

## **Su Kalitesi Analizi ve Öneriler: Veri Odaklı Bir Yaklaşım**

### **Proje Amaçları**

- Su kalitesini etkileyen faktörleri belirlemek ve anlamak.
- Su kaynaklarının korunması ve su arıtma süreçlerinin iyileştirilmesi için veri odaklı çözümler geliştirmek.
- Su kalitesinin izlenmesi ve değerlendirilmesi için öneriler sunmak.
- Toplum sağlığını korumak ve çevresel sürdürülebilirliği desteklemek için su kalitesinin artırılmasına katkıda bulunmak.

Bu amaçlar doğrultusunda, projenin veri analizi ve sonuçları, su kaynaklarının korunması ve su kalitesinin iyileştirilmesine yönelik karar alınmasına ve politika oluşturulmasına katkıda bulunmayı hedeflemektedir.

### **Proje Kapsamı**

Bu proje, su kalitesi üzerine veri odaklı bir analiz sunmaktadır. Ana veri kaynağı olarak, çeşitli su numunelerinden elde edilen kimyasal ve mikrobiyolojik veri seti kullanılmıştır. Veri seti, suyun farklı bileşenlerinin (örneğin, kloramin, bakteri, virüsler vb.) konsantrasyonlarını içermektedir.

Analiz yöntemleri arasında, veri görselleştirme ve istatistiksel analizler bulunmaktadır. Veri setinin incelenmesi ve çözümlenmesi için Python programlama dili kullanılmıştır. Pandas, Seaborn ve Matplotlib gibi Python kütüphaneleri analizlerde kullanılan araçlar arasındadır.

Bu proje, su kalitesi analizinde veri odaklı bir yaklaşım sunarak su kaynaklarının yönetimi ve korunması için karar vericilere ve ilgili paydaşlara değerli bir kaynak sağlamayı hedeflemektedir.

# 1.Giriş

Bu dokümantasyon, suyun kimyasal bileşimini ve güvenliğini anlamak amacıyla oluşturulan waterQuality1 veri seti hakkında bilgi sağlamak üzere hazırlanmıştır. Veri seti, farklı kimyasal bileşenlerin ve mikroorganizmaların su numunelerindeki konsantrasyonunu içermektedir. Bu veri seti, suyun içilebilirliğini ve insan sağlığına olan etkilerini değerlendirmek için toplanmıştır.

Veri seti, toplamda 7998 su numunesinden elde edilen verileri içermektedir. Her bir su numunesi, 20 farklı özellik veya değişkeni temsil etmektedir. Bu özellikler arasında, alüminyum, amonyak, arsenik, bakır, kurşun gibi kimyasal elementlerin yanı sıra, bakteri ve virüs gibi mikrobiyal bileşenler de bulunmaktadır.

Bu dokümantasyon, suyun kimyasal bileşimi ve güvenliği ile ilgili bilgi edinmek isteyen araştırmacılar, mühendisler ve kamu sağlığı uzmanları için bir kaynak oluşturmayı amaçlamaktadır. Veri setinin içeriği, yapısı ve potansiyel kullanımı hakkında genel bir anlayış sağlamak için aşağıda verilen bilgiler ve analizler sunulmuştur.

## 2.Verii Tanımı

Bu veri seti, suyun kimyasal bileşimini ve güvenliğini değerlendirmek amacıyla toplanmıştır. Veri seti, farklı kimyasal elementlerin ve mikroorganizmaların farklı su numunelerindeki konsantrasyonunu içermektedir. Her bir su numunesi, bir dizi özellik veya değişkenle temsil edilir.

- **Alüminyum (aluminium):** Su numunelerindeki alüminyum konsantrasyonunu (mg/L) temsil eder. Alüminyum, suyun kirliliğiyle ilişkilendirilebilecek bir metaldir. Yüksek alüminyum seviyeleri, suyun tadını, kokusunu ve rengini etkileyebilir. Ayrıca, uzun vadeli maruziyet ciddi sağlık sorunlarına neden olabilir.
- **Amonyak (ammonia):** Su numunelerindeki amonyak konsantrasyonunu (mg/L) temsil eder. Amonyak, organik maddelerin ayrışması sırasında ortaya çıkan bir bileşiktir. Yüksek amonyak seviyeleri, suyun kalitesini olumsuz yönde etkileyebilir ve balıklar ve diğer su yaşamı için toksik olabilir.
- **Arsenik:** Su numunelerindeki arsenik konsantrasyonunu (mg/L) temsil eder. Arsenik, çeşitli sağlık sorunlarına neden olabilecek toksik bir elementtir. Yüksek arsenik seviyeleri, kanser riskini artırabilir ve ciddi deri lezyonlarına yol açabilir.
- **Barium:** Su numunelerindeki barium konsantrasyonunu (mg/L) temsil eder. Barium, endüstriyel faaliyetler ve maden atıkları gibi kaynaklardan suya sızabilir. Yüksek barium seviyeleri, kalp ritmi bozukluklarına ve diğer sağlık sorunlarına neden olabilir.
- **Kadmiyum:** Su numunelerindeki kadmiyum konsantrasyonunu (mg/L) temsil eder. Kadmiyum, toksik bir ağır metaldir ve sigara dumanı ve endüstriyel atıklar gibi kaynaklardan suya karışabilir. Yüksek kadmiyum seviyeleri, böbrek hasarı, kemik erimesi ve kanser gibi sağlık sorunlarına yol açabilir.
- **Kloramin:** Su numunelerindeki kloramin konsantrasyonunu (mg/L) temsil eder. Kloramin, su dezenfeksiyonunda yaygın olarak kullanılan bir kimyasaldır. Yüksek kloramin seviyeleri, sudaki organik maddelerin parçalanmasıyla trihalometanlar gibi zararlı yan ürünlerin oluşumuna yol açabilir.
- **Krom:** Su numunelerindeki krom konsantrasyonunu (mg/L) temsil eder. Krom, endüstriyel faaliyetlerden, madencilikten veya doğal kaynaklardan suya karışabilir. Yüksek krom seviyeleri, solunum yolu hastalıkları, cilt tahrişi ve kanser riski gibi sağlık sorunlarına neden olabilir.

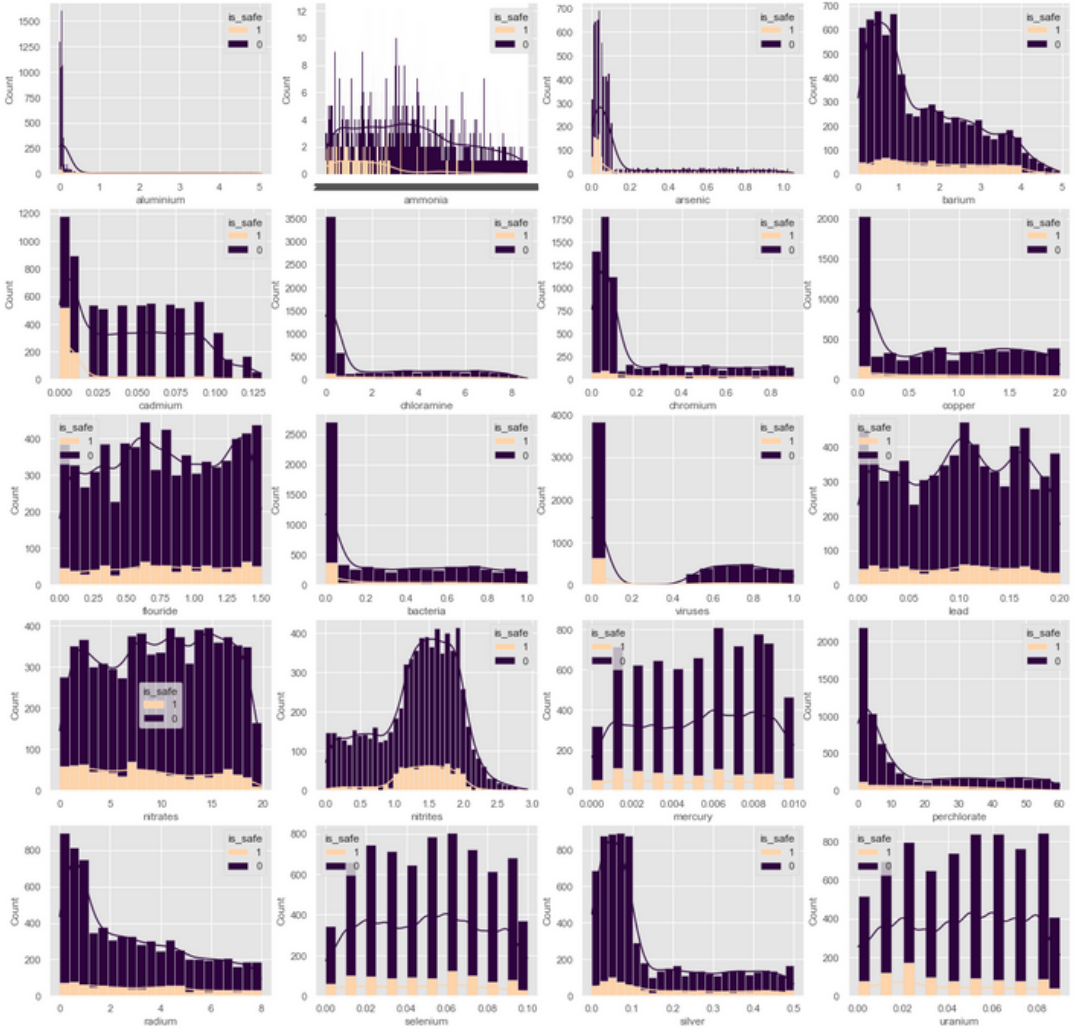
- **Bakır:** Su numunelerindeki bakır konsantrasyonunu (mg/L) temsil eder. Bakır, su borularından ve endüstriyel atıklardan kaynaklanabilir. Yüksek bakır seviyeleri, mide bulantısı, kusma ve karaciğer hasarı gibi sağlık sorunlarına yol açabilir.
- **Florür:** Su numunelerindeki florür konsantrasyonunu (mg/L) temsil eder. Florür, diş sağlığı için önemli bir mineraldir, ancak yüksek seviyeleri diş lekelerine ve kemik hastalıklarına neden olabilir.
- **Bakteri:** Su numunelerindeki bakteri miktarını (CFU/mL) temsil eder. Yüksek bakteri seviyeleri, suyun mikrobiyal kirliliğine işaret edebilir ve sindirim sistemi enfeksiyonlarına neden olabilir.
- **Virüs:** Su numunelerindeki virüs miktarını (PFU/mL) temsil eder. Virüsler, suyun mikrobiyal kirliliğinin bir göstergesi olabilir ve suyun güvenli olup olmadığını değerlendirmek için önemlidir.
- **Kurşun:** Su numunelerindeki kurşun konsantrasyonunu (mg/L) temsil eder. Kurşun, eski boruların ve endüstriyel atıkların bir sonucu olarak suya karışabilir. Yüksek kurşun seviyeleri, zehirlenme, sinir sistemi hasarı ve çocuklarda gelişim sorunları gibi ciddi sağlık etkilerine neden olabilir.
- **Nitratlar:** Su numunelerindeki nitrat konsantrasyonunu (mg/L) temsil eder. Nitratlar, gübrelerin ve hayvan atıklarının yıkımıyla suya karışabilir. Yüksek nitrat seviyeleri, bebeklerde mavi bebek sendromu riskini artırabilir ve sudaki nitratlar yoluyla tüketildiğinde sağlık sorunlarına neden olabilir.
- **Nitritler:** Su numunelerindeki nitrit konsantrasyonunu (mg/L) temsil eder. Nitritler, suyun kirlenmesinin bir göstergesi olabilir ve aşırı miktarda tüketildiğinde sağlık sorunlarına neden olabilir.
- **Cıva:** Su numunelerindeki cıva konsantrasyonunu (mg/L) temsil eder. Cıva, endüstriyel atıklar ve madencilik faaliyetleri gibi kaynaklardan suya karışabilir. Yüksek cıva seviyeleri, sinir sistemi hasarı, böbrek hasarı ve doğmamış bebeklerde gelişim sorunları gibi ciddi sağlık sorunlarına neden olabilir.
- **Perklorat:** Su numunelerindeki perklorat konsantrasyonunu (mg/L) temsil eder. Perklorat, gübreler, roket yakıtları ve patlayıcılar gibi endüstriyel kaynaklardan suya sızabilir. Yüksek perklorat seviyeleri, tiroid hormonlarını etkileyebilir ve sağlık sorunlarına neden olabilir.
- **Radon:** Su numunelerindeki radon konsantrasyonunu (pCi/L) temsil eder. Radon, toprak ve kayalarından gelen doğal bir gazdır ve yeraltı suyuna sızabilir. Yüksek radon seviyeleri, solunum yolu kanseri riskini artırabilir.
- **Selenyum:** Su numunelerindeki selenyum konsantrasyonunu (mg/L) temsil eder. Selenyum, bazı mineraller ve endüstriyel faaliyetlerden kaynaklanabilir. Yüksek selenyum seviyeleri, saç dökülmesi, sinir sistemi hasarı ve diğer sağlık sorunlarına neden olabilir.
- **Gümüş:** Su numunelerindeki gümüş konsantrasyonunu (mg/L) temsil eder. Gümüş, endüstriyel ve ticari faaliyetlerden kaynaklanabilir. Yüksek gümüş seviyeleri, cilt tahrişi ve solunum sorunları gibi sağlık sorunlarına neden olabilir.
- **Uranyum:** Su numunelerindeki uranyum konsantrasyonunu (mg/L) temsil eder. Uranyum, doğal kaynaklardan suya karışabilir. Yüksek uranyum seviyeleri, böbrek hasarı ve kanser riski gibi sağlık sorunlarına neden olabilir.

### 3.Keşifsel Veri Analizi (EDA)

#### Özellik Dağılımları

Öncelikle, veri setimizdeki özelliklerin dağılımlarını inceleyerek başlayalım. Histogramlar, özelliklerin frekans dağılımlarını görselleştirmek için kullanışlı bir araçtır. Aşağıdaki histogramlar, su numunelerinin farklı özelliklerinin dağılımını göstermektedir.

Her bir özellik için histogramlar incelendiğinde, bazı özelliklerin normal dağılıma yakın olduğunu, bazılarının ise sağa çarpık (positively skewed) olduğunu gözlemleyebiliriz. Bu bilgi, veri setimizdeki özelliklerin genel dağılımları hakkında bize önemli ipuçları vermektedir.

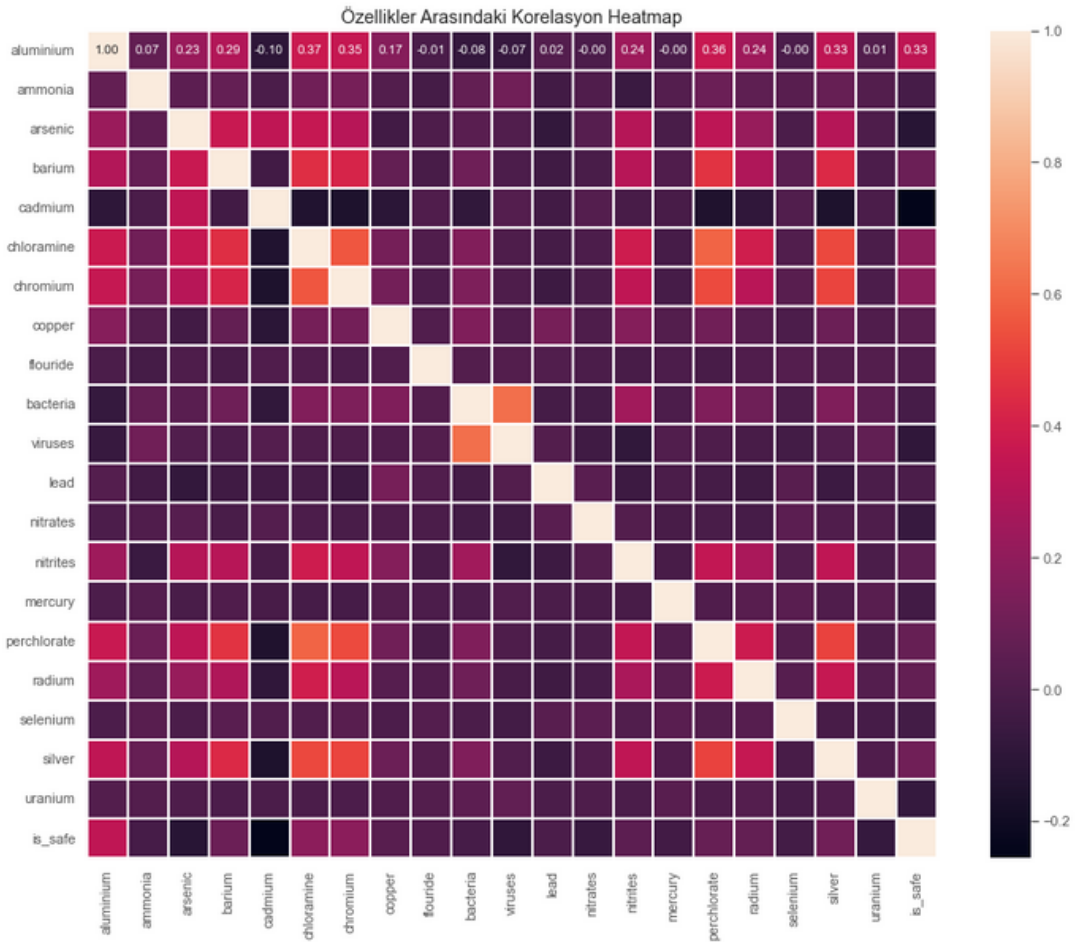


### Özellikler Arasındaki Korelasyon

Ayrıca, özellikler arasındaki korelasyonu da incelemeliyiz. Korelasyon, bir özelliğin diğerine olan bağımlılığını ölçer. Korelasyon matrisi, özellikler arasındaki korelasyonları görselleştirmek için kullanılan bir araçtır. Aşağıdaki heatmap, özellikler arasındaki korelasyonu göstermektedir.

Bu ısı haritasından, bazı özellikler arasında güçlü bir pozitif veya negatif korelasyon olduğunu gözlemleyebiliriz. Örneğin, bazı kimyasal bileşenler arasında pozitif bir korelasyon, diğer özellikler arasında ise negatif bir korelasyon olduğunu görebiliriz.

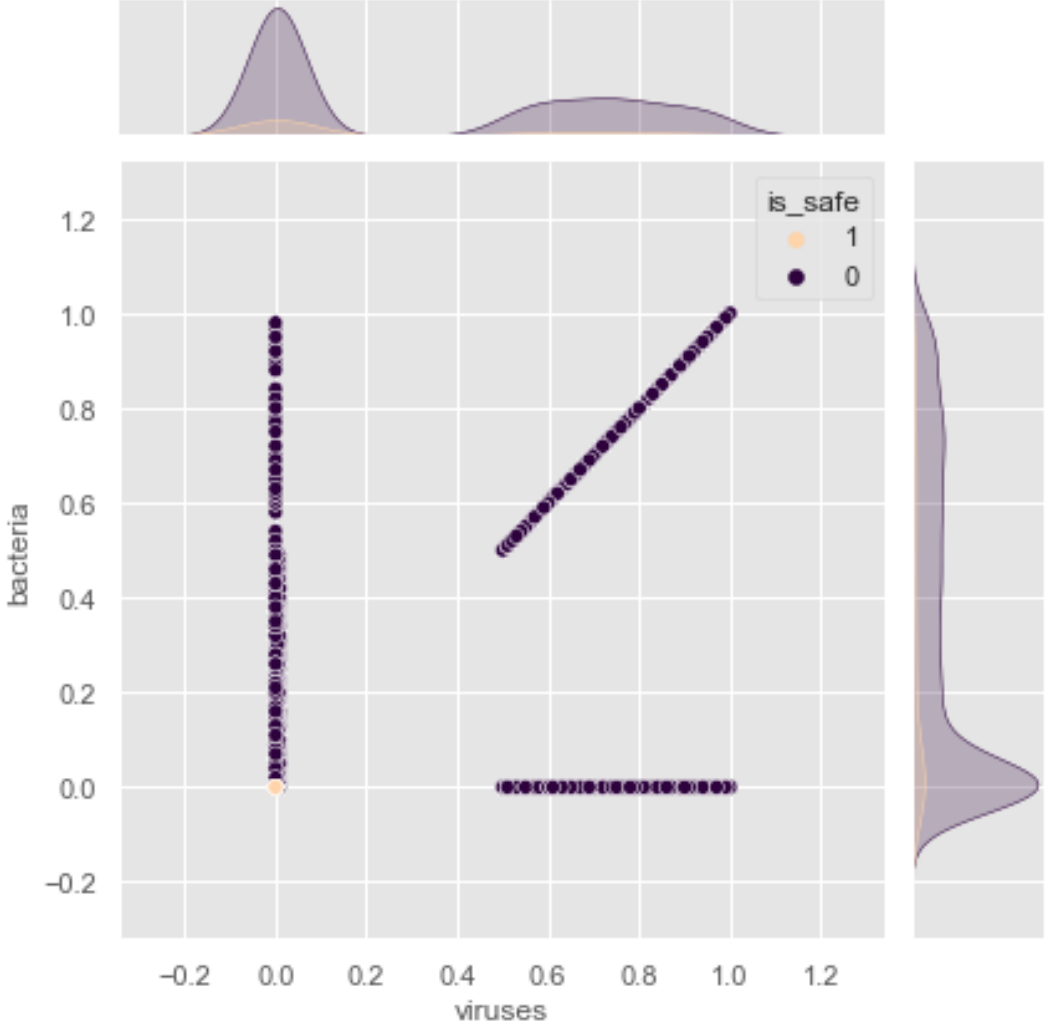
Bu keşifsel veri analizi sonuçları, veri setimizin daha derinlemesine analiz edilmesine ve gelecekteki modelleme çalışmalarına rehberlik etmek için önemlidir.



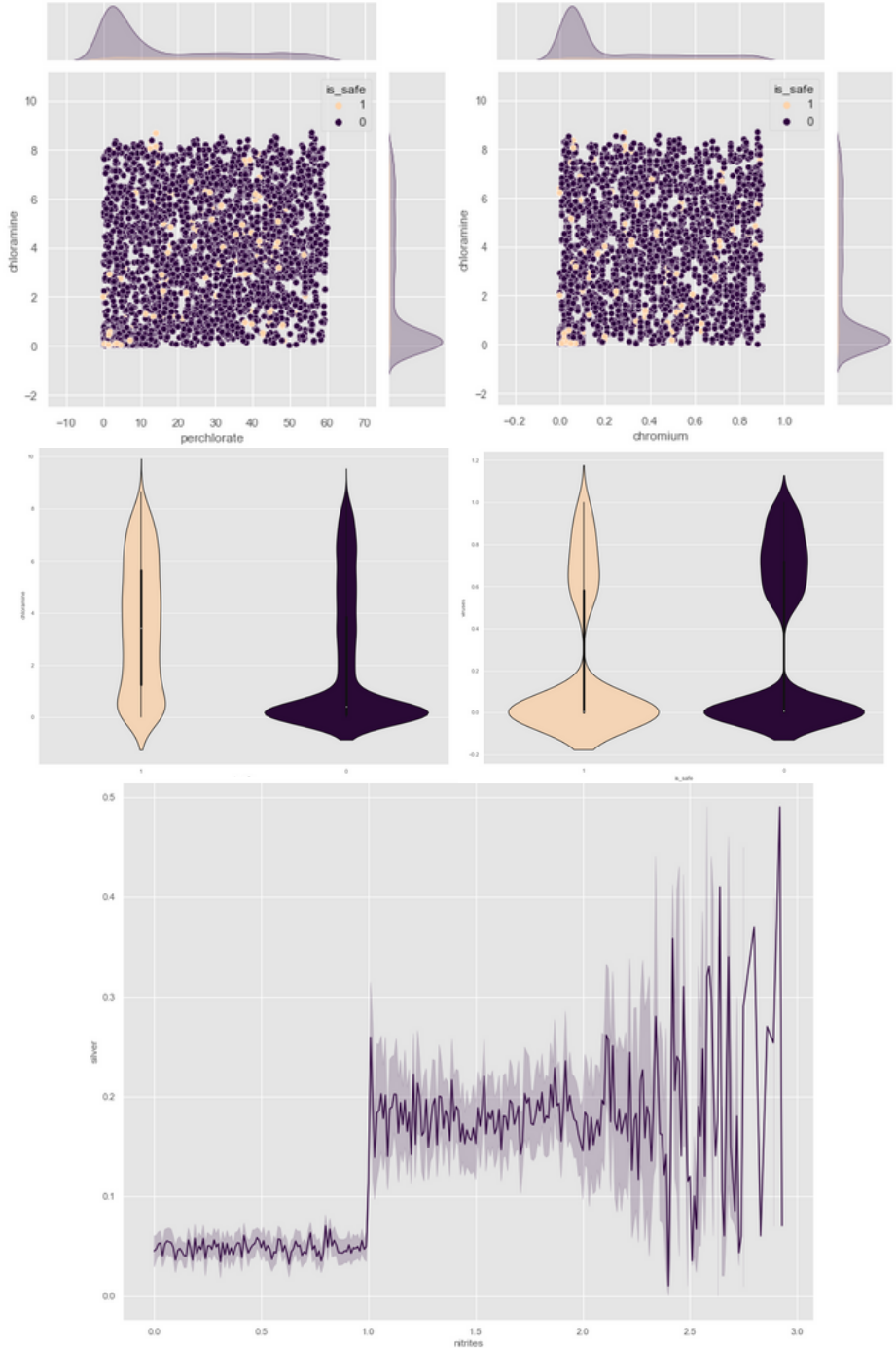
## Özellikler Arasındaki İlişkiler

Özellikler arasındaki ilişkileri anlamak da önemlidir. Örneğin, su numunelerindeki bakteri miktarı ile virüs miktarı arasındaki ilişkiyi inceleyebiliriz. Aşağıdaki scatter plot, bu ilişkiyi görselleştirmektedir.

Bu grafikten, bakteri miktarı ile virüs miktarı arasında bir ilişki olduğu, yüksek bakteri miktarına sahip numunelerin genellikle yüksek virüs miktarına sahip olduğu görülmektedir.



Analiz sürecinde oluşturulan diğer grafikler, verinin çeşitli yönlerini anlamak için kullanılmıştır. Bu grafikler, veri setindeki değişkenler arasındaki ilişkileri, dağılımları ve örüntüleri görselleştirerek verinin derinlemesine incelenmesine olanak sağlamıştır. Böylece, su kalitesi ile ilgili önemli özellikler ve potansiyel ilişkiler daha iyi anlaşılmış ve analiz sonuçlarına katkıda bulunmuştur.



## 4.Modelleme ve Analiz

Veri seti üzerinde çeşitli analiz tekniklerinin ve modellerinin uygulanabileceği çeşitli alanlar bulunmaktadır. Öncelikle, sınıflandırma analizi, veri setindeki özelliklerin belirli bir hedef değişkeni tahmin etmek için kullanılabilir. Örneğin, 'is\_safe' değişkeninin belirlenmesi için sınıflandırma modelleri uygulanabilir. Bunun yanı sıra, regresyon analizi, bir değişkenin diğer değişkenlerle olan ilişkisini belirlemek için kullanılabilir. Örneğin, su kalitesini etkileyen faktörleri incelemek için regresyon modelleri kullanılabilir.

Ayrıca, kümeleme analizi, veri setindeki benzer özelliklere sahip gözlemleri gruplamak için kullanılabilir. Su kalitesi üzerinde benzer etkilere sahip özelliklere sahip su numunelerini tanımlamak için kümeleme algoritmaları kullanılabilir. Bununla birlikte, derin öğrenme gibi daha gelişmiş yapay zeka teknikleri de su kalitesi tahmini için kullanılabilir.

Tüm bu analiz teknikleri ve modelleri, su kalitesi veri setindeki önemli desenleri, ilişkileri ve özellikleri keşfetmek ve gelecekteki su kalitesi tahminlerini geliştirmek için kullanılabilir.

## 5.Sonuçlar ve Öneriler

Analizlerimiz sonucunda su kalitesini belirleyen çeşitli faktörler üzerinde detaylı bir inceleme gerçekleştirildi. Özellikle, su numunelerindeki kloramin, virüsler ve bakteri konsantrasyonlarının su kalitesi üzerinde belirgin bir etkiye sahip olduğu gözlemlendi. Bu bulgular, su arıtma tesislerinin ve su kaynaklarının yönetimi konusunda önemli bir anlayış sağlamaktadır.

Kloramin, su arıtma proseslerinde yaygın olarak kullanılan bir dezenfektan olarak bilinmektedir. Yüksek kloramin seviyeleri, su kalitesinin düşük olabileceğini işaret edebilir ve su tüketiminde olumsuz etkilere neden olabilir. Aynı şekilde, su numunelerinde yüksek bakteri ve virüs konsantrasyonları, sağlık risklerinin artabileceğini göstermektedir. Analizler ayrıca su kalitesini değerlendirmek için kullanılan 'is\_safe' değişkeni ile diğer özellikler arasındaki ilişkileri de inceledi. Bu ilişkiler, suyun güvenliği açısından kritik öneme sahiptir ve su kaynaklarının korunması için alınacak tedbirlerde rehberlik edici olabilir.

Sonuç olarak, elde edilen bulgular su kaynaklarının korunması ve su arıtma tesislerinin verimliliğinin artırılması için stratejiler geliştirmek için önemli bir temel oluşturmaktadır. Su kalitesinin artırılmasıyla birlikte, toplum sağlığının korunması ve çevresel sürdürülebilirliğin desteklenmesi hedeflenmektedir.