Practical Part of the exercise – Tal Getz 212554026

3.1

1

The price is positive, bedrooms are non-negative and discrete, so are bathrooms, view and floors. The waterfront is a Boolean. A living room size is over 290sqft.

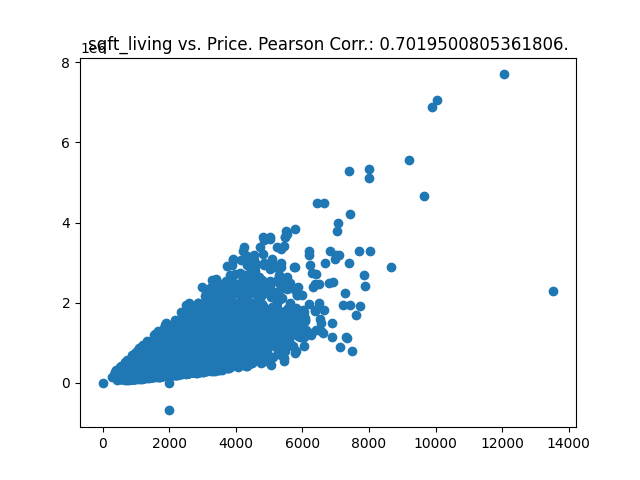
I decided to remove many features among them the categorical ones as I deemed them irrelevant to the fit, due to their unusual values and bad linear correlation to the price.

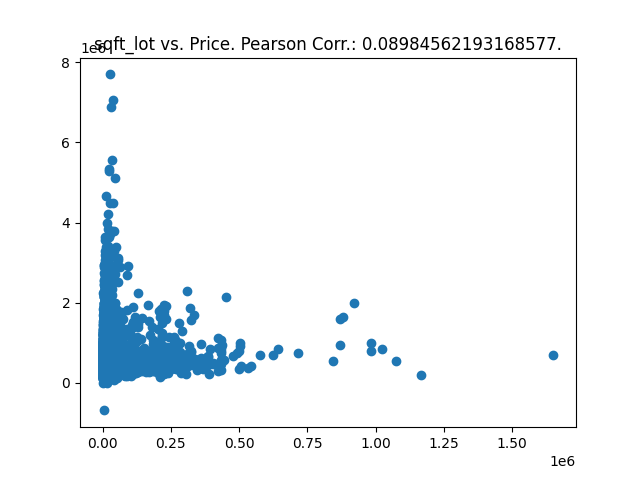
I also created a new “renovated” Boolean value.

I removed rows with missing values as I was not without lots of data.

I split into X and Y as it was stated as optional and I chose to.

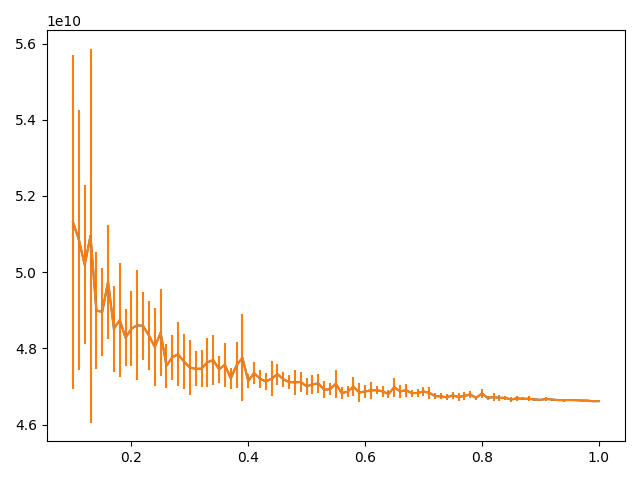
2

the sqft living is quite close to a y=x line in its graph meaning a 1to1 correlation.



The sqftlot has little to no connection to the price. Many points with similar sqftlot and similar price

4

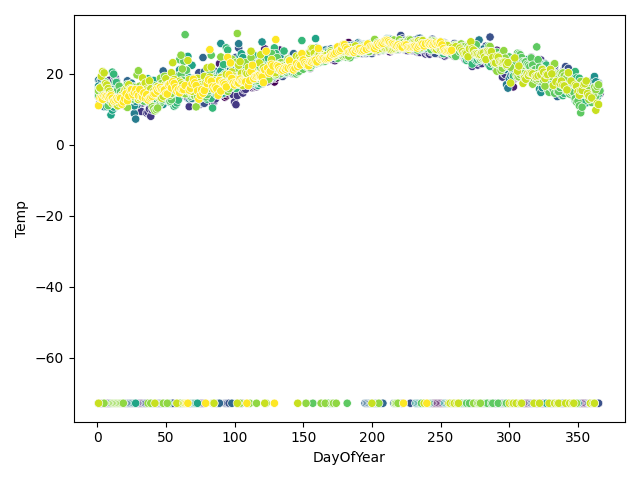


We can see the mean error is slowly limiting a static value of ~4.4 \* 10^10.

