Final Group Presentation Script

Have you ever tried cooking a meal that you didn’t have a recipe for? You just had a ton of ingredients and basically started doing stuff until the idea became more concrete? That’s how we built our Machine Learning model.

We had data on weather, on quality of life, on government spending, and travel destinations, we just had to create a model. So, we decided to create an arbitrary percentile for quality of life, and gave each city a 1 or a 0 based on it.

Once the ‘high\_q’ measure was created, we accounted for overfitting by dropping the features that would be far too obvious for a ML model, like the quality of life index and the city\_rank. And now it was time to start cooking.

We created both a Logistic Regression Model and a Random Forest Model. The Random Forest Model worked way too well, with 100% accuracy. So, we dropped more features and n\_estimators, but still came back at 100%.

The Logistic Regression Model had a 98% accuracy, so we had a good idea our data was valid enough.

A couple of issues surfaced: the quality of life measures were merged by state, so we dropped the raw quality of life index measure as a feature. The government expenses were flat amounts, not per capita, so the hope was StandardScalar would account for that.

Regardless of which model was more accurate, the beauty was in the Feature Importances measure. It turns out that the features with the highest impact were Safety, Healthcare, and Government Expenses on Outdoor Activities (including travel and tourism). Keep that in mind when looking to invest in rental properties.