* Description of preliminary data preprocessing
  + In our preprocessing, we cleaned and filtered out any columns that would not be necessary for the model. Specifically, we took out 'state', 'location', 'address', 'chip\_latitude', 'chip\_longitude' columns. Additionally, we converted some columns into int or float variables that were orginally registered as string variable i.e. 'National\_rank', 'People/Sq.Mile', and 'Population'. Finally, before we encoded all object types into zeros and ones, we created a bin called 'Other' where the primary type count was less than 1000.
* Description of preliminary feature engineering and preliminary feature selection, including their decision-making process
  + For our features we used all the columns after they were encoded and dropped the target column which was 'chipotle'.
* Description of how data was split into training and testing sets
  + To split our data into training and testing sets, we used the train\_test\_split function from sklearn package. By default, the package splits the dataset into 75% training data and 25% testing data.
* Explanation of model choice, including limitations and benefits
  + For our model, we decided to use a supervised machine learning model with resampling and ensemble techniques.
  + In oversampling, its limitation is the algorithm can be heavily influenced by outliers and lead to noisy data. With under-sampling, the it does not work with small dataset. To overcome this limitation, we used a technique that is a combination of oversampling and under-sampling techniques.
  + For ensemble technique the main benefit is that bias/variance can be reduced and most of the times thus the model is not underfitted/overfitted. However, this technique is also less interpretable so the output is hard to predict and explain.