
THE GEORGE
WASHINGTON
UNIVERSITY

WASHINGTON, DC



Introduction

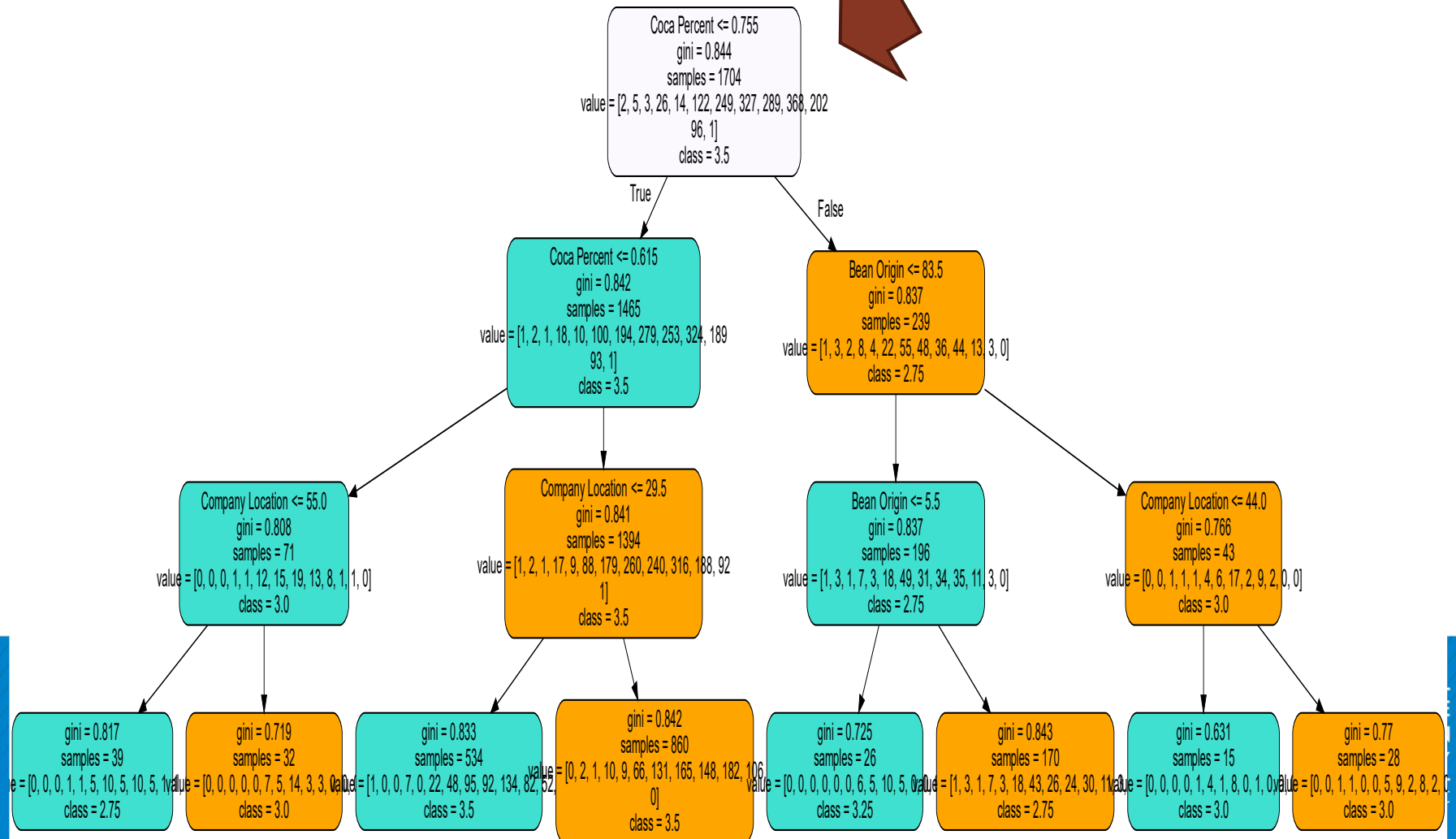
- Using chocolate bars' renting to build a classification model
 - Decision Tree is introduced in this project to set up a classification model
 - Parameter: cocoa percentage, bean origin, bean type, company location
 - Target: chocolate bars' renting
 - Model with entropy and gini
 - Mass missing issue on one variable, bean type
 - Using SVM to predict and fill in the missing data

