BİL470 ARA RAPOR FİTOLL

Atakan Ekiz - Bahadır İhsan Herdem

1. State of Data Collection (Veri Toplama Durumu)

Bu proje kapsamında beslenme ve sağlık odaklı dört farklı veri seti bir araya getirilmiştir.
Tüm veri setleri Kaggle ve açık veri kaynaklarından temin edilmiş olup, veri toplama süreci tamamlanmıştır.
Veriler, bireylerin beslenme alışkanlıklarını, öğün bileşimlerini, vücut ölçüleriyle ilişkilerini ve sağlık hedeflerine (örneğin kilo verme, kas kazanımı) yönelik örüntüleri incelemeyi amaçlamaktadır.

Kullanılan veri setleri:

1. **daily_food_nutrition_dataset.csv:** Günlük yiyeceklerin kalori, protein, yağ, karbonhidrat, sodyum ve şeker oranları gibi besin değerlerini içermektedir.
2. **Food_and_Nutrition_new.csv:** Makro besin öğelerine (protein, yağ, karbonhidrat) göre detaylı besin dağılımlarını ve porsiyon bazlı kalori değerlerini sunmaktadır.
3. **GYM.csv:** Kullanıcıların hedefleri (fat burn, muscle gain, maintain weight) ve BMI, yaş, cinsiyet gibi fiziksel özellikleri yer almaktadır.
4. **healthy_meal_plans.csv:** Öğünlerin “healthy (1)” veya “not healthy (0)” olarak etiketlendiği sınıflandırma verisidir.

Tüm veri setleri temizlenmiş, eksik değer kontrolleri yapılmış ve gerekli durumlarda veri tipleri dönüştürülmüştür.
Ayrıca ortak sütunlar (örneğin *calories*, *protein*, *fat*, *carbs*) üzerinde veri standardizasyonu sağlanarak analizlerin birbiriyle uyumlu çalışması hedeflenmiştir.

Ön analizlerde **“is_healthy”** değişkeninde ciddi bir sınıf dengesizliği tespit edilmiştir (yaklaşık %90 “not healthy”, %10 “healthy”).
Bu dengesizliği gidermek amacıyla, modelleme aşamasında **SMOTE (Synthetic Minority Oversampling Technique)** yöntemiyle veri dengelemesi uygulanmıştır.
Böylece her iki sınıf da eşit temsil edilerek modelin adil, tarafsız ve yüksek genelleme yeteneğine sahip sonuçlar üretmesi hedeflenmektedir.

2. Exploratory Data Analysis (EDA)

2.1 Daily Food Nutrition Dataset Analizi

Bu veri seti, günlük yiyeceklerin besin içeriklerini kapsamaktadır. Her bir örnek; kalori (kcal), protein (g), karbonhidrat (g), yağ (g), lif (g), şeker (g), sodyum (mg), kolesterol (mg) ve su alımı (ml) gibi temel besin değerlerini içermektedir.
Bu analiz, veri dağılımlarını, olası uç değerleri ve değişkenler arası ilişkileri belirlemeyi amaçlamaktadır.

Histogram Analizleri

Yapılan histogram analizleri (Protein, Fat, Carbohydrates, Fiber, Sugars, Sodium, Cholesterol, Water Intake) besin bileşenlerinin **geniş ve dengeli bir dağılım** gösterdiğini ortaya koymaktadır.

Her değişken yaklaşık olarak **üniform (düzgün) dağılım** göstermekte, yani örnekler tüm değer aralığına homojen bir şekilde dağılmıştır.

Bu durum, modelin belirli bir aralıkta yoğunlaşan (örneğin düşük kaloriye aşırı eğilimli) bir veri yapısından etkilenmeden genelleme yapabileceğini gösterir.

- **Carbohydrates (g):** Karbonhidrat miktarları 0–100 g aralığında eşit yayılım göstermektedir. Bu durum hem düşük karbonhidratlı hem de yüksek karbonhidratlı besinlerin veri setinde dengeli şekilde yer aldığını gösterir.
- **Protein (g):** Protein değerleri benzer biçimde 0–50 g aralığında yayılmıştır; bu da hem bitkisel hem hayvansal protein kaynaklarının dengeli temsil edildiğini düşündürür.
- **Fat (g) & Cholesterol (mg):** Yağ ve kolesterol değerleri arasında genel bir paralellik vardır; yüksek yağ içeriği yüksek kolesterolle ilişkilidir.
- **Fiber (g):** Lif içeriği genellikle düşük seviyelerde yoğunlaşmakla birlikte bazı yüksek lifli örnekler de veri içinde bulunmaktadır.
- **Sodium (mg):** 0–1000 mg aralığında geniş bir dağılım göstermektedir; bu, hem düşük tuzlu hem de yüksek sodyumlu gıdaların bulunduğunu gösterir.
- **Water Intake (ml):** Su alımı da dengeli bir dağılım sergilemektedir.
- **Calories (kcal):** 50–600 aralığında dengeli bir dağılım gözlenmiştir. Bu, düşük kalorili ve yüksek kalorili öğünlerin dengeli biçimde temsil edildiğini göstermektedir.

Korelasyon Analizi

“Besin Değerleri Korelasyon Isı Haritası” incelendiğinde, değişkenler arasında **çok düşük(0) korelasyon değerleri** gözlenmiştir.

Bu durum, her besin bileşeninin bağımsız olarak değiştiğini ve veri setinde **multicollinearity (çoklu doğrusal ilişki)** probleminin olmadığını göstermektedir.

Yani kaloriyi belirleyen faktörler arasında güçlü bir doğrusal ilişki bulunmamaktadır; örneğin protein artışı her zaman yağ veya karbonhidrat artışıyla paralel ilerlememektedir.

Bunun sebebi de datasetin her değeri birbirinden bağımsız olarak almış olmasıdır.Yani her bir besin değeri ayrı bir feature olarak incelenmiştir zaten o yüzden birbirleriyle bağları yoktur.

Meal Type Analizi

“Meal Type Bazında Ortalama Kalori” grafiğine göre kahvaltı, öğle ve akşam yemeklerinin ortalama kalori değerleri **yaklaşık 325–330 kcal** civarındadır.
Ara öğünler (“Snack”) ise ortalama **320 kcal** ile biraz daha düşük enerji değerine sahiptir.
Bu da veri setinde öğünler arasında kalori farkının çok yüksek olmadığını ve her öğünün dengeli şekilde temsil edildiğini göstermektedir.

Boxplot (Dağılım ve Ölçek Analizi)

Veri setindeki tüm besin değişkenleri için dağılımı ve olası uç değerleri incelemek amacıyla iki tür boxplot oluşturulmuştur:
biri ham değerler, diğeri ise log(1+x) dönüşümü uygulanmış ölçek üzerindedir.

▫ **Normal Ölçekli Boxplot:**

Ham değerlerle çizilen boxplot grafiğinde, besin değişkenlerinin birbirine göre ölçek farkı çok yüksek olduğu için çoğu sütun görselde tabanda sıkışmış şekilde görünmektedir.
Bu, kalori, sodyum ve su alımı gibi yüksek ölçekli değişkenlerin diğer bileşenlerin (örneğin lif veya yağ) dağılımlarını gölgelemesine neden olmuştur.
Bu nedenle bu grafikte uç değer gözlemi yapmak zordur; dağılımın gerçek yapısı net bir şekilde seçilememektedir.
Ancak bu durum, veride bir problem olduğu anlamına gelmez — sadece ölçek farklılığının görsel analizi zorlaştırdığı anlamına gelir.

▫ **Log Ölçekli Boxplot:**

Logaritmik dönüşüm sonrası çizilen grafikte ise dağılımlar net bir biçimde ortaya çıkmıştır.
Log(1+x) dönüşümü sayesinde her değişken benzer ölçekte değerlendirilmiş ve gözlemler arasındaki farklar görünür hale gelmiştir.

- Calories, Sodium, Cholesterol ve Water Intake gibi yüksek ölçekli değişkenler normalize edilerek diğer değişkenlerle aynı düzleme taşınmıştır.
- Protein, Carbohydrates, Fat ve Sugars değişkenleri log dönüşümü sonrası orta aralıklarda dengeli dağılım göstermektedir.
- Herhangi bir değişkende belirgin uç değer kümesi bulunmamıştır; bu da verinin iyi temizlenmiş, dengeli ve uç değerlerden arındırılmış olduğunu göstermektedir.

Bu gözlemler, modelleme aşamasında ölçek farklarını minimize etmek için log transform veya standardizasyon gibi tekniklerin uygun olacağını göstermektedir.
Ayrıca verinin bu haliyle istatistiksel açıdan kararlı ve homojen bir yapıya sahip olduğu sonucuna varılabilir.

2.1Genel Değerlendirme

Bu veri setine ilişkin gözlemler şu şekilde özetlenebilir

- Veri dağılımları homojen ve geniş aralıklara yayılmıştır.
- Uç değerler özellikle sodyum ve kolesterol değişkenlerinde yoğunlaşmıştır.
- Log dönüşümü, verinin daha dengeli bir yapıya kavuşmasını sağlamıştır.
- Değişkenler arası korelasyon düşüktür, bu da modelin açıklayıcı gücünü artırmaktadır.
- Veri yapısı, besin değeri tahmini, sağlıklı/sağlıksız sınıflandırma veya kalori optimizasyonu gibi makine öğrenimi görevleri için uygundur.

2.2 Diet Recommendation Dataset Analizi

Bu veri seti, bireylerin boy, kilo, yaş, aktivite seviyesi, günlük kalori hedefi ve makro besin tüketim değerleri (protein, karbonhidrat, yağ, lif, su) gibi temel sağlık ve beslenme değişkenlerini içermektedir.
Ana amaç, bu değişkenler üzerinden bireylerin enerji dengelerini analiz etmek, kalori hedeflerinden sapma durumlarını belirlemek ve kişiselleştirilmiş diyet öneri modelleri geliştirmektir.

Ham veri bu temel parametreleri içermekle birlikte, analiz sürecinde modelleme kabiliyetini artırmak amacıyla çok sayıda yeni türetilmiş değişken (engineered features) oluşturulmuştur.
Bu adım, veri kümesini yalnızca betimsel analiz için değil, aynı zamanda tahmine dayalı (predictive) modelleme için de uygun hale getirmiştir.

Aşağıda oluşturulan yeni değişkenler ve bu değişkenlerin işlevleri yer almaktadır:

2.2.1 **BMI (Body Mass Index – Vücut Kitle İndeksi)**

Kilo ve boy verilerinden hesaplanarak bireyin vücut kompozisyonunu temsil eder.

Bu değişken, bireyin kilolu, obez ya da zayıf olma durumunu belirlemek için temel göstergedir.

BMI_Class

BMI değerleri kategorik olarak sınıflandırılmıştır:

- *Underweight: BMI < 18.5*
- *Normal: 18.5 ≤ BMI < 25*
- *Overweight: 25 ≤ BMI < 30*
- *Obese: BMI ≥ 30*

Bu sınıflandırma, modelin BMI'a göre kalori açığı veya beslenme davranışlarını analiz edebilmesine olanak tanır.

Macro_kcal

Besinlerden alınan toplam enerji hesaplanmıştır:

- *Pozitif değer: hedefin üstünde alım*
- *Negatif değer: kalori açığı (hedefin altında tüketim)*

Bu değişken, enerji dengesinin en doğrudan ölçütüdür ve analizlerde hedef uyumunu değerlendirmek için ana metrik olarak kullanılmıştır.

2.2.2 Makro Besin Yüzdeleri

Toplam enerjinin hangi oranda protein, karbonhidrat ve yağdan geldiğini gösterir:

Bu yüzdeler, bireyin diyet dengesini analiz etmekte ve yaş/aktivite gruplarına göre makro dağılım karşılaştırmalarında kullanılmıştır.

2.2.3 Genel Dağılım Analizleri

BMI (Vücut Kitle İndeksi) Dağılımı

Histogram grafiğine göre BMI değerleri 25 civarında yoğunlaşmaktadır; bu da katılımcıların önemli bir kısmının normal ile hafif kilolu (overweight) aralığında olduğunu göstermektedir.
Dağılım hafif sağa çarpıktır (right-skewed), yani aşırı kilolu bireyler az sayıda da olsa veri setinde bulunmaktadır.

Makro Besin Yüzdeleri (Protein_%, Carb_%, Fat_%)

- *Protein_% dağılımı 20-30 % aralığında, normalde çift tepe (bimodal) yapıya sahiptir. Bazı örneklerde 30 % civarında sabit bir değer görülmekte, bu da modelleme öncesi veri oluşturma veya yuvarlama etkisini düşündürmektedir.*
- *Carb_% (karbonhidrat yüzdesi) 40-55 % arasında dağılmış olup, genellikle 50 % civarında merkezlenmiştir.*
- *Fat_% dağılımı 25-30 % civarında toplanmış, dengeli bir sağa çarpıklık göstermektedir.*

Bu üç makro bileşenin toplamı yaklaşık 100 % olup, veri genel olarak makro denge açısından geçerli ve tutarlı görünmektedir.

2.2.2 Korelasyon Analizi

Korelasyon ısı haritasına göre:

- *Weight, Height ve BMI arasında yüksek pozitif korelasyon vardır.*
- *Calories, Macro_kcal ve Daily Calorie Target değişkenleri de güçlü şekilde birbiriyle ilişkili olup enerji dengesinin modelde ana belirleyicileridir.*
- *Protein, Carbohydrates ve Fat arasındaki korelasyonlar orta düzeydedir; ancak yüzdelik halleri (Protein_%, Carb_%, Fat_%) birbirini tamamlayan ters ilişkiler*

göstermektedir (örneğin Protein_% artarsa → Fat_% azalır).

- *Calorie_Gap, kalori hedefinden sapmayı ifade eder ve BMI ile düşük pozitif korelasyon göstermektedir; yani BMI yükseldikçe kalori fazlası verme eğilimi artmaktadır.*

2.2.3 Yaş ve Aktivite Düzeyine Göre Kalori Analizi

Yaş Gruplarına Göre Calorie Gap

Boxplot analizine göre genç yaş gruplarında (< 18) hedef kaloriye ulaşamama eğilimi daha yüksektir (negatif Calorie_Gap).
19-60 yaş arası bireylerde fark daha dengelidir, ancak 46 yaş sonrası tekrar negatif yönde sapma görülmektedir.
Bu durum, yaş arttıkça metabolik hızın azalmasına ve enerji dengesinin negatifleşmesine bağlanabilir.

Aktivite Seviyesine Göre Calorie Gap (Log Ölçeğinde)

Log-ölçekli boxplot'a göre "Extremely Active" bireylerin Calorie_Gap değerleri en düşük varyansa sahiptir; yani enerji dengesi daha stabildir.
Diğer gruplarda özellikle Moderately Active ve Sedentary bireylerde sapma aralığı geniştir.
Bu da düşük fiziksel aktivite düzeylerinde beslenme ile harcama arasındaki dengesizliğin arttığını göstermektedir.

2.2.4 Makro Dağılımının Yaş Gruplarına Göre İncelenmesi

Yaş gruplarına göre makro oranlarının boxplot'ları incelendiğinde:

- *Karbonhidrat oranı tüm yaş gruplarında en yüksek bileşendir (ortalama ~45 %).*
- *Protein oranı 20-30 % civarında seyretmektedir; 36-45 yaş aralığında en dengeli düzeydedir.*
- *Yağ oranı yaş arttıkça hafif artış göstermektedir, bu da metabolik adaptasyonla ilişkilidir.*

Bu bulgular, diyet önerisi oluştururken yaşa bağlı makro ayarlamaları yapılması gerektiğini vurgulamaktadır.

2.2.5 BMI Sınıfına Göre Ortalama Calorie Gap

Bar grafiğinde görüldüğü üzere:

- *Overweight (fazla kilolu) bireylerin ortalama Calorie_Gap değeri en düşük (yaklaşık −100 kcal),*
- *Obese (obez) bireylerde ise biraz daha yüksek (yaklaşık −50 kcal) çıkmıştır.*
  Bu, fazla kilolu bireylerin genellikle enerji fazlası oluşturduklarını, obez bireylerin ise diyet uygulaması nedeniyle kısmen enerji açığı yaşadıklarını gösterebilir.

2.2.6 Genel Değerlendirme

Bu analizler sonucunda veri seti:

- *BMI, yaş ve aktivite değişkenlerinin Calorie_Gap üzerinde anlamlı etkisi olduğunu,*
- *Makro bileşenlerin dağılımının genel olarak tutarlı ancak bazı sabit değer kümeleri içerdiğini,*
- *Enerji dengesizliğinin özellikle düşük aktivite ve genç yaş gruplarında belirginleştiğini göstermektedir.*

Modelleme aşamasında bu içgörüler, kişiselleştirilmiş diyet öneri sisteminde özelleştirilmiş makro oranları ve kalori hedefleri belirlemek için kullanılacaktır.
Ek olarak, veri log-dönüşümü ve standardizasyon ile normalize edilerek makine öğrenmesi tabanlı "kalori dengeleme" modeline girdi olarak kullanılacaktır.

2.3 Meal-Ex Planner Dataset (GYM.csv)

Bu veri seti, kullanıcıların hedefleri (Goal), BMI kategorileri, egzersiz planları, beslenme planları ve cinsiyet gibi değişkenleri içermektedir.
Amaç, bireylerin kilo kontrol hedeflerine (yağ yakımı veya kas kazanımı) göre diyet ve egzersiz düzenlerini analiz etmek, değişkenler arasındaki olası ilişkileri ortaya koymaktır.

2.3.1 Hedef (Goal) Dağılımı

Veri kümesinde iki ana hedef türü bulunmaktadır:

- *Muscle gain (kas kazanımı)*
- *Fat burn (yağ yakımı)*

Bar grafiğine göre iki hedef türü de yaklaşık eşit frekansta görülmektedir. Bu durum, veri setinin dengeli bir dağılıma sahip olduğunu ve modelleme aşamasında hedef değişkeninin sınıf dengesizliği problemi yaratmayacağını göstermektedir.

2.3.2 BMI Category Dağılımı

Katılımcıların vücut kitle indeksine (BMI) göre dağılımları şu şekildedir:
Underweight, Normal weight, Overweight ve Obesity kategorilerinin oranları birbirine oldukça yakındır.
Bu dağılım, veri setinin farklı vücut tiplerini eşit temsil ettiğini göstermekte ve diyet/egzersiz analizlerinde genel popülasyon çeşitliliğini koruduğunu ortaya koymaktadır.

2.3.3 Exercise Schedule Dağılımı

Egzersiz planları arasında şu rutinler yer almaktadır:

- *Light weightlifting, Yoga, and 2000 steps walking*
- *Moderate cardio, Strength training, and 5000 steps walking*
- *HIIT, Cardio, and 8000 steps walking*

- *Low-impact cardio, Swimming, and 10000 steps walking*

Tüm plan türlerinin yaklaşık eşit frekansta seçildiği görülmektedir.
Bu durum kullanıcıların farklı kondisyon seviyelerine göre çeşitliliğe sahip olduklarını, ancak belirli bir planın baskın olmadığını göstermektedir.

2.3.4 Meal Plan Dağılımı

Dört temel beslenme planı incelenmiştir:

1. *High-calorie, protein-rich diet*
2. *Balanced diet with moderate protein and carbohydrates*
3. *Low-carb, high-fiber diet*
4. *Low-calorie, nutrient-dense diet*

Her bir plan benzer sıklıkta gözlemlenmiştir; bu durum modelin farklı beslenme türlerini dengeli biçimde temsil ettiğini göstermektedir.

2.3.5 Cinsiyet (Gender) Dağılımı ve Hedef İlişkisi

Erkek ve kadın katılımcı oranları birbirine çok yakın olup cinsiyet dağılımı dengelidir.
Her iki cinsiyet grubunda da kas kazanımı ve yağ yakımı hedefleri benzer oranlarda görülmektedir.
Bu bulgu, hedef belirleme sürecinde cinsiyetin belirleyici bir faktör olmadığını düşündürmektedir.

2.3.6 Gender vs BMI Category

Cinsiyet ve BMI kategorileri arasındaki ilişkiyi gösteren ısı haritasına göre:

- *Kadınlar arasında Underweight (zayıf) kategorisi oranı biraz daha yüksektir.*
- *Erkeklerde ise Overweight (fazla kilolu) kategorisi biraz daha baskındır.*
Genel olarak farklar küçük olup, cinsiyetin BMI üzerinde marjinal bir etkisi vardır.

2.3.7 BMI Category vs Goal

Bu analizde BMI kategorilerine göre hedef dağılımı incelenmiştir.
Sonuçlar, BMI değeri düşük olan bireylerin genellikle muscle gain (kas kazanımı),
BMI değeri yüksek olan bireylerin ise fat burn (yağ yakımı) hedefi belirlediklerini göstermektedir.
Bu ilişki, kullanıcıların hedef belirleme süreçlerinin fizyolojik gerçeklikle uyumlu olduğunu doğrulamaktadır.

2.3.8 Goal vs Exercise Schedule

Hedefler ve egzersiz planları arasındaki ısı haritası, hedefe göre tercih edilen antrenman türlerini göstermektedir:

- *Fat burn hedefi olan kullanıcılar, daha çok light weightlifting ve cardio ağırlıklı programları tercih etmektedir.*

- *Muscle gain hedefi olan kullanıcılar ise HIIT ve weightlifting odaklı rutinleri daha sık seçmektedir.*

Bu bulgu, kullanıcıların hedeflerine uygun egzersiz planlarını seçme konusunda bilinçli olduklarını göstermektedir.Ayrıca yine set yine çok benzer dağılmıştır.

2.3.9 BMI Category vs Meal Plan

BMI kategorileri ile meal plan türleri arasındaki ilişki oldukça belirgindir:

- *Underweight bireyler → High-calorie, protein-rich diet*
- *Normal weight bireyler → Balanced diet*
- *Overweight bireyler → Low-carb, high-fiber diet*
- *Obese bireyler → Low-calorie, nutrient-dense diet*

Tüm kategorilerde ilişki %100'e yakın bir eşleşme göstermektedir.
Bu durum, meal plan önerilerinin BMI durumuna göre otomatik ve tutarlı şekilde eşleştirildiğini göstermektedir.

2.3.10 Kategorik Değişkenler Arası Korelasyon

Son olarak, Gender, Goal, BMI Category, Exercise Schedule ve Meal Plan değişkenleri arasındaki korelasyon matrisi incelenmiştir.
Sonuçlar:

- *BMI Category – Exercise Schedule arasında negatif yüksek korelasyon (-0.80),*
- *BMI Category – Meal Plan arasında pozitif korelasyon (0.39) gözlemlenmiştir.*

Bu da, egzersiz sıklığı arttıkça BMI değerinin azalma eğiliminde olduğunu ve meal plan türünün BMI ile doğrudan ilişkili olduğunu göstermektedir.
Diğer değişkenler arasında anlamlı bir korelasyon bulunmamıştır.

2.3 Genel Değerlendirme

Bu analiz, Meal-Ex Planner veri setinin hedef, beslenme, egzersiz ve vücut kompozisyonu arasındaki ilişkileri açık biçimde ortaya koymuştur.
Elde edilen bulgular, sistemin kişiselleştirilmiş öneri yapısının tutarlı olduğunu göstermekte ve makine öğrenimi aşamasında kullanılacak etiketleme ve öznitelik seçimlerinin doğruluğunu desteklemektedir.

2.4 Healthy Meal Plans Dataset

2.4.1 Temel Besin Değerleri ve Dağılımlar

Bu bölümde veri setinde yer alan sayısal değişkenlerin (kalori, yağ, karbonhidrat, protein, hazırlık süresi, malzeme sayısı) genel dağılımı incelenmiştir.

Boxplot Analizleri

- *Değişkenlerin hiçbirinde belirgin uç değer (outlier) bulunmamaktadır.*
- *Tüm değişkenler 0–1 aralığına normalize edilmiştir, bu da verilerin ölçeklenmiş olduğunu gösterir.*
- *Medyan değerlerin genellikle 0.5 civarında olması, verilerin dengeli bir şekilde dağıldığını ve aşırı sağa/sola kaymanın olmadığını göstermektedir.*

Histogram Analizleri

Histogramlarda tüm değişkenler neredeyse uniform (düzgün) dağılım göstermektedir:

- *Bu durum, her aralıkta benzer sayıda gözlem bulunduğunu ve verinin dengeli olduğunu ortaya koymaktadır.*
- *Özellikle num_ingredients (malzeme sayısı) değişkeninde küçük varyasyonlar olsa da genel dağılım homojendir.*

Sonuç:
Veri seti temiz, normalize edilmiş ve dengelidir. Bu yapı, sonraki istatistiksel analizlerin ve modellemelerin güvenilir sonuçlar üretmesini sağlar.

2.4.2 Sağlıklı / Sağlıksız Yemek Karşılaştırması ve Korelasyon Analizi

Bu bölümde "is_healthy" değişkenine göre veri seti iki sınıfa ayrılmış ve karşılaştırmalar yapılmıştır.

Sınıf Dağılımı

*Grafikten açıkça görülmektedir ki **sağlıksız yemekler (0)**, **sağlıklı yemeklere (1)** göre oldukça fazladır.*

*Bu durum bir **class imbalance (sınıf dengesizliği)** problemidir ve model eğitimi sırasında SMOTE veya class_weight gibi yöntemlerle dengelenmelidir.*

Değişken Bazlı Karşılaştırmalar

Boxplot analizleri sonucunda:

- *Kalori: Sağlıksız yemeklerin kalori değeri belirgin şekilde yüksektir.*
- *Protein: Sağlıksız yemeklerde genellikle protein oranı daha fazladır.*
- *Hazırlık Süresi: Sağlıklı ve sağlıksız yemekler arasında belirgin fark yoktur.*
- *Malzeme Sayısı: Sağlıksız yemekler genellikle daha fazla malzeme içermektedir, bu da daha karmaşık tariflerin çoğunlukla sağlıksız olduğunu göstermektedir.*

Korelasyon Analizi

Sürekli Değişkenler + is_healthy

- *is_healthy ile fat (yağ) arasında zayıf negatif korelasyon (-0.19) vardır.*

- *Diğer değişkenler (kalori, protein, carbs, prep_time) ile güçlü bir ilişki yoktur (<0.1).*

  Bu durum, sağlıklılık etiketinin tek bir faktör yerine çoklu faktörlerin birleşimiyle belirlendiğini gösterir.

  Diyet Tipleri Arası Korelasyon

- *Keto, Paleo ve Gluten-Free diyetleri arasında yüksek pozitif korelasyonlar (0.72–0.78) vardır.*
- *Vegan ve Vegetarian diyetleri arasında da orta düzeyde pozitif ilişki (0.61) görülmektedir.*
- *Vegan/Vegetarian diyetleri ile Keto/Paleo arasında ise negatif ilişki (-0.5 ila -0.7), yani iki grup birbirine zıttır.*
- *Mediterranean (Akdeniz) diyeti diğerlerinden bağımsız bir profil sergilemektedir.*

Diyet Etiketi Frekansları

- *Vegetarian ve Gluten-Free tarifler en yüksek sayıda örneğe sahiptir (~280).*
- *Mediterranean ve Vegan tarifler ise daha az (~150 civarı).*

  Bu dağılım, veri setinde bitkisel ve glutensiz tariflerin baskın olduğunu göstermektedir.

  Sonuç:

- *Sağlıksız yemekler yüksek kalori, yağ ve protein içerme eğilimindedir.*
- *Yağ miktarı, sağlıklılık durumunu belirlemede en güçlü negatif göstergedir.*
- *Diyet türleri arasında anlamlı kümelenmeler mevcuttur (ör. Keto–Paleo–GlutenFree).*

2.4.3 Makrobesin Farklılıkları ve Diyet Türlerine Göre Sağlıklılık Oranları

Bu bölümde “is_healthy” etiketiyle birlikte makrobesin (fat, carbs) değişkenlerinin etkisi ve diyet bazlı sağlık oranları analiz edilmiştir.

Makrobesin Bazında Karşılaştırma

Yağ (Fat)

- *Sağlıksız yemeklerde medyan yağ değeri ~0.55, sağlıklı yemeklerde ~0.35’tir.*
- *Yağ oranı arttıkça sağlıklı olma olasılığı azalır.*

Karbonhidrat (Carbs)

- *Sağlıksız yemeklerde medyan karbonhidrat değeri ~0.55, sağlıklı yemeklerde ~0.4’tür.*
- *Yüksek karbonhidrat içeren tariflerin genellikle sağlıksız olarak sınıflandırıldığı görülmektedir.*

Diyet Türlerine Göre Sağlıklılık Oranı

Grafikte diyet etiketleri bazında “flag = 1” iken sağlıklı olma oranları gösterilmiştir:

- *Vegan ve Gluten-Free tarifler en yüksek sağlıklılık oranına (~%10) sahiptir.*
- *Vegetarian ve Paleo tariflerde oran %9 civarındadır.*
- *Keto tariflerinde bu oran biraz düşüktür (~%8.5).*
- *Mediterranean (Akdeniz) tarifler en düşük orana sahiptir (~%6.5).*

Sonuç:

- *Sağlıklı yemeklerin genellikle düşük yağ ve karbonhidrat içerdiği,*
- *Vegan ve Gluten-Free diyetlerin sağlıklı olarak etiketlenme olasılığının daha yüksek olduğu,*
- *Mediterranean diyet örneklerinin ise nispeten daha az sağlıklı sınıfta yer aldığı belirlenmiştir.*

2.4 Genel Sonuç ve Yorum

Healthy Meal Plans veri setine ilişkin genel bulgular şunlardır:

a. *Veri seti temiz, normalize edilmiş ve dengeli yapıda.*
b. *“Sağlıklı” ve “sağlıksız” sınıfları arasında belirgin dengesizlik vardır.*
c. *Yağ ve karbonhidrat miktarları sağlıklılık durumunu en fazla etkileyen değişkenlerdir.*
d. *Vegan, Gluten-Free ve Vegetarian diyetleri sağlık açısından pozitif yönde öne çıkmaktadır.*
e. *Keto ve Paleo beslenme biçimleri birbiriyle ilişkili ve orta düzeyde sağlıklı; Mediterranean ise bağımsız ancak nispeten düşük orandadır.*

Genel Değerlendirme:
Veri seti, farklı diyet türlerinin besin değerleriyle olan ilişkilerini anlamak için güçlü bir temel sunmaktadır. Makrobesin dağılımları, kalori yoğunluğu ve diyet etiketi kombinasyonları, gelecekte beslenme öneri sistemleri ve kişisel diyet planlama modelleri için kullanılabilir niteliktedir.

