

NutriSage: Smart Nutrition Assistant for Food Analysis and Personalized Meal Guidance

by

AADITYA KUMAR DHAKA (2448001)

NEELANJAN DUTTA (2448040)

Under the guidance of


Dr. Sathya P



OCTOBER – 2025



CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE | DELHI NCR | PUNE

Department of Statistics and Data Science	
PROJECT DIARY - Sep 2025	
Class: MSc Data Science	Trimester: V
Course : MDS581 - PROJECT - II	
Student Reg. No: 2448001	Name: Aaditya kumar Dhaka
Guide Name: Dr Sathya P	
Date: From : 29/09/2025	To: 04/10/2025
<p>Task Completed:</p> <p>In the 3rd week, I started working on data preprocessing which included cleaning the dataset, handling missing values, and normalizing key attributes like calories and nutrients. Along with this, I performed EDA to study nutrients distributions, category balance and detect any outliers in the data.</p> <p>Task Planned:</p> <p>For the upcoming week I plan to complete preprocessing by finalizing feature engineering steps such as creating new derived attributes. I will also start preparing the dataset splits to get ready for model training.</p> <p style="text-align: right;"> Guide Signature</p>	

Meeting with our alumni (Sneha S)

Meet - kkh-azmy-thz

00:06:34

AADITYA KUMAR DHAKA 2448001 (Presenting)

localhost:8888/notebooks/Desktop/Net/Sage/JupyterLab

JupyterLab

Columns and dtypes:

```
image_url      object
camera_or_phone_prob  float64
dish_name       object
food_type       object
ingredients      object
portion_size    object
nutritional_profile  object
cooking_method  object
sub_id         int64
dtype: object
```

First 5 rows:

	image_url	camera_or_phone_prob	food_prob	dish_name	food_type	Ingredients
0	https://fileblob.azure/78433225950104680_e03546.jpg	0.7	0.95	Fried Chicken	Restaurant food	['chicken', 'breeding', 'oil']
1	https://fileblob.azure/783332747700103732_674873.jpg	0.7	1.00	Pho	Restaurant food	['noodles', 'beef', 'beef', 'lime', 'green onions', 'lime']
2	https://fileblob.azure/7832600501600103585_654234.jpg	0.8	0.95	Tan food Dumplings	Restaurant food	['dumplings', 'chili oil', 'any sauce']
3	https://fileblob.azure/783906800700101108_805153.jpg	0.7	1.00	Bananas	Raw vegetables and fruits	['Bananas']
4	https://fileblob.azure/7837964773700103218_173175.jpg	0.8	0.90	Noodle Stir-fry	Restaurant food	['noodles', 'chicken', 'vegetables', 'sauce']

19:19 | kkh-azmy-thz

AADITYA KUMAR DHAKA 2448001

Sneha S

Neelanjana Dutta

Meet - kkh-azmy-thz

00:11:54

AADITYA KUMAR DHAKA 2448001 (Presenting)

localhost:8888/notebooks/Desktop/Net/Sage/JupyterLab

JupyterLab

Plotting nutritional profiles:

```
Printed nutrition columns. Normalized counts:
calories_kcal: 20000
protein_g: 20000
fat_g: 20000
carbohydrate_g: 20000
```

calories_kcal

fat_g

protein_g

carbohydrate_g

19:24 | kkh-azmy-thz

AADITYA KUMAR DHAKA 2448001

Sneha S

Neelanjana Dutta

Meet - kkh-azmy-thz

https://meet.google.com/kkh-azmy-thz

Neelanjan Dutta (You, presenting)

00:20:47

Stop presenting

jupyter Data Preprocessing and EDA Last checkpoint 2 days ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

Descriptive Statistics

	count	mean	std	min	25%	50%	75%	max
camera_or_phone_prob	100000	0.740385	0.420751	0.0	0.7	0.7	0.8	0.8
food_prob	100000	0.837804	0.342095	0.0	0.8	0.8	1.0	1.0
fat_g	100000	17.580940	15.874091	0.0	16.0	15.0	25.0	250.0
protein_g	100000	27.390590	15.191140	0.0	3.0	20.0	30.0	250.0
calories_kcal	100000	411.345600	254.453407	0.0	250.0	350.0	400.0	1000.0
carbohydrate_g	100000	40.543090	20.891460	0.0	20.0	35.0	40.0	60.0
total_portion_g	100000	360.276040	197.219170	0.0	250.0	350.0	450.0	1000.0
num_ingredients	100000	3.537770	1.426804	0.0	3.0	4.0	5.0	11.0

Correlation Matrix

	camera_or_phone_prob	food_prob	fat_g	protein_g	calories_kcal	carbohydrate_g	total_portion_g	num_ingredients
camera_or_phone_prob	1.000000	0.191196	-0.016883	0.070767	0.011791	0.051105	0.040848	-0.071044
food_prob	0.191196	1.000000	0.204163	0.255419	0.279746	0.221184	0.260366	0.215881
fat_g	-0.016883	0.204163	1.000000	0.841000	0.891480	0.286725	0.564033	0.357319
protein_g	0.070767	0.255419	0.841000	1.000000	0.895663	0.322513	0.624274	0.300763
calories_kcal	0.011791	0.279746	0.891480	0.895663	1.000000	0.627791	0.601096	0.477010
carbohydrate_g	0.051105	0.221184	0.286725	0.322513	0.627791	1.000000	0.445883	0.397979
total_portion_g	0.040848	0.260366	0.564033	0.624274	0.601096	0.445883	1.000000	0.397779
num_ingredients	-0.071044	0.215881	0.357319	0.300763	0.477010	0.397979	0.397779	1.000000

Visualizing Correlations and Distributions

meet.google.com is sharing a window.

Stop sharing Hide

19:33 | kkh-azmy-thz

AADITYA KUMAR DHAKA 2448001

Sneha S

Neelanjan Dutta

Meet - kkh-azmy-thz

https://meet.google.com/kkh-azmy-thz

Neelanjan Dutta (You, presenting)

00:27:23

Stop presenting

jupyter Data Preprocessing and EDA Last checkpoint 2 days ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

```
plt.figure(figsize=(10,8))
plt.scatter(calories_kcal, protein_g)
plt.title('Scatter: calories_kcal vs protein_g')
plt.show()
```

Scatter: calories_kcal vs protein_g

meet.google.com is sharing a window.

Stop sharing Hide

19:40 | kkh-azmy-thz

AADITYA KUMAR DHAKA 2448001

Sneha S

Neelanjan Dutta