What’s In Your Food?

Inna Munroe, July 18 2018

U.S. Department of Health & Human Services

- Data from 2009-2010 indicates that over 78 million U.S. adults and about 12.5 million (16.9%) children and adolescents are obese
- Recent reports project that by 2030, half of all adults (115 million adults) in the United States will be obese.
- The annual cost of being overweight is $524 for women and $432 for men; annual costs for being obese are even higher: $4,879 for women and $2,646 for men.

https://www.hhs.gov/fitness/resource-center/facts-and-statistics/index.html

Potential Uses of Machine Learning Models

- Voice assistants (interactive Calorie prediction based on ingredients)
- Diet Apps (diet optimization)
- Food recommendation and automated food classification

What is *Open Food Facts*?

- It is a non-profit organization that maintains an open international database of products

- They also have data available in a csv format on: Kaggle.com
- Information about ingredients, origins, brands, retailers, categories and nutritional facts

- Quick demo: https://world.openfoodfacts.org/product/0000020039127/butter-croissants-fresh-easy

Two Problems

1. **Regression problem**: can we predict energy content of a product if we only know its ingredients, category, and serving size?

2. **Classification:** can we correctly identify food categories based on nutritional information and serving size?

Energy is measured **in kJ**

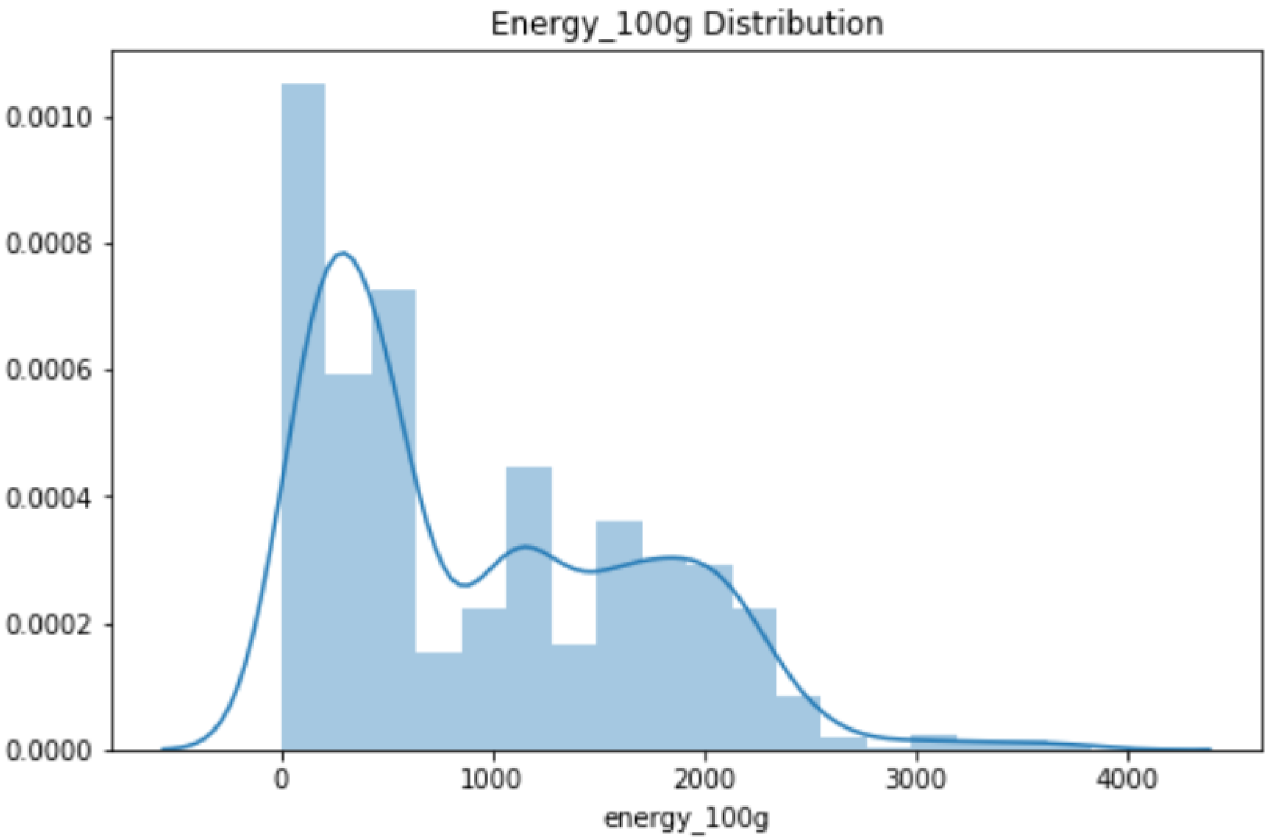
Metrics

1 Calorie = 4.184 kJ

Regression

- Two countries dominated the dataset: France and the US

- Eventually, the subset from the **US** was selected to narrow down the number of languages used for the ingredients



The target variable is not normally distributed

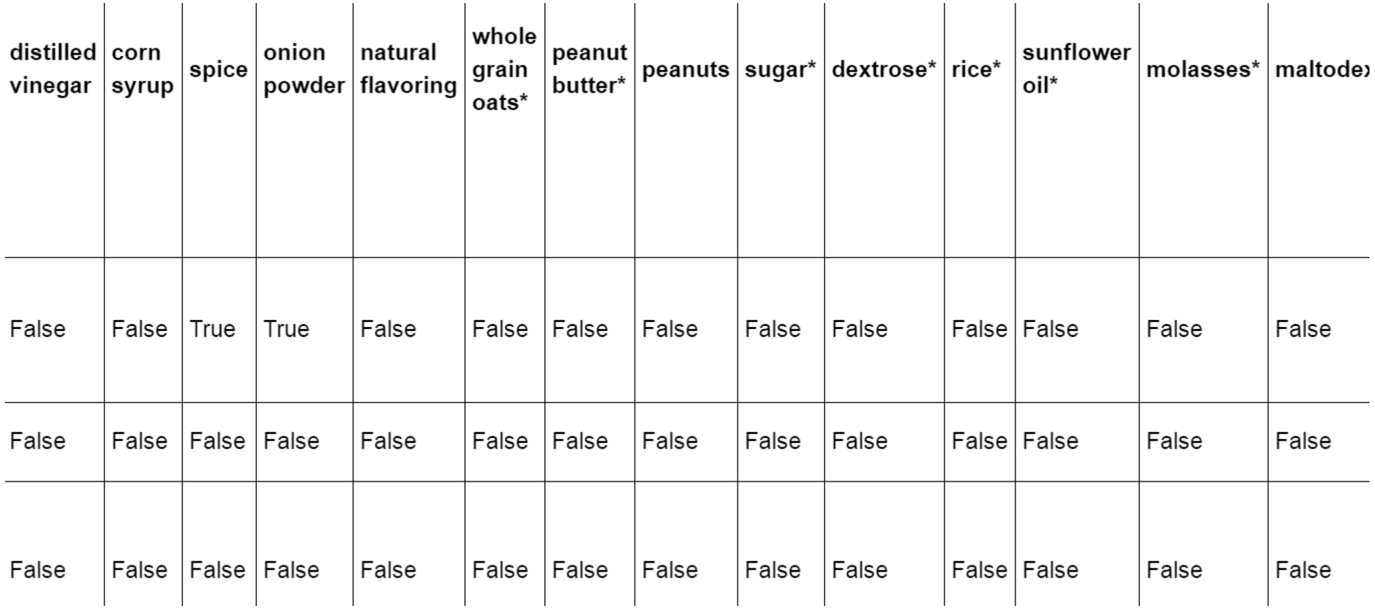
Non-linear transformations were attempted in the trial stages and not implemented in the final models

They did not lead to improvements in the best models

Energy per 100 g

Features

- 3650 binary features were for ingredient items that appear in the training set (ingredients text, one single string)



distilled vinegar	corn syrup	spice	onion powder	natural flavoring	whole grain oats*	peanut butter*	peanuts	sugar*	dextrose*	rice*	sunflower oil*	molasses*	maltodex
False	False	True	True	False	False	False	False	False	False	False	False	False	False
False	False	False	False	False	False	False	False	False	False	False	False	False	False
False	False	False	False	False	False	False	False	False	False	False	False	False	False

Additional Features

- Serving size (high variation in measurements, e.g. grams, mls, cups, oz, table spoons etc.).

