Prediction of rate of success of the Starbucks stores in particular area

Capstone project

The following code will predict customer rating (i.e. the rate of success) of Starbucks store in particular location:

1. This code uses the data from foursquare.com to get the store number, latitude and longitude and customer rating.
2. Using this lat and long values, the Pandas code retrieves the attributes with for starbucks location around Dayton OH within distance 100km
3. As an additional parameter the "cost of living" and "livability" rates were introduced to characterize particular neighborhood. That information was taken from the realtors database: https://www.areavibes.com.
4. The classification and clustering were performed to classify data into two sets, one with high customer rating (1) and one with lower response (0). The corresponding Map is provided.
5. The Decision Tree and Gaussian NB algorithms were applied to train model on this data set.
6. Next step was to pick (randomly) other set of stores around Columbus OH within distance 100km.
7. Again, using this lat and long values, the Pandas code retrieves the attributes for all starbucks location.
8. The classification and clustering again were performed to classify data into two sets: one with high customer rating (1) and one with lower response (0). The corresponding Map is provided.
9. The obtained "features" data was used on initially trained Decision Tree and Gaussian NB models to predict customer rating for this new set of stores. It seems prediction works well (100%).