Daily Food Nutrition Analysis

Project Title: Exploring Nutritional Patterns & Health Insights with Data Analysis

Overview:

This project explores daily food consumption patterns using a dataset of 10,000 food entries, analyzing their nutritional content. I conducted data cleaning, exploratory data analysis (EDA), and visualization to uncover trends in macronutrient intake, high-calorie foods, and unhealthy dietary habits.

Through Python (Pandas, Matplotlib, Seaborn), I extracted valuable insights into caloric discrepancies, macronutrient distribution, and high-risk food items, enabling better dietary awareness.

Key Objectives:

- **Macronutrient Breakdown**: How protein, carbohydrates, and fats vary across different food categories and meal types.
- **Food Category Insights**: Identifying the food groups contributing the most to daily calorie intake.
- **High-Risk Foods**: Finding foods with excessive sodium, cholesterol, and sugar levels.
- **Trend Analysis Over Time**: Investigating how calorie and nutrient intake fluctuate monthly.
- **Comparison of Provided vs. Calculated Calories**: Assessing data accuracy in reported nutritional values.

Data Cleaning & Preprocessing

- Removed **duplicates and missing values**.
- Converted **dates into datetime format** for trend analysis

	Date	User_ID	Food_Item	Category	Calories (kcal)	Protein (g)	Carbohydrates (g)	Fat (g)	Fiber (g)	Sugars (g)	Sodium (mg)	Cholesterol (mg)	Meal_Type	Water_Intake (ml)
0	2024-09-11	496	Eggs	Meat	173	42.4	83.7	1.5	1.5	12.7	752	125	Lunch	478
1	2024-12-17	201	Apple	Fruits	66	39.2	13.8	3.2	2.6	12.2	680	97	Lunch	466
2	2024-06-09	776	Chicken Breast	Meat	226	27.1	79.1	25.8	3.2	44.7	295	157	Breakfast	635
3	2024-08-27	112	Banana	Fruits	116	43.4	47.1	16.1	6.5	44.1	307	13	Snack	379
4	2024-07-28	622	Banana	Fruits	500	33.9	75.8	47.0	7.8	19.4	358	148	Lunch	471
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9995	2024-09-18	455	Salmon	Meat	346	29.8	55.6	4.6	1.7	0.5	976	87	Breakfast	793
9996	2024-12-13	913	Grapes	Fruits	174	22.9	54.9	32.1	2.5	5.9	255	56	Lunch	665
9997	2024-01-31	943	Strawberry	Fruits	63	36.5	23.8	21.6	0.8	48.9	757	63	Snack	876
9998	2024-09-28	571	Spinach	Vegetables	564	26.2	58.9	11.9	3.3	43.0	482	33	Breakfast	319
9999	2024-09-07	33	Banana	Fruits	442	20.9	27.3	29.6	9.9	30.9	919	22	Dinner	914

```
df.describe()
```

	Date	User_ID	Calories (kcal)	Protein (g)	Carbohydrates (g)	Fat (g)	Fiber (g)	Sugars (g)	Sodium (mg)	Cholesterol (mg)
count	10000	10000.000000	10000.000000	10000.000000	10000.000000	10000.00000	10000.000000	10000.000000	10000.000000	10000.000000
mean	2024-07-01 02:49:37.920000	498.706300	327.693900	25.523050	52.568550	25.43735	4.986940	25.052570	497.969000	151.886600
min	2024-01-01 00:00:00	1.000000	50.000000	1.000000	5.000000	1.00000	0.000000	0.000000	0.000000	0.000000
25%	2024-04-02 00:00:00	245.000000	190.000000	13.200000	28.800000	13.30000	2.500000	12.500000	249.750000	76.000000
50%	2024-06-30 12:00:00	492.000000	328.000000	25.500000	52.800000	25.30000	5.000000	25.000000	495.000000	153.000000
75%	2024-09-29 00:00:00	748.000000	464.000000	37.700000	76.400000	37.60000	7.500000	37.700000	749.000000	228.000000
max	2024-12-31 00:00:00	1000.000000	600.000000	50.000000	100.000000	50.00000	10.000000	50.000000	1000.000000	300.000000
std	NaN	289.123477	158.194716	14.131993	27.387152	14.14532	2.864984	14.480605	287.988001	87.360643

