

Project Report

This project analyzes visual food data to identify key trends in dietary choices and nutritional intake. By uncovering patterns in food selection, we aim to provide insights for personalized wellness programs and dietary analysis. This analysis can inform the development of targeted interventions to promote healthier eating habits. Our findings will empower individuals to make more informed food choices and improve overall well-being through better dietary guidance.

# Goals

**Understand Dietary Patterns and Habits:** Identify and analyze dietary patterns, habits, and trends from the food image data to gain insights into eating behaviors

**Explore Factors Influencing Food Choices:** Investigate factors influencing food choices, such as cuisine type, meal location, mood, and dietary restrictions

**Explore Dietary Diversity**

**Investigate the Impact of Food Choices on Mood**

# Metadata

1. **Image Metadata**

* **Image Filename**
* **Date**

**II. Features**

* **Type of Cuisine**
  + *Description:* The likely cuisine or culinary style of the dish.
  + *Possible Values:* Thai, Italian, Mexican, American, Fusion, Combination, Unknown, etc.
* **Happiness Level (1-5)**
  + *Description:* A subjective rating of how appealing or satisfying the meal appears.
  + *Possible Values:* Integer scale from 1 (lowest) to 5 (highest).
* **Course of the Meal**
  + *Description:* The part of the meal represented by the dish.
  + *Possible Values:* Starter, Main Course, Side Dish, Dessert, Snack, Drink, Unknown.
* **Flag for Sugar (High/Medium/Low)**
  + *Description:* A relative indication of the sugar content of the meal.
  + *Possible Values:* High, Medium, Low, Unknown.
* **Flag for Salt (High/Medium/Low)**
  + *Description:* A relative indication of the salt content of the meal.
  + *Possible Values:* High, Medium, Low, Unknown.
* **Flag for Healthy (Yes/No) or e.g., healthy, unhealthy, balanced, carb-heavy, protein-rich?**
  + *Description:* A subjective assessment of the overall healthfulness of the meal.
  + *Possible Values:* Yes, No, Unknown.
* **Level of Processing**
  + *Description:* Subjective assessment of how processed the food is.
  + *Possible Values:* Unprocessed, Minimally Processed, Processed, Highly Processed.
* **Preparation Method:**
  + *Description:* How the food was prepared.
  + *Possible Values:* Fried, Baked, Grilled, Steamed, Raw.
* **Dominant Color:**
  + *Description:* The most prominent color in the dish.
  + *Possible Values:* Red, Green, Yellow, Brown, White, etc
* **Food Diversity:**
  + *Description:* A measure of how many different types of food are present in the meal.
  + *Possible Values:* High, Medium, Low.

# Data Processing:

**Workflow:**

Image Folder → Extract Basic Image Info→ Analyze Image with Gemini AI →

Extract & Clean AI Response (Features) → Store Features in JSON Lines File →

Amend json data based on pictures → Load into Pandas DataFrame for Analysis

**Key Processes:**

**Image Data Ingestion & Preprocessing:**

* Collect images in a folder of food images (JPG, PNG, JPEG) for processing

**AI-Powered Image Analysis (Google Gemini):**

* Leverages the Google Gemini Pro Vision model (gemini-1.5-flash) for in-depth image analysis.
* A detailed prompt instructs the AI to act as a "food analyst" and extract specific features from each image.
* Requires a Gemini API key for operation, with a function included to configure this key.

**Feature Extraction:**

* The AI model is prompted to identify and categorize various aspects of the food in the image, including:
  + **Cuisine:** (e.g., Thai, Indian, Italian, American)
  + **Happiness Level:** (Subjective appeal, 1-5 scale)
  + **Meal Course:** (e.g., Starter, Main Course, Dessert)
  + **Sugar & Salt Content:** (High, Medium, Low)
  + **Healthiness:** (Subjective assessment, e.g., healthy, unhealthy, balanced)
  + **Processing Level:** (e.g., Unprocessed, Minimally Processed, Highly Processed)
  + **Preparation Method:** (e.g., Fried, Baked, Grilled)
  + **Dominant Color:**
  + **Food Diversity:** (High, Medium, Low)
* Results are stored in a structured dictionary format, including the pre-extracted image name, size, and date.

