Multiple Linear Regression to Predict House Prices

What is R-squared? Why does it matter?

Tells us how good our model is at predicting.

R-squared of .95 means that 95% of the variability of home prices can be explained (predicted) by our model.

Results: a model to accurately predict King County, WA house prices

Original model achieved an R-squared of .954 (95.4% of home price variability can be predicted by our model)

Very accurately predicts home prices in King County. . .

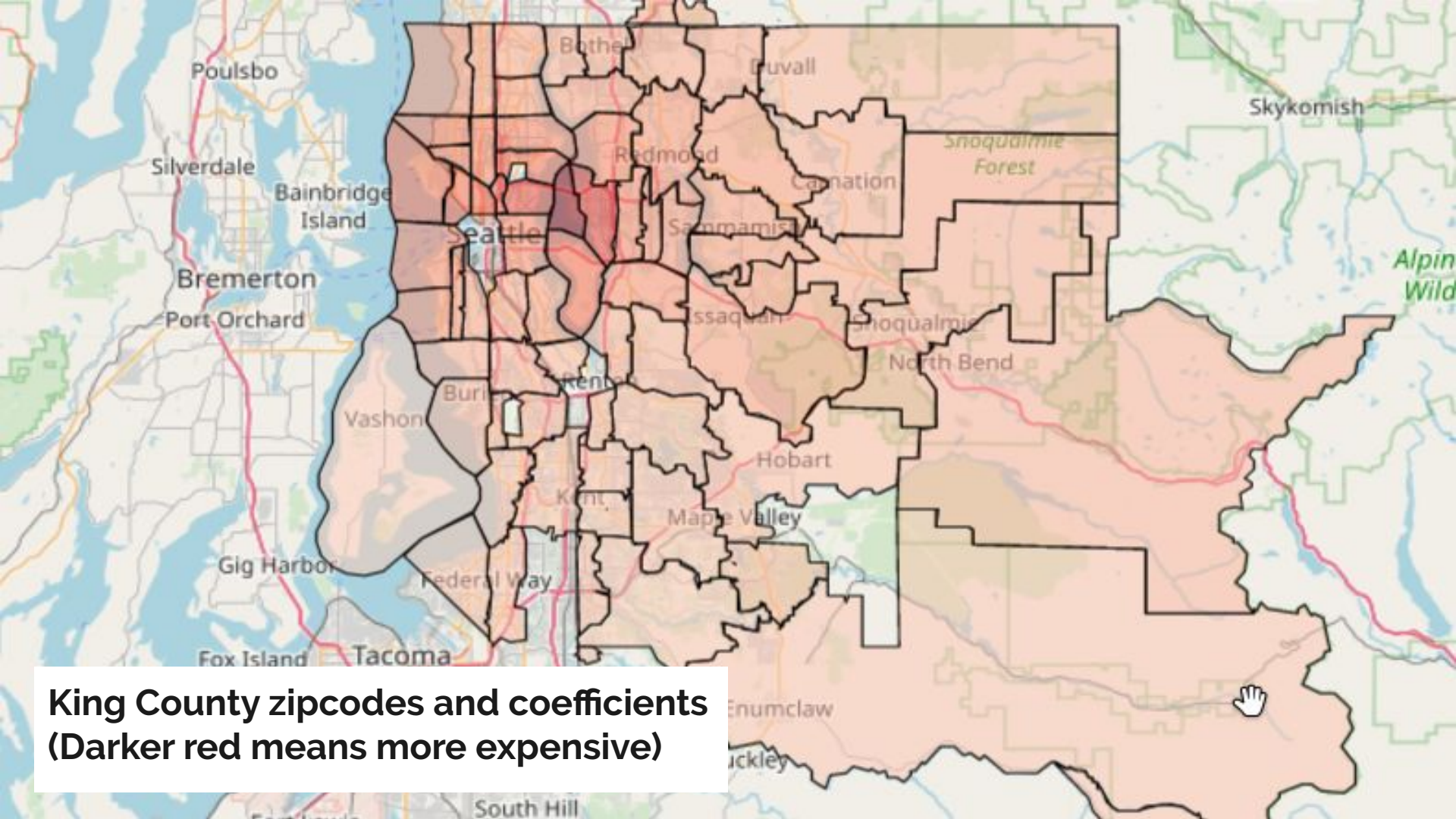
But what if you wanted to predict house prices in a smaller area, like Seattle Metro Area?

Why not just stick with zipcodes as predictors?

