Recipes for Muscle Retention on a Cut

Name(s): Kenneth Xu

Website Link: https://kennethxu922.github.io/recipe-nutrition-analysis/

In [1]:
```
import pandas as pd
import numpy as np

import plotly.express as px
pd.options.plotting.backend = 'plotly'

from lec_utils import * # Feel free to uncomment and use this. It'll make yo
```

In [2]:
```
recipes = pd.read_csv('RAW_recipes.csv')
interactions = pd.read_csv('RAW_interactions.csv')
interactions
```

Out[2]:

	user_id	recipe_id	date	rating	review
0	1293707	40893	2011-12-21	5	So simple, so delicious! Great for chilly fall...
1	126440	85009	2010-02-27	5	I made the Mexican topping and took it to bunk...
2	57222	85009	2011-10-01	5	Made the cheddar bacon topping, adding a sprin...
...	...	...	...	...	...
731924	157126	78003	2008-06-23	5	WOW! Sometimes I don't take the time to rate ...
731925	53932	78003	2009-01-11	4	Very good! I used regular port as well. The ...
731926	2001868099	78003	2017-12-18	5	I am so glad I googled and found this here. Th...

731927 rows × 5 columns

In [3]:
```
# Left merge recipes and interactions on recipe ID
merged_recipes = pd.merge(
    recipes,
    interactions,
    left_on='id',
    right_on='recipe_id',
    how='left'
)

merged_recipes['rating'] = merged_recipes['rating'].replace(0, np.nan)

# Group by recipe_id and compute mean rating (ignoring NaN)
avg_rating = merged_recipes.groupby('recipe_id')['rating'].mean().reset_inde
```

```
avg_rating.rename(columns={'rating': 'avg_rating'}, inplace=True)

# Merge avg_rating back into the main DataFrame
merged_recipes = pd.merge(
    merged_recipes,
    avg_rating,
    on='recipe_id',
    how='left'
)

merged_recipes
```

Out[3]:

	name	id	minutes	contributor_id	...	date	rating	review	a
0	1 brownies in the world best ever	333281	40	985201	...	2008-11-19	4.0	These were pretty good, but took forever to ba...	
1	1 in canada chocolate chip cookies	453467	45	1848091	...	2012-01-26	5.0	Originally I was gonna cut the recipe in half ...	
2	412 broccoli casserole	306168	40	50969	...	2008-12-31	5.0	This was one of the best broccoli casseroles t...	
...	...	...	...	...	...	...	...	...	
234426	cookies by design sugar shortbread cookies	298509	20	506822	...	2008-06-19	1.0	This recipe tastes nothing like the Cookies by...	
234427	cookies by design sugar shortbread cookies	298509	20	506822	...	2010-02-08	5.0	yummy cookies, i love this recipe me and my sm...	
234428	cookies by design sugar shortbread cookies	298509	20	506822	...	2014-11-01	NaN	I work at a Cookies By Design and can say this...	

234429 rows × 18 columns

In [4]:

```
merged_recipes['nutrition'] = merged_recipes['nutrition'].apply(
    lambda x: eval(x) if isinstance(x, str) else x
)

nutrition_cols = [
    'calories',
    'total_fat_pdv',
    'sugar_pdv',
    'sodium_pdv',
    'protein_pdv',
```

```
    'saturated_fat_pdv',
    'carbohydrates_pdv'
]

merged_recipes[nutrition_cols] = pd.DataFrame(
    merged_recipes['nutrition'].tolist(),
    index=merged_recipes.index
)

# Convert all nutrition columns to numeric
merged_recipes[nutrition_cols] = merged_recipes[nutrition_cols].apply(pd.to_


merged_recipes[['name'] + nutrition_cols].head()
```

Out[4]:

	name	calories	total_fat_pdv	sugar_pdv	sodium_pdv	protein_pdv	saturated_
0	1 brownies in the world best ever	138.4	10.0	50.0	3.0	3.0	
1	1 in canada chocolate chip cookies	595.1	46.0	211.0	22.0	13.0	
2	412 broccoli casserole	194.8	20.0	6.0	32.0	22.0	
3	412 broccoli casserole	194.8	20.0	6.0	32.0	22.0	
4	412 broccoli casserole	194.8	20.0	6.0	32.0	22.0	

In [5]:
```
merged_recipes.head()
```

Out[5]:

	name	id	minutes	contributor_id	...	sodium_pdv	protein_pdv	saturate
0	1 brownies in the world best ever	333281	40	985201	...	3.0	3.0	
1	1 in canada chocolate chip cookies	453467	45	1848091	...	22.0	13.0	
2	412 broccoli casserole	306168	40	50969	...	32.0	22.0	
3	412 broccoli casserole	306168	40	50969	...	32.0	22.0	
4	412 broccoli casserole	306168	40	50969	...	32.0	22.0	