3. List of Decided Models and Reasoning Behind

3.1 Regresyon Modelleri (Tahminleme)

Kullanıcının yaş, kilo, boy, cinsiyet, aktivite düzeyi gibi özelliklerinden yola çıkarak günlük kalori ihtiyacını ve makrobesin oranlarını tahmin etmek için kullanılacaktır.

- ***Linear Regression:** Basit ve yorumlanabilir bir temel model sağlar.*
- ***Random Forest Regressor:** Doğrusal olmayan ilişkileri yakalar, gürültüye dayanıklıdır.*
- ***XGBoost Regressor:** Karmaşık etkileşimleri otomatik öğrenir, yüksek performans sağlar.*

Bu modeller, diyet hedeflerine uygun kalori ve makrobesin oranlarını üretmek için temel tahmin motorunu oluşturur.

3.2 Sınıflandırma Modelleri (Etiketleme)

Kullanıcının fiziksel durumu (ör. BMI, aktivite düzeyi, hedef) ve yemek içerikleri üzerinden "sağlıklı/sağlıksız" tarif ayrımı veya "kas yapma/yağ yakma/dengeli" gibi hedef sınıfları tahmin etmek için kullanılacaktır.

- ***Logistic Regression:*** *Basit ve açıklanabilir bir referans modeldir.*
- ***Support Vector Machine (SVM):*** *Karmaşık sınıf sınırlarını belirlemede etkilidir.*
- ***Gradient Boosting / XGBoost Classifier:*** *Ensemble yapısıyla yüksek doğruluk sağlar.*

Bu modeller, tahmin edilen makrobesin oranlarına uygun olarak kullanıcıya önerilecek diyet kategorilerini belirlemede rol oynar.

3.3 Model Seçim Süreci

Belirtilen tüm modeller proje kapsamında test edilip karşılaştırılacaktır.
Her modelin performansı uygun metriklerle (regresyon için RMSE–MAE, sınıflandırma için Accuracy–F1–AUC) ölçülecek,
ve en iyi sonuç veren modeller sistemin nihai versiyonuna entegre edilecektir.

Performans değerlendirmesinde yalnızca doğruluk değil, yorumlanabilirlik, veriyle uyum ve genelleme kabiliyeti de dikkate alınacaktır.
Bu süreç sonucunda yalnızca işlevsel ve istikrarlı sonuç veren modeller proje sisteminde aktif olarak kullanılacaktır.

4.Paper Section

4.1 Abstract (Özet)

Bu proje, bireylerin fiziksel özellikleri ve yaşam tarzı verilerinden yola çıkarak kişiselleştirilmiş beslenme önerileri sunan bir yapay zekâ tabanlı sistem geliştirmeyi amaçlamaktadır. Sistem, kullanıcıdan alınan temel bilgiler (yaş, cinsiyet, boy, kilo, aktivite seviyesi, hedef) aracılığıyla günlük enerji ve makrobesin ihtiyaçlarını tahmin eden modellerden oluşur. Bu tahminler, sistemin ilerleyen aşamalarında kişisel beslenme planlarının oluşturulmasında kullanılacaktır. Fitol, veri analizi ve makine öğrenmesini bir araya getirerek kullanıcıların sağlık hedeflerine uygun şekilde beslenme düzenlerini optimize etmeyi hedefleyen bir akıllı beslenme asistanı olarak tasarlanmaktadır.

4.2 Related Work (İlgili Çalışmalar)

Son yıllarda yapay zekâ destekli sağlık ve beslenme sistemleri üzerine yapılan çalışmalar, kişiselleştirilmiş diyet önerilerinin kullanıcı davranışları üzerindeki etkisini artırmıştır.

Örneğin "Personalized Nutrition Recommendation Using Machine Learning" (IEEE, 2020) çalışmasında, kullanıcı verilerinden besin ihtiyaçları tahmin edilerek bireysel planlama yapılmıştır.

Benzer şekilde "Hybrid AI Models for Health Recommendation" (Expert Systems, 2021) araştırmasında, kullanıcı profiline göre enerji ihtiyacı tahmin eden modellerin, beslenme yönetiminde verimliliği artırdığı rapor edilmiştir.

Bu proje, literatürdeki bu yaklaşımları temel alarak çoklu veri setlerini birleştiren, öğrenme tabanlı bir beslenme analiz sistemini uygulamayı hedeflemektedir.

4.3 Brief Planned Methodology (Planlanan Yöntem)

Fitol projesi, dört farklı veri setinden yararlanılarak oluşturulan çok katmanlı bir modelleme sürecine dayanmaktadır:

1. ***Food_and_Nutrition Dataset:*** *Kullanıcı profiline göre günlük kalori ve makrobesin oranlarını tahmin eden regresyon modelleri (Linear Regression, Random Forest, XGBoost).*
2. ***GYM Dataset:*** *Kullanıcının fiziksel hedefi ve aktivite düzeyine göre kategorik sınıflandırma yapan modeller (Logistic Regression, SVM, Gradient Boosting)(opsiyonel zaman içerisinde analiz edip faydalıysa kullanılacaktır.)*
3. ***Healthy_Meal_Plans Dataset:*** *Tariflerin besin profiline göre sağlıklı/sağlıksız olarak etiketlendiği sınıflandırma modelleri.*
4. ***Daily_Food_Nutrition Dataset:*** *Model doğrulama ve veri analizi için kullanılan destekleyici veri kümesi.*

Modellerin performansı doğruluk, hata oranı (MAE, RMSE) ve F1 gibi ölçütlerle değerlendirilecektir.

Elde edilen sonuçlara göre en başarılı modeller sonraki geliştirme aşamalarında kullanılacaktır.

4.4 Proposed Implementation

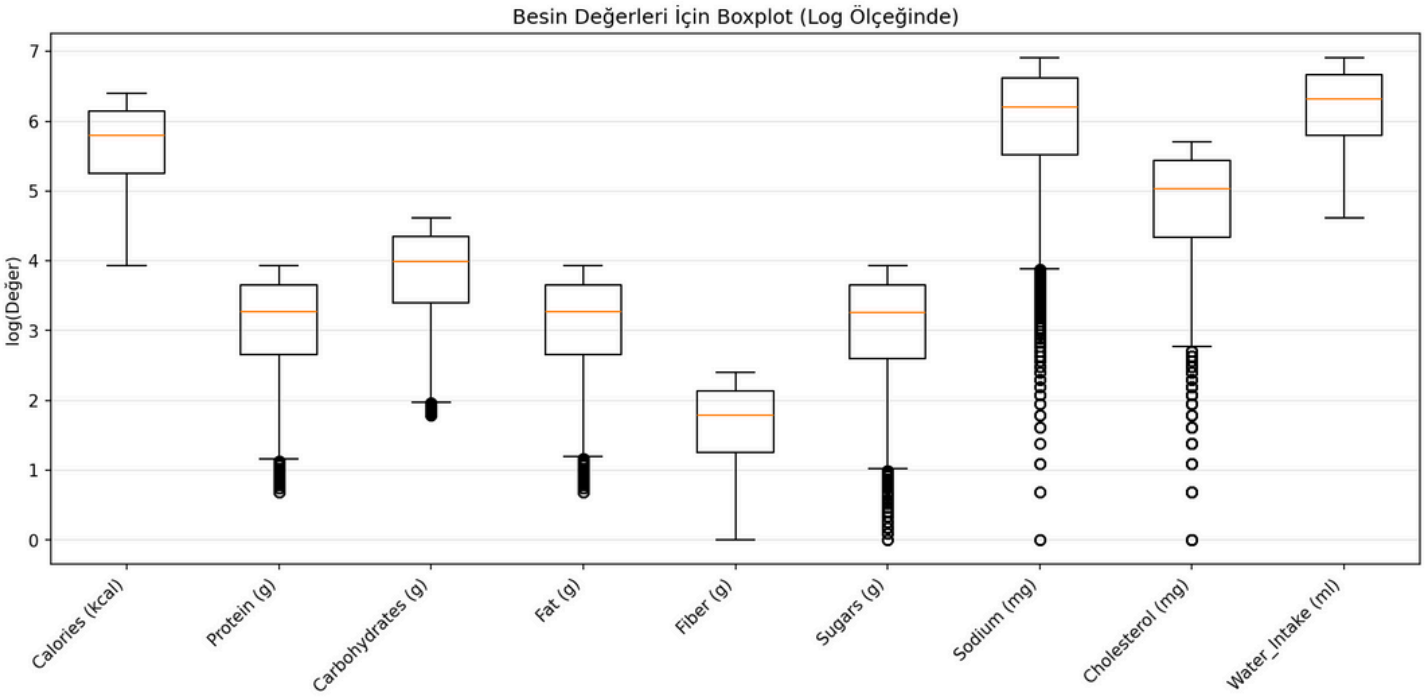
Proje uygulaması dört temel aşamada yürütülecektir:

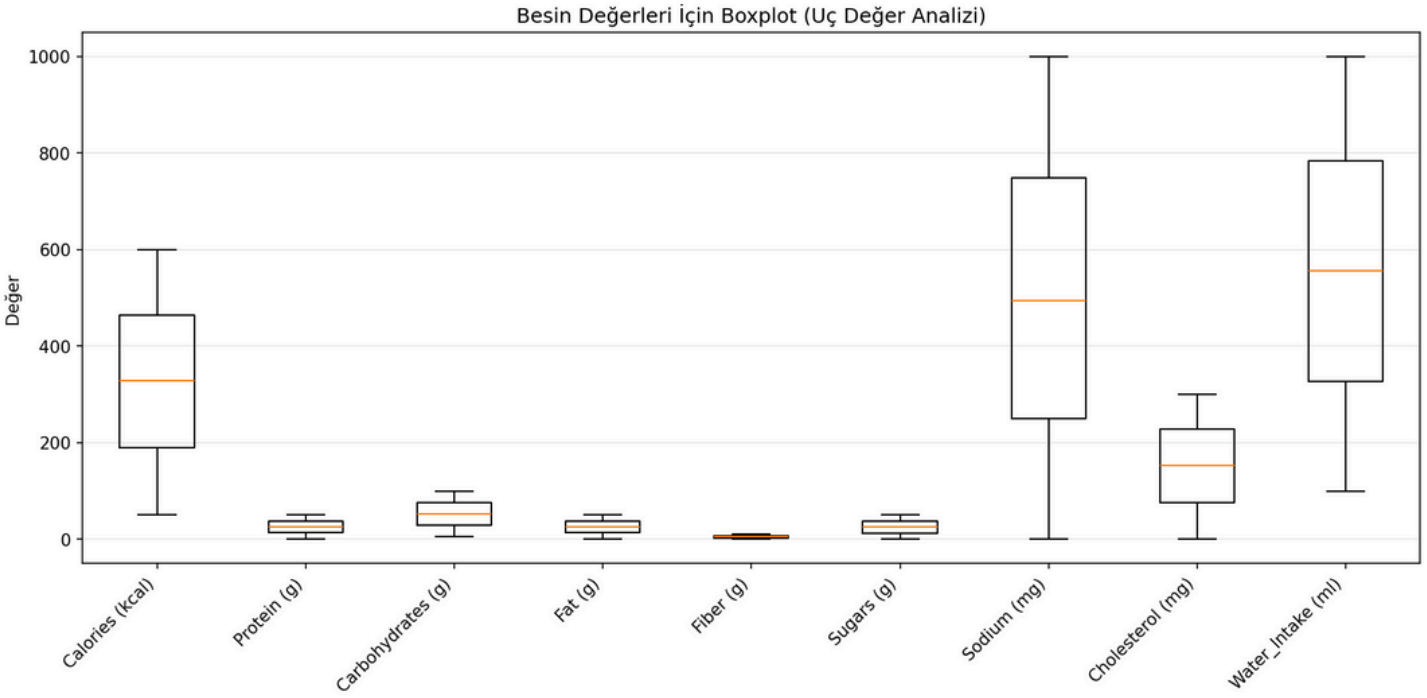
1. ***Veri Hazırlama:*** *Tüm veri setlerinin birleştirilmesi, temizlenmesi ve gerekli özniteliklerin (BMI, CalorieGap, MacroRatio vb.) üretilmesi.*
2. ***Model Eğitimi ve Testi:*** *Regresyon ve sınıflandırma modellerinin farklı kombinasyonlarının denenmesi ve performanslarının karşılaştırılması.*
3. ***Model Entegrasyonu:*** *En iyi sonuç veren modellerin sistemin çekirdeğine entegre edilmesi.*

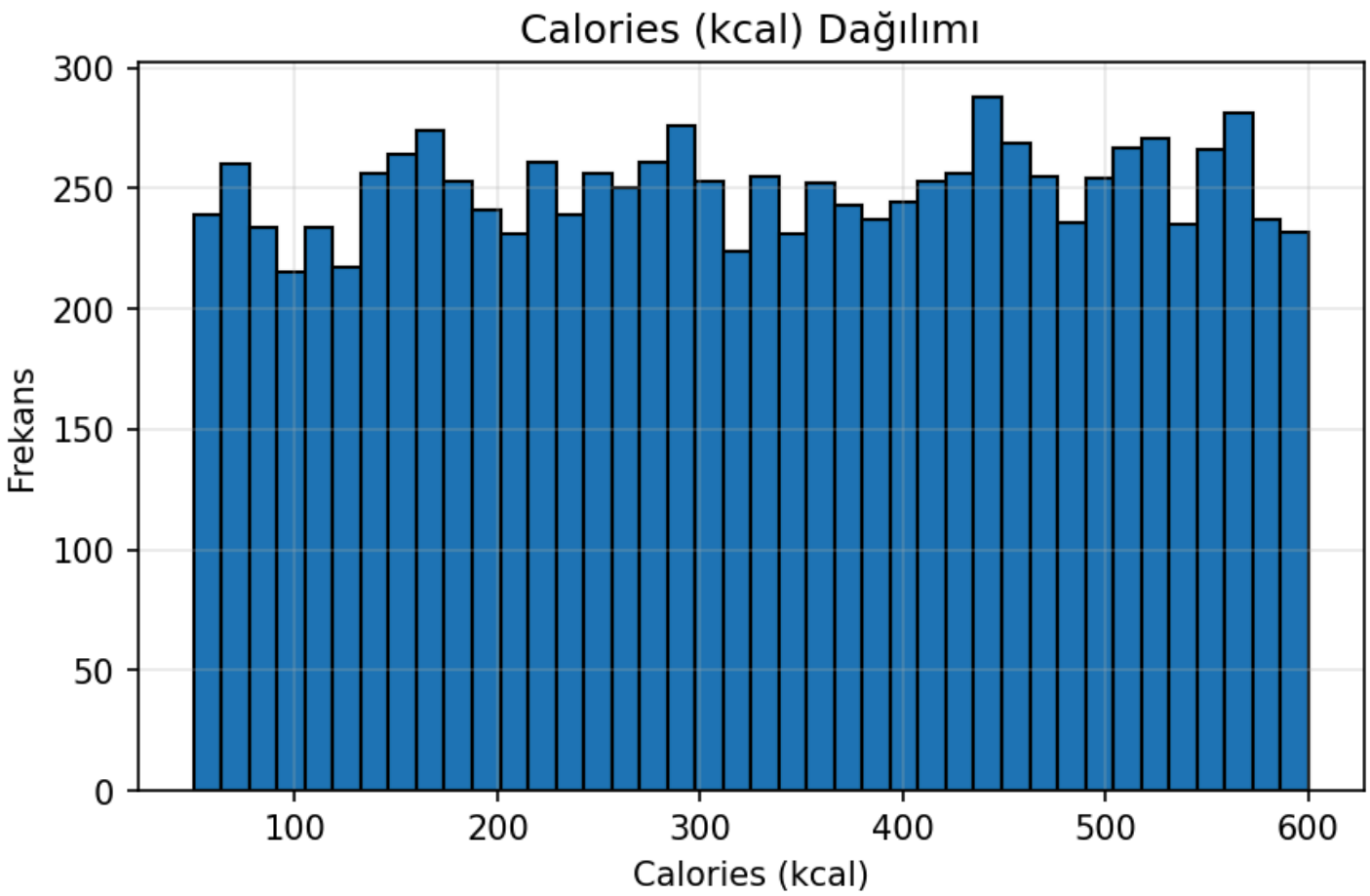
4. ***Sonuç Analizi ve Geliştirme:*** *Model çıktılarının incelenmesi, doğruluk değerlendirmesi ve ilerleyen süreçte öneri modülüne dönüştürülmesi.*

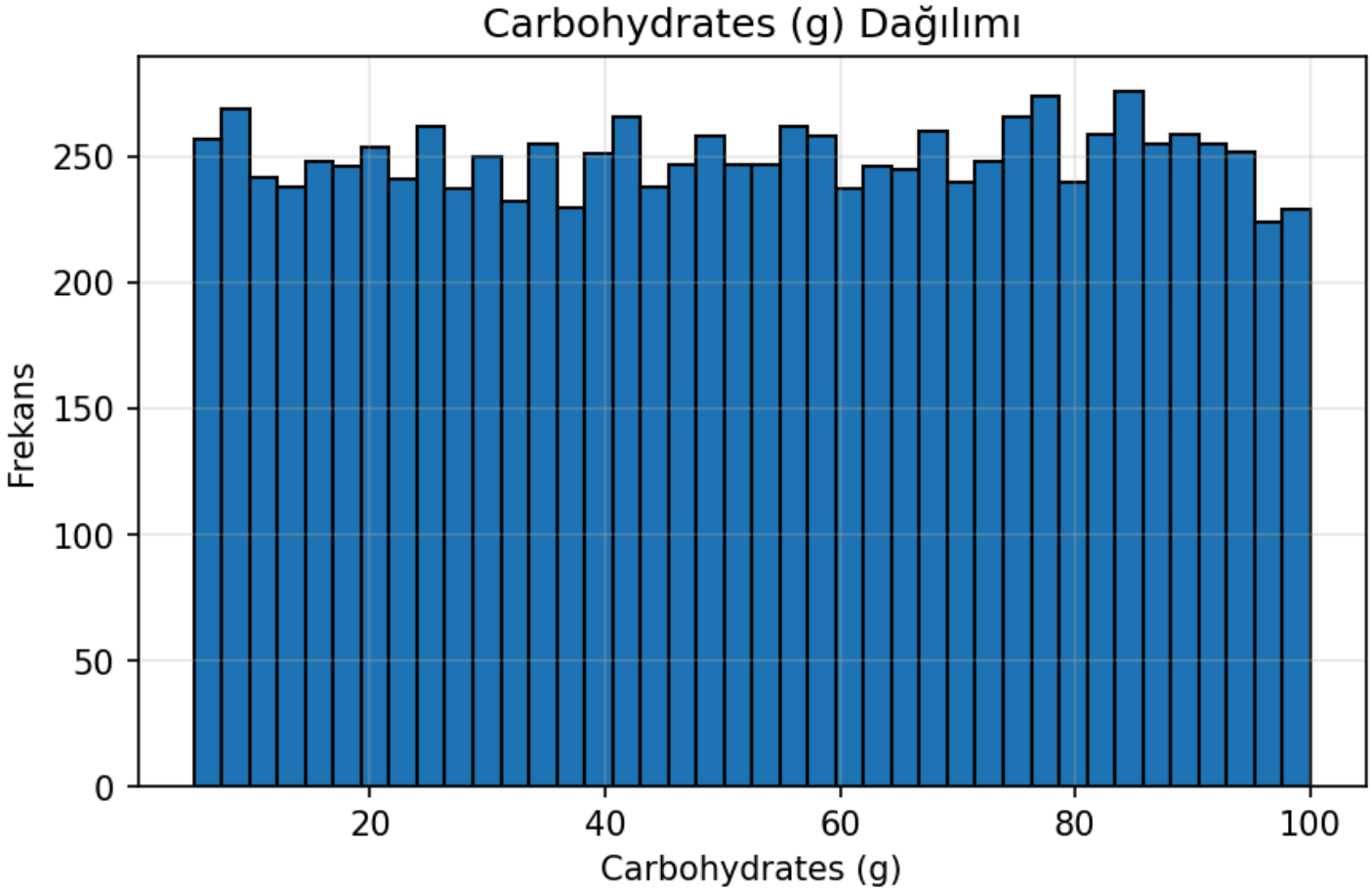
ANALİZ ÇIKTILARI

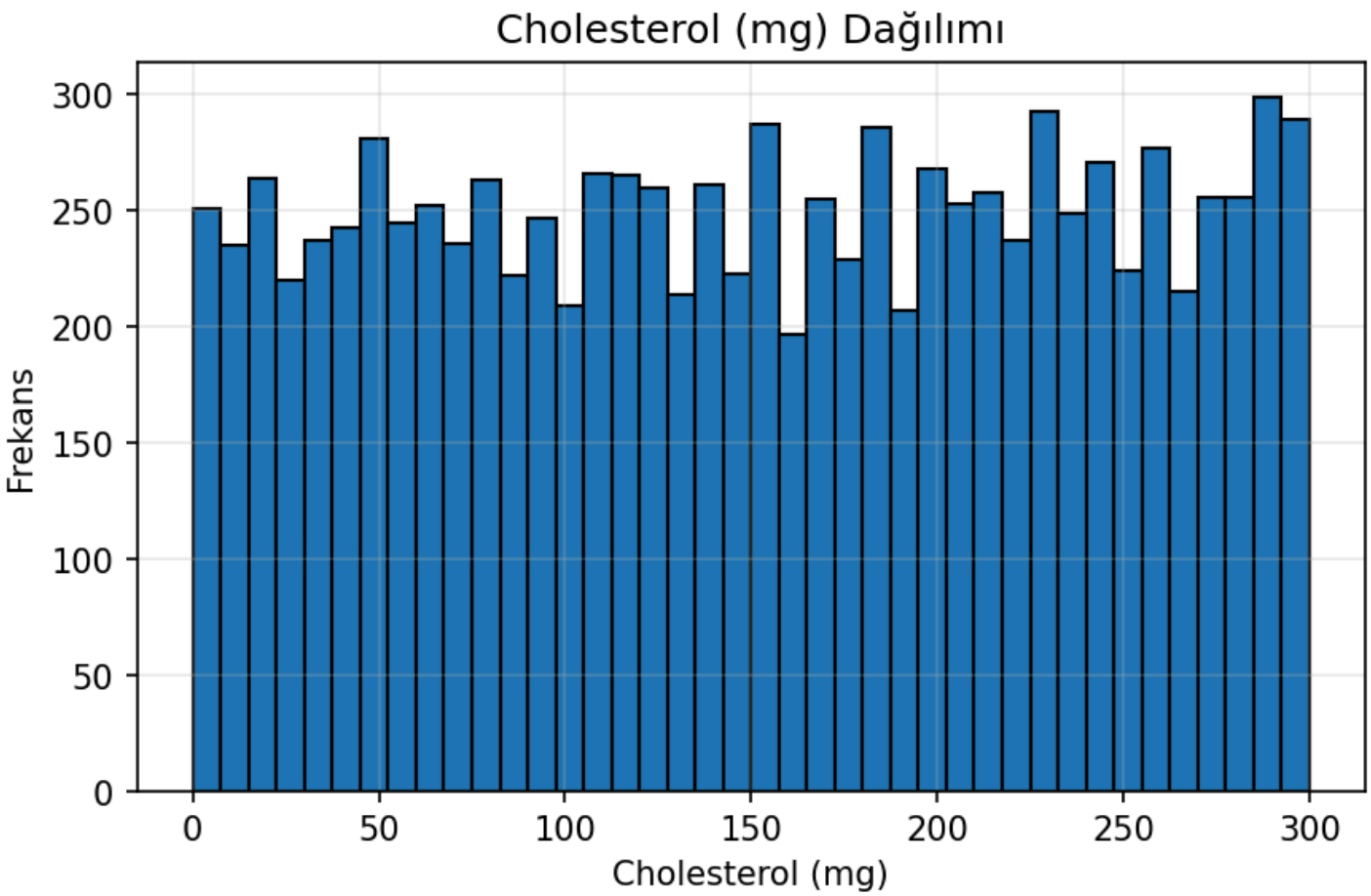
daily_food:

