Rex Partner Report

Dear Rex,

It has been our pleasure collaborating with you on this project. This email is to give an overview summary of our current stage of work, the challenges we are navigating and where we are heading next. Additionally, we have also attached our technical report which gives a more detailed breakdown.

As discussed, the aim of the project is to predict supply and demand for homes in the Denver area. We are approaching this with a Bayesian hierarchical model that will output probabilities of a house being listed or sold on a per home basis and these predictions will be aggregated to output market level predictions. Initially, these predictions will be for a fixed time period but, given adequate performance of the model, we hope to stretch to time series data by the end of the project.

Current Stage:

In the past few weeks, we spent much of our time cleaning up the data and set up baseline models for predicting supply and demand in the Denver metro area. We use data sources from MLS, REX and Census Bureau to construct a dataset that consists of transaction, housing profile and residential community features. After cleaning up the data, we built a simple Logistic Regression model to predict the probability that a property will be sold in Q2, 2019, and achieve 73% accuracy.

Another two areas that we worked on include experimenting with predicting days on market and researching on the Bayesian model. The reason we chose to predict days on market beside our main target variables is because 1) we think days on market can also be a good proxy for supply and demand and 2) we want to see the strength of our model in different settings. We also started working on the Bayesian model by implementing the hierarchical model on PyMC3, and we hope to provide more concrete results from this model in the upcoming weeks.

Future Plan:

In the subsequent weeks, we intend to merge the updated dataset and complete data cleaning for both baseline (non-Bayesian ML) and developed (Bayesian hierarchical) models. We will also explore further exploratory data analysis with true DOM data and

implement SMOTE (generate synthesized data to tackle the imbalanced classification) for the baseline model. Lastly, the Bayesian hierarchical model will be fully implemented with fine-tuned prior parameters.

Please let us know if you have any suggestions or concerns regarding our project, and we look forward to our next chat on Zoom!

Sincerely,