

## **A Predictive Analysis of Food Nutrient Density Research Questions/Answers**

David Berberena

Bellevue University

DSC 680 Applied Data Science

Amirfarrokh Iranitalab

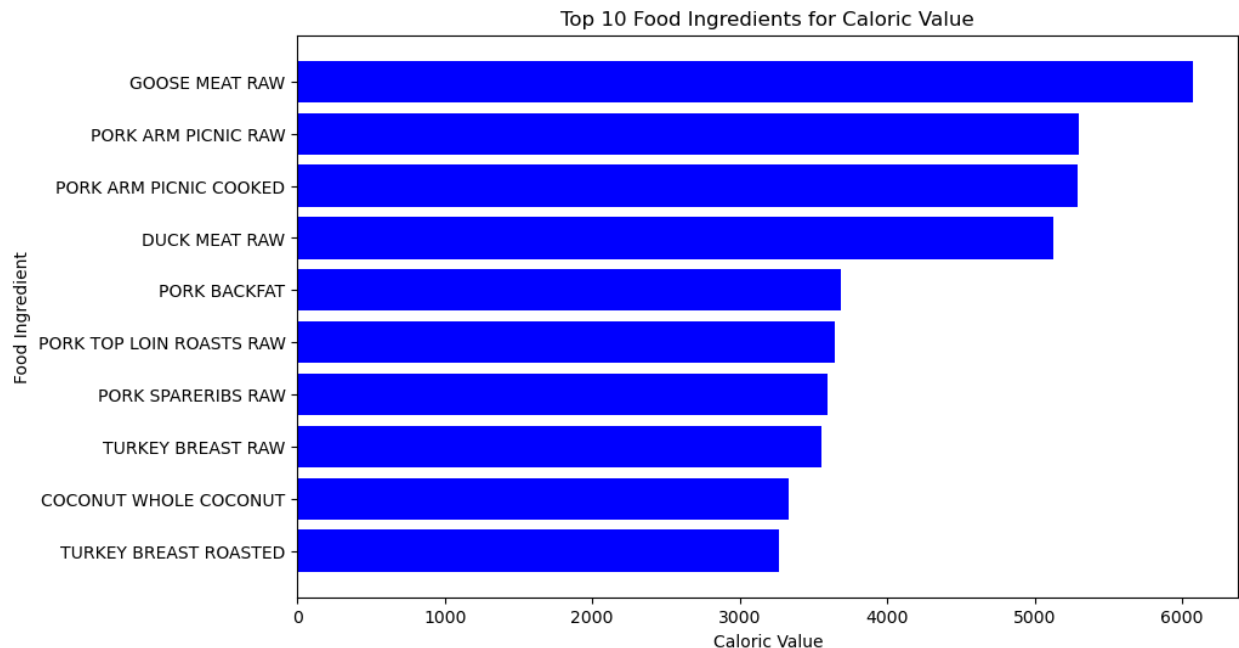
September 22, 2024

## A Predictive Analysis of Food Nutrient Density Research Questions/Answers

Regarding my study on the prediction of nutrition density, the questions that I believe end users would be interested in knowing about the study are vast, but there are a few key ideas that come to mind that may benefit the general public the most:

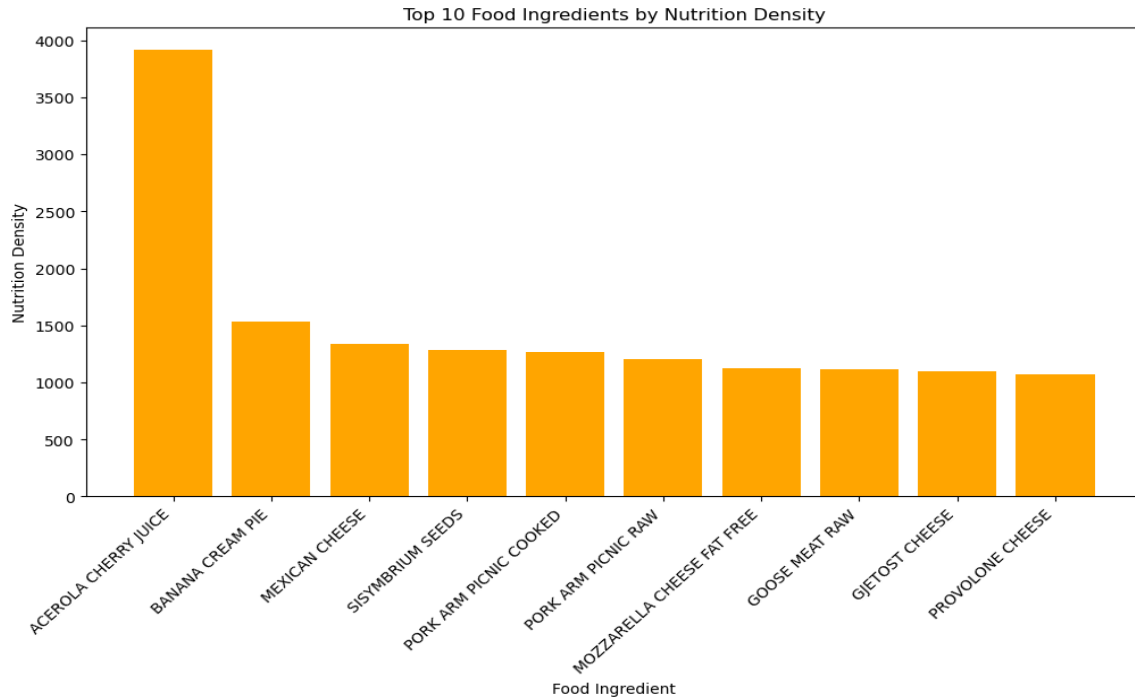
1. What food ingredient has the most calories?

**ANSWER:** As seen in the bar chart below, the ingredient that provides the most caloric intake according to the data is raw goose meat.



2. What food is the most nutrient-dense?

**ANSWER:** As seen in the bar chart below, the food ingredient with the highest nutrition density value according to the data is acerola cherry juice.



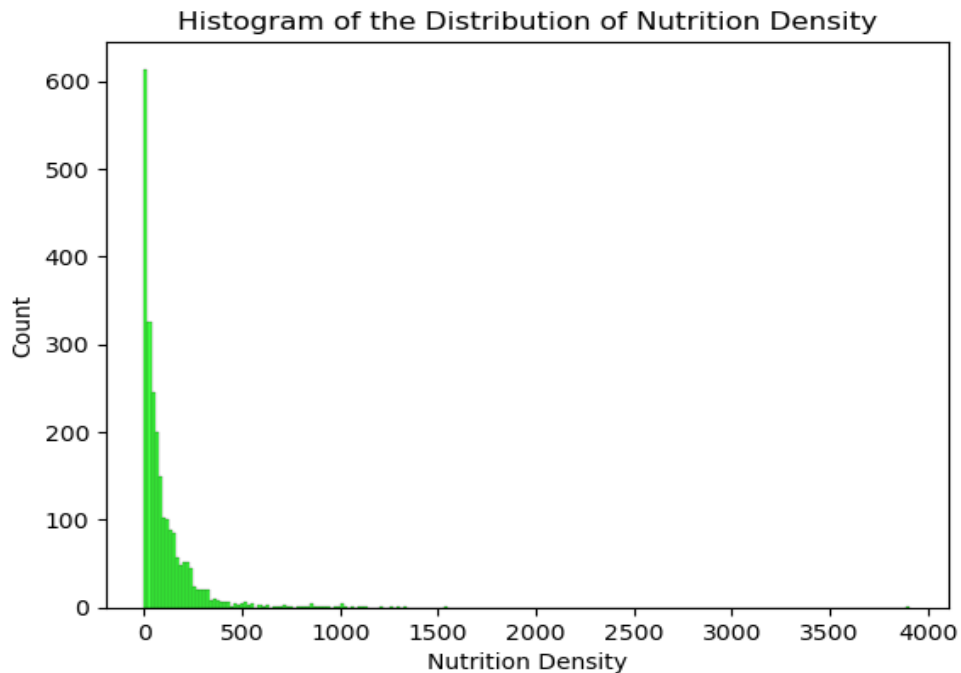
3. Which nutrients are highly correlated with nutrition density?

**ANSWER:** In the search for this answer, a correlation matrix was generated and displayed using a heatmap. The chart is very large with many variables, so I have extracted the top twenty highest positively correlated nutrients and placed them into the table below.