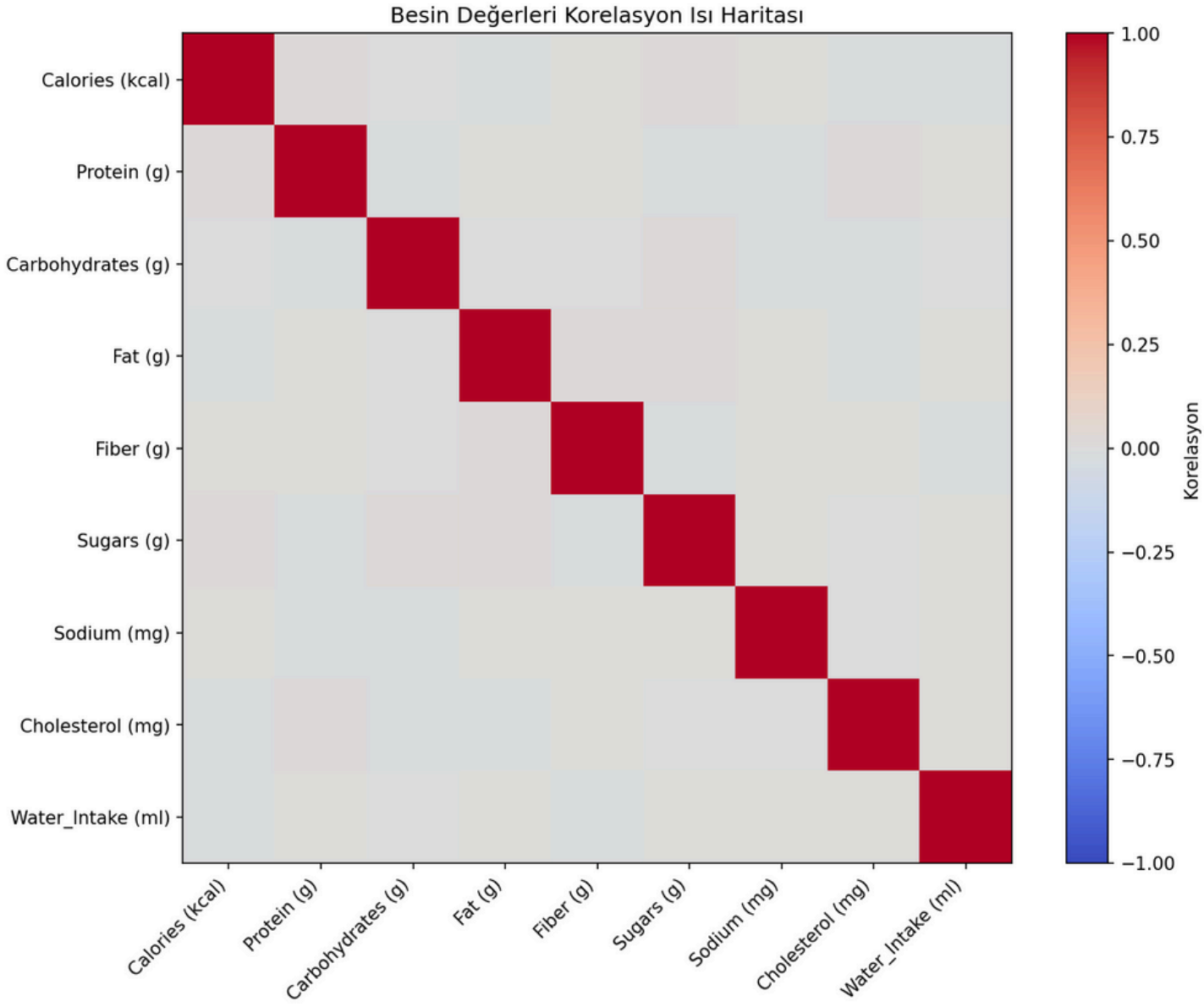


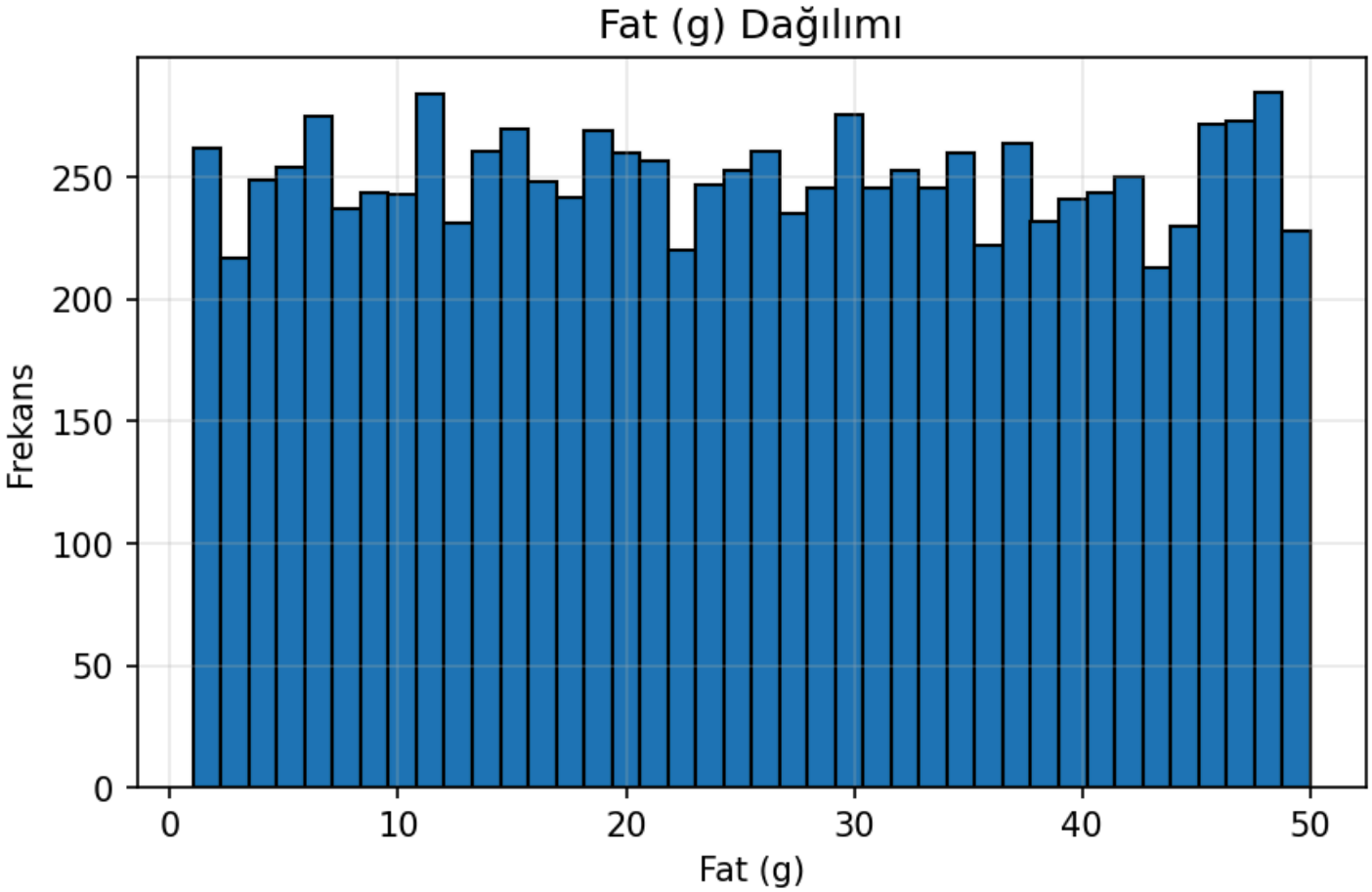


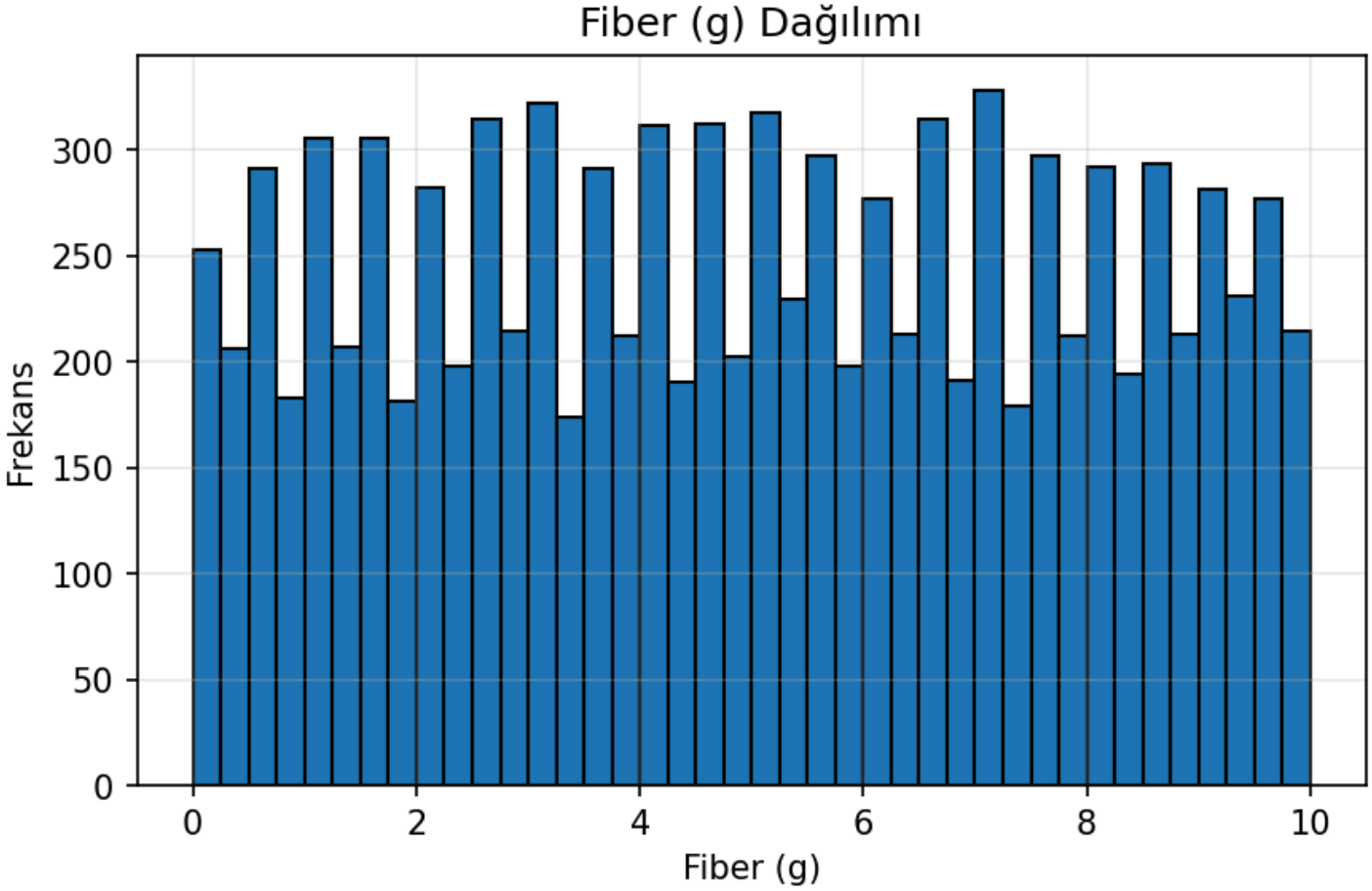

Calories (kcal) Dağılımı
Frekans
Calories (kcal)
0
50
100
150
200
250
300
100
200
300
400
500
600

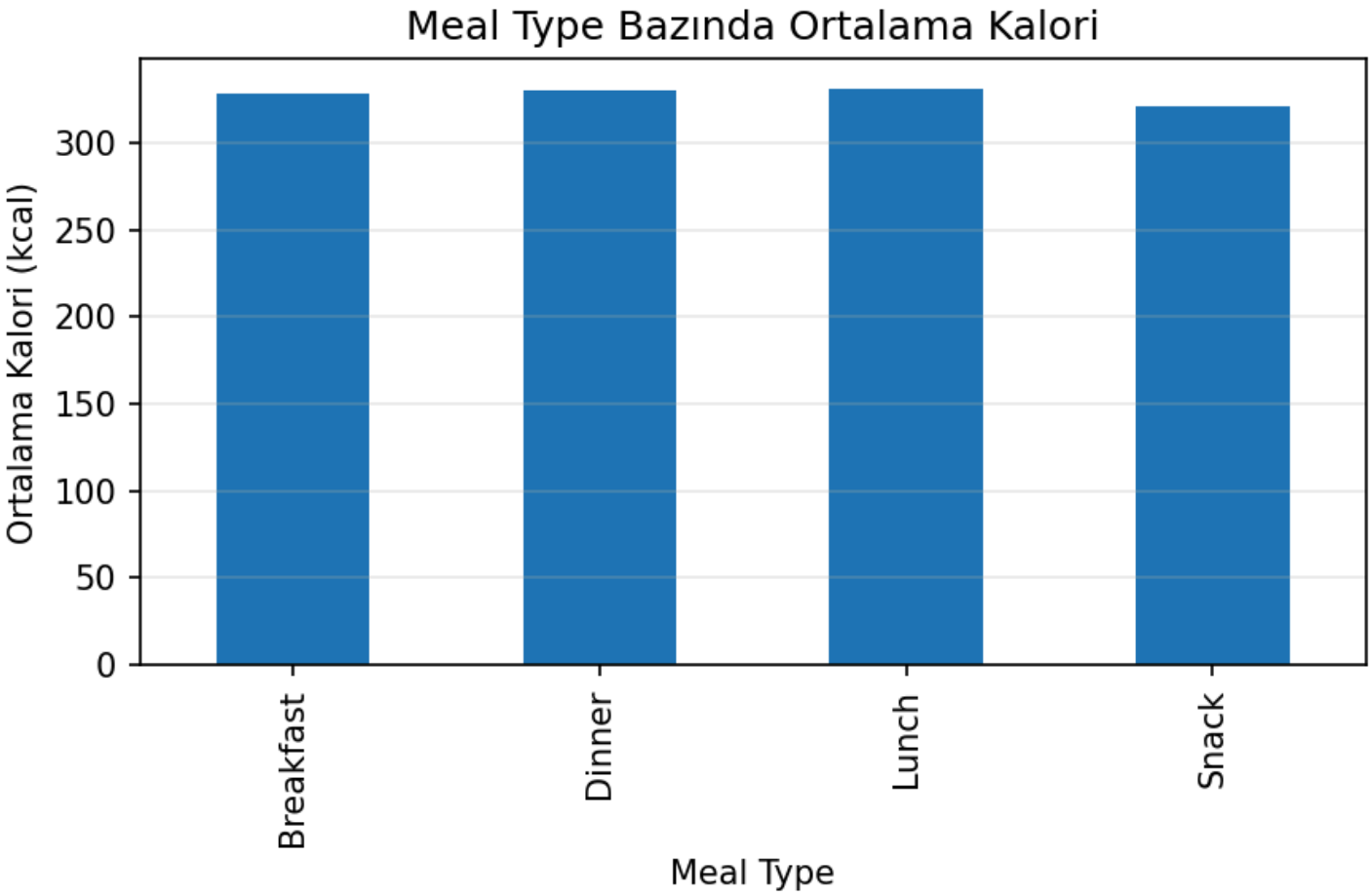

Carbohydrates (g) Dağılımı
Frekans
Carbohydrates (g)
0
50
100
150
200
250
20
40
60
80
100

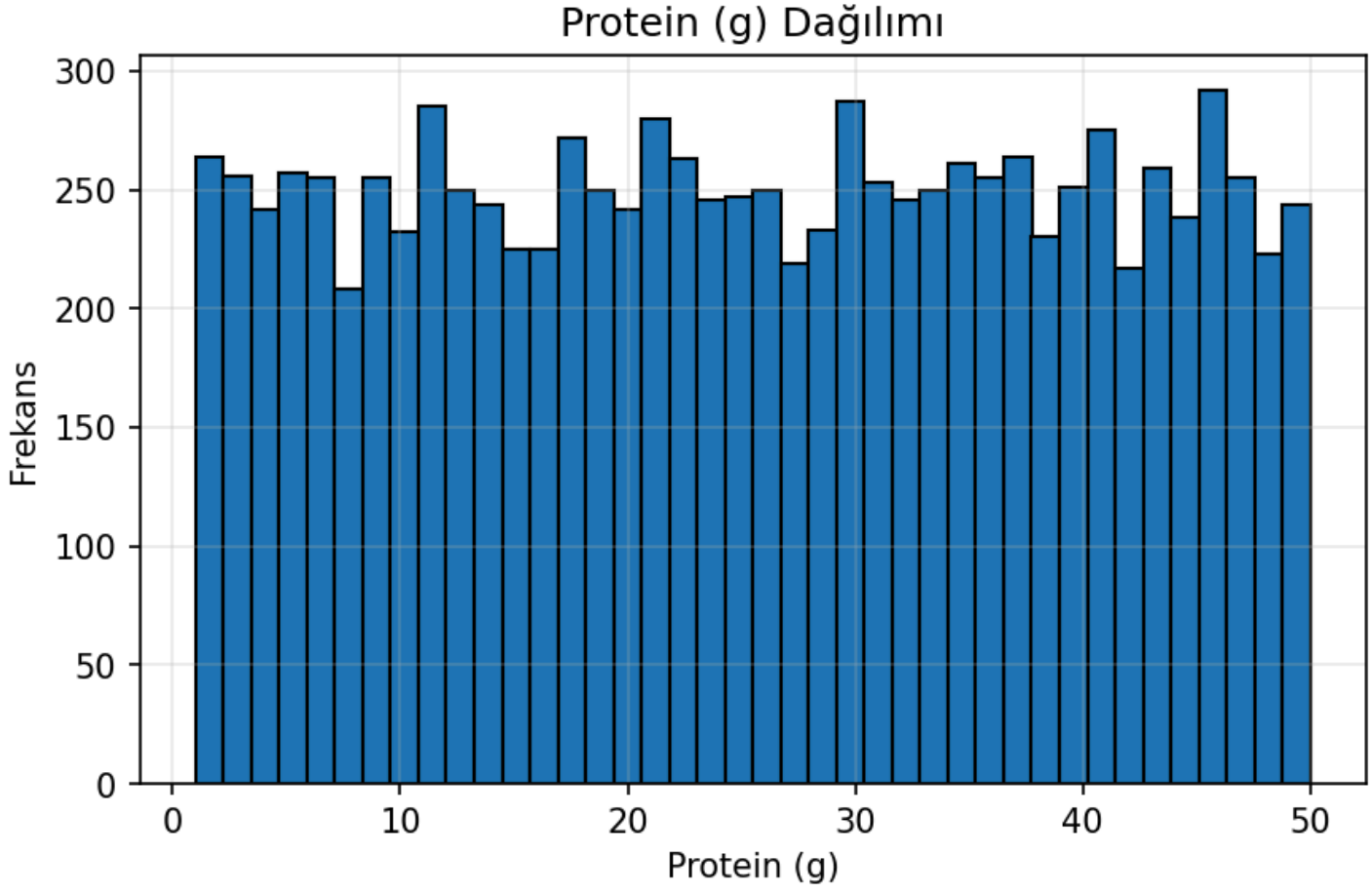

Cholesterol (mg) Dağılımı
Frekans
Cholesterol (mg)
0
50
100
150
200
250
300

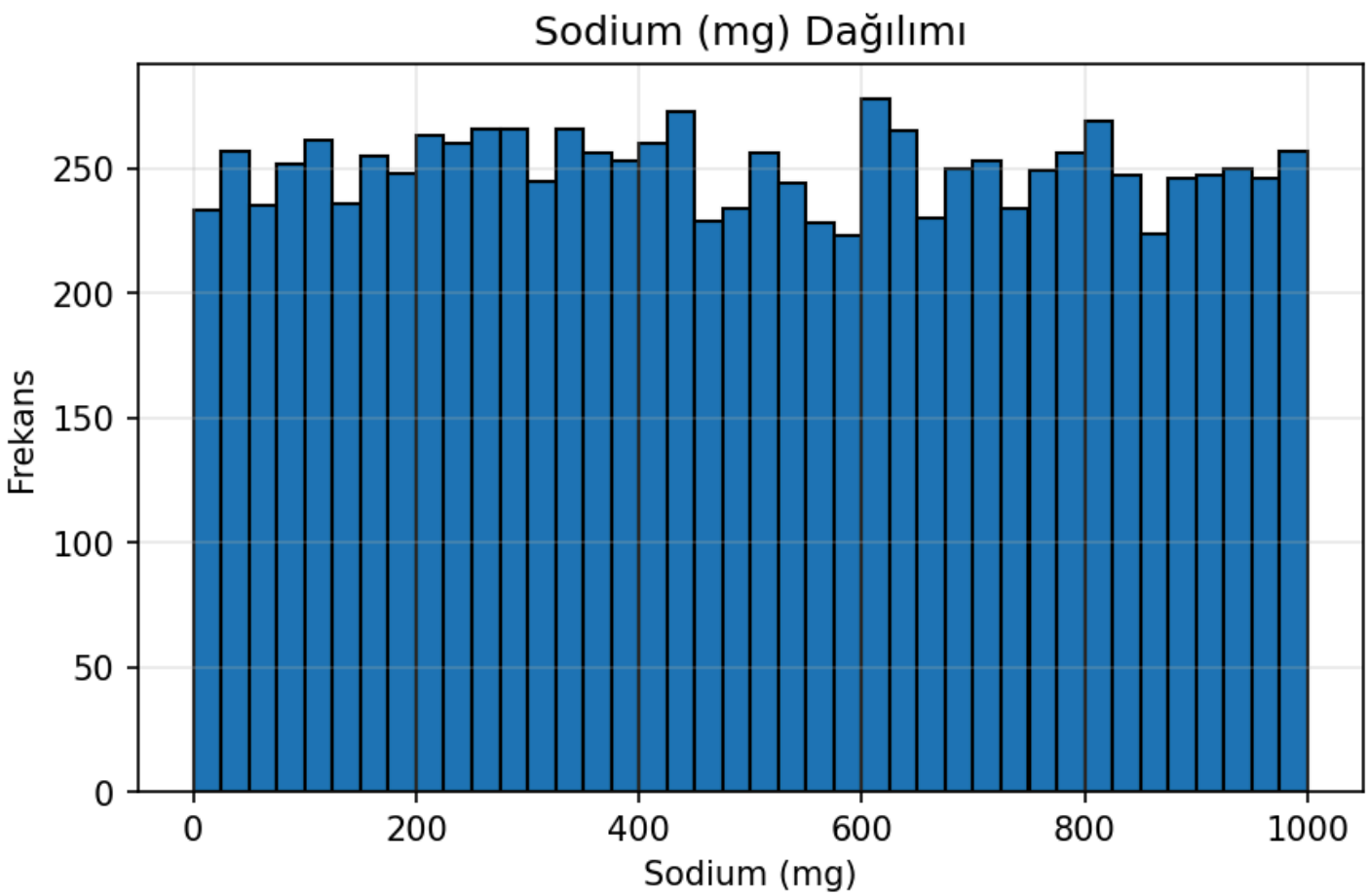

Besin Değerleri Korelasyon Isı Haritası
Calories (kcal)
Protein (g)
Carbohydrates (g)
Fat (g)
Fiber (g)
Sugars (g)
Sodium (mg)
Cholesterol (mg)
Water_Intake (ml)
Korelasyon
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

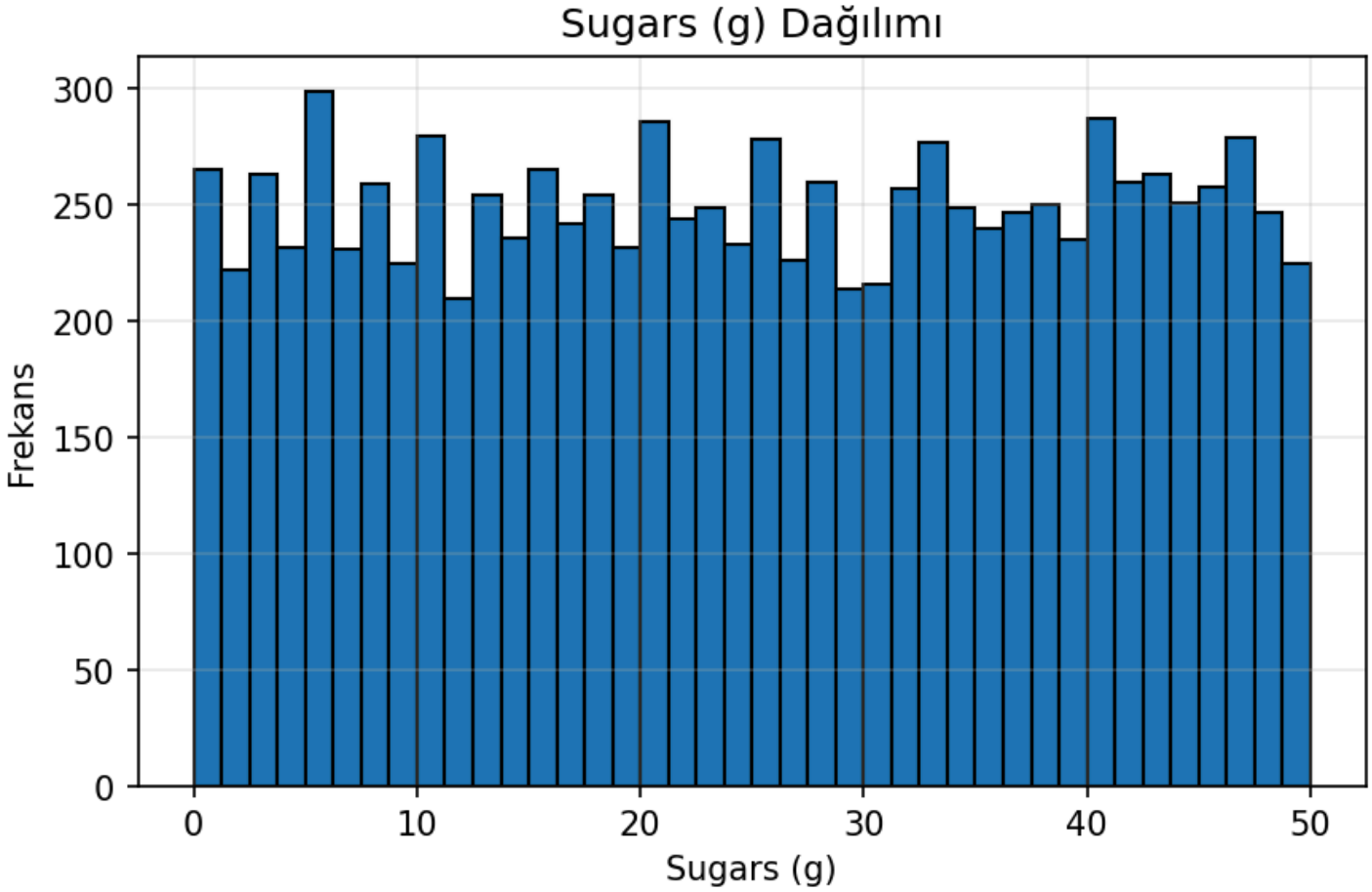

Fat (g) Dağılımı
Frekans
0
50
100
150
200
250
0
10
20
30
40
50
Fat (g)

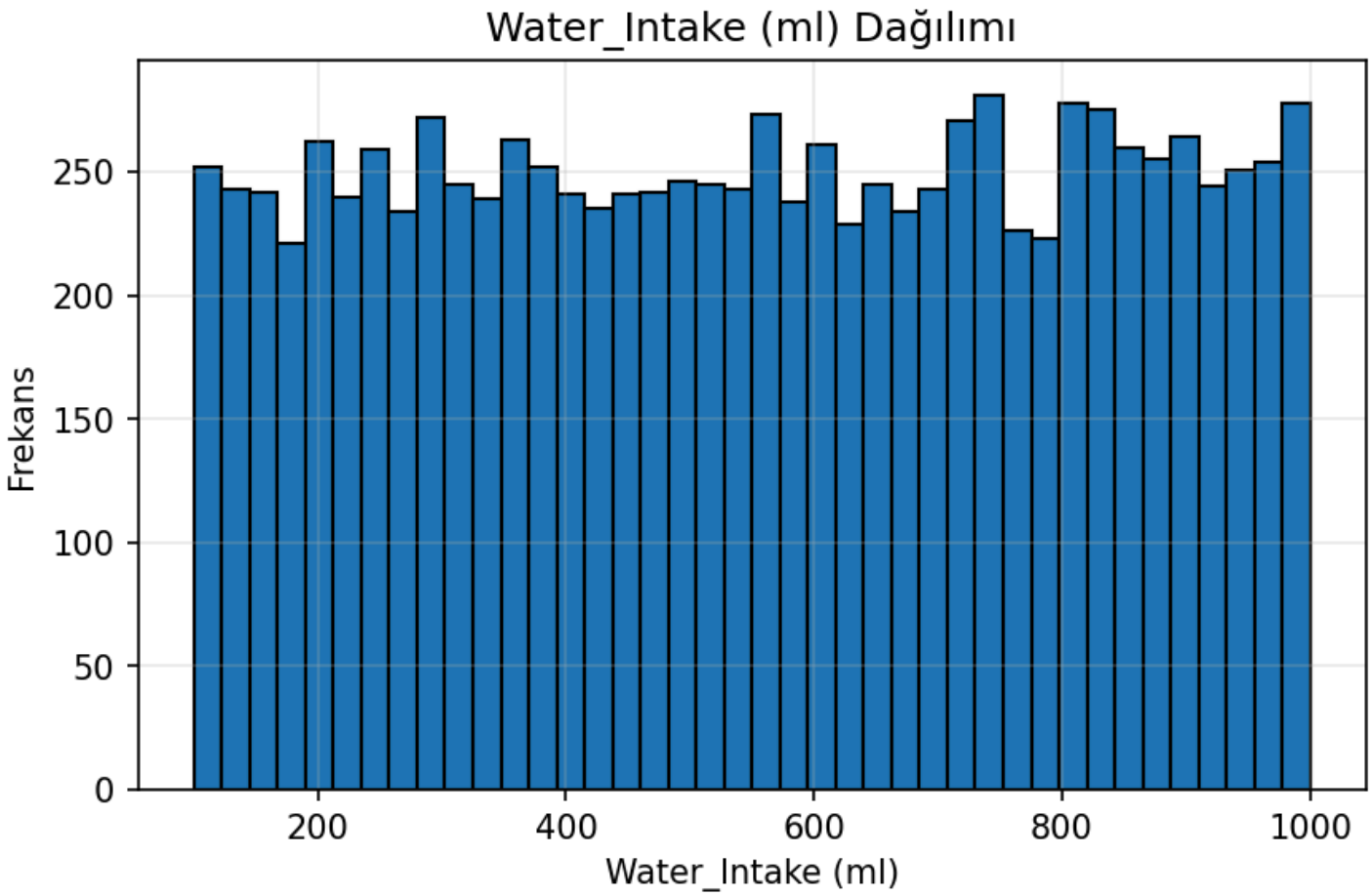

Fiber (g) Dağılımı
Frekans
0
50
100
150
200
250
300
0
2
4
6
8
10
Fiber (g)


Meal Type Bazında Ortalama Kalori
Ortalama Kalori (kcal)
0
50
100
150
200
250
300
Breakfast
Dinner
Lunch
Snack
Meal Type

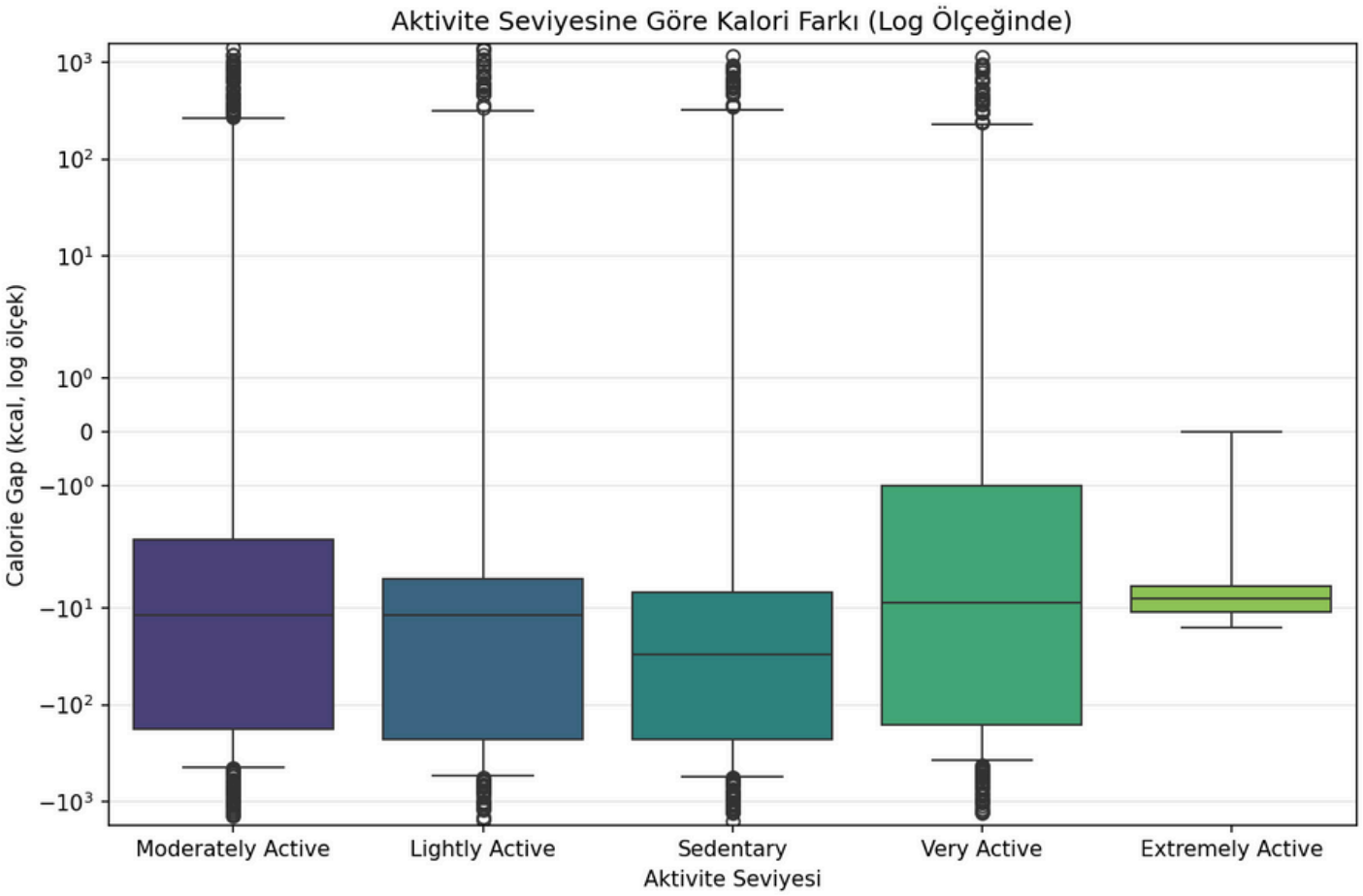

Protein (g) Dağılımı
Frekans
0
50
100
150
200
250
300
0
10
20
30
40
50
Protein (g)