- Food categories: dummies for categorical data

Metrics Used

- RMSE (root mean square error)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

- MAE (mean absolute error)

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

- R^2

Models Attempted

• **Linear Models**:	**Training Set Cross Validation**	**Test Set**
Ridge Regression:	408.3 kJ	388.21
Lasso Regression:	448.62 kJ	428.78
Elastic Net:	405.91 kJ	387.73

Models Attempted

• **Tree Models**:	Training Set Cross Validation	Test Set
Random Forest:	334.76 kJ	320.92 kJ
XGBoost Regression:	326.05 kJ	330.70 kJ

Mean Absolute Error and R^2

XGBoost:	179.85 kJ	0.8149
Random Forest:	169.22 kJ	0.8257

Go! Go! Go... Slim!!
with SoySteak Jerky
™
Made from NON-GMO Soy Protein
A Wellness Alternative to Meat
One serving helps you feel full longer!
Just one serving before a meal prevents over-eating
V Vegan
All Natural
NON GMO
97% Fat Free
Peppered w/ Black Pepper
Why SoySteak?
• 11 Grams Protein per serving.
• 180mg Sodium per serving. (except Grilled flavor, 335mg, Light with Tomato, 85mg).
• 9 Grams Total Carbs per serving.
• No Cholesterol.
• 97% Fat Free.
• 86 Calories per serving.
• Isoflavones, the Naturally Occurring Valuable Nutrients of Soy.
• All Natural. 100% Vegan. No Preservative. Non GMO. No Coloring.
Revolutionary!
Ready to Enjoy!
Tender! Very Delicious!
Taste Better than Real Meat!
Even Tastier when Heated or in Hot Soup.
May Reduce Heart Disease!
FDA has concluded that 25 grams of soy protein per day with diets low in saturated fat and cholesterol have a cholesterol-lowering effect and may reduce the risk of heart disease. One serving of SoySteak supplies 11 grams of soy protein. Note: This statement does not apply to Grilled Flavor due to sodium content

Best Result

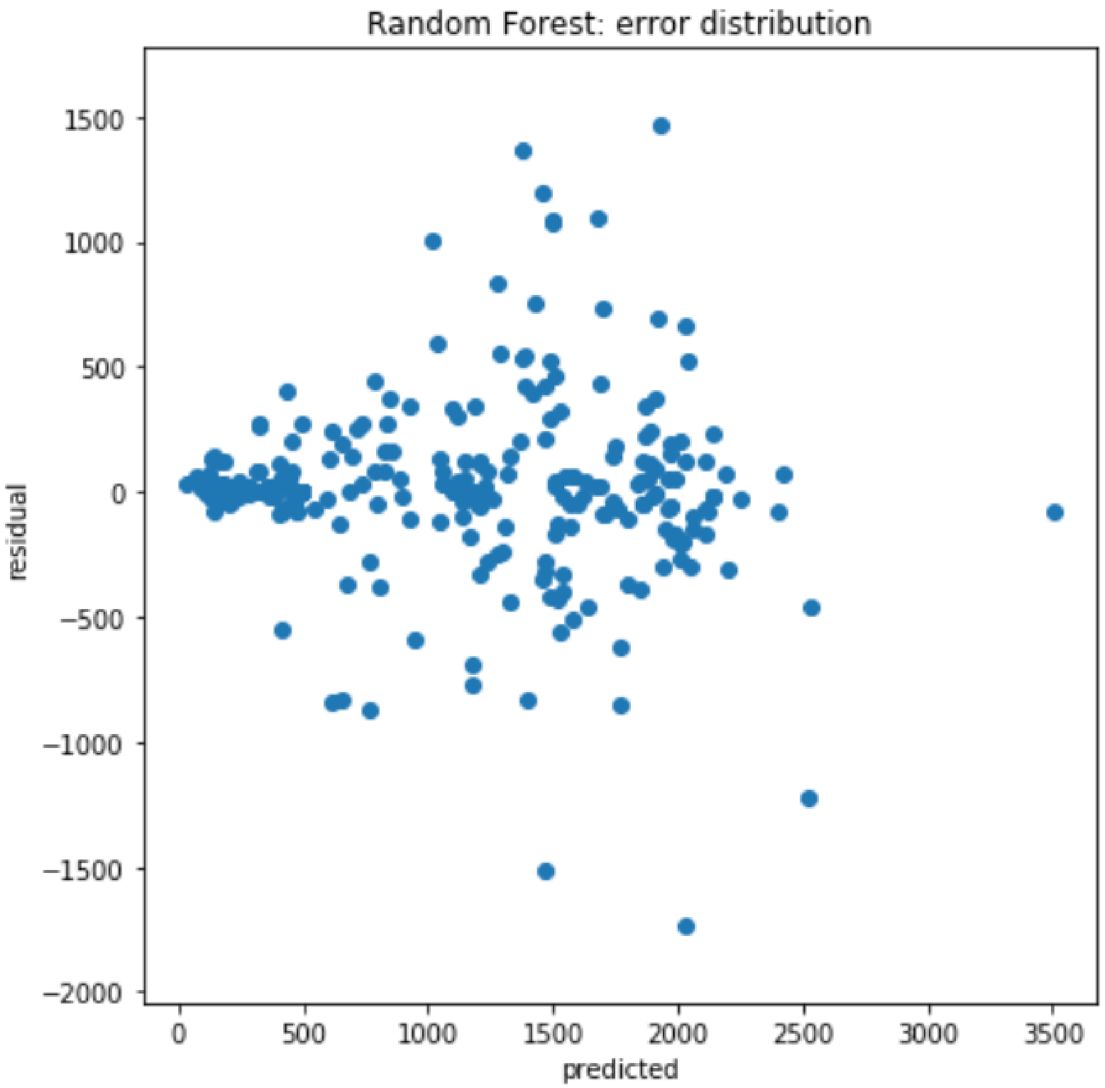
- RMSE: 320.92 kJ (76 Calories) per 100 g
- MAE: 169.22 kJ (or 40.4 Calories) per 100 g