5 rows × 25 columns

Step 1: Introduction

In [6]:
```
Which high protein, low calorie recipes have the best ratings to help mainta
```

```
Object `cut` not found.
```

Step 2: Data Cleaning and Exploratory Data Analysis

In [7]:
```
# TODO
```

In [ ]:
```

```

In [ ]:
```

```

In [8]:
```
print(merged_recipes[['avg_rating', 'protein_pdv', 'calories']].isna().sum()
```

```
avg_rating     2777
protein_pdv       0
calories          0
dtype: int64
```

In [9]:
```
cleaned_recipes = merged_recipes.dropna(subset=['avg_rating'])

# Verify
print(cleaned_recipes[['avg_rating', 'protein_pdv', 'calories']].isna().sum(
```

```
avg_rating     0
protein_pdv    0
calories       0
dtype: int64
```

In [10]:
```
cleaned_recipes.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 231652 entries, 0 to 234428
Data columns (total 25 columns):
 #   Column             Non-Null Count   Dtype  
---  ------             --------------   -----  
 0   name               231652 non-null  object 
 1   id                 231652 non-null  int64  
 2   minutes            231652 non-null  int64  
 3   contributor_id     231652 non-null  int64  
 4   submitted          231652 non-null  object 
 5   tags               231652 non-null  object 
 6   nutrition          231652 non-null  object 
 7   n_steps            231652 non-null  int64  
 8   steps              231652 non-null  object 
 9   description        231542 non-null  object 
 10  ingredients        231652 non-null  object 
 11  n_ingredients      231652 non-null  int64  
 12  user_id            231652 non-null  float64
 13  recipe_id          231652 non-null  float64
 14  date               231652 non-null  object 
 15  rating             219393 non-null  float64
 16  review             231595 non-null  object 
 17  avg_rating         231652 non-null  float64
 18  calories           231652 non-null  float64
 19  total_fat_pdv      231652 non-null  float64
 20  sugar_pdv          231652 non-null  float64
 21  sodium_pdv         231652 non-null  float64
 22  protein_pdv        231652 non-null  float64
 23  saturated_fat_pdv  231652 non-null  float64
 24  carbohydrates_pdv  231652 non-null  float64
dtypes: float64(11), int64(5), object(9)
memory usage: 46.0+ MB
```

In [11]:
```
!conda install -c conda-forge tabulate --yes
```

```
Channels:
 - conda-forge
 - defaults
Platform: osx-arm64
Collecting package metadata (repodata.json): done
Solving environment: done


==> WARNING: A newer version of conda exists. <==
    current version: 24.11.2
    latest version: 25.3.1

Please update conda by running

    $ conda update -n base -c conda-forge conda



# All requested packages already installed.
```

In [12]:
```
print(
    cleaned_recipes.head()
    .to_markdown(index=False)
)
cleaned_recipes.head()
```

name	id	minutes	contributor_id	submitted	tags	nutrition	n_steps	steps	description	ingredients	n_ingredients	user_id	recipe_id	date	rating	review	avg_rating	calories	total_fat_pdv	sugar_pdv	sodium_pdv	protein_pdv	saturated_fat_pdv	carbohydrates_pdv
1 brownies in the world best ever	333281	40	985201	2008-10-27	['60-minutes-or-less', 'time-to-make', 'course', 'main-ingredient', 'preparation', 'for-large-groups', 'desserts', 'lunch', 'snacks', 'cookies-and-brownies', 'chocolate', 'bar-cookies', 'brownies', 'number-of-servings']	[138.4, 10.0, 50.0, 3.0, 3.0, 19.0, 6.0]	10	['heat the oven to 350f and arrange the rack in the middle', 'line an 8-by-8-inch glass baking dish with aluminum foil', 'combine chocolate and butter in a medium saucepan and cook over medium-low heat , stirring frequently , until evenly melted', 'remove from heat and let cool to room temperature', 'combine eggs , sugar , cocoa powder , vanilla extract , espresso , and salt in a large bowl and briefly stir until just evenly incorporated', 'add cooled chocolate and mix until uniform in color', 'add flour and stir until just incorporated', 'transfer batter to the prepared baking dish', 'bake until a tester inserted in the center of the brownies comes out clean , about 25 to 30 minutes', 'remove from the oven and cool completely before cutting']	these are the most; chocolatey, moist, rich, dense, fudgy, delicious brownies that you'll ever make.....sereiously! there's no doubt that these will b