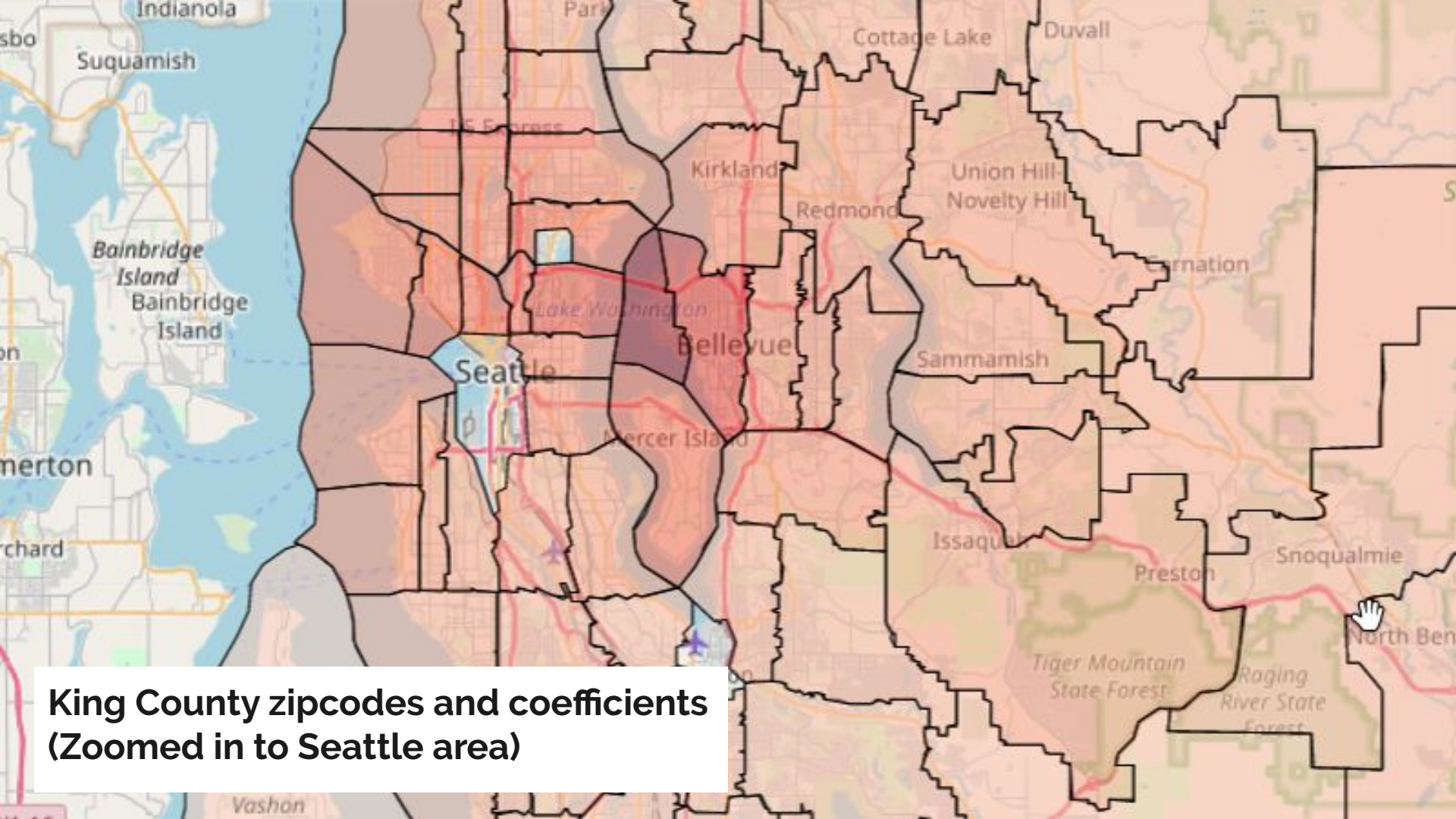
1. Zipcodes are non-standard shapes and sizes. This makes comparing them difficult.

