A3 Tucson Housing Regressor

Contributors: Jason Spaw

Objective:

- Develop price predictor for home prices in Tuscon.
- Models:
    - SLR
    - KNN Classifier
    - KNN Regressor
    - Multivariate Regressor

Feature Engineering

```
clean_house['age'] = 2024 - clean_house.year_built ← Provide better correlation than year_built

clean_house['inv_age'] = 1 / clean_house.age  ← Provide better correlation than year_built

clean_house['price_sqft'] = clean_house.sold_price/clean_house.sqrt_ft  ← For Bins

clean_house['house_bin'] = clean_house.price_sqft//100  ← Bins for KNN Classifier

clean_house['log_price'] = np.log(clean_house.sold_price)  ← Heteroskedastic

clean_house['log_sqft'] = np.log(clean_house.sqrt_ft)  ← Heteroskedastic
```

Results:

Simple/Fast Linear Regression (SLR):

```
[24] eval_metrics(yc_test, y_hat)

     {'OLS': 0.7274, 'R-squared': -3.6875, 'MAPE': 62.245}
```

KNN Classifier:

```
[79] y_hat1 = knn_loc.predict(X_test2.astype(int), 17)


[80] print(accuracy(yc_test2, y_hat1))

     0.46509129967776586
```

Results:

KNN Regressor:

```
# played with various values of k all gave R2 of about 0.33
y_hat2 = knnr.predict(X_test2, 37)
print(eval_metrics(y_test2, y_hat2))
```

```
{'OLS': 0.0101, 'R-squared': 0.3484, 'MAPE': 23.405}
```

- Trial set:
```
y_hat4 = knnr.predict(X_test3, 37)
print(eval_metrics(y_test3, y_hat4))
```

```
{'OLS': 0.0073, 'R-squared': 0.1761, 'MAPE': 18.3402}
```

Predicted vs Actual Home Price - KNN Regressor

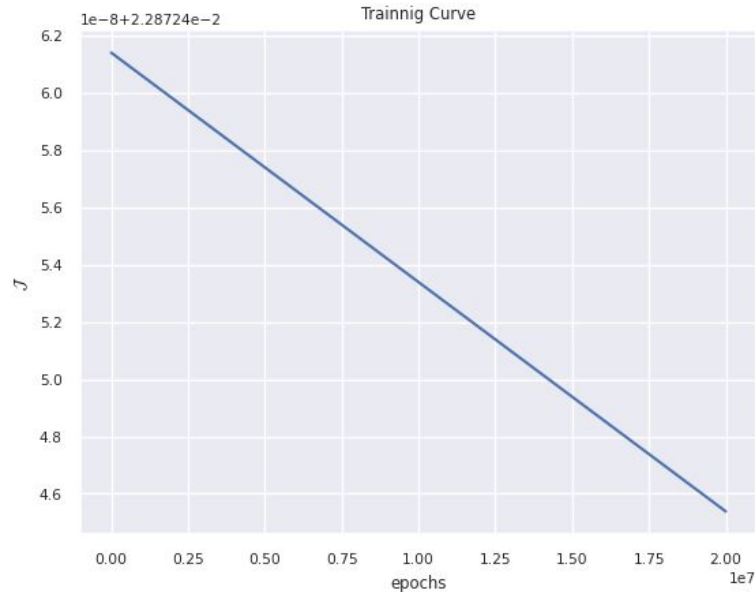# Results:

KNN Regressor (Continued):

```
2282      actual log(price): 0.3926895650585171
990       actual log(price): 0.6434604295776502
608       actual log(price): 0.6707002790677188
1721      actual log(price): 0.5185730384706583
3439      actual log(price): 0.3334901174062088
Name: log_price, dtype: object
------------------------------------------------------
predicted log(price):
[0.37176285 0.48582101 0.47440606 0.54396363 0.42589962]
------------------------------------------------------
actual price:
2282      1.480959
990       1.903055
608       1.955606
1721      1.679629
3439      1.395831
Name: log_price, dtype: float64
------------------------------------------------------
predicted price:
[1.450289   1.62550902 1.60705942 1.72282198 1.53096709]
```
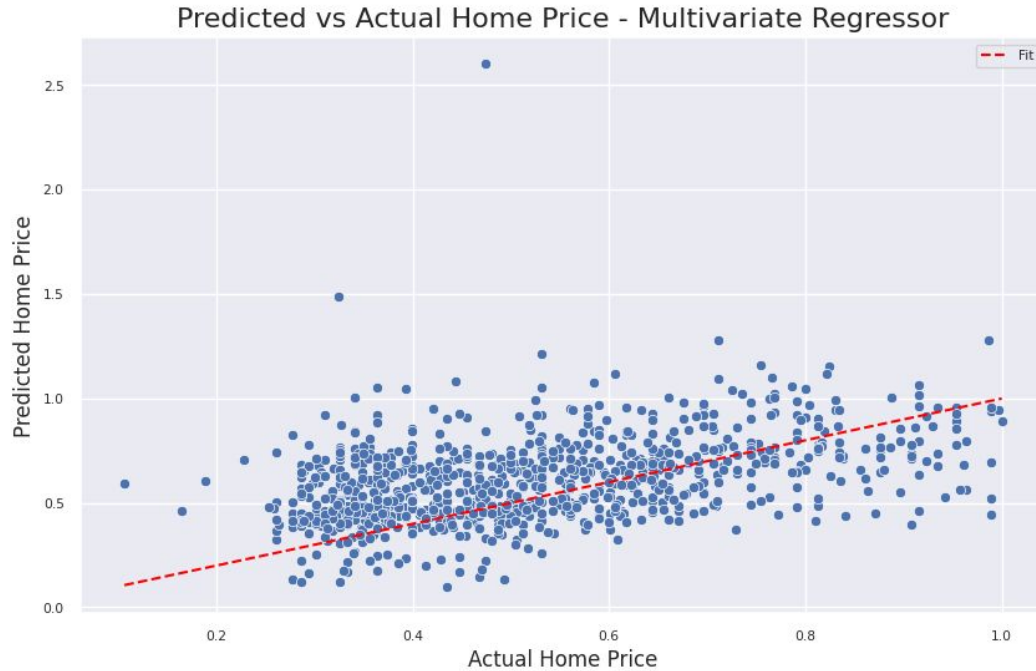
Results:

Multivariate Regressor:

```
eval_metrics(y_test2, y_hat3)
```

```
{'OLS': 0.0237, 'R-squared': -0.528, 'MAPE': 36.6968}
```

Results:

Multivariate Regressor (Continued):

Conclusion

Feature Engineering:

- age
- Inv_age
- Price_sqft
- House_bin
- Log_price
- log_sqft

Models Explored:

- SLR
- KNN Classifier
- KNN Regressor
- Multivariate Regressor

Found Most Viable: KNN Regressor

Questions?

Thank You