Variable Name	Correlation Value	Variable Name	Correlation Value
Calcium	0.796	Monounsaturated Fats	0.358
Phosphorus	0.558	Zinc	0.331
Caloric Value	0.535	Iron	0.325
Potassium	0.529	Carbohydrates	0.323
Vitamin C	0.490	Vitamin B6	0.303
Magnesium	0.471	Polyunsaturated Fats	0.285
Protein	0.455	Dietary Fiber	0.274
Fat	0.422	Vitamin B1	0.233
Vitamin B3	0.393	Vitamin B2	0.209
Water	0.359	Vitamin B5	0.209

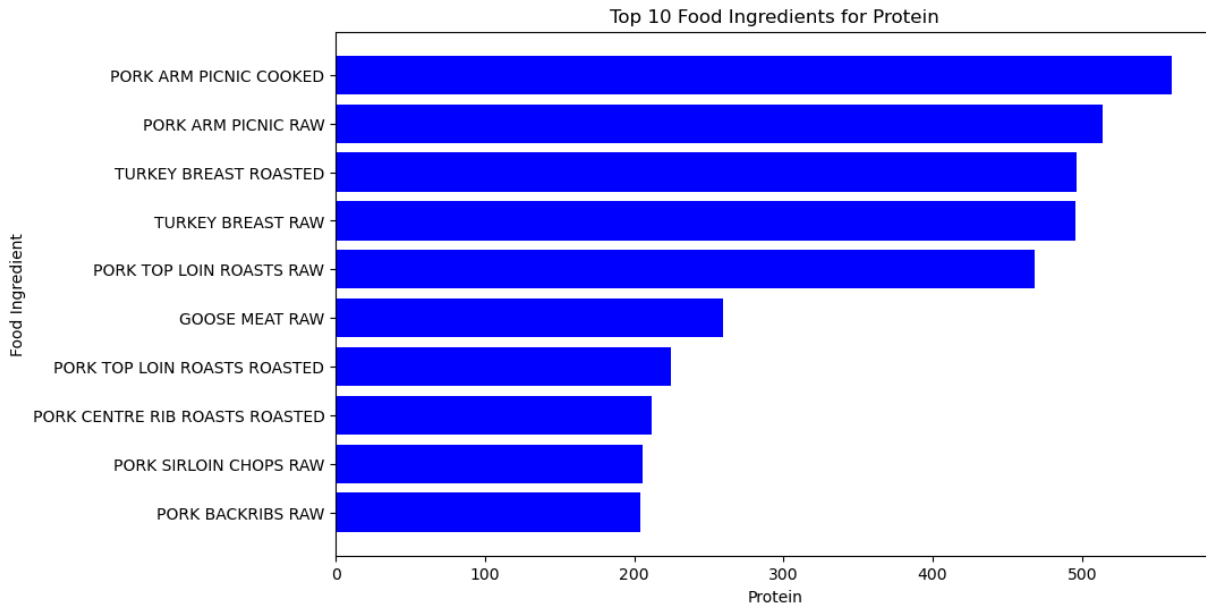
4. What nutrition density measure do most food ingredients fall between?

**ANSWER:** As seen in the histogram below, the nutrition density range in which most food ingredients fall is between zero and five hundred.



5. What are the top ten foods rich in protein?

**ANSWER:** As seen in the bar chart below, they are, according to the data, as follows (in descending order): cooked pork arm picnic, raw pork arm picnic, roasted turkey breast, raw turkey breast, raw pork top loin roast, raw goose meat, roasted pork top loin roast, roasted pork center rib roast, raw pork sirloin chops, and raw pork back ribs.



