Write a summary of the work in this notebook. Capture the fact that you gained a baseline idea of performance by simply taking the average price and how well that did. Then highlight that you built a linear model and the features that found. Comment on the estimate of its performance from cross-validation and whether its performance on the test split was consistent with this estimate. Also highlight that a random forest regressor was tried, what preprocessing steps were found to be best, and again what its estimated performance via cross-validation was and whether its performance on the test set was consistent with that. State which model you have decided to use going forwards and why. This summary should provide a quick overview for someone wanting to know quickly why the given model was chosen for the next part of the business problem to help guide important business decisions.

A: To start this process we began with average prices. Trying to predict with this shows that there is significantly more to the ticket pricing than just the prices hanging around the average. After testing that out how correlative the average price was, we moved on to a basic linear model. The basic model showed some predictive value with an R2 score of great than 0.7. there was some incomplete data and to get sort that out we imputed the values comparing both median and mean. We found the predictive value to be similar in both cases. We moved on from there to refining the model by building pipelines. With the pipelines we were able to deduce that the model is over fitting to the training set. Using pipelines, we were able to find a set of features that more closely predict ticket prices. We moved on from there to building a Random Forest model. Using the median imputation for the Random Forest using cross-validation we learned that its performance wasn’t significantly better than that of Linear Regression. Using a mix of mean and median for imputation only improved slightly from there. Ultimately the Random Forest model outperformed the Linear model by enough for us to settle on using it moving forward. It had lower error and varied less than the other model used. One key take away the business should consider is that more data is needed if they want more accurate models with greater predictive capacity.