Summary

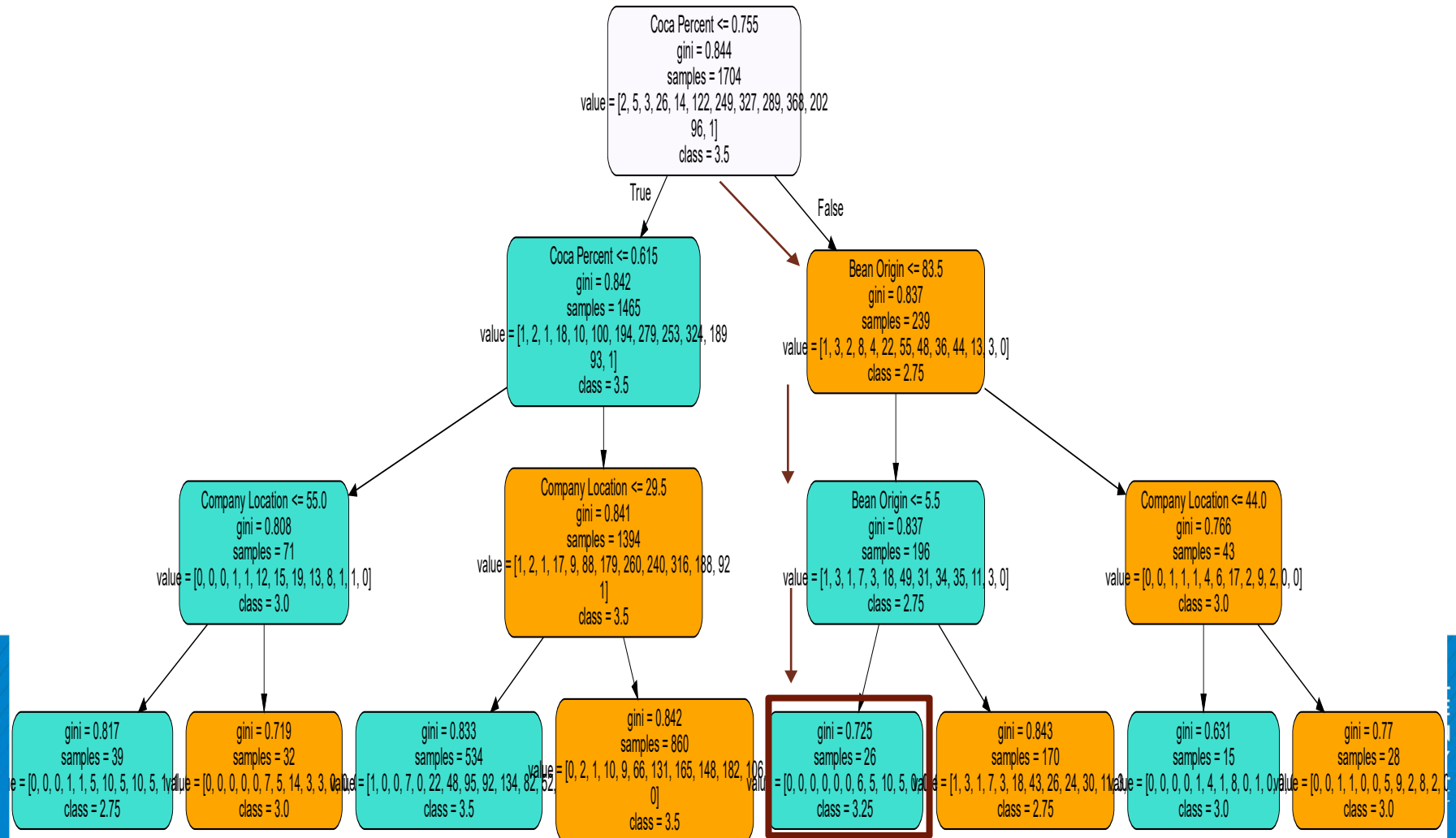
- 75% cocoa percent is key to make a good chocolate
- Secret to make a better dark chocolate
 - If cocoa percent is higher than 75%, where the bean originate plays an important role
- Classification model discussion
 - Metrics: entropy VS gini
 - Cocoa percent >75%
 - Important factor: company location vs bean origin
 - Non-linear SVM has better accuracy to do prediction
 - Linear-SVM: 43.5 % V.S. Non-linear SVM: 67.4%