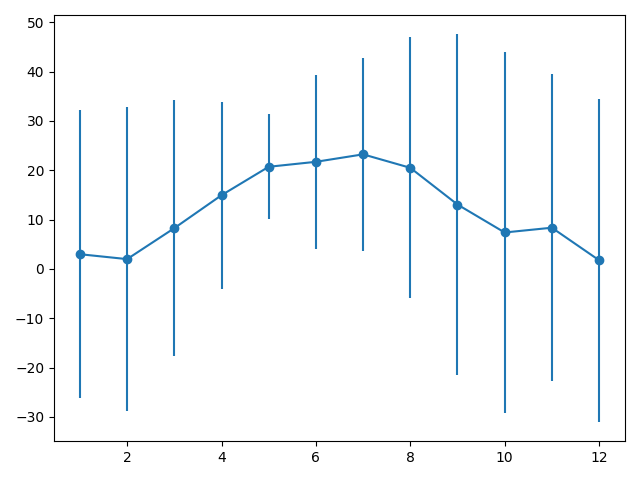
By the smaller and smaller error bar we can infer the stderr is reaching 0, which means the estimator is consistent.

3.2

2

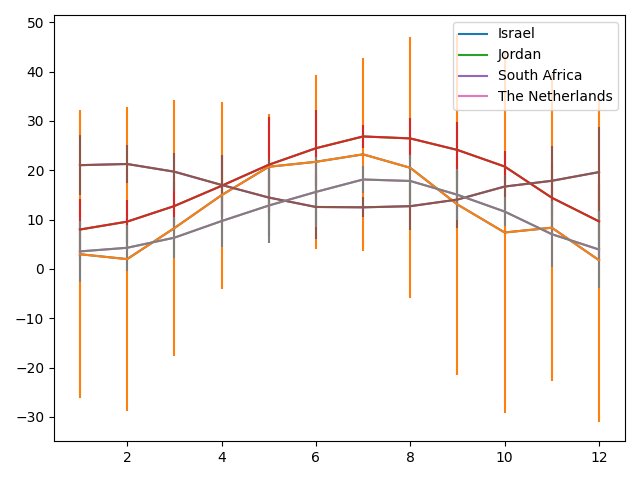


The polynomial degree looks to be around 2-4



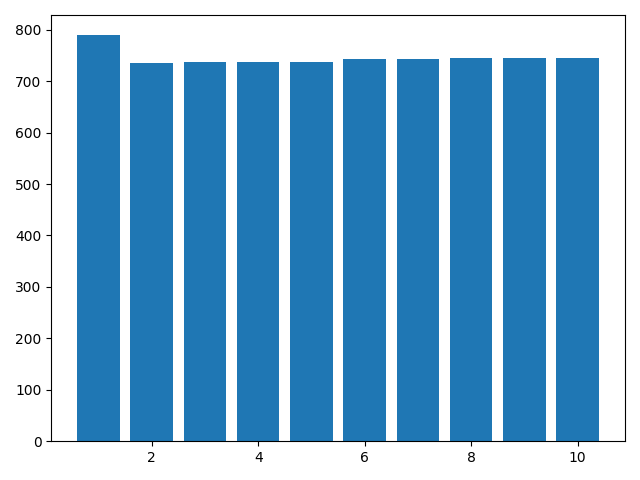
I would assume it would predict better on May (5th month) due to its small error bar.

3

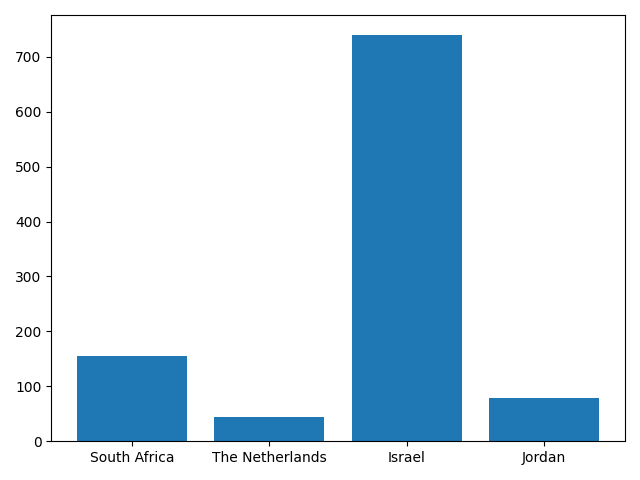


Israel Jordan and the Netherlands share an overall curve appearance therefore, I would assume the model for Israel would work well for Jordan and the Netherlands but not south Africa

\*also I know I have a problem with the graphs here, I’m afraid I don’t have time to fix it ):

4. 

The lowest error is achieved at a second degree polynomial with a close call by 3rd, 4th and 5th degree polynomials. We could consider those, and also a 0th degree polynomial although I believe it would function badly in this case.

5

As predicted in question 3 the graph shows a low error for both Jordan and the Netherlands but a higher error for south Africa.

From this graph I would assume israel’s temperatures aren’t polynomial-like in nature.

4