e your fav brownies ever for you can add things to them or make them plai
n.....either way they're pure heaven!
| ['bittersweet chocolate', 'unsalted butter', 'eggs', 'granulated sugar',
'unsweetened cocoa powder', 'vanilla extract', 'brewed espresso', 'kosher sa
lt', 'all-purpose flour'] | 9 | 386585 | 333281
| 2008-11-19 | 4 | These were pretty good, but took forever to bake.
I would send it ended up being almost an hour! Even then, the brownies stuc
k to the foil, and were on the overly moist side and not easy to cut. They
did taste quite rich, though! Made for My 3 Chefs.
| 4 | 138.4 | 10 | 50 | 3 |
3 | 19 | 6 |
| 1 in canada chocolate chip cookies | 453467 | 45 | 18480
91 | 2011-04-11 | ['60-minutes-or-less', 'time-to-make', 'cuisine', 'prepar
ation', 'north-american', 'for-large-groups', 'canadian', 'british-columbia
n', 'number-of-servings']
| [595.1, 46.0, 211.0, 22.0, 13.0, 51.0, 26.0] | 12 | ['pre-heat oven
the 350 degrees f', 'in a mixing bowl , sift together the flours and baking
powder', 'set aside', 'in another mixing bowl , blend together the sugars ,
margarine , and salt until light and fluffy', 'add the eggs , water , and va
nilla to the margarine / sugar mixture and mix together until well combine
d', 'add in the flour mixture to the wet ingredients and blend until combine
d', 'scrape down the sides of the bowl and add the chocolate chips', 'mix un
til combined', 'scrape down the sides to the bowl again', 'using an ice crea
m scoop , scoop evenly rounded balls of dough and place of cookie sheet abou
t 1 - 2 inches apart to allow for spreading during baking', 'bake for 10 - 1
5 minutes or until golden brown on the outside and soft & chewy in the cente
r', 'serve hot and enjoy !'] | this is the recipe that we use at my school c
afeteria for chocolate chip cookies. they must be the best chocolate chip co
okies i have ever had! if you don't have margarine or don't like it, then ju
st use butter (softened) instead.
| ['white sugar', 'brown sugar', 'salt', 'margarine', 'eggs', 'vanilla', 'wa
ter', 'all-purpose flour', 'whole wheat flour', 'baking soda', 'chocolate ch
ips'] | 11 | 424680 | 453467
| 2012-01-26 | 5 | Originally I was gonna cut the recipe in half (jus
t the 2 of us here), but then we had a park-wide yard sale, & I made the who
le batch & used them as enticements for potential buyers ~ what the hey, a f
ree cookie as delicious as these are, definitely works its magic! Will be ma
king these again, for sure! Thanks for posting the recipe! | 5 |
595.1 | 46 | 211 | 22 | 13 |
51 | 26 |
| 412 broccoli casserole | 306168 | 40 | 509
69 | 2008-05-30 | ['60-minutes-or-less', 'time-to-make', 'course', 'main-in
gredient', 'preparation', 'side-dishes', 'vegetables', 'easy', 'beginner-coo
k', 'broccoli']
| [194.8, 20.0, 6.0, 32.0, 22.0, 36.0, 3.0] | 6 | ['preheat oven
to 350 degrees', 'spray a 2 quart baking dish with cooking spray , set asid
e', 'in a large bowl mix together broccoli , soup , one cup of cheese , garl
ic powder , pepper , salt , milk , 1 cup of french onions , and soy sauce',
'pour into baking dish , sprinkle remaining cheese over top', 'bake for 25 m
inutes or until cheese is lightly browned', 'sprinkle with rest of french fr
ied onions and bake until onions are browned and cheese is bubbly , about 10
more minutes']
| since there are already 411 recipes for broccoli casserole posted to "zaa
r" ,i decided to call this one #412 broccoli casserole.i don't think there
are any like this one in the database. i based this one on the famous "green
bean casserole" from campbell's soup. but i think mine is better since i do