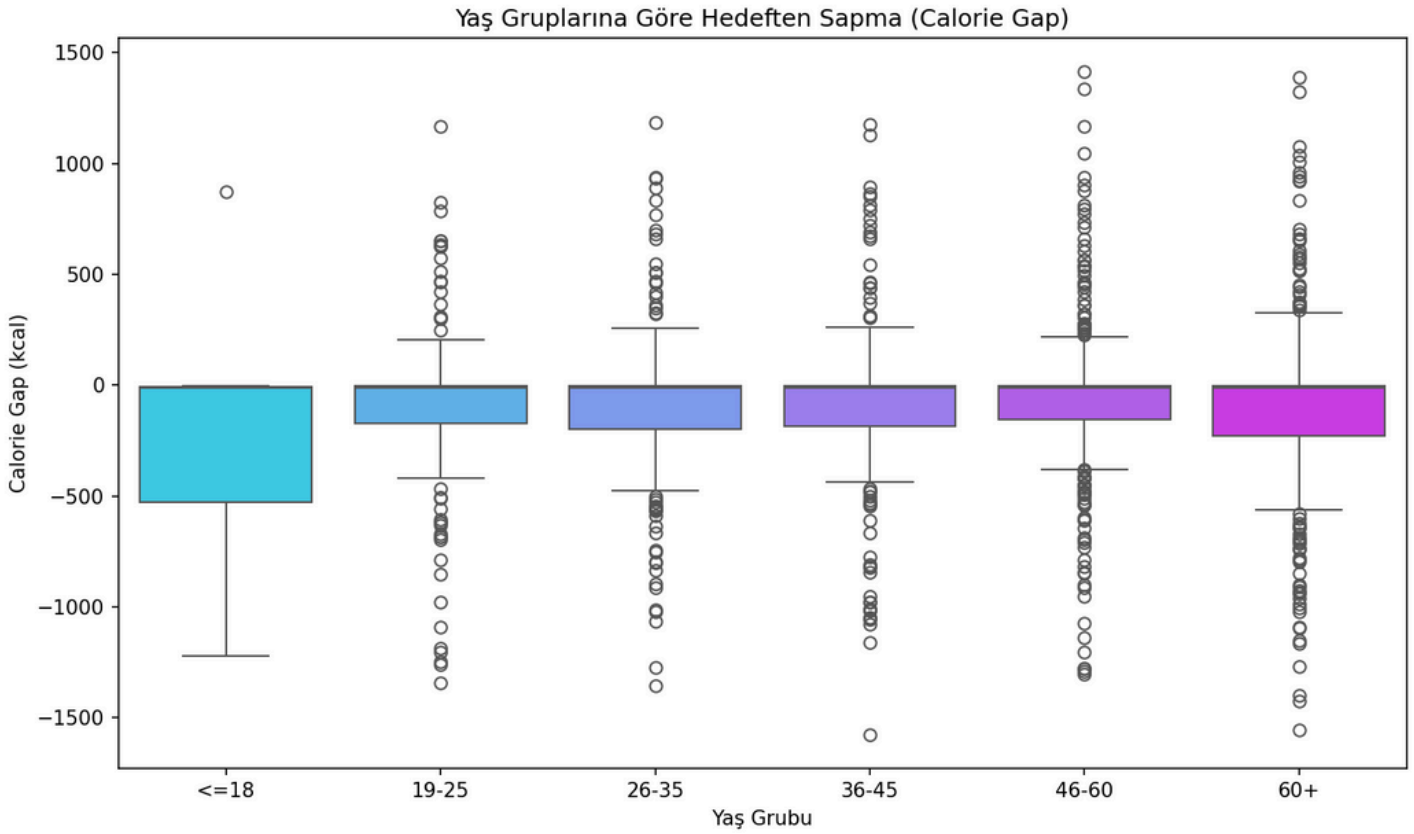

Sodium (mg) Dağılımı
Frekans
0
50
100
150
200
250
0
200
400
600
800
1000
Sodium (mg)

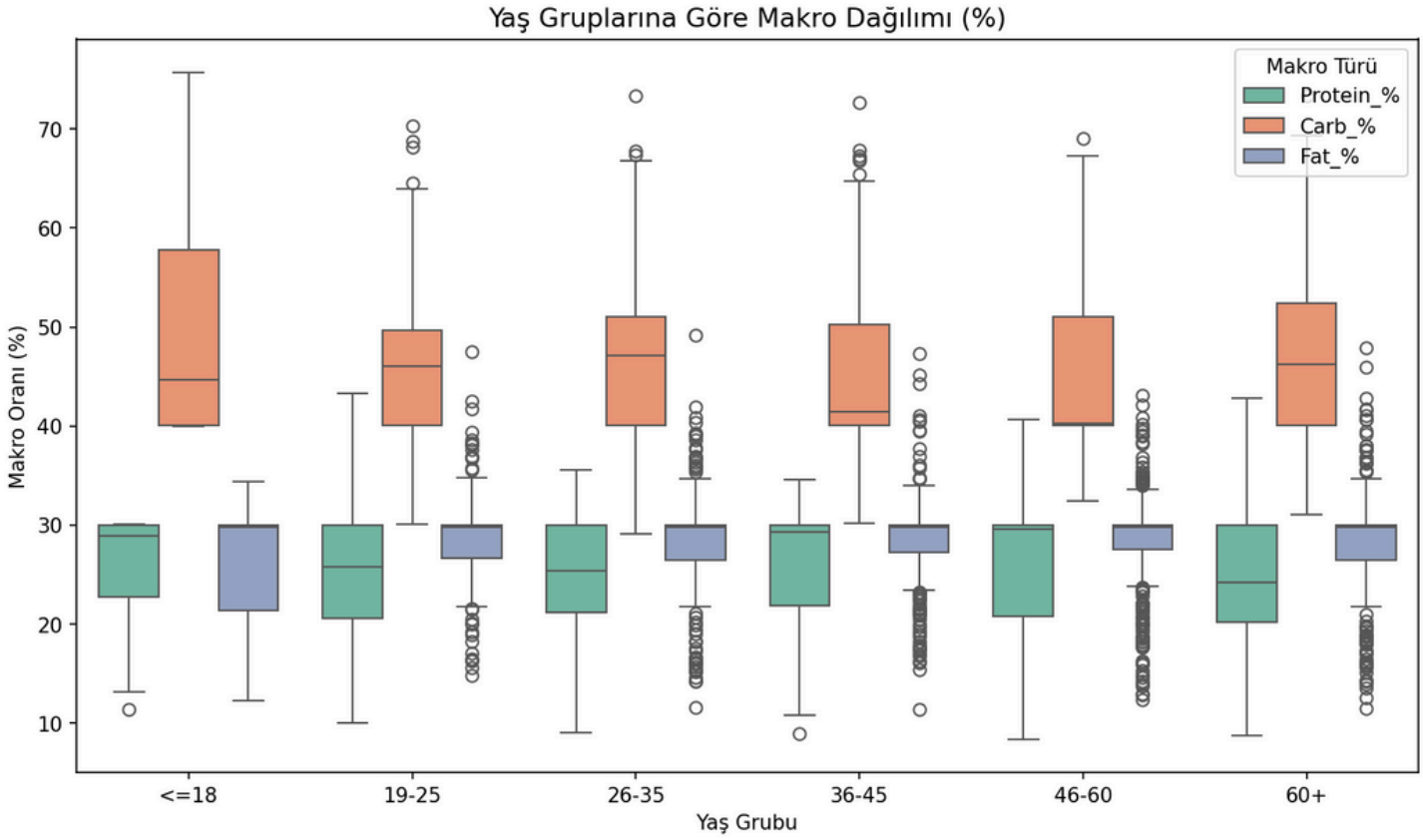

Sugars (g) Dağılımı
Frekans
0
50
100
150
200
250
300
0
10
20
30
40
50
Sugars (g)

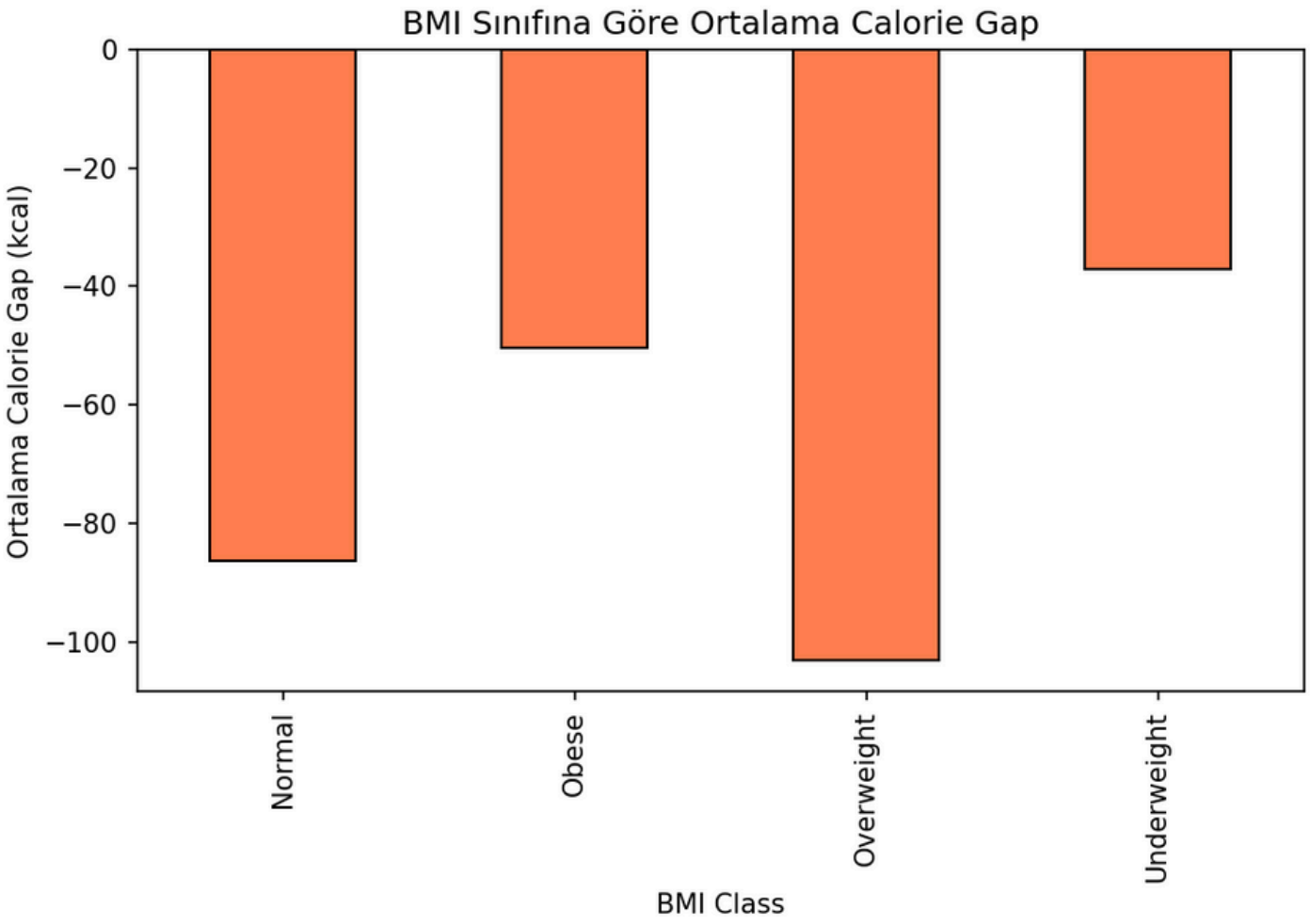


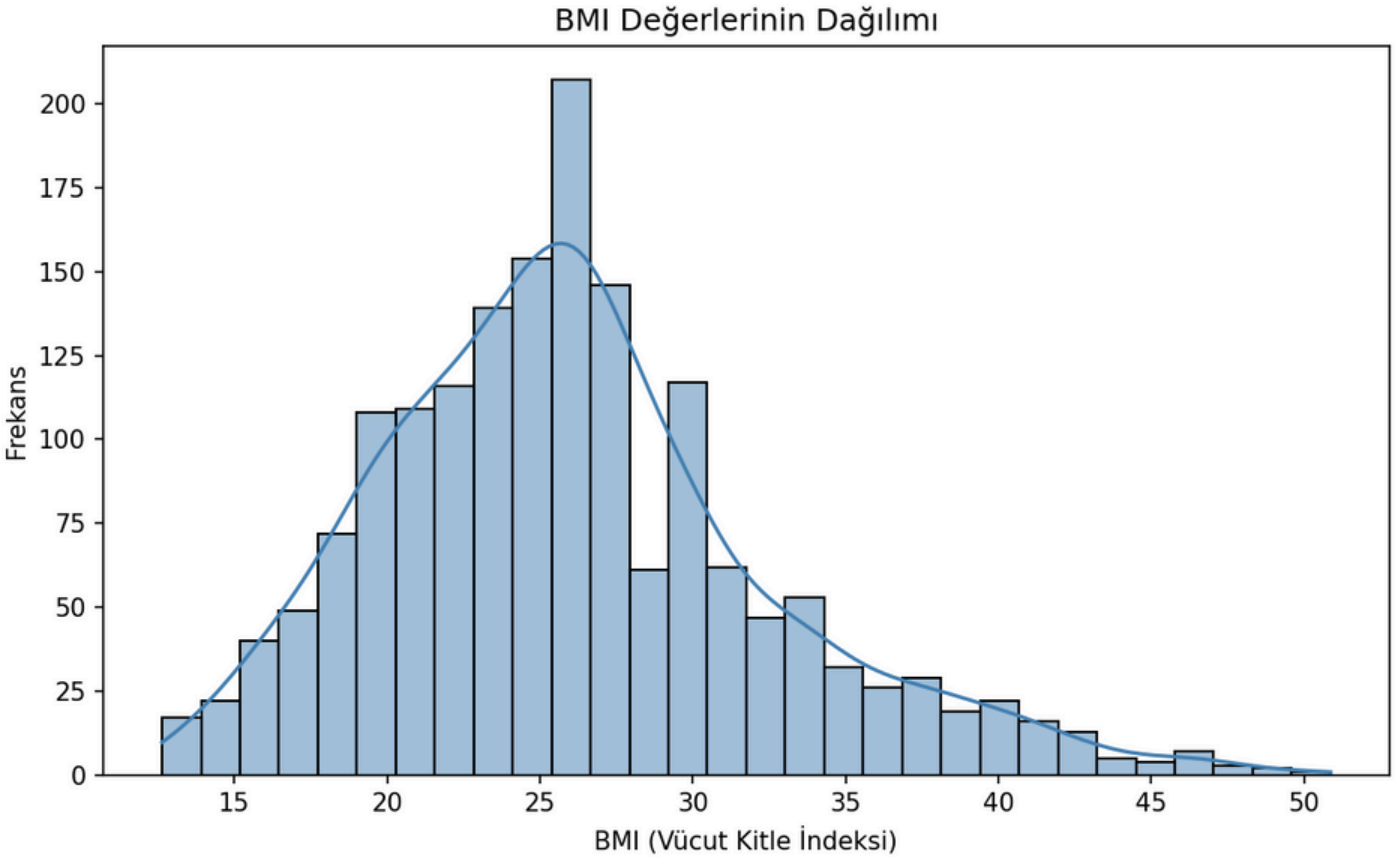
diet_recommendation:

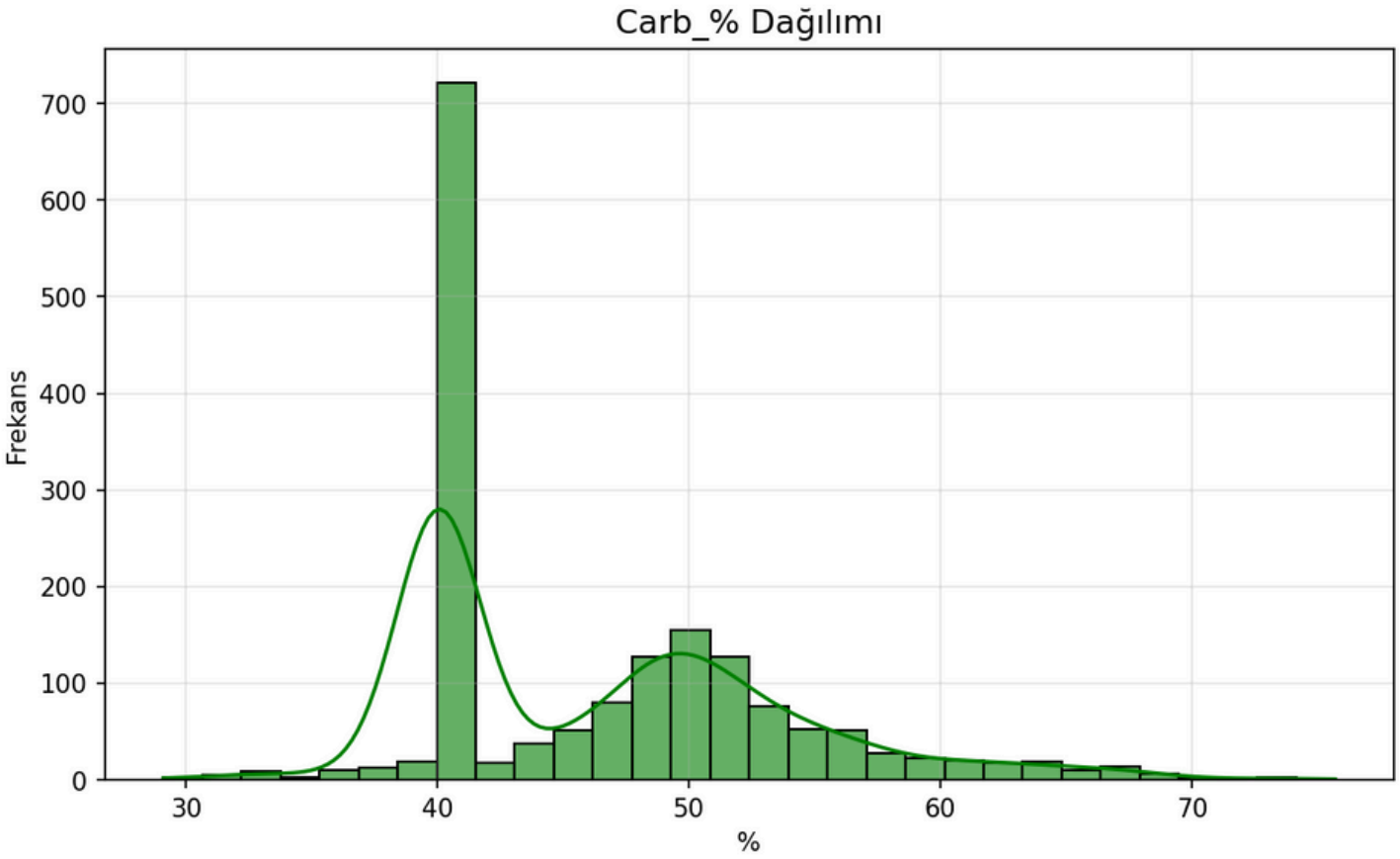


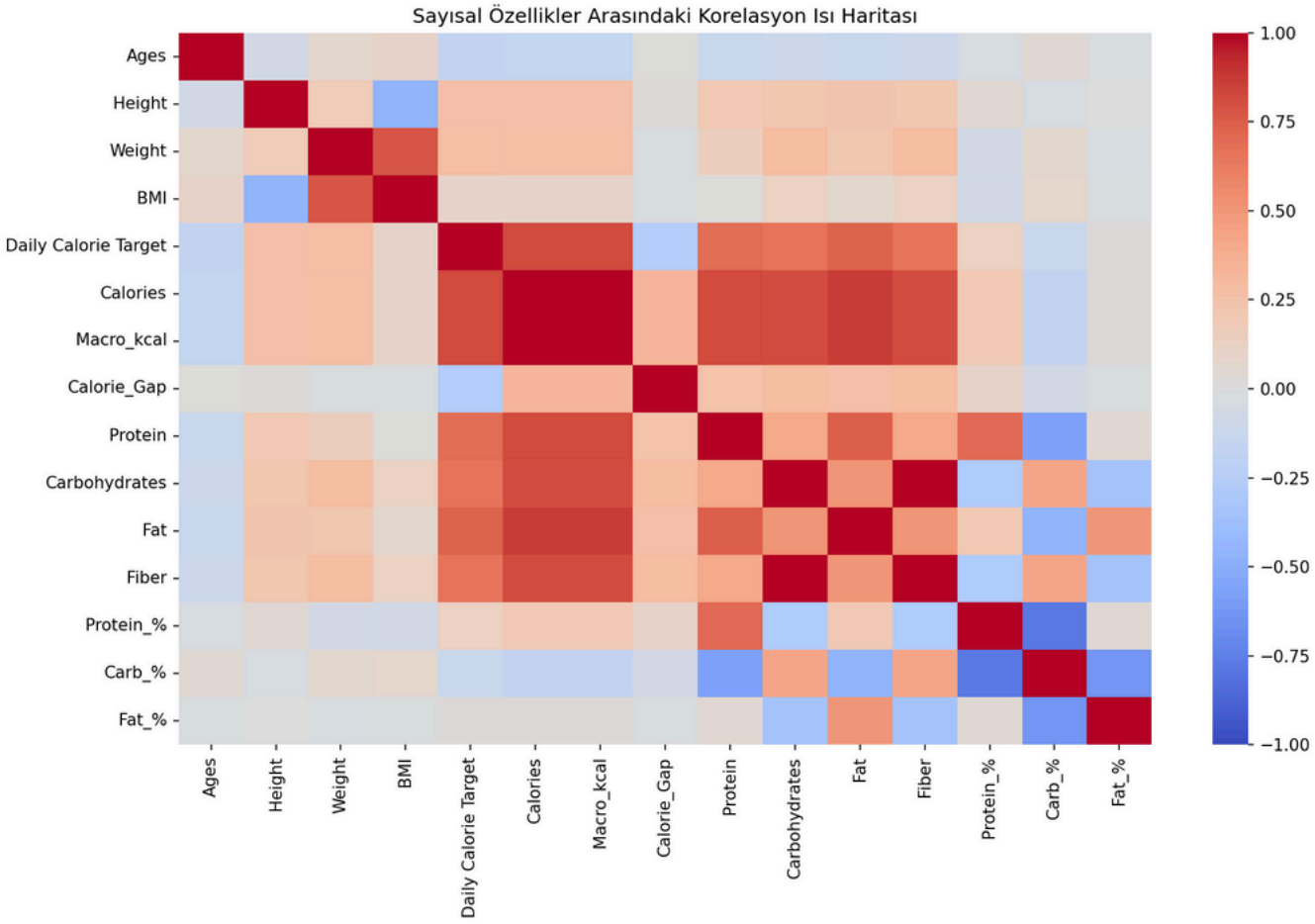

Yaş Gruplarına Göre Hedeften Sapma (Calorie Gap)
Calorie Gap (kcal)
1500
1000
500
0
−500
−1000
−1500
<=18
19-25
26-35
36-45
46-60
60+
Yaş Grubu

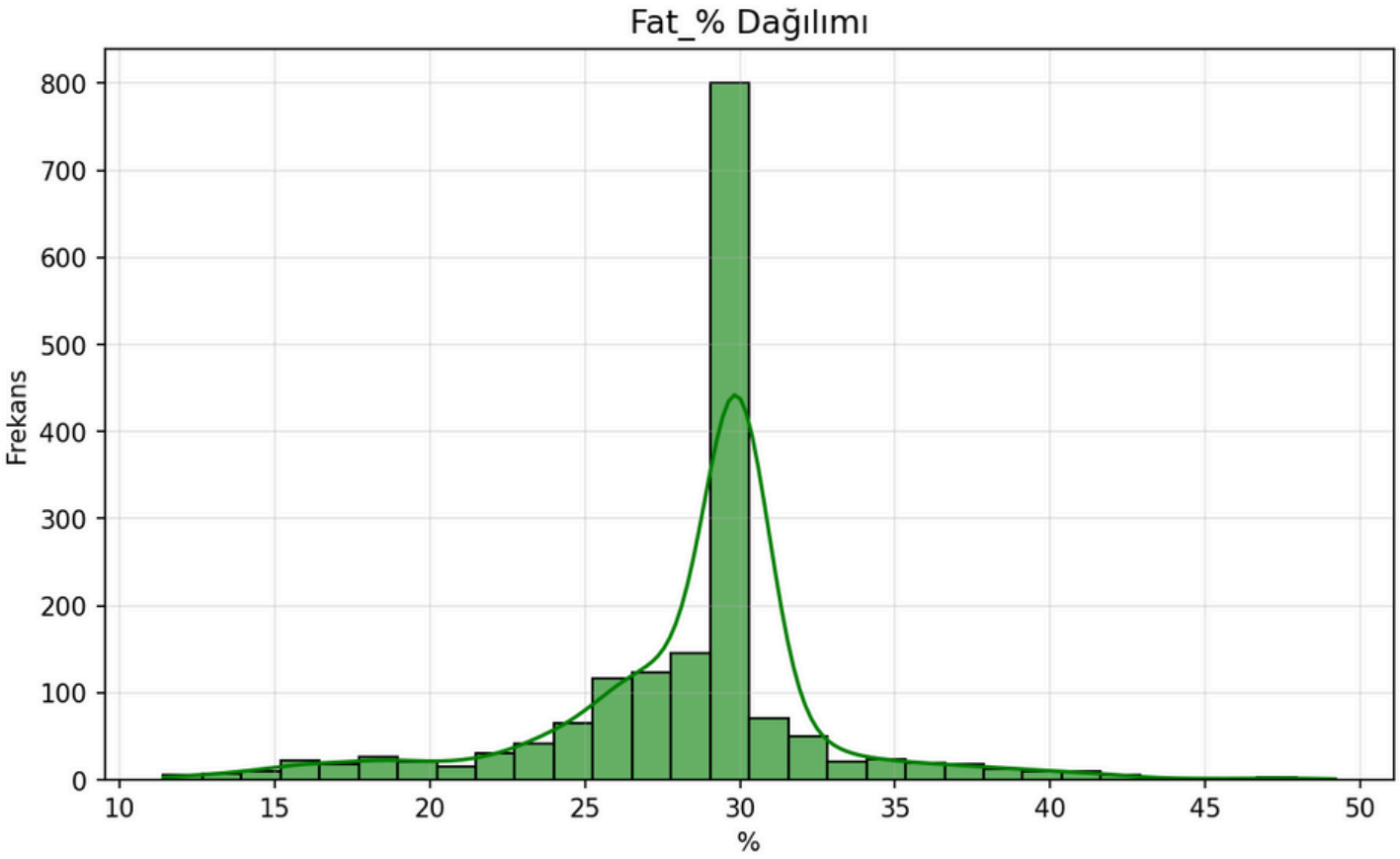

Yaş Gruplarına Göre Makro Dağılımı (%)
Makro Türü
Protein_%
Carb_%
Fat_%
Makro Oranı (%)
70
60
50
40
30
20
10
<=18
19-25
26-35
36-45
46-60
60+
Yaş Grubu

