



KONAKLAMA ÖNERİ SİSTEMİ: VERİ ANALİZİ VE KİŞİSELLEŞTİRİLMİŞ TAVSİYELER

Yazılım Mühendisliği Ana Bilim Dalı

Dönem Projesi

Nergis Durmazgezer Çakmak

ORCID 0009-0003-3660-0219

Proje Danışmanı: Dr. Öğr. Üyesi Osman Gökalp

Şubat 2024

Konaklama Öneri Sistemi: Veri Analizi ve Kişiselleştirilmiş Tavsiyeler

ÖZ

Seyahat ve konaklama sektörü, dijital platformların hızlı yükselişiyle birlikte önemli bir dönüşüm geçirmiştir. Bu sektördeki dönüşümün öncülerinden biri olan Airbnb platformu, seyahat etmek isteyen insanlara dünya genelinde çeşitli konaklama deneyimleri sunmuştur.

Projenin temel amacı, Airbnb platformunun geniş veri setlerini kullanarak kişiselleştirilmiş konaklama önerileri sunan bir öneri sistemi geliştirmektir. Bu kapsamda, 'Ülke/Şehir' bazlı güncel veri setlerinden South Aegean Greece (Ege Yunan Adaları) veri setleri analiz edilmiş, konaklama türleri, fiyatlar, konum bazlı özellikler incelenerek kullanıcıların seyahat amaçlarına ve bütçelere uygun önerilerde bulunulması hedeflenmiştir. Geliştirilen öneri sisteminde, metin madenciliği ve benzerlik analizlerini içeren bir yaklaşım baz alınarak, kullanıcıların terim girişlerine en uygun konaklama yerlerini belirlemek için TF-IDF vektörleme ve Cosine Similarity kullanılmıştır. Böylece, kullanıcıların kişisel tercihlerine ve bütçelerine en uygun konaklama önerileri sunulmaktadır.

Proje, konaklama sektöründeki büyük veri setlerinin detaylı bir analizini içerir. Airbnb platformunda bulunan geniş veri setleri temizlenir, birleştirilir ve çeşitli analizlerle konaklama yerlerinin lokasyonları, fiyatları, türleri ve müşteri değerlendirmeleri incelenir. Bu analizler, konaklama türlerinin coğrafi dağılımından fiyat trendlerine, popülerlikten ulaşım hatlarına kadar birçok faktörü ele almaktadır.

Anahtar Sözcükler: Konaklama, Veri Analitiği, Öneri Sistemi, Metin Madenciliği, Coğrafi Veri İşleme

Accommodation Recommendation System: Data Analytics and Personalized Recommendations

Abstract

The travel and accommodation industry has undergone a significant transformation with the rapid rise of digital platforms. AirBnb, one of the pioneers of this transformation in this sector, offers travelers a variety of accommodation experiences around the world.

The main aim of the project is to develop a recommendation system that provides personalized accommodation recommendations using the large datasets of the AirBnb platform. In this context, South Aegean Greece (Aegean Greek Islands) datasets, one of the current datasets based on 'Country/City', were analyzed, accommodation types, prices, location-based features were examined and it was aimed to make recommendations suitable for users' travel purposes and budgets. In the developed recommendation system, based on an approach that includes text mining and similarity analysis, TF-IDF vectorization and Cosine Similarity calculations are used to determine the most suitable accommodation places for users' term entries. Thus, accommodation recommendations that best suit users' personal preferences and budgets are provided.

The project includes a detailed analysis of large data sets in the hospitality industry. The large datasets available on the AirBnb platform are cleaned, merged and analyzed through various analyses to examine the locations, prices, types and customer reviews of accommodations. These analyses range from the geographical distribution of accommodation types to price trends, popularity and accessibility.

Keywords: Accommodation, Data Analytics, Recommendation System, Text Mining, Geographic Data Processing

*Sevgili eřim Hasan AKMAK ve kızlarımız Elif DEMİRCİN ile Hazal
DEMİRCİN'e...*

Teşekkür

Projeme verdiği destek ve olumlu yaklaşımları için proje danışmanım *Dr. Öğr. Üyesi Osman GÖKALP*'e; bütün hedeflerimde yanımda olan ve beni her zaman sonuna kadar beni destekleyen sevgili eşim *Mak.Müh. Hasan ÇAKMAK*'a; beni bu alanla tanıştıran ve başarılı olabileceğime inandıran değerli arkadaşım *EE. Müh. Hakan SELEK*'e; ve *tüm dönem arkadaşlarıma* ayrı ayrı teşekkür ederim.

İçindekiler

Öz	i
Abstract	ii
Teşekkür	iv
Şekiller Listesi.....	vii
Tablolar Listesi.....	ix
1 Giriş	1
2 İlgili Çalışmalar	2
3 Veri İncelemesi	5
3.1 Kullanılan Veriler	5
3.1.1 “listings.csv” Veri Seti	5
3.1.2 “neighbourhoods.geojson” Veri Seti	9
3.2 Verilerin İncelenmesi.....	12
3.2.1 Type (Veri Tipi) İncelemesi	12
3.2.2 Satır ve Sütun Sayıları İncelemesi.....	12
3.2.3 Columns (Sütun İsimleri) İncelemesi	12
3.2.4 Discribe (Temel İstatistiksel Bilgiler) İncelemesi	13
3.2.5 İfo (-Çeşitli- Bilgi) İncelemesi	14
3.2.6 Numerik (Sayısal) Veriler Arasındaki Korelasyon İncelemesi	15
3.2.7 'last_review' Tarih Verisi İncelemesi.....	16
3.3 Verilerin Hazırlanması	18
3.3.1 Verilerin Temizlenmesi	18
3.3.1.1 NaN/None (Boş) Değerlerin İncelenmesi	18

3.3.1.2	Boş Sütunların Silinmesi	19
3.3.1.3	Kullanılmayacak Sütunların Silinmesi.....	20
3.3.2	Kullanılacak Sütunlarda Eksik Değerlerin Doldurulması	20
3.4	Verilerin Birleştirilmesi	23
3.4.1	Ortak Sütunların Bulunması	23
3.4.2	Ortak Sütun Hazırlanması	23
3.4.3	Birleştirme	24
4	Veri Analizi ve Görselleştirilmesi.....	25
4.1	Konakların Harita Üzerindeki Gösterimi.....	25
4.1.1	CRS ve WGS 84 Sistemi.....	25
4.1.2	Anahtar Kelimelere Göre Konak Belirleme	26
4.2	Veriler Arasında İlişki Grafikleri.....	28
4.2.1	Lokasyon/Ada Seçiminde Kullanılabilecek Bazı Analizler	28
4.2.1.1	Adalar Bazında Günlük ve Ortalama Fiyatlar	28
4.2.1.2	Adalar Bazında Sunulan Toplam Konak Sayıları	30
4.2.2	Konak Türü seçiminde Kullanılabilecek Bazı Analizler	31
4.2.2.1	Konaklama Türlerinin Yaygınlığı	31
4.2.2.2	Konaklama Türlerinin Ortalama Fiyatları	32
4.2.2.3	En Çok Yorum Alan Konaklar	33
5	Projenin Geliştirilmesi	35
5.1	Konakların Harita Üzerindeki Gösterimi.....	35
5.2	TF-IDF Matrisinin Oluşturulması.....	36
5.3	Cosine Similarity Hesaplama.....	38
5.4	Önerilerin Alınması ve Benzerlik Sıralama	39
5.5	Sonuçların Görüntülenmesi	41
6	Sonuç.....	42
	Kaynaklar	43

Şekiller Listesi

Şekil 3.1	neighbourhood.geojson (ikinci) veri seti.....	19
Şekil 3.2	listing.csv (birinci) veri seti.....	20
Şekil 3.3	Veri setlerinde tip incelemesi.....	21
Şekil 3.4	Satır ve sütun sayıları incelemesi.....	21
Şekil 3.5	Birinci veri seti için columes incelemesi.....	21
Şekil 3.6	İkinci veri seti için columes incelemesi.....	22
Şekil 3.7	Ortak columes incelemesi.....	22
Şekil 3.8	Birinci veri seti için describe incelemesi.....	22
Şekil 3.9	İkinci veri seti için describe incelemesi.....	23
Şekil 3.10	Birinci veri seti için info incelemesi.....	23
Şekil 3.11	İkinci veri seti için info incelemesi.....	23
Şekil 3.12	Sayısal sütunlar arasındaki korelasyon ilişkisi.....	24
Şekil 3.13	Sayısal sütunlar arasındaki korelasyon matrisi.....	25
Şekil 3.14	Son yorum tarihlerini içeren liste.....	25
Şekil 3.15	En eski ve en yeni son yorum tarihlerini içeren liste.....	26
Şekil 3.16	Son yorumların aylara göre dağılımı.....	26
Şekil 3.17	NaN/Boş değerler.....	27
Şekil 3.18	Birinci veri setinde yer alan boş sütunlar.....	28
Şekil 3.19	İkinci veri setinde yer alan boş sütunlar.....	28
Şekil 3.20	Projede kullanılmayacak sütunlar.....	29
Şekil 3.21	Tüm sütunlardaki eksik değerlerin sayısı.....	29
Şekil 3.22	Price sütunundaki eksik değer sayısı.....	30
Şekil 3.23	Ada bazında fiyat ortalamaları listesi.....	31
Şekil 3.24	Ada bazında fiyat ortalamaları grafiği.....	31
Şekil 3.25	Veri setleri arasındaki ortak sütunlar.....	32
Şekil 3.26	Oluşturulan ortak “geometry” sütunu.....	32
Şekil 4.1	Tüm konakların harita üzerindeki coğrafi konumları.....	34

Şekil 4.2	“breakfast” ve “5.0” anahtar kelimelerini içeren konakların coğrafi konumları	35
Şekil 4.3	“20 bed” anahtar kelimesini içeren konakların coğrafi konumları	36
Şekil 4.4	Adaların günlük ortalama fiyatları grafiği (sıralı)	37
Şekil 4.5	Adaların günlük ortalama fiyatları (scatter plot)	37
Şekil 4.6	Adaların günlük ortalama fiyatları (box plot)	38
Şekil 4.7	Ada bazında sunulan toplam konak sayıları	39
Şekil 4.8	Adalara göre toplam ev sayısı dağılımı.....	40
Şekil 4.9	Konaklama türleri	40
Şekil 4.10	Konaklama türlerinin yaygınlığı	41
Şekil 4.11	Konaklama türlerine göre günlük ortalama fiyatlar	41
Şekil 4.12	En çok yorum alan konaklar listesi	42
Şekil 4.13	En çok yorum alan konaklar ve yorum sayısı grafiği	42
Şekil 5.1	En çok yorum alan konakların coğrafi konumları.....	44
Şekil 5.2	TF-IDF matrisi	46
Şekil 5.3	İhtiyaca en uygun 5 konak	46
Şekil 5.4	Cosine similarity kullanımı	47
Şekil 5.5	Cosine similarity ve uzay vektör modeli.....	47
Şekil 5.6	Cosine similarity formülü.....	48
Şekil 5.7	Anahtar kelimelere göre önerilerin alınması	48
Şekil 5.8	Önerilen konakların coğrafi konumları	50

Tablolar Listesi

Tablo 3.1	Birinci veri setine ait özellikler	14
Tablo 3.2	İkinci veri setine ait özellikler	18

Bölüm 1

Giriş

Konaklama seçimi yapmak, seyahat edecek kişiler için oldukça karmaşık bir iştir. Kişilerin sosyal ve ekonomik durumlarına, arzularına, karakterlerine ve hedeflerine göre bu seçimler sürekli değişiklik göstermektedir. Özellikle hiç bilinmeyen bir ülkeye ya da şehre seyahat etmek bazen oldukça güçtür. Bu nedenle, bilinmezliğin mümkün olduğunca azaltılması gerekir.

Bu projede temel amaç, güncel konaklar arasından kişilere en uygun konağı en kısa sürede önermek; ve diğer seçeneklerin genel durumlarını yansıtarak yapılan tercihin optimal bir tercih olup olmadığının değerlendirmesini objektif olarak sunabilmektir. Bu proje, Yunan adalarından birine seyahat etmeyi planlayan fakat hangi adanın kendisi için daha uygun olacağına henüz karar verememiş, nerede konaklayacağını, yaklaşık ne kadar bütçe ayırması gerektiğini bilemeyen birisine ideal seçim konusunda rehberlik etmek hedefi üzerine kurulmuştur.

Bu hedef için, AirBnb güncel veri setleri kullanılmıştır. Projenin yapısı aynı kalmakla birlikte, veri setinin değiştirilmesi ile birlikte, proje arzu edilen her ülkede ya da şehirde konak önerisi sunmayı mümkün kılmaktadır. Ayrıca, önerilen konakları harita üzerinde göstererek konakların fiiliyatta daha kolay bulunmasını ve belli lokasyonlara uzaklıklarının daha iyi anlaşılmasını sağlamaktadır.

Projenin gelişim safhasında öncelikle, veri setlerinin incelenmesi, amaca uygun temizlenmesi, boşlukların doldurulması ve veriler arasındaki ilişkilerin analiz edilmesi gerçekleştirilmiştir. Ardından, metin madenciliği ve bilgi algoritmalarında kullanılan vektörleme ile, kullanıcıdan istek ve arzularının alındığı; bunlara en yakın sonuçları veren bir sistem oluşturulmuştur.

Bölüm 2

İlgili Çalışmalar

Öneri sistemleri, kullanıcıların bilgi arayışı süreçlerinde arzu ettikleri nesnelere temsil eden ve aynı zamanda bu nesnelere ile ilgili kullanıcı gereksinimlerini ve tercihlerini göz önünde bulundurarak öneriler sunan akıllı uygulamalar olarak tanımlanır (Aggarwal, 2016; Akca, 2024; Mahmood ve Ricci, 2009; Yücebaş, 2019). Öneri sistemleri, izleyicilerin web üzerindeki platformlar üzerinden yalnızca kimlik bilgilerini değil, duygu ve deneyimlerini de veriye dönüştürmek suretiyle beğenebilecekleri film önerileri sunmayı ve böylece müşteri memnuniyetini en üst düzeye çıkarmayı amaçlar (Felfernig vd., 2021; Gasparetti vd., 2021; Monti vd., 2021; Thomas ve John, 2021; Zhang vd., 2021).

Konaklama öneri sistemlerinin çoğu, “fiyata göre sıralama” veya “yoruma göre sıralama” ile ilişkilidir ve genellikle bu mantığa dayanmaktadır. Karaman vd. (2015), istenilen hizmetin ve nesnenin nokta atışı olarak kullanıcının karşısına çıkarılmasını, öneri sistemlerini bir nevi karar destek sistemi olarak adlandırmışlardır. Benzer odaklı çalışmalar aşağıda özetlenmiştir.

Sayar ve Turdaliev (2018) yaptıkları çalışmada, kullanıcıların memnuniyetini arttırmayı amaçlamış; bu amaçla otomatik otel rezervasyon sistemi yapmışlardır. Geliştirdikleri sistemde, halihazırda yapılmış bir rezervasyonun yeni alternatifler geldikçe sistem tarafından otomatik olarak değiştirilmesi (rezervasyon yükseltici) sağlanmıştır. Otomatik otel yükseltme sistemi iki durumda devreye girmiş; birinci durumda, aynı fiyata daha iyi otel bulunduğu sistem kullanıcıyla iletişim kurmadan otel otomatik yükseltilmiştir. İkinci durumda ise, bu işlem aynı kategorideki otel daha düşük fiyata bulunduğu yapılmıştır. Aradaki ücret farkının da kullanıcıya aide edilmesi planlanmıştır. Bir otelin diğer otelden daha iyi olma olasılığı yüzde 51-55%’i bulmadığı müddetçe, otel değiştirme ya da öneride bulunmaya gerek duyulmamıştır.

Hatta yazarlar, otomatik deęiřtirme yapmaktansa müşteriye sadece öneride bulunmanın ve nihai kararı müşterinin kendi isteęine bırakmanın müşteri memnuniyeti açısından daha olumlu olacaęı kanaatine varmışlardır.

Yalçın (2016), çoklu-ölçüt tabanlı ortak filtreleme sistemleri (OF) üzerine yoğunlaşmıştır. Zira, internet kullanıcılarının, büyük bir veri kümesi içerisinde ihtiyacına uygun ürünleri/hizmetleri seçmekte zorlanabildikleri, binlerce ürün arasından ihtiyaç duyduęu ürünleri doęru ve kolay bir şekilde tespit edebilmesinin ciddi olumsuz bir durum olduęu kanısındadır. Kullanıcı ve ürün deęerlendirmelerini ikili-verilerden oluşan vektörler gibi düşünerek, kullanıcılar/ürünler arasındaki benzerlikleri, çeřitli ikili benzerlik metrikleri yardımıyla belirleyerek ürün deęerlendirme yelpazesini sayısal olarak yapan bir sistem kurmuştur. Bu sistemde, kullanıcılara öneriler üretmek için OF algoritmalarını kullanmıştır Ghazanfar vd., (2012), Karypis, (2001), Su ve Khoshgoftaar (2009) da benzer çalışmalarda bulunmuştur..

Cingiz ve Marangoz (2021), öneri sistemi oluştururlarken model tabanlı filtreleme yaklaşımlarından olan işbirlikçi filtreleme yöntemlerini kullanmışlardır. Çalışmalarında Jester veri seti içerisinde bulunan şakaları kullanıcı tabanlı ve öge tabanlı işbirlikçi filtreleme yöntemleri ile kümelemiřlerdir. Budak ve Gümüřtař (2022) da aynı şekilde, kullanıcı-ürün skorları üzerinden, kullanıcı bazlı işbirlikçi filtreleme yöntemlerine iliřkin farklı bilgileri de kullanabilen bir yaklařım önermiştir. Türker vd. (2019), öneri sitelerinin işbirlikçi filtrelemeye dayalı sistemler ve içerik tabanlı öneri sistemleri olmak üzere iki temel kategoriden olduęunu belirterek her iki yöntemi de birleřtiren melez bir çalışma yapmışlardır.

Walek vd. (2016), otel rezervasyon sistemlerinde otel servislerinin dıřında, otel etrafındaki veya şehirdeki etkinlik, aktivite veya ilgi noktalarıyla baę kurmak gerektięini, tarihi noktaların, gezi yerlerinin ön plana çıkarılması gerektięini savunmuşlardır. Fakat bu verilerin toplanması büyük iş yükü getireceęinden modeli yalnızca iyileřtirilebilecek bir model olarak sunmuşlardır.

Ghose vd. (2010), otel sıralama sonucunu arama sorgusuna göre en uygun deęil, müşteriye göre en uygun şekilde sıralama yöntemini geliřtirmişlerdir. Çalışma sonucunda, otel sıralama sonucunun bir iş seyahati için bařka, turistik bir seyahat için

başka olduğunu kanıtlamışlardır. İş için seyahat edenlerin fiyatlara en az duyarlı; turistik seyahat edenerin ise fiyatlara en çok duyarlı olduğunu tespit etmişlerdir. Öte yandan, iş seyahati için konaklayanların otelin toplu taşımaya ya da, otoyola yakın olup olmamasını önemsemedikleri; oysaki turistler için bu ve benzeri yerlere yakın olmanın önemli olduğu sonucuna varmışlardır.

Gelemet vd. (2022), çalışmalarında TF-IDF (term frequency–inverse document frequency) algoritması ve kosinüs benzerliği ile doğal dil işleme (NLP) kullanmış, Netflix kullanıcı verileri üzerinde bir içerik öneri sistemi geliştirmişlerdir. Yaptıkları analizler ile benzerlik yöntemleri ve uygun eşleşme verilerini kullanıcılara sunup, kişisel bazda öneri yapılmasını sağlamışlardır.

Doğal dil işleme kapsamında yapılan başka TF-IDF ile yapılmış çalışmalar da mevcuttur. Ahmed (2017) ve Korkmaz (2021) TF-IDF yöntemini, haberlerin gerçekliğinin tespitinde; Martineau ve Finin (2009) duygu analizi tespitinde; Sjarif vd. (2019), istenmeyen mesajların filtrelenmesinde; Meltem ve Çamurcu (2011) web sayfalarında bulunan dokümanlara erişmede ve ekrana getirmede kullanmışlardır. Belgelerin yakınlıklarının tespit edilmesinde vektör uzayında kosinüs benzerliği de dahil edilmiştir. Bu proje, bahsi geçen TF-IDF ve kosinüs benzerliği çalışmalarına benzer bir mantık ile geliştirilmiştir.

Bölüm 3

Veri İncelemesi

3.1 Kullanılan Veriler

Bu proje için kullanılan veri setleri AirBnb Data sitesinde bulunan güncel veri setlerinden elde edilmiştir. Ülke/Şehir bazlı olarak hazırlanan bu veri setleri içerisinde tercih Ege Yunan Adaları (South Aegean Greece) veri setleri seçilmiştir. Her ne kadar yapılan bu projede Ege Yunan Adaları verileri tercih edilmiş ise de, projenin istenilen her lokasyona uygulanması mümkündür. Bunun için ilgili lokasyonun veri setlerinin indirilmesi yeterlidir.

AirBnb Data sitesinde her lokasyonun kendine ait birkaç veri seti bulunmaktadır. Proje kapsamında Ege Yunan Adalarına ait veri setlerinden aşağıda yer alan iki veri seti kullanılmıştır.

3.1.1 “listing.csv” Veri Seti

Tablo 3.1: Birinci veri setine ait özellikler

	Sütun Başlıkları
1	id
2	name
3	host_id
4	host_name
5	neighbourhood_group
6	neighbourhood
7	latitude
8	longitude
9	room_type
10	price
11	minimum_nights
12	number_of_reviews
13	last_review

14	reviews_per_month
15	calculated_host_listings_count
16	availability_365
17	number_of_reviews_ltm
18	license

Birinci veri setinde yer alan sütun başlıklarının anlamı, AirBnB'nin kendi internet sitesinde yer alan "Inside Airbnb Data Dictionary" dokümanında belirtilmiştir. Buna göre;

- id : "AirBnb's unique identifier for the listing"

AirBnb'nin her bir ilan için atamış olduğu benzersiz kayıt numarasıdır.

- name : "Name of the listing"

İlanın adıdır ve içerisinde ilanın içeriği, bulunduğu yer, yıldız skoru, yatak ve oda sayıları gibi bilgiler yer almaktadır.

- host_id : "Airbnb's unique identifier for the host/user"

AirBnb'nin ilan veren/ev sahibi/kullanıcı için atamış olduğu benzersiz kayıt numarasıdır.

- host_name : "Name of the host. Usually just the first name(s)"

İlan veren/ev sahibi/kullanıcıların adıdır. Genellikle sadece adları içermekte, soyadları içermemektedir.

- neighbourhood :

Her ne kadar birebir tercümesi "komşu" olarak anlaşılrsa da; aslında ilanın hangi ada içerisinde bulunduğunu ifade etmektedir.

- neighbourhood_group_cleansed : "The neighbourhood group as geocoded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles"

Açık veya genel dijital şekil dosyalarıyla tanımlanan bölgelere göre enlem ve boylam kullanılarak coğrafi olarak kodlanan bölge grubu şeklinde tanımlanmasına rağmen,

tüm sütun içeriği NaN'dır. Bu nedense, projenin ilerleyen aşamalarında bu sütun silinmiştir.

- latitude : "Uses the World Geodetic System (WGS84) projection for latitude and longitude"

Enlem değeridir. Dünya Jeodezik Sistemi (WGS84) projeksiyonu baz alınarak ilanın bulunduğu enlem yazılmıştır.

- longitude : "Uses the World Geodetic System (WGS84) projection for latitude and longitude"

Boylam değeridir. Dünya Jeodezik Sistemi (WGS84) projeksiyonu baz alınarak ilanın bulunduğu boylam yazılmıştır.

- room_type : "[Entire home/apt|Private room|Shared room|Hotel]. All homes are grouped into the following three room types: Entire place: Entire places are best if you're seeking a home away from home. With an entire place, you'll have the whole space to yourself. This usually includes a bedroom, a bathroom, a kitchen, and a separate, dedicated entrance. Hosts should note in the description if they'll be on the property or not (ex: ""Host occupies first floor of the home""), and provide further details on the listing. Private rooms: Private rooms are great for when you prefer a little privacy, and still value a local connection. When you book a private room, you'll have your own private room for sleeping and may share some spaces with others. You might need to walk through indoor spaces that another host or guest may occupy to get to your room. Shared rooms: Shared rooms are for when you don't mind sharing a space with others. When you book a shared room, you'll be sleeping in a space that is shared with others and share the entire space with other people. Shared rooms are popular among flexible travelers looking for new friends and budget-friendly stays."

Veri seti içerisinde dört tip konaklama türü mevcuttur. Tüm Ev/Daire: Genellikle bir yatak odası, bir banyo, bir mutfak ve ayrı, özel bir girişi olan ev ya da dairelerdir. Bütün alan ziyaretçilere ait olacaktır. İlanı veren ev sahipleri mülkte bulunup bulunmayacaklarını açıklamada belirtmelidir. Örneğin, ev sahibi evin birinci katında

oturuyor ve ziyaretçileri bizzat karşılamayı tercih ediyor ise bunu ilanda belirtmelidir. Özel Oda: Uyumak için özel bir odanın rezerve edildiği fakat bazı alanların (örneğin; banyo, mutfak gibi) başkalarıyla paylaşılabilirdiği yerlerdir. Odaya ulaşmak için başka misafirin kullanabileceği kapalı alanlardan geçilmesi gerekebilir. Otel Odası: Genellikle konforlu yataklar, banyo, televizyon, telefon, klima, mini buzdolabı, güvenlik kasası gibi temel olanaklara sahip odalardır. Paylaşımlı Oda: Kiralanan alan tamamen başkalarıyla paylaşılmaktadır. Yeni arkadaşlar arayan ve bütçe dostu konaklama arayan esnek gezginler arasında popüler bir seçenektir.

- price : “daily price in local currency”

Yerel para birimi cinsinden günlük fiyattır. Bu proje için para birimi “Euro”dur.

- minimum_nights : “minimum number of night stay for the listing (calendar rules may be different)”

İlandaki konağı rezerve edebilmek için gereken minimum gecesi sayısıdır.

- availability365 : “availability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host”

İlandaki konağın rezerve edilebileceği maksimum gün sayısıdır. Elbette ki, diğer misafirlerin rezervasyonu bu süreyi kısıtlamaktadır.

- number_of_reviews : “The number of reviews the listing has”

İlandaki konağın aldığı tüm yorumların sayısıdır.

- number_of_reviews_1tm : “The number of reviews the listing has (in the last 12 months)”

İlandaki konağın sadece son 12 ay içinde aldığı yorumların sayısıdır.

- last_review : “The date of the last/newest review”

İlandaki konağın aldığı en son/en yeni yorumun tarihidir.

- license : “The licence/permit/registration number”

Lisans/izin/kayıt numarasıdır.

- calculated_host_listings_count : “The number of listings the host has in the current scrape, in the city/region geography”

İlan veren/ev sahibi/kullanıcıların o şehir/bölge coğrafyasındaki tüm konaklarının (ilana verilen tüm konaklarının) sayısıdır.

- reviews_per_month : “The number of reviews the listing has over the lifetime of the listing”

İlandaki konağın ilan süresince sahip olduğu yorumların aylık bazdaki sayısıdır.

3.1.2 “neighbourhood.geojson” Veri Seti

Tablo 3.2: İkinci veri setine ait özellikler

	Sütun Başlıkları
1	id
2	name
3	host_id

Bu ikinci veri setinde yalnızca üç sütun başlığı bulunmaktadır. Buna göre;

- neighbourhood :

Her ne kadar birebir tercümesi “komşu” olarak anlaşılrsa da; aslında ilanın hangi ada içerisinde bulunduğunu ifade etmektedir.

- neighbourhood_group :

Açık veya genel dijital şekil dosyalarıyla tanımlanan bölgelere göre enlem ve boylam kullanılarak coğrafi olarak kodlanan bölge grubu şeklinde tanımlanmasına rağmen, tüm sütun içeriği NaN’dır. Bu nedenle, projenin ilerleyen aşamalarında bu sütun silinmiştir.

➤ geometry:

İlanın bulunduğu enlem ve boylam değerleridir.

	neighbourhood	neighbourhood_group	geometry
0	Αγαθονησίου (Agathonisi)	None	MULTIPOLYGON (((26.94942 37.48437, 26.94943 37...
1	Αμοργού (Amorgos)	None	MULTIPOLYGON (((25.89415 36.88140, 25.89415 36...
2	Ανάφης (Anafi)	None	MULTIPOLYGON (((25.88678 36.26739, 25.88671 36...
3	Άνδρου (Andros)	None	MULTIPOLYGON (((24.95917 37.69039, 24.95897 37...
4	Αντιπάρου (Antiparos)	None	MULTIPOLYGON (((25.01459 36.98395, 25.01460 36...
5	Αστυπαλαίας (Astypalea)	None	MULTIPOLYGON (((26.22293 36.55997, 26.22292 36...
6	Θήρας (Santorini)	None	MULTIPOLYGON (((25.22866 36.22052, 25.22851 36...
7	Ιητών (Ios)	None	MULTIPOLYGON (((25.38813 36.65075, 25.38815 36...
8	Καλυμνίων (Kalymnos)	None	MULTIPOLYGON (((26.93429 36.93263, 26.93439 36...
9	Καρπάθου (Karpathos)	None	MULTIPOLYGON (((27.10520 35.42192, 27.10505 35...
10	Κάσου (Kasos)	None	MULTIPOLYGON (((26.88592 35.44574, 26.88576 35...
11	Κιμώλου (Kimolos)	None	MULTIPOLYGON (((24.64532 36.74151, 24.64509 36...
12	Κύθνου (Kythnos)	None	MULTIPOLYGON (((24.41172 37.31715, 24.41151 37...
13	Κω (Kos)	None	MULTIPOLYGON (((26.98787 36.75052, 26.98799 36...
14	Λειψίων (Lipsi)	None	MULTIPOLYGON (((26.72729 37.33687, 26.72729 37...
15	Λέρου (Leros)	None	MULTIPOLYGON (((26.81229 37.14553, 26.81222 37...
16	Μεγίστης (Kastellorizo)	None	MULTIPOLYGON (((29.59218 36.15328, 29.59219 36...
17	Μήλου (Milos)	None	MULTIPOLYGON (((24.12955 36.51814, 24.12942 36...
18	Μυκόνου (Mykonos)	None	MULTIPOLYGON (((25.57824 37.42351, 25.57833 37...
19	Νάξου & Μικρών Κυκλάδων (Naxos & Small Cyclades)	None	MULTIPOLYGON (((25.39513 36.82821, 25.39533 36...
20	Νισύρου (Nisyros)	None	MULTIPOLYGON (((27.18046 36.68282, 27.18050 36...
21	Πάρου (Paros)	None	MULTIPOLYGON (((25.12375 36.97042, 25.12382 36...
22	Πάτμου (Patmos)	None	MULTIPOLYGON (((26.74932 37.36511, 26.74930 37...
23	Ρόδου (Rhodes)	None	MULTIPOLYGON (((27.70291 35.93652, 27.70310 35...
24	Σερίφου (Seriphos)	None	MULTIPOLYGON (((24.52833 37.18381, 24.52833 37...
25	Σικίνου (Sikinos)	None	MULTIPOLYGON (((25.03033 36.63400, 25.03032 36...
26	Σίφνου (Sifnos)	None	MULTIPOLYGON (((24.63827 37.02483, 24.63823 37...
27	Σύμης (Symi)	None	MULTIPOLYGON (((27.75171 36.57131, 27.75167 36...
28	Σύρου - Ερμούπολης (Syros)	None	MULTIPOLYGON (((25.05938 37.36631, 25.05954 37...
29	Τήλου (Tilos)	None	MULTIPOLYGON (((27.37987 36.38239, 27.37976 36...
30	Τήνου (Tinos)	None	MULTIPOLYGON (((25.02924 37.61517, 25.02937 37...
31	Φοληγάνδρου (Folegandros)	None	MULTIPOLYGON (((24.98686 36.61302, 24.98699 36...
32	Χάλκης (Halki)	None	MULTIPOLYGON (((27.66654 36.25762, 27.66662 36...
33	Κέας (Kea)	None	MULTIPOLYGON (((24.14582 37.74900, 24.14582 37...

Şekil 3.1: neighbourhood.geojson (ikinci) veri seti

id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month	calculated_host_listings_count	availability_365	number_of_reviews_ltm	license
0	Windmill in Imerigli - #5.0 - 2 bedrooms - 3...	50838.0	Nikos	NaN	Επίπεδο (Santorini)	36.452810	25.432060	Entire home/apt	280.0	2.0	8.0	2022-05-21	0.08	3.0	235.0	0.0	116709100185701
1	Windmill in Imerigli - #5.0 - 2 bedrooms - 3...	50838.0	Nikos	NaN	Επίπεδο (Santorini)	36.452640	25.432250	Entire home/apt	280.0	2.0	11.0	2023-10-29	0.10	3.0	249.0	1.0	116709100185701
2	Home in Káa - 4 bedrooms - 5 beds - 4 baths	74288.0	Christina	NaN	Κάλας (Káa)	37.580502	24.274979	Entire home/apt	575.0	5.0	0.0	NaN	NaN	1.0	0.0	0.0	00000337602
3	Home in Sífnos - #4.81 - 2 bedrooms - 2 beds - ...	2219407.0	Alexandra	NaN	Σίφνος (Sífnos)	36.972540	24.731700	Entire home/apt	90.0	4.0	37.0	2023-09-06	0.29	1.0	332.0	2.0	00001015017
4	Townhouse in Rhodes - #4.80 - 1 bedroom - 1 be...	135660.0	Claudia	NaN	Ρόδου (Rhodes)	36.441110	28.230600	Entire home/apt	80.0	3.0	88.0	2023-09-19	0.59	19.0	269.0	7.0	00000939953
5	Rental unit in Ios - #4.83 - 1 bedroom - 1 bed...	659287.0	María	NaN	Ιππύσι (Ios)	36.786970	25.294990	Entire home/apt	90.0	4.0	67.0	2016-10-05	0.53	2.0	363.0	0.0	;
6	Home in Ios - #5.0 - 2 bedrooms - 4 beds - 2.5...	659287.0	María	NaN	Ιππύσι (Ios)	36.788060	25.296780	Entire home/apt	210.0	3.0	3.0	2015-10-04	0.02	2.0	364.0	0.0	;
7	Farm stay in Karpathos - #4.80 - 1 bedroom - 1...	133185.0	Marianna	NaN	Καρπάθου (Karpathos)	35.481640	27.167870	Entire home/apt	40.0	2.0	42.0	2023-09-21	0.32	1.0	322.0	5.0	000000493446
8	Cycladic house in Paros - #4.70 - 3 bedrooms - ...	2248079.0	Elina	NaN	Πάρος (Paros)	37.010970	25.246510	Entire home/apt	250.0	4.0	30.0	2023-06-18	0.26	1.0	385.0	3.0	00000117471
9	Apartment in Santorini - #4.72 - 1 bedroom - ...	408247.0	Vassiliki	NaN	Επίπεδο (Santorini)	36.371010	25.480540	Entire home/apt	49.0	1.0	81.0	2023-06-28	0.64	19.0	385.0	4.0	1170100
10	Apartment in Santorini - #4.74 - 1 bedroom - ...	408247.0	Vassiliki	NaN	Επίπεδο (Santorini)	36.371010	25.480540	Entire home/apt	49.0	1.0	141.0	2023-06-28	1.13	19.0	385.0	5.0	1170100
11	Apartment in Santorini - #4.81 - 1 bedroom - ...	408247.0	Vassiliki	NaN	Επίπεδο (Santorini)	36.371040	25.480500	Entire home/apt	49.0	1.0	149.0	2023-11-06	1.17	19.0	263.0	14.0	1170100
12	Cycladic house in Mandraki - #4.83 - 3 bedrooms - ...	1418201.0	D & Y Boutique Athens	NaN	Νεάπολη (Neyros)	36.609620	27.131630	Entire home/apt	NaN	3.0	30.0	2023-06-19	0.26	1.0	0.0	1.0	Empty
13	Townhouse in Káos - #4.20 - 1 bedroom - 4 bed...	243468.0	Eleni	NaN	Κάου (Kaos)	35.419010	26.916670	Entire home/apt	40.0	3.0	6.0	2016-06-10	0.05	1.0	0.0	0.0	;
14	Villa in Oia - #4.74 - 2 bedrooms - 7 beds - 2...	3867239.0	María	NaN	Επίπεδο (Santorini)	36.460650	25.374630	Entire home/apt	300.0	1.0	264.0	2023-10-25	2.08	37.0	0.0	32.0	114461231062400
15	Rental unit in Kolíbia - #4.88 - 2 bedrooms - ...	331438.0	Margara	NaN	Ρόδου (Rhodes)	36.457080	28.146170	Entire home/apt	95.0	1.0	10.0	2023-07-06	0.07	10.0	136.0	2.0	1476612310798400
16	Serviced apartment in Kanezados - #4.87 - 1 be...	239946204.0	Vivi	NaN	Επίπεδο (Santorini)	36.410889	25.446930	Entire home/apt	120.0	2.0	191.0	2023-06-08	1.49	4.0	164.0	2.0	1113852
17	Condo in Kanezados - #4.74 - 1 bedroom - 2 be...	4091774.0	Helen	NaN	Επίπεδο (Santorini)	36.421200	25.469880	Private room	90.0	1.0	23.0	2023-07-30	0.20	6.0	167.0	1.0	1185528
18	Boutique hotel in Kanezados - #4.82 - 1 bedroo...	239946204.0	Vivi	NaN	Επίπεδο (Santorini)	36.410889	25.446930	Private room	150.0	1.0	130.0	2023-09-19	1.11	4.0	173.0	6.0	1185528
19	Bed and breakfast in Kanezados - #4.80 - 1 be...	239946204.0	Vivi	NaN	Επίπεδο (Santorini)	36.410889	25.446930	Private room	130.0	1.0	259.0	2023-10-21	2.02	4.0	175.0	8.0	1142719
20	Condo in Kanezados - #4.81 - 1 bedroom - 1 be...	4091774.0	Helen	NaN	Επίπεδο (Santorini)	36.420170	25.469320	Entire home/apt	80.0	1.0	148.0	2023-09-27	1.17	6.0	181.0	3.0	1185528

Şekil 3.2: listing.csv (birinci) veri seti

3.2 Verilerin İncelenmesi

3.2.1 Type (Veri Tipi) İncelemesi

1. Type (Veri Tipi) İncelemesi

```
type(airbnb_yunan_adalari)
pandas.core.frame.DataFrame

type(airbnb_yunan_adalari_geo)
geopandas.geodataframe.GeoDataFrame
```

Şekil 3.3: Veri setlerinde tip incelemesi

3.2.2 Satır ve Sütun Sayıları İncelemesi

2. Satır ve Sütun Sayıları İncelemesi

```
airbnb_yunan_adalari.shape
(32698, 18)

airbnb_yunan_adalari_geo.shape
(34, 3)
```

Şekil 3.4: Satır ve sütun sayıları incelemesi

3.2.3 Columns (Sütun İsimleri) İncelemesi

3. Columns (Sütun İsimleri) İncelemesi

```
airbnb_yunan_adalari.columns

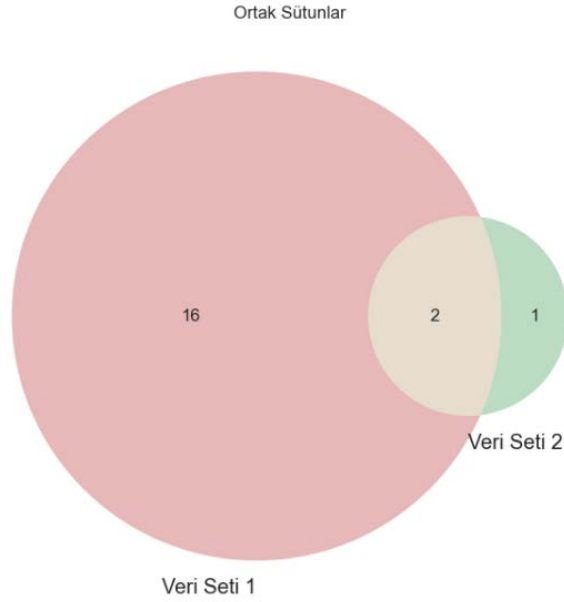
Index(['id', 'name', 'host_id', 'host_name', 'neighbourhood_group',
       'neighbourhood', 'latitude', 'longitude', 'room_type', 'price',
       'minimum_nights', 'number_of_reviews', 'last_review',
       'reviews_per_month', 'calculated_host_listings_count',
       'availability_365', 'number_of_reviews_ltm', 'license'],
      dtype='object')
```

Şekil 3.5: Birinci veri seti için columns incelemesi

```
airbnb_yunan_adalari_geo.columns
```

```
Index(['neighbourhood', 'neighbourhood_group', 'geometry'], dtype='object')
```

Şekil 3.6: İkinci veri seti için cols inceleme



```
ortak_sütunlar = airbnb_yunan_adalari.columns.intersection(airbnb_yunan_adalari_geo.columns)  
print(ortak_sütunlar)
```

```
Index(['neighbourhood_group', 'neighbourhood'], dtype='object')
```

Şekil 3.7: Ortak cols inceleme

3.2.4 Discribe (Temel İstatistiksel Bilgiler) İnceleme

```
airbnb_yunan_adalari.describe()
```

	id	host_id	neighbourhood_group	latitude	longitude	price	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_listings_count	availability_365	number_of_reviews_1tm
count	3.269800e+04	3.269800e+04	0.0	32698.000000	32698.000000	30385.000000	32698.000000	32698.000000	25343.000000	32698.000000	32698.000000	32698.000000
mean	2.945333e+17	2.178388e+08	NaN	36.890291	25.837337	880.677209	3.984097	19.193731	0.625817	13.858340	181.093339	4.840899
std	3.940625e+17	1.756000e+08	NaN	0.501118	1.182954	2116.134356	32.539036	36.910799	0.719842	43.805992	126.307974	8.624693
min	1.313100e+04	3.717700e+04	NaN	35.396107	34.265990	10.000000	1.000000	0.000000	0.010000	1.000000	0.000000	0.000000
25%	2.480244e+07	5.312050e+07	NaN	36.427258	25.165273	80.000000	1.000000	1.000000	0.170000	1.000000	67.000000	0.000000
50%	4.564548e+07	1.820767e+08	NaN	36.994530	25.377670	150.000000	2.000000	6.000000	0.380000	3.000000	186.000000	2.000000
75%	6.801509e+07	3.929350e+08	NaN	37.380160	26.538423	367.000000	3.000000	22.000000	0.820000	8.000000	282.000000	6.000000
max	1.050051e+18	5.515323e+08	NaN	37.970260	29.598050	54113.000000	1125.000000	1884.000000	23.440000	336.000000	365.000000	423.000000

Şekil 3.8: Birinci veri seti için describe inceleme

```
airbnb_yunan_adalari_geo.describe()
```

	neighbourhood	neighbourhood_group	geometry
count	34	0	34
unique	34	0	34
top	Αγαθονησίου (Agathonisi)	NaN	MULTIPOLYGON (((26.949424 37.484373, 26.949426...
freq	1	NaN	1

Şekil 3.9: İkinci veri seti için describe incelemesi

3.2.5 Info (-Çeşitli- Bilgi) İncelemesi

```
airbnb_yunan_adalari.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32698 entries, 0 to 32697
Data columns (total 18 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   id                                     32698 non-null  int64
1   name                                  32698 non-null  object
2   host_id                               32698 non-null  int64
3   host_name                             32698 non-null  object
4   neighbourhood_group                   0 non-null     float64
5   neighbourhood                          32698 non-null  object
6   latitude                              32698 non-null  float64
7   longitude                             32698 non-null  float64
8   room_type                             32698 non-null  object
9   price                                 30385 non-null  float64
10  minimum_nights                        32698 non-null  int64
11  number_of_reviews                     32698 non-null  int64
12  last_review                           25343 non-null  object
13  reviews_per_month                     25343 non-null  float64
14  calculated_host_listings_count        32698 non-null  int64
15  availability_365                       32698 non-null  int64
16  number_of_reviews_ltm                 32698 non-null  int64
17  license                                31056 non-null  object
dtypes: float64(5), int64(7), object(6)
memory usage: 4.5+ MB
```

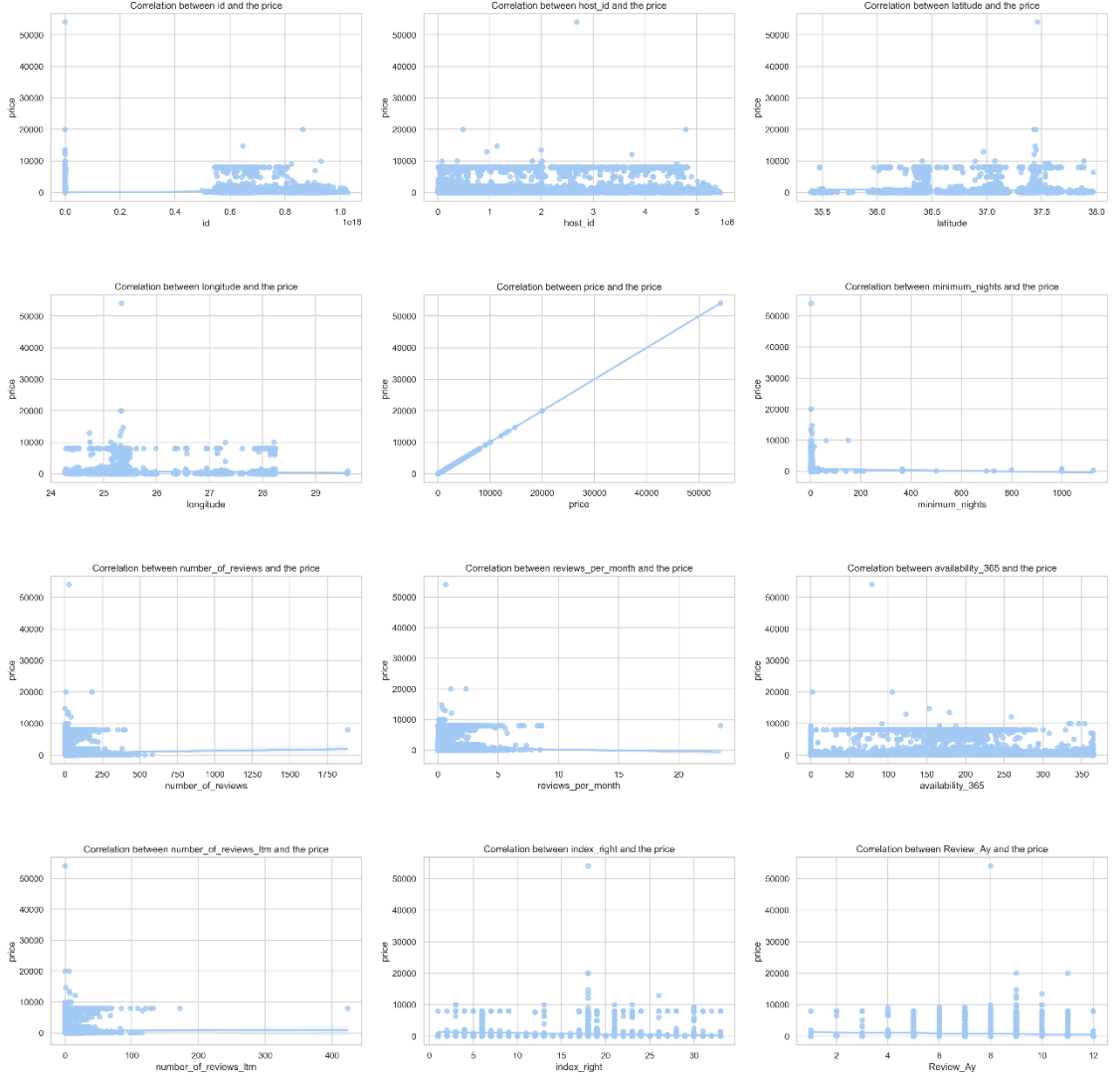
Şekil 3.10: Birinci veri seti için info incelemesi

```
airbnb_yunan_adalari_geo.info()

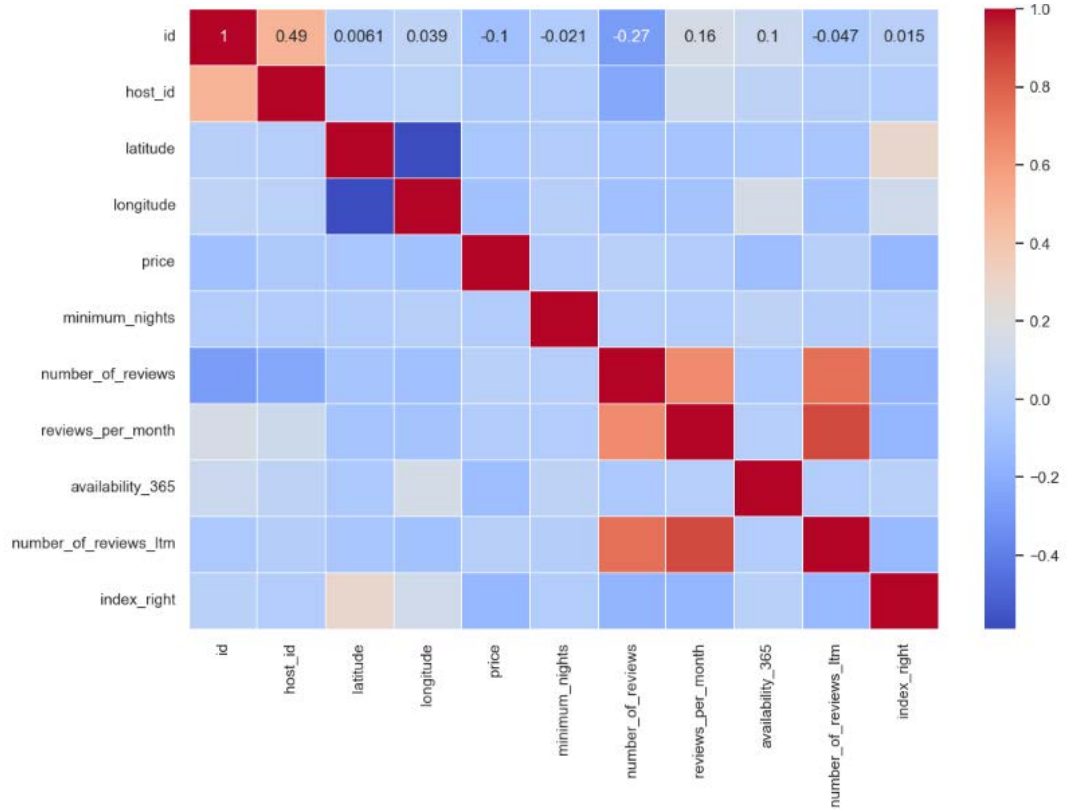
<class 'geopandas.geodataframe.GeoDataFrame'>
RangeIndex: 34 entries, 0 to 33
Data columns (total 3 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   neighbourhood                          34 non-null     object
1   neighbourhood_group                    0 non-null     object
2   geometry                               34 non-null     geometry
dtypes: geometry(1), object(2)
memory usage: 948.0+ bytes
```

Şekil 3.11: İkinci veri seti için info incelemesi

3.2.6 Numerik (Sayısal) Veriler Arasındaki Korelasyon İnceleme



Şekil 3.12: Sayısal Sütunlar Arasındaki Korelasyon İlişkisi



Şekil 3.13: Sayısal Sütunlar Arasındaki Korelasyon Matrisi

Korelasyon tablosu fiyat ile diğer özellikler arasında güçlü bir ilişki olmadığını göstermektedir.

3.2.7 'last_review' Tarih Verisi İncelemesi

```
<DatetimeArray>
['2022-05-21 00:00:00', '2023-10-29 00:00:00', '2023-09-06 00:00:00',
 '2023-09-19 00:00:00', '2018-10-05 00:00:00', '2015-10-04 00:00:00',
 '2023-09-21 00:00:00', '2023-08-18 00:00:00', '2023-06-28 00:00:00',
 '2023-11-06 00:00:00',
 ...
 '2023-04-16 00:00:00', '2022-12-29 00:00:00', '2023-03-08 00:00:00',
 '2022-11-20 00:00:00', '2023-02-13 00:00:00', '2023-03-27 00:00:00',
 '2022-11-18 00:00:00', '2023-02-28 00:00:00', '2023-03-15 00:00:00',
 '2023-12-21 00:00:00']
Length: 1268, dtype: datetime64[ns]
```

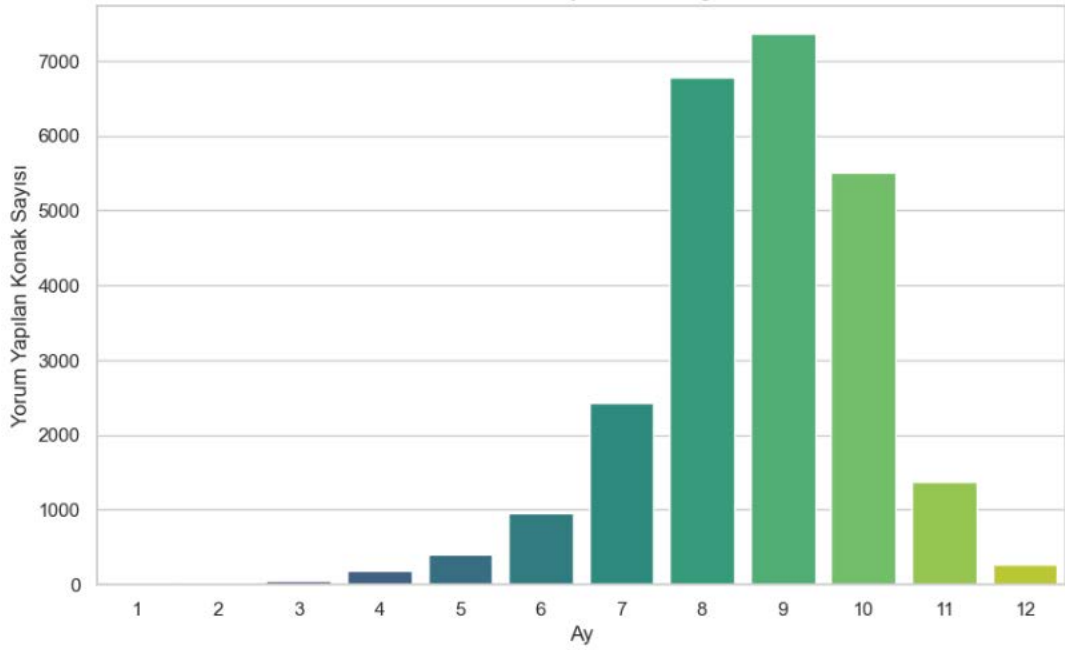
Şekil 3.14: Son yorum tarihlerini içeren liste

En eski tarih: 2012-09-28 00:00:00
En yeni tarih: 2023-12-21 00:00:00

Şekil 3.15: En eski ve en yeni son yorum tarihlerini içeren liste

Veri setinde yer alan tarih verileri, ilanlar hakkında yapılan son yorum tarihleridir. En yeni yorum tarihinin 21.12.2023 olması veri setinin güncel bir veri seti olduğunu göstermektedir.

Öte yandan, en eski son yorum tarihi 2012 yılına dayanmaktadır. Kanımca, bir konağın 2012 yılından bu yana yorum almaması konağın tercih edilme olasılığını azaltan bir durumdur.



Şekil 3.16: Son yorumların aylara göre dağılımı

Son yorum tarihi verisi, konakların hangi sezonda tercih edilmesi gerektiği hakkında da genel bir bilgi verebilir düzeydedir. Buna göre, trend zamanın yaz mevsimi olduğu anlaşılmaktadır.

3.3 Verilerin Hazırlanması

3.3.1 Verilerin Temizlenmesi

3.3.1.1 NaN/None (Boş) Değerlerin İncelenmesi

```
airbnb_yunan_adalari.isnull().sum()

id                0
name              0
host_id          0
host_name        0
neighbourhood_group    32698
neighbourhood      0
latitude          0
longitude         0
room_type         0
price            2313
minimum_nights    0
number_of_reviews  0
last_review      7355
reviews_per_month 7355
calculated_host_listings_count  0
availability_365  0
number_of_reviews_1tm  0
license          1642
dtype: int64
```

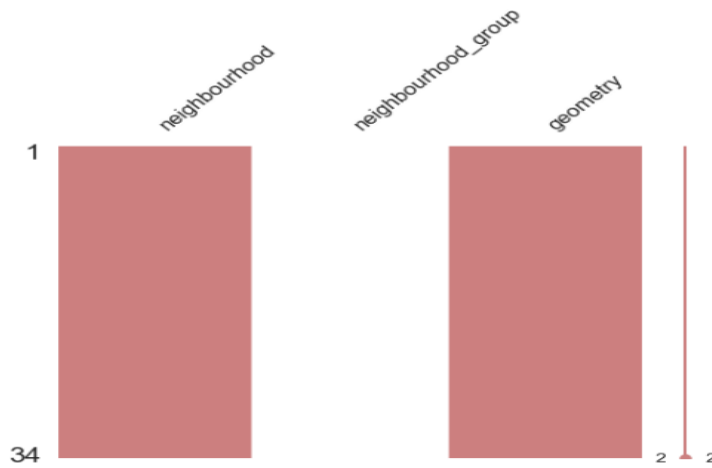
Şekil 3.17: NaN/Boş Değerler

3.3.1.2 Boş Sütunların Silinmesi



Şekil 3.18: Birinci veri setinde yer alan boş sütunlar

```
airbnb_yunan_adalari_geo.isnull().sum()\n\nneighbourhood      0\nneighbourhood_group 34\ngeometry           0\ndtype: int64\n\nsns.set(style="whitegrid")\nsns.set_palette("pastel")\n\nmsno.matrix(airbnb_yunan_adalari_geo, figsize=(6, 4), fontsize=12, color=(0.8, 0.5, 0.5))\nplt.show()
```



Şekil 3.19: İkinci veri setinde yer alan boş sütunlar

3.3.1.3 Kullanılmayacak Sütunların Silinmesi

```
selected_columns = ['host_name', 'calculated_host_listings_count', 'license']
subset_airbnb_yunan_adalari = airbnb_yunan_adalari[selected_columns]
print(subset_airbnb_yunan_adalari)
```

	host_name	calculated_host_listings_count	license
0	Nikos	3	1167K91001265701
1	Nikos	3	1167K91001265701
2	Christina	1	00000337602
3	Alexandra	1	00001615017
4	Claudia	15	00000903953
...
32693	Holidu	96	00000129884
32694	Pantelis	2	00000410152
32695	Καλλιόπη	1	00002382131
32696	Holidu	106	01044001350
32697	Holidu	68	00001184827

```
[32698 rows x 3 columns]
```

Şekil 3.20: Projede kullanılmayacak sütunlar

Listede yer alan host_name, calculated_host_listings_count ve license değerlerinin projede hiçbir katkısı yoktur. Bu nedenle veri kümesinden çıkarılmıştır.

3.3.2 Kullanılacak Sütunlarda Eksik Değerlerin Doldurulması

```
airbnb_yunan_adalari.isnull().sum()
```

id	0
name	0
host_id	0
neighbourhood	0
latitude	0
longitude	0
room_type	0
price	2313
minimum_nights	0
number_of_reviews	0
last_review	7355
reviews_per_month	7355
availability_365	0
number_of_reviews_ltm	0
dtype:	int64

Şekil 3.21: Tüm sütunlardaki eksik değerlerin sayısı

Listede yer alan last_review ile reviews_per_month değerleri 0 (sıfır) ile doldurulabilir değerlerdir.

```
airbnb_yunan_adalari['last_review'].fillna(0, inplace=True)
airbnb_yunan_adalari['reviews_per_month'].fillna(0, inplace=True)
airbnb_yunan_adalari.isnull().sum()
```

```
id                0
name              0
host_id           0
neighbourhood     0
latitude          0
longitude         0
room_type         0
price             2313
minimum_nights    0
number_of_reviews 0
last_review       0
reviews_per_month 0
availability_365  0
number_of_reviews_ltm 0
dtype: int64
```

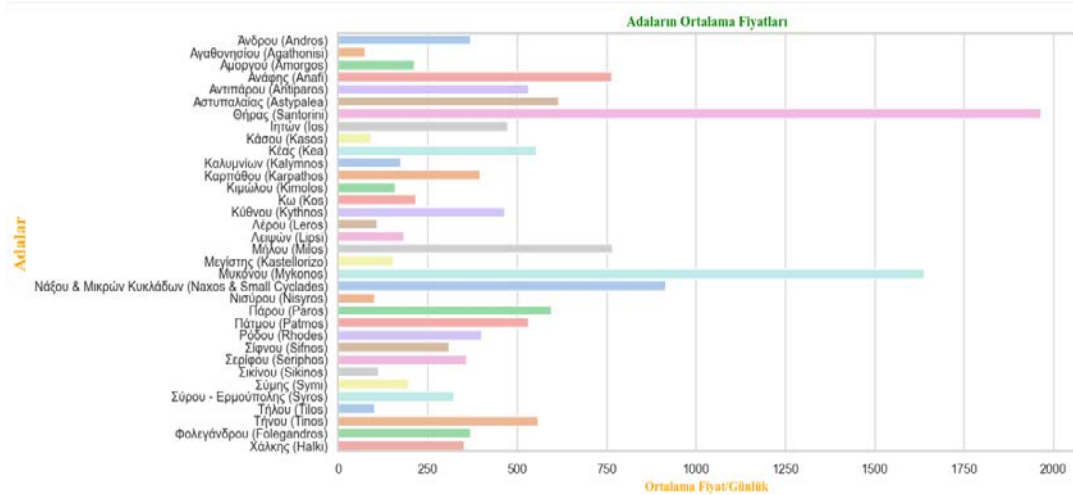
Şekil 3.22: Price sütunundaki eksik değer sayısı

Listede yer alan price değeri ise, ortalama fiyat ile doldurulması gereken bir değerdir. Price değerlerini 0 (sıfır) ile doldurmak ortalamaları çok aşağı çekeceğinden price sütunundaki boşlukların ortalama fiyat ile doldurulması daha doğrudur. Bu doğrultuda, tüm konak fiyatlarının ortalamasını almak yerine, adaları gruplandırarak her adaya ait ortalama fiyatları çıkarmak gerekir.

```
neighbourhood_gruplama = airbnb_yunan_adalari.groupby('neighbourhood')
grup_means = neighbourhood_gruplama['price'].mean()
grup_means
```

```
neighbourhood
Άνδρου (Andros)                370.650106
Αγαθονησίου (Agathonisi)      75.000000
Αμοργού (Amorgos)            211.647799
Ανάφης (Anafi)               765.195122
Αντιπάρου (Antiparos)        531.472441
Αστυπαλαίας (Astypalea)     616.149321
Θήρας (Santorini)           1964.572970
Ιητών (Ios)                  473.870523
Κάσου (Kasos)                91.285714
Κέας (Kea)                   553.134000
Καλυμνίων (Kalymnos)        175.289130
Καρπάθου (Karpathos)        396.851852
Κιμώλου (Kimos)             158.065421
Κω (Kos)                     215.950837
Κύθνου (Kythnos)            464.148867
Λέρου (Leros)                108.653631
Λειψών (Lipsi)              184.026316
Μήλου (Milos)               766.807571
Μεγίστης (Kastellorizo)     155.400000
Μυκόνου (Mykonos)           1638.533431
Νάξου & Μικρών Κυκλάδων (Naxos & Small Cyclades) 914.905384
Νισύρου (Nisyros)            100.512195
Πάρου (Paros)                596.788468
Πάτμου (Patmos)              531.425714
Ρόδου (Rhodes)               401.094517
Σίφνου (Sifnos)              308.738916
Σερίφου (Seriphos)          358.862903
Σικίνου (Sikinos)           111.457143
Σύμης (Symi)                 196.316667
Σύρου - Ερμούπολης (Syros)  322.592878
Τήλου (Tilos)                101.548387
Τήνου (Tinos)                558.760933
Φολεγάνδρου (Folegandros)   368.878049
Χάλκης (Halki)              351.061538
Name: price, dtype: float64
```

Şekil 3.23: Ada bazında fiyat ortalamaları listesi



Şekil 3.24: Ada bazında fiyat ortalamaları grafiği

3.4 Verilerin Birleştirilmesi

3.4.1 Ortak Sütunların Bulunması

```
ortak_sütunlar2 = airbnb_yunan_adalari.columns.intersection(airbnb_yunan_adalari_geo.columns)
print(ortak_sütunlar2)

Index(['neighbourhood'], dtype='object')
```

Şekil 3.25: Veri setleri arasındaki ortak sütunlar

3.4.2 Ortak Sütun Hazırlanması

İlandaki konakların bulunduğu enlem ve boylam aralıklarından, bu konakların hangi adaya tekabül ettiklerini bulmak avantaj sağlayacaktır. Bu nedenle, ortak bir ‘neighbourhood’ sütunu ortak olmasına rağmen amaca uygun ortak bir ‘geometry’ sütunu da hazırlanmalıdır. Bunun, “listing.csv” veri setinde bulunan ‘latitude’ ve ‘longitude’ sütunlarının birleşimi ile gerçekleştirilmesi mümkündür. Nitekim, geometry sütunu, coğrafi şekilleri içeren bir sütundur ve haritalandırma için oldukça önemlidir.

```
geometry
POINT (25.43206 36.45281)
POINT (25.43225 36.45264)
```

Şekil 3.26: Oluşturulan ortak “geometry” sütunu

3.4.3 Birleřtirme

Verilerin İncelenmesi bařlıklı bölümde incelendiđi üzere; “listing.csv” veri setinde yer alan veriler `pandas.core.frame.DataFrame`’dir. Bu, Pandas kütüphanesinde temel veri yapısıdır. Nitelik itibariyle iki boyutlu, deđişken uzunluklu bir veri yapısı vardır ve tablo şeklindeki verileri temsil eder. Satırlar ve sütunlar arasında etiketlenmiş verileri içerir. Proje kapsamında; veri okuma ve yazma, veri filtreleme, gruplama, sıralama amaçlarına uygun bir yapıdadır.

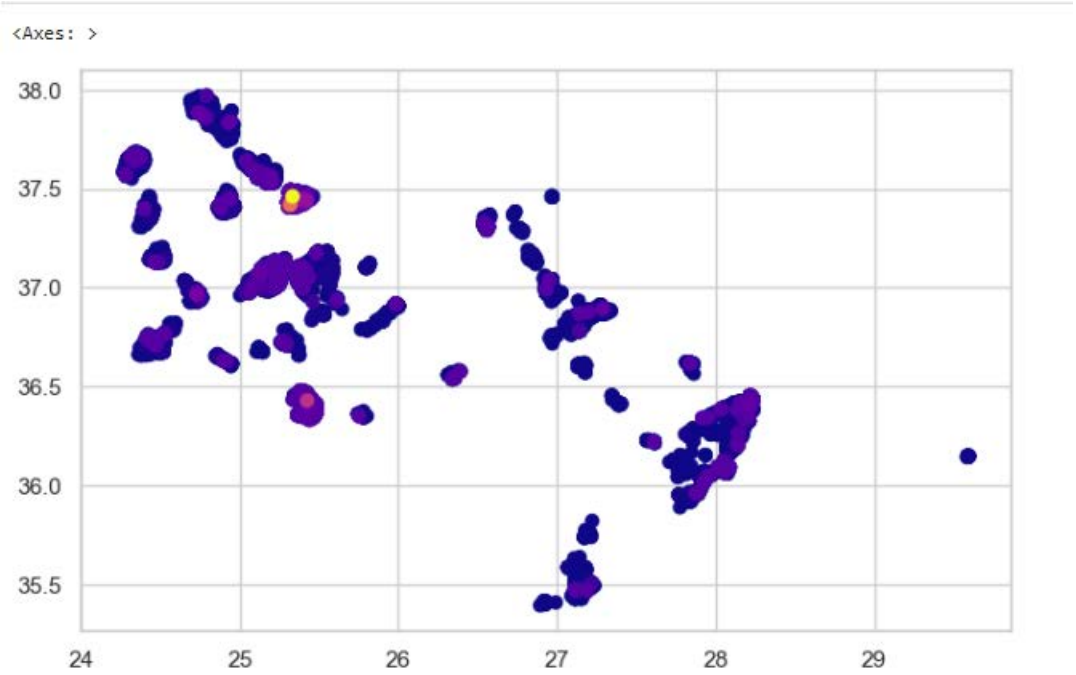
“neighbourhood.geojson” veri setinde yer alan veriler ise, `geopandas.geodataframe.GeoDataFrame`’dir. Bu, GeoPandas kütüphanesindeki özel bir `DataFrame` türüdür. Ve her iki kütüphane arasındaki temel fark, GeoPandas’ın cođrafi verilerle çalışabilme yeteneđine sahip olmasıdır. `GeoDataFrame`, cođrafi verileri temsil etmek, haritalamak ve analiz etmek için tasarlanmıştır ve `pandas DataFrame`’in genişletilmiş bir sürümüdür. `GeoDataFrame`, cođrafi şekillere sahip sütunlar içermektedir.

Bölüm 4

Veri Analizi ve Görselleştirilmesi

4.1 Konakların Harita Üzerindeki Gösterimi

4.1.1 CRS ve WGS 84 Sistemi



Şekil 4.1: Tüm konakların harita üzerindeki coğrafi konumları

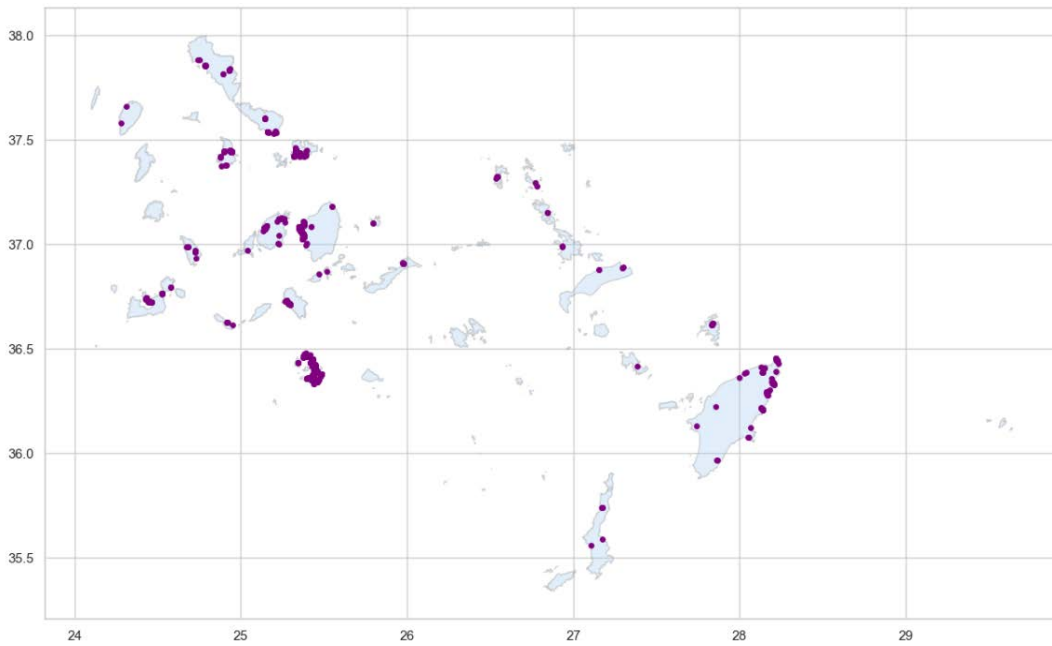
CRS (Coordinate Reference System), bir uzayda noktaların ve nesnelerin konumlarını belirtmek için kullanılan bir sistemdir. Bu sistemde bir noktanın konumu, bir referans noktasına veya referans doğrusuna göre belirlenir. Yani, bir noktanın koordinatları, belirli bir başlangıç noktasına veya doğruya olan uzaklıklarını ifade eder.

CRS'nin Projeksiyonlu CRS (Projected CRS) ve Coğrafi CRS (Geographic CRS) gibi farklı türleri vardır. Projeksiyonlu CRS, dünya yüzeyini bir düzlem üzerine yansıtmak için kullanılırken, coğrafi CRS enlem ve boylam gibi küresel koordinatları temsil eder. Bir harita veya coğrafi veri seti kullanılıyorsa, veri setinin hangi CRS'ye dayandığını bilmek önemlidir. Veri setleri arasında uyumsuzluk olmaması ve doğru analizler yapabilmek için, kullanılan CRS'leri dikkatlice yönetmek önemlidir. Eğer kullanılan veri setleri farklı CRS'ye sahipse, bu verileri aynı CRS'ye dönüştürmek gerekir.

WGS 84 (World Geodetic System 1984) ise, dünya üzerindeki coğrafi konumları belirlemek için kullanılan bir coğrafi referans sistemidir. 1984 yılında Ulusal Jeodezi ve Jeofizik Merkezi (National Geospatial-Intelligence Agency-NGA) tarafından tanıtılmıştır. Dünya yüzeyini elipsoid bir modelle temsil eder ve GPS (Global Positioning System) gibi küresel konumlandırma sistemlerinin temelini oluşturur. Koordinat birimleri genellikle enlem ve boylam cinsindedir.

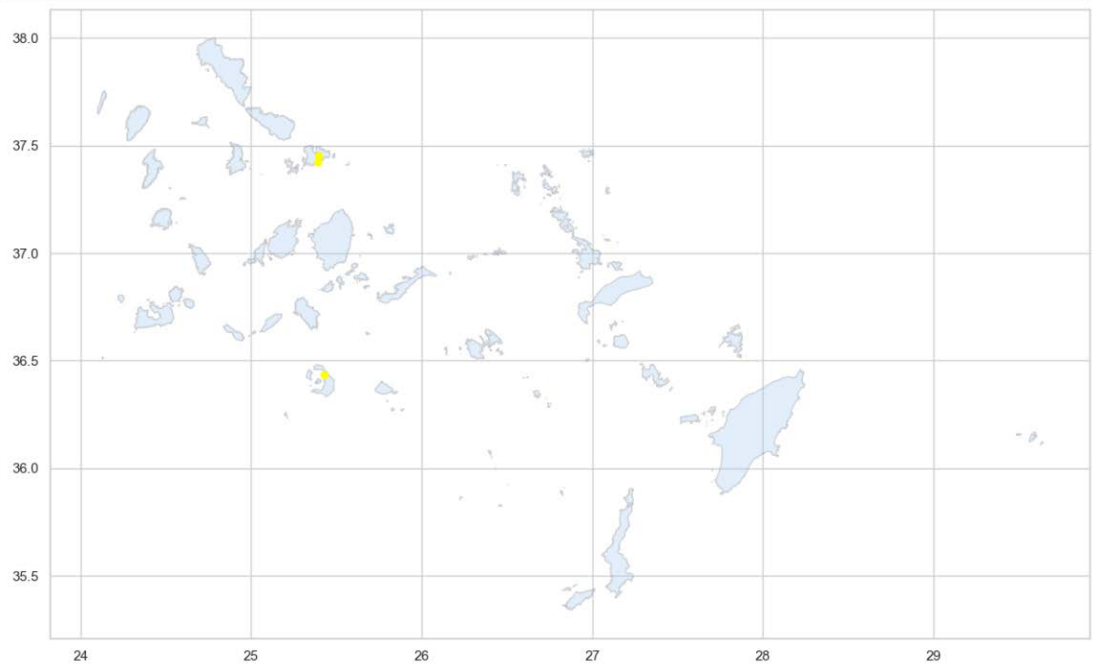
“listings.csv” veri seti ile “neighbourhood.geojson” veri setinin “geometry” sütununda birleştirilmesinin temel nedeni enlem ve boylamları verilen konaklardan bu haritayı elde edebilmektir.

4.1.2 Anahtar Kelimelere Göre Konak Belirleme



Şekil 4.2: “breakfast” ve “5.0” anahtar kelimelerini içeren konakların coğrafi konumları

Ada sınırları yukarıdaki harita üzerinde daha belirgindir. Şekil 4.2’de yer alan harita, anahtar kelime olarak girilen ‘Kahvaltı ve 5 yıldız yorum’ ifadelerini içeren ilanlara aittir. Şekil 4.3’te yer alan harita ise, anahtar kelime olarak girilen ‘20 yatak’ ifadesini içeren ilanlara aittir.



Şekil 4.3: “20 bed” anahtar kelimesini içeren konakların coğrafi konumları

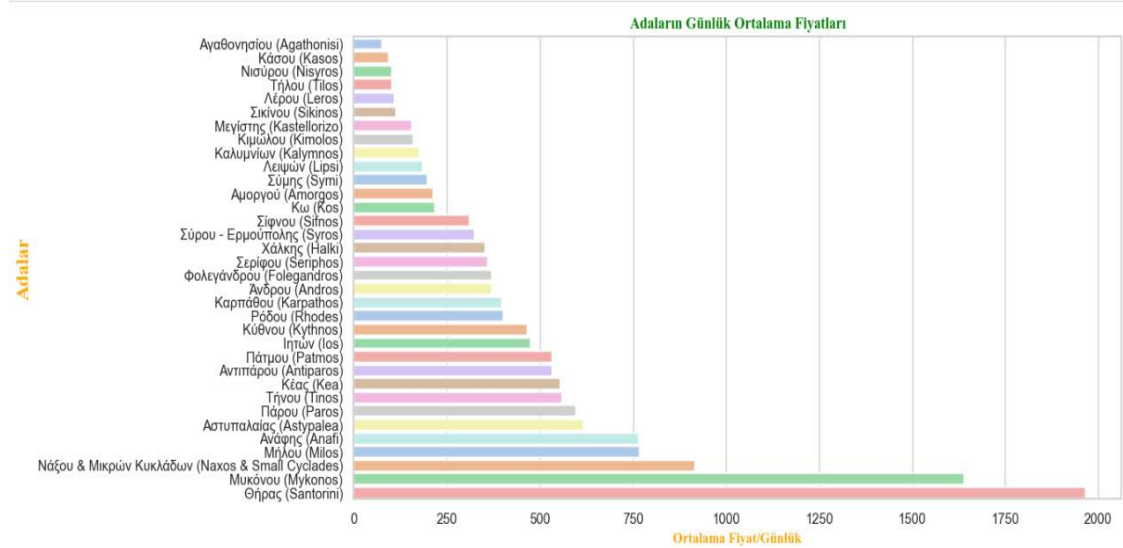
Kişilerin farklı tercih ve arzularına uygun giriş yapmak/konak filtrelemek mümkündür. Böylece, kişilerin hangi Yunan adalarına öncelik vermesi, hangi adaları seyahat planlarından çıkarması gerektiği belirginleşecektir.

4.2 Veriler Arasında İlişki Grafikleri

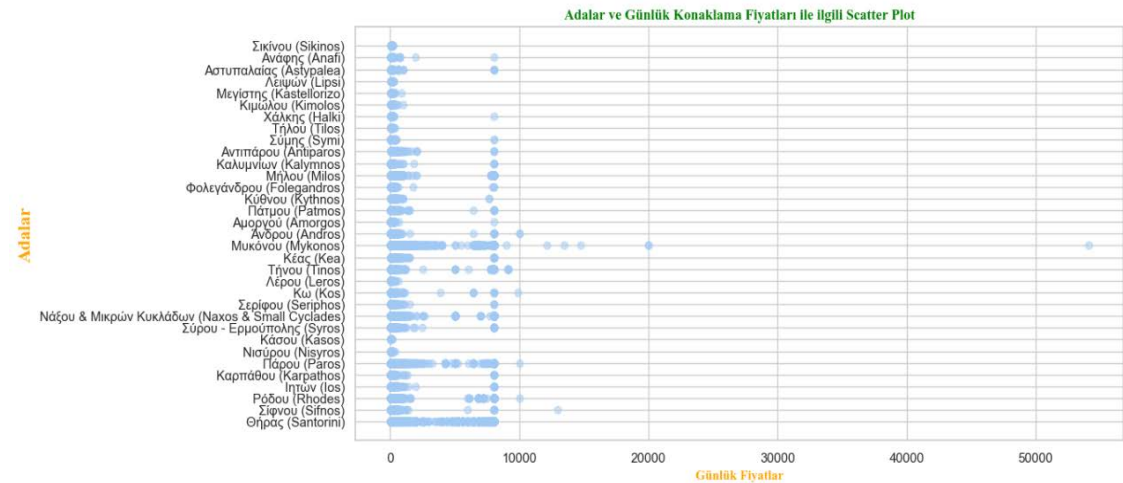
4.2.1 Lokasyon/Ada Seçiminde Kullanılabilecek Bazı Analizler

4.2.1.1 Adalar Bazında Günlük ve Ortalama Fiyatlar

Adalar bazında günlük ve ortalama fiyatlar, seyahat talebi ve trendlerini anlamak için kullanılabilir niteliktedir. Nitekim, adalar arası fiyat karşılaştırması, seyahat talebindeki değişiklikleri gösterebilir.

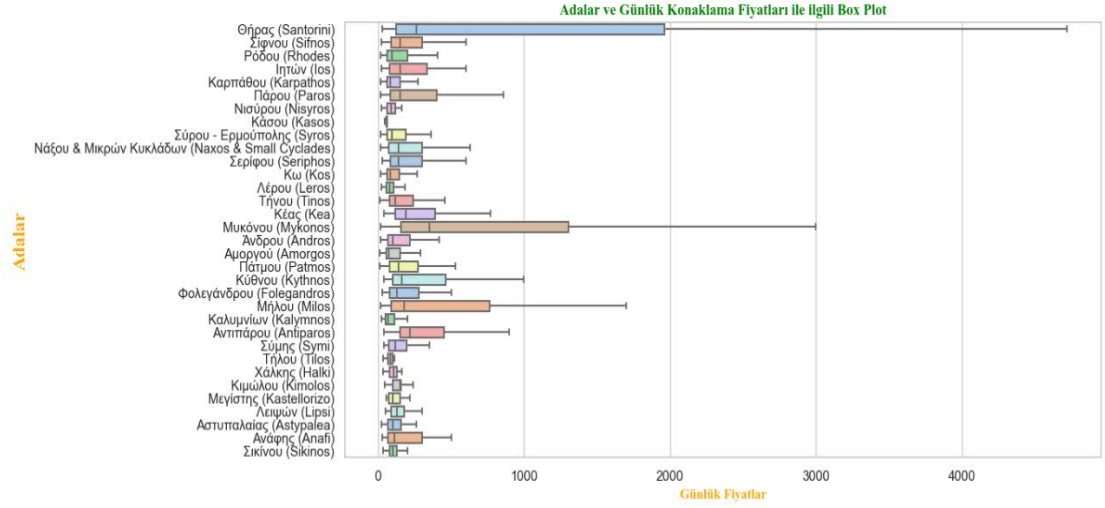


Şekil 4.4: Adaların günlük ortalama fiyatları grafiği (sıralı)



Şekil 4.5: Adaların günlük ortalama fiyatları (scatter plot)

Öte yandan, adalar bazında günlük ve ortalama fiyatlara hâkim olmak, hangi adanın bütçeye daha uygun olduğunu ve nerede konaklanacağını belirlemede de büyük rol oynayabilir.



Şekil 4.6: Adaların günlük ortalama fiyatları (box plot)

Bu kapsamda ilk önce, adalar bazında günlük ve ortalama fiyatlar grafiği çizdirilmiş, sıralanmış ve ardından adalar bazındaki ortalama fiyatların daha çok hangi ücretlerde yoğunlaştığı analiz edilmiştir.

4.2.1.2 Adalar Bazında Sunulan Toplam Konak Sayıları

	Ada	Toplam Konak Sayısı
5	Θήρας (Santorini)	4129
23	Ρόδου (Rhodes)	3792
21	Πάρου (Paros)	2781
19	Νάξου & Μικρών Κυκλάδων (Naxos & Small Cyclades)	2672
18	Μυκόνου (Mykonos)	2550
30	Τήνου (Tinos)	1195
16	Μήλου (Milos)	1064
28	Σύρου - Ερμούπολης (Syros)	871
12	Κω (Kos)	838
0	Άνδρου (Andros)	762
24	Σίφνου (Sifnos)	592
8	Κέα (Kea)	420
9	Καλυμνίων (Kalymnos)	381
10	Καρπάθου (Karpathos)	379
6	Ιητών (Ios)	371
25	Σερίφου (Serifhos)	361
13	Κύθνου (Kythnos)	317
1	Αμοργού (Amorgos)	316
22	Πάτμου (Patmos)	284
3	Αντιπάρου (Antiparos)	205
27	Σύμης (Syml)	198
4	Αστυπαλοίκας (Astypalea)	148
11	Κιμώλου (Kimos)	136
31	Φολεγάνδρου (Folegandros)	135
14	Λέρου (Leros)	125
20	Νισύρου (Nisyros)	80
32	Χάλκης (Halki)	54
2	Ανάφης (Anafi)	41
17	Μεγίστης (Kastellorizo)	41
29	Τήλου (Tilos)	39
26	Σικίνου (Sikinos)	28
15	Λειψίων (Lipsi)	25
7	Κάσου (Kasos)	13

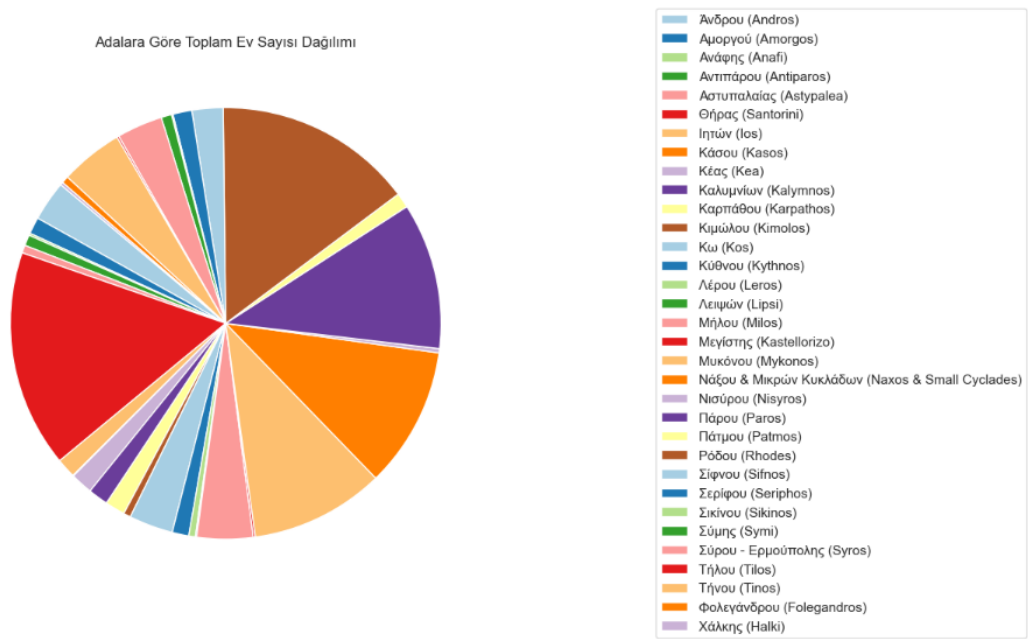
Şekil 4.7: Ada bazında sunulan toplam konak sayıları

Adalar bazında sunulan toplam konak sayıları, bir lokasyonun turistik potansiyelini değerlendirmek, ekonomik katkısını anlamak, işletmelerin planlamalarını yapmak ve turist deneyimini yönetmek için önemli bir bilgi kaynağıdır.

Adalar bazında sunulan toplam konak sayıları, adaların turizme olan yatkınlıklarının tespiti için önemli rol oynayabilir. Örneğin; adalar üzerindeki toplam konaklama

kapasitesi, turizm potansiyelini değerlendirmek için kullanılabilir. Bu, bir lokasyonun ne kadar büyük bir ziyaretçi akınına karşı hazır olduğunu gösterir. Ya da, sektördeki istihdam olanaklarının analizini yapabilmek için de kullanılabilir. Konaklama kapasitesinin fazlalığı, daha fazla iş imkanı demektir. Öte yandan, toplam konaklama kapasitesi, seçenekleri fazla olan bir yer bir keşifçi için daha cazip iken; bir doğa insanı için cazip olmayabilir.

Bu nedenle veri setinin adalar bazında sunulan toplam konak sayıları da analiz edilmiştir.



Şekil 4.8: Adalara göre toplam ev sayısı dağılımı

4.2.2 Konak Türü seçiminde Kullanılabilecek Bazı Analizler

4.2.2.1 Konaklama Türlerinin Yaygınlığı



Şekil 4.9: Konaklama türleri

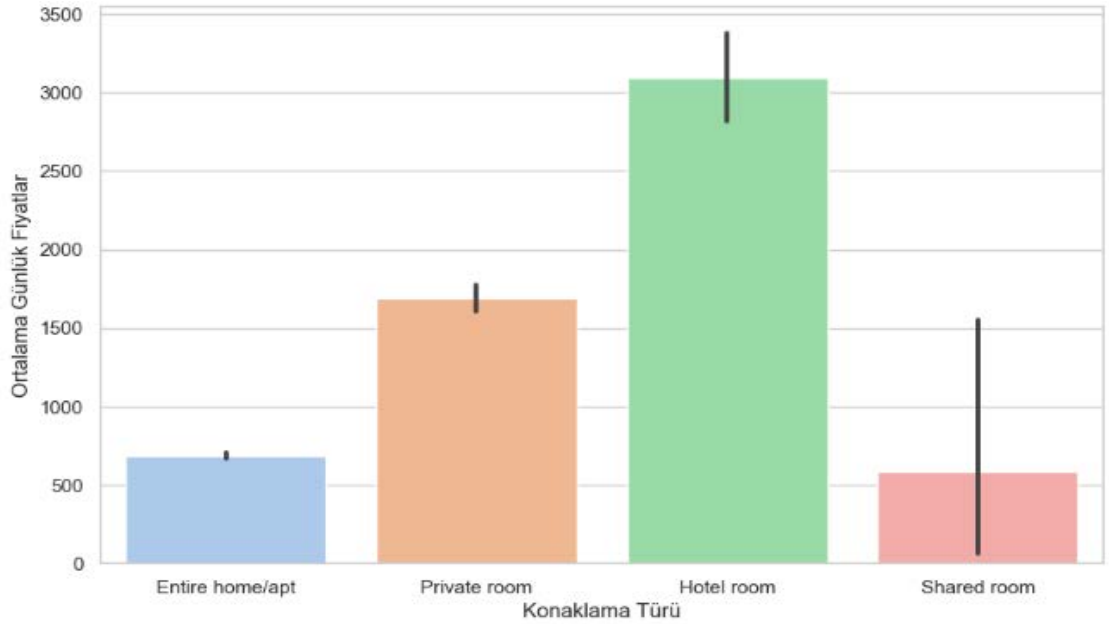
```
konaklama_türlerinin_yayginligi = birlesik['room_type'].value_counts()
konaklama_türlerinin_yayginligi

room_type
Entire home/apt    21846
Private room      2954
Hotel room         532
Shared room        11
Name: count, dtype: int64
```

Şekil 4.10: Konaklama türlerinin yaygınlığı

4.2.2.2 Konaklama Türlerinin Ortalama Fiyatları

Gayrimenkul türünün de fiyatlar üzerinde önemli bir etkisi vardır. Her mülk türü için günlük fiyatlar aşağıda incelenmiştir.



Şekil 4.11: Konaklama türlerine göre günlük ortalama fiyatlar

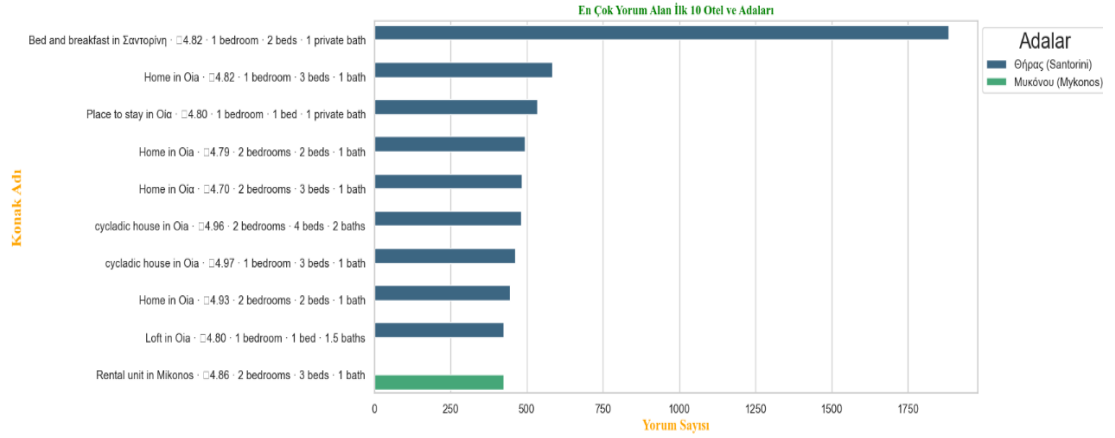
Şekil 4.11’de hangi mülk türlerinin daha pahalı olduğu görülmektedir. Otel ve özel odaların ortalama fiyatları daha yüksek; paylaşımlı odaların ortalama fiyatları ise bütçeye en uygun olanıdır. Yine de, Tüm Ev/Dairelerin sayısının fazla oluşu, daha çok tercih edilen bir tür olduğu yorumunu da birlikte getirmektedir.

4.2.2.3 En Çok Yorum Alan Konaklar

```
en_cok_yorum_alanlar = birlesik.sort_values(by='number_of_reviews', ascending=False).head(10)
en_cok_yorum_alanlar
```

	id	name	host_id	neighbourhood_left	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review
4822	17020360	Bed and breakfast in Zavróplivn · ★4.82 · 1 bed...	48229568	Θήραç (Santorini)	36.34745	25.45756	Private room	8000.0	1	1884	2023-10-15
338	773406	Home in Oia · ★4.82 · 1 bedroom · 3 beds · 1 bath	3867239	Θήραç (Santorini)	36.46242	25.37432	Entire home/apt	180.0	1	583	2023-11-20
1759	6116427	Place to stay in Oia · ★4.80 · 1 bedroom · 1 b...	3867239	Θήραç (Santorini)	36.46059	25.37449	Private room	100.0	1	534	2023-11-25
519	1395834	Home in Oia · ★4.79 · 2 bedrooms · 2 beds · 1 ...	3867239	Θήραç (Santorini)	36.46051	25.37615	Entire home/apt	130.0	1	495	2023-11-14
1895	6513290	Home in Oia · ★4.70 · 2 bedrooms · 3 beds · 1 ...	3867239	Θήραç (Santorini)	36.45988	25.37409	Entire home/apt	95.0	1	484	2023-11-30
341	970170	cycladic house in Oia · ★4.96 · 2 bedrooms · 4...	5044938	Θήραç (Santorini)	36.46069	25.37368	Entire home/apt	581.0	3	481	2023-11-29
358	974291	cycladic house in Oia · ★4.97 · 1 bedroom · 3 ...	5044938	Θήραç (Santorini)	36.46068	25.37380	Entire home/apt	465.0	3	463	2023-12-07
6467	20168768	Home in Oia · ★4.93 · 2 bedrooms · 2 beds · 1 ...	143542643	Θήραç (Santorini)	36.46170	25.37851	Entire home/apt	83.0	2	446	2023-12-17
168	887770	Loft in Oia · ★4.80 · 1 bedroom · 1 bed · 1.5 ...	1923369	Θήραç (Santorini)	36.46177	25.37376	Entire home/apt	188.0	3	425	2023-10-24
579	1556394	Rental unit in Mikonos · ★4.86 · 2 bedrooms · ...	7894137	Muxóvου (Mykonos)	37.44761	25.32801	Entire home/apt	23.0	2	424	2023-10-24

Şekil 4.12: En çok yorum alan konaklar listesi



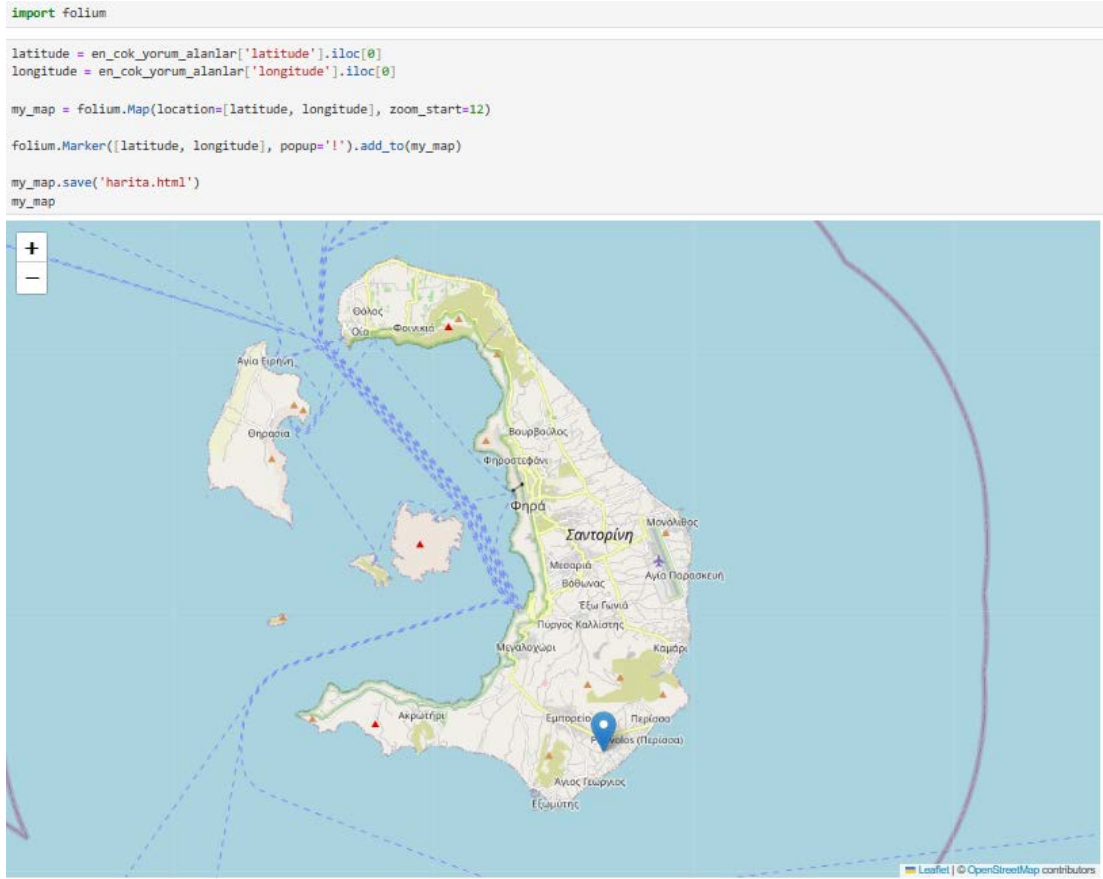
Şekil 4.13: En çok yorum alan konaklar ve yorum sayısı grafiği

Bu analiz ile, en çok yorum alan konakların bir listesi oluşturulmuştur. Bu analiz sonucunda 10. sırada yer alan konak, fiyat-performans değerlendirmesinde adeta bir ışık gibi parlamaktadır. Zira, hem en popüler adalardan olan Mikanos'ta yer almaktadır, hem de günlüğü 23 Euro'dur. Hakkında