n't like cream of mushroom soup.submitted to "zaar" on may 28th,2008 | ['fro
zen broccoli cuts', 'cream of chicken soup', 'sharp cheddar cheese', 'garlic
powder', 'ground black pepper', 'salt', 'milk', 'soy sauce', 'french-fried o
nions'] | 9 | 29782 | 306168 | 2008-1
2-31 | 5 | This was one of the best broccoli casseroles that I have e
ver made. I made my own chicken soup for this recipe. I was a bit worried a
bout the tsp of soy sauce but it gave the casserole the best flavor. YUM!
| 5 | 194.8 | 20 | 6 | 32 |
22 | 36 | 3 |
| | | |
| |
| | |
|
|
| | | | | |
The photos you took (shapeweaver) inspired me to make this recipe and it act
ually does look just like them when it comes out of the oven.
Thanks so much for sharing your recipe shapeweaver. It was wonderful! Going
into my family's favorite Zaar cookbook :)
412 broccoli casserole	306168	40	509
69	2008-05-30	['60-minutes-or-less', 'time-to-make', 'course', 'main-in
gredient', 'preparation', 'side-dishes', 'vegetables', 'easy', 'beginner-coo
k', 'broccoli']
[194.8, 20.0, 6.0, 32.0, 22.0, 36.0, 3.0]	6	['preheat oven
to 350 degrees', 'spray a 2 quart baking dish with cooking spray , set asid
e', 'in a large bowl mix together broccoli , soup , one cup of cheese , garl
ic powder , pepper , salt , milk , 1 cup of french onions , and soy sauce',
'pour into baking dish , sprinkle remaining cheese over top', 'bake for 25 m
inutes or until cheese is lightly browned', 'sprinkle with rest of french fr
ied onions and bake until onions are browned and cheese is bubbly , about 10
more minutes']
since there are already 411 recipes for broccoli casserole posted to "zaa
r" ,i decided to call this one #412 broccoli casserole.i don't think there
are any like this one in the database. i based this one on the famous "green
bean casserole" from campbell's soup. but i think mine is better since i do
n't like cream of mushroom soup.submitted to "zaar" on may 28th,2008	['fro
zen broccoli cuts', 'cream of chicken soup', 'sharp cheddar cheese', 'garlic
powder', 'ground black pepper', 'salt', 'milk', 'soy sauce', 'french-fried o
nions']	9	1.19628e+06	306168	2009-0
4-13	5	I made this for my son's first birthday party this weeken
d. Our guests INHALED it! Everyone kept saying how delicious it was. I was I
could have gotten to try it.
5	194.8	20	6	32
22 | 36 | 3 |

```
| 412 broccoli casserole                | 306168 |        40 |           509
69 | 2008-05-30  | ['60-minutes-or-less', 'time-to-make', 'course', 'main-in
gredient', 'preparation', 'side-dishes', 'vegetables', 'easy', 'beginner-coo
k', 'broccoli']
| [194.8, 20.0, 6.0, 32.0, 22.0, 36.0, 3.0]    |         6 | ['preheat oven
to 350 degrees', 'spray a 2 quart baking dish with cooking spray , set asid
e', 'in a large bowl mix together broccoli , soup , one cup of cheese , garl
ic powder , pepper , salt , milk , 1 cup of french onions , and soy sauce',
'pour into baking dish , sprinkle remaining cheese over top', 'bake for 25 m
inutes or until cheese is lightly browned', 'sprinkle with rest of french fr
ied onions and bake until onions are browned and cheese is bubbly , about 10
more minutes']
| since there are already 411 recipes for broccoli casserole posted to "zaa
r" ,i decided to call this one  #412 broccoli casserole.i don't think there
are any like this one in the database. i based this one on the famous "green
bean casserole" from campbell's soup. but i think mine is better since i do
n't like cream of mushroom soup.submitted to "zaar" on may 28th,2008 | ['fro
zen broccoli cuts', 'cream of chicken soup', 'sharp cheddar cheese', 'garlic
powder', 'ground black pepper', 'salt', 'milk', 'soy sauce', 'french-fried o
nions']          |               9 | 768828          |      306168 | 2013-0
8-02 |         5 | Loved this.  Be sure to completely thaw the broccoli.  I d
idn&#039;t and it didn&#039;t get done in time specified.  Just cooked it a
little longer though and it was perfect.  Thanks Chef.
|            5 |      194.8 |              20 |           6 |           32 |
22 |                 36 |                  3 |
```

Out[12]:

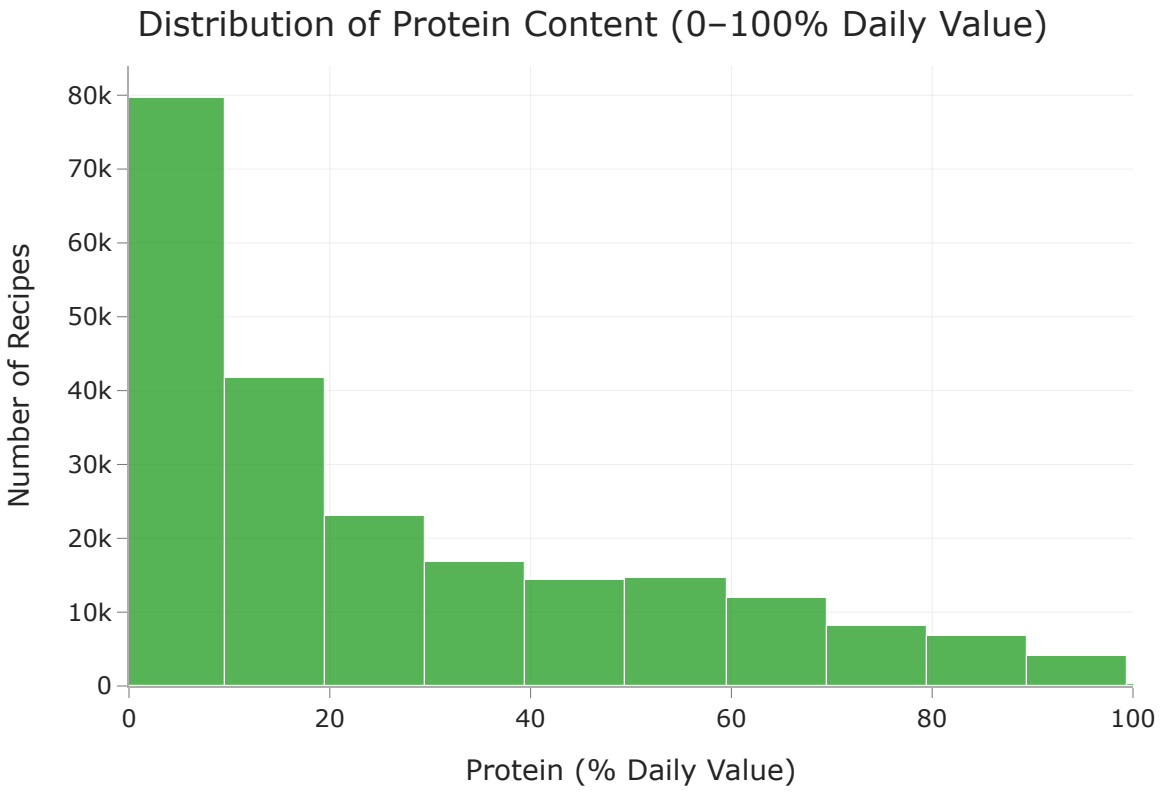
	name	id	minutes	contributor_id	...	sodium_pdv	protein_pdv	saturate
0	1 brownies in the world best ever	333281	40	985201	...	3.0	3.0	
1	1 in canada chocolate chip cookies	453467	45	1848091	...	22.0	13.0	
2	412 broccoli casserole	306168	40	50969	...	32.0	22.0	
3	412 broccoli casserole	306168	40	50969	...	32.0	22.0	
4	412 broccoli casserole	306168	40	50969	...	32.0	22.0	