2. Zipcodes are too big for detailed analysis, like the effect of location within Seattle.

3. Zipcodes change, latitudes and longitudes don’t.

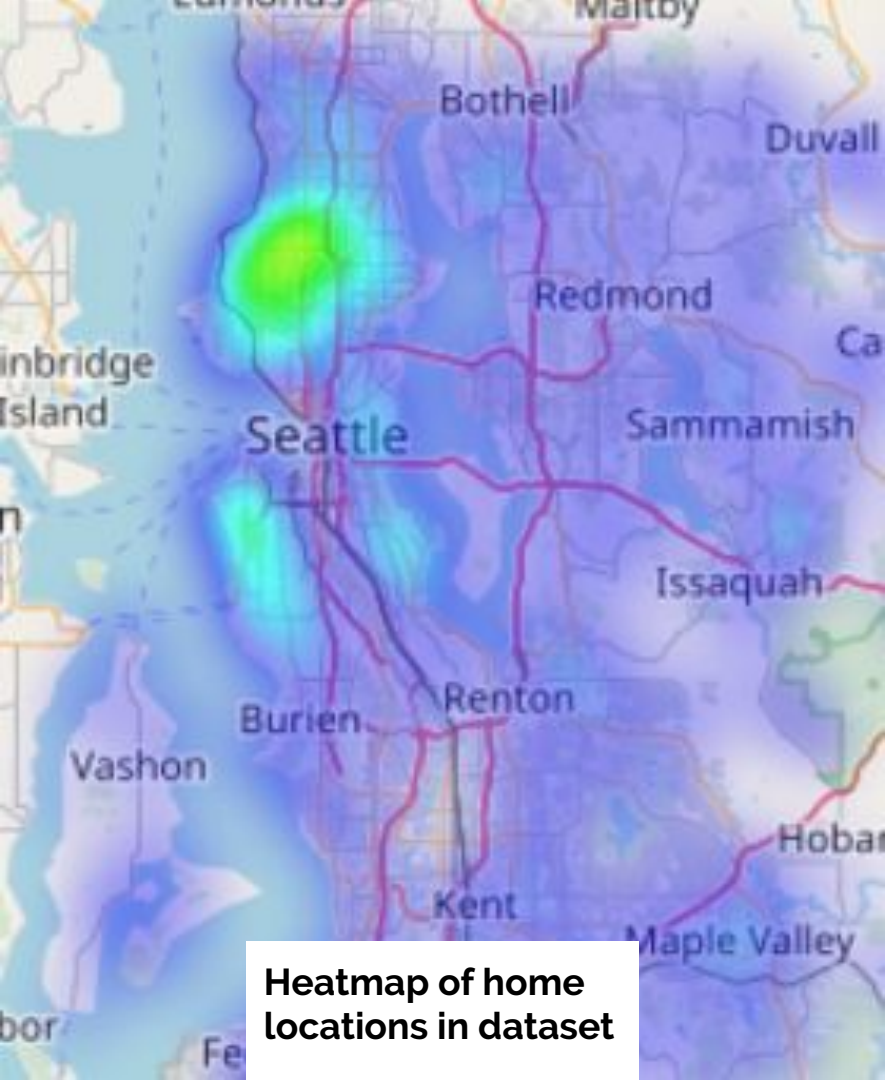


King County zipcodes and coefficients (Darker red means more expensive)

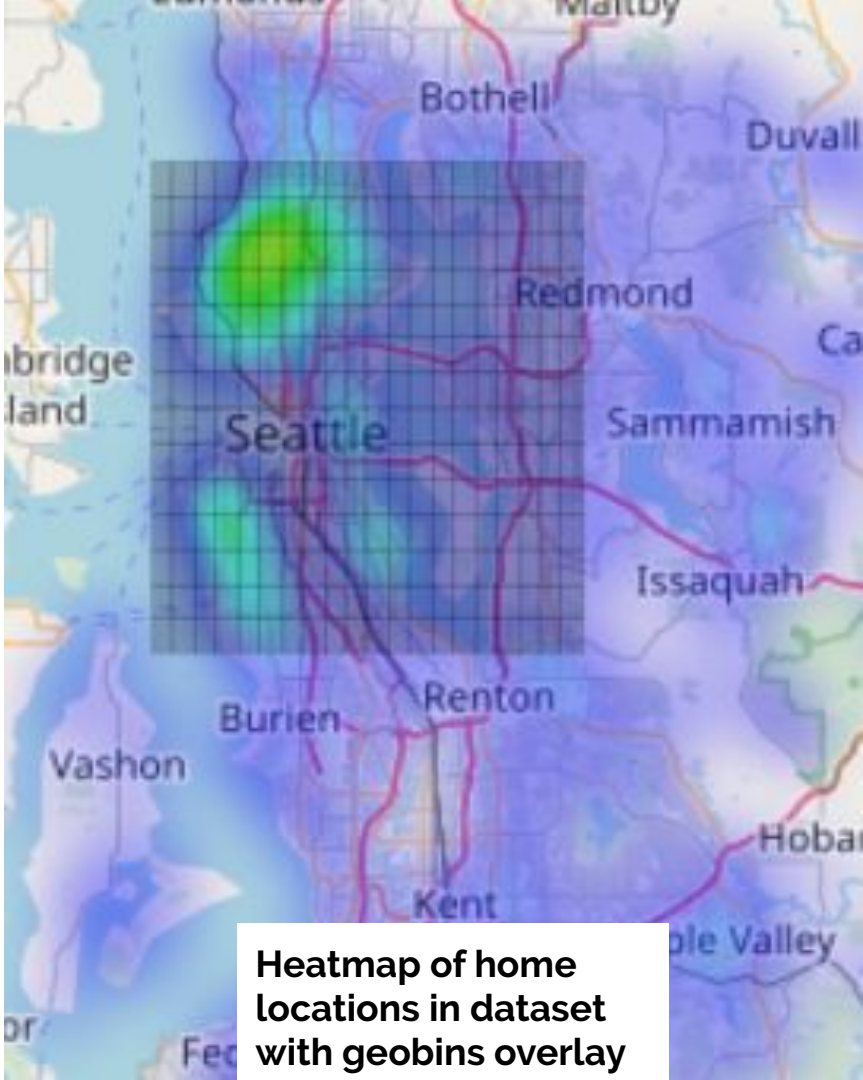


King County zipcodes and coefficients (Zoomed in to Seattle area)

So let’s create our own location-based predictors called “geobins”!