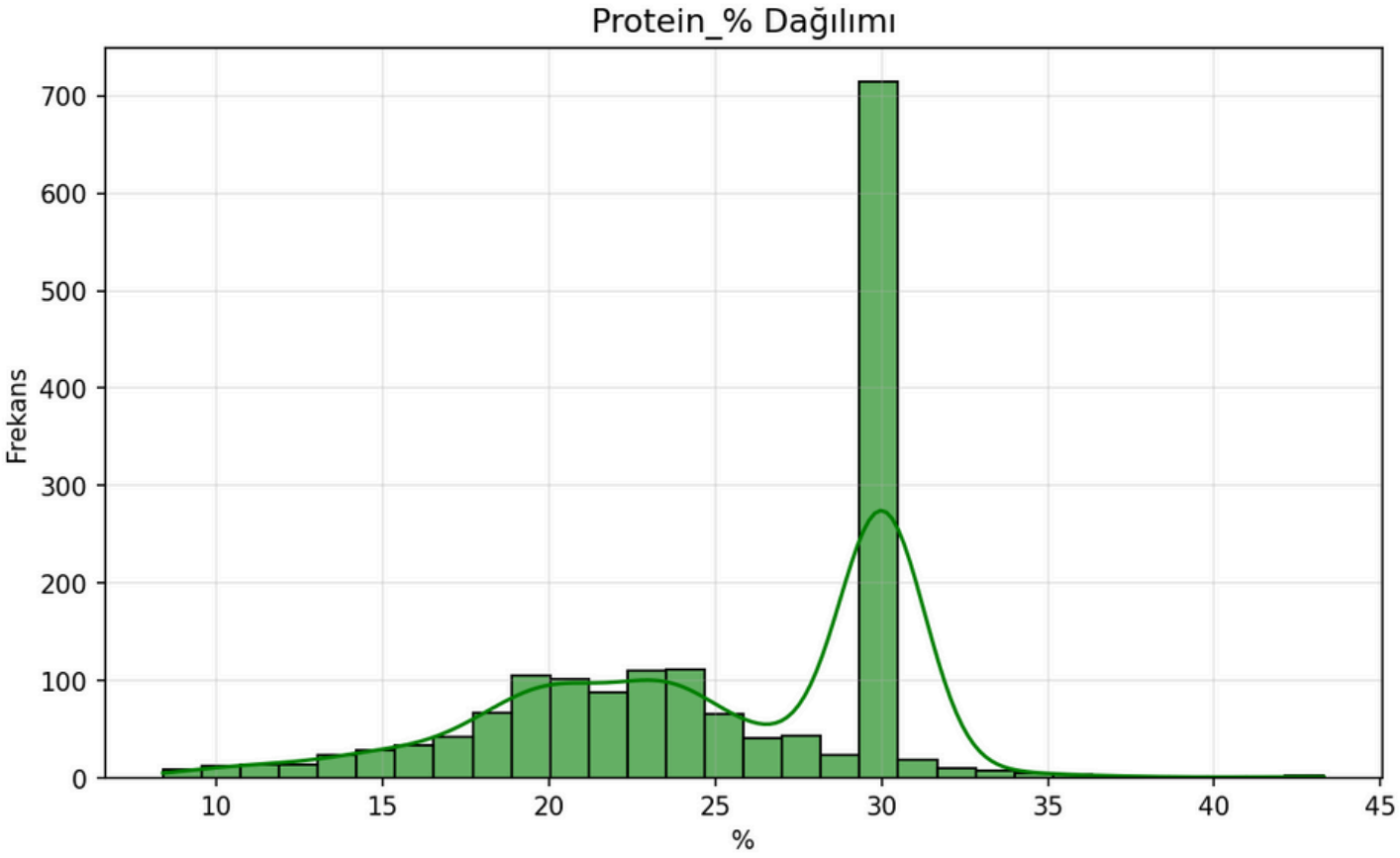

BMI Sınıfına Göre Ortalama Calorie Gap
Ortalama Calorie Gap (kcal)
0
−20
−40
−60
−80
−100
Normal
Obese
Overweight
Underweight
BMI Class

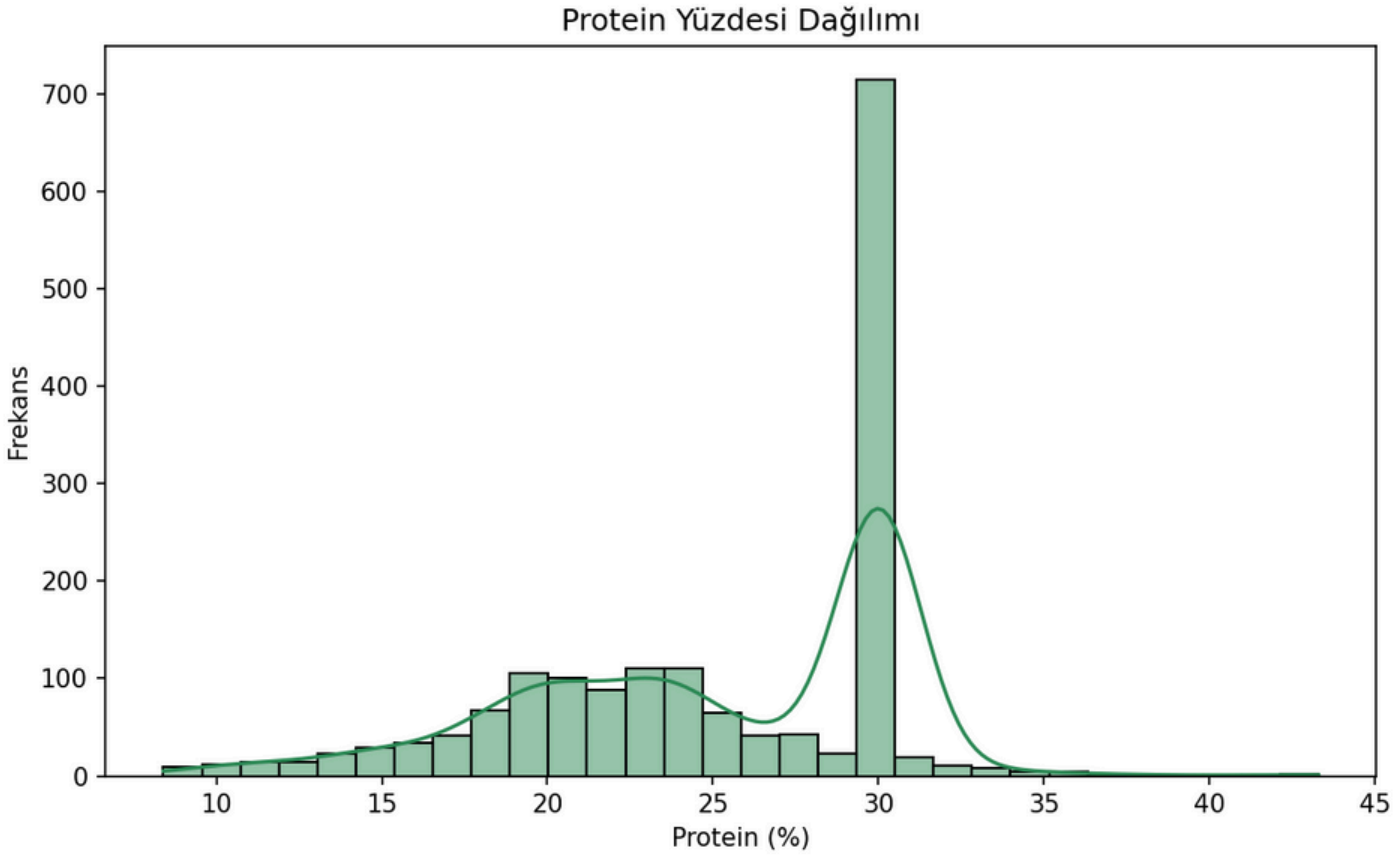

BMI Değerlerinin Dağılımı
Frekans
200
175
150
125
100
75
50
25
0
15
20
25
30
35
40
45
50
BMI (Vücut Kitle İndeksi)


Carb_% Dağılımı
Frekans
%
0
100
200
300
400
500
600
700
30
40
50
60
70

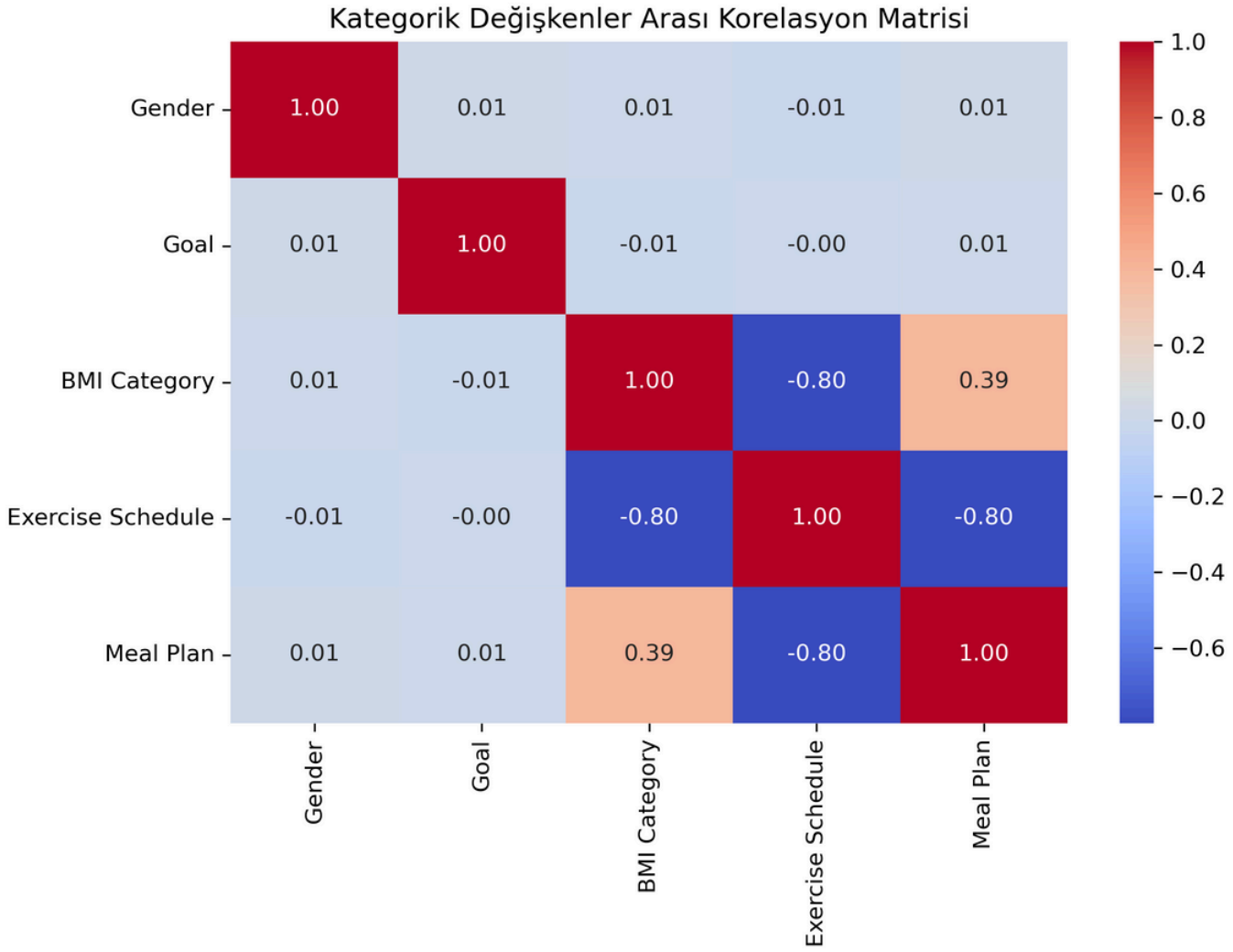

Sayısal Özellikler Arasındaki Korelasyon Isı Haritası
Ages
Height
Weight
BMI
Daily Calorie Target
Calories
Macro_kcal
Calorie_Gap
Protein
Carbohydrates
Fat
Fiber
Protein_%
Carb_%
Fat_%
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

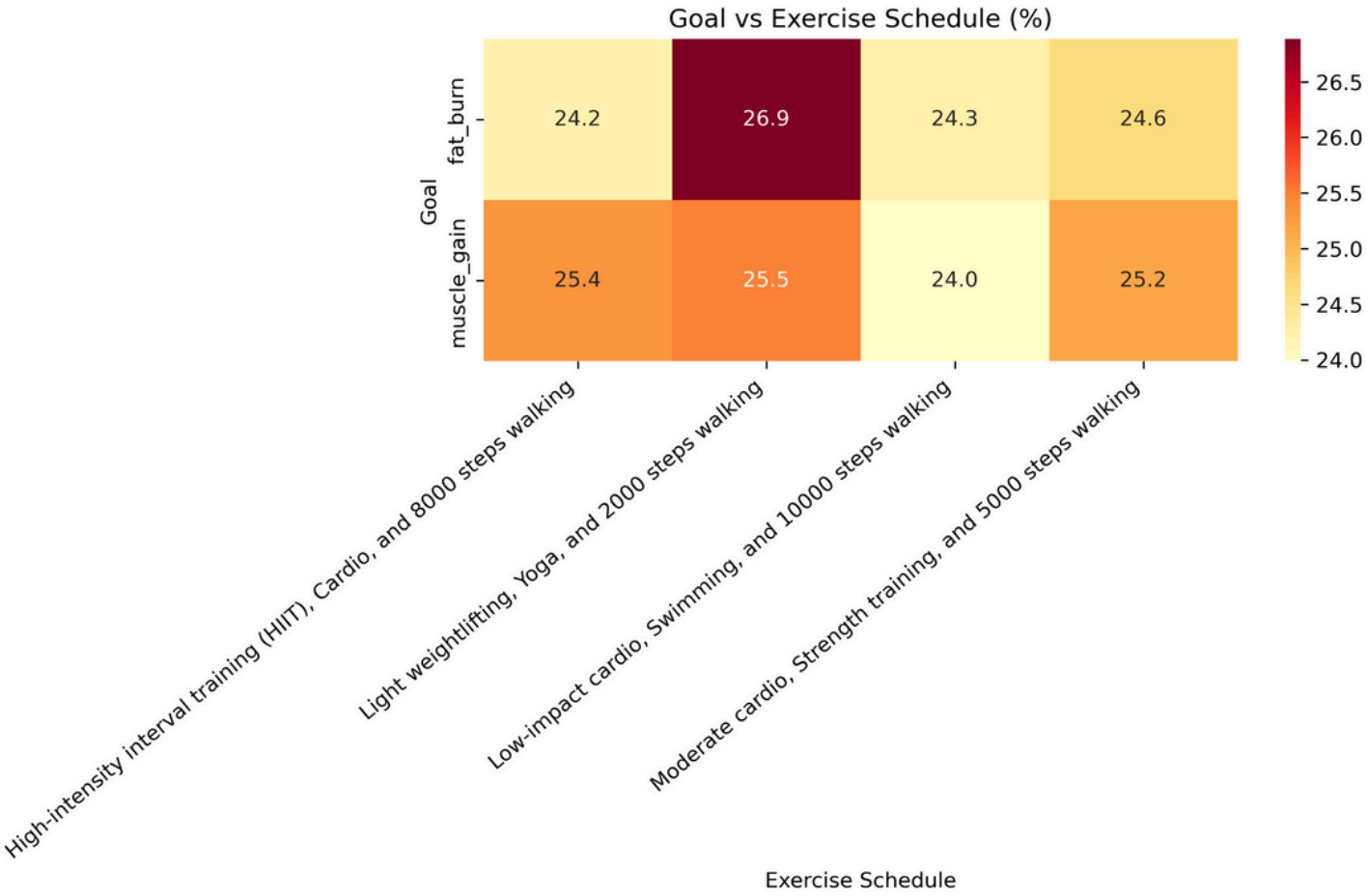

Fat_% Dağılımı
Frekans
%
0
100
200
300
400
500
600
700
800
10
15
20
25
30
35
40
45
50

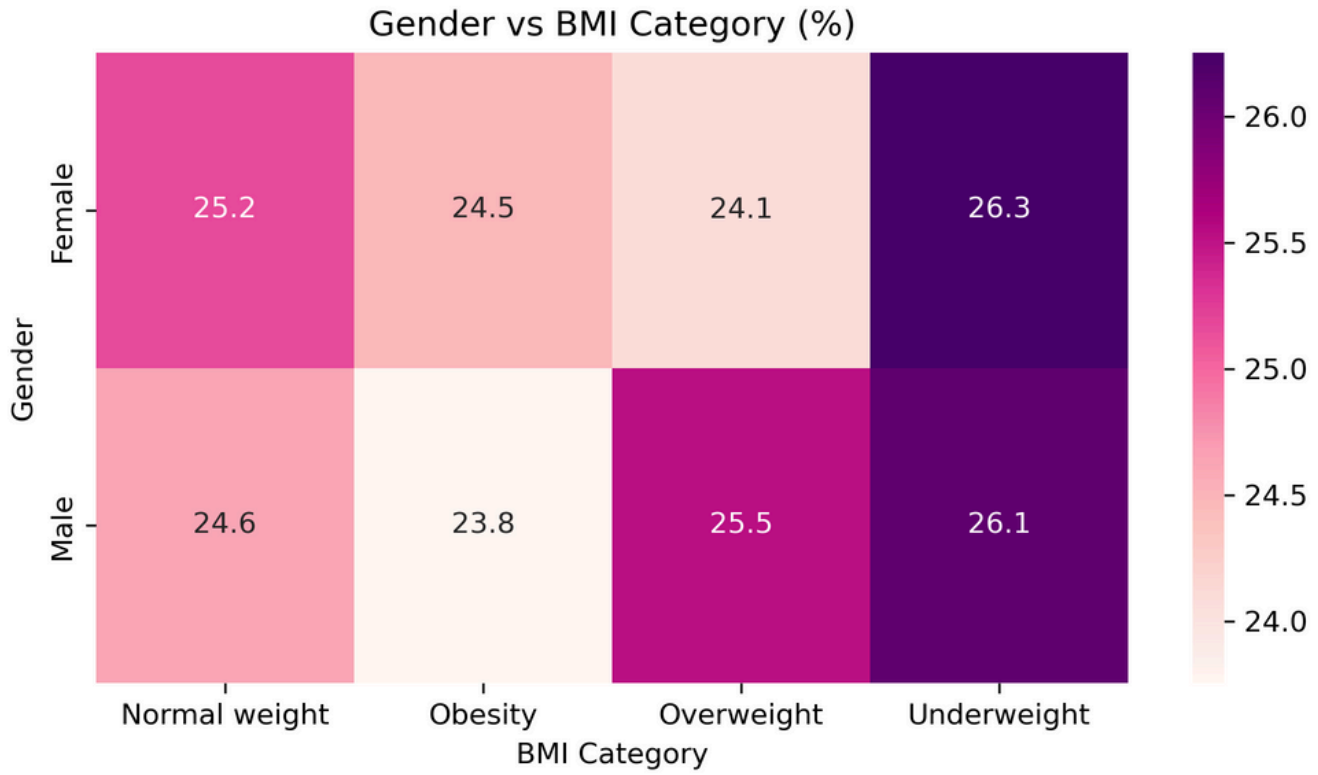

Protein_% Dağılımı
Frekans
%
0
100
200
300
400
500
600
700
10
15
20
25
30
35
40
45

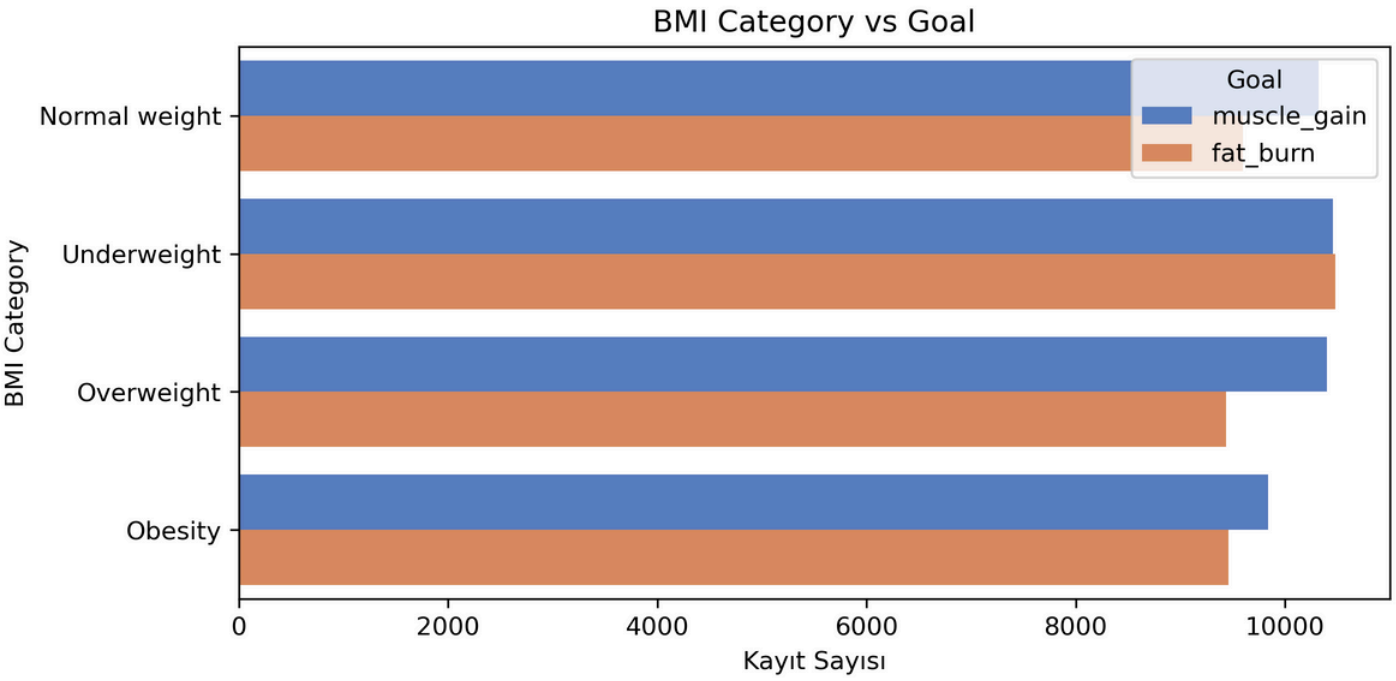


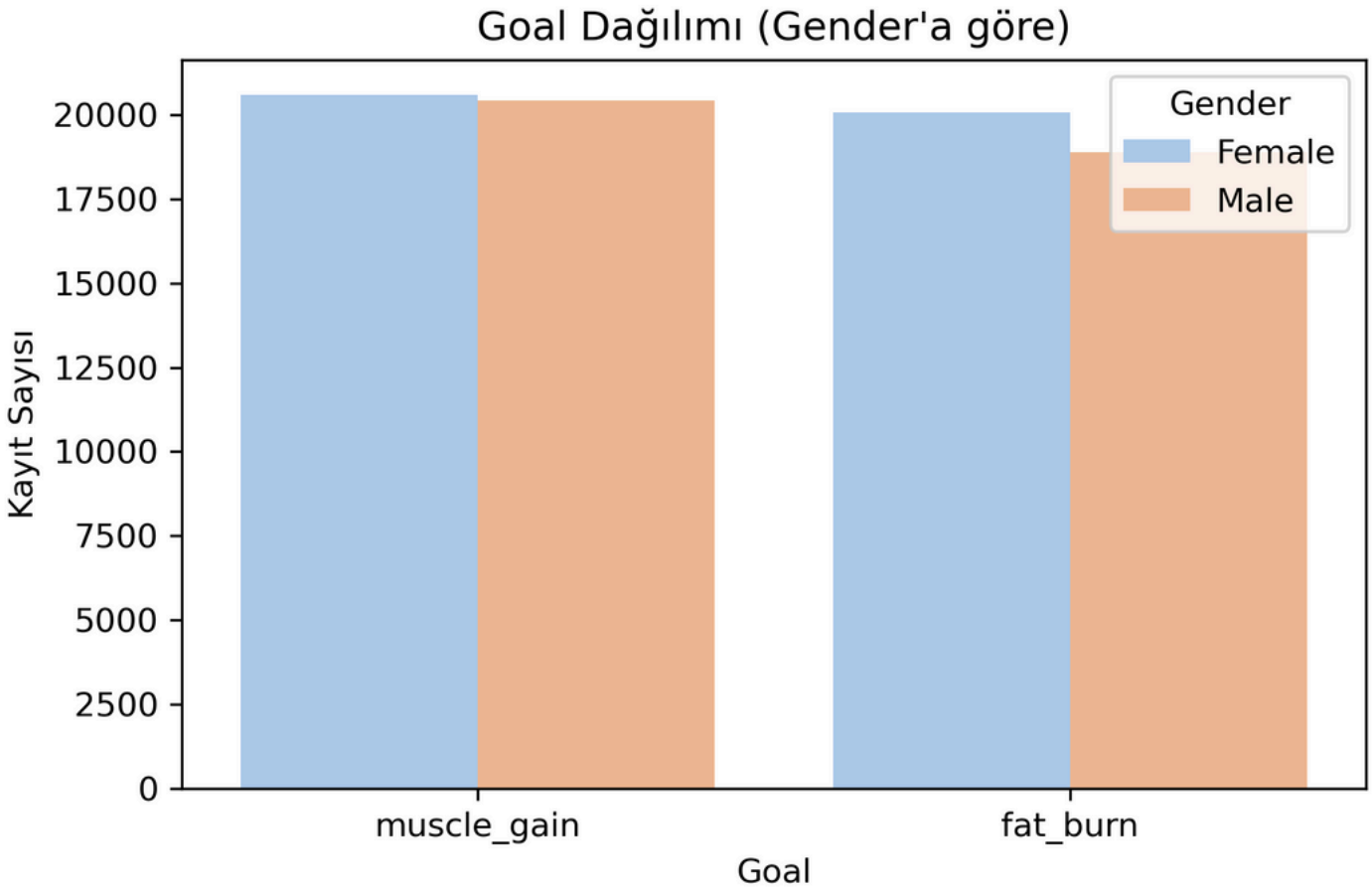
meal_ex_planner:

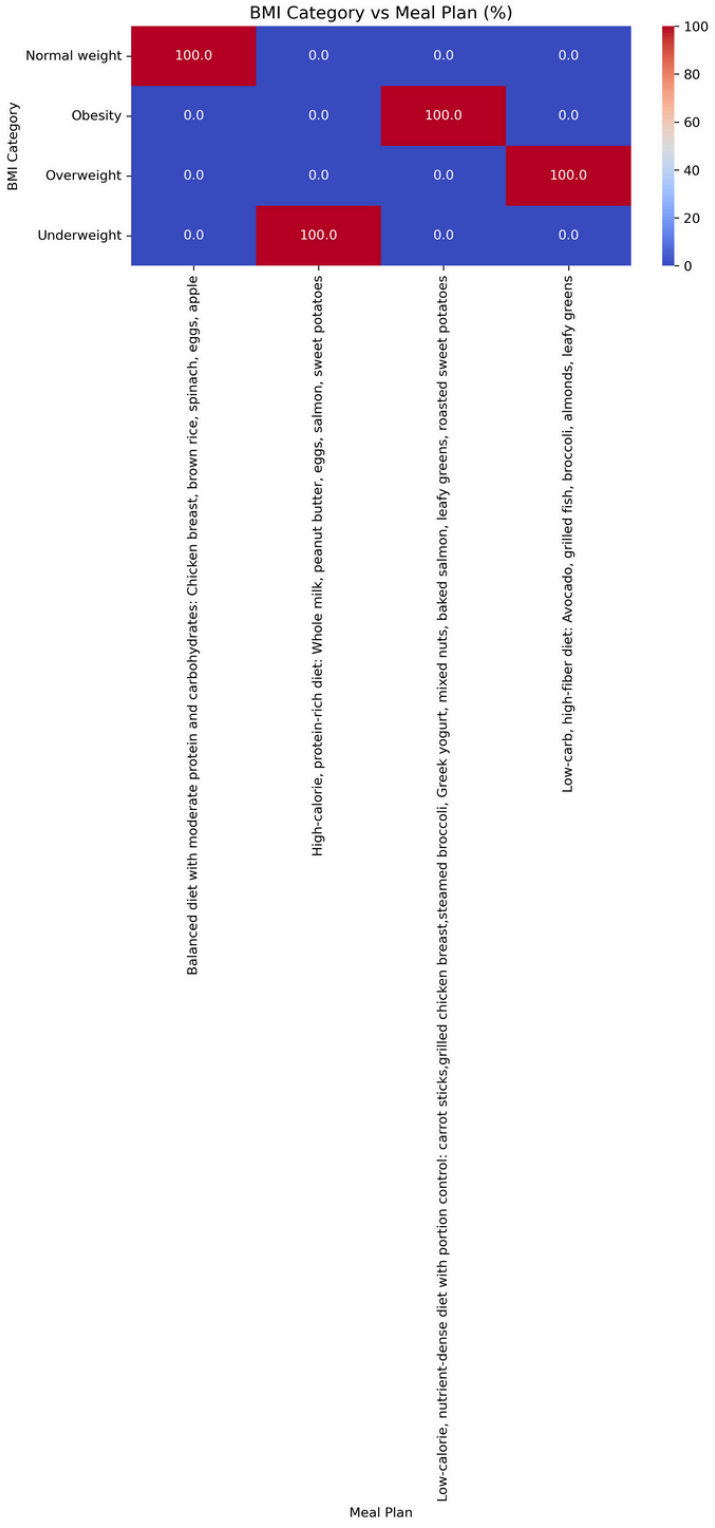

Kategorik Değişkenler Arası Korelasyon Matrisi
Gender
Goal
BMI Category
Exercise Schedule
Meal Plan
1.00 0.01 0.01 -0.01 0.01
0.01 1.00 -0.01 -0.00 0.01
0.01 -0.01 1.00 -0.80 0.39
-0.01 -0.00 -0.80 1.00 -0.80
0.01 0.01 0.39 -0.80 1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