5 rows × 25 columns

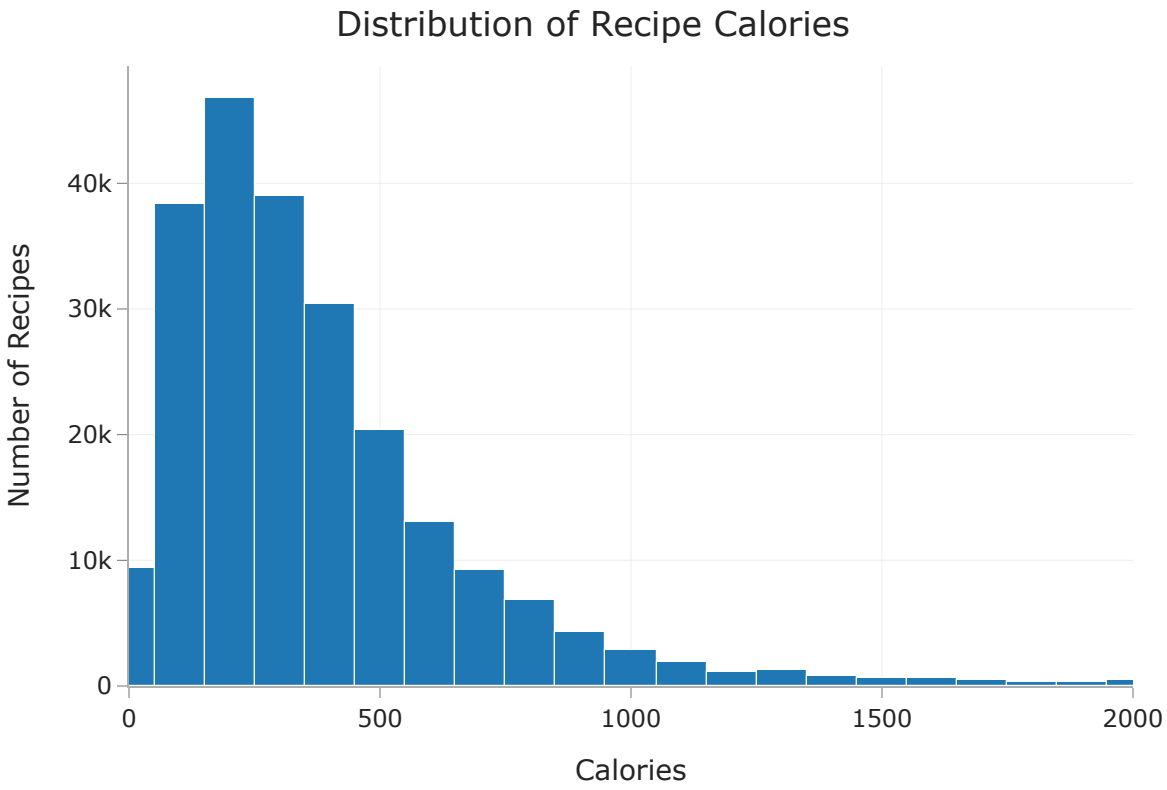
In [13]:

```
#univariate 1
import plotly.express as px

# Filter protein_pdv between 0–100%
```

```
protein_filtered = merged_recipes[
    (merged_recipes['protein_pdv'].between(0, 100)) &
    (merged_recipes['protein_pdv'].notna())
]

# Create histogram
fig = px.histogram(
    protein_filtered,
    x='protein_pdv',
    title='Distribution of Protein Content (0–100% Daily Value)',
    labels={'protein_pdv': 'Protein (% Daily Value)'},
    nbins=20,
    opacity=0.8,
    color_discrete_sequence=['#2ca02c']
)

fig.update_layout(
    xaxis_title='Protein (% Daily Value)',
    yaxis_title='Number of Recipes',
    hovermode='x unified',
    xaxis_range=[0, 100]
)

fig.show()
fig.write_html(
    'assets/protein-distribution.html',
    include_plotlyjs='cdn',
    full_html=True
)
```