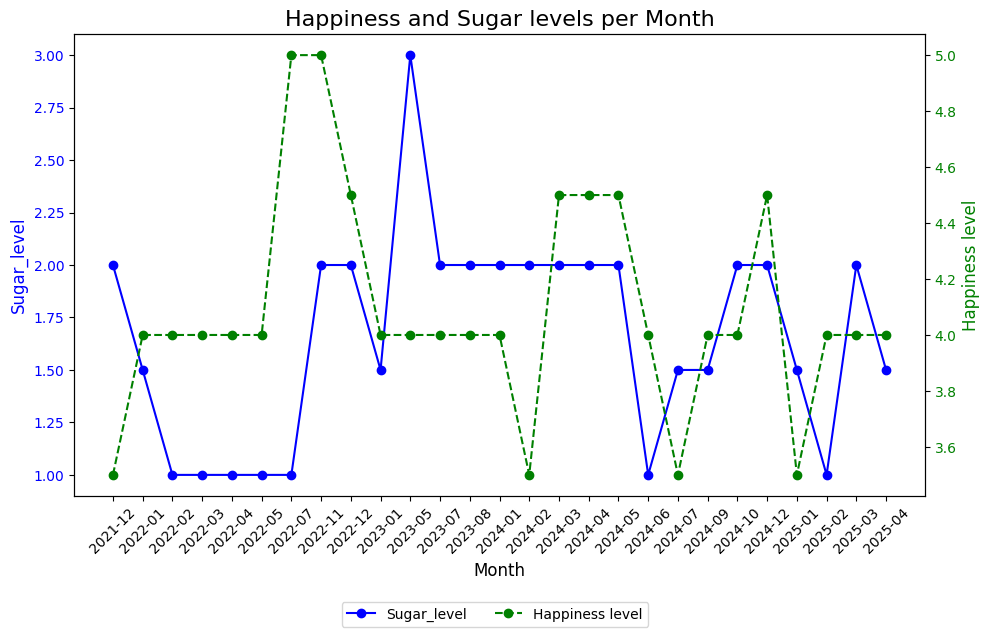
**Output & Data Storage:**

* The system iterates through all images in a specified folder.
* For each image, the analysis result (the extracted feature dictionary) is collected.
* The collected results (a list of dictionaries) are written to a JSON Lines (.json) file, with each dictionary representing an image on a new line. This allows for efficient storage and subsequent loading.
* A function is provided to read this JSON Lines file into a Pandas DataFrame for further data manipulation and analysis (e.g., converting date strings to datetime objects).

# Data Exploration:

* Identifies the most frequently occurring cuisines, meal courses, preparation methods, and dominant colors in the food images.
* Shows the distribution of perceived happiness levels, sugar/salt content, healthiness, processing levels, and food diversity.
* Explores potential correlations between numerical features like happiness, sugar/salt, healthiness, processing level, and food diversity.
* Visualizes trends in the number of images and average characteristics (like cuisine diversity or sugar/happiness levels) over months.

e.g.



# Hypothesis Testing:

A p-value less than 0.05 was considered to indicate a statistically significant difference in features.

**Hypothesis: Food Diversity vs. Food Characteristics**

* **Null Hypothesis (H0):** There is no difference in the mean levels of 'Happiness\_Level', 'Sugar', 'Salt', 'Healthy', and 'Processing\_level' between foods with high diversity (Food\_Diversity=3) and foods with low diversity (Food\_Diversity=1).
* **Results:**
  + **Happiness\_Level (p-value=0.18):** No significant difference in happiness levels was found based on food diversity.
  + **Sugar, Salt, Healthy, Processing\_level (p-values < 0.0001):** Significant differences were observed for these features, suggesting that food diversity is related to these characteristics.
* **Conclusion:** While happiness levels appear similar regardless of food diversity, characteristics like sugar, salt, healthiness, and processing level differ significantly between high and low diversity foods. This is because low diversity food e.g. coffee is biased on a few features e.g. sugar but not others.

## Chi-squared Test

Performed to assess whether there is a statistically significant association between two categorical variables. A subset of the data (happy\_food) focusing on major cuisines was used for these tests.

**Hypothesis: Happiness\_Level vs. Other Categorical Features**

* **Null Hypothesis (H0):** Happiness\_Level is independent of Sugar level (Salt level, Healthy level, Processing\_level).
* **Results:**
  + Happiness\_Level vs. Sugar (p-value ≈ 1.43e-18)
  + Happiness\_Level vs. Salt (p-value ≈ 6.61e-12)
  + Happiness\_Level vs. Healthy (p-value ≈ 2.58e-02)
  + Happiness\_Level vs. Processing\_level (p-value ≈ 3.14e-25)
* **Conclusion:** Extremely low p-values indicate very strong evidence to reject the null hypothesis. There are statistically significant associations between 'Happiness\_Level' and 'Sugar', 'Salt', 'Healthy', and 'Processing\_level'.

**Hypothesis: Meal\_Course vs. Other Categorical Features**

* **Null Hypothesis (H0):** Meal\_Course is independent of Sugar level (Salt level, Healthy level, Happiness\_Level, Processing\_level).
* **Results:**
  + Meal\_Course vs. Sugar (p-value ≈ 4.60e-42)
  + Meal\_Course vs. Salt (p-value ≈ 2.66e-31)
  + Meal\_Course vs. Healthy (p-value ≈ 1.58e-11)
  + Meal\_Course vs. Happiness\_Level (p-value ≈ 1.57e-18)
  + Meal\_Course vs. Processing\_level (p-value ≈ 1.65e-11)
* **Conclusion:** Extremely low p-values provide strong evidence that 'Meal\_Course' is significantly associated with 'Sugar', 'Salt', 'Healthy', 'Happiness\_Level', and 'Processing\_level'.

## ANOVA Test

Hypothesis: Meal Course vs. Happiness\_Level

* **Null Hypothesis (H0):** There is no significant difference in the 'Happiness\_Level' across Main Course, Dessert, and Drink.
* **Results** (p-value ≈ 1.73e-10; F-statistic ≈ 24.49)**:** Rejected the null hypothesis.
* **Conclusion:** There is a statistically significant difference in the 'Happiness\_Level' across different meal courses (Main Course, Dessert, Drink).

Hypothesis: Cuisine Type vs. Sugar

* **Null Hypothesis (H0):** There is no significant difference in the 'Sugar' level across Indian, Italian, and Fusion cuisines.
* **Results** (p-value ≈ 0.028; F-statistic ≈ 3.62)**:** Rejected the null hypothesis.
* **Conclusion:** There is a statistically significant difference in the ‘Sugar’ across different Cuisines.

# Summary:

Chi-squared tests confirmed strong associations between happiness levels/meal courses and other food attributes like sugar, salt, healthiness, and processing level.

ANOVA results indicate that while happiness levels **may not differ significantly** **across selected cuisines** (Indian, Italian, Fusion), they *do* differ significantly when compared across different meal courses (Main Course, Dessert, Drink) and Sugar levels.

# Public References:

https://www.eurofins.de/food-analysis/analytical-testing/nutritional-values/