6. Are any vitamins and minerals impactful to nutrition density?

**ANSWER:** As shown in the below outline of feature importance to nutrition density for the decision tree, random forest, and gradient boosting regressor models, we can see that calcium, vitamin C, phosphorus, and potassium are all established as impactful according to the data and the models generated.

```

Best performing base estimators and their performance metrics:

Decision Tree Best Estimator:

{'ccp_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': None, 'max_leaf_nodes': None, 'min_impurity_decrease': 0.0, 'min_samples_leaf': 1, 'min_samples_split': 2, 'min_weight_fraction_leaf': 0.0, 'random_state': 123, 'splitter': 'best'}

Performance Metrics for Decision Tree:

RMSE: 41.49383966411136
R-squared: 0.9388757928138893

Feature importances:

Calcium: 0.5451
Vitamin C: 0.2703
Caloric Value: 0.1070
Phosphorus: 0.0321
Dietary Fiber: 0.0120

Random Forest Best Estimator:

{'bootstrap': True, 'ccp_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, 'max_samples': None, 'min_impurity_decrease': 0.0, 'min_samples_leaf': 1, 'min_samples_split': 2, 'min_weight_fraction_leaf': 0.0, 'n_estimators': 200, 'n_jobs': None, 'oob_score': False, 'random_state': 123, 'verbose': 0, 'warm_start': False}

Performance Metrics for Random Forest:

RMSE: 34.89358787335249
R-squared: 0.9567747767624285

Feature importances:

Calcium: 0.5793
Vitamin C: 0.2414
Caloric Value: 0.0850
Saturated Fats: 0.0178
Carbohydrates: 0.0106

Gradient Boosting Best Estimator:

{'alpha': 0.9, 'ccp_alpha': 0.0, 'criterion': 'friedman_mse', 'init': None, 'learning_rate': 0.1, 'loss': 'squared_error', 'max_depth': 3, 'max_features': None, 'max_leaf_nodes': None, 'min_impurity_decrease': 0.0, 'min_samples_leaf': 1, 'min_samples_split': 2, 'min_weight_fraction_leaf': 0.0, 'n_estimators': 200, 'n_iter_no_change': None, 'random_state': 123, 'subsample': 1.0, 'tol': 0.0001, 'validation_fraction': 0.1, 'verbose': 0, 'warm_start': False}

Performance Metrics for Gradient Boosting:

RMSE: 27.04256135158072
R-squared: 0.9740377912504109

Feature importances:

Calcium: 0.5559
Vitamin C: 0.2746
Caloric Value: 0.0980
Phosphorus: 0.0160
Potassium: 0.0068

```

7. How important is calcium to overall nutrition density?

**ANSWER:** As shown above in the outline of feature importances, calcium is shown in all three of the individual models that are a part of the ensemble stacking regressor model to be between 54.5% and 57.9% impactful to the model's generation of predictions regarding overall nutrition density.

8. How can we use this study's results in everyday life?

**ANSWER:** This study's results and implications can span to everyday life by prompting the end user to reevaluate what foods they intake based on the increased access to nutrition density information. They can decide to integrate more nutrient-dense foods into their diet for a

healthier lifestyle, whether for weight loss, muscle gain, detoxification, medical reasons, or even simply to try different foods. If an end user is conscious about the foods they wish to eat and those they do not, the results here can provide an enhanced knowledge of foods that do not match the user's nutrition density needs. Organizations and private companies can also use this information to streamline their diet regimens and meal plans available to their customers. Two companies that come to mind are Weight Watchers and Hello Fresh.

9. Do we have to eat the foods that are found to be the most nutrient-dense?

**ANSWER:** No! The increased access to the results of this study does not mean anyone will have to eat different foods or be forced to change their diet in any way. The same can be said for food manufacturing businesses in the sense that they will not be ordered to change anything about the food items they produce provided these results. These results are meant simply to influence recommendations on how we choose our dietary habits, from the individual to large organizations like the FDA and the USDA which provide countrywide recommendations for how we should partake in our dieting strategy. However, if anyone should choose not to eat these nutrient-dense foods, they still maintain the freedom to choose whatever foods they wish to eat, regardless of the effects of their diet on their overall health and well-being.

10. Who are the results of this study directly impacting?

**ANSWER:** The results of this study are intended to impact organizations like the FDA and the USDA which recommend dieting regimens to the masses, private dietitians and nutritionists who create dieting regimens for individuals based on their specific needs, and the general public who would be able to independently choose what foods they would like to eat provided the new knowledge displayed with these results. Nutrition apps and websites could also be

impacted by these results as well, as nutrition logging apps like MyFitnessPal and ingredient scanning apps like Yuka can be regularly updated with the most accurate nutrition density scores and nutrient values to provide the best ingredient recommendation and the most accurate nutrient intake for their users.