Key Findings & Visualizations

1. Macronutrient Breakdown

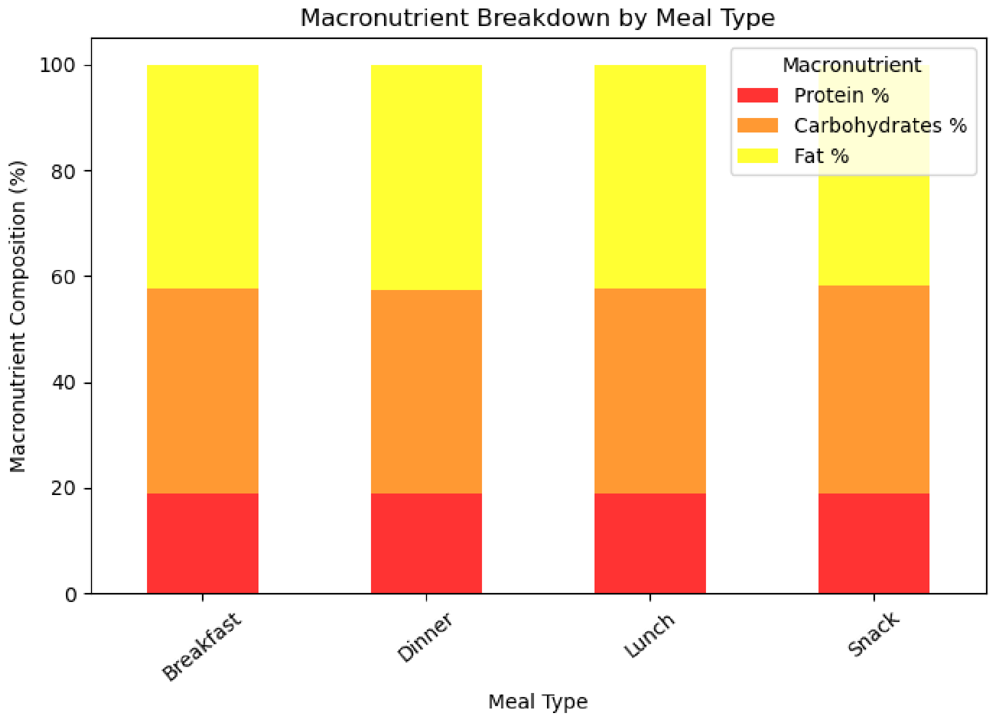
meal_macronutrient_ratios = df.groupby("Meal_Type")[["Protein (g)","Carbohydrates (g)","Fat (g)"]].mean()

Meal_Type	Protein (g)	Carbohydrates (g)	Fat (g)
Breakfast	25.334428	52.362564	25.484408
Dinner	25.686736	52.163843	25.702197
Lunch	25.407077	52.639928	25.446321
Snack	25.670502	53.124480	25.108650

```
meal_macronutrient_ratios["Protein %"] = meal_macronutrient_ratios["Protein (g)"] * 4 / (meal_macronutrient_ratios["Protein (g)"] * 4 + meal_macronutrient_ratios["Carbohydrates (g)"] * 4 + meal_macronutrient_ratios["Fat (g)"] * 9) * 100

meal_macronutrient_ratios["Carbohydrates %"] = meal_macronutrient_ratios["Carbohydrates (g)"] * 4 / (meal_macronutrient_ratios["Protein (g)"] * 4 + meal_macronutrient_ratios["Carbohydrates (g)"] * 4 + meal_macronutrient_ratios["Fat (g)"] * 9) * 100

meal_macronutrient_ratios["Fat %"] = meal_macronutrient_ratios["Fat (g)"] * 9 / (meal_macronutrient_ratios["Protein (g)"] * 4 + meal_macronutrient_ratios["Carbohydrates (g)"] * 4 + meal_macronutrient_ratios["Fat (g)"] * 9) * 100

meal_macronutrient_ratios[["Protein %","Carbohydrates %","Fat %"]]


meal_macronutrient_ratios = meal_macronutrient_ratios[["Protein %","Carbohydrates %","Fat %"]]
```

Meal_Type	Protein %	Carbohydrates %	Fat %
Breakfast	18.761113	38.776483	42.462404
Dinner	18.931778	38.446081	42.622141
Lunch	18.778157	38.905728	42.316115
Snack	18.974505	39.267277	41.758219

```
# Stacked bar chart: Macronutrient Breakdown by Meal Type

import matplotlib.pyplot as plt

meal_macronutrient_ratios.plot(kind = "bar" , stacked = True, figsize = (8,5) , colormap = "autumn" , alpha = 0.8)

plt.xlabel("Meal Type")

plt.ylabel("Macronutrient Composition (%)")

plt.title("Macronutrient Breakdown by Meal Type")

plt.xticks(rotation = 40)

plt.legend(title = "Macronutrient", loc = "upper right")

plt.show()
```



Insight: Different meal types and food categories have distinct macronutrient compositions.

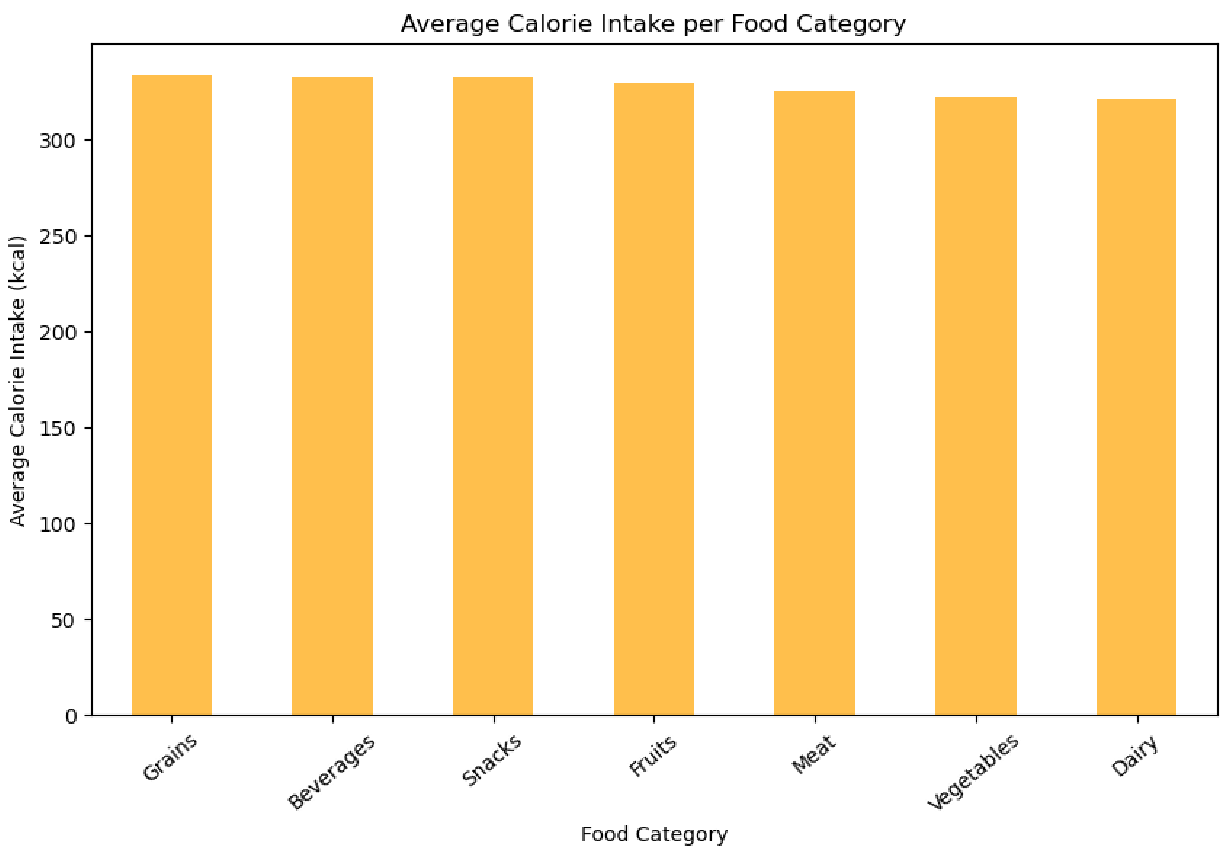
2. Food Category Contributions to Calorie Intake

```
food_category_calories = df.groupby("Category")["Calories (kcal)"].mean().sort_values(ascending = False)
```

```
Category
Grains        333.234104
Beverages     332.231142
Snacks        332.157821
Fruits        329.162423
Meat          325.011989
Vegetables    321.518466
Dairy         320.671918
Name: Calories (kcal), dtype: float64
```

```
# Average calorie intake per food category

plt.figure(figsize=(10,6))
```

```
food_category_calories.plot(kind = "bar",color = "orange" , alpha = 0.7)

plt.xlabel("Food Category")

plt.ylabel("Average Calorie Intake (kcal)")

plt.title("Average Calorie Intake per Food Category")

plt.xticks(rotation = 40)

plt.show()
```

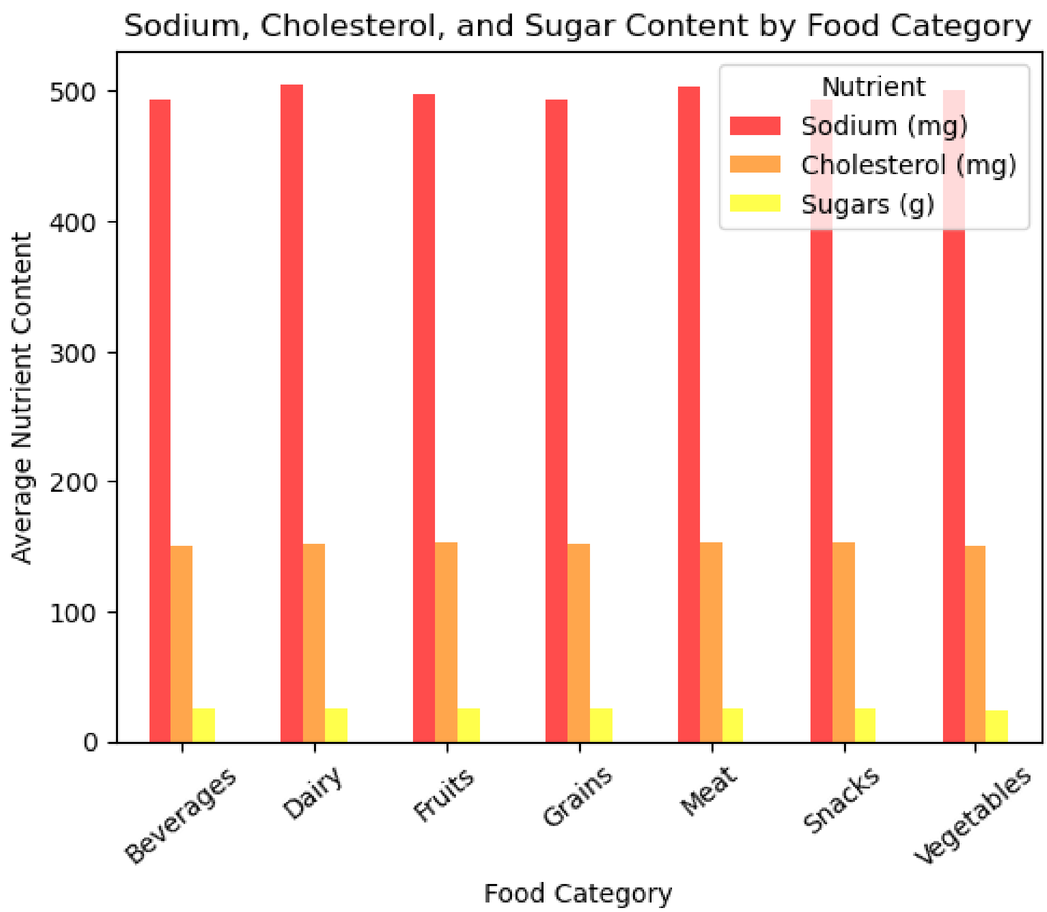



```
food_category_nutrient = df.groupby("Category")[["Sodium (mg)","Cholesterol (mg)","Sugars (g)"]].mean()
```

Category			
Beverages	493.152941	149.925952	25.509896
Dairy	505.434932	151.412329	25.372945
Fruits	497.543703	153.171370	25.028217
Grains	493.075867	151.884393	24.851662
Meat	503.227786	153.408322	24.872567
Snacks	493.217179	153.342877	25.125559
Vegetables	499.955256	150.053267	24.580682

```
# Sodium, Cholesterol, and S

plt.figure(figsize=(12,6))
```

```
food_category_nutrient.plot(kind = "bar", colormap = "autumn" , alpha = 0.7)
plt.xlabel("Food Category")
plt.ylabel("Average Nutrient Content")
plt.title("Sodium, Cholesterol, and Sugar Content by Food Category")
plt.xticks(rotation = 40)
plt.legend(title = "Nutrient" , loc = "upper right")
plt.show()
```



Insights:

- Some categories, such as grains and dairy, contribute significantly more to daily calorie intake than vegetables or fruits.
- High sodium content is a concern across all food categories, potentially indicating a reliance on processed or salty foods.

- Cholesterol levels are relatively balanced but higher in animal-based foods, reinforcing the importance of moderation in dairy and meat consumption.
- Sugar intake is lower but still present in some categories, especially Beverages and Fruits, which may indicate added sugars in certain products.

3. High-Sodium, High-Cholesterol, and High-Sugar Foods

```
# Identifying Top High-Sodium, High-Cholesterol, and High-Sugar Foods
top_sodium_foods = df.sort_values(by = "Sodium (mg)" , ascending= False).head(10)
top_cholesterol_foods = df.sort_values(by = "Cholesterol (mg)" , ascending = False).head(10)
top_sugar_foods = df.sort_values(by="Sugars (g)", ascending=False).head(10)
```

```
Top High-Sodium Foods:
```

tegory	Calories (kcal)	Protein (g)	Carbohydrates (g)	Fat (g)	Fiber (g)	Sugars (g)	Sodium (mg)	Cholesterol (mg)	Meal_Type	Water_Intake (ml)	Calculated_Calories	macronutrient
erages	576	18.3	67.1	29.9	5.4	8.6	1000	89	Lunch	844	610.7	610.7
Snacks	535	4.5	29.0	30.6	6.5	5.6	1000	26	Snack	617	409.4	409.4
Dairy	337	48.8	25.9	13.0	1.4	24.2	1000	281	Dinner	434	415.8	415.8
etables	131	29.1	84.7	30.6	0.7	14.3	1000	55	Breakfast	838	730.6	730.6
Snacks	160	1.4	26.5	40.2	8.6	41.5	1000	36	Breakfast	540	473.4	473.4
Grains	71	7.0	83.3	25.7	2.9	0.6	1000	79	Lunch	760	592.5	592.5
Snacks	212	26.8	29.3	15.4	9.0	25.2	1000	296	Dinner	618	363.0	363.0
etables	469	4.9	28.7	20.6	5.8	24.1	1000	107	Lunch	552	319.8	319.8
Dairy	124	31.7	42.2	7.8	9.4	27.7	1000	18	Dinner	114	365.8	365.8
Meat	148	23.9	92.5	48.0	8.1	7.2	1000	209	Breakfast	932	897.6	897.6

```
Top High-Cholesterol Foods:
```

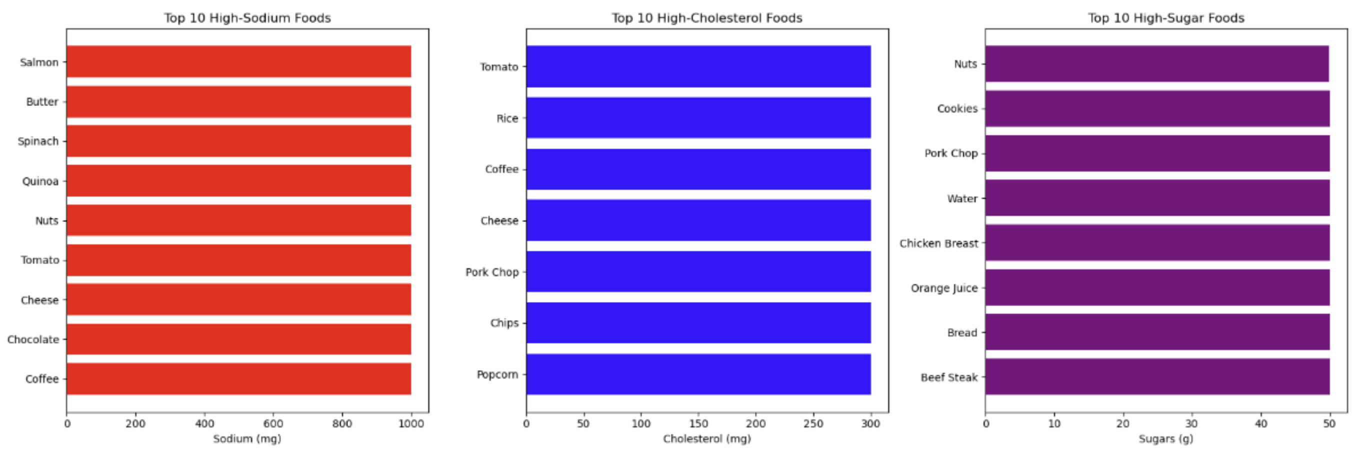
	Date	User_ID	Food_Item	Category	Calories (kcal)	Protein (g)	Carbohydrates (g)	Fat (g)	Fiber (g)	Sugars (g)	Sodium (mg)	Cholesterol (mg)	Meal_Type	Water_Intake (ml)
2337	2024-07-06	69	Popcorn	Snacks	183	43.7	79.6	7.5	8.1	1.0	115	300	Breakfast	852
9132	2024-03-23	239	Chips	Snacks	215	42.9	37.9	13.4	8.5	49.2	542	300	Snack	746
4205	2024-11-03	930	Pork Chop	Meat	180	25.4	97.8	23.1	8.9	38.6	289	300	Lunch	933
6104	2024-10-18	906	Cheese	Dairy	390	36.9	84.6	27.7	1.9	36.3	339	300	Dinner	278
5761	2024-11-15	308	Coffee	Beverages	439	19.0	10.1	8.4	5.3	33.7	754	300	Breakfast	472
8844	2024-01-29	857	Pork Chop	Meat	539	45.2	15.4	18.2	3.5	21.4	536	300	Dinner	514
5772	2024-06-27	292	Rice	Grains	185	12.1	64.0	10.4	2.9	19.1	182	300	Breakfast	955
4747	2024-07-20	823	Rice	Grains	183	34.2	9.5	37.4	8.7	36.6	30	300	Snack	958
3202	2024-03-02	931	Pork Chop	Meat	346	31.6	11.4	30.7	2.8	12.7	950	300	Snack	963
3750	2024-02-19	44	Tomato	Vegetables	415	33.1	8.6	21.7	4.0	36.7	360	300	Lunch	409

```
Top High-Sugar Foods:
```

	Date	User_ID	Food_Item	Category	Calories (kcal)	Protein (g)	Carbohydrates (g)	Fat (g)	Fiber (g)	Sugars (g)	Sodium (mg)	Cholesterol (mg)	Meal_Type	Water_Intake (ml)
1071	2024-09-22	158	Beef Steak	Meat	461	43.1	15.5	36.6	6.2	50.0	609	161	Dinner	966
3954	2024-04-01	668	Beef Steak	Meat	333	17.5	56.1	13.4	2.2	50.0	525	253	Breakfast	760
3786	2024-06-14	931	Bread	Grains	540	48.2	68.8	49.1	5.1	50.0	77	203	Lunch	789
6055	2024-07-14	422	Orange Juice	Beverages	556	48.8	5.0	38.1	3.9	50.0	54	121	Dinner	594
6040	2024-06-11	250	Chicken Breast	Meat	372	6.4	50.7	44.1	5.6	50.0	255	222	Breakfast	941
4286	2024-11-13	372	Water	Beverages	163	25.8	33.7	17.1	5.1	50.0	409	189	Dinner	243
5668	2024-08-18	772	Pork Chop	Meat	130	18.3	84.3	18.6	2.7	50.0	938	244	Snack	742
8834	2024-09-20	88	Cookies	Snacks	159	20.6	35.9	34.1	9.6	50.0	634	127	Snack	999
7985	2024-05-16	506	Nuts	Snacks	578	18.8	89.6	29.0	2.7	49.9	826	144	Dinner	211
2313	2024-06-29	188	Chicken Breast	Meat	331	29.8	77.8	31.6	7.2	49.9	16	125	Breakfast	692

```
fig, axes = plt.subplots(1, 3, figsize=(18,6))

# High Sodium Foods

axes[0].barh(top_sodium_foods["Food_Item"], top_sodium_foods["Sodium (mg)"], color="red")

axes[0].set_xlabel("Sodium (mg)")

axes[0].set_title("Top 10 High-Sodium Foods")

# High Cholesterol Foods
```

```
axes[1].barh(top_cholesterol_foods["Food_Item"], top_cholesterol_foods["Cholesterol (mg)"], color="blue")

axes[1].set_xlabel("Cholesterol (mg)")

axes[1].set_title("Top 10 High-Cholesterol Foods")

# High Sugar Foods

axes[2].barh(top_sugar_foods["Food_Item"], top_sugar_foods["Sugars (g)"], color="purple")

axes[2].set_xlabel("Sugars (g)")

axes[2].set_title("Top 10 High-Sugar Foods")

plt.tight_layout()

plt.show()
```



Insights:

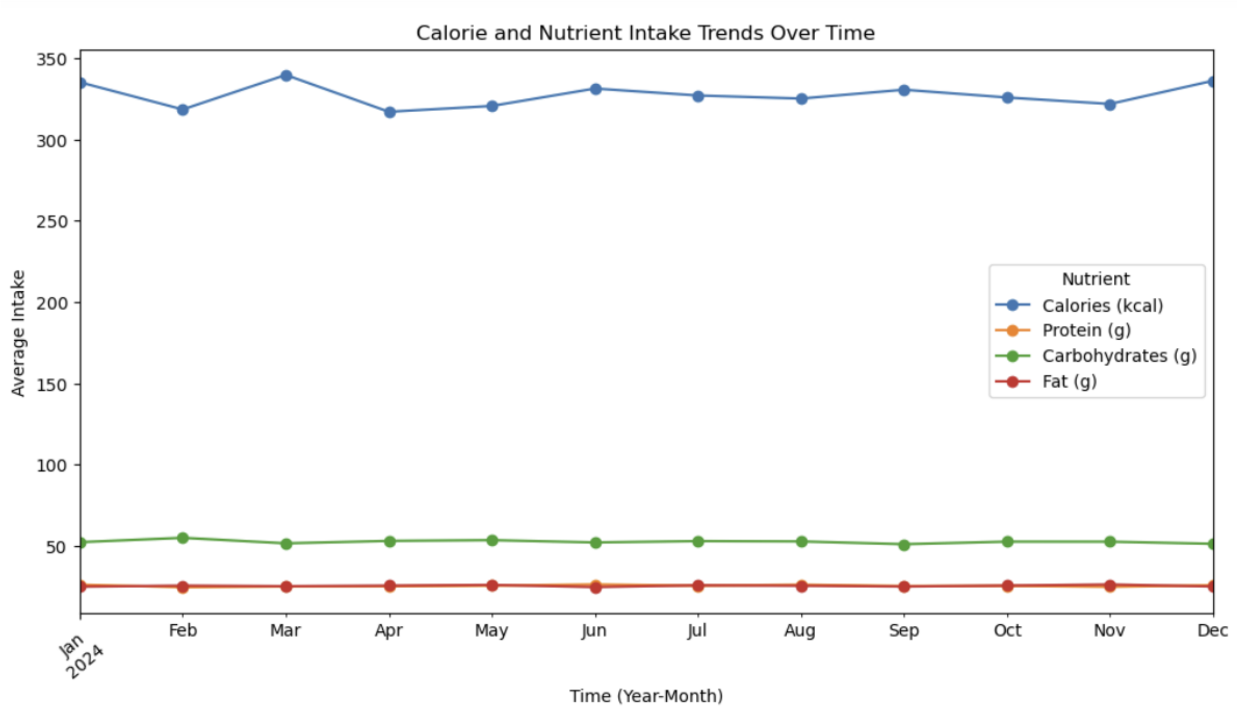
- Certain natural foods like **salmon, cheese, and spinach** have high sodium content, alongside some processed items like butter and chocolate.
- While dairy (cheese) and meat (pork chop) contain cholesterol, some unexpected items like **tomato and coffee** also appear on the list, possibly due to data inconsistencies.
- Sugary beverages like **orange juice and desserts (cookies)** contain high sugar levels, though some listed items (like pork chop and water) may require further investigation.

4. Trend Analysis Over Time

```
df["Year-Month"] = df["Date"].dt.to_period("M")

trend_analysis = df.groupby("Year-Month")[["Calories (kcal)", "Protein (g)", "Carbohydrates (g)", "Fat (g)"]].mean()
```

Year-Month	Calories (kcal)	Protein (g)	Carbohydrates (g)	Fat (g)
2024-01	335.301606	26.140482	52.321216	24.951491
2024-02	318.535912	24.621271	54.994613	25.567403
2024-03	339.788684	25.089261	51.608776	25.216975
2024-04	317.116029	25.172488	53.105981	25.577512
2024-05	320.698824	25.581294	53.556588	26.015529
2024-06	331.414319	26.322887	52.161854	24.710915
2024-07	327.134571	25.558585	52.987703	25.783295
2024-08	325.244946	26.147562	52.791082	25.517122
2024-09	330.686905	25.320119	51.067857	25.055595
2024-10	325.902728	25.453974	52.655160	25.612574
2024-11	321.921189	24.875065	52.630749	26.277519
2024-12	336.202381	25.792857	51.305119	25.065952

```
# Calorie and nutrient trends over time

plt.figure(figsize=(12,6))

trend_analysis.plot(marker = 'o' , figsize = (12,6))

plt.xlabel("Time (Year-Month)")

plt.ylabel("Average Intake")

plt.title("Calorie and Nutrient Intake Trends Over Time")

plt.xticks(rotation = 40)

plt.legend(title = "Nutrient")

plt.show()
```



Insights:

- Some months have higher calorie consumption, possibly due to holidays, seasonal food availability, or lifestyle changes.
- The overall diet composition remains stable, implying no major dietary shifts in macronutrient intake.

5. Comparison of Provided vs. Calculated Calories

Calories are derived from macronutrients using the following standard values:

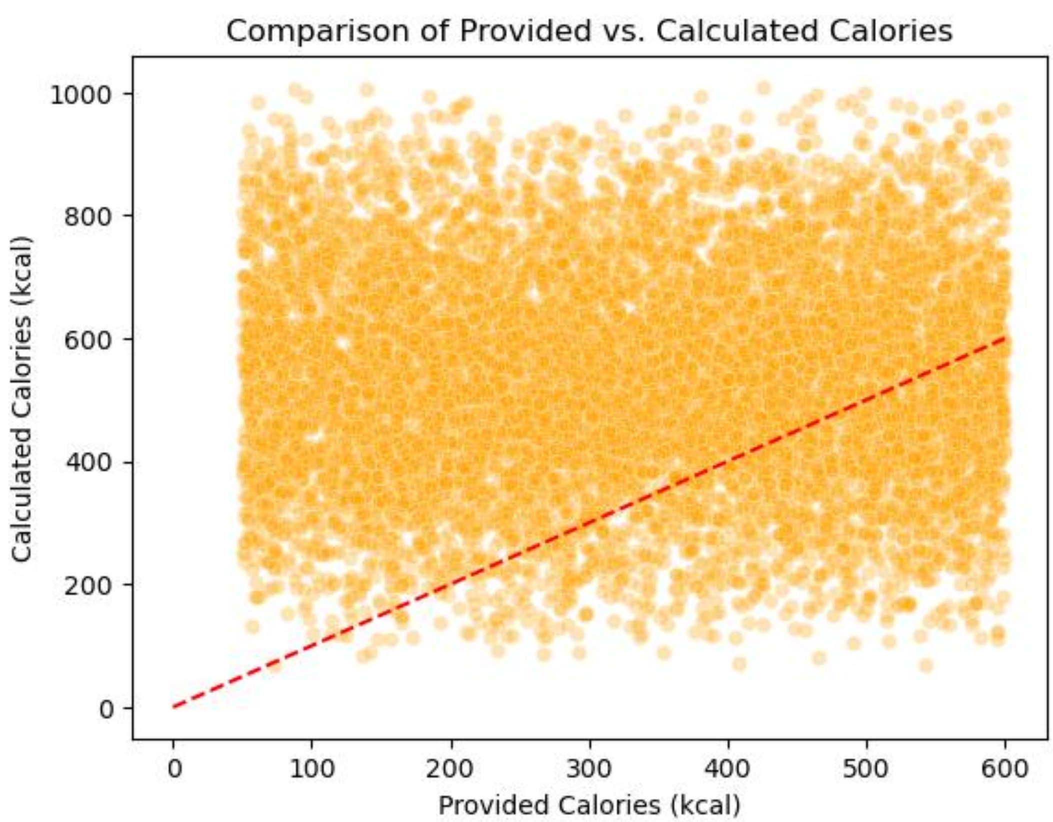
- **Protein:** 4 kcal per gram
- **Carbohydrates:** 4 kcal per gram
- **Fat:** 9 kcal per gram

Thus, the calculated calorie formula is:

Calories = (Protein×4) + (Carbohydrates×4) + (Fat×9)

```
df['Calculated_Calories'] = df['Protein (g)'] * 4 + df['Carbohydrates (g)'] * 4 + df['Fat (g)'] * 9

df['Macronutrient_Percentage'] = df['Calculated_Calories'] / df['Calories (kcal)'] * 100
```

```
import seaborn as sns

plt.figure(figsize=(8,6))

sns.scatterplot(x = df['Calories (kcal)'] , y = df['Calculated_Calories'] , alpha = 0.3 , color = 'orange')

plt.plot([0, df['Calories (kcal)'].max()], [0, df['Calories (kcal)'].max()], color='red', linestyle='--', label="Ideal Line")

plt.xlabel("Provided Calories (kcal)")

plt.ylabel("Calculated Calories (kcal)")

plt.title("Comparison of Provided vs. Calculated Calories")
```



Insights:

- There is a pattern of underreported calorie values, suggesting potential inaccuracies in the dataset.
- Food items with lower reported calories tend to have a much higher calculated calorie range, which could indicate missing data or incorrect food labeling.
- **Further investigation is needed** to determine if specific food categories have the highest discrepancies and whether certain macronutrients are consistently missing or underestimated.

Technical Tools and Skills

- Data Cleaning & Preprocessing
- Data Visualization & Storytelling
- Python (Pandas, Matplotlib, Seaborn)
- Jupyter Notebook

Next Steps

- **Further Investigate High Discrepancy Foods:** Identify specific food categories or items with the largest calorie misreporting.
- **Expand Analysis to Dietary Impact:** Investigate how these trends affect health metrics like BMI and nutritional balance.
- **Develop a Predictive Model:** Use machine learning to estimate missing nutritional values.

Conclusion

This project highlights the importance of accurate nutritional data and its impact on dietary choices. The analysis uncovered significant discrepancies in reported calorie values, emphasizing the need for better data integrity in food tracking. Additionally, identifying high-risk foods and nutrient trends provides valuable insights for health-conscious decision-making.