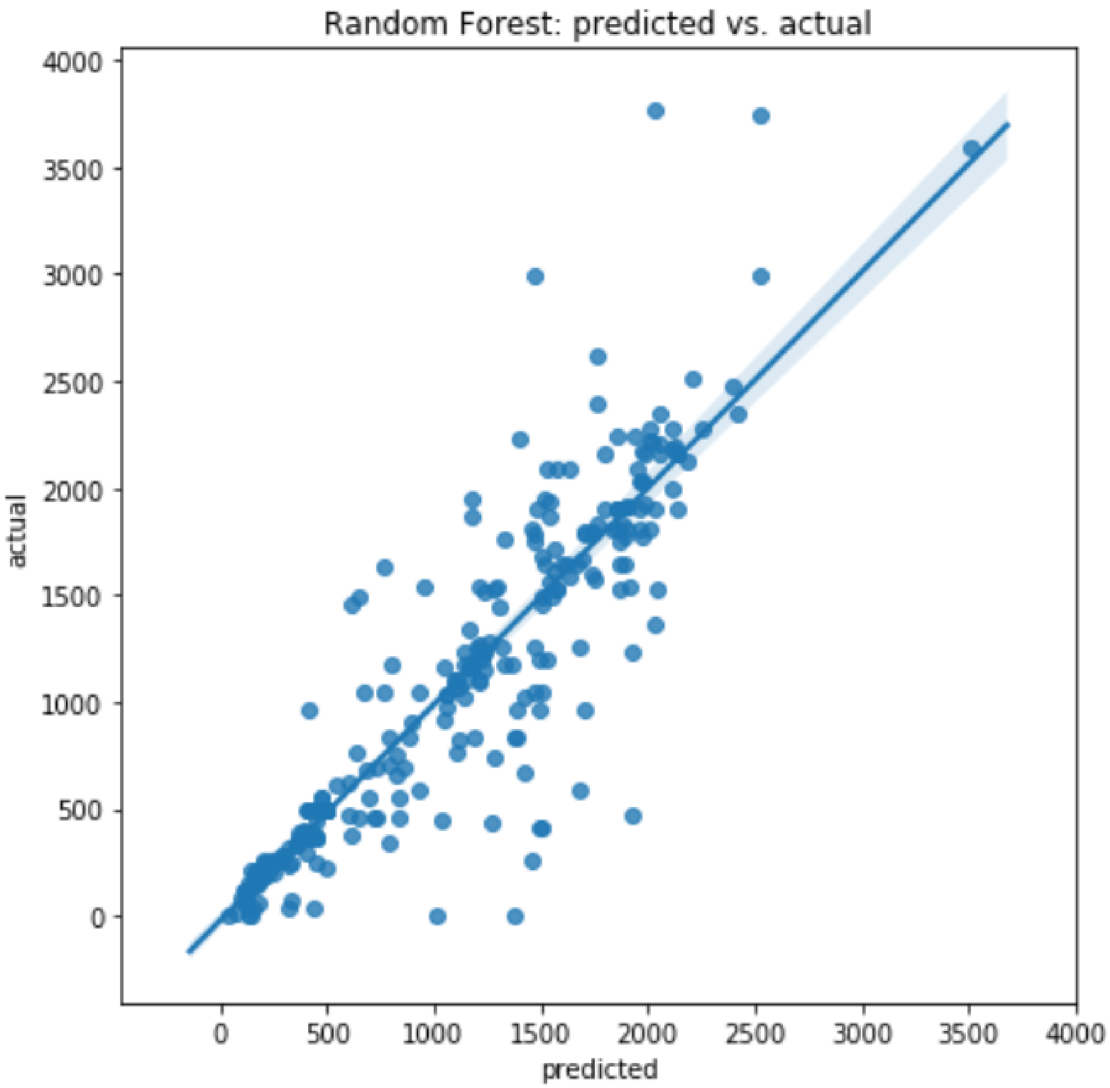
* Nutella – 2255 kJ per (539 Calories) per 100 g

* Mixed-berry granola bar – 1506 kJ (364 Calories) per 100 g

* Iced green tea - 71 kJ (17 Calories) per 100 g

Errors Analysis





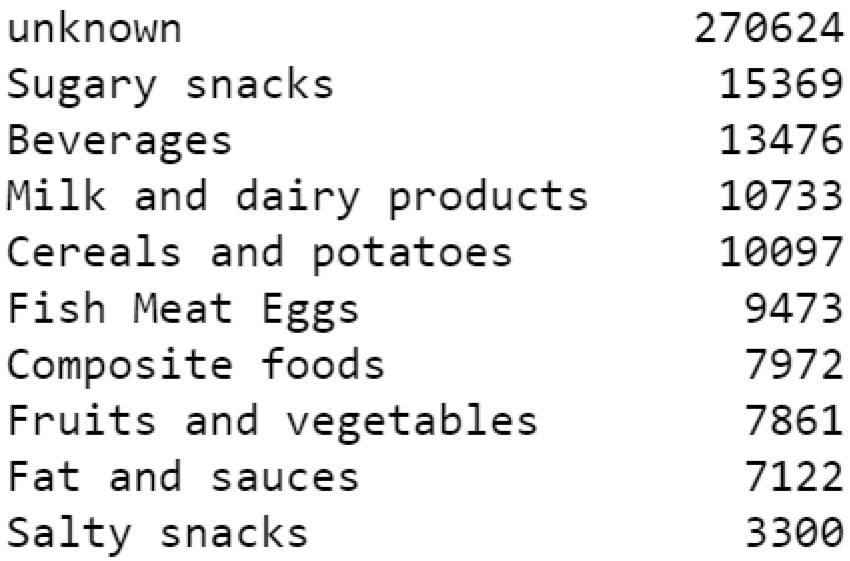
Improvements

- Better Text Parsing (time limit to handle all the punctuation complexities)
- More model precise parameter finetuning
- Information about the weight of each ingredient might be needed for more predictive power

Classifier

- Currently, the dataset has 270624 uncategorized food items

- We could use machine learning to classify items based on their nutritional content and serving size