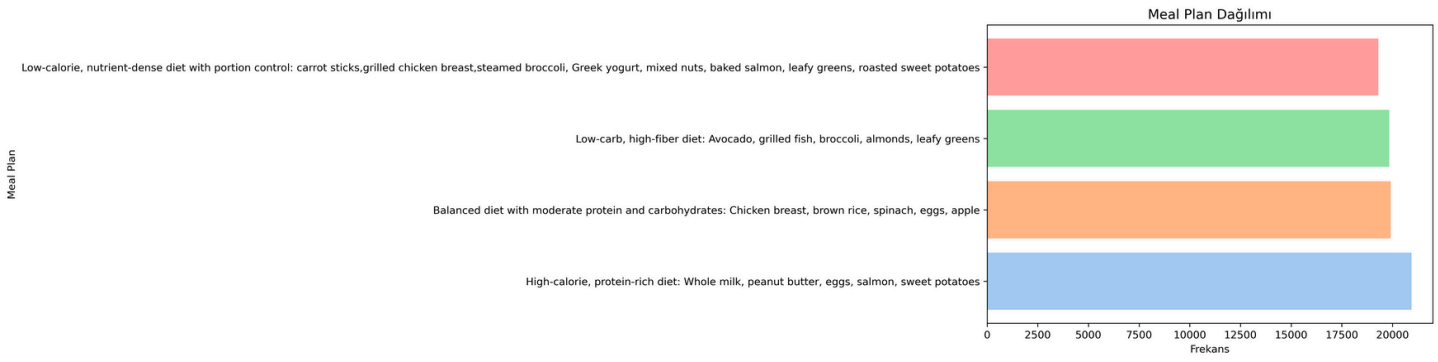

Goal vs Exercise Schedule (%)
Goal
fat_burn
muscle_gain
24.2 26.9 24.3 24.6
25.4 25.5 24.0 25.2
26.5
26.0
25.5
25.0
24.5
24.0
High-intensity interval training (HIIT), Cardio, and 8000 steps walking
Light weightlifting, Yoga, and 2000 steps walking
Low-impact cardio, Swimming, and 10000 steps walking
Moderate cardio, Strength training, and 5000 steps walking
Exercise Schedule

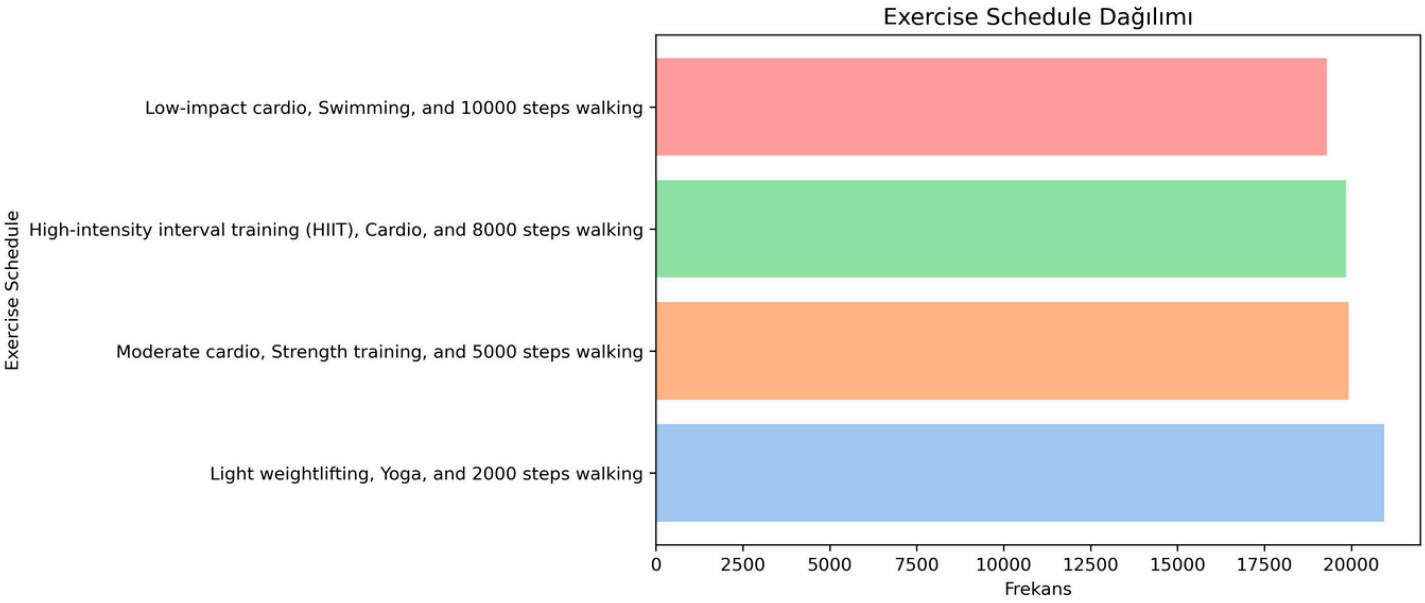

Gender vs BMI Category (%)
Female
Male
Gender
25.2
24.5
24.1
26.3
24.6
23.8
25.5
26.1
26.0
25.5
25.0
24.5
24.0
Normal weight
Obesity
Overweight
Underweight
BMI Category

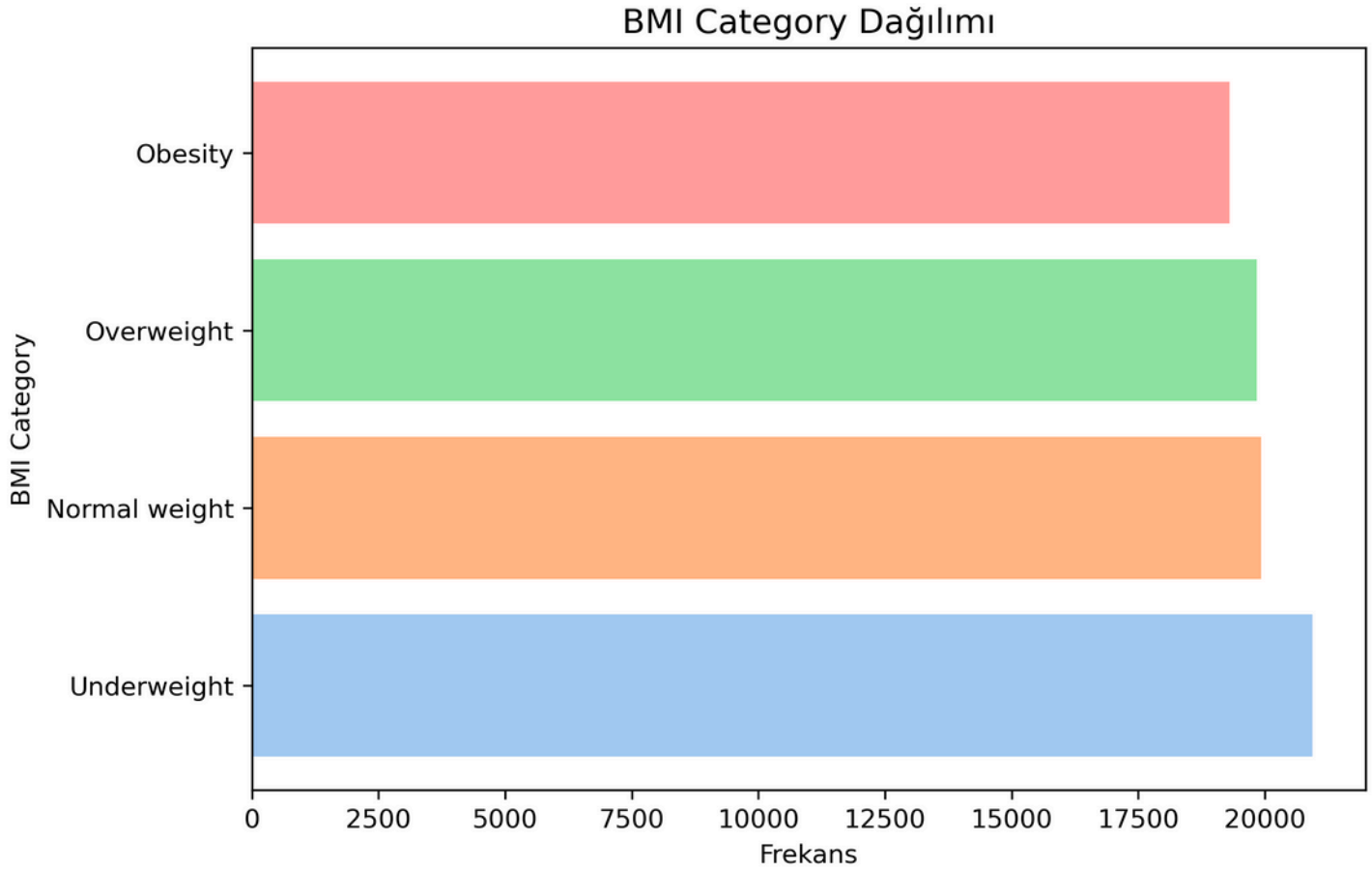

BMI Category vs Goal
Goal
muscle_gain
fat_burn
Normal weight
Underweight
Overweight
Obesity
BMI Category
0
2000
4000
6000
8000
10000
Kayıt Sayısı

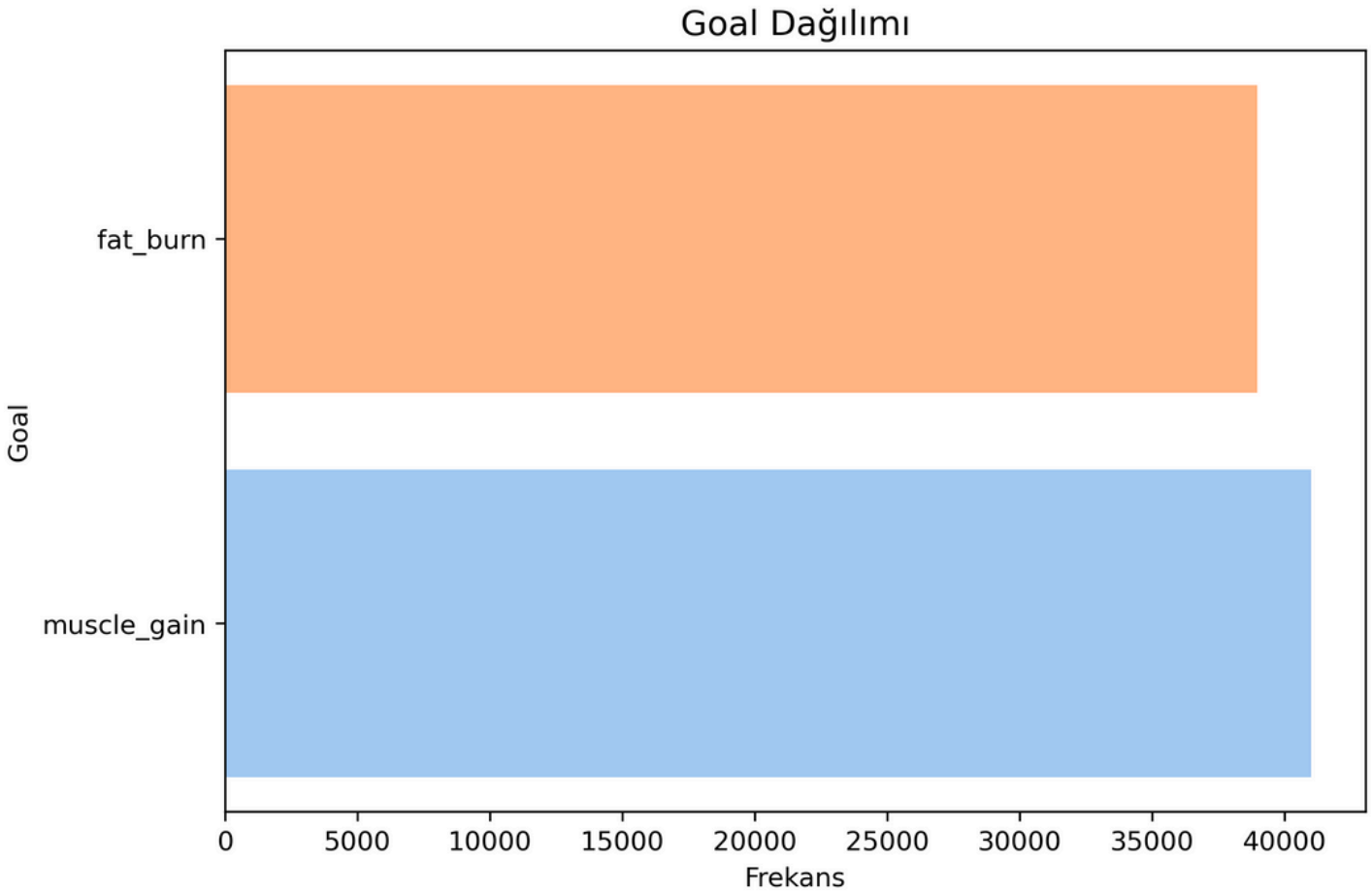

Goal Dağılımı (Gender'a göre)
Gender
Female
Male
Kayıt Sayısı
20000
17500
15000
12500
10000
7500
5000
2500
0
muscle_gain
fat_burn
Goal

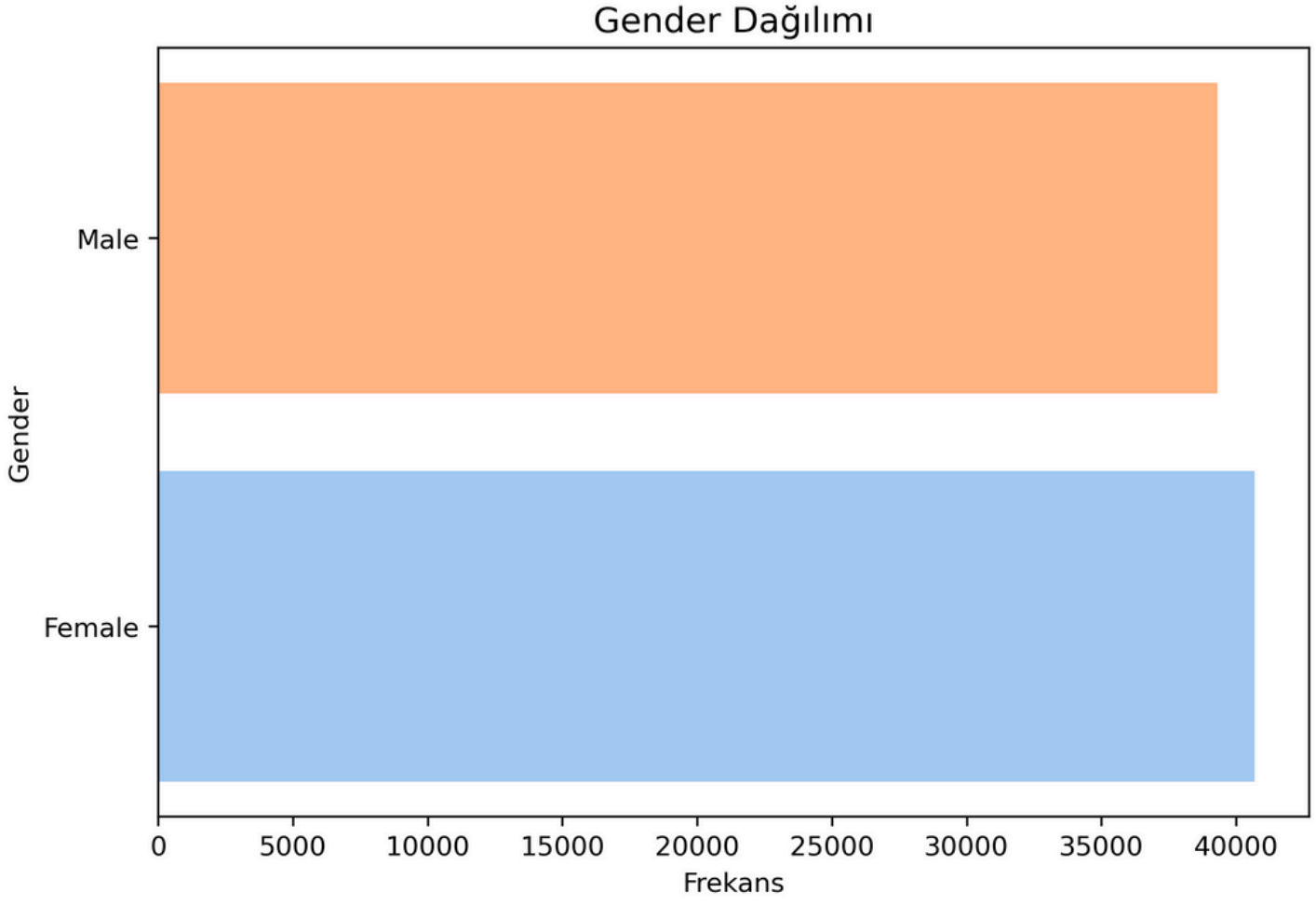

BMI Category vs Meal Plan (%)
BMI Category
Normal weight
Obesity
Overweight
Underweight
100.0
0.0
100
80
60
40
20
0
Balanced diet with moderate protein and carbohydrates: Chicken breast, brown rice, spinach, eggs, apple
High-calorie, protein-rich diet: Whole milk, peanut butter, eggs, salmon, sweet potatoes
Low-calorie, nutrient-dense diet with portion control: carrot sticks,grilled chicken breast,steamed broccoli, Greek yogurt, mixed nuts, baked salmon, leafy greens, roasted sweet potatoes
Low-carb, high-fiber diet: Avocado, grilled fish, broccoli, almonds, leafy greens
Meal Plan


Meal Plan Dağılımı
Meal Plan
Low-calorie, nutrient-dense diet with portion control: carrot sticks,grilled chicken breast,steamed broccoli, Greek yogurt, mixed nuts, baked salmon, leafy greens, roasted sweet potatoes
Low-carb, high-fiber diet: Avocado, grilled fish, broccoli, almonds, leafy greens
Balanced diet with moderate protein and carbohydrates: Chicken breast, brown rice, spinach, eggs, apple
High-calorie, protein-rich diet: Whole milk, peanut butter, eggs, salmon, sweet potatoes
0
2500
5000
7500
10000
12500
15000
17500
20000
Frekans

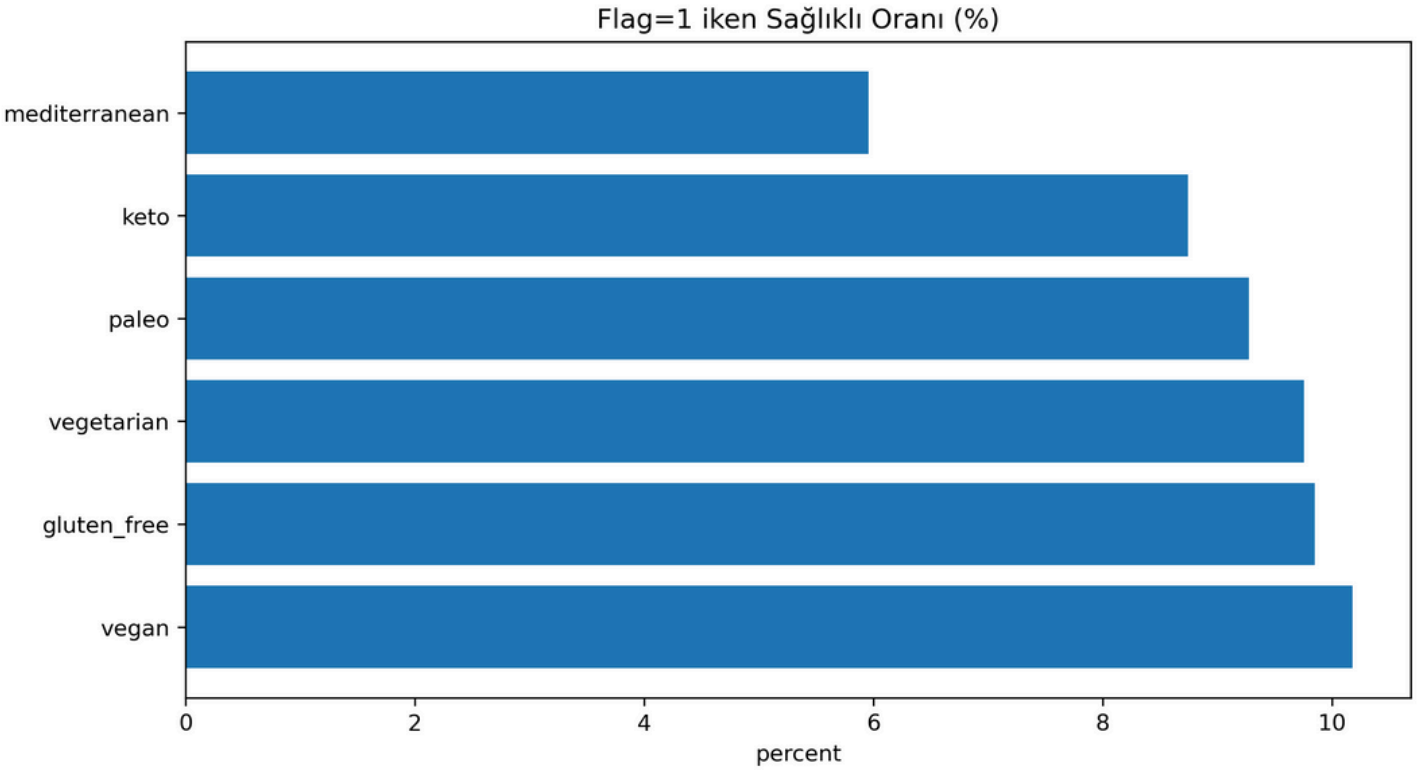

Exercise Schedule Dağılımı
Exercise Schedule
Low-impact cardio, Swimming, and 10000 steps walking
High-intensity interval training (HIIT), Cardio, and 8000 steps walking
Moderate cardio, Strength training, and 5000 steps walking
Light weightlifting, Yoga, and 2000 steps walking
0
2500
5000
7500
10000
12500
15000
17500
20000
Frekans


BMI Category Dağılımı
BMI Category
Obesity
Overweight
Normal weight
Underweight
0
2500
5000
7500
10000
12500
15000
17500
20000
Frekans

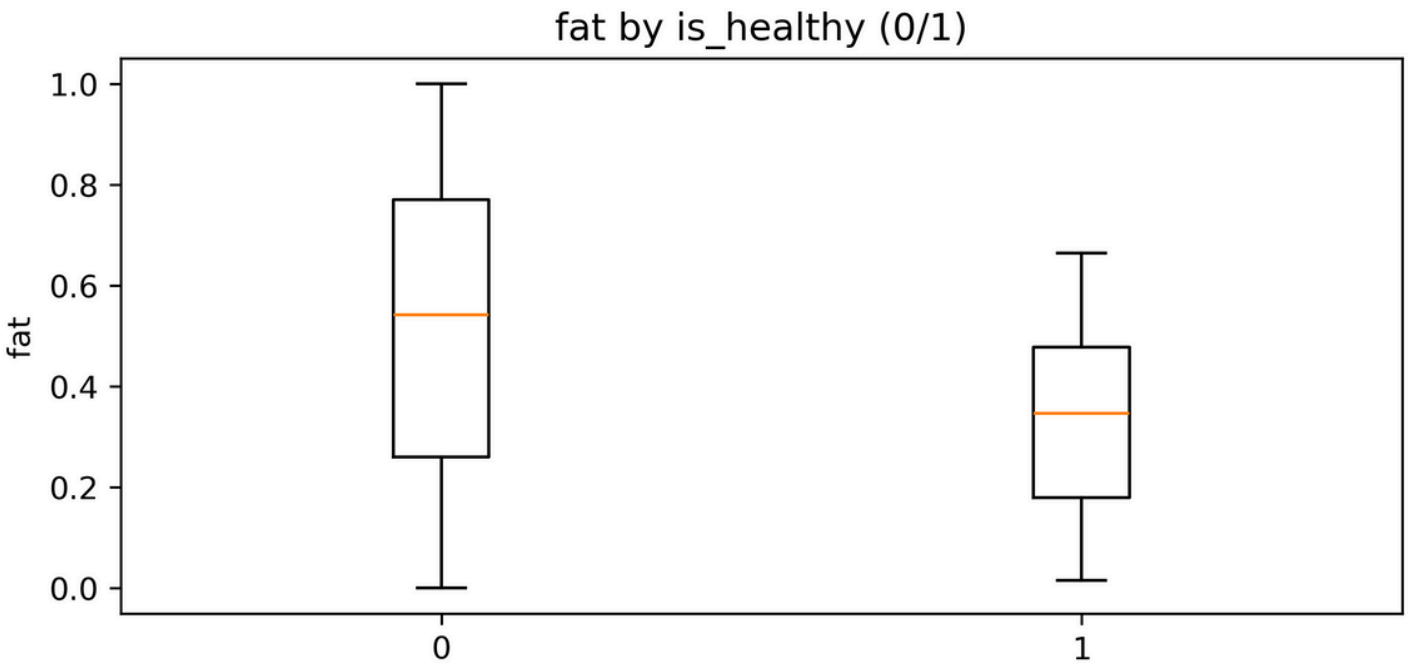


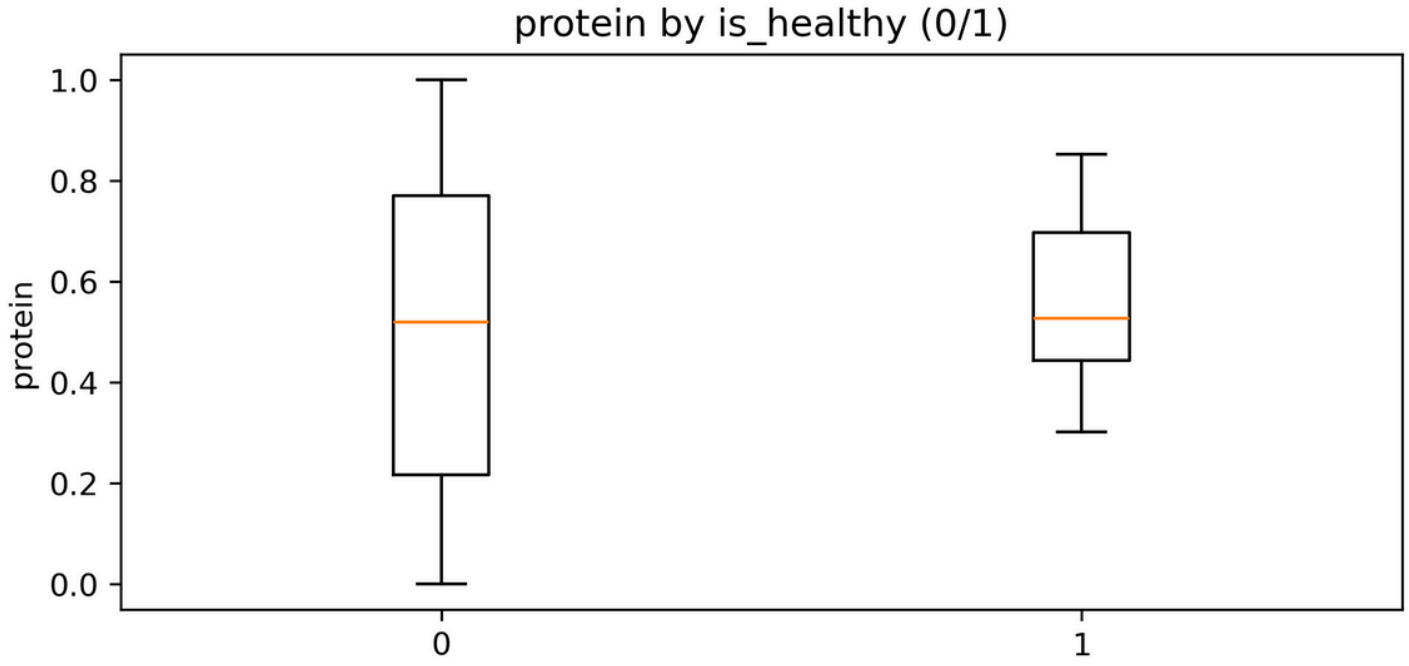


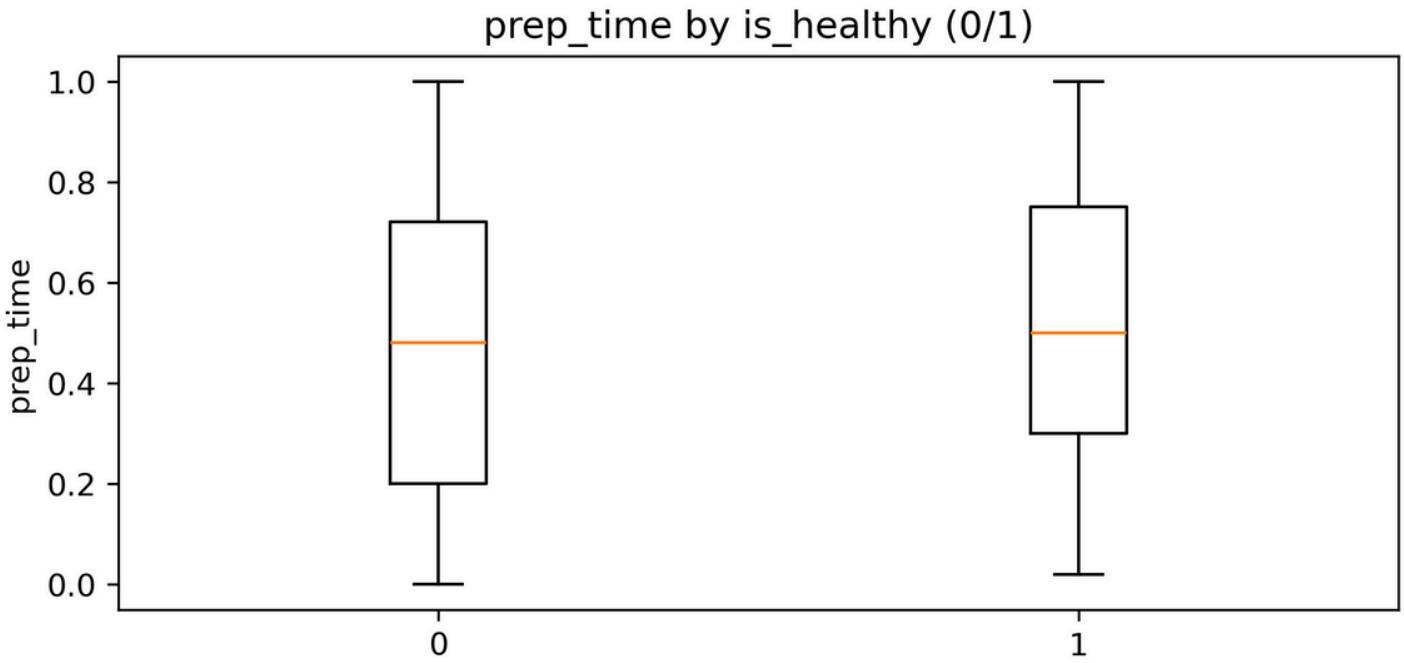
healthy_meal:

