FINAL THOUGHTS

Challenges/ Obstacles Faced

1. Challenges of processing the Austin listings dataset include the presence of many outliers, the distinction between important and less important features, etc.
2. Challenges of processing the US weekly trend dataset include a considerable amount of missing data, selection and aggregation of the right amount of historical data for prediction, etc.
3. The decision to use a regular ML library like sklearn or to opt for a big data analytics framework like pyspark
4. Challenges to choose the machine learning algorithms and models that are best suited for the prediction of house prices
5. Data linking between the two main datasets and how we would combine the national and local trend to tell a convincing story

Potential Next Steps/ Future Direction

1. Addressing model performance issues. Some of our models suffer from either suboptimal scoring or overfitting. There's room for improvement given more in-depth and sophisticated approaches of pre-processing and modeling
2. Data linking. Although our two datasets do differ moderately, we still like to make some connections between them. Future study can be done comparing cities within Texas, or Austin vs. other major housing markets in the U.S., etc.
3. Advanced analysis. Our datasets also come with a huge amount of text and image data. It would be an interesting direction to add in NLP and CNN components to our existing ML modeling to paint a bigger and clearer picture