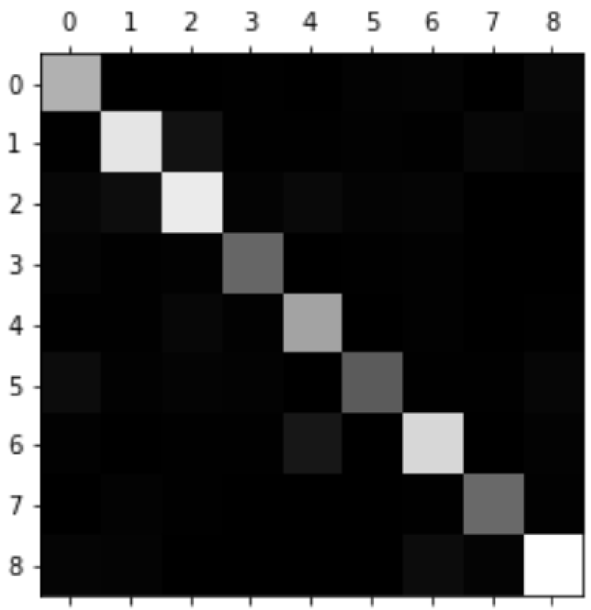


```
unknown                       270624
Sugary snacks                  15369
Beverages                      13476
Milk and dairy products        10733
Cereals and potatoes           10097
Fish Meat Eggs                  9473
Composite foods                 7972
Fruits and vegetables           7861
Fat and sauces                  7122
Salty snacks                    3300
```

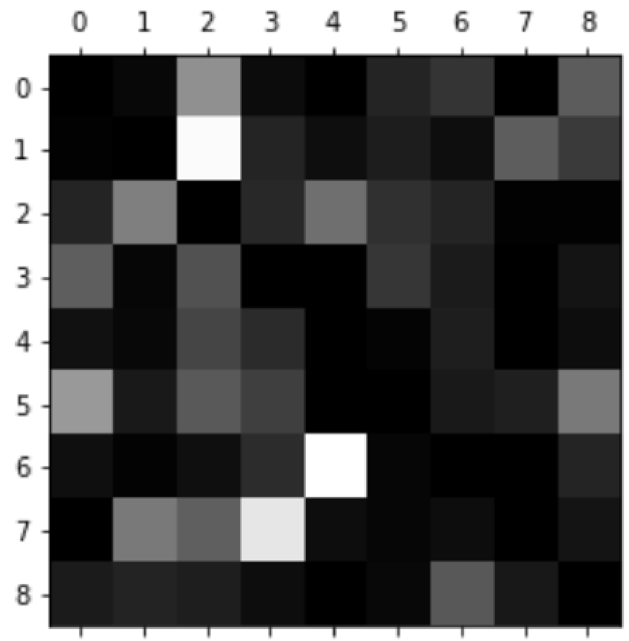
Tree Models Accuracy

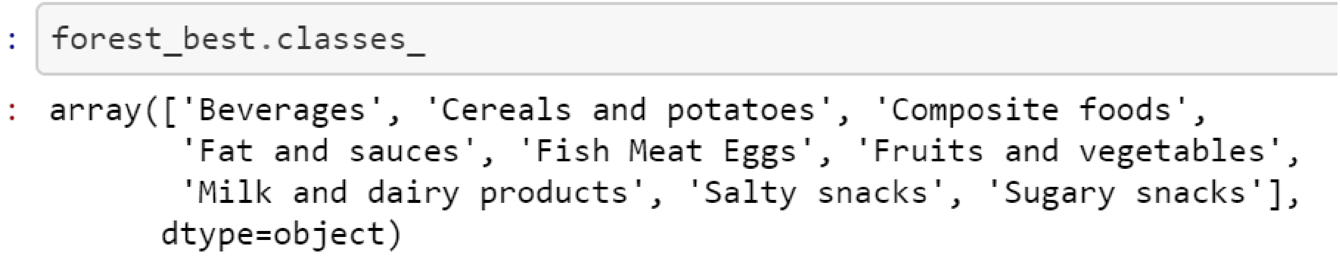
Tree Models:	**Training Set Cross Validation**	**Test Set**
Random Forest:	0.9027	0.87
XGBoost Classifier:	0.903	0.84

Confusion Matrix



Confusion Matrix with the diagonal zeroed out and normalized rows





```
: forest_best.classes_

: array(['Beverages', 'Cereals and potatoes', 'Composite foods',
         'Fat and sauces', 'Fish Meat Eggs', 'Fruits and vegetables',
         'Milk and dairy products', 'Salty snacks', 'Sugary snacks'],
        dtype=object)
```

Improvements

- Initial categorization might not be precise:

e.g. fish meat eggs and milk and dairy products were frequently confused by the model

- Other categories often get confused with composite foods

Photography Credits

All photographs used in the PowerPoint were taken by **Marc Bell**