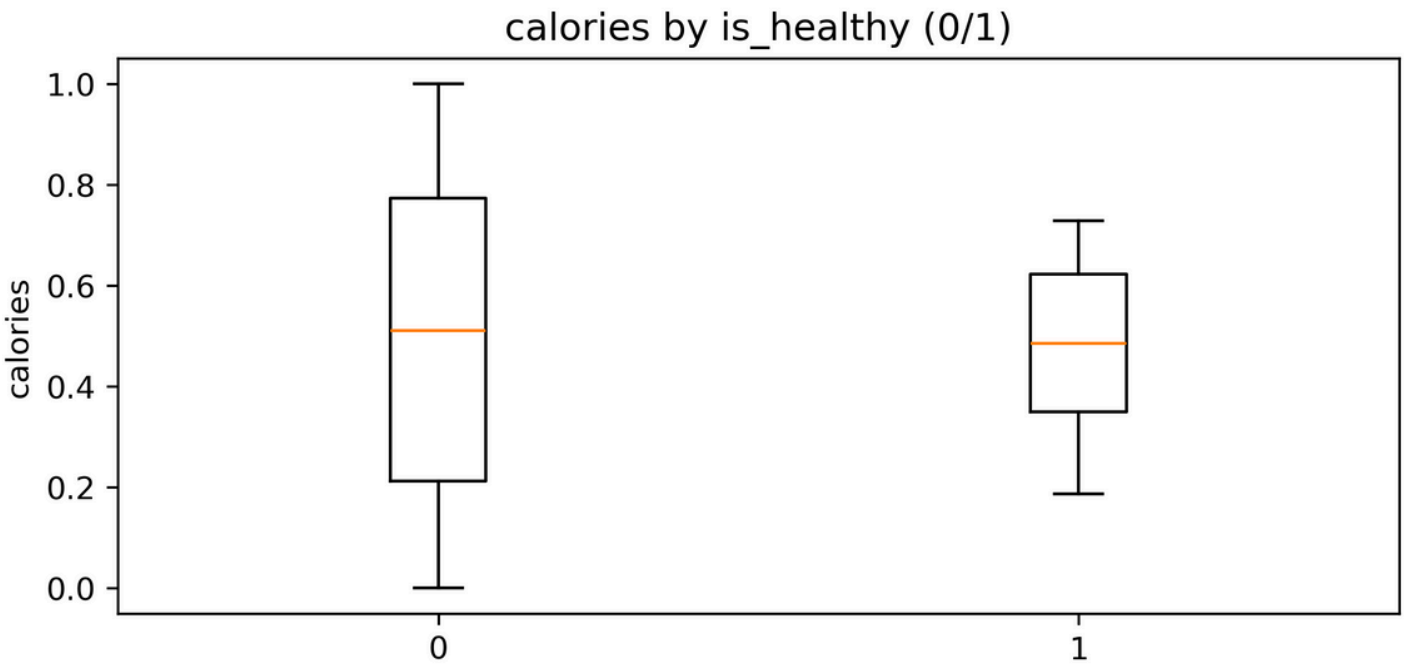

Flag=1 iken Sağlıklı Oranı (%)
mediterranean
keto
paleo
vegetarian
gluten_free
vegan
0
2
4
6
8
10
percent

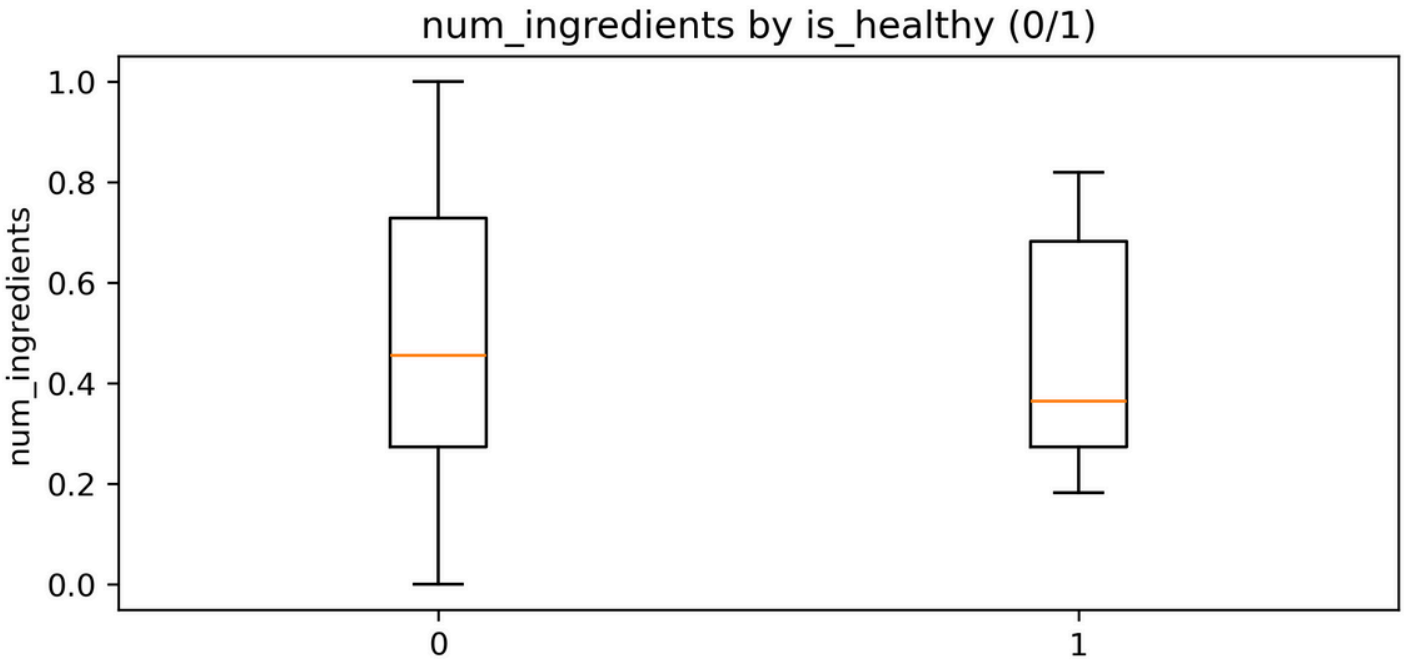

carbs by is_healthy (0/1)
carbs
1.0
0.8
0.6
0.4
0.2
0.0
0
1

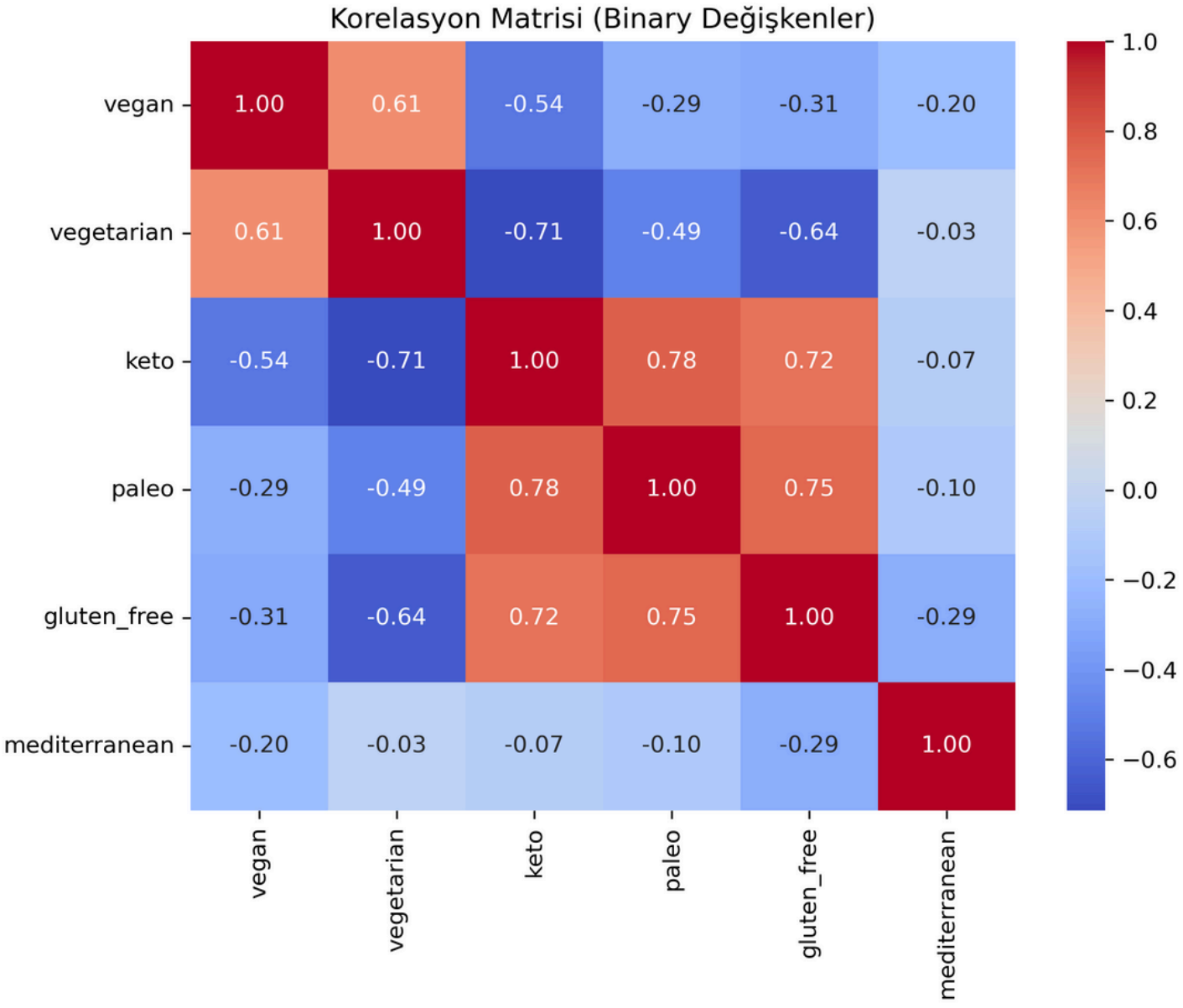

fat by is_healthy (0/1)
fat
1.0
0.8
0.6
0.4
0.2
0.0
0
1

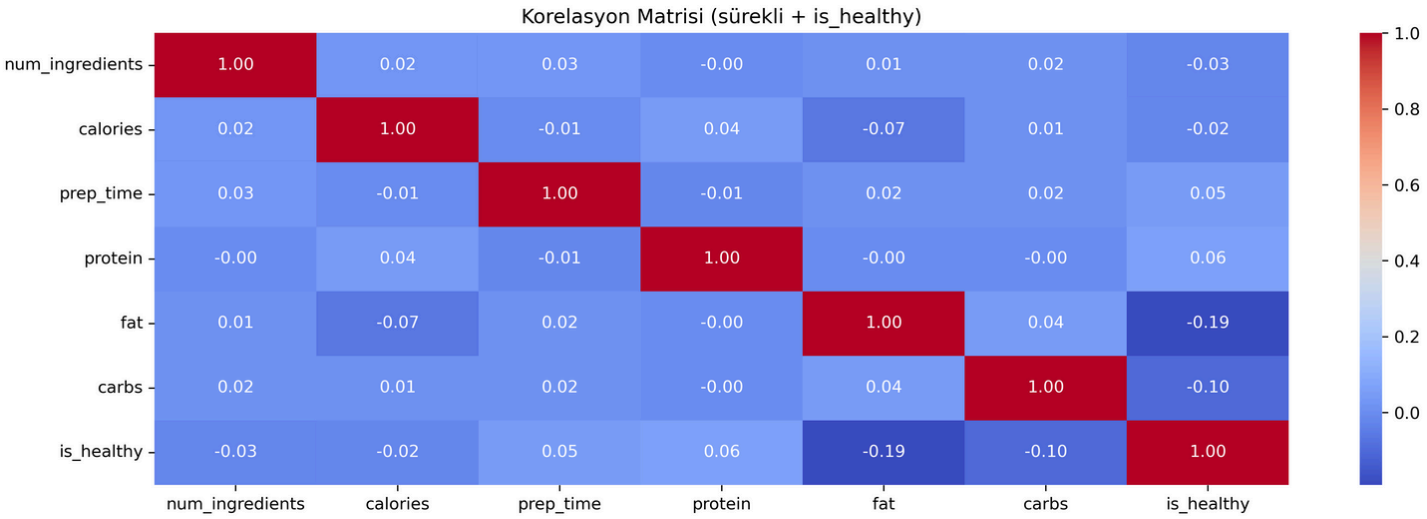

protein by is_healthy (0/1)
protein
1.0
0.8
0.6
0.4
0.2
0.0
0
1

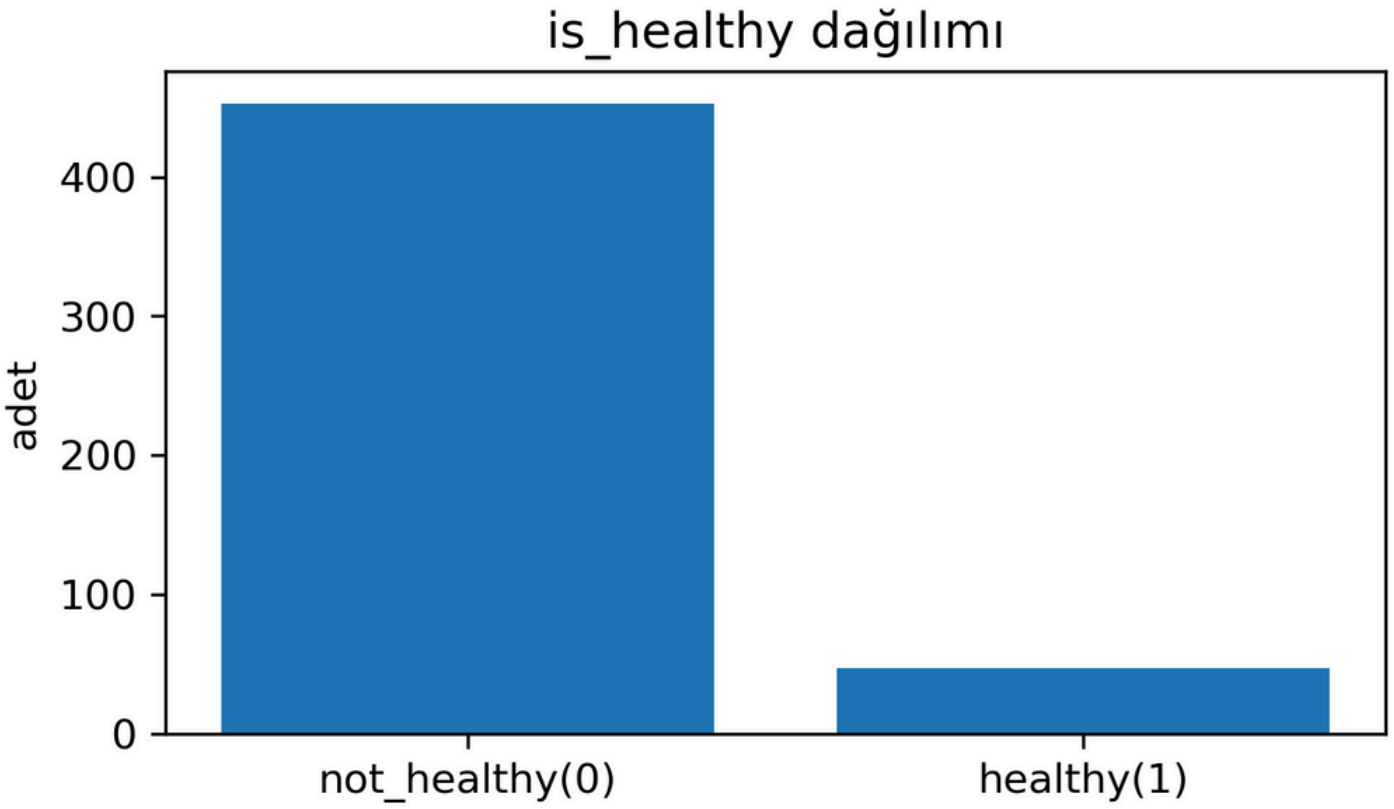

prep_time by is_healthy (0/1)
prep_time
1.0
0.8
0.6
0.4
0.2
0.0
0
1

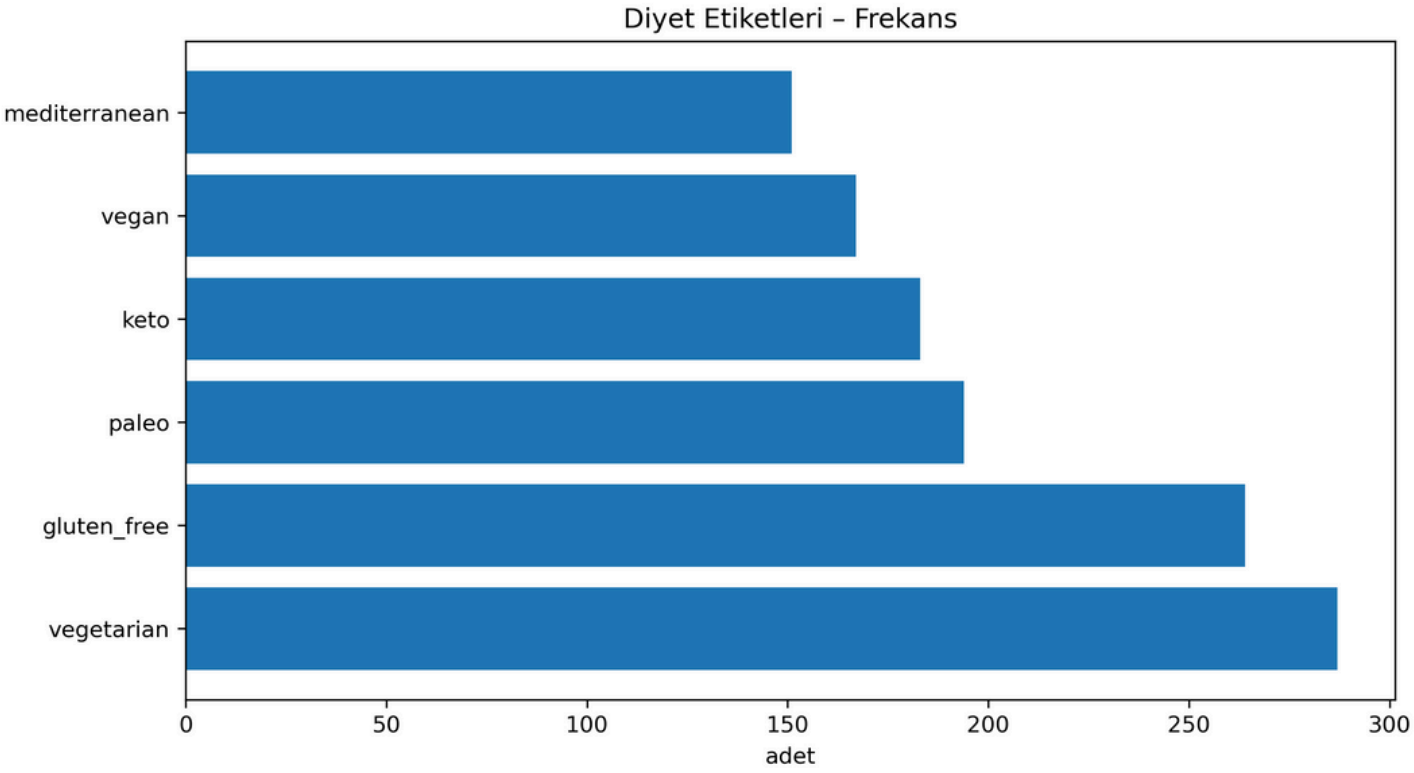

calories by is_healthy (0/1)
calories
1.0
0.8
0.6
0.4
0.2
0.0
0
1

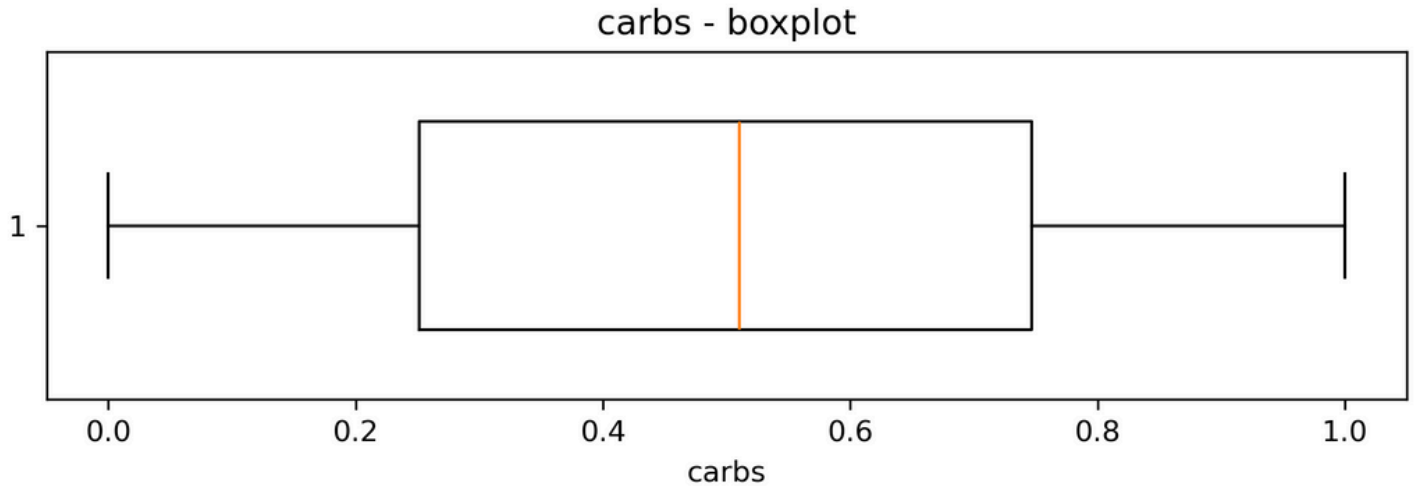

num_ingredients by is_healthy (0/1)
num_ingredients
1.0
0.8
0.6
0.4
0.2
0.0
0
1

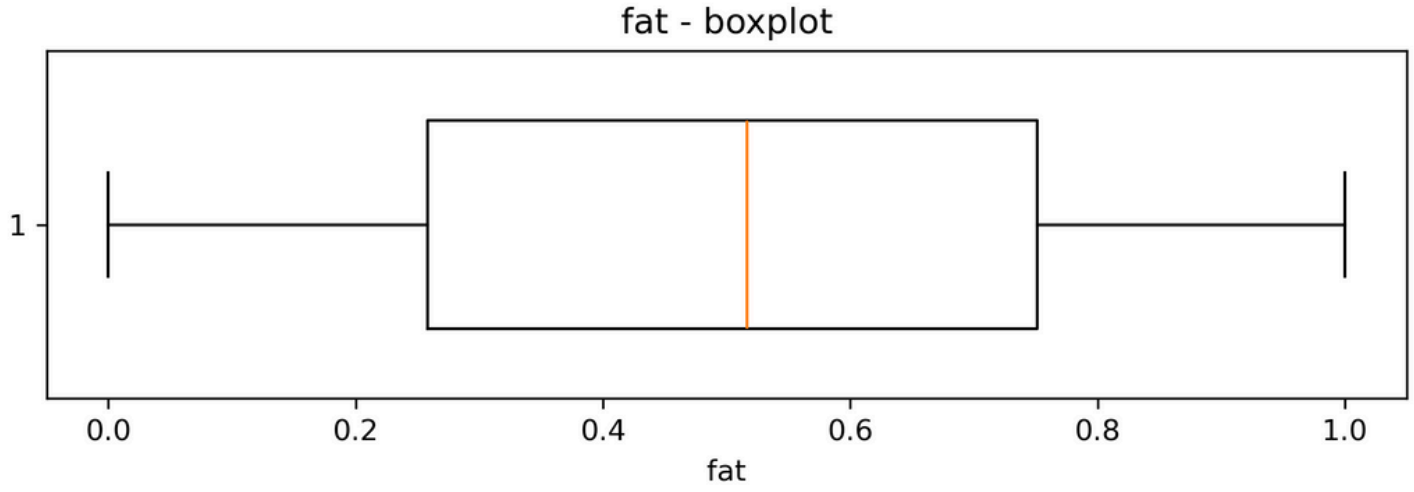

Korelasyon Matrisi (Binary Değişkenler)
vegan
vegetarian
keto
paleo
gluten_free
mediterranean
1.00 0.61 -0.54 -0.29 -0.31 -0.20
0.61 1.00 -0.71 -0.49 -0.64 -0.03
-0.54 -0.71 1.00 0.78 0.72 -0.07
-0.29 -0.49 0.78 1.00 0.75 -0.10
-0.31 -0.64 0.72 0.75 1.00 -0.29
-0.20 -0.03 -0.07 -0.10 -0.29 1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

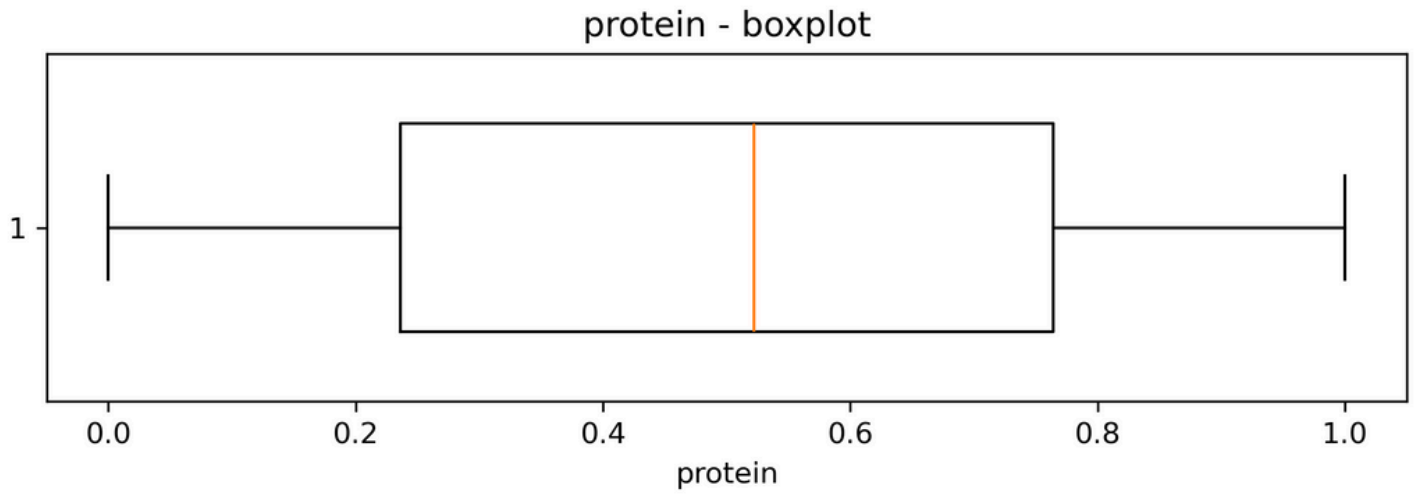

Korelasyon Matrisi (sürekli + is_healthy)
num_ingredients
calories
prep_time
protein
fat
carbs
is_healthy
1.00 0.02 0.03 -0.00 0.01 0.02 -0.03
0.02 1.00 -0.01 0.04 -0.07 0.01 -0.02
0.03 -0.01 1.00 -0.01 0.02 0.02 0.05
-0.00 0.04 -0.01 1.00 -0.00 -0.00 0.06
0.01 -0.07 0.02 -0.00 1.00 0.04 -0.19
0.02 0.01 0.02 -0.00 0.04 1.00 -0.10
-0.03 -0.02 0.05 0.06 -0.19 -0.10 1.00
1.0
0.8
0.6
0.4
0.2
0.0

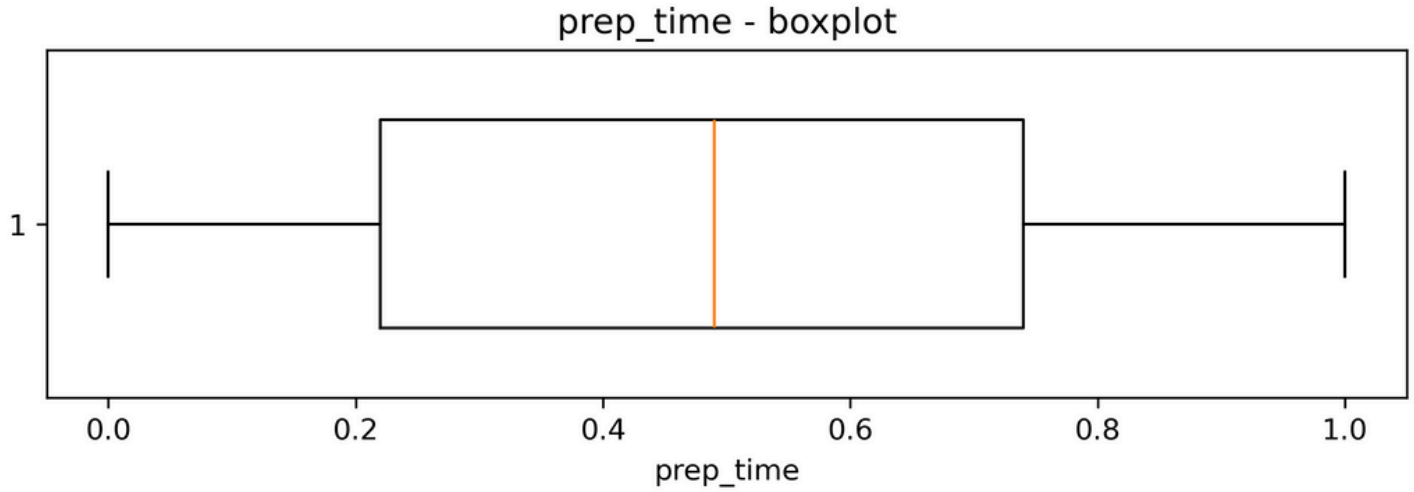

is_healthy dağılımı
adet
0
100
200
300
400
not_healthy(0)
healthy(1)

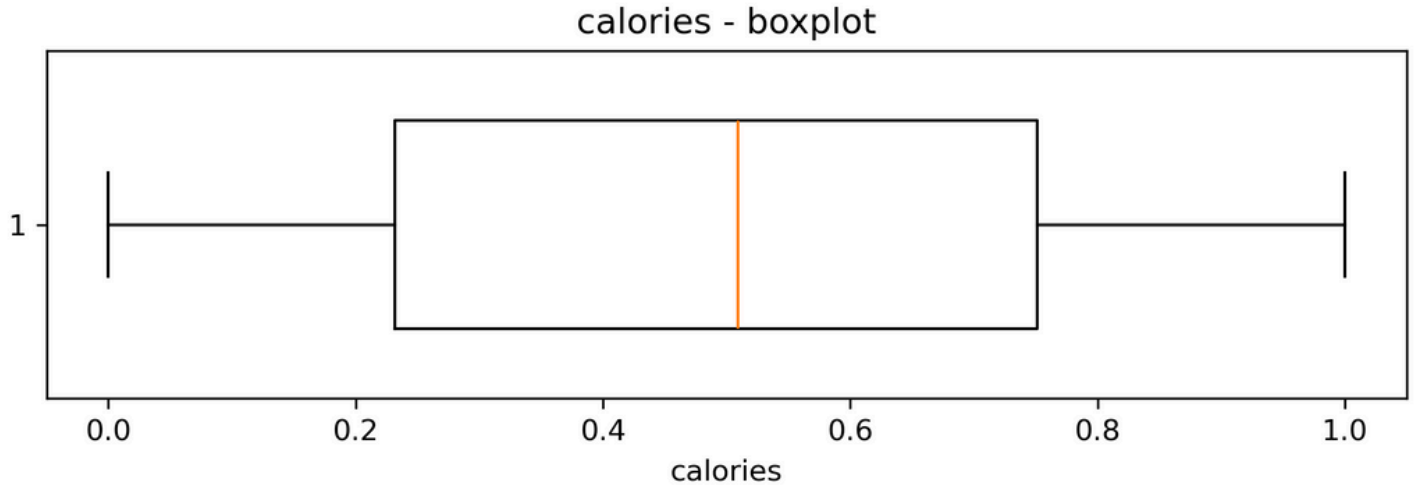

Diyet Etiketleri – Frekans
mediterranean
vegan
keto
paleo
gluten_free
vegetarian
0
50
100
150
200
250
300
adet

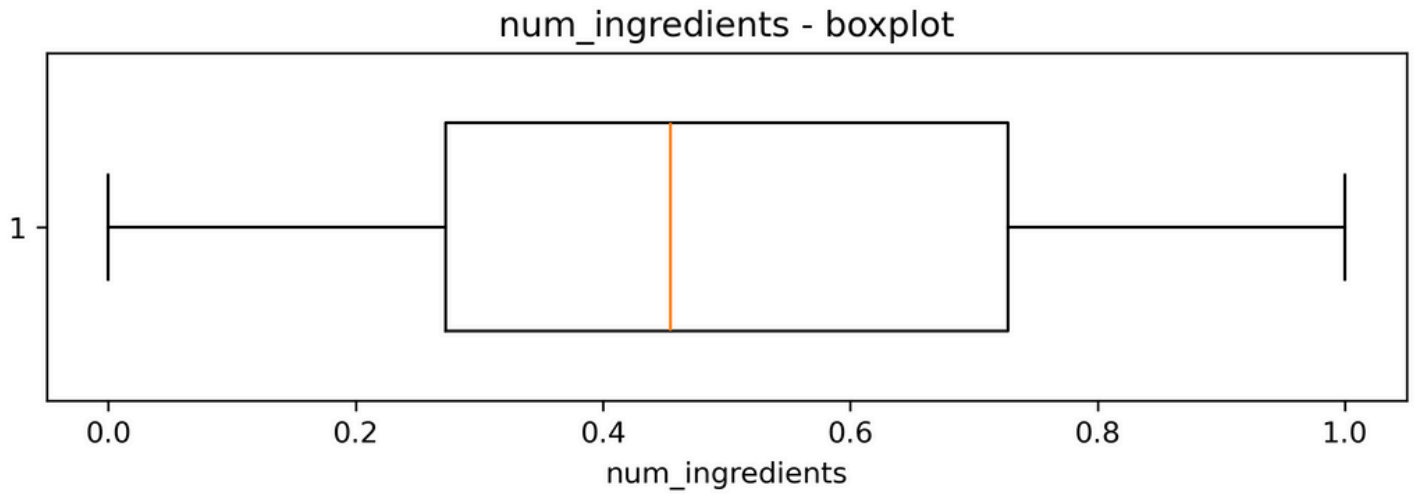

carbs - boxplot
1
0.0
0.2
0.4
0.6
0.8
1.0
carbs

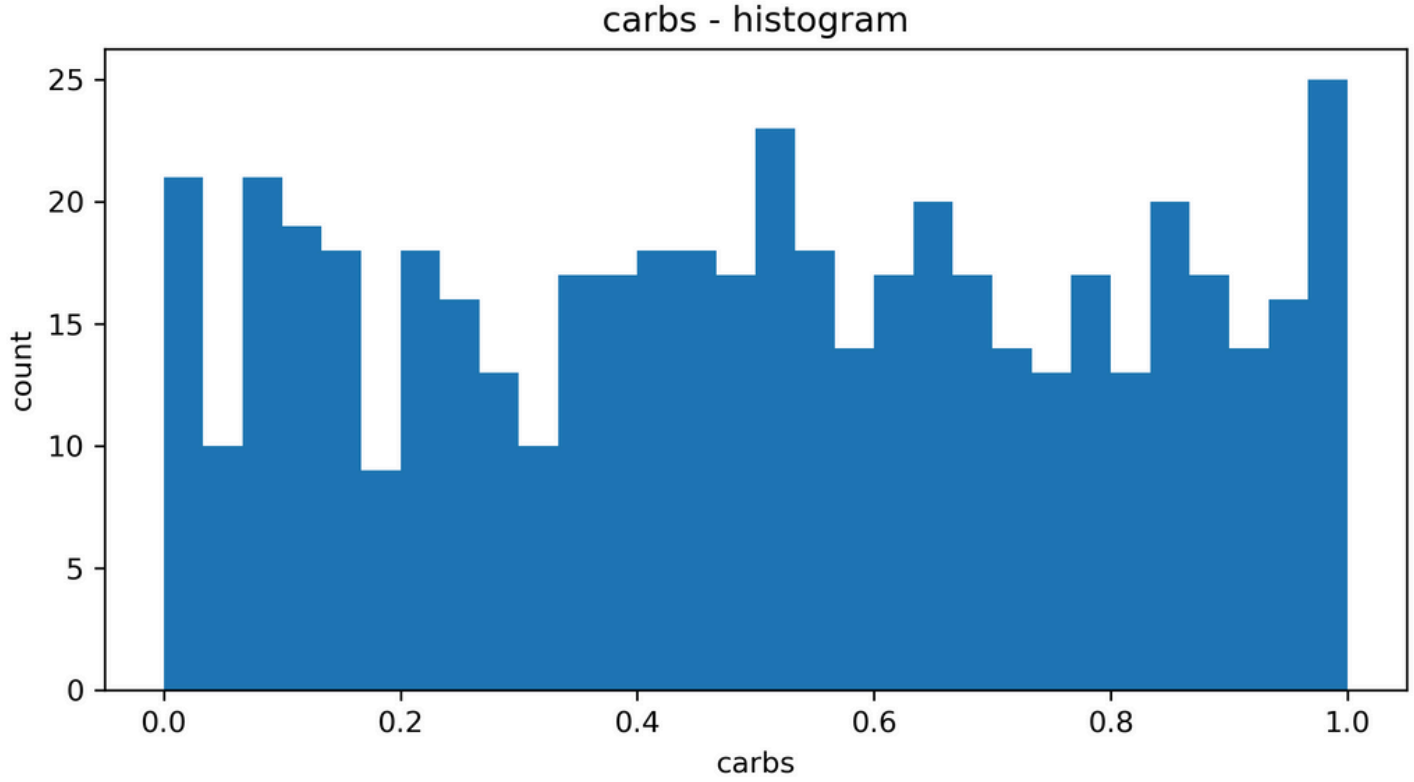

fat - boxplot
1
0.0
0.2
0.4
0.6
0.8
1.0
fat

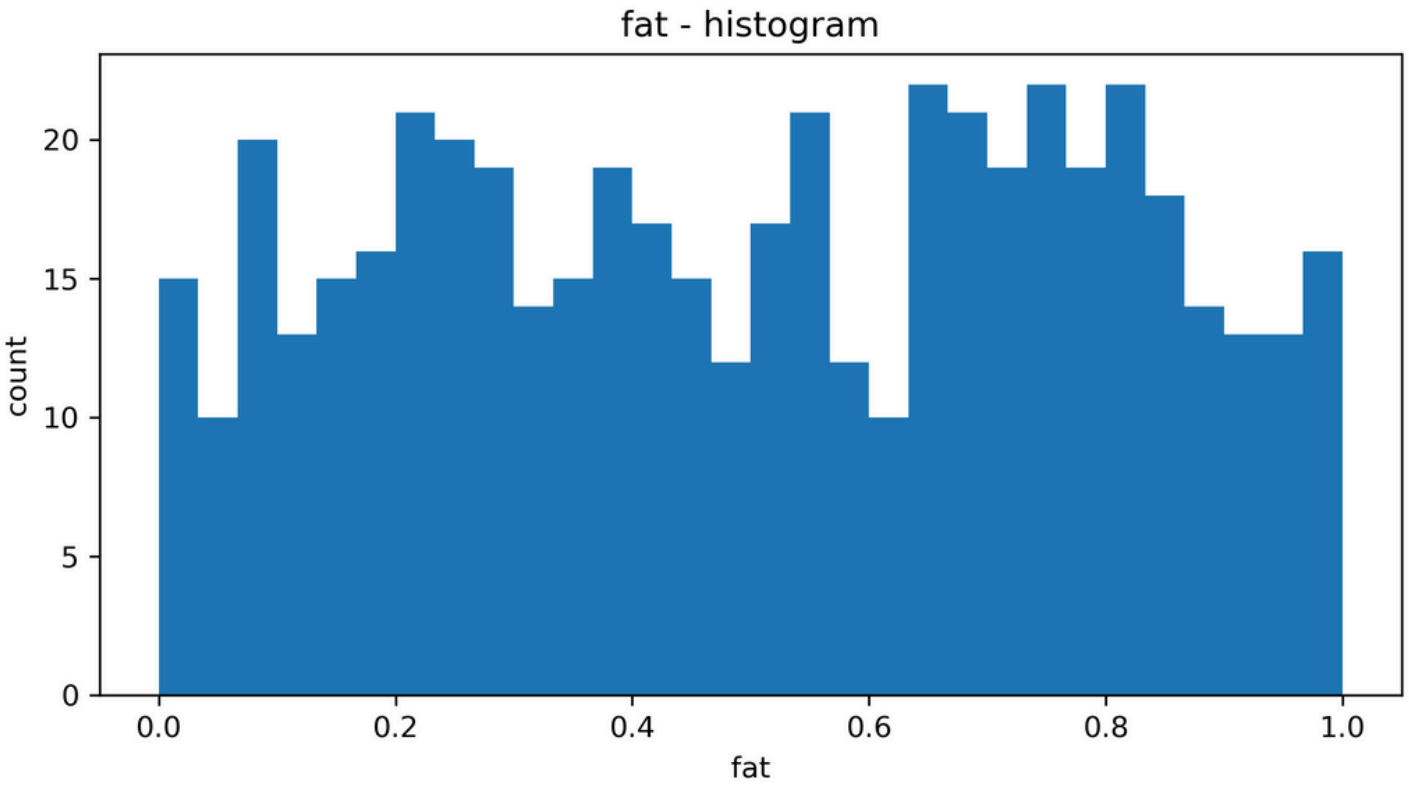

protein - boxplot
1
0.0
0.2
0.4
0.6
0.8
1.0
protein

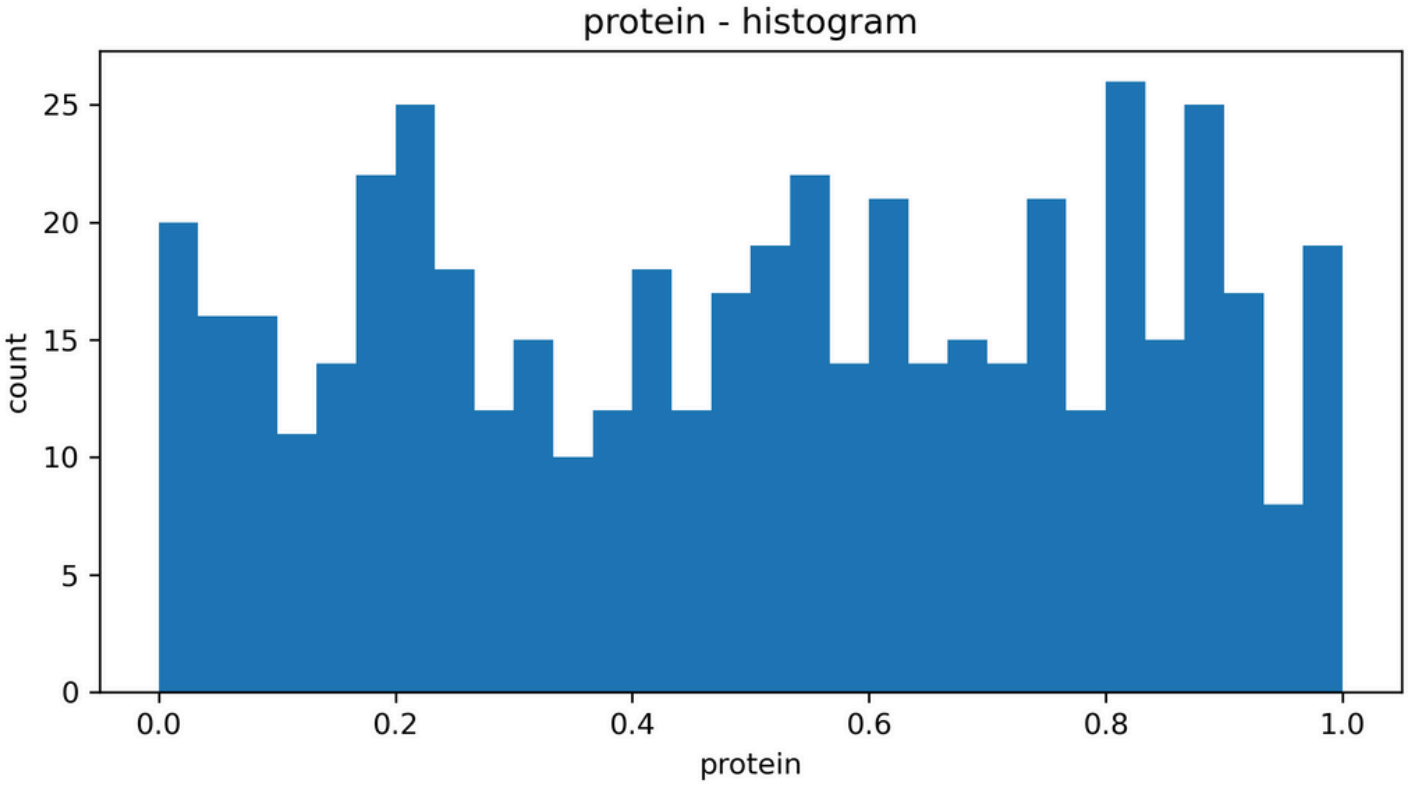

prep_time - boxplot
1
0.0
0.2
0.4
0.6
0.8
1.0
prep_time

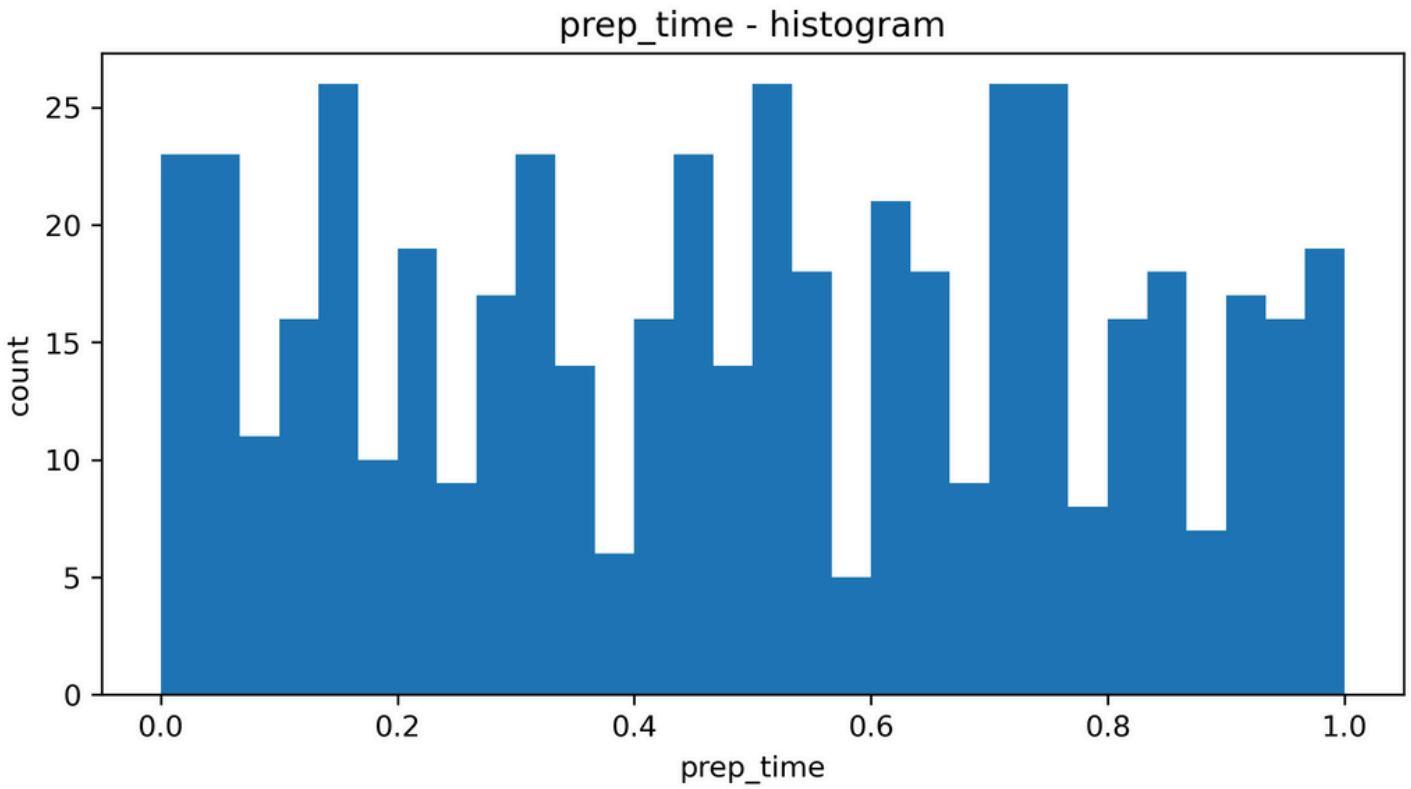

calories - boxplot
1
0.0
0.2
0.4
0.6
0.8
1.0
calories

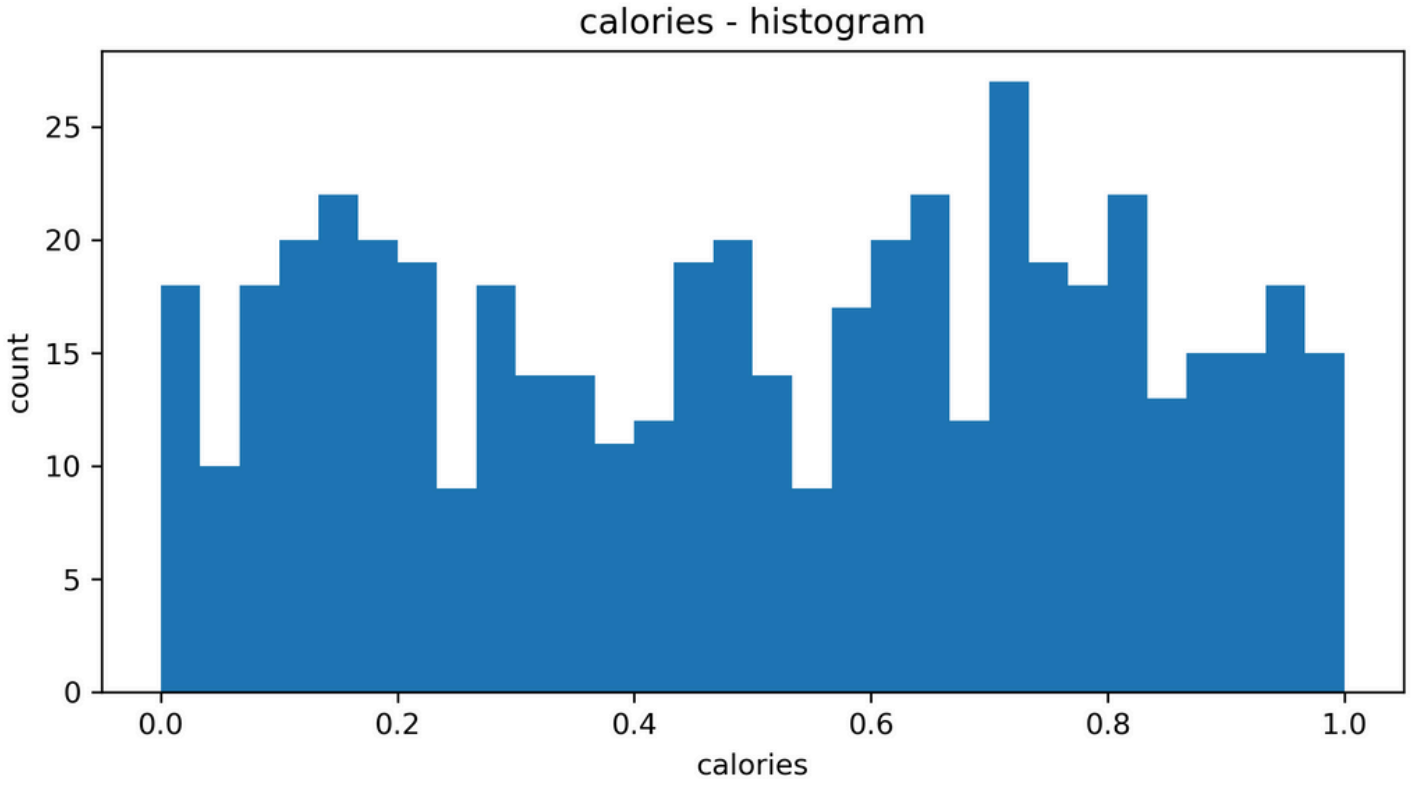

num_ingredients - boxplot
1
0.0
0.2
0.4
0.6
0.8
1.0
num_ingredients

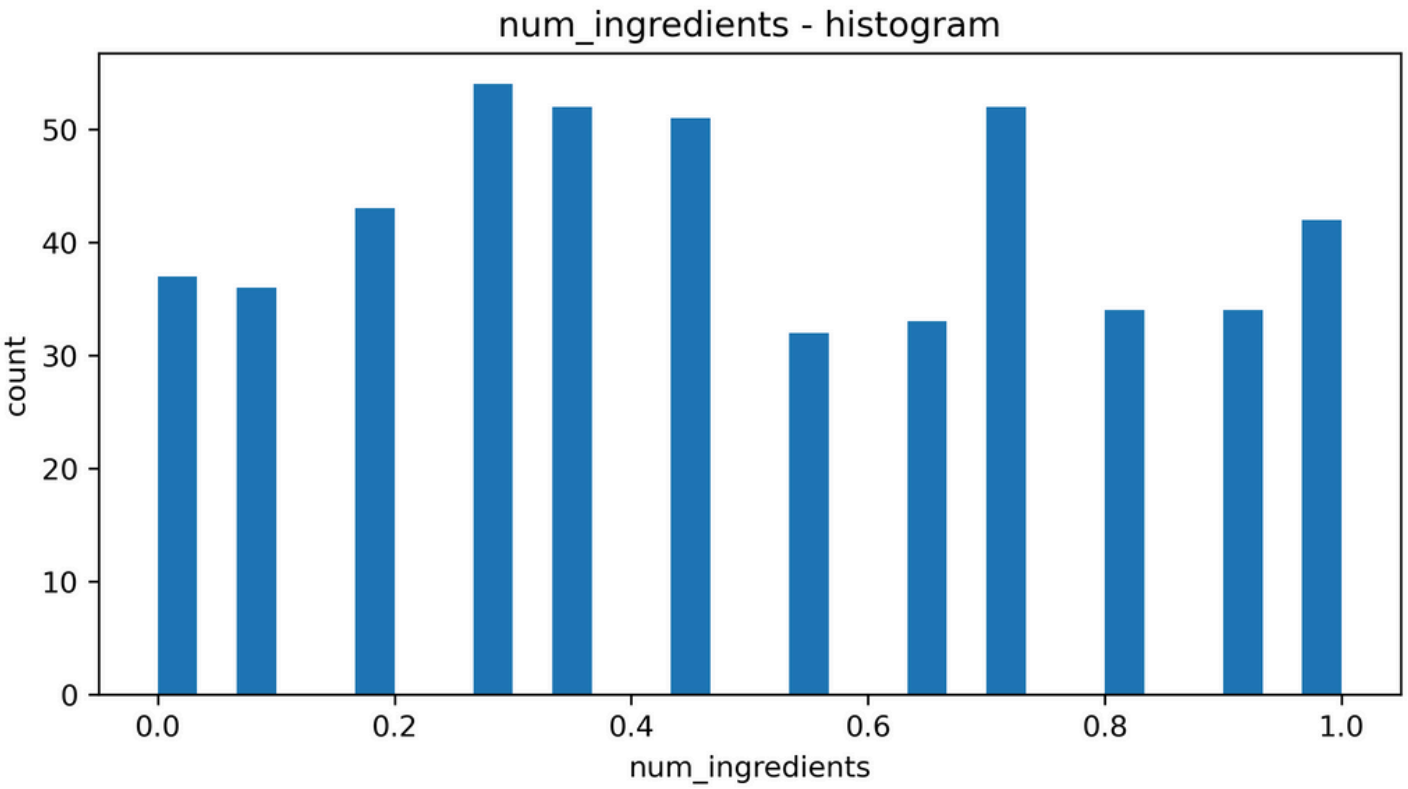

carbs - histogram
25
20
15
10
5
0
count
0.0
0.2
0.4
0.6
0.8
1.0
carbs


fat - histogram
count
fat
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
1.0


protein - histogram
count
protein
0
5
10
15
20
25
0.0
0.2
0.4
0.6
0.8
1.0


prep_time - histogram
count
prep_time
0
5
10
15
20
25
0.0
0.2
0.4
0.6
0.8
1.0


calories - histogram
count
calories
0
5
10
15
20
25
0.0
0.2
0.4
0.6
0.8
1.0


num_ingredients - histogram
count
num_ingredients
0
10
20
30
40
50
0.0
0.2
0.4
0.6
0.8
1.0