75% cocoa percent is key

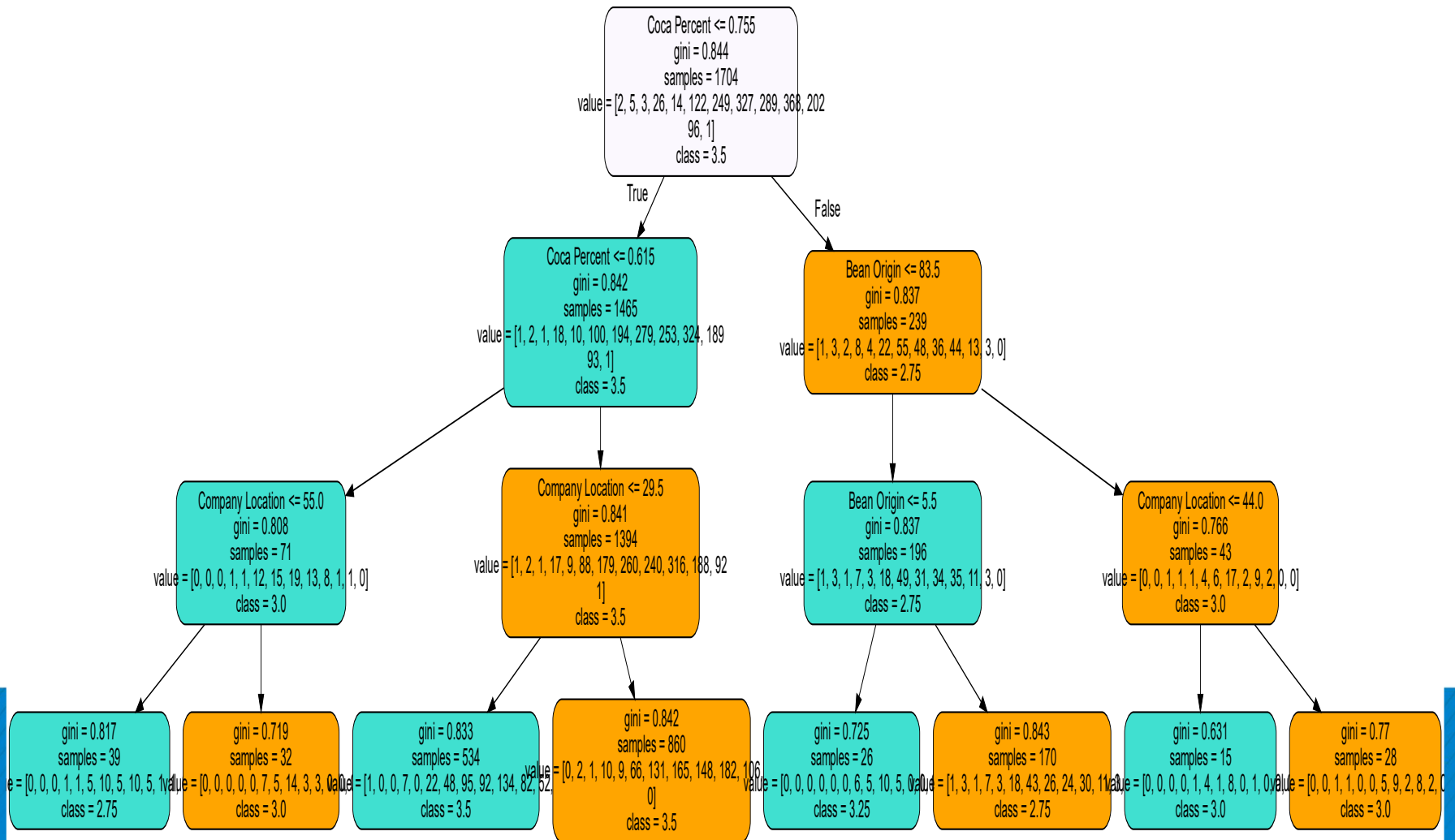
The most important attribute



Secret to make dark chocolates



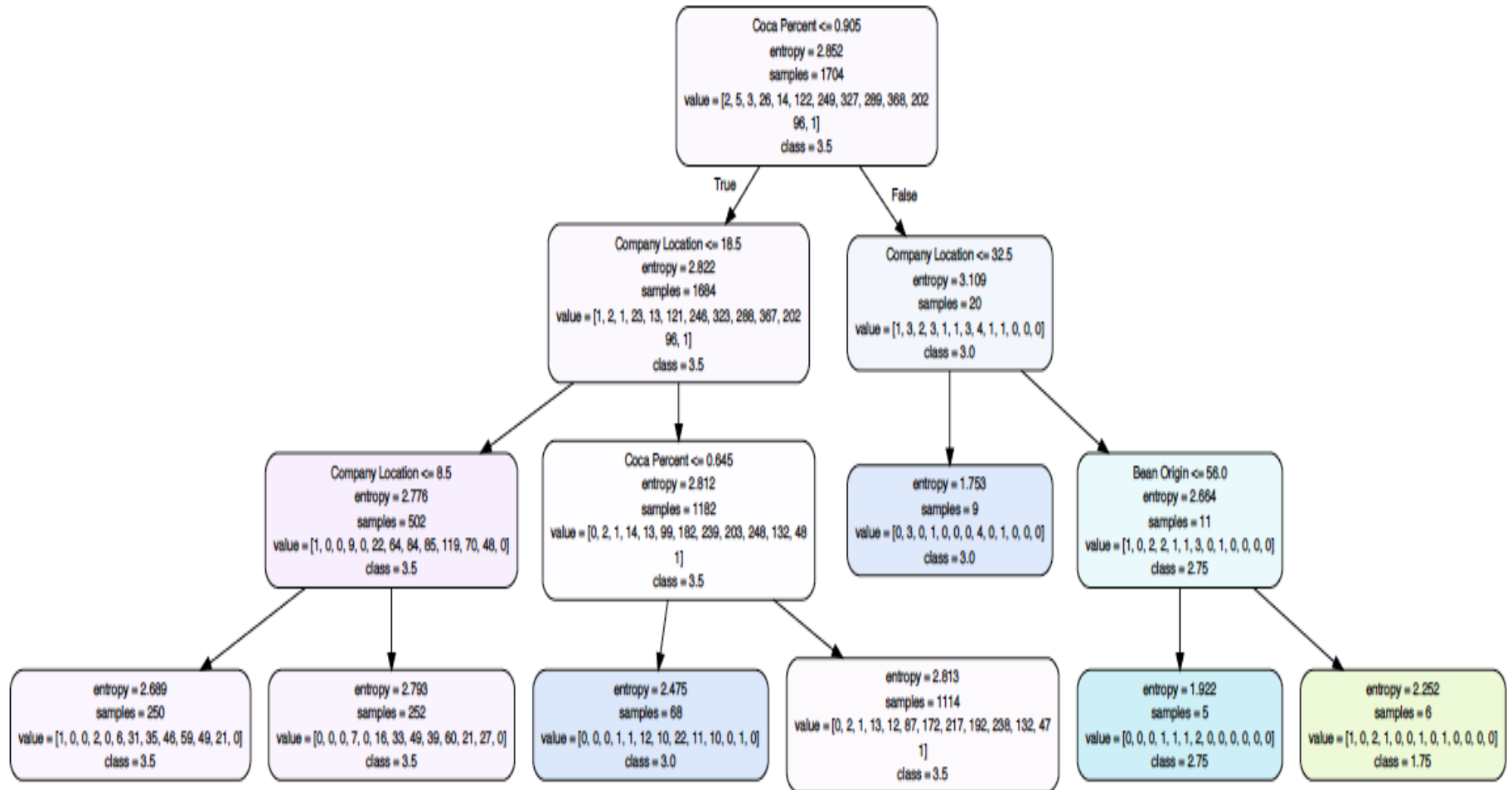
Model with gini



Model with gini

The above gini method also shows cocoa percent is the most important factor and the ideal percentage is equal or less than 75.5. If the cocoa percent is larger than 90.5%, then bean origin plays a key role in the classification. Companies whose bean origin is less than 5.5 is still able to produce good chocolate bars. Then if bean origin is larger than 5.5, company location is the important factor.

Model with entropy



Secret to make dark chocolates

The above decision tree shows cocoa percent is the most important factor and the ideal percentage is equal or less than 90.5. If the cocoa percent is larger than 90.5%, then the company location plays a key role in the classification. Companies whose location is less than 32.5 is still able to produce good chocolate bars. The third essential factor is bean origin when company location is larger than 32.5, and as for bean type, its importance contribution is zero.

Result

- Classification model for dark chocolate
 - Model with gini
 - Bean origin with higher rating: 6 countries
 - Africa, Carribean, C. Am. ,Australia, Belize, Bolivia,Brazil,Burma
 - Model with entropy
 - Company location with higher rating : 36 countries
 - Amsterdam, Argentina, Australia, Austria, Belgium, Bolivia, and others

Enjoy dark chocolate!

THE GEORGE
WASHINGTON
UNIVERSITY

WASHINGTON, DC