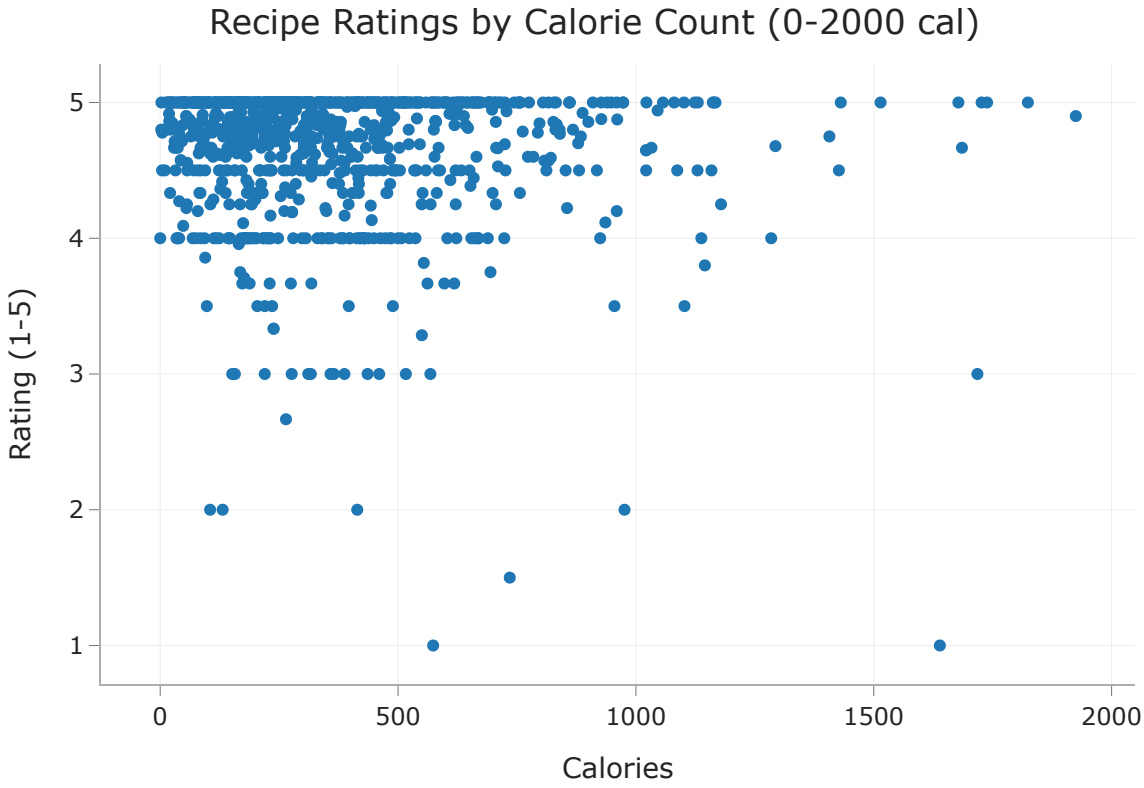


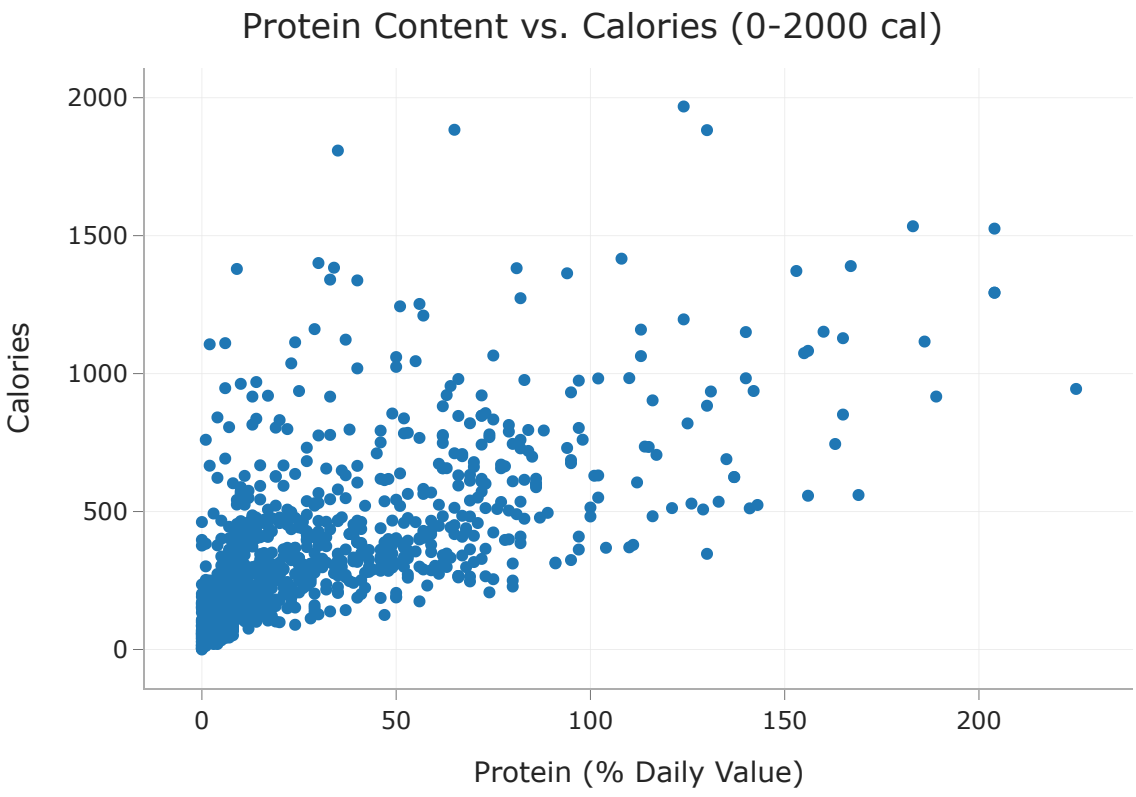
In [14]:
```
#univariate 2
fig = px.histogram(
    cleaned_recipes,
    x='calories',
    title='Distribution of Recipe Calories',
    labels={'calories': 'Calories per Serving'},
    range_x=[0, 2000],
    nbins=500,
    color_discrete_sequence=['#1f77b4']
)
fig.update_layout(
    xaxis_title='Calories',
    yaxis_title='Number of Recipes'
)
fig.show()
```



In [15]:
```
#Bivariate Plots
fig1 = px.scatter(
    cleaned_recipes[cleaned_recipes['calories'] <= 2000].sample(1000),
    x='calories',
    y='avg_rating',
    title='Recipe Ratings by Calorie Count (0-2000 cal)',
    labels={'calories': 'Calories', 'avg_rating': 'Rating (1-5)'}
)
fig1.show()

fig2 = px.scatter(
    cleaned_recipes[cleaned_recipes['calories'] <= 2000].sample(1000),
    x='protein_pdv',
    y='calories',
    title='Protein Content vs. Calories (0-2000 cal)',
```

```
    labels={'protein_pdv': 'Protein (% Daily Value)', 'calories': 'Calories'
)
fig2.show()
fig1.write_html(
    'assets/calories-vs-rating.html',
    include_plotlyjs='cdn',
    full_html=True
)
```





