



Sunucu Günlüklerine Dayalı E-Ticaret Web Sitesi Trafiki Analizi

Yazılım Mühendisliği Ana Bilim Dalı

Dönem Projesi

Halil İbrahim ÇİT

Y220240122

Proje Danışmanı: Dr. Emre ŞATIR

Haziran, 2024

Sunucu Gnlklerine Dayalı E-Ticaret Web Sitesi Trafıęi Analizi

z

Bu projede, e-ticaret web sitelerinin kullanıcı davranıřlarını ve trafik dinamiklerini anlamak amacıyla sunucu eriřim gnlkleri analiz edilmiřtir. ECLog veri seti kullanılarak yapılan analizler, web sitesinin performansını deęerlendirmek, kullanıcı eęilimlerini belirlemek ve gelecekteki alıřmalara yol gstermek iin gerekleřtirilmiřtir.

Analizler kapsamında gnlk ve saatlik trafik daęılımını, en sık yapılan aramalar, ynlendiren siteler, bařarılı ve bařarısız istekler, tekrar ziyaret oranları, kullanıcı trleri, trafik analizleri ve bot trafięi gibi eřitli unsurlar incelenmiřtir. Bu sayede, web sitesine hangi saatlerde daha fazla eriřim olduęu, kullanıcıların en ok ne aradıęı, hangi sitelerden ynlendirildikleri ve karřılařtıkları teknik sorunlar tespit edilmiřtir.

Elde edilen bulgular, web sitesi sahiplerinin ve pazarlama ekiplerinin stratejik kararlar almasına yardımcı olmayı amalamaktadır. Bu bilgiler, web sitesinin performansını optimize etmek, kullanıcı deneyimini iyileřtirmek ve pazarlama stratejilerini daha etkili hale getirmek iin kullanılacaktır.

Anahtar Szckler: Log analizi, Sunucu eriřim gnlkleri, Kullanıcı davranıřları, Trafik dinamikleri, E-ticaret.

E-Commerce Website Traffic Analysis Based on Server Logs

Abstract

In this project, server access logs were analyzed to understand user behaviors and traffic dynamics on e-commerce websites. Using the ECLog dataset, the analyses were conducted to evaluate the website's performance, identify user trends, and guide future efforts.

The analyses encompassed various elements such as daily and hourly traffic distribution, most frequent searches, referring sites, successful and failed requests, revisit rates, user types, traffic analyses, and bot traffic. This allowed the identification of peak access times, the most common user searches, the referring sites, and the technical issues encountered by users.

The findings are intended to assist website owners and marketing teams in making strategic decisions. These insights will be utilized to optimize website performance, enhance user experience, and improve marketing strategies.

Keywords: Log analysis, Server access logs, User behaviors, Traffic dynamics, E-commerce.

İçindekiler

Öz	1
Abstract	2
Şekiller Listesi.....	5
Tablolar Listesi.....	6
Kısaltmalar Listesi	7
1 GİRİŞ	8
2 MATERYAL VE YÖNTEM	10
3 LİTERATÜR TARAMA.....	12
3.1 Sunucu Günlükleri Analizi	12
3.2 Sunucu Günlükleri Analizinin Önemi	13
3.3 Sunucu Günlükleri Analizi İle İlgili Çalışmalar	14
3.4 Sunucu Günlükleri Analizinde Kullanılan Yöntemler.....	17
4 SUNUCU GÜNLÜKLERİNE DAYALI E-TİCARET WEB SİTESİ TRAFİĞİ ANALİZİ.....	18
4.1 Veri Setinin Özellikleri	18
4.2 Veri Hazırlığı ve Temizleme	19
4.3 Trafik Analizleri.....	24
4.3.1 Genel Trafik Dağılımı.....	24
4.3.2 Trafik Dağılımı Karşılaştırması	33
4.3.3 En Sık Yapılan Aramalar	40
4.3.4 Yönlendiren Siteler	42
4.3.5 Başarılı ve Başarısız İstekler.....	44
4.3.6 Tekrar Ziyaret Oranı	46
4.3.7 Kullanıcı Segmentasyonu ve Kümeleme Analizi	48
4.3.8 Bot ve Kullanıcı Trafik Karşılaştırması.....	51

4.3.9 Anomali Tespiti	54
4.3.10 İşletim Sistemi ve Tarayıcı Tercihleri	56
4.3.11 En Çok Ziyaret Edilen Ürünler ve Kategoriler.....	60
4.3.12 En Çok Satılan Ürünler ve En Popüler Kategoriler.....	63
4.3.13 Hafta Günlerine ve Aylara Göre Satın Alımlar	65
4.3.14 Kategori ve Ürün Bazında Satın Alımlar	68
4.3.15 Giriş ve Çıkış İşlemleri Analizi	69
5 SONUÇ ve ÖNERİLER.....	75
6 GELECEK ÇALIŞMALAR.....	77
7 KAYNAKÇA	78

Şekiller Listesi

Şekil 4.1	Günlük Trafik Verileri	26
Şekil 4.2	Haftanın Günlerine Göre Trafik Verileri	28
Şekil 4.3	Aylık Trafik Verileri	30
Şekil 4.4	Ülkelere göre trafik dağılımı	31
Şekil 4.5	İnsan ve bot istekleri için saatlik istek sayıları ile trafik dağılımı.....	34
Şekil 4.6	İnsan ve bot istekleri için haftalık istek sayıları trafik dağılımı.....	36
Şekil 4.7	İnsan ve bot istekleri için aylık istek sayıları trafik dağılımı	39
Şekil 4.8	En sık yapılan aramalar	41
Şekil 4.9	Yönlendiren siteler	43
Şekil 4.10	Başarılı ve başarısız İstekler.....	45
Şekil 4.11	Tekrar ziyaret oranı	47
Şekil 4.12	Kullanıcı segmentasyonu ve kümeleme analizi	50
Şekil 4.13	Bot ve kullanıcı trafiği karşılaştırma	53
Şekil 4.14	Anomali tespiti	55
Şekil 4.15	Kullanıcıların tercih ettikleri işletim sistemleri ve tarayıcı tercihleri	57
Şekil 4.16	İşletim sistemleri ve tarayıcı tercihleri	59
Şekil 4.17	En çok ziyaret edilen ürünler ve kategoriler	62
Şekil 4.18	En çok satılan ürünler ve en popüler kategoriler.....	64
Şekil 4.19	Haftanın günlerine göre ve aylara göre satın almalar	67
Şekil 4.20	Kategori ve ürün bazında satın alımlar	69
Şekil 4.21	Giriş işlemleri dağılımı.....	70
Şekil 4.22	Çıkış işlemleri dağılımı	71
Şekil 4.23	Giriş çıkış işlemleri karşılaştırılması.....	73

Tablolar Listesi

Tablo 4.1	Veri seti genel bilgilerinin gösterildiği tablo.....	19
Tablo 4.2	Örnek beş satırın gösterildiği tablo	20
Tablo 4.3	Kategori tipinin dönüşümünün gösterildiği tablo.....	22
Tablo 4.4	Günlük trafik verilerinin gösterildiği tablo	25
Tablo 4.5	Haftanın günlerine göre trafik verilerinin gösterildiği tablo	27
Tablo 4.6	Aylık trafik verilerinin gösterildiği tablo	29
Tablo 4.7	Ülkelere göre trafik dağılımının gösterildiği tablo.....	31
Tablo 4.8	Saatlik trafik dağılımının gösterildiği tablo	33
Tablo 4.9	Hafta günleri trafik dağılımının gösterildiği tablo	36
Tablo 4.10	Aylık trafik dağılımının gösterildiği tablo	38
Tablo 4.11	En sık yapılan aramaların gösterildiği tablo.....	40
Tablo 4.12	Yönlendiren sitelerin dağılımının gösterildiği tablo	42
Tablo 4.13	Başarılı başarısız isteklerin gösterildiği tablo	44
Tablo 4.14	Tekrar ziyaret oranının gösterildiği tablo.....	46
Tablo 4.15	Kullanıcı segmentasyonu ve kümeleme analizinin gösterildiği tablo.....	49
Tablo 4.16	Bot ve kullanıcı trafiği karşılaştırmasının gösterildiği tablo.....	52
Tablo 4.17	Anomali tespitinin gösterildiği tablo.....	54
Tablo 4.18	İşletim sistemi tercihlerinin gösterildiği tablo.....	56
Tablo 4.19	Tarayıcı tercihlerinin gösterildiği tablo.....	57
Tablo 4.20	İşletim sistemleri ve tarayıcı tercihlerinin gösterildiği tablo.....	59
Tablo 4.21	En çok ziyaret edilen ürünlerin gösterildiği tablo	61
Tablo 4.22	En çok ziyaret edilen kategorilerin gösterildiği tablo	61
Tablo 4.23	En çok satılan ürünlerin ürün koduna göre gösterildiği tablo	63
Tablo 4.24	En popüler kategorilerin gösterildiği tablo.....	64
Tablo 4.25	Hafta günlerine göre satın alımların gösterildiği tablo.....	66
Tablo 4.26	Aylara göre satın alımların gösterildiği tablo.....	66
Tablo 4.27	Kategori ve ürün bazında satın alımların gösterildiği tablo	68

Kısaltmalar Listesi

HTTP	Hyper Text Transfer Protocol
SEO	Search Engine Optimization
SIEM	Security Information and Event Management
OS	Operating System
EClog	E-Commerce Log
Wroclaw	Wroclaw University of Science and Technology
URI	Uniform Resource Identifier
IP	Internet Protocol
PANDAS	Python Data Analysis Library
DATETIME	Python Date and Time Library
RE	Regular Expressions
PARSE	User Agents Parsing Library
URLPARSE	URL Parsing Library
PYCOUNTRY	Python Country Library
STANDART SCALER	Scikit-Learn Standard Scaler
KMEANS	Scikit-Learn K-Means Clustering
ISOLATION FOREST	Scikit-Learn Isolation Forest

Bölüm 1

Giriş

E-ticaret sektörü, internet kullanımının yaygınlaşmasıyla birlikte büyük bir dönüşüm yaşamış ve geleneksel alışveriş yöntemleri hızla dijital platformlara yerini bırakmıştır. Bu dönüşüm hem işletmeler hem de tüketiciler için yeni fırsatlar ve zorluklar yaratmış, e-ticaret siteleri müşterilere geniş ürün yelpazesi sunmanın yanı sıra alışveriş deneyimlerini kişiselleştirme ve anlık geri bildirimlerle hizmet kalitesini artırma imkânı sağlamıştır. Ancak, bu avantajları etkin bir şekilde kullanabilmek için web trafiği analizi kritik hale gelmiştir. Web trafiği analizi, kullanıcı davranışlarını anlamak, pazarlama stratejilerini optimize etmek ve web sitesinin performansını iyileştirmek için büyük öneme sahiptir.

Bu projenin temel amacı, sunucu günlüklerine dayalı olarak bir e-ticaret web sitesinin trafiğini analiz etmektir. Sunucu günlükleri, bir web sitesine yapılan tüm isteklerin kaydedildiği dosyalar olup, bu günlükler üzerinden yapılan analizler, kullanıcıların web sitesinde nasıl etkileşimde bulunduğunu detaylı bir şekilde ortaya koyabilir. Bu analizler sayesinde, web sitesinin hangi günlerde daha fazla ziyaret edildiği, kullanıcıların hangi ürünlere daha çok ilgi gösterdiği ve web sitesine hangi saatlerde daha yoğun ilgi olduğu belirlenebilecektir. Ayrıca, kullanıcıların web sitesine hangi dış kaynaklardan yönlendirme aldıkları ve sitenin farklı bölümlerine birden fazla kez mi yoksa tek bir ziyaretle mi eriştikleri gibi bilgiler de ortaya çıkarılabilir.

Projenin kapsamı, sunucu günlüklerinin belirlenen kriterler doğrultusunda detaylı bir şekilde incelenmesini içermektedir. Bu inceleme ile kullanıcıların web sitesinde nasıl etkileşimde bulunduğu, hangi ürünlerin daha fazla ziyaret edildiği ve hangi aramaların yapıldığı gibi bilgiler elde edilecektir. Ayrıca, sayfaya hangi diğer sitelerden yönlendirme yapıldığı ve kullanıcıların sayfayı bir veya daha fazla kez ziyaret edip etmediği gibi veriler de analiz edilecektir. Kullanıcı davranışlarını daha detaylı anlamak ve belirli kriterlere göre kullanıcı verilerini gruplandırmak, projenin önemli bir parçasıdır. Bu gruplandırmalar, farklı kullanıcı türlerinin ihtiyaçlarını ve beklentilerini daha iyi anlamamıza yardımcı olacaktır.

Proje kapsamında kullanılan veri seti [1], Wroclaw Bilim ve Teknoloji Üniversitesi tarafından sağlanan EClog veri setidir. Bu veri seti, 1 Aralık 2019-31 Mayıs 2020 tarihleri arasında bir e-ticaret sitesinde kaydedilen HTTP düzeyindeki sunucu erişim günlüklerini içermektedir. EClog veri seti, toplam 35.157.691 HTTP isteğini kapsamakta olup, kullanıcıların web sitesine yaptıkları isteklerle ilgili ayrıntılı bilgiler sunmaktadır. Bu bilgiler arasında IpId, UserId, TimeStamp, HttpMethod, Uri, HttpVersion, ResponseCode, Bytes, Referrer ve UserAgent gibi çeşitli veriler bulunmaktadır. Bu zengin veri seti, bazı temizleme ve küçültme işlemlerinden geçirilerek e-ticaret sitesinin trafik analizi için kullanılacaktır. Analiz süreci, e-ticaret sitesinin kullanıcı deneyimini artırmak, satışları optimize etmek ve genel performansını iyileştirmek için değerli bilgiler sağlayacaktır. Bu sayede, e-ticaret sitesi rekabet avantajını artırarak, daha geniş bir kullanıcı kitlesine ulaşabilir ve müşteri memnuniyetini artırabilir.

Bölüm 2

Materyal ve Yöntem

Bu çalışmada, Wrocław Bilim ve Teknoloji Üniversitesi tarafından sağlanan EClog veri seti kullanılmıştır. Veri seti, e-ticaret sitesinde kaydedilen HTTP düzeyindeki sunucu erişim günlüklerini ve HTTP isteklerini içermekte ve kullanıcıların web sitesine yaptıkları isteklerle ilgili IpId, UserId, TimeStamp, HttpMethod, Uri, HttpVersion, ResponseCode, Bytes, Referrer ve UserAgent gibi çeşitli bilgileri barındırmaktadır.

Veri ön işleme aşamasında, EClog veri seti hatalı veya eksik kayıtların temizlenmesi için incelenmiş, gereksiz veriler kaldırılmış ve eksik veriler uygun yöntemlerle tamamlanmıştır. Ayrıca, zaman damgası verileri, daha kolay analiz edilebilmesi için standart bir zaman formatına dönüştürülmüştür.

Veri analizi aşamasında, veri setinin genel özelliklerini anlamak için temel istatistiksel analizler yapılmış, bu analizler ziyaret sayısı ve en çok ziyaret edilen sayfalar gibi metrikleri içermiştir. Kullanıcı davranışlarını daha iyi anlamak için çeşitli segmentasyon yöntemleri uygulanmış, örneğin coğrafi konum (ülke), işletim sistemi (OS) ve ziyaret sıklığı gibi kriterlere göre segmentler oluşturulmuştur.

Görselleştirme aşamasında, elde edilen veriler kullanıcı davranışlarını ve trafik eğilimlerini görselleştirmek için çeşitli grafikler ve tablolar kullanılarak sunulmuştur. Görselleştirmeler ziyaretçi akışı ve dönüşüm oranlarını içermektedir.

Modelleme aşamasında zaman serisi analizleri yapılmış, bu analizler web trafiğindeki trendleri belirlemek için kullanılmıştır. Ayrıca, kullanıcı davranışlarını ve web sitesi performansını optimize etmek için çeşitli makine öğrenimi algoritmaları uygulanmış, bu algoritmalar kullanıcı segmentasyonu ve anomali tespiti gibi alanlarda kullanılmıştır.

Bu süreçlerde, veri işleme, analiz ve görselleştirme için Python programlama dili kullanılmış, veri manipülasyonu ve analizi için pandas, datetime, re, user_agents parse,

urllib. parse urlparse, pycountry kütüphaneleri, veri görselleştirme için hvplot. pandas kütüphanesi, makine öğrenimi modelleri için Scikit-Learn StandardScaler, KMeans ve IsolationForest kütüphanesi ve analiz ile görselleştirme işlemlerini yürütmek için Jupyter Notebook ortamı kullanılmıştır. Bu yöntemler, sunucu günlüklerine dayalı olarak e-ticaret web sitesinin trafik analizini kapsamlı bir şekilde gerçekleştirmeyi ve elde edilen sonuçlarla web sitesinin performansını ve kullanıcı deneyimini iyileştirmeyi amaçlamaktadır.

Bölüm 3

Literatür Tarama

3.1 Sunucu Günlükleri Analizi

Yaşanan teknolojik gelişmeler ve maliyetlerin düşmesiyle internet, günlük yaşamın vazgeçilmez bir unsuru haline gelmiş ve insanlar bilgi ihtiyaçlarının büyük bir kısmını internet üzerinden karşılamaya başlamıştır. Bu gelişmeler kurum ve kuruluşların hizmet verme biçim ve anlayışlarında da değişikliklerin oluşmasını sağlamıştır.

Kurum ve kuruluşlar, interneti reklam, e-ticaret, iletişim, bilgi-belge paylaşımı, bankacılık işlemleri ve eğitim gibi çeşitli alanlarda kullanmakta ve bu amaçlarla web siteleri oluşturmaktadır. Her gün kurumlara ait bu web sitelerinde çok sayıda log üretilmektedir. Bununla birlikte sunucu ve istemcilerin logları da eklendiğinde bu veriler arasından önemli ve gerekli bilgilerin süzülmesi, verilerin anlamlandırılarak bilgiye dönüştürülmesi, çeşitli işlem ve trafik bilgileri ile ilişkilendirilmesi, web sitelerini ziyaret eden kullanıcıların davranışlarının incelenerek bu web sitelerinin güncellenmesi veya geliştirilmesi için analizinin yapılması gerekmektedir. [1]

Kullanıcıların site içerisinde hangi sayfaya girdikleri, siteyi hangi aşamadayken terk ettikleri, sitede ne kadar (daha az – daha çok) zaman harcadıkları gibi bilgilere kolaylıkla erişilebilmekte ve kullanıcıların zaman geçirdiği, alışveriş yaptığı, işlem yürüttüğü veya gönderim yaptığı alanlardan (bankacılık işlemleri vb. gibi) hareketle ilgi alanlarının belirlenmesinde, sayfa kullanım dağılımlarının belirlenmesinde, alışveriş temelinde hazırlanan bir web sitesi ise daha etkili pazarlama stratejilerinin oluşturulmasında, son olarak reklam alma gibi birçok konuda doğru karar verilmesine katkı sağlayan bilgiler bu analizler ile elde edilebilmektedir.

Başka bir çalışmada sunucu günlükleri analizi şu şekilde ifade edilmiştir: “Çok sayıda kullanıcı hakkında davranışsal veri toplamaya yönelik uygun ve basit bir yol” [2]

Sunucu günlükleri analizi, genellikle ilk başta fark edilmeyen insan davranışlarını gözlemlenme, kullanıcı arayüzlerinin değerlendirilmesi ve geliştirilmesi, ek olarak sistem performansı ve kullanımı ile bilgi arama davranışlarını istatistiksel olarak analiz etme ve bu süreçte ortaya çıkan sorunları belirleyip önlem alma ve ilgili konudaki web sitesinin geliştirilmesi amacıyla önemli rol oynamaktadır.

3.2 Sunucu Günlükleri Analizinin Önemi

Sunucu günlükleri analizi, bir organizasyonun bilişim altyapısının güvenliğini sağlamak, performansını izlemek ve verimliliğini artırmak için kritik bir öneme sahiptir. Bu analiz, sunucuların işletim durumunu, kullanıcı etkileşimlerini, sistem kaynaklarının kullanımını ve diğer önemli metrikleri izlemeye olanak tanımaktadır.

Sunucu günlüklerinin analizi, potansiyel tehditlerin tespit edilmesi, sistemin dengesizliklerinin belirlenmesi ve hata ayıklama süreçlerinin yönlendirilmesi gibi bir dizi alanda organizasyonlara değerli bilgiler sağlamaktadır. Ayrıca, bu analiz, yasal uyumluluk gerekliliklerini yerine getirme, olayların kökenini belirleme ve kriz durumlarına hızlı bir şekilde yanıt verme gibi önemli güvenlik ve yönetim gereksinimlerini desteklemektedir.

Hem sistem hem de kullanım geri bildirimleri açısından büyük hacimli ve önemli bir bilgi kaynağı olan ve sunuculardan gelen log kayıtlarının sadece saklanması, organizasyonlar için risk yönetimi, stratejik karar verme ve faaliyet programlama gibi çeşitli boyutlarda herhangi bir avantaj sağlamamaktadır. Önemli olan, log kayıtlarının analitik yöntemlerle incelenmesi ve verilerin kurumsal ve stratejik karar alma süreçlerinde önemli bir bilgi kaynağına dönüştürülmesidir. Bu şekilde, organizasyonlar faaliyetlerini yönetme konusunda daha bilinçli kararlar alabilirler. [3]

Farklı formatlardaki log kayıtları, üzerinde buldukları sistem hakkında birçok bilgiyi tutmasından ve bu bilgiler ile adli olayların aydınlatılması, vb. gibi durumlarda önemli sonuçların alınmasına etki etmesi gibi durumlardan hareketle bir uçağın kara kutusuna benzetilebilmektedir.

Sunucu üzerinde saklanan erişim kayıtları, başlangıçta herhangi bir müdahale veya düzenleme olmadan incelendiğinde, anlamsız görünebilen kayıtlardır. Bu kayıtların

karmaşık ve düzensiz bir yapıda olduğu gözlemlenir. Ancak, bu veriler kullanıcı etkileşimlerini ve davranışlarını dikkate alarak ele alındığında, anlamlı bilgiler elde edilir ve veriler analiz edilerek anlam kazanır. Bu süreçte, ilgili erişim kayıtları incelenirken; erişim kayıtlarının genel bilgileri, site ziyaretçilerinin kullandığı tarayıcıların dağılımı, kullanıcıların ziyaret ettiği ülkelerin coğrafi dağılımı, web sitesine erişim dağılımları (günlük, haftalık, aylık), kullanıcıların sayfalarda geçirdikleri süre, en popüler giriş yapılan sayfalar, en çok ziyaret edilen sayfalar, log trafiğinin dağılımı, yönlendirme ile gerçekleşen sayfa ziyaretleri gibi istatistiksel bilgiler elde edilebilmektedir.

Başka bir çalışmada Sunucu günlüklerinin analiz edilmesinin önemi ile ilgili şu görüş paylaşılmıştır:

“Log kayıtları; kullanıcıların sistemdeki etkileşimlerini, gerçekleştirdikleri işlemleri ve diğer faaliyetleri detaylı olarak kaydeden veri dosyaları olmasından hareketle; işlemsel profil tespiti için önemli bir kaynak oluşturur çünkü kullanıcıların davranışlarını ve alışkanlıklarını analiz etmek için kritik bilgiler içermektedirler. Log kayıtlarının analizi, organizasyonlara kullanıcıların sistemle nasıl etkileşimde bulunduğunu, hangi faaliyetleri gerçekleştirdiklerini ve hangi eğilimleri sergilediklerini daha iyi anlama fırsatı sunmaktadır. Bu sayede organizasyonlar kullanıcı deneyimini iyileştirebilmekte, güvenlik önlemlerini güçlendirerek, iş süreçlerini daha verimli hale getirebilmektedir. Dolayısıyla, log kayıtları analizi, bu gibi işlemlerin yürütülmesinde kritik bir öneme sahiptir ve organizasyonların karar alma süreçlerinde temel bir rol oynamaktadır.” [4]

3.3 Sunucu Günlükleri Analizi İle İlgili Çalışmalar

Sunucu günlükleri analizi, çeşitli alanlarda yapılan çalışmalarla yoğun bir şekilde ele alınmaktadır. Bu çalışmalar genellikle güvenlik analizi, performans izleme, hata ayıklama, yasal uyumluluk ve kullanım analizi gibi konulara odaklanmaktadır. Güvenlik analizi, sunucu günlüklerini kötü amaçlı yazılımların, saldırıların ve yetkilendirilmemiş erişim girişimlerinin belirlenmesi ve önlenmesi için kullanılırken; performans izleme çalışmaları, sunucu günlüklerini sistem performansının izlenmesi ve performans sorunlarının tanımlanması için kullanılmaktadır. Hata ayıklama ve

sorun giderme arařtırmaları sunucu gnlklerini sistemdeki hataları tanımlamak ve zmek iin kullanılırken yasal uyumluluk ve denetimler zerine yapılan alıřmalar, sunucu gnlklerini yasal dzenlemelere uygunluk saęlamak ve denetim gerekliliklerini karřılamak iin kullanılmaktadır. Son olarak, kullanım analizi arařtırmaları, sunucu gnlklerini kullanıcı etkileřimlerini izlemek ve kullanıcı davranıřlarını anlamak iin kullanılmaktadır. Bu alıřmalar, sunucu gnlklerinin farklı kullanım alanlarını ve potansiyelini ortaya koyar ve iřletmelerin bilgi teknolojisi altyapılarını daha etkin bir řekilde ynetmelerine yardımcı olmaktadır. [5]

Gvenlik analizi, performans izleme, hata ayıklama, yasal uyumluluk ve kullanım analizi gibi konulara odaklanıp, eřitli alanlarda yoęun bir řekilde incelemelerle sunucu gnlklerinin farklı kullanım alanlarını ve potansiyelini ortaya koyarak, iřletmelerin bilgi teknolojisi altyapılarını daha etkin bir řekilde ynetmelerine yardımcı olmak maksadıyla sunucu trafięinin ve kullanıcı davranıřlarının analizinin gerekleřtirildięi aralar bulunmaktadır. alıřmada bu aralardan Google Analytics ve Clickstream Analizi mantıęıyla Log Dosyası Analizi ele alınmıřtır.

Google Analytics; web sitelerine gelen trafięi izleyerek kullanıcıların siteye nasıl ulařtıęını, hangi sayfaları ziyaret ettięini ve sitede ne kadar sre kaldıęını analiz etmektedir. Bu veriler web sitesinin performansını lmek, kullanıcı davranıřlarını anlamak ve pazarlama stratejilerini optimize etmek iin kritik neme sahiptir. [6]

Google Analytics, web geliřtirme ve iyileřtirme, SEO (Arama Motoru Optimizasyonu), pazarlama, satıř ve reklam alanlarında kullanıcı davranıřlarına dayalı stratejiler belirlemek iin kullanılan ok ynl bir analiz aracıdır. Google tarafından sunulan bu cretsiz ara, web siteleri ve mobil uygulamalar zerinden kullanıcı etkileřimlerine ynelik kapsamlı veriler saęlamaktadır. Google Analytics, kullanıcıların siteye nasıl ulařtıęını, hangi sayfaları ziyaret ettięini ve sitede ne kadar sre kaldıęını analiz ederek, web sitesinin performansını lmeye ve kullanıcı deneyimini iyileřtirmeye ynelik deęerli igrler sunar. Bu baęlamda, dijital pazarlama stratejilerinin optimize edilmesi ve karar alma srelerinin desteklenmesi aısından kritik bir rol oynamaktadır. [7]

Clickstream Analizi, web sitesi ziyaretilerinin hangi sayfaları ve hangi sırayla ziyaret ettiklerine iliřkin verilerin toplanması, analiz edilmesi ve raporlanması srecidir.

Ziyaretçilerin bir web sitesinde izlediği yol "tıklama akışı" olarak adlandırılmaktadır ve bu süreç, tıklama akışı verileri ile tıklama akışı analizinden oluşmaktadır. Tıklama akışı verileri, kullanıcının web sitesindeki etkileşimlerini içerirken, tıklama akışı analizi bu verileri izleme, analiz etme ve raporlama sürecini kapsamaktadır.

Bu analiz, kullanıcıların site üzerindeki davranışlarını inceleyerek kullanıcı deneyimini iyileştirme, pazarlama stratejilerini optimize etme ve site performansını artırma amaçlarına hizmet etmektedir. Kullanıcıların tıklamaları ve etkileşimleri toplanarak analiz edilmekte; bu sayede gezinme yolları, en çok ziyaret edilen sayfalar ve siteyi terk ettikleri noktalar gibi önemli metrikler belirlenmektedir. Sonuçlar, işletmelerin kullanıcı davranışlarını daha iyi anlamalarına ve stratejik kararlar almalarına yardımcı olacak şekilde raporlanmakta ve görselleştirilmektedir.

Web siteleri, kullanıcıların ilk arama sayfasından ürün veya hizmet satın almaya kadar olan süreçlerini izlemek için tıklama akışı verilerini kullanmaktadır. Bu veriler, arama motorları, internet servis sağlayıcıları, reklam ağları ve telekomünikasyon kuruluşları tarafından da toplanır ve analiz edilmektedir. Tıklama akışı analizi, dijital pazarlama ve e-ticaret dünyasında kritik bir araç olup, kullanıcı davranışlarını anlamak ve site performansını optimize etmek için vazgeçilmez kabul edilmektedir. [8]

Log dosyası analizi, bilgisayar sistemlerinde ve ağlarda gerçekleşen olayların kaydedildiği log dosyalarının derinlemesine incelenmesi ve değerlendirilmesi sürecidir. [9] Log dosyası analizi, bilgisayar korsanlığı tespiti, güvenlik izleme, uyumluluk denetimi ve performans izleme gibi çeşitli disiplinlerde kritik bir rol oynamaktadır. Örneğin, bir güvenlik bilgilendirme sistemi (SIEM), log dosyalarını toplar, depolar, analiz eder ve anormal etkinlikleri belirlemek için karmaşık algoritmalar veya makine öğrenimi modelleri kullanır. Bu analiz süreci, log dosyalarında kaydedilen olayların derinlemesine incelenmesini içermektedir ve bu, belirli bir olayın kökenini, etkisini ve sonuçlarını anlamak için gerekmektedir. Bu süreçte, log dosyalarında kaydedilen verileri anlamlı bilgilere dönüştürmek için çeşitli teknikler kullanılmakta, bunlar arasında desen tanıma, istatistiksel analiz, veri madenciliği ve yapay zekâ teknikleri yer almaktadır.

Log dosyası analizi; bilgisayar sistemlerinin güvenliği ve performansının sürekli olarak izlenmesi ve iyileştirilmesi için önemli bir araç olarak kabul edilmektedir. [10]

Bu araç potansiyel güvenlik tehditlerinin erken tespitinde, uyumluluk gereksinimlerinin karşılanmasında ve sistem performansının optimize edilmesinde önemli amaçlar için hizmet etmektedir.

3.4 Sunucu Günlükleri Analizinde Kullanılan Yöntemler

Web sitesi trafiği analizi, çeşitli istatistiksel ve makine öğrenimi yöntemlerinin kullanıldığı karmaşık bir süreçtir. Bu analizler, web sitelerinin ziyaretçi trafiğinin analiz edilmesi, anomali tespiti ve pazarlama stratejilerinin optimize edilmesi gibi amaçlarla gerçekleştirilmektedir.

Zaman serisi analizi, web trafiğinin zaman içindeki değişimlerini anlamak için temel bir yöntemdir. Web trafiği, belirli dönemlerde ve mevsimlerde farklılık gösterebilmektedir. Bu nedenle zaman serisi analiz yöntemleri, trendleri ve diğer zamanla değişen faktörleri belirlemek için kullanılmaktadır.

Makine öğrenimi, web trafiği verilerinin daha karmaşık analizlerini gerçekleştirmek için kullanılan bir diğer önemli araçtır. Bu algoritmalar, büyük veri kümelerindeki gizli desenleri tanımlamak, kullanıcı davranışlarını sınıflandırmak ve gelecekteki trafiği tahmin etmek için kullanılmaktadır. Örneğin, k-means kümeleme algoritması, web trafiğinin farklı segmentlerini belirlemek için kullanılabilir.

Bu analizler, web sitesi sahiplerine değerli bilgiler sağlamaktadır. Bu bilgiler, trafiği anlama, pazarlama stratejilerini optimize etme, kullanıcı deneyimini geliştirme ve işletmelerin hedeflerine ulaşma konularında yol gösterici olabilmektedir. Ancak, bu analizlerin doğru ve güvenilir sonuçlar üretebilmesi için veri kalitesinin ve analiz yöntemlerinin dikkatlice seçilmesi gerekmektedir.

Bölüm 4

Sunucu Günlüklerine Dayalı E-Ticaret Web Sitesi Trafığı Analizi

4.1 Veri Setinin Özellikleri

Veri seti, kullanıcıların web sitesine yaptıkları isteklerin yanı sıra tarayıcı türleri, cihaz türleri, işletim sistemleri, arama terimleri, yönlendiren siteler ve yanıt kodları gibi çeşitli bilgileri içermekte ve bu özellikler web trafiği analizinde kullanılmaktadır.

Her bir veri satırı, bir kullanıcının belirli bir zaman diliminde web sitesine yaptığı isteği temsil etmektedir. İlgili bilgiler arasında;

- IP adresi,
- Kullanıcı kimliği,
- Zaman damgası,
- HTTP yöntemi,
- İstenen URI,
- HTTP sürümü,
- Yanıt kodu,
- İletilen bayt miktarı,
- Yönlendiren site,
- Kullanıcı ajanı bilgileri bulunmaktadır.

Veri setinde bulunan özellikler, kullanıcı davranışlarını anlamak ve web trafiğini analiz etmek için kritik öneme sahiptir. Ayrıca, veri setinin çoğu sütununun kategori veri tipine dönüştürülmesi, bellek kullanımını optimize etmekte ve veri işlemlerini hızlandırmaktadır. Bu dönüşümlerle, büyük boyuttaki veri setlerinin işlenmesi sırasında performans artışı sağlanmaktadır.

4.2 Veri Hazırlığı ve Temizleme

```
import pandas as pd

# Veri setinin yüklenilmesi
df = pd.read_csv('eclog.csv')

# Veri setinin genel bilgileri
df.info()
```

Veri seti, eclog.csv dosyasından pandas kütüphanesi kullanılarak yüklenmiştir. Bu veri seti, 35.157.691 satır ve 10 sütundan oluşmaktadır. Sütunlar, web sitesine yapılan isteklerin çeşitli özelliklerini temsil etmektedir.

Tablo 4.1: Veri setinin genel bilgilerinin gösterildiği tablo

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35157691 entries, 0 to 35157690
Data columns (total 10 columns):
#   Column          Dtype
---  ---
0   IpId            object
1   UserId         object
2   TimeStamp      int64
3   HttpMethod     object
4   Uri            object
5   HttpVersion    object
6   ResponseCode  int64
7   Bytes          int64
8   Referrer       object
9   UserAgent      object
dtypes: int64(3), object(7)
memory usage: 2.6+ GB
```

Tablo 4.1’de veri setinin genel bilgileri gösterilmiştir. Her bir sütun şu şekildedir:

- IpId: Kullanıcının IP adresi ID'si
- UserId: Kullanıcı kimliği
- TimeStamp: Zaman damgası
- HttpMethod: HTTP yöntemi
- Uri: İstenen URI
- HttpVersion: HTTP sürümü
- ResponseCode: HTTP yanıt kodu
- Bytes: İletilen veri miktarı,
- Referrer: Yönlendiren site,
- UserAgent: Kullanıcı ajanı bilgisi

Veri setinin bellekte kapladığı yer 2.6 GB olup, büyük bir veri seti olduğu görülmektedir. Verinin büyüklüğü ve çeşitliliği, web sitesine gelen trafiği analiz etmek için kapsamlı bir kaynak sağlamaktadır. İlk değerlendirme, veri setinin genel yapısını ve her bir sütunun veri tipini göz önünde bulundurarak, veri temizleme ve ön işleme adımlarının gerektiğini göstermektedir. Özellikle TimeStamp sütununun datetime formatına dönüştürülmesi, IpId ve UserId gibi sütunların daha uygun veri tiplerine dönüştürülmesinin, analizlerde verimliliği artıracakı düşünülmektedir. Ayrıca, UserAgent ve Referrer gibi sütunların detaylı analizinin yapılmasının, kullanıcı davranışlarının daha iyi anlaşılmasına katkı sağlayacağı düşünülmektedir.

Veri setinin örnek 5 satırı, web sitesine yapılan çeşitli isteklerin farklı özelliklerini içermektedir. Bu örneklerle, veri setinin çeşitliliği ve her bir sütunun içerdiği bilgilerin kapsamı daha iyi anlaşılacaktır.

Örnek 5 satırın görüntülenmesi
df.sample (5)

Tablo 4.2: Örnek beş satırın gösterildiği tablo

	IpId	UserId	TimeStamp	HttpMethod	Uri	HttpVersion	ResponseCode	Bytes	Referrer	UserAgent
13994267	138669PL	-	637168327400000000	POST	/inne/informacja_online.php	HTTP/1.1	200	0	https://shop.our-internet-company.pl/c-24_202...	Mozilla/5.0 (Linux; Android 8.0.0; RNE-L21 Bui...
31791978	60808US	-	637252369590000000	GET	/p-9776.html	HTTP/1.1	304	0	-	Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X Bu...
25425511	17892PL	-	637224666640000000	GET	/images/1_A_marzec2/mini/100px_image1199.jpg	HTTP/1.1	200	12108	https://shop.our-internet-company.pl/c-33_343...	Mozilla/5.0 (Mobile; Windows Phone 8.1; Androi...
27645349	252880PL	-	637235043000000000	GET	/images/mini/100px_image11105.jpg	HTTP/1.1	200	12972	https://shop.our-internet-company.pl/c-21.html...	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv7...
20633056	81PL	-	637201225180000000	GET	/szablony/shop_11.rwd/obrazki/szablon/tlo_gorn...	HTTP/1.1	200	2810	https://shop.our-internet-company.pl/szablony/...	Mozilla/5.0 (Windows NT 10.0; Win64; x64) Appl...

Tablo 4.2’de yer alan başlıkların her birinin içeriği ve önemi şu şekilde açıklanabilmektedir:

IpId: Kullanıcının IP adresi ID’si. Bu örneklerde farklı IP adresleri yer almakta ve her bir satır farklı bir kullanıcıya ait olabilir.

UserId: Kullanıcı kimliği sütunu boş bırakılmıştır. Bu sütunun doldurulması veya çıkarılması değerlendirilmelidir.

TimeStamp: İsteklerin yapıldığı zaman damgası. Bu değerler, datetime formatına dönüştürülerek anlamlı hale getirilebilir.

HttpMethod: HTTP yöntemi. Bu örneklerde GET ve POST yöntemleri kullanılmıştır. GET yöntemi veri çekme işlemlerini, POST yöntemi ise veri gönderme işlemlerini temsil eder.

Uri: İstenen kaynak. Bu sütun, kullanıcının erişmek istediği URL'yi içerir ve kullanıcı davranışını anlamada önemlidir.

HttpVersion: HTTP sürümü. Bu örneklerde tüm istekler HTTP/1.1 protokolü kullanılarak yapılmıştır.

ResponseCode: HTTP Yanıt Kodu. Bu değer, isteğin başarılı olup olmadığını gösterir. 200 kodu, isteğin başarıyla işlendiğini belirtir. 304 kodu ise içeriğin değişmediğini ve önbellekten kullanılabilirliğini belirtir.

Bytes: İletilen veri miktarı, bayt cinsinden. Bu, her istekte ne kadar veri gönderildiğini veya alındığını gösterir.

Referrer: Yönlendiren site. Kullanıcıların hangi sayfadan geldiklerini gösterir ve kullanıcı akışını analiz etmede kullanılır.

UserAgent: Kullanıcı ajanı bilgisi. Bu bilgi, kullanıcının hangi tarayıcıyı ve işletim sistemini kullandığını belirtir.

Bu örnekler, kullanıcıların web sitesine yaptıkları isteklerin çeşitliliğini ve veri setindeki her bir sütunun içerdiği bilgilerin kapsamını göstermektedir. Bu veriler, kullanıcı davranışlarını analiz etmek, web sitesinin performansını değerlendirmek ve kullanıcı deneyimini iyileştirmek için kullanılabilir.

```
# object değerlerin category tipine dönüştürülme işlemleri
df[df.columns.difference(['TimeStamp', 'ResponseCode', 'Bytes'])] =
df[df.columns.difference(
    ['TimeStamp', 'ResponseCode', 'Bytes'])].astype('category')

df.info()
```

Veri setindeki birçok sütunun category veri tipine dönüştürülmesinin, bellek kullanımını önemli ölçüde azaltacağı düşüncesiyle category veri tipine dönüşüm

gerçekleştirilmiştir. Kategori veri tipi, aynı değeri tekrar eden verileri daha verimli bir şekilde saklar ve büyük veri setlerinde önemli miktarda bellek tasarrufu sağlamaktadır.

Tablo 4.3: Kategori tipi dönüşümünün gösterildiği tablo

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35157691 entries, 0 to 35157690
Data columns (total 10 columns):
#   Column      Dtype
---  ---
0   IpId        category
1   UserId      category
2   TimeStamp   int64
3   HttpMethod  category
4   Uri         category
5   HttpVersion category
6   ResponseCode int64
7   Bytes       int64
8   Referrer    category
9   UserAgent   category
dtypes: category(7), int64(3)
memory usage: 1.4 GB
```

Tablo 4.3'te veri setinin kategori değişimi gösterilmektedir. İlk veri setinin boyutu 2.6 GB iken, dönüşümlerden sonra boyut 1.4 GB'a düşmüştür. Bu dönüşümler bellek kullanımını optimize ederek veri işlemlerini hızlandırmakta ve daha verimli hale getirmektedir. Bu sayede, büyük veri setlerinin işlenmesi sırasında performans artışı sağlanmaktadır.

```
import datetime

# Zaman işlemleri
df['TimeStamp'] = df['TimeStamp'].apply(lambda x: datetime.datetime(1, 1,
1, 1) + datetime.timedelta(microseconds=x / 10))
df['Date'] = df['TimeStamp'].dt.date
df['Hour'] = df['TimeStamp'].dt.hour
df['DayOfWeek'] = df['TimeStamp'].dt.day_name().astype('category')

df['Month'] = df['TimeStamp'].dt.month_name().astype('category')
```

Zaman damgasını anlamlı ve işlenebilir bir forma dönüştürmek amacıyla, TimeStamp sütunundaki mikro saniye cinsinden olan değerler datetime nesnesine dönüştürülmüştür. Ardından, tarih (Date), saat (Hour), haftanın günü (DayOfWeek) ve ay (Month) gibi yeni özellikler çıkarılmıştır. Bu bilgiler, zaman bazlı analizler yapabilmek için yararlı olmaktadır.

```
import re

import pycountry

# Kullanıcı kimliği ve ülke işlemleri
```

```

df['UserId'] = df['IpId'].apply(lambda x: re.findall(r'\d+', x)[0] if
re.findall(r'\d+', x) else None)
df['Country'] = df['IpId'].str.extract('([a-zA-Z]+)').astype('category')
df['CountryName'] = df['Country'].apply(lambda x:
pycountry.countries.get(alpha_2=x).name if x and
pycountry.countries.get(alpha_2=x) else None).astype('category')

```

Kullanıcıların kimlik bilgilerini ve ülke bilgilerini belirlemek için Kullanıcı Kimliği ve Ülke Bilgisi IpId sütunundan sayısal değerler çıkarılarak UserId oluşturulmaktadır. Ülke kodları IpId sütunundan çıkarılarak Country sütunu oluşturulmaktadır. Ardından pycountry kütüphanesi kullanılarak ülke kodları ülke isimlerine dönüştürülmektedir (CountryName).

```

# URI işlemleri
df['Category'] = df['Uri'].str.extract(r'cPath=(\d+)').astype('category')
df['ProductID'] = pd.to_numeric(df['Uri'].str.extract(r'products_id=(\d+)',
expand=False), errors='coerce')
df['Payment'] = df['Uri'].str.contains('action=buy_now')
df['Login'] = df['Uri'].str.contains('login.')
df['Logout'] = df['Uri'].str.contains('wylogowanie')
df['Search'] = df['Uri'].apply(lambda x: re.search(r'query\d+', x).group(0)
if re.search(r'query\d+', x) else None).astype('category')

```

Uri sütunundan kategori bilgisi (Category), ürün ID'si (ProductID), ödeme işlemi olup olmadığı (Payment), giriş (Login) ve çıkış (Logout) bilgileri çıkartılmaktadır. Bu bilgiler elde edilerek trafik kaynaklarını analiz etmek daha da kolaylaşmaktadır. Web sitesi üzerindeki kullanıcı aktivitelerini analiz etmek için arama terimleri Search sütununda saklanmaktadır.

```

from urllib.parse import urlparse

# Yönlendiren site işlemleri
df['ReferrerSite'] = df['Referrer'].apply(lambda x:
urlparse(x).netloc.split('.')[2] if len(urlparse(x).netloc.split('.')) > 1
else urlparse(x).netloc).astype('category')

```

Bu adımla, Referrer sütunundaki URL'den yönlendiren sitenin alan adı çıkarılarak ReferrerSite sütununa eklenmektedir.

```

from user_agents import parse

# Kullanıcı ajan işlemleri
df['UserAgentParsed'] = df['UserAgent'].apply(parse).astype('category')
df['Browser'] = df['UserAgentParsed'].apply(lambda x:
x.browser.family).astype('category')
df['OS'] = df['UserAgentParsed'].apply(lambda x:
x.os.family).astype('category')

```



```
df['Bot'] = df['UserAgentParsed'].apply(lambda x:
x.is_bot).astype('category')
```

Kullanıcıların hangi tarayıcıları ve yazılımları kullandığını belirlemek için UserAgent sütunundaki bilgiler user_agents kütüphanesi kullanılarak analiz edilmekte ve tarayıcı (Browser), işletim sistemi (OS) ve bot olup olmadığı (Bot) bilgileri yeni sütunlara çıkarılmaktadır.

```
#Gereksiz olan sütunların temizlenmesi
df.drop(columns=['UserAgentParsed', 'UserAgent', 'Referrer'], inplace=True)

# Tekrarlı verilerin temizlenmesi
df = df.drop_duplicates()
```

Veri setini temiz tutmak ve bellek kullanımını optimize etmek için gereksiz sütunlar kaldırılmakta; veri setinde tekrarlı ve hatalı veriler bulunabileceği, bu verilerin bellekte fazla yer kaplayacağı ve analiz sonuçlarını olumsuz ve yanlış etkileyebileceği düşüncesinden hareketle tekrarlı veriler kontrol edilerek temizlenmektedir.

4.3 Trafik Analizleri

Trafik analizleri, günlük, hafta günlerine ve aylık olarak, kullanıcı davranışlarını ve trafik dinamiklerini detaylı bir şekilde anlamak için yapılmaktadır. Bu analizler, web sitesinin hangi günlerde ve hangi aylarda daha fazla ziyaretçi aldığını belirleyerek, pazarlama stratejilerini ve kampanyaları optimize etmek için kritik veriler sağlamaktadır. Ayrıca, yoğun trafik dönemlerinde site performansını artırmak ve kullanıcı deneyimini iyileştirmek için de önemlidir.

4.3.1 Genel Trafik Dağılımı

Günlük Trafik

```
import hvplot.pandas

# Günlük trafik sayıları
daily_traffic = df.groupby('Date').size().reset_index(name='TrafficCount')
daily_traffic.set_index('Date')
```

Tablo 4.4: Günlük trafik verilerinin gösterildiği tablo

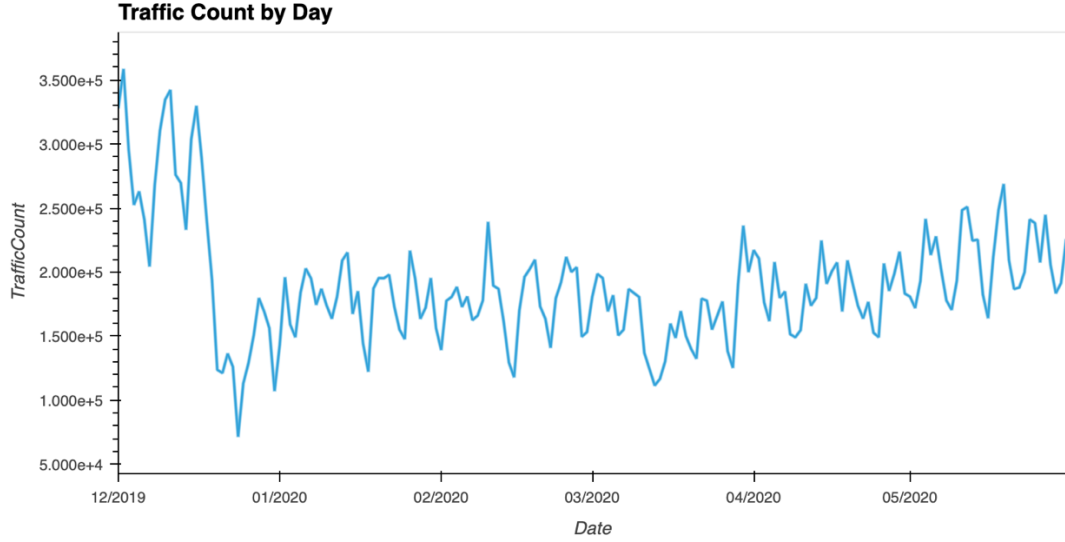
TrafficCount	
Date	
2019-12-01	328092
2019-12-02	358712
2019-12-03	295361
2019-12-04	252572
2019-12-05	262983
...	...
2020-05-27	244752
2020-05-28	205571
2020-05-29	183479
2020-05-30	191333
2020-05-31	225627

183 rows × 1 columns

Günlük trafik verilerinin yer aldığı 183 satır ve 1 sütun olarak veri setinde ele alınarak oluşturulan tablodan bir kısım Tablo 4.4'te gösterilmektedir.

```
# Günlük trafik grafiği
hv_plot_daily_traffic = daily_traffic.hvplot.line(x='Date',
y='TrafficCount', title='Traffic Count by Day', width=800, height=400)

hv_plot_daily_traffic
```



Şekil 4.1: Günlük trafik verileri

Şekil 4.1’de web sitesine yapılan günlük isteklerin zaman içindeki dağılımı gösterilmektedir. İnceleme dönemi boyunca, günlük istek sayılarında belirgin dalgalanmalar gözlemlenmektedir.

Grafiğin başlangıcında yüksek bir trafik seviyesi görülmektedir, bu durum muhtemelen tatil sezonuna denk gelmektedir ve bu dönemlerde kullanıcıların alışveriş aktivitelerinin arttığı anlaşılmaktadır. Zaman ilerledikçe, günlük istek sayılarında bir düşüş yaşandığı fark edilmektedir. Tatil sezonunun bitimiyle birlikte kullanıcıların alışveriş aktivitelerinde bir azalma meydana geldiği görülmektedir. Ancak, bu düşüşün ardından trafik seviyeleri daha dengeli bir çizgide seyretmekte ve bu da web sitesinin normal kullanım dönemine girdiğini göstermektedir.

Mart 2020’de ise günlük istek sayılarında belirgin bir artış yaşanmaktadır. Bu artışın, COVID-19 pandemisinin başlangıcından kaynaklandığı düşünülmektedir. Çünkü pandeminin etkisiyle insanların fiziksel mağazalardan uzaklaşıp online alışverişe yöneldiği görülmektedir. Bu dönem, online alışverişte genel bir artışın yaşandığı bir döneme işaret etmektedir. Grafikte ayrıca, günlük istek sayılarında dalgalanmalar dikkat çekmektedir. Hafta içi ve hafta sonu günleri arasında belirgin farklılıklar görülmektedir. Örneğin, hafta sonları insanların daha fazla alışveriş yapma eğiliminde olduğu, hafta içi ise iş saatleri nedeniyle trafik yoğunluğunun daha düşük olduğu gözlemlenmektedir. Grafikte görülen günlük trafik dağılımı, web sitesinin kullanım

dönemlerine, tatil sezonlarına ve pandemi gibi olağanüstü durumlara bağlı olarak değişiklik göstermektedir. Bu tür analizler, web sitesi yöneticilerinin ve pazarlama ekiplerinin kullanıcı davranışlarını daha iyi anlamalarına ve stratejik kararlar almasına yardımcı olmaktadır.

Haftanın Günlerine Göre Trafik

```
# Haftanın günlerine göre trafik sayıları
dayofweek_traffic = df.groupby('DayOfWeek',
observed=False).size().reset_index(name='TrafficCount')

# Haftanın günlerini sıralamak için kategorik dönüşüm
days = ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday',
'Sunday']
dayofweek_traffic['DayOfWeek'] =
pd.Categorical(dayofweek_traffic['DayOfWeek'], categories=days,
ordered=True)

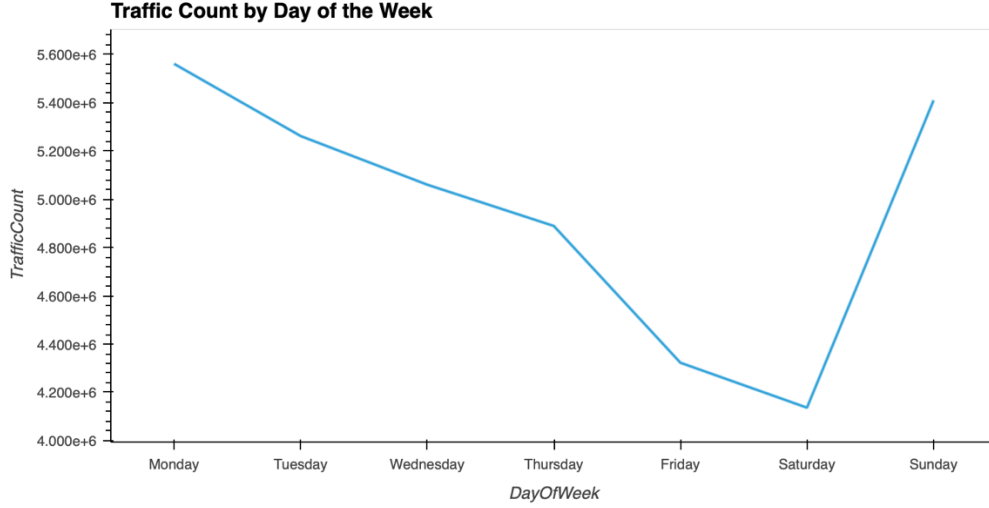
# Haftanın günlerine göre trafik tablosu
dayofweek_traffic = dayofweek_traffic.sort_values('DayOfWeek')
dayofweek_traffic.set_index('DayOfWeek')
```

Tablo 4.5: Haftanın günlerine göre trafik verilerinin gösterildiği tablo

TrafficCount	
DayOfWeek	
Monday	5560512
Tuesday	5261494
Wednesday	5060082
Thursday	4889101
Friday	4322862
Saturday	4136708
Sunday	5409007

Tablo 4.5'te haftanın günlerine göre web sitesine yapılan trafik sayısı gösterilmektedir.

```
# Haftanın günlerine göre trafik grafiği
hv_plot_dayofweek_traffic = dayofweek_traffic.hvplot.line(x='DayOfWeek',
y='TrafficCount', title='Traffic Count by Day of the Week', width=800,
height=400)
hv_plot_dayofweek_traffic
```



Şekil 4.2: Haftanın günlerine göre trafik verileri

Şekil 4.2’de haftanın günlerine göre web sitesine yapılan trafik sayısı gösterilmektedir. X-ekseni haftanın günlerini, Y-ekseni ise trafik sayısını temsil etmektedir. Bu analiz, kullanıcıların haftanın hangi günlerinde web sitesini daha yoğun ziyaret ettiklerini anlamak için yapılmaktadır.

- Pazartesi günü, haftanın en yoğun trafik günü olarak dikkat çekmektedir. Kullanıcılar, hafta başında web sitesine daha fazla ilgi göstermekte ve aktif olmaktadır. Bu, haftalık iş ve alışveriş rutinlerinin başlamasıyla trafiğin zirve yaptığı günü işaret etmektedir.
- Salı ve çarşamba günleri, pazartesi gününe göre hafif bir düşüş gösterse de hala yüksek trafik seviyeleri görülmektedir. Bu durum, kullanıcıların hafta başındaki aktivitelerine devam ettiğini göstermektedir.
- Perşembe ve cuma günleri, haftanın ortasında başlayan trafik düşüşünün daha belirgin hale geldiği günler olarak görülmektedir. Bu günlerde, kullanıcıların hafta sonuna yaklaşırken web sitesini daha az ziyaret ettiği görülmektedir.
- Cumartesi günü, haftanın en düşük trafik günü olduğu görülmektedir. Kullanıcılar, hafta sonu aktiviteleri nedeniyle web sitesini daha az ziyaret etmektedir.
- Pazar günü, cumartesi gününe kıyasla trafik tekrar artış göstermektedir. Kullanıcılar, hafta sonunun sonunda tekrar siteye dönüş yapmaktadır. Bu artış, pazartesi gününe hazırlık veya hafta sonu bitmeden yapılan alışverişlerle açıklanabilmektedir.

Bu analiz ile birlikte, web sitesinin hangi günlerde daha yoğun kullanıldığını ve hangi günlerde daha az ziyaret edildiğini anlamak kolaylaşmaktadır. Pazartesi günleri yüksek trafik oranları ile dikkat çekerken, cumartesi günleri düşük trafik oranları göstermektedir. Bu bilgiler, pazarlama stratejilerini zamanlamak, site bakım ve güncellemelerini planlamak için kullanılabilir. Pazartesi ve pazar günleri, yüksek kullanıcı trafiğinden faydalanarak yapılan kampanyalar ve promosyonlarla daha etkili sonuçlar alınabilmektedir. Perşembe ve cuma günlerinde ise kullanıcı ilgisini artırmak için özel içerikler ve teşvik edici kampanyalar düzenlenmesi uygun görülmektedir. Cumartesi gününün, düşük trafik göz önünde bulundurularak teknik bakım ve güncellemeler için uygun bir zaman dilimi olabileceği düşünülmektedir.

Aylık Trafik

```
# Aylık trafik sayıları
monthly_traffic = df.groupby('Month',
observed=False).size().reset_index(name='TrafficCount')

# Ayları sıralamak için kategorik dönüşüm
months = ['December', 'January', 'February', 'March', 'April', 'May']
monthly_traffic['Month'] = pd.Categorical(monthly_traffic['Month'],
categories=months, ordered=True)

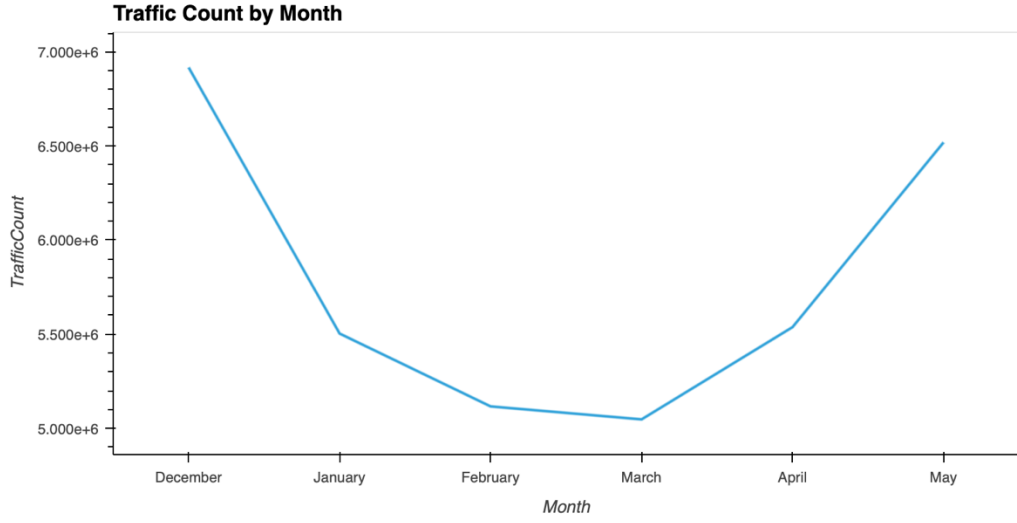
# Aylık trafik tablosu
monthly_traffic = monthly_traffic.sort_values('Month')
monthly_traffic.set_index('Month')
```

Tablo 4.6: Aylık trafik verilerinin gösterildiği tablo

TrafficCount	
Month	
December	6916804
January	5503210
February	5116189
March	5047228
April	5537493
May	6518842

Tablo 4.6'da aylara göre (6 aylık olarak) web sitesine yapılan trafik sayısı gösterilmektedir.

```
# Aylık trafik grafiği
hv_plot_monthly_traffic = monthly_traffic.hvplot.line(x='Month',
y='TrafficCount', title='Traffic Count by Month', width=800, height=400)
hv_plot_monthly_traffic
```



Şekil 4.3: Aylık trafik verileri

Şekil 4.3'te aylık trafik sayısı gösterilmektedir. Bu analiz, kullanıcıların yılın hangi aylarında web sitesini daha yoğun ziyaret ettiklerini anlamak için yapılmaktadır. Yatay eksen ilgili ayları (Aralık, Ocak, Şubat, Mart, Nisan, Mayıs) temsil ederken, dikey eksen web sitesine yapılan trafik sayısını (TrafficCount) göstermektedir. Şekil 4.3 incelendiğinde;

Aralık ayında web trafiğinin yaklaşık 7 milyon ziyaretle en yüksek seviyede olduğu görülmektedir. Bu durumun, yıl sonu alışveriş sezonu ve tatil dönemleri nedeniyle artan e-ticaret faaliyetlerinden kaynaklanabileceği düşünülmektedir. Ocak ayında ise trafik önemli ölçüde azalarak yaklaşık 5,5 milyon seviyesine düşmektedir. Bu düşüş, tatil döneminin sona ermesi ve insanların daha az alışveriş yapmasından kaynaklanmaktadır.

Şubat ve Mart aylarına gelindiğinde trafik, Ocak ayına göre daha düşük bir seviyede kalmaya devam eder ve yaklaşık 5 milyon seviyesinde sabitlendiği görülebilmektedir. Bu aylar, genellikle alışveriş açısından daha sakin geçmektedir. Nisan ayında trafik yeniden artış göstermeye başladığı ve yaklaşık 5,5 milyon seviyesine ulaştığı görülmektedir. Bu artış, ilkbahar mevsiminin başlaması ve insanların daha fazla alışveriş yapma eğiliminde olmasıyla açıklanabilir. Mayıs ayında ise trafik, Nisan ayına kıyasla önemli bir artış göstererek yaklaşık 6,5 milyon seviyesine ulaşmaktadır.

Bu artışın, yaz sezonunun yaklaşması ve anneler günü gibi özel günlerin de etkisiyle olabileceği düşünülmektedir.

Ülkelere Göre Trafik Dağılımı

Ülkelere Göre Trafik Dağılımı analizi, web sitesine hangi ülkelerden daha fazla ziyaretçi geldiğini belirlemek için yapılmaktadır. Bu analiz, farklı ülkelerdeki kullanıcıların ilgi ve ihtiyaçlarını anlamak ve bölgesel pazarlama stratejilerini optimize etmek için önemlidir.

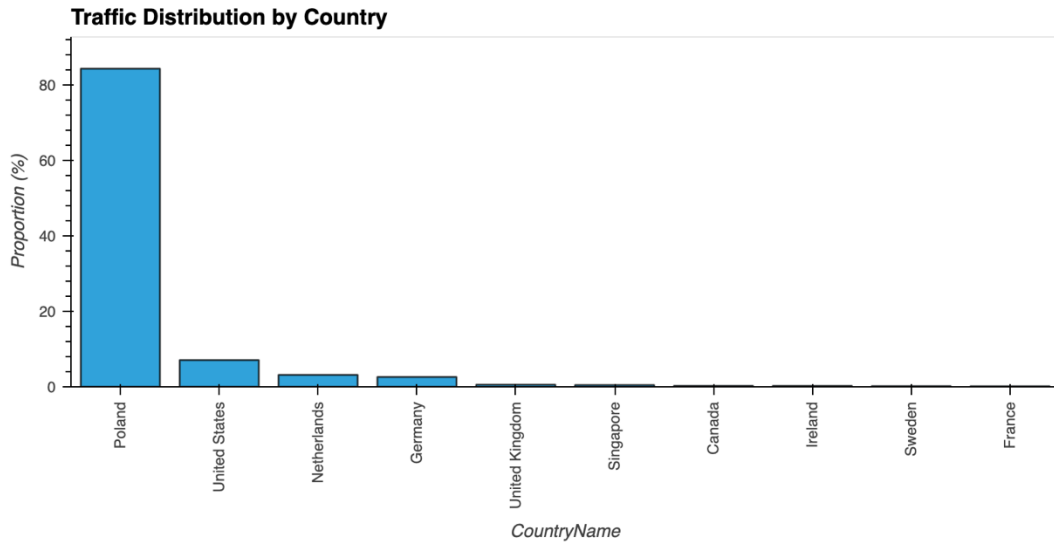
```
# Ülkelere göre trafik sayısı
country_traffic =
df['CountryName'].value_counts(normalize=True).reset_index(name='Proportion (%)').head(10)
country_traffic['Proportion (%)']*100
country_traffic.set_index('CountryName').map('{:.2f}%'.format).T
```

Tablo 4.7: Ülkelere göre trafik dağılımının gösterildiği tablo

CountryName	Poland	United States	Netherlands	Germany	United Kingdom	Singapore	Canada	Ireland	Sweden	France
Proportion (%)	84.29%	7.07%	3.15%	2.60%	0.56%	0.47%	0.26%	0.25%	0.19%	0.16%

Tablo 4.7’de ülkelere göre trafik dağılımı gösterilmektedir.

```
# Ülkelere göre trafik dağılımı grafiği
hv_country_traffic = country_traffic.hvplot.bar(x='CountryName',
y='Proportion (%)', title='Traffic Distribution by Country', rot=90,
width=800, height=400)
hv_country_traffic
```



Şekil 4.4: Ülkelere göre trafik dağılımı

Polonya'dan gelen trafik %84,29 gibi çok yüksek bir orana sahiptir. Bu, web sitesinin büyük bir kısmının Polonya'dan gelen ziyaretçiler tarafından kullanıldığını göstermektedir. Polonya'nın ardından gelen diğer ülkeler ise çok daha düşük oranlara sahiptir.

Amerika Birleşik Devletleri (United States) %7,07 oranında trafikle ikinci sırada yer almaktadır. Hollanda (Netherlands) %3,15 oranında trafikle üçüncü sıradadır. Almanya (Germany) %2,60 oranında trafikle dördüncü sırada bulunmaktadır. Birleşik Krallık (United Kingdom) %0,56 oranında trafikle beşinci sırada yer almaktadır.

Şekil 4.4'te ayrıca, Singapur, Kanada, İrlanda, İsveç ve Fransa gibi ülkelerden gelen trafik oranları da gösterilmektedir. Bu ülkelerden gelen trafik, %1'in altında kalmaktadır.

Bu dağılım, web sitesinin Polonya'da çok popüler olduğunu ve ziyaretçilerin büyük bir çoğunluğunun Polonya'dan geldiğini açıkça ortaya koymaktadır. Diğer ülkelerden gelen trafik oranları ise oldukça düşüktür. Bu durum, web sitesinin pazarlama stratejilerini ve içerik yönetimini Polonya'daki kullanıcıların ihtiyaçlarına ve tercihlerine göre optimize etmesi gerektiğini işaret etmektedir. Aynı zamanda, web sitesinin global erişimini artırmak için Amerika Birleşik Devletleri, Hollanda ve Almanya gibi diğer yüksek trafik sağlayan ülkelerde de daha fazla tanıtım yapılabileceğini göstermektedir.

Ülkelere göre trafik dağılımı analizi, web sitesinin hangi bölgelerde daha fazla kullanıcıya sahip olduğunu ve bu bölgelerdeki kullanıcıların ihtiyaçlarını daha iyi anlamak için kritik veriler sağlamaktadır. Bu bilgiler, web sitesinin stratejik planlamalarını ve pazarlama kampanyalarını optimize etmesine yardımcı olabilir.

4.3.2 Trafik Dağılımı Karşılaştırması

Saatlik Trafik Dağılımı

Saatlik trafik dağılımı analizi, web sitesine yapılan isteklerin gün içerisindeki saatlik dağılımını inceleyerek kullanıcıların hangi saatlerde daha aktif olduğunu belirlemeye çalışmaktadır.

```
# insan ve bot trafiği için saatlik istek sayıları
human_hourly_traffic = df[df['Bot']==False].groupby('Hour',
observed=False).size().reset_index(name='RequestCount')
bot_hourly_traffic = df[df['Bot']==True].groupby('Hour',
observed=False).size().reset_index(name='RequestCount')

# insan ve bot trafiği verilerinin birleştirilmesi
hourly_traffic=pd.merge(human_hourly_traffic, bot_hourly_traffic,
on='Hour', how='outer').set_index('Hour')
hourly_traffic.columns = ['HumanRequestCount', 'BotRequestCount']
hourly_traffic
```

Tablo 4.8: Saatlik trafik dağılımının gösterildiği tablo

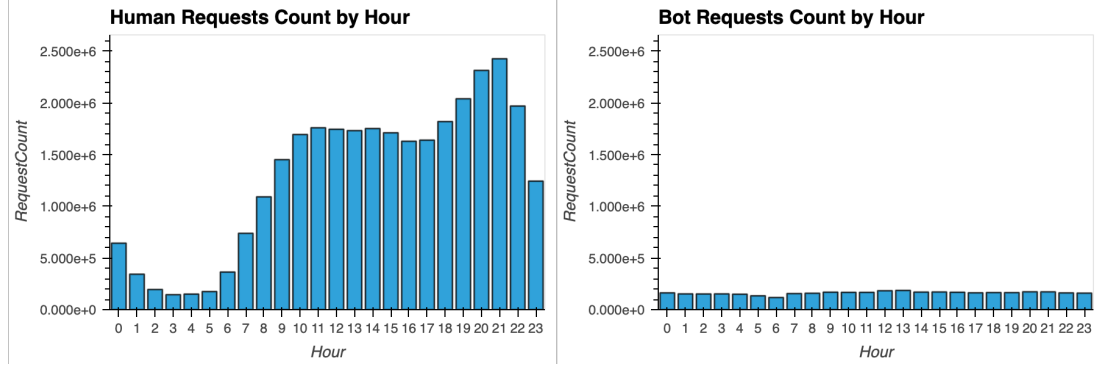
Hour	HumanRequestCount	BotRequestCount
0	642856	162139
1	342819	152497
2	194897	151835
3	144328	152519
4	150414	149281
5	174573	133844
6	363741	117098
7	738266	155456
8	1091045	158565
9	1450583	169067
10	1693831	167400
11	1759230	167318
12	1744877	182986
13	1731590	185905
14	1751479	169979
15	1710563	171011
16	1628670	168306
17	1639704	162784
18	1819154	165261
19	2039403	164743
20	2313940	172554
21	2426759	172270
22	1969402	161830
23	1243216	159778

Tablo 4.8’de insan ve bot trafiği için saatlik istek sayıları gösterilmektedir.

```
# Saatlik trafik dağılımı grafiği
```

```
hv_plot_human_hourly_traffic = human_hourly_traffic.hvplot.bar(x='Hour',  
y='RequestCount', title='Human Requests Count by Hour', width=450,  
height=300)
```

```
hv_plot_bot_hourly_traffic = bot_hourly_traffic.hvplot.bar(x='Hour',  
y='RequestCount', title='Bot Requests Count by Hour', width=450, height=300)  
(hv_plot_human_hourly_traffic+hv_plot_bot_hourly_traffic)
```



Şekil 4.5: İnsan ve bot trafiği için saatlik istek sayılarının karşılaştırması

Şekil 4.5'te insan ve bot isteklerinin saatlik dağılımlarını göstermektedir. Sol tarafta, insan isteklerinin saatlere göre dağılımı gösterilmekte ve sağ tarafta ise bot isteklerinin saatlere göre dağılımı yer almaktadır.

İnsan isteklerinin incelenmesi, gece yarısından sabah saatlerine kadar (0-7 saatleri arası) düşük bir trafik seviyesi olduğunu göstermektedir. Sabah saat 8'den itibaren istek sayısının arttığı ve öğleden sonra 12-17 saatleri arasında nispeten yüksek ve dengeli bir seviyede seyrettiği gözlemlenmektedir. Akşam saatlerinde (18-21 arası) trafik daha da artmakta ve en yüksek seviyeye ulaşmaktadır. Bu saat diliminde istek sayısı 2,5 milyona kadar çıkmaktadır. Saat 22'den itibaren ise trafik yeniden azalmaya başlamaktadır. Bu durum, insanların gün içinde e-ticaret sitesiyle yoğun olarak etkileşimde bulunduğunu ve akşam saatlerinde bu etkileşimin zirveye çıktığını göstermektedir.

Bot isteklerinin dağılımına baktığımızda ise gün boyunca (0-23 saatleri arası) sabit ve düşük seviyede bir trafik gözlemlenmektedir. Bot istekleri, günün hiçbir saatinde insan istekleri kadar yoğun değildir ve genellikle aynı seviyede kalmaktadır. Bu, botların siteye düzenli aralıklarla ve sürekli olarak erişim sağladığını, ancak bu erişimin insan trafiğine kıyasla çok daha az olduğunu göstermektedir.

İnsan istekleri gün içindeki saatlere göre belirgin dalgalanmalar gösterirken, bot istekleri oldukça sabit bir dağılım sergilemektedir. Bu farklılık, kullanıcıların günlük rutinlerinin e-ticaret sitesi trafiğine etkisini açıkça ortaya koymaktadır. Akşam saatlerindeki yoğun insan trafiği, pazarlama stratejileri ve site optimizasyonu açısından önemli veriler sunarken, bot trafiğinin düşük ve sabit olması, siteye yönelik otomatik erişimlerin düzenli ve kontrollü olduğunu göstermektedir.

Hafta Günleri Trafik Dağılımı

Hafta Günleri Trafik Dağılımı analizi, web sitesine hangi günlerde daha fazla ziyaretçi geldiğini belirlemek için yapılmaktadır. Bu analiz, kullanıcıların alışveriş alışkanlıklarını ve belirli günlerdeki trafik yoğunluğunu anlamak için önemlidir.

```
# insan ve bot trafiği için günlere göre istek sayıları

human_dayofweek_traffic = df[df['Bot']==False].groupby('DayOfWeek',
observed=False).size().reset_index(name='HumanRequestCount')

bot_dayofweek_traffic = df[df['Bot']==True].groupby('DayOfWeek',
observed=False).size().reset_index(name='BotRequestCount')

# Günleri sıralamak için kategorik dönüşüm

days = ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday',
'Sunday']

human_dayofweek_traffic['DayOfWeek'] =
pd.Categorical(human_dayofweek_traffic['DayOfWeek'], categories=days,
ordered=True)

bot_dayofweek_traffic['DayOfWeek'] =
pd.Categorical(bot_dayofweek_traffic['DayOfWeek'], categories=days,
ordered=True)

# insan ve bot trafiği verilerinin birleştirilmesi

dayofweek_traffic=pd.merge(human_dayofweek_traffic, bot_dayofweek_traffic,
on='DayOfWeek', how='outer').set_index('DayOfWeek')

dayofweek_traffic
```

Tablo 4.9: Hafta günleri trafik dağılımının gösterildiği tablo

	HumanRequestCount	BotRequestCount
DayOfWeek		
Monday	4979841	580671
Tuesday	4747457	514037
Wednesday	4505027	555055
Thursday	4347525	541576
Friday	3774416	548446
Saturday	3589345	547363
Sunday	4821729	587278

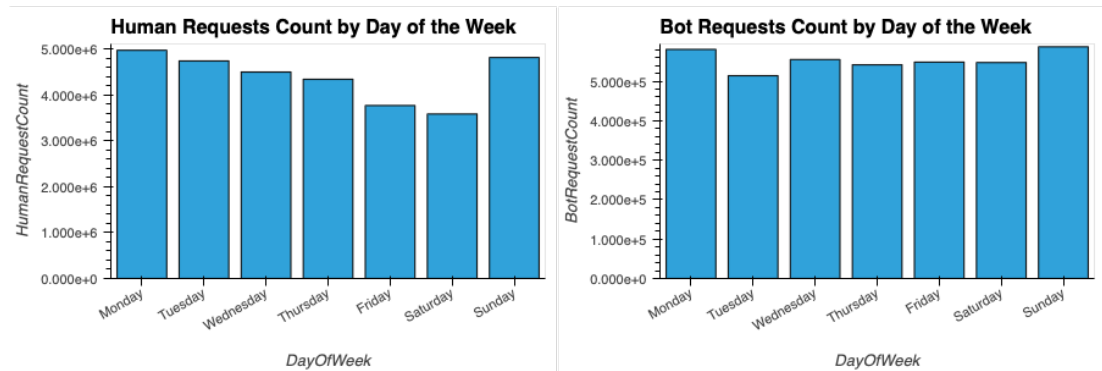
Tablo 4.9'da insan ve bot istekleri için haftalık istek sayıları trafik dağılımı karşılaştırması gösterilmektedir.

Hafta günleri trafik dağılımı grafiği

```
hv_plot_human_dayofweek_traffic = human_dayofweek_traffic.sort_values('DayOfWeek').hvplot.bar(x='DayOfWeek', y='HumanRequestCount', title='Human Requests Count by Day of the Week', rot=20, width=450, height=300)
```

```
hv_plot_bot_dayofweek_traffic = bot_dayofweek_traffic.sort_values('DayOfWeek').hvplot.bar(x='DayOfWeek', y='BotRequestCount', title='Bot Requests Count by Day of the Week', rot=20, width=450, height=300)
```

```
(hv_plot_human_dayofweek_traffic+hv_plot_bot_dayofweek_traffic)
```



Şekil 4.6: İnsan ve bot istekleri için haftalık istek sayıları trafik dağılımı

Şekil 4.6’da insan ve bot istekleri için haftalık istek sayıları gösterilmektedir. İnsan istekleri ve bot istekleri ayrı grafiklerde sunulmuş olup, y-ekseni istek sayılarını, x-ekseni ise haftanın günlerini temsil etmektedir.

Şekil 4.6’da insan isteklerinin haftanın her günü oldukça düzenli bir dağılım sergilediği, ancak belirli günlerde trafiğin daha yoğun, Pazar gününden çarşamba gününe kadar en yüksek istek sayısına sahip olduğu ve bu günlerde siteye olan ilgilerinin de arttığı Şekil 4.6’dan görülebilmektedir. Çarşamba ve Perşembe günlerinde istek sayısında hafif bir düşüş gözlemlenmekteyse de trafiğin hala yüksek seviyelerde olduğu, cuma günü ise trafiğin biraz daha azaldığı söylenebilir. Cumartesi gününde, trafik haftanın diğer günlerine göre belirgin bir şekilde düşmüş olduğu görülmektedir. Bu durum, kullanıcıların hafta sonu aktiviteleri nedeniyle siteyi daha az ziyaret ettiklerini göstermektedir. Pazar günü kullanıcıların siteyi daha fazla ziyaret ettiği ve trafiğin tekrar artış gösterdiği görülmektedir.

Bot isteklerinin haftanın günlerine göre dağılımına baktığımızda, insan isteklerine göre daha sabit bir dağılım sergilediği görülmektedir. Bot trafiği, haftanın her günü benzer seviyelerde olup, insan trafiği gibi belirgin dalgalanmalar göstermemektedir. Her gün için bot trafiğinin oldukça yüksek seyrettiği ve özellikle Pazar ve Pazartesi günlerinde de yüksek seviyelerde kaldığı görülmektedir.

İnsan trafiğinin zirve günlerinin Pazar, Pazartesi ve Salı olduğu, bu günlerde kullanıcıların siteyi yoğun olarak kullandıkları söylenebilir. Çarşamba ve Perşembe günlerinde trafik bir miktar azalırken, Cuma günü biraz daha düşmektedir. Cumartesi gününde trafik önemli ölçüde azalmakta, Pazar günü ise tekrar artış göstermektedir.

Aylık Trafik Dağılımı

Aylık Trafik Dağılımı analizi, web sitesine hangi aylarda daha fazla ziyaretçi geldiğini belirlemek için yapılmaktadır. Bu analiz, sezonluk trendleri ve kullanıcı davranışlarını anlamak için kritik bir yöntemdir. E-ticaret siteleri gibi mevsimsel değişikliklerden etkilenen platformlarda, bu tür analizler pazarlama stratejilerini optimize etmek için büyük önem taşımaktadır.

```

# insan ve bot trafiği için aylık istek sayıları

human_monthly_traffic = df[df['Bot'] == False].groupby('Month',
observed=False).size().reset_index(name='HumanRequestCount')

bot_monthly_traffic = df[df['Bot'] == True].groupby('Month',
observed=False).size().reset_index(name='BotRequestCount')

# Ayları sıralamak için kategorik dönüşüm

months = ['December', 'January', 'February', 'March', 'April', 'May']

human_monthly_traffic['Month'] =
pd.Categorical(human_monthly_traffic['Month'],
categories=months,
ordered=True)

bot_monthly_traffic['Month'] = pd.Categorical(bot_monthly_traffic['Month'],
categories=months, ordered=True)

# insan ve bot trafiği verilerinin birleştirilmesi

monthly_traffic = pd.merge(human_monthly_traffic, bot_monthly_traffic,
on='Month', how='outer').set_index('Month')

monthly_traffic

```

Tablo 4.10: Aylık trafik dağılımının gösterildiği tablo

	HumanRequestCount	BotRequestCount
Month		
December	6296354	620450
January	4945350	557860
February	4406398	709791
March	4376821	670407
April	4890930	646563
May	5849487	669355

Tablo 4.10'da insan ve bot istekleri için aylık trafik dağılımı ayrı grafiklerde sunulmaktadır.

```

# Aylık trafik dağılımı grafiği

```

```

hv_plot_human_month_traffic =
human_monthly_traffic.sort_values('Month').hvplot.bar(x='Month',
y='HumanRequestCount', title='Human Requests Count by Month',width=450,
height=300)

```

```

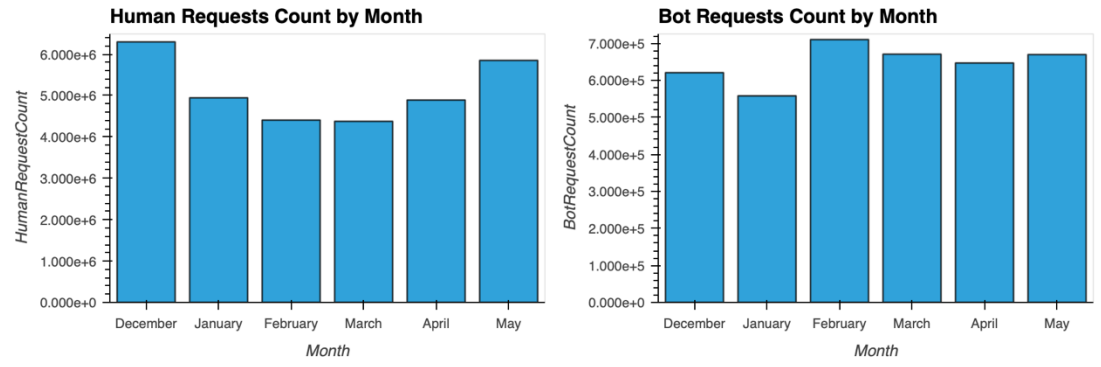
hv_plot_bot_month_traffic =
bot_monthly_traffic.sort_values('Month').hvplot.bar(x='Month',
y='BotRequestCount', title='Bot Requests Count by Month', width=450,
height=300)

```

```

(hv_plot_human_month_traffic+hv_plot_bot_month_traffic)

```



Şekil 4.7: İnsan ve bot istekleri için aylık trafik dağılımı

Şekil 4.7’de insan ve bot istekleri için aylık istek sayıları ayrı grafiklerde sunulmaktadır. İnsan istekleri ve bot istekleri y-ekseni üzerinde istek sayılarıyla, x-ekseni üzerinde ise aylarla gösterilmektedir. Şekil 4.7’de insan isteklerinin aylar boyunca değişen bir dağılım sergilediği görülmektedir.

Aralık ayında en yüksek istek sayısına sahip olduğu dikkat çekmektedir. Bu ay, yılın sonunda gerçekleşen tatil ve alışveriş sezonu olması nedeniyle yüksek trafik almaktadır. Ocak ayında istek sayısında belirgin bir düşüş gözlemlenmektedir. Yılbaşı tatilinin bitmesi ve normal iş düzenine dönülmesi nedeniyle trafik azalmaktadır. Şubat ayında trafik daha da azalmaktadır, bu da kış aylarının etkisini ve kullanıcıların siteyi daha az ziyaret ettiğini göstermektedir. Mart ayında trafik tekrar artış göstermekte ve Nisan ayında bu artış devam etmektedir. Bu dönemde kullanıcıların siteyi daha fazla ziyaret ettiği ve etkinliklerin arttığı gözlemlenmektedir. Mayıs ayında ise trafik, Nisan ayına göre biraz daha artarak yüksek seviyelerde kalmaktadır.

Bot isteklerinin aylık dağılımına bakıldığında, insan isteklerine göre daha sabit bir dağılım sergilediği görülmektedir. Bot trafiği, yıl boyunca belirli bir seviyede kalmakta ve insan trafiği gibi büyük dalgalanmalar göstermemektedir. Her ay için bot trafiğinin oldukça yüksek olduğu görülmekte ve belirgin bir artış veya azalma görülmemektedir.

İnsan trafiğinin zirve yaptığı aylar Aralık ve Mayıs olup, bu dönemlerde kullanıcıların siteyi yoğun olarak kullandıkları görülmektedir. Ocak ve Şubat aylarında trafik düşerken, Mart ve Nisan aylarında yeniden artış göstermektedir. Bot trafiği ise, web sitesinin sürekli izlenmesi ve güvenlik önlemlerinin alınması gerektiğini göstermektedir.

4.3.3 En Sık Yapılan Aramalar

En sık yapılan aramalar kullanıcıların web sitesinde hangi ürün veya hizmetlere daha fazla ilgi gösterdiğinin anlaşılmasını kolaylaştırmaktadır.

```
# En sık yapılan arama sayıları
```

```
search_terms = df['Search'].value_counts().reset_index(name='SearchCount').head(10)[1:]
search_terms.set_index('Search')
```

Tablo 4.11: En sık yapılan aramaların gösterildiği tablo

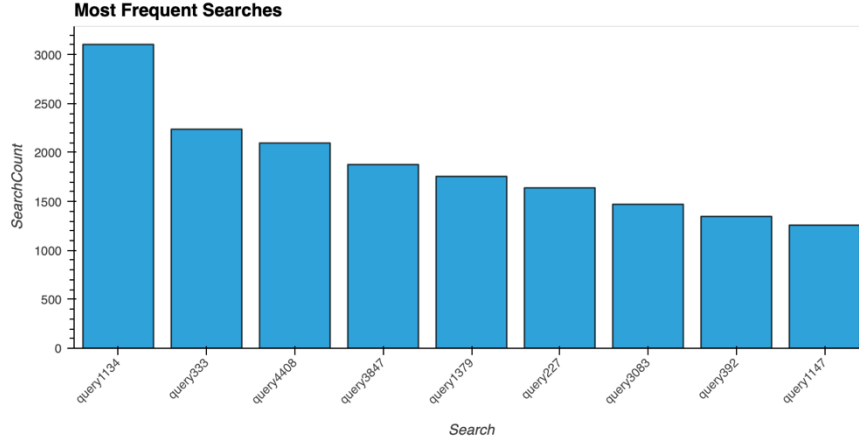
	SearchCount
Search	
query1134	3104
query333	2238
query4408	2097
query3847	1877
query1379	1756
query227	1639
query3083	1471
query392	1347
query1147	1258

Tablo 4.11’de en sık yapılan aramalar analizi ile kullanıcıların web sitesinde en çok hangi aramaları gerçekleştirdiği gösterilmiştir.

```
# En sık yapılan aramalar grafiği
```

```
hv_plot_search_terms = search_terms.hvplot.bar(x='Search', y='SearchCount',
title='Most Frequent Searches', width=800, height=400, rot=45)
```

```
hv_plot_search_terms
```



Şekil 4.8: En sık yapılan aramalar

Şekil 4.8’de en sık aranan terimlere karşılık gelen arama sayıları gösterilmektedir. Arama terimleri x-ekseni üzerinde, arama sayıları ise y-ekseni üzerinde yer almaktadır. Analiz, kullanıcıların hangi ürün veya hizmetlere ilgi gösterdiğini ve web sitesinde en çok hangi konularda arama yaptıklarını ortaya koymaktadır.

Grafikte, query1134 teriminin en sık aranan terim olduğu ve diğer terimlere kıyasla belirgin bir şekilde öne çıktığı görülmektedir. Bu terimi, sırasıyla query333, query4408, query3847 ve diğerleri takip etmektedir. Arama terimlerinin sıklıklarına bakıldığında, bazı terimlerin diğerlerine göre daha fazla arandığı anlaşılmaktadır.

En sık aranan terimlerin belirlenmesi, web sitesi sahiplerine ve pazarlama ekiplerine, kullanıcı taleplerini ve ilgi alanlarını belirleme konusunda yardımcı olmaktadır. Örneğin, en popüler arama terimleri arasında yer alan ürünlerin stoklarının artırılması veya bu ürünler için özel promosyonlar düzenlenmesinin, satışları artırmada etkili olabileceği düşünülmektedir. Ayrıca, kullanıcıların en çok aradığı terimler doğrultusunda yeni ürün veya hizmetlerin eklenmesi de kullanıcı memnuniyetini ve siteye olan bağlılığını artırmakta etkili olmaktadır.

En sık yapılan aramalar, kullanıcıların web sitesinde ne tür bilgilere veya ürünlere ulaşmak istediklerini açıkça göstermekte ve bu veriler, kullanıcı deneyimini iyileştirmek için stratejik kararlar alınmasına yardımcı olmaktadır.

4.3.4 Yönlendiren Siteler

Yönlendiren siteler analizi, bir web sitesine gelen trafiğin hangi dış sitelerden geldiğini belirlemeye yönelik yapılan bir çalışmadır.

```
# Yönlendiren site sayıları
```

```
referrer_sites =  
df['ReferrerSite'].value_counts().reset_index(name='SiteCount').head(10)  
referrer_sites.set_index('ReferrerSite')
```

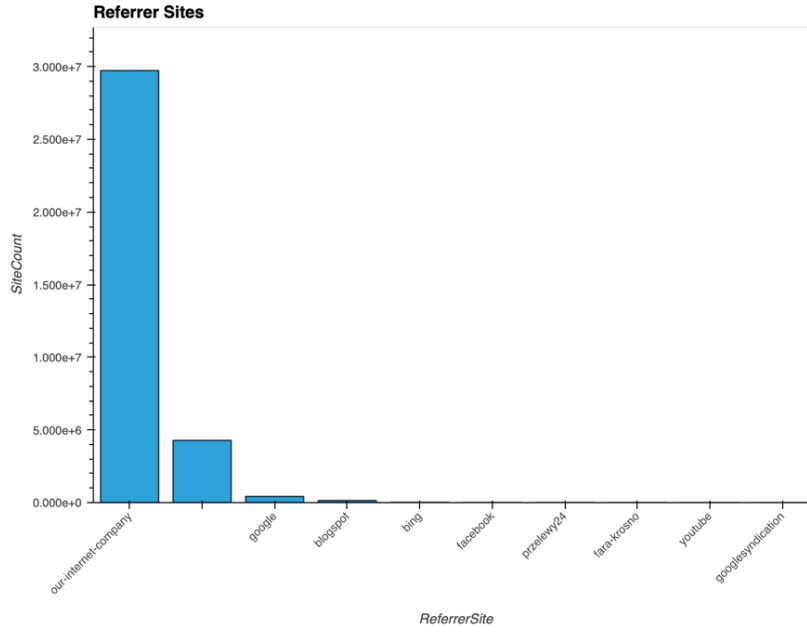
Tablo 4.12: Yönlendiren sitelerin dağılımının gösterildiği tablo

ReferrerSite	SiteCount
our-internet-company	29727357
	4281314
google	421733
blogspot	143701
bing	19225
facebook	10039
przelewy24	4662
fara-krosno	4345
youtube	2252
googlesyndication	2206

Web sitesine gelen trafiğin hangi dış sitelerden geldiği ve bu sitelerin dağılımının nasıl olduğu Tablo 4.12'deki bilgilerden görülmektedir.

```
# Yönlendiren siteler grafiği
```

```
hv_plot_referrer_sites = referrer_sites.hvplot.bar(x='ReferrerSite',  
y='SiteCount', title='Referrer Sites', rot=45, width=800, height=600)  
hv_plot_referrer_sites
```



Şekil 4.9: Yönlendiren siteler

Şekil 4.9’da en çok yönlendirme sağlayan siteler x-ekseni üzerinde, yönlendirme sayıları ise y-ekseni üzerinde gösterilmektedir.

En yüksek yönlendirme sayısını our-internet-company sitesi sağlamaktadır. Bu site analiz edilen web sitesinin kendisi olabilir. Bu durumda, kendi kendine yapılan yönlendirmelerin sayısının yüksek olması, site içi dolaşımın yoğun olduğunu ve kullanıcıların sitede farklı sayfalar arasında geçiş yaptığını göstermektedir.

İkinci sırada yer alan site ise google olup, önemli bir yönlendirme kaynağı olduğu görülmektedir. Bu da kullanıcıların Google arama motoru üzerinden siteye yönlendirildiğini göstermektedir. Diğer yönlendiren siteler arasında blogspot, bing, facebook, przelewy24, fara-krosno, youtube ve googlesyndication gibi siteler bulunmaktadır. Bu siteler, daha düşük yönlendirme sayıları ile katkı sağlamaktadır.

Şekil 4.9’da bazı sitelerin çok düşük yönlendirme sayısına sahip olması da dikkat çekmektedir. Bu durum, bu sitelerden gelen trafiğin nispeten önemsiz olduğunu ve web sitesine önemli bir katkı sağlamadığını göstermektedir.

Yönlendiren siteler analizi, özellikle web sitesi sahiplerine en çok yönlendirme yapan sitelerle ilişkilerin güçlendirilmesi konusunda, bu sitelerde daha fazla pazarlama

çabalarının artırılmasında ve web sitesinin trafiğini ve potansiyel satışlarını artırmada katkı sağlamaktadır.

4.3.5 Başarılı ve Başarısız İstekler

Başarılı ve başarısız istekler analizi, web sitesine yapılan HTTP isteklerinin yanıt kodlarına göre dağılımını incelemektedir.

```
# Başarılı ve başarısız istek sayıları

response_code_counts =
df['ResponseCode'].value_counts(normalize=True).reset_index(name='Proportion (%)')

response_code_counts['Proportion (%)'] = response_code_counts['Proportion (%)'] * 100

response_code_counts.set_index('ResponseCode').map('{:.2f}%'.format).T
```

Tablo 4.13: Başarılı ve başarısız isteklerin gösterildiği tablo

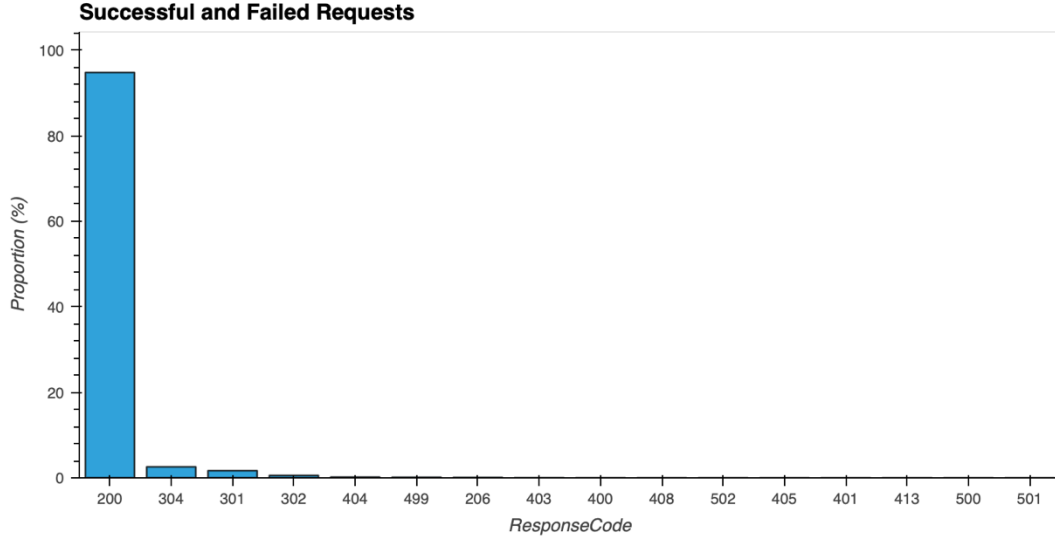
ResponseCode	200	304	301	302	404	499	206	403	400	408	502	405	401	413	500	501
Proportion (%)	94.74%	2.58%	1.68%	0.56%	0.18%	0.14%	0.09%	0.02%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Tablo 4.13’de başarılı ve başarısız istekler yüzdelerle dağılımlı olarak görülmektedir.

```
# Başarılı ve başarısız istekler grafiği

hv_plot_response_codes = response_code_counts.hvplot.bar(x='ResponseCode',
y='Proportion (%)', title='Successful and Failed Requests', width=800,
height=400)

hv_plot_response_codes
```



Şekil 4.10: Başarılı ve başarısız istekler

Başarılı ve başarısız istekler analizi, web sitesine yapılan HTTP isteklerinin yanıt kodlarına göre dağılımını incelemektedir. Şekil 4.10’da x-ekseni üzerinde HTTP yanıt kodları, y-ekseni üzerinde ise her bir kodun aldığı istek oranı (%) gösterilmektedir.

Bu analize göre, 200 yanıt kodu (OK) isteklerin büyük bir çoğunluğunu oluşturmaktadır (%94,74). Bu, web sitesinin kullanıcı taleplerini büyük ölçüde başarıyla karşıladığını ve genel olarak doğru çalıştığını göstermektedir.

Diğer önemli yanıt kodları ise 304 (Not Modified), 301 (Moved Permanently) ve 302 (Found) kodlarıdır. 304 kodu, istemcinin önbelleğindeki kaynağın değişmediğini gösterir ve sunucu kaynaklarının verimli kullanımını işaret etmektedir. 301 ve 302 kodları, istemcinin başka bir URL’ye yönlendirildiğini gösterir ve çeşitli yeniden yönlendirmelerin web sitesinde yaygın olduğunu belirtir.

Dikkate değer diğer kodlar arasında 404 (Not Found) ve 499 (Client Closed Request) bulunmaktadır. 404 kodu, istenen kaynağın bulunamadığını belirtir ve bu kodun oranının %0.18 olması, düzeltilmesi gereken kırık bağlantıların varlığına işaret etmektedir. 499 kodu ise istemcinin isteği kapattığını göstermekte ve bu durumun kullanıcı deneyimini olumsuz etkileyebileceğini ifade etmektedir.

Diğer hata kodlarının çok düşük oranlarda olduğu gözlemlenmektedir. Örneğin, 400 (Bad Request), 403 (Forbidden), 408 (Request Timeout), 500 (Internal Server Error) ve 502 (Bad Gateway) gibi kodların, çeşitli hata durumlarını temsil ettiği

bilinmektedir. Bu tür hata kodlarının düşük oranlarda olması, sunucu tarafında ciddi sorunların nadir olduğunu göstermektedir.

Başarılı isteklerin yüksek oranı, web sitesinin genel olarak doğru çalıştığını gösterirken, başarısız isteklerin oranı ve türleri, iyileştirilmesi gereken alanları belirlemektedir.

Özellikle, yüksek 404 hatası oranı, kullanıcıların istedikleri sayfalara ulaşamadığını ve bu durumun düzeltilmesi gerektiğini göstermektedir. Benzer şekilde, sunucu hataları gibi daha ciddi sorunların hızlı bir şekilde çözülmesi, web sitesinin güvenilirliğini artırmaktadır.

4.3.6 Tekrar Ziyaret Oranı

Tekrar ziyaret oranı analizi, kullanıcıların bir web sitesine ne sıklıkla geri döndüğünü belirlemek için yapılır. Bu analiz, kullanıcı sadakatini ve web sitesine olan ilgiyi ölçmek için önemli bir göstergedir.

```
# Tekrar ziyaret oranı
```

```
repeat_visits = df.groupby('UserId', observed=False).size().reset_index(name='VisitCount')
```

```
repeat_visits['IsRepeatVisitor'] = repeat_visits['VisitCount'] > 1
```

```
repeat_visit_counts = repeat_visits['IsRepeatVisitor'].value_counts(normalize=True).reset_index(name='Proportion (%)')
```

```
repeat_visit_counts['Proportion (%)'] = repeat_visit_counts['Proportion (%)'] * 100
```

```
pd.DataFrame(repeat_visit_counts).set_index('IsRepeatVisitor').map('{:.2f}%'.format)
```

Tablo 4.14: Tekrar ziyaret oranının gösterildiği tablo

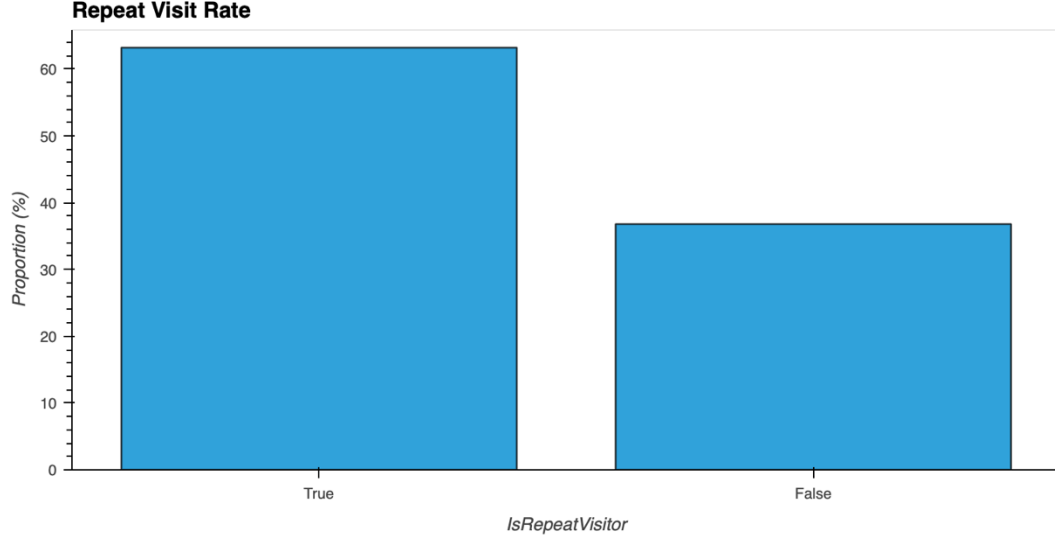
Proportion (%)	
IsRepeatVisitor	
True	63.21%
False	36.79%

Tablo 4.14'te tekrar ziyaret oranı ve dağılımı gösterilmektedir.

```
# Tekrar ziyaret oranı grafiđi
```

```
hv_plot_repeat_visits = repeat_visit_counts.hvplot.bar(x='IsRepeatVisitor',  
y='Proportion (%)', title='Repeat Visit Rate', width=800, height=400)
```

```
hv_plot_repeat_visits
```



Şekil 4.11: Tekrar ziyaret oranı

Şekil 4.11’de tekrar ziyaret eden kullanıcıların oranı ile sadece bir kez ziyaret eden kullanıcıların oranı karşılaştırılmaktadır. Şekil 4.11’den görüldüğü üzere, True sütunu, siteyi birden fazla kez ziyaret eden kullanıcıları temsil ederken, False sütunu, siteyi sadece bir kez ziyaret eden kullanıcıları temsil etmektedir.

Analiz sonuçlarına göre, kullanıcıların yaklaşık %63’ü siteyi tekrar ziyaret ettiği, yani siteye geri dönmüş oldukları görülmektedir. Bu durum, kullanıcıların siteye olan ilgisinin ve sadakatinin yüksek olduğunu göstermektedir. Kullanıcıların büyük bir kısmının siteye tekrar gelmesi, web sitesinin kullanıcıların ihtiyaçlarına ve beklentilerine uygun içerik sunduğunu, kullanıcı deneyiminin olumlu olduğunu ve yüksek ihtimalle pazarlama stratejilerinin etkili olduğunu göstermektedir.

Kullanıcıların yaklaşık %37’sinin siteyi sadece bir kez ziyaret ettiği Şekil 4.11’den görülebilmektedir. Bu oran göz önünde tutularak değerlendirildiğinde, web sitesinin bazı kullanıcılar için yeterince ilgi çekici veya yararlı olmadığı söylenebilmektedir.

Kullanıcıların tekrar ziyaret etmemesinin nedenleri incelenerek web sitesinin bu noktada iyileştirilmesi gereken yönlerinin belirlenmesi gerekmektedir.

Analiz ile, web sitesi sahiplerinin kullanıcı sadakatini artırmak için stratejiler geliştirmesine yardımcı olmak hedeflenmektedir. Tekrar ziyaret oranını artırmak için, kullanıcıların ilgisini çekecek ve onları siteye geri döndürecek kampanyalar, kişiselleştirilmiş içerikler ve kullanıcı dostu deneyimler sunulması gerekmektedir. Siteyi sadece bir kez ziyaret eden kullanıcıların geri bildirimleri alınarak site üzerindeki eksiklikler giderilebilmekte ve kullanıcı deneyimi iyileştirilebilmesi mümkün görünmektedir.

4.3.7 Kullanıcı Segmentasyonu ve Kümeleme Analizi

Kullanıcı segmentasyonu ve kümeleme analizi, e-ticaret siteleri için hayati öneme sahip bir analizdir. Analiz, kullanıcıları benzer davranış ve özelliklerine göre gruplandırarak, her bir segmentin belirgin özelliklerini anlamayı ve pazarlama stratejilerini optimize etmeyi amaçlamaktadır.

```
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

# Kullanıcı metrikleri
user_metrics = df[df['Bot']==False].groupby('UserId',observed=False).agg({
    'Payment': 'sum',
    'Search': 'count',
    'Login': 'sum',
    'Logout': 'sum',
    'Bytes': 'sum',
    'HttpMethod': 'count',
    'Category': 'nunique',
    'Browser': 'nunique',
    'OS': 'nunique',
    'ReferrerSite': 'nunique',
    'Country': 'nunique',
```

```

        'ProductID': 'count'}).reset_index()

# Özelliklerin ölçeklenmesi

scaler = StandardScaler()

scaled_features = scaler.fit_transform(user_metrics.drop(columns=['UserId']))

# Kümeleme analizi

kmeans = KMeans(n_clusters=7, random_state=42)

user_metrics['Cluster'] = kmeans.fit_predict(scaled_features)

# Küme sayıları

cluster_counts = user_metrics['Cluster'].value_counts(normalize=True).reset_index(name='Proportion (%)')

cluster_counts['Proportion (%)'] = cluster_counts['Proportion (%)'] * 100

cluster_counts.set_index('Cluster', inplace=True)

cluster_counts.sort_index().map('{:.2f}%'.format).T

```

Tablo 4.15: Kullanıcı segmentasyonu ve kümeleme analizinin gösterildiği tablo

Cluster	0	1	2	3	4	5	6
Proportion (%)	75.94%	0.00%	11.82%	0.00%	0.05%	10.66%	1.52%

Tablo 4.15'te kullanıcı türleri ve kümeleme analizi yüzdelerle dağılımlarıyla gösterilmektedir.

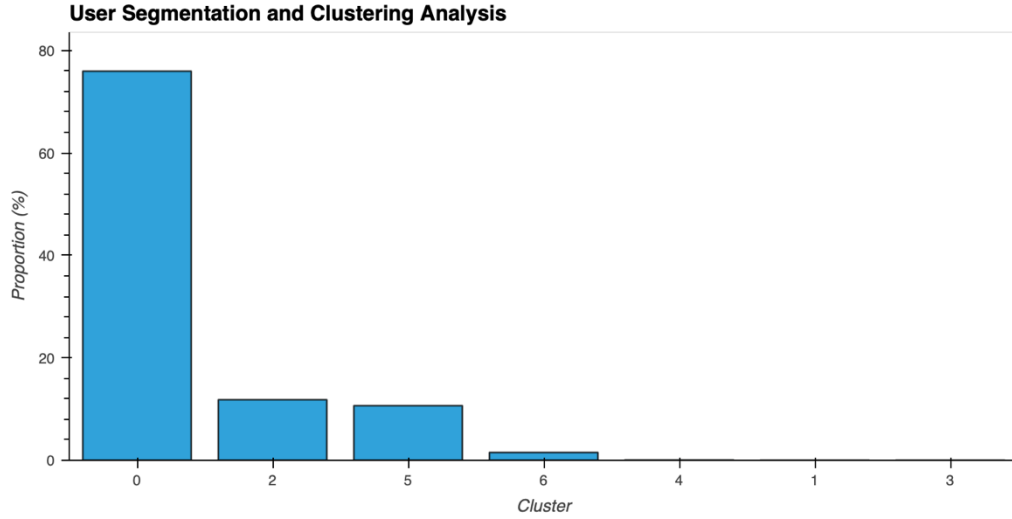
```

# Kullanıcı segmentasyonu ve kümeleme grafiği

hv_plot_clusters = cluster_counts.hvplot.bar(x='Cluster', y='Proportion (%)', title='User Segmentation and Clustering Analysis', width=800, height=400)

hv_plot_clusters

```



Şekil 4.12: Kullanıcı segmentasyonu ve kümeleme analizi

Şekil 4.12’de k-means kümeleme algoritması kullanılarak elde edilen kullanıcı kümeleri ve bu kümelerdeki kullanıcı oranları gösterilmektedir. Analiz sonucunda yedi farklı kullanıcı kümesi oluşturulmaktadır. Her küme, kullanıcıların web sitesindeki davranışlarına ve etkileşimlerine göre belirlenmiştir:

Küme 0: En büyük kullanıcı grubunu içermektedir (%75,94). Bu kullanıcılar, genel olarak web sitesiyle yüksek etkileşimde bulunan ve çeşitli ürün veya hizmetlere ilgi gösteren geniş bir kullanıcı kitlesini temsil etmektedir.

Küme 1, 3 ve 4: Çok küçük oranlarda kullanıcı içermektedir (%0.00). Bu segmentler muhtemelen düşük etkileşimde bulunan veya çok spesifik davranışlar sergileyen kullanıcıları temsil etmektedir.

Küme 2: İkinci en büyük segmenttir (%11,82). Bu kullanıcıları, belirli ürünlere veya kategorilere yüksek ilgi gösteren ve nispeten daha büyük bir grup oluşturmaktadır.

Küme 5: Orta büyüklükte bir kullanıcı grubunu içermektedir (%10,66). Bu segment, belirli davranışlar ve etkileşimler sergileyen kullanıcıları barındırmaktadır.

Küme 6: Daha küçük bir kullanıcı grubu içermektedir (%1.52). Bu kullanıcıların, spesifik ürünlere veya hizmetlere yüksek ilgi gösteren bir grup olabileceği düşünülmektedir.

Bu segmentlerin her biri, farklı pazarlama stratejileri ve kullanıcı deneyimi iyileştirmeleri için önemli ipuçları sunmaktadır. Örneğin, en büyük segment olan Küme 0, genel kullanıcı kitlesini hedefleyen kampanyalar ve promosyonlarla desteklenebilirken, daha küçük segmentler (Küme 2 ve Küme 5) özel teklifler veya kişiselleştirilmiş içeriklerle motive edilebilmektedir.

Kullanıcı segmentasyonu, pazarlama stratejilerinin daha etkin bir şekilde planlanmasına ve uygulanmasına olanak tanımaktadır. Bu sayede, her bir kullanıcı segmentinin ihtiyaçları ve beklentileri doğrultusunda özelleştirilmiş stratejiler geliştirilerek müşteri memnuniyeti artırılabilir ve dönüşüm oranları iyileştirilebilmektedir.

Kullanıcı segmentasyonu ve kümeleme analizi, e-ticaret sitelerinin kullanıcı davranışlarını daha iyi anlamalarına ve daha etkili pazarlama stratejileri geliştirmelerine yardımcı olmaktadır. Analiz, kullanıcı bağlılığını artırmak ve satışları optimize etmek için değerli bilgiler sağlamaktadır.

4.3.8 Bot ve Kullanıcı Trafik Karşılaştırması

Bot ve Kullanıcı Trafik Karşılaştırması, web sitesine gelen bot trafiği ile gerçek kullanıcı trafiği arasındaki farkları göstermektedir.

```
# Bot ve kullanıcı trafiği sayıları

bot_traffic = df[df['Bot']] == True].groupby('Date').size().reset_index(name='BotRequests')

user_traffic = df[df['Bot']] == False].groupby('Date').size().reset_index(name='UserRequests')

# Verilerin birleştirilmesi

traffic_comparison = pd.merge(bot_traffic, user_traffic, on='Date',
how='outer').fillna(0)

traffic_comparison.set_index('Date')
```

Tablo 4.16: Bot ve kullanıcı trafiği karşılaştırmasının gösterildiği tablo

	BotRequests	UserRequests
Date		
2019-12-01	34864	293228
2019-12-02	27021	331691
2019-12-03	18692	276669
2019-12-04	16233	236339
2019-12-05	15525	247458
...
2020-05-27	18061	226691
2020-05-28	20045	185526
2020-05-29	19837	163642
2020-05-30	16670	174663
2020-05-31	18695	206932

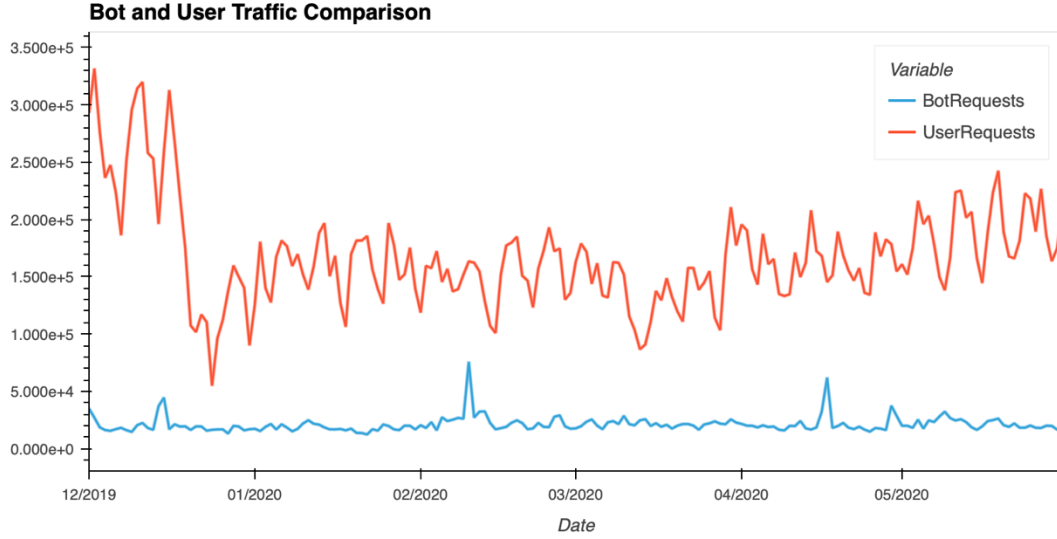
Tablo 4.16’da bot ve kullanıcı trafiği karşılaştırması gösterilmektedir.

Bu analiz ile botların ve gerçek kullanıcıların siteye yaptıkları isteklerin zaman içindeki dağılımını inceleyerek bot trafiğinin yoğun olduğu dönemleri belirlemeye yardımcı olmaktadır. Bot ve kullanıcı trafiği karşılaştırmasının grafiksel şekilde gösterimi Şekil 4.13’de yer almaktadır:

Bot ve kullanıcı trafiği grafiği

```
hv_plot_traffic_comparison = traffic_comparison.hvplot.line(x='Date',  
y=['BotRequests', 'UserRequests'], title='Bot and User Traffic Comparison',  
width=800, height=400).opts(legend_position='top_right')
```

```
hv_plot_traffic_comparison
```



Şekil 4.13: Bot ve kullanıcı trafik karşılaştırma

Şekil 4.13’de kırmızı çizgi, kullanıcıların siteye yaptığı istekleri (UserRequests), mavi çizgi ise botların siteye yaptığı istekleri (BotRequests) temsil etmektedir. Kullanıcı istekleri, bot isteklerine göre belirgin bir şekilde daha yoğundur ve günlük olarak dalgalanmalar göstermektedir. Kullanıcı trafiği, belirli zaman aralıklarında artış ve azalışlar sergileyerek, kullanıcıların belirli dönemlerde siteye daha fazla ilgi gösterdiğini ortaya koymaktadır.

Özellikle yılın başında kullanıcı trafiğinde belirgin bir artış gözlemlenmektedir. Kullanıcı istekleri, genel olarak bot isteklerinden çok daha yüksektir ve bu durum, web sitesinin kullanıcılar tarafından aktif olarak kullanıldığını göstermektedir. Bot trafiği ise kullanıcı trafiğine göre oldukça düşük kalmaktadır ve zaman içinde düzenli bir seyir izlemektedir.

Bot trafiği, kullanıcı trafiğine kıyasla daha sabit bir dağılım göstermektedir. Bot istekleri, belirli zaman dilimlerinde hafif artışlar göstermekle birlikte, genel olarak düşük seviyelerde seyretmektedir. Bu durum, botların belirli bir program dahilinde siteyi taradığını veya belirli zamanlarda aktif olduklarını göstermektedir. Web sitesinin güvenlik önlemlerini optimize etmek için de analizden yararlanılmaktadır. Bot trafiğinin yoğun olduğu dönemlerin belirlenmesi, güvenlik duvarları ve diğer önlemler aracılığıyla zararlı bot aktivitelerinin engellenmesine yardımcı olmaktadır.

Ayrıca, kullanıcı trafiğinin yüksek olduğu dönemlerde sitenin performansını iyileştirmek için gerekli optimizasyonların yapılması gerektiğini ortaya koymaktadır. Bot ve kullanıcı trafiği karşılaştırması, web sitesine gelen isteklerin niteliğini anlamak ve site güvenliği ile performansını optimize etmek için kritik bir analiz olarak kabul edilmektedir. Bu tür analizler, e-ticaret sitelerinin kullanıcı deneyimini iyileştirmesine ve potansiyel güvenlik tehditlerine karşı proaktif önlemler almasına yardımcı olmaktadır. Bu şekilde hem kullanıcı memnuniyeti artırılmakta hem de sitenin genel güvenliği sağlanabilmektedir.

4.3.9 Anomali Tespiti

Anomali tespiti, web sitesinin trafik verilerinde anormallikleri belirlemek için yapılan bir analizdir. Bu analiz, web sitesinde yaşanan beklenmedik olayları veya olası güvenlik tehditlerini tespit etmek için kritik bir yöntem olarak bilinmektedir. E-ticaret siteleri gibi yüksek trafiğe sahip platformlarda, bu tür analizler büyük önem taşımaktadır.

```
from sklearn.ensemble import IsolationForest

# Trafik verisinin hazırlanması
traffic_data = df.groupby('Date').size().reset_index(name='RequestCount')
traffic_data['Date'] = pd.to_datetime(traffic_data['Date'])
traffic_data.set_index('Date', inplace=True)

# IsolationForest ile anomali tespiti
model = IsolationForest(contamination=0.05)
traffic_data['Anomaly'] = model.fit_predict(traffic_data[['RequestCount']])
traffic_data[traffic_data['Anomaly'] == -1].T
```

Tablo 4.17: Anomali tespitinin gösterildiği tablo

Date	2019-12-01	2019-12-02	2019-12-03	2019-12-09	2019-12-10	2019-12-11	2019-12-15	2019-12-16	2019-12-24	2019-12-31
RequestCount	328092	358712	295361	310541	334713	342254	303679	329865	71337	107087
Anomaly	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

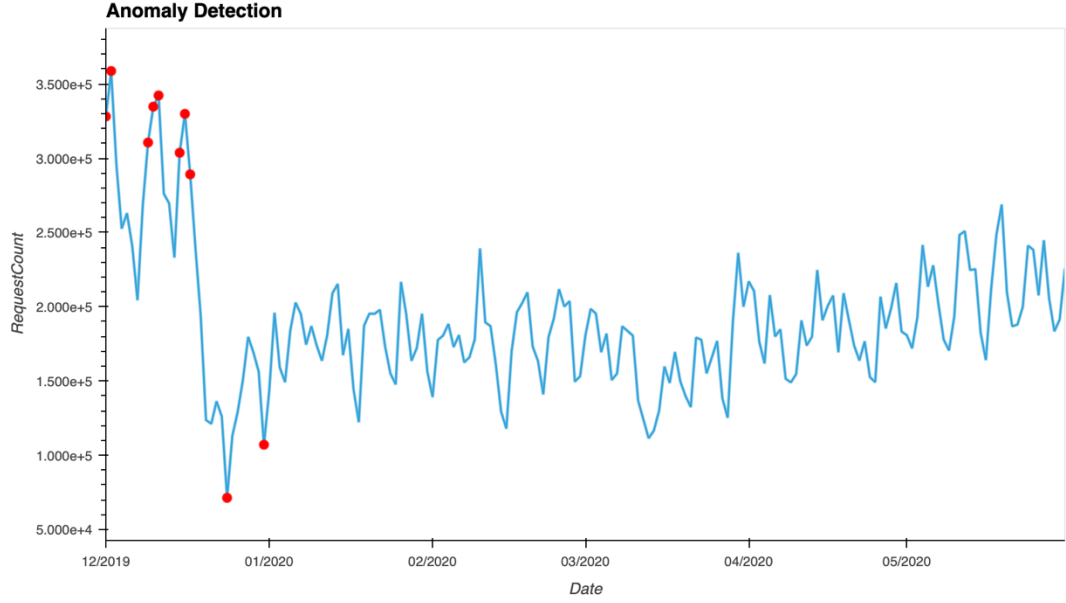
Tablo 4.17’de web sitesinin trafik verilerinde var olan anormallikler gösterilmeye çalışılmaktadır.

```
# Anomali tespiti grafiđi
```

```
anomalies = traffic_data[traffic_data['Anomaly'] == -1]
```

```
hv_plot_anomalies = traffic_data.hvplot.line(y='RequestCount',  
title='Anomaly Detection', width=900, height=500) *  
anomalies.hvplot.scatter(y='RequestCount', color='red', size=50)
```

```
hv_plot_anomalies
```



Şekil 4.14 Anomali tespiti

Şekil 4.14'te günlük istek sayılarının zaman içindeki deđişimi mavi bir çizgi ile gösterilmektedir. Kırmızı noktalar ise tespit edilen anomalileri temsil etmektedir. Anomaliler, trafik verilerinde normalden sapmalar olarak tanımlanmakta ve genellikle olađandışı yüksek veya düşük istek sayıları ile kendini göstermektedir. Bu analizde, anomali tespiti için İzolasyon Ormanı (Isolation Forest) algoritması kullanılmaktadır. Bu algoritma, veri setindeki olađandışı noktaları belirlemek için yaygın olarak kullanılan bir yöntemdir. İzolasyon Ormanı, veri noktalarını rastgele alt kümelere ayırarak her bir veri noktasının izole edilme süresini ölçmekte ve kısa sürede izole edilebilen noktalar, genellikle anomali olarak kabul edilmektedir.

Şekil 4.14'te görüldüğü gibi, özellikle belirli dönemlerde trafik verilerinde anormallikler tespit edilmiştir. Bu anormallikler, trafik akışında olađandışı deđişiklikler olduğunu göstermektedir. Örneđin, bazı dönemlerde istek sayılarında ani artışlar veya düşüşler gözlemlenmektedir. Bu tür anomaliler, web sitesinde

beklenmedik olaylar, kullanıcı davranışlarındaki ani değişiklikler veya potansiyel güvenlik tehditlerine işaret etmektedir. Analiz, web sitesinin güvenliğini sağlamak ve performansını iyileştirmek için önemli bilgiler sunmaktadır. Anomalilerin tespit edilmesi, web sitesinin yönetim ekibine olası sorunları hızlı bir şekilde tanımlama ve çözüme imkânı sağlamaktadır.

Ayrıca, anormal trafik davranışlarının belirlenmesi, gelecekte benzer durumların önlenmesi için stratejik önlemler alınmasına yardımcı olmaktadır. Anomali tespiti, web sitesinin sağlıklı ve güvenli bir şekilde çalışmasını sağlamak için kritik bir araçtır. Bu tür analizlerde e-ticaret sitelerinin kullanıcı deneyimini iyileştirmesine ve potansiyel tehditlere karşı proaktif önlemler almasına katkı sağlamaktadır. Bu sayede, web sitesinin genel performansı artırılmakta ve kullanıcı memnuniyeti sağlanmaktadır.

4.3.10 İşletim Sistemi ve Tarayıcı Tercihleri

İşletim Sistemi ve Tarayıcı Tercihleri analizi, kullanıcıların hangi cihaz ve yazılımları kullandığını belirlemek için yapılan bir incelemedir. Bu analiz, web sitesinin uyumluluğunu ve kullanıcı deneyimini optimize etmek için yardımcı bir yöntem olarak bilinmektedir.

```
# İşletim sistemi sayıları
```

```
os_preferences = df[df['Bot']] ==  
False]['OS'].value_counts().reset_index(name='OSCount').head(10)
```

```
os_preferences.set_index('OS').T
```

Tablo 4.18: İşletim sistemi tercihlerinin gösterildiği tablo

OS	Windows	Android	iOS	Mac OS X	Ubuntu	Linux	Windows Phone	Fedora	Other	Chrome OS
OSCount	14500812	13450559	1857595	626001	110073	96685	44149	36384	30133	8552

```
# Tarayıcı sayıları
```

```
browser_preferences = df[df['Bot']] ==  
False]['Browser'].value_counts().head(10).reset_index(name='BrowserCount')
```

```
browser_preferences.set_index('Browser').T
```

Tablo 4.19: Tarayıcı tercihlerinin gösterildiği tablo

Browser	Chrome Mobile	Chrome	Firefox	Samsung Internet	Mobile Safari	Opera	Facebook	Edge	IE	Safari
BrowserCount	10242751	7794618	5610087	1582233	1533863	1023645	658758	605398	357738	323320

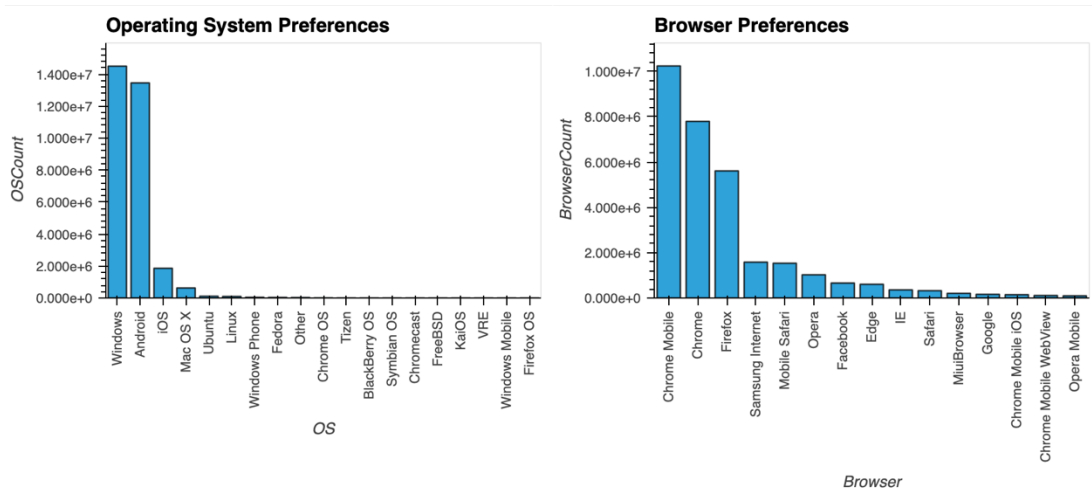
Tablo 4.18 ve Tablo 4.19’da web sitesine erişim sağlayan kullanıcıların tercih ettikleri işletim sistemleri ve tarayıcı tercihleri gösterilmektedir.

İşletim sistemi ve tarayıcı grafiği

```
hv_plot_os_preferences = os_preferences.hvplot.bar(x='OS', y='OSCount',  
title='Operating System Preferences', xlabel='OS', width=450, height=400,  
rot=90)
```

```
hv_plot_browser_preferences = browser_preferences.hvplot.bar(x='Browser',  
y='BrowserCount', title='Browser Preferences', width=450, height=400, rot=90)
```

```
hv_plot_os_preferences + hv_plot_browser_preferences
```



Şekil 4.15: Kullanıcıların tercih ettikleri işletim sistemleri ve tarayıcı tercihleri

Şekil 4.15’te görüldüğü gibi, kullanıcıların en çok tercih ettiği işletim sistemi Windows’tur. Windows’u, Android işletim sistemi takip etmektedir. Bu iki işletim sistemi, kullanıcıların büyük bir çoğunluğunu temsil etmektedir. iOS, Mac OS X ve Linux gibi diğer işletim sistemleri ise daha az kullanılmaktadır.

Şekil 4.15’te Chrome tarayıcısının açık ara en çok kullanılan tarayıcı olduğu görülmektedir. Chrome Mobile ve Chrome tarayıcıları, toplamda en yüksek kullanıcı sayısına sahiptir. Bu, Chrome’un hem masaüstü hem de mobil platformlarda yaygın olarak tercih edildiğini göstermektedir. Firefox, Samsung Internet ve Mobile Safari

gibi diğer tarayıcılar da kullanıcılar arasında popülerdir, ancak Chrome'un gerisinde kalmaktadır. Opera, Edge, IE ve Safari gibi tarayıcılar ise daha az kullanılmaktadır.

Bu analizler web sitesi optimizasyonu ve kullanıcı deneyimini iyileştirmek için önemli bilgiler sunmaktadır. Örneğin, en yaygın kullanılan tarayıcı ve işletim sistemlerine göre web sitesi optimizasyonları yapılabilmekte ve test süreçleri bu platformlar üzerinde yoğunlaştırılabilir. Bu bilgiler ayrıca, pazarlama stratejilerinin hedef kitlenin teknolojik tercihlerine uygun olarak şekillendirilmesine yardımcı olabilmektedir. Bu şekilde, kullanıcı deneyimi artırılabilir ve web sitesinin erişilebilirliği maksimize edilebilir.

Genel olarak bakıldığında; kullanıcıların büyük çoğunluğunun Windows işletim sistemi ve Chrome tarayıcısını tercih ettiği anlaşılmaktadır. Bu veriler, web sitesi geliştirme ve optimizasyon çalışmalarında dikkate alınarak, daha geniş bir kullanıcı kitlesine daha iyi hizmet sunulabilmek için yapılması düşünülen hareketlerde ciddi oranda etkili olmaktadır.

```
# İşletim sistemi ve tarayıcı verilerinin birleştirilmesi
```

```
os_browser_preferences=pd.DataFrame(df[df['Bot'] == False][['OS',  
'Browser']].value_counts(normalize=True)[:10])
```

```
os_browser_preferences=os_browser_preferences['proportion']*100
```

```
# Stil ayarları
```

```
styles = [
```

```
    {'selector': 'th.col_heading.level0', 'props': [('border', '1px solid  
black'), ('text-align', 'center')]},
```

```
    {'selector': 'th.col_heading.level1', 'props': [('border', '1px solid  
black'), ('text-align', 'center')]},
```

```
    {'selector': 'th', 'props': [('border', None), ('text-align',  
'center')]}]
```

```
# İşletim sistemi ve tarayıcı bilgilerinin stil ile gösterilmesi
```

```
pd.DataFrame(os_browser_preferences.map('{:.2f}%'.format)).T.style.set_table_styles(styles)
```

Tablo 4.20: İşletim sistemleri ve tarayıcı tercihlerinin gösterildiği tablo

OS	Android		Windows		iOS		Windows		Windows	
	Chrome Mobile	Chrome	Firefox	Samsung Internet	Mobile Safari	Opera	Edge	Facebook	Chrome	IE
proportion	33.29%	22.96%	17.66%	5.13%	4.99%	3.25%	1.97%	1.96%	1.34%	1.16%

Tablo 4.20’de işletim sistemleri ve tarayıcı tercihlerinin yüzdelik dağılımları gösterilmektedir. Tablodan hareketle Chrome ve Chrome Mobile tarayıcılarının daha çok tercih edildiği görülmektedir.

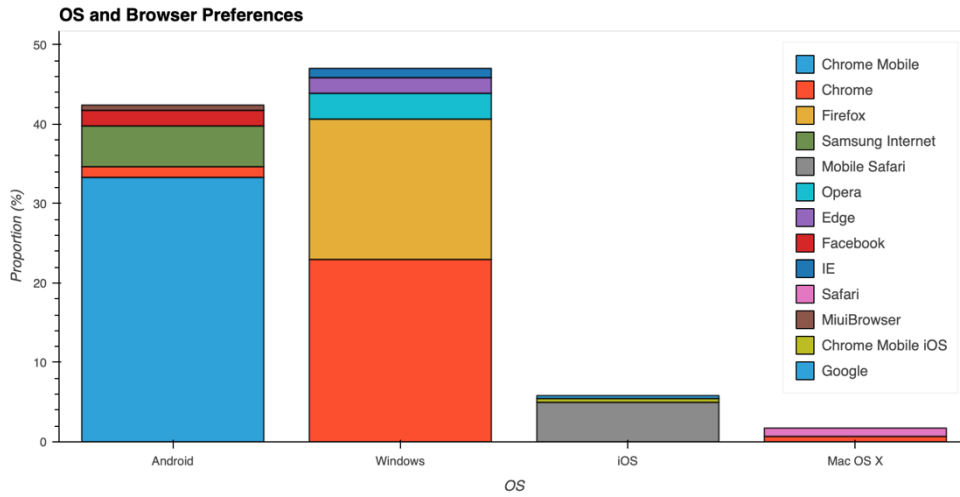
```
import warnings
```

```
warnings.filterwarnings('ignore', category=FutureWarning)
```

İşletim sistemi ve tarayıcı tercihleri grafiği

```
hv_os_browser_preferences=pd.DataFrame(os_browser_preferences).hvplot.bar(s  
tacked=True,title='OS and Browser Preferences', width=900,ylabel='Proportion  
(%)', legend="top_right", height=450)
```

```
hv_os_browser_preferences
```



Şekil 4.16: İşletim sistemleri ve tarayıcı tercihleri

Şekil 4.16’da web sitesine erişim sağlayan kullanıcıların tercih ettikleri işletim sistemleri ve bu sistemlerde kullanılan tarayıcıların dağılımı gösterilmektedir. Bu analiz, kullanıcıların hangi platformları ve tarayıcıları tercih ettiklerini belirlemek için yapılmaktadır.

İşletim Sistemi Tercihleri:

Şekil 4.16'da görüldüğü gibi, kullanıcıların en çok tercih ettiği işletim sistemleri arasında Windows, Android ve iOS öne çıkmaktadır. Windows işletim sistemi en yüksek orana sahip olup, kullanıcıların yaklaşık %40'ı tarafından tercih edilmektedir.

Android işletim sistemi de benzer bir oranla ikinci sırada yer almaktadır. iOS ve Mac OS X gibi diğer işletim sistemleri daha düşük oranlarda kullanılmaktadır. Bu veriler, web sitesinin kullanıcı kitlesinin büyük ölçüde Windows ve Android işletim sistemlerini tercih ettiğini göstermektedir.

Yine Şekil 4.16'dan tarayıcı tercihlerinin işletim sistemlerine göre farklılık gösterdiği görülmektedir:

- Windows kullanıcıları arasında en çok tercih edilen tarayıcılar Chrome ve Firefox'dur.
- Android kullanıcıları ise ağırlıklı olarak Chrome Mobile ve Samsung Internet tarayıcılarını tercih etmektedir.
- iOS kullanıcıları ise genellikle Mobile Safari tarayıcısını kullanmaktadır. Diğer tarayıcılar, tüm işletim sistemleri genelinde daha az tercih edilmektedir.

Şekil 4.16'daki yığılılı çubuklar, her bir işletim sisteminde kullanılan tarayıcıların oranlarını göstermektedir. Örneğin, Windows işletim sistemi kullanıcılarının büyük bir kısmı Chrome tarayıcısını kullanırken, küçük bir kısmı Firefox ve Internet Explorer gibi diğer tarayıcıları tercih etmektedir.

Android kullanıcılarının çoğu Chrome Mobile tarayıcısını kullanırken, Samsung Internet ve diğer tarayıcılar daha az tercih edilmektedir. Kullanıcıların büyük çoğunluğunun Windows ve Android işletim sistemlerini ve Chrome tarayıcısını tercih ettiği anlaşılmaktadır.

4.3.11 En Çok Ziyaret Edilen Ürünler ve Kategoriler

En çok ziyaret edilen ürünler ve kategoriler analizi, web sitesinin kullanıcı davranışlarını anlamak ve bu doğrultuda önemli kararlar almak için değerli öngörüler kazandırmaktadır. Bu tür analizlerle e-ticaret sitelerinin daha etkili pazarlama ve satış stratejileri geliştirmesine yardımcı olunmaktadır.

```
# En çok ziyaret edilen ürün oranı

top_selling_products
df['ProductID'].value_counts(normalize=True).reset_index(name='Proportion (%)').head(10)

top_selling_products['ProductID']=top_selling_products['ProductID'].astype(int)

top_selling_products['Proportion (%)']*100

top_selling_products.set_index('ProductID').map('{:.2f}%'.format).T
```

Tablo 4.21: En çok ziyaret edilen ürünlerin gösterildiği tablo

ProductID	634	290	80	1161	1195	1529	1750	570	1158	194
Proportion (%)	4.35%	2.14%	0.98%	0.90%	0.74%	0.72%	0.64%	0.57%	0.49%	0.48%

Tablo 4.21’de web sitesi ziyaretçilerinin en çok ziyaret ettiği ürünler ürün koduna göre ve ziyaret edilme sıklığına göre yüzdelik dağılımlarıyla gösterilmektedir.

```
# En çok ziyaret edilen kategori oranı

top_categories
df['Category'].value_counts(normalize=True).reset_index(name='Proportion (%)').head(10)

top_categories['Proportion (%)']*100

top_categories.set_index('Category').map('{:.2f}%'.format).T
```

Tablo 4.22: En çok ziyaret edilen kategorilerin gösterildiği tablo

Category	30	33	21	23	39	34	25	24	35	79
Proportion (%)	35.91%	19.55%	19.18%	11.10%	2.20%	2.16%	1.45%	0.99%	0.99%	0.96%

Tablo 4.22’de de web sitesi ziyaretçilerinin en çok ziyaret ettiği ürün kategorileri kategori numarasına ve yüzdelik dağılımlarına göre gösterilmektedir.

```
# En çok ziyaret edilen ürünler ve kategoriler grafiği

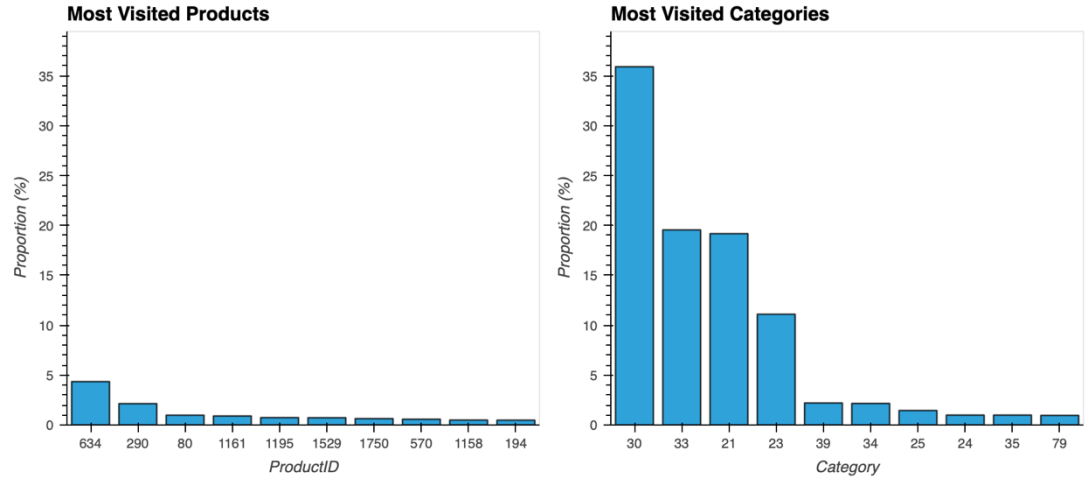
hv_plot_top_selling_products
top_selling_products.head(10).hvplot.bar(x='ProductID', y='Proportion (%)',
title='Most Visited Products', width=450, height=400)
```

```

hv_plot_top_categories = top_categories.head(10).hvplot.bar(x='Category',
y='Proportion (%)', title='Most Visited Categories', width=450, height=400)

hv_plot_top_selling_products + hv_plot_top_categories

```



Şekil 4.17: En çok ziyaret edilen ürünler ve kategoriler

Şekil 4.17’de web sitesinde en çok ziyaret edilen ürünler ve kategoriler gösterilmektedir. Analiz ile kullanıcıların en çok ilgi gösterdiği ürünler ve bu ürünlerin ait olduğu kategoriler belirlenmeye çalışılmaktadır.

Şekil 4.17’de görüldüğü gibi, bazı ürünler diğerlerine kıyasla çok daha fazla ziyaret edilmektedir. Örneğin, ProductID 634, 290 ve 80 en çok ziyaret edilen ürünler arasında yer almaktadır. Bu ürünlerin, kullanıcıların büyük bir kısmının ilgisini çekmiş olduğu ve sıkça ziyaret edildiği görülmektedir. Diğer ürünlerin ise nispeten daha az ziyaret edildiği söylenebilmektedir.

Bu veriler, kullanıcıların belirli ürünlere yönelik yüksek bir ilgi gösterdiğini ve bu ürünlerin satış potansiyelinin yüksek olduğunu göstermektedir.

Kategoriler bazında yapılan analizde, bazı kategorilerin diğerlerine göre daha popüler olduğu görülmektedir. Şekil 4.16’da en çok ziyaret edilen kategoriler gösterilmektedir. Örneğin, Category 30, 33 ve 21 en çok ziyaret edilen kategoriler arasında yer almaktadır. Bu veriler, belirli kategorilerin kullanıcılar tarafından daha fazla tercih edildiğini ve bu kategorilere yönelik pazarlama stratejilerinin güçlendirilmesi gerektiğini göstermektedir.

Şekil 4.17'deki çubuklar, her bir ürün ve kategori için ziyaret oranlarını göstermektedir. En çok ziyaret edilen ürün ve kategorilerin belirlenmesi, web sitesinin kullanıcı deneyimini optimize etmek ve satış stratejilerini geliştirmek için önemli bilgiler sunmaktadır.

4.3.12 En Çok Satılan Ürünler ve En Popüler Kategoriler

En çok satılan ürünler ve en popüler kategoriler analizi, e-ticaret sitesinin performansını artırmak için önemli veriler sunmaktadır. Bu tür analizlerle, hem kullanıcı memnuniyeti sağlanmakta hem de ticari başarıyı maksimize etmek için stratejik kararlar alınmasına yardımcı olunmaktadır.

```
# En çok satılan ürün sayıları
```

```
top_selling_products = df[df['Payment'] == True]['ProductID'].value_counts().reset_index(name='PaymentCount').head(15)

top_selling_products['ProductID']=top_selling_products['ProductID'].astype(int)

top_selling_products.set_index('ProductID').T
```

Tablo 4.23: En çok satılan ürünlerin ürün koduna göre gösterildiği tablo

ProductID	28	33	97	70	79	69	2017	98	474	1158	570	569	1910	71	640
PaymentCount	36	33	28	27	26	26	25	24	20	20	20	19	19	19	18

Tablo 4.23'te en çok satılan ürünler ürün koduna ve satış sayılarının yüzdelik dağılımlarına göre gösterilmektedir

```
# En çok satılan kategori oranı
```

```
top_categories = df[df['Payment'] == True]['Category'].value_counts(normalize=True).reset_index(name='Proportion (%)').head(10)

top_categories['Proportion (%)']*100

top_categories.set_index('Category').map('{:.2f}%'.format).T
```


Tablo 4.24: En popüler kategorilerin gösterildiği tablo

Category	21	25	30	60	78	75	74	71	70	67
Proportion (%)	98.81%	0.79%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

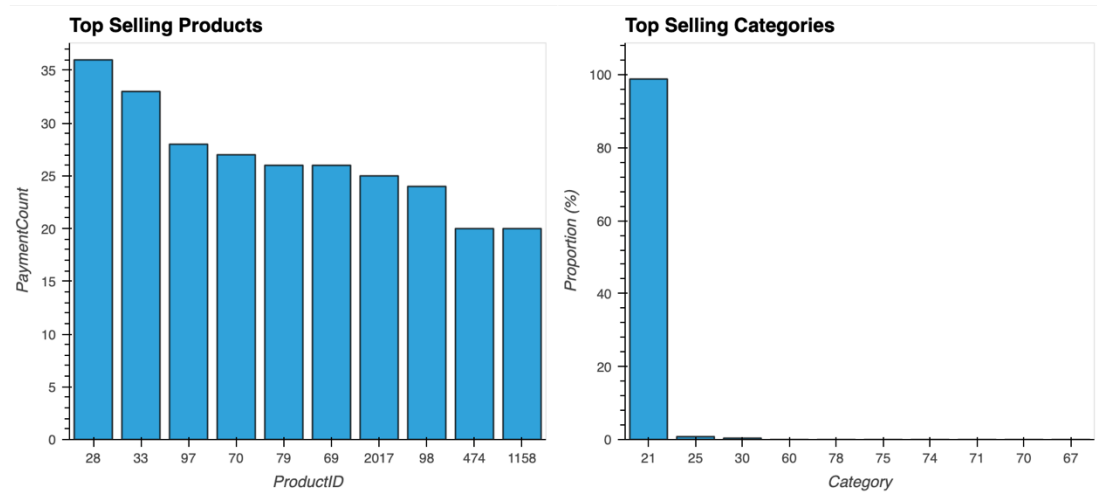
Tablo 4.24'te de en popüler kategorilerin neler olduğu gösterilmektedir. 21 numaralı kategorinin açık farkla en çok satılan ürün kategorisi olduğu görülmektedir. Tablolardan yararlanarak sayfa kullanıcılarının hangi ürün ve hangi kategoriye daha fazla ilgi gösterdiği yorumlanabilmektedir. En popüler kategoriler belirlenerek; pazarlama stratejilerini ve stok yönetimini optimize etmek için kullanılabilir.

En çok satılan ürünler ve en popüler kategoriler grafiği

```
hv_plot_top_selling_products = top_selling_products.head(10).hvplot.bar(x='ProductID', y='PaymentCount', title='Top Selling Products', width=450, height=400)

hv_plot_top_categories = top_categories.head(10).hvplot.bar(x='Category', y='Proportion (%)', title='Top Selling Categories', width=450, height=400)

hv_plot_top_selling_products + hv_plot_top_categories
```



Şekil 4.18: En çok satılan ürünler ve en popüler kategoriler

Şekil 4.18'de e-ticaret web sitesinde en çok satılan ürünler ve en popüler kategoriler belirlenmeye çalışılmaktadır. Analizler, kullanıcıların hangi ürün ve kategorilere ilgi gösterdiğini anlamak ve stok yönetimi ile pazarlama stratejilerini optimize etmek amacıyla gerçekleştirilmektedir.

Şekil 4.18’de en çok satılan ürünlerin ProductID’leri ve satış sayıları gösterilmektedir. Bu kullanıcıların en çok satın aldığı ürünleri belirlemek için yapılmaktadır. En çok satılan ürünlerin belirlenmesi, satış stratejilerini geliştirmek açısından önemli bilgiler sağlamaktadır. ProductID 28, 33 ve 97 en çok satılan ürünler arasında yer almaktadır. Bu ürünler, web sitesinin en popüler ve en çok talep gören ürünleridir. Bu bilgilerle satış kampanyaları ve envanter yönetimi daha etkin bir şekilde planlanabileceği düşünülmektedir. Diğer ürünler ise nispeten daha az satılmıştır, bu da bu ürünlerin tanıtımının artırılması veya stok yönetiminde farklı stratejiler uygulanması gerektiğini göstermektedir.

Şekil 4.18’de kullanıcıların en çok tercih ettiği kategoriler ve bu kategorilerin kullanım sayıları gösterilmektedir. Bu analiz, kullanıcıların hangi ürün kategorilerine daha fazla ilgi gösterdiğini belirlemek için yapılmaktadır. En popüler kategoriler, pazarlama stratejilerini ve stok yönetimini optimize etmek için kullanılabilir. Category 21, 25 ve 30 en popüler kategoriler arasında yer almaktadır ve bu kategorilerdeki ürünler, kullanıcılar tarafından daha fazla tercih edilmektedir. Bu bilgiler, web sitesinin ürün çeşitliliğini ve kategori bazında stratejilerini belirlerken önemli bir rehberlik sağlamaktadır.

4.3.13 Hafta Günlerine ve Aylara Göre Satın Alımlar

Analiz, kullanıcıların hangi günlerde ve aylarda daha fazla satın alma gerçekleştirdiğini belirlemek için yapılmaktadır. Bu pazarlama stratejilerini optimize etmek ve satışları artırmak için yapılmaktadır.

Haftalık Satın Alımlar

#Haftalık satın alım sayıları

```
daily_payments = df[df['Payment'] == True].groupby('DayOfWeek').size().reset_index(name='PaymentCount')
```

```
daily_payments['DayOfWeek'] = pd.Categorical(daily_payments['DayOfWeek'], categories=days, ordered=True)
```

```
daily_payments.sort_values('DayOfWeek').set_index('DayOfWeek').T
```

Tablo 4.25: Hafta günlerine göre satın alımların gösterildiği tablo

DayOfWeek	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
PaymentCount	673	1522	413	743	929	813	1411

Tablo 4.25'te hafta günlerine göre yapılan satın alımlar gösterilmektedir.

Aylık Satın Alımlar

```
#Aylık satın alım sayıları
```

```
monthly_payments = df[df['Payment'] == True].groupby('Month').size().reset_index(name='PaymentCount')
monthly_payments['Month'] = pd.Categorical(monthly_payments['Month'],
categories=months, ordered=True)
monthly_payments.sort_values('Month').set_index('Month').T
```

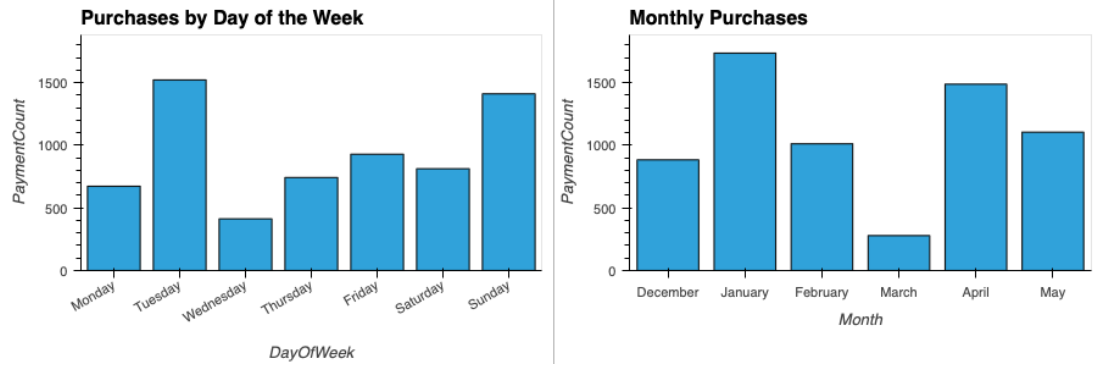
Tablo 4.26: Aylara göre satın alımların gösterildiği tablo

Month	December	January	February	March	April	May
PaymentCount	884	1736	1012	279	1488	1105

Tablo 4.26'da aylara göre yapılan satın alımlar gösterilmektedir.

```
# Hafta günlerine ve aylara göre satın alma grafiği
```

```
hv_daily_payments =
daily_payments.sort_values('DayOfWeek').hvplot.bar(x='DayOfWeek',
y='PaymentCount', title='Purchases by Day of the Week', rot=30, width=450,
height=300)
hv_monthly_payments =
monthly_payments.sort_values('Month').hvplot.bar(x='Month',
y='PaymentCount', title='Monthly Purchases', width=450, height=300)
hv_daily_payments + hv_monthly_payments
```



Şekil 4.19: Haftanın günlerine ve aylara göre satın alımlar

Şekil 4.19’da haftanın günlerine ve aylara göre web sitesinde yapılan satın alma işlemlerinin dağılımı incelenmektedir. Analiz, kullanıcıların hangi günlerde ve aylarda daha fazla satın alma gerçekleştirdiğini belirlemek için yapılmaktadır. Pazar ve salı günleri, en yüksek ödeme sayısına sahiptir. Bu günlerde kullanıcıların daha aktif olup, alışveriş yapma eğiliminde olduğu görülmektedir. Pazartesi ve çarşamba günü alım sayısında belirgin bir düşüş gözlemlenirken, Perşembe ve cuma günleri tekrar artış görülmektedir. Çarşamba günü alışverişlerin en düşük seviyede olduğu gün olarak görülürken, Pazar günü, kullanıcıların hafta sonu alışverişlerini yapması nedeniyle ödeme sayısında bir artış görülmektedir.

Aylara göre yapılan satın alma işlemleri incelendiğinde Ocak ayında, diğer aylara kıyasla en yüksek ödeme sayısı kaydedilmektedir. Bu durum, tatil sezonu sonrası indirimlerin etkisinin olabileceği düşünülmektedir. Aralık ayındaki yüksek alışveriş sayısı ise yılbaşı alışverişlerinden kaynaklanmaktadır. Şubat ve mart ayındaki düşük alışveriş sayısı ise bir duraklama veya sezon dışı dönemi işaret edebileceği düşünülmektedir. Nisan ayında ödemeler nispeten dengeli ve yüksek seviyelerde seyrederken, mayıs ayında bir miktar azalma olduğu gözlemlenmektedir.

Analiz ile web sitesindeki satış performansını optimize etmek için değerli bilgiler elde edilebilmektedir. Haftanın belirli günlerinde ve yılın belirli aylarında yapılan kampanyalarla, kullanıcıların alışveriş eğilimlerine göre zamanlanabilir şekilde olduğu düşünülmektedir. Benzer şekilde, Ocak ayı gibi yoğun alışveriş dönemlerinde yapılan özel kampanyalarla, kullanıcı katılımının ve satışların çok fazla hale gelebileceği düşünülmektedir.

4.3.14 Kategori ve Ürün Bazında Satın Alımlar

Kullanıcıların en çok hangi kategorilerde ve ürünlerde satın alma gerçekleştirdiği belirlenmeye çalışılmaktadır. Bu analizle; hangi ürünlerin ve kategorilerin popüler olduğunu anlamak kolaylaşmaktadır.

```
# Kategori ve ürün bazında satın alma sayıları
```

```
category_product_purchases = df[df['Payment']==True].groupby(['Category',  
'ProductID']).size().reset_index(name='PurchaseCount')
```

```
category_product_purchases = category_product_purchases.sort_values(by='PurchaseCount',  
ascending=False).head(20)
```

```
category_product_purchases['ProductID']=category_product_purchases['Product  
ID'].astype('int')
```

```
category_product_purchases.set_index('ProductID').T
```

Tablo 4.27: Kategori ve ürün bazında satın almaların gösterildiği tablo

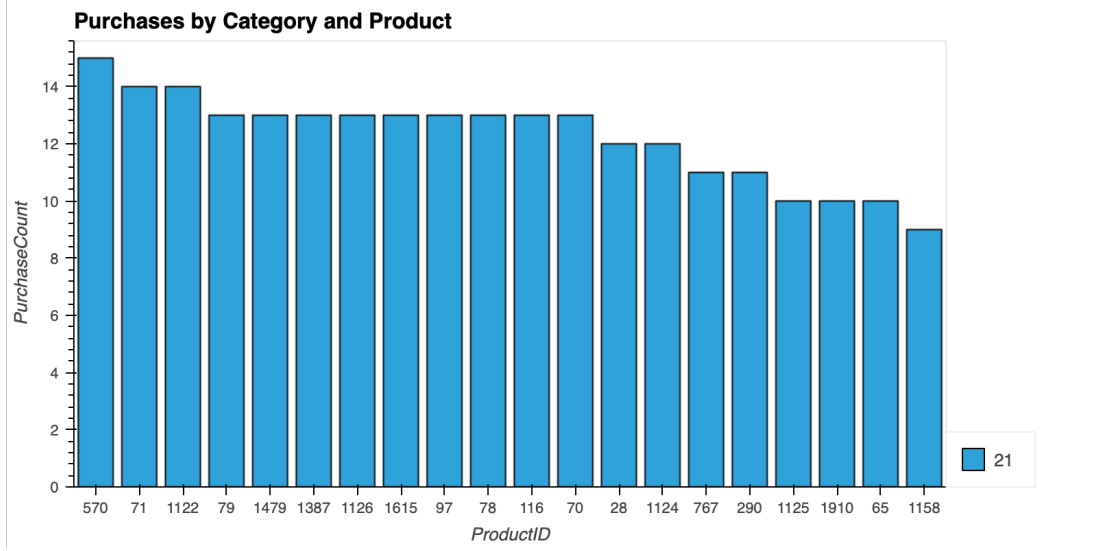
ProductID	570	71	1122	79	1479	1387	1126	1615	97	78	116	70	28	1124	767	290	1125	1910	65	1158
Category	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
PurchaseCount	15	14	14	13	13	13	13	13	13	13	13	13	12	12	11	11	10	10	10	9

Tablo 4.27’de kategori ve ürün bazında yapılan satın alma analizi, kullanıcıların en çok hangi kategorilerde ve ürünlerde satın alma gerçekleştirdiği görülebilmektedir.

```
# Kategori ve ürün bazında satın almalar grafiği
```

```
hv_category_product_purchases = category_product_purchases.hvplot.bar  
(x='ProductID', y='PurchaseCount', by='Category', stacked=True,  
title='Purchases by Category and Product', width=800, height=400)
```

```
hv_category_product_purchases
```



Şekil 4.20: Kategori ve ürün bazında satın alımlar

Şekil 4.20’de kategori ve ürün bazında yapılan satın alma analizi gösterilmektedir. Her kategoride en çok satın alınan ürünler ve bu ürünlerin satın alma sayıları gösterilmektedir. Satın alma sayıları y-ekseni üzerinde, ürün ID'leri ise x-ekseni üzerinde yer almaktadır. Ayrıca, her bir sütunun rengi, ürünlerin ait olduğu kategoriyi temsil etmektedir. Görüldüğü gibi, kategori 21'deki ürünler en yüksek satın alma sayısına sahiptir. ProductID'si 570 olan ürün, en fazla satın alınan üründür ve bu ürünün yüksek talep gördüğü anlaşılmaktadır. Diğer popüler ürünler arasında ProductID'leri 71 ve 1122 olanlar bulunmaktadır. Bu ürünler, kullanıcıların sıkça tercih ettiği ve satın aldığı ürünlerdir.

Kategori 21'deki ürünlerin yüksek satın alma sayıları, bu kategorideki ürünlerin müşteriler arasında popüler olduğunu göstermektedir. Bu bilgi, stok yönetimi ve pazarlama kampanyalarının popüler ürünler ve kategoriler üzerine yoğunlaştırılması gerektiğini işaret etmektedir. Düşük talep gören ürünler ve kategoriler için ise etkili kararlar alınarak satışların artırılmasına katkı sağlayacak yolların araştırılması gerekmektedir.

4.3.15 Giriş ve Çıkış İşlemleri Analizi

Günlük login ve logout işlemlerinin dağılımı, kullanıcıların siteyle olan etkileşimlerini ve davranışlarını anlamak için önemli veriler sağlar. Bu veriler, web sitesinin performansını optimize etmek ve kullanıcı memnuniyetini artırmak için kullanılabilir.

Giriş İşlemleri (Login) Dağılımı

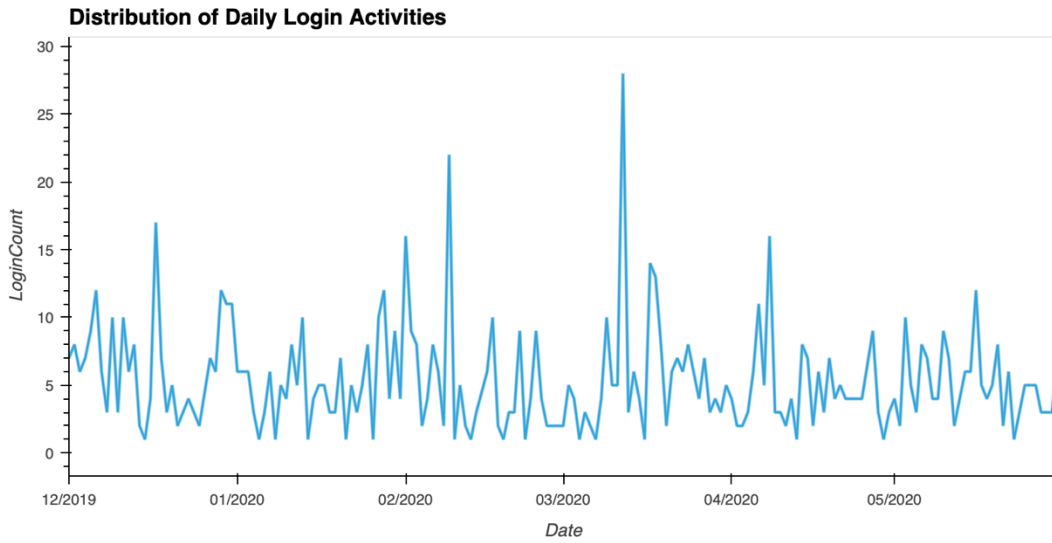
```
# Login işlemlerinin dağılım sayıları
```

```
daily_login_data = df[df['Login']].groupby('Date').size().reset_index(name='LoginCount')
```

```
# Login işlemlerinin dağılım grafiği
```

```
hv_daily_login_data = daily_login_data.hvplot.line(x='Date', y='LoginCount',  
title='Distribution of Daily Login Activities', xlabel='Date',  
ylabel='LoginCount', width=800, height=400)
```

```
hv_daily_login_data
```



Şekil 4.21: Giriş işlemleri dağılımı

Şekil 4.21’de kullanıcıların web sitesine giriş yapma davranışlarının zaman içerisindeki dağılımı gösterilmektedir. Günlük login işlemlerinin sayıları, zaman serisi olarak y-ekseni üzerinde ve tarihler x-ekseni üzerinde yer almaktadır.

Belirli zaman dilimlerinde giriş sayılarında önemli dalgalanmalar ve zirveler olduğu görülmektedir. Özellikle Mart 2020’nin sonlarına doğru giriş sayılarında belirgin artış dikkat çekmektedir. Bu dönemlerde, site tarafından özel kampanyalar, indirimler veya tanıtım etkinlikleri düzenlenmiş olabileceği, bunun da kullanıcıların siteye olan ilgisini artırmış olabileceği düşünülmektedir.

Giriş işlemlerinin düzenli olarak devam ettiği ve kullanıcıların siteye düzenli olarak giriş yaptığı gözlemlenmektedir. Bu durum, kullanıcıların siteye olan bağlılığından ve

sürekli etkileşimde olmalarından kaynaklanmaktadır. Kullanıcıların siteye sık sık giriş yapması, web sitesinin kullanıcılar için değerli ve ilgi çekici olduğunu ifade etmektedir.

Siteye giriş sayılarının bazı dönemlerde oldukça düşük olduğu dikkat çekmektedir. Bu dönemlerde, site trafiğini artırmak için stratejilerin gözden geçirilmesi ve gerekli önlemlerin alınması gereklidir.

Düşük giriş dönemleri, kullanıcıların ilgisinin azaldığı veya siteye erişimde sorunlar yaşandığı zaman dilimlerini işaret etmektedir. Bu durumlar, site yönetimi tarafından çok iyi ve dikkatli bir biçimde analiz edilerek, kullanıcıların siteye olan ilgisini artıracak stratejilerin geliştirilmesi gerekmektedir.

Çıkış İşlemleri (Logout) Dağılımı

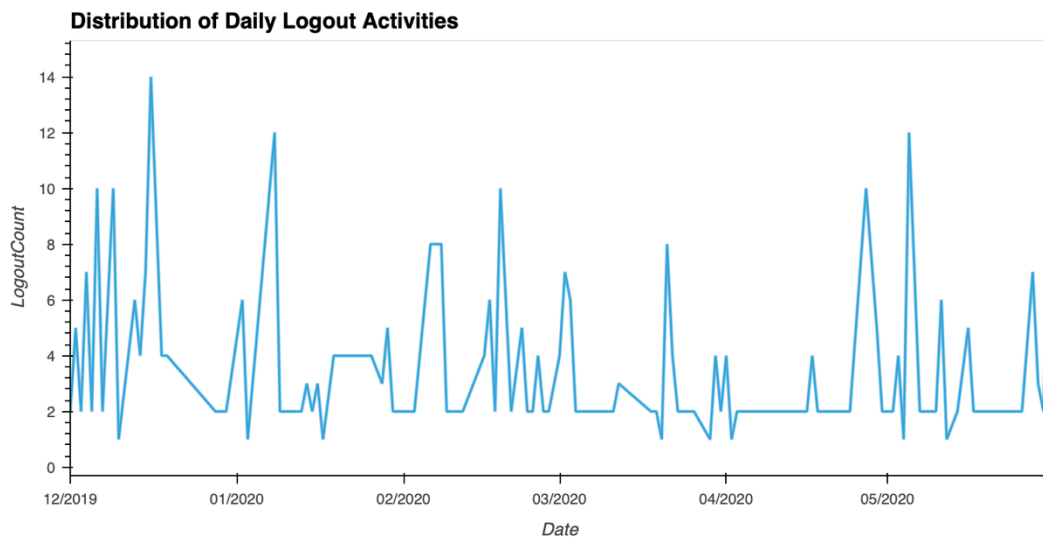
```
# Logout işlemlerinin dağılım sayıları
```

```
daily_logout_data = df[df['Logout']].groupby('Date').size().reset_index(name='LogoutCount')
```

```
# Logout işlemlerinin dağılım grafiği
```

```
hv_plot_logout = daily_logout_data.hvplot.line(x='Date', y='LogoutCount', title='Distribution of Daily Logout Activities', xlabel='Date', ylabel='LogoutCount', width=800, height=400)
```

```
hv_plot_logout
```



Şekil 4.22: Çıkış işlemleri dağılımı

Şekil 4.22’de kullanıcıların web sitesinden çıkış (logout) yapma davranışlarının zaman içerisindeki dağılımı gösterilmektedir. Günlük logout işlemlerinin sayıları, y-ekseni üzerinde ve tarihler x-ekseni üzerinde yer almaktadır.

Belirli dönemlerde logout işlemlerinin sayısında belirgin artışlar olduğu görülmektedir. Özellikle Ocak 2020'nin başlarında logout işlemlerinde belirgin zirve dikkat çekmektedir. Bu dönemlerde, kullanıcıların siteye giriş yaptıktan sonra kısa süre içinde çıkış yaptıkları gözlemlenmektedir. Bu durum, kullanıcıların siteyi kullanırken yaşadığı deneyimlerin veya karşılaştıkları sorunların bir göstergesi olarak değerlendirilmektedir.

Logout işlemlerinin düzenli olarak devam ettiği ve kullanıcıların siteye giriş yapıp çıkış yaptığı genel bir eğilim görülmektedir. Bu, kullanıcıların siteyi aktif olarak kullandığını ve belirli bir etkileşim süresi geçirdikten sonra çıkış yaptıklarını göstermektedir. Kullanıcıların düzenli olarak logout yapması, siteye olan bağlılıklarını ve etkileşimde bulduklarını göstermektedir.

Bazı dönemlerde logout işlemlerinin sayısının oldukça düşük olması; bu dönemlerde, kullanıcıların siteyi daha uzun süre kullanmış olabileceğinden veya siteye giriş yapmadan siteyi terk etmiş olabileceği ihtimalinden kaynaklanmaktadır. Düşük logout sayıları, kullanıcıların siteye olan ilgisinin arttığı veya giriş işlemlerinden sonra siteyi daha fazla keşfettikleri dönemlere işaret etmektedir. Logout işlemlerinin zaman içindeki dağılımı, kullanıcıların siteye olan bağlılığı ve kullanım davranışları hakkında önemli bilgiler sunmaktadır. Bu veriler, site yönetimi tarafından analiz edilerek, kullanıcı deneyimini iyileştirmek ve kullanıcıların site üzerinde daha uzun süre kalmalarını sağlamak için etkili olmaktadır.

Giriş ve Çıkış İşlemlerinin Karşılaştırılması

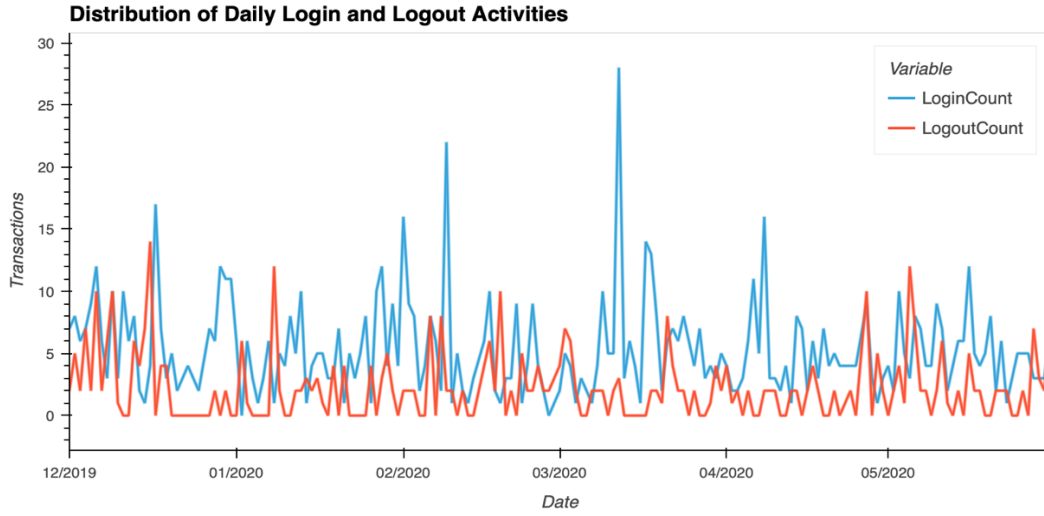
```
# Login ve Logout verilerinin birleştirilmesi
```

```
login_logout_data = pd.merge(daily_login_data, daily_logout_data, on='Date',  
how='outer').fillna(0)
```

```
# Giriş ve çıkış işlemlerinin karşılaştırılma grafiği
```

```
hv_plot_login_logout = login_logout_data.hvplot.line(x='Date',  
y=['LoginCount', 'LogoutCount'], title='Distribution of Daily Login and  
Logout Activities', xlabel='Date', ylabel='Transactions', width=800,  
height=400).opts(legend_position='top_right')
```

```
hv_plot_login_logout
```



Şekil 4.23: Giriş çıkış işlemlerinin karşılaştırılması

Şekil 4.23'te e-ticaret web sitesine yapılan günlük giriş (login) ve çıkış (logout) işlemlerinin zaman içindeki dağılımı gösterilmektedir. Mavi çizgi login işlemlerini, kırmızı çizgi ise logout işlemlerini temsil etmektedir.

Login işlemlerinin sayısının logout işlemlerinin sayısından daha yüksek olduğu görülmektedir. Bu durum, kullanıcıların siteye giriş yaptıktan sonra çıkış yapmama eğiliminde olduklarını veya tarayıcıyı kapatarak oturumlarını sonlandırdıklarını gösterebilir. Mart 2020'nin sonlarına doğru login işlemi belirgin zirve dikkat çekmektedir. Bu dönemlerde login sayılarının artması, kullanıcıların bu dönemlerde siteye olan ilgisinin arttığını veya belirli kampanyalar veya etkinlikler nedeniyle siteye daha fazla giriş yaptıklarını gösterebilir.

Logout işlemleri, login işlemlerine kıyasla daha düşük sayılarda gerçekleşmiştir, ancak yine de belirli dönemlerde logout sayılarında artış gözlemlenmektedir. Bu artışlar, kullanıcıların siteyi belirli bir süre kullandıktan sonra çıkış yaptıklarını gösterebilir. Login ve logout işlemleri arasındaki farklar, kullanıcıların site üzerinde ne kadar süre kaldıklarına ve siteye olan bağlılıklarına dair ipuçları verebilir. Login sayılarının logout sayılarından fazla olması, kullanıcıların siteye giriş yaptıktan sonra daha uzun süre aktif olduklarını gösterebilir.

Bu analiz, kullanıcıların siteye giriş ve çıkış yapma eğilimlerini anlamak için önemlidir. Kullanıcı davranışlarının incelenmesi, site yönetimi tarafından kullanıcı deneyimini iyileştirmek için stratejiler geliştirilmesine yardımcı olabilir. Örneğin,

belirli dnemlerde login sayılarındaki artışlar dikkate alınarak bu dnemlerde daha fazla kaynak ayırmak veya kullanıcıların siteyi daha uzun süre kullanmalarını sağlamak için çeşitli iyileştirmeler yapılabilir. Benzer şekilde, logout sayılarındaki artışların nedenleri araştırılarak kullanıcıların siteyi terk etme nedenleri tespit edilip gerekli önlemler alınabilir.

Bölüm 5

Sonuç ve Öneriler

Bu çalışmada, bir e-ticaret web sitesine ait sunucu günlüklerinin analizi gerçekleştirilmiş ve bu analizler sonucunda web sitesinin performansı, kullanıcı davranışları ve trafik dinamikleri detaylı bir şekilde incelenmiştir. Wrocław Bilim ve Teknoloji Üniversitesi tarafından sağlanan EClog veri seti kullanılarak yapılan bu analizler, web sitesinin farklı yönlerini ortaya koyarak kullanıcı etkileşimlerini daha iyi anlamamıza olanak tanımıştır. Elde edilen bulgular, web sitesinin mevcut durumunu değerlendirmek ve gelecekteki stratejik kararları şekillendirmek için değerli içgörüler sağlamıştır.

Genel trafik dağılımında, Pazar ve pazartesi günleri ile Aralık ve Mayıs aylarında yoğun bir trafik gözlemlenmiştir. Bu bulgular, pazarlama stratejilerinin optimize edilmesi için önemli ipuçları sunmaktadır. Ükelere göre trafik dağılımında, Polonya'dan gelen trafik %84,29 gibi çok yüksek bir orana sahiptir. Amerika Birleşik Devletleri %7.07, Hollanda %3.15, Almanya %2,60 oranında trafikle sıralanmaktadır. Kullanıcı davranışları analizinde, kullanıcıların en çok aradıkları terimler, ziyaret ettikleri ürünler ve yönlendiren siteler gibi veriler, stok yönetimi ve ürün geliştirme çalışmalarında dikkate alınmalıdır.

HTTP yanıt kodlarının analizi, web sitesinin performansını ve olası sorunlarını değerlendirmek için kritik bilgiler sağlamaktadır. Özellikle, 404 ve 500 hata kodlarının oranları incelenerek iyileştirme alanları belirlenmiştir. Bu doğrultuda web sitesinin performansını iyileştirmek amacıyla, bu hata kodlarının detaylı bir analizinin yapılması ve potansiyel sorunların belirlenip çözümlenmesi büyük önem taşımaktadır. Ayrıca, tekrar ziyaret oranı analizleri, kullanıcıların büyük bir kısmının siteyi tekrar ziyaret ettiğini ve yüksek bir kullanıcı sadakati olduğunu ortaya koymaktadır.

Kullanıcı segmentasyonu ve kümeleme analizi, pazarlama stratejilerinin kişiselleştirilmesi açısından büyük önem taşımaktadır. Farklı kullanıcı gruplarına özel

kampanyalar ve ürün önerileri sunularak, kullanıcı memnuniyeti artırılabilir ve satışlar yükseltilebilir.

Analizlerden elde edilen sonuçlar doğrultusunda, e-ticaret web sitesinin performansını artırmaya yönelik çeşitli stratejik öneriler sunulabilir. İlk olarak, pazarlama stratejilerinin optimize edilmesi gerekmektedir. Bu bağlamda, özellikle Pazar ve Pazartesi günleri ile tatil sezonlarında özel indirim kampanyaları düzenlenerek satışlar artırılabilir. Ayrıca, stok yönetimi ve ürün geliştirme çalışmalarının titizlikle ele alınması, kullanıcı taleplerine daha etkin ve hızlı bir şekilde yanıt verilmesini sağlayarak müşteri memnuniyetini artırabilir.

Güvenlik ve anomali tespiti çalışmaları, web sitesinin güvenliğini ve performansını optimize etmek için önemli bilgiler sunmuştur. Bot trafiğinin tespiti ve engellenmesi, web sitesinin performansını artırmada kritik bir rol oynar. Güvenlik önlemlerinin artırılması da göz ardı edilmemelidir. Bot trafiğinin tespiti ve engellenmesi gibi güvenlik tedbirleri, web sitesinin güvenliğini sağlamak ve kullanıcı deneyimini olumsuz etkileyen otomatik trafiği minimize etmek adına kritik bir rol oynamaktadır.

Tekrar ziyaret oranı analizleri, sadık müşterilere yönelik özel kampanyaların düzenlenmesi gerekliliğini vurgulamakla birlikte, kullanıcı geri bildirimlerinin sistematik bir şekilde toplanması gerekir ve bu geri bildirimlere dayalı olarak iyileştirme çalışmalarının da yapılması gerekmektedir.

Son olarak, mobil uyumluluğun artırılması, en çok kullanılan işletim sistemine ait mobil uygulama geliştirilmesi ve çok kanallı pazarlama stratejilerinin kullanılması, web sitesinin erişilebilirliğini ve etkinliğini artırabilir. Bu önerilerin uygulanması, e-ticaret web sitesinin performansını ve kullanıcı memnuniyetini artırarak rekabet avantajını sürdürmesine katkı sağlayabilir.

Bölüm 6

Gelecek Çalışmalar

Bu çalışmanın sonuçları, web sitesinin mevcut performansını ve kullanıcı davranışlarını anlamada önemli bilgiler sağlamıştır. Ancak, ileri düzey veri ve analiz teknikleriyle daha derinlemesine incelemeler de yapılabilir. Gelecekteki araştırmalar aşağıdaki alanlara odaklanabilir:

- Gerçek zamanlı analizler için gerçek zamanlı veri analitiği araçları kullanılarak anlık trafik dalgalanmaları ve kullanıcı davranışları izlenebilir, böylece hızlı müdahale ve optimizasyon sağlanabilir.
- Makine öğrenimi ve yapay zekâ teknikleri kullanılarak kullanıcı davranışları ve trafik dinamikleri daha iyi anlaşılabilir, segmentasyon, tahmin modelleri ve kişiselleştirilmiş öneri sistemleri geliştirilebilir.
- Kullanıcıların web sitesinde geçirdikleri süre, ziyaret ettikleri sayfalar ve yaptıkları işlemler detaylı olarak analiz edilerek kullanıcı yolculuğu daha iyi anlaşılabilir. Bu analizler, kullanıcı deneyimini iyileştirmek için önemli içgörüler sağlar.
- Kullanıcıların web sitesine farklı kanallardan (örneğin sosyal medya, e-posta, arama motorları) nasıl geldikleri ve bu kanalların etkinliği analiz edilerek çok kanallı pazarlama stratejileri geliştirilebilir.
- Bot trafiğinin detaylı analizi ve zararlı bot aktivitelerinin tespiti için gelişmiş güvenlik önlemleri ve analitik yöntemler uygulanabilir.

Bu araştırmalar, web sitesinin performansını ve kullanıcı deneyimini geliştirmek için stratejik kararlar almak için önemli bilgiler sağlayacaktır. Bu da web sitesinin kullanıcı tabanını genişleterek rekabet avantajını korumasına ve kullanıcı memnuniyetini artırmasına yardımcı olur.

Bölüm 7

Kaynakça

- [1] Chodak, G., Suchacka, G., & Chawla, Y. (2020). HTTP-level e-commerce data based on server access logs for an online store. *Computer networks*, 183, 107589.
- [2] ÖZSEVEN, T., & DÜĞENCİ, M. (2011). LOG Analiz: Erişim Kayıt Dosyaları Analiz Yazılımı ve GOP Üniversitesi Uygulaması. *Bilişim Teknolojileri Dergisi*, 4(2).
- [3] Yalçınkaya, B., Gedikli, M. E., & Cibaroğlu, M. O. (2019). Elektronik Belge Yönetim Sisteminde Log Analizi: İstatistiksel Bir Değerlendirme. *Bilgi Yönetimi ve Bilgi Güvenliği: eBelge-eArşiv-eDevlet-Bulut Bilişim-Büyük Veri-Yapay Zekâ, Yayına Hazırlayanlar: Bahattin Yalçınkaya, M. Altay Ünal, Burcu Yılmaz, Fahrettin Özdemirci, Ankara: Ankara Üniversitesi*, 61-77.
- [4] Jansen, B. J., Spink, A., & Taksa, I. (Eds.). (2008). *Handbook of research on web log analysis*. IGI Global.
- [5] Çınar, I., & Bilge, H. Ş. (2016). Web Madenciliği Yöntemleri ile Web Loglarının İstatistiksel Analizi ve Saldırı Tespiti. *Bilişim Teknolojileri Dergisi*, 9(2), 125.
- [6] Çelik, S. (2017). Web günlük dosyalarının analizi için web kullanım madenciliğinin uygulanması. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 46(1), 62-75.
- [7] Tyler, M. E., & Ledford, J. L. (2006). *Google analytics*. John Wiley & Sons.
- [8] Ürkek, Y. (2020). Google Analytics' in ölçümlene ile birlikte internet gazeteciliğine getirdiği yenilikler. *Medya ve Kültürel Çalışmalar Dergisi*, 2(2), 15-26.

[9] İnternet: Clickstream analysis, (2024). What is Clickstream Analysis. Alındığı tarih: 07.06.2024, adresi: <https://www.macrometa.com/articles/what-is-clickstream-analysis>

[10] AKBAŞ, E. Bilgi Güvenliği ve Log Yönetimi Sistemlerinin Analizi. *Web Sayfası*: https://www.researchgate.net/profile/Ertugrul-Akbas-3/publication/356592793_Bilgi_Guvenligi_ve_Log_Yonetimi_Sistemlerinin_Analizi_Ertugrul_AKBAS/links/61a35da67323543e210f062d/Bilgi-Guevenligi-ve-Log-Yoenetimi-Sistemlerinin-Analizi-Ertugrul-AKBAS.pdf Erişim Tarihi: Haziran 2024).

[11] Yalçınkaya, B. A. H. A. T. T. İ. N., Ünal, M., Yılmaz, B., & Özdemirci, F. (2019). Bilgi Yönetimi Ve Bilgi Güvenliği: eBelge-eArşiv-eDevlet-Bulut Bilişim-Büyük Veri-Yapay Zekâ.