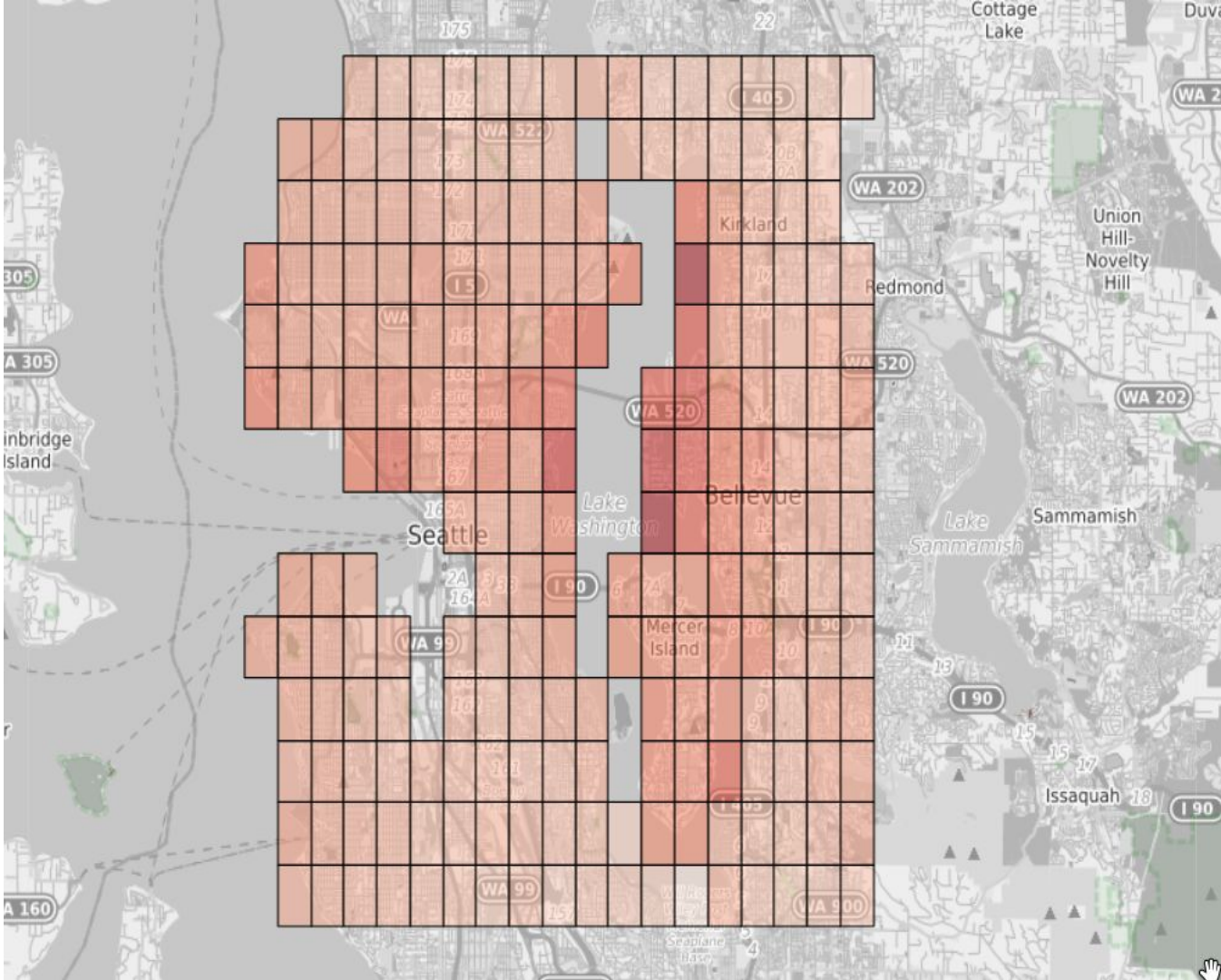
Heatmap of home locations in dataset



Heatmap of home locations in dataset with geobins overlay

Final Result

- Model with location-based predictors that won't change
- Standardized size and shape
- Can be as granular as needed
- Same R-squared as zipcode-based model, predicts just as well



THANK YOU!

Jupyter Notebook at:
https://github.com/benratkin/module_2_project