In [16]:
```
# Create expanded protein tiers
protein_groups = cleaned_recipes.groupby(
    pd.cut(cleaned_recipes['protein_pdv'],
           bins=[0, 20, 30, 40, 60, 100, 200],
           labels=['Very Low (0-20%)',
                   'Low (20-30%)',
                   'Moderate (30-40%)',
                   'High (40-60%)',
                   'Very High (60-100%)',
                   'Extreme (100%+)'])
).agg({
    'avg_rating': 'mean',
    'calories': 'median',
    'name': 'count'
}).rename(columns={
    'avg_rating': 'Avg Rating',
    'calories': 'Median Calories',
    'name': 'Recipe Count'
})

protein_groups.style.format({
    'Avg Rating': '{:.1f}',
    'Median Calories': '{:.0f}'
}).applymap(lambda x: 'background-color: lightgreen'
            if x in ['High (40-60%)','Very High (60-100%)']
            else '')
```

Out[16]:

protein_pdv	Avg Rating	Median Calories	Recipe Count
Very Low (0-20%)	4.7	197	111693
Low (20-30%)	4.7	348	21844
Moderate (30-40%)	4.7	374	16334
High (40-60%)	4.7	413	28758
Very High (60-100%)	4.7	568	29616
Extreme (100%+)	4.6	884	11113

Step 3: Framing a Prediction Problem

In [17]:
```
print("Can we predict a recipe’s average rating (1–5 stars) based solely on
```

```
Can we predict a recipe’s average rating (1–5 stars) based solely on its nut
ritional profile and preparation complexity, to identify high-protein, low-c
alorie recipes that are likely to taste good?
```

Step 4: Baseline Model

In [18]:
```
from sklearn.linear_model import LinearRegression
from sklearn.pipeline import make_pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score

# Step 4: Baseline Model
# Select simplest features (protein + calories)
features = ['protein_pdv', 'calories']
X = cleaned_recipes[features]
y = cleaned_recipes['avg_rating']

# Train-test split (80-20)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, rar

# Create and fit pipeline (scaling + linear regression)
baseline_model = make_pipeline(
    StandardScaler(),
    LinearRegression()
)
baseline_model.fit(X_train, y_train)

# Evaluate
train_score = baseline_model.score(X_train, y_train)
test_score = baseline_model.score(X_test, y_test)

print(f"Baseline Model R² Scores:")
print(f"- Train: {train_score:.3f}")
```

```
print(f"- Test: {test_score:.3f}")

# Check coefficients
print("\nFeature Coefficients:")
for feat, coef in zip(features, baseline_model.named_steps['linearregressior
    print(f"- {feat}: {coef:.3f}")

print(baseline_model.named_steps['standardscaler'].mean_)
```

```
Baseline Model R² Scores:
- Train: 0.000
- Test: 0.000

Feature Coefficients:
- protein_pdv: -0.005
- calories: -0.003
[ 33.14 417.93]
```

Step 5: Final Model

In [21]:

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.pipeline import Pipeline
from sklearn.model_selection import GridSearchCV, train_test_split

# 1) Feature engineering
cleaned_recipes['protein_per_calorie'] = cleaned_recipes['protein_pdv'] / cl
cleaned_recipes['is_quick_meal']       = (cleaned_recipes['minutes'] <= 30).
features = ['protein_pdv', 'calories', 'protein_per_calorie', 'is_quick_meal

# 2) Train/test split
X = cleaned_recipes[features]
y = cleaned_recipes['avg_rating']
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# 3) Define pipeline and hyperparameter grid
final_pipeline = Pipeline([
    ('model', RandomForestRegressor(random_state=42))
])
param_grid = {
    'model__n_estimators': [50],
    'model__max_depth':   [10]
}

# 4) Grid search (2-fold CV on a 10k subset for speed)
grid_search = GridSearchCV(final_pipeline, param_grid, cv=2, scoring='r2')
X_sub = X_train.sample(10000, random_state=42)
y_sub = y_train.loc[X_sub.index]
grid_search.fit(X_sub, y_sub)

# 5) Extract best params and retrain on full set
rf_params = {k.split("__")[1]: v for k, v in grid_search.best_params_.items(
final_model = RandomForestRegressor(random_state=42, **rf_params)
final_model.fit(X_train, y_train)
```

```
# 6) Evaluate
final_test_score = final_model.score(X_test, y_test)
print(f"Final Model Test R²: {final_test_score:.3f}")
```

```
Final Model Test R²: 0.042
```

In [ ]: