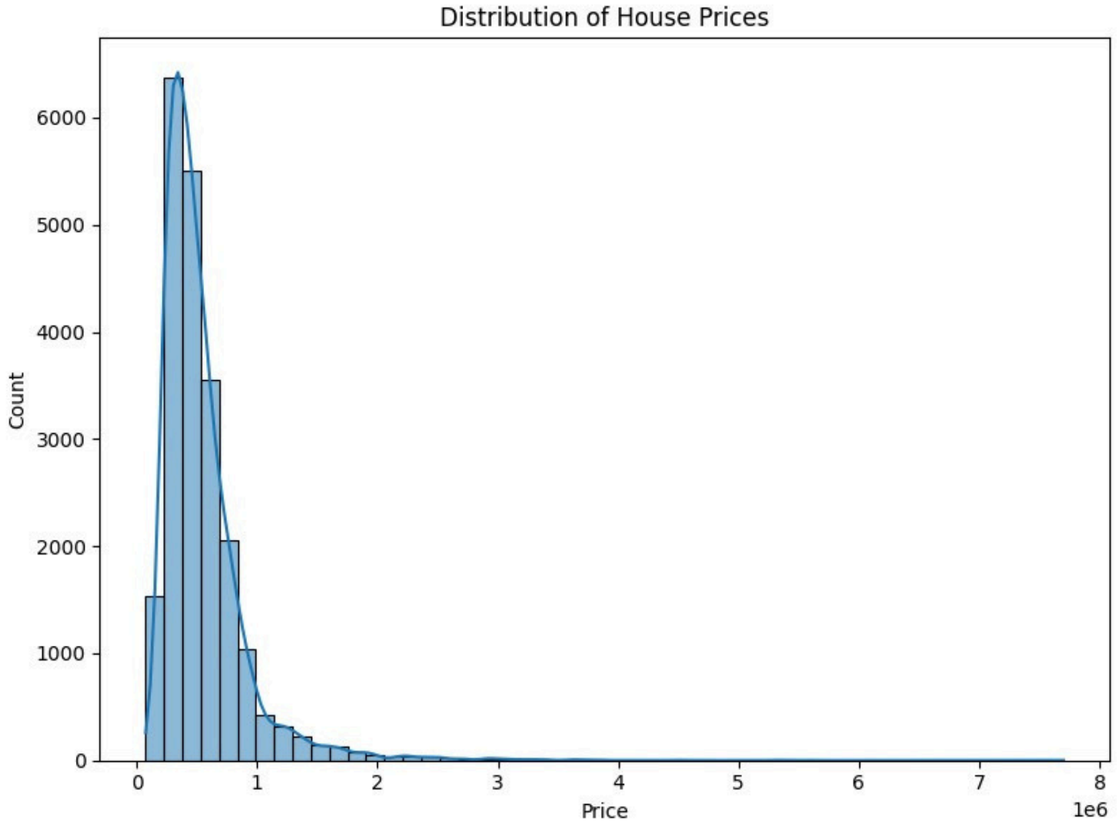
THE SALES PRICE OF HOUSES IN KING COUNTY, SEATTLE.

2014 - 2015

presented by
Ekaterina Kuznetsova

ACTION PLAN

Defining the problem, exploratory data analysis, data processing, feature engineering, model selection, model training, model evaluation and interpretation.



MARKET TRENDS

PERFORMANCE

MODEL COMPARISON TABLE

```
1  evaluate_models(X_train, X_test, y_train, y_test)
```

✓ 1m 10.2s

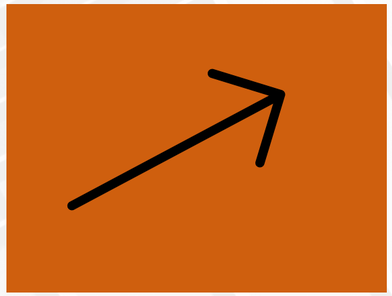
	Model	R2 Train	MAE Train	RMSE Train	R2 Test	MAE Test	RMSE Test
0	Linear Regression	0.7013	124566.5251	197539.5918	0.7034	126807.8815	211755.1798
1	Decision Tree	1.0000	0.0000	0.0000	0.7048	104482.8223	211246.3140
2	Random Forest	0.9835	25340.5261	46441.5539	0.8507	72895.6975	150222.4806
3	KNN	0.6758	126567.8734	205809.9577	0.5038	164991.8859	273877.9857
4	XGBoost	0.9774	38719.6207	54387.6444	0.8658	69609.6814	142414.6545

PERFORMANCE METRICS OF TWO MACHINE LEARNING MODELS - RANDOM FOREST AND XGBOOST

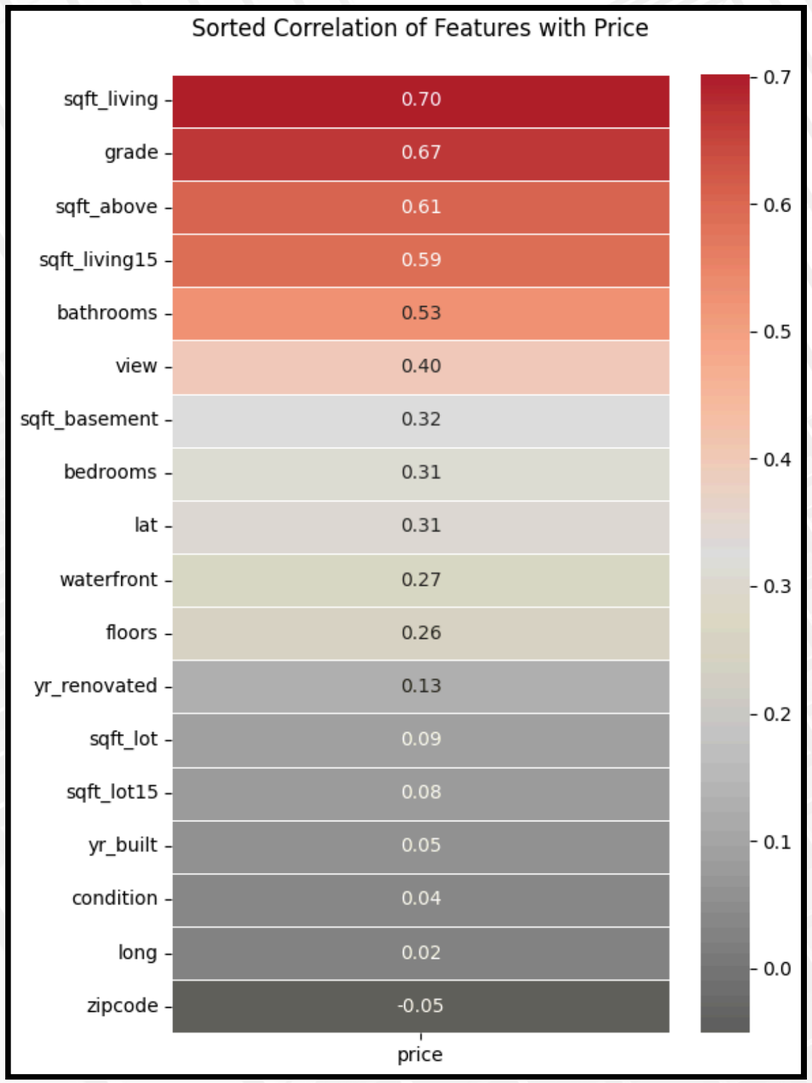
```
1  evaluate_scalers_models(X, y)
2  evaluate_scalers_models
```

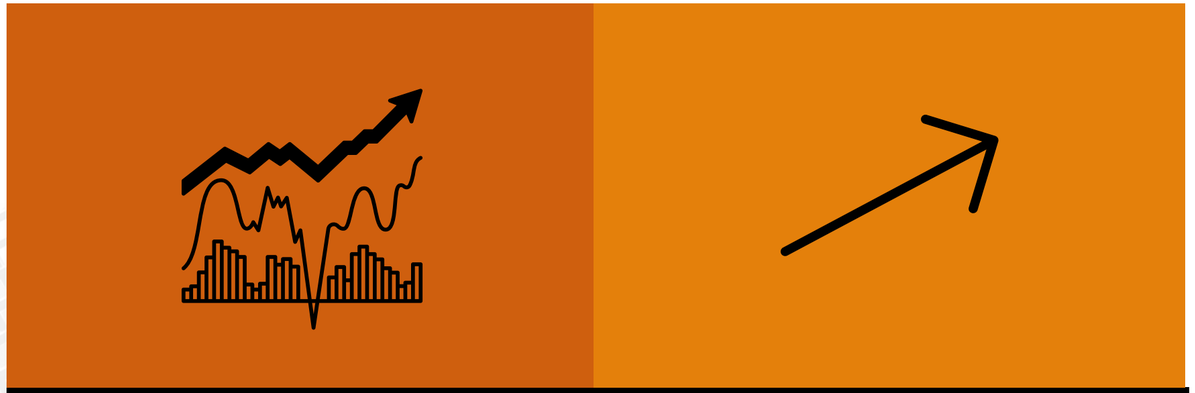
✓ 6m 6.1s

Scaler	Model	R2 Train	MAE Train	RMSE Train	R2 Test	MAE Test	RMSE Test
StandardScaler	Random Forest	0.9835	25347.6259	46448.6285	0.8508	72885.4348	150192.7324
StandardScaler	XGBoost	0.9572	50722.4556	74804.1208	0.8803	69257.9083	134544.8991
MinMaxScaler	Random Forest	0.9835	25343.1833	46440.1085	0.8507	72881.0188	150213.9967
MinMaxScaler	XGBoost	0.9572	50722.4556	74804.1208	0.8803	69257.9083	134544.8991
RobustScaler	Random Forest	0.9835	25338.8235	46428.0294	0.8507	72894.2670	150228.4161
RobustScaler	XGBoost	0.9572	50722.4556	74804.1208	0.8803	69257.9083	134544.8991
MaxAbsScaler	Random Forest	0.9835	25336.8611	46419.1955	0.8506	72928.1637	150279.4238
MaxAbsScaler	XGBoost	0.9572	50722.4556	74804.1208	0.8803	69257.9083	134544.8991
QuantileTransformer	Random Forest	0.9835	25335.8467	46453.2465	0.8503	72993.8638	150445.7362
QuantileTransformer	XGBoost	0.9571	50705.2358	74891.6561	0.8807	69247.7771	134312.5454



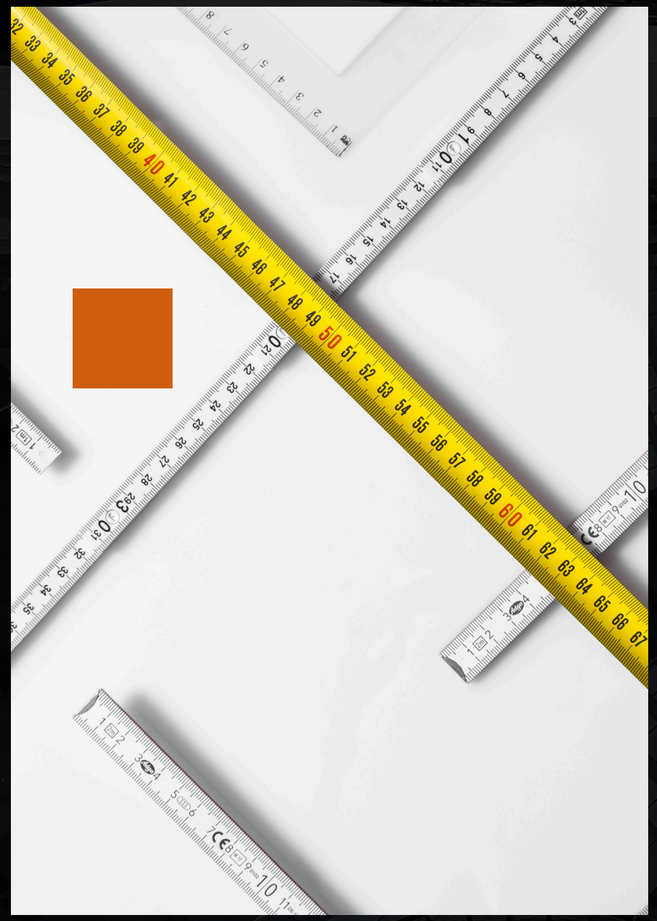
MATRIX CORRELATION





Model	R2 Train	MAE Train	RMSE Train	R2 Test	MAE Test	RMSE Test
XGBoost	0.8484	106431.7068	140743.0711	0.524	153401.256	268245.052

TUNE HYPEPARAMETERS



Model	Max Depth	R2 Train	MAE Train	RMSE Train	R2 Test	MAE Test	RMSE Test
XGBoost	1	0.7998	96024.1616	161738.1544	0.7309	104118.2581	201715.3048
XGBoost	2	0.8932	74565.2742	118096.4766	0.8093	85152.4917	169798.5660
XGBoost	4	0.9484	55080.5514	82113.9086	0.8370	73343.0944	156985.4173
XGBoost	6	0.9774	38627.9935	54294.1478	0.8706	69256.8090	139880.6699
XGBoost	8	0.9926	22212.1003	31182.4475	0.8676	69655.8024	141459.5124
XGBoost	10	0.9988	8354.7403	12479.0243	0.8377	72576.1759	156662.3606

	Model	Max Depth	R2 Train	MAE Train	RMSE Train	R2 Test	MAE Test	RMSE Test
0	XGBoost	1	0.7998	96024.1616	161738.1544	0.7309	104118.2581	201715.3048
1	XGBoost	2	0.8932	74565.2742	118096.4766	0.8093	85152.4917	169798.5660
2	XGBoost	4	0.9484	55080.5514	82113.9086	0.8370	73343.0944	156985.4173
3	XGBoost	6	0.9774	38627.9935	54294.1478	0.8706	69256.8090	139880.6699
4	XGBoost	8	0.9926	22212.1003	31182.4475	0.8676	69655.8024	141459.5124
5	XGBoost	10	0.9988	8354.7403	12479.0243	0.8377	72576.1759	156662.3606
6	XGBoost	10	0.9762	39528.9834	55819.2451	0.8722	65794.4124	139009.5082
7	XGBoost	10	0.9762	39528.9834	55819.2451	0.8722	65794.4124	139009.5082
8	XGBoost	10	0.9762	39528.9834	55819.2451	0.8722	65794.4124	139009.5082

GRID SEARCH WITH CROSS-VALIDATION

OPTIMIZE XGBOOST MODEL

```
Fitting 5 folds for each of 144 candidates, totalling 720 fits
Best Parameters: {'colsample_bytree': 0.7, 'learning_rate': 0.1, 'max_depth': 4, 'n_estimators': 500, 'subsample': 0.7}
Best CV R2 Score: 0.9040034340537584
```

TEST MODEL EVALUATION

```
Test R²: 0.88550617399330 61
Test MAE: 67955.39348036665
Test RMSE: 131562.91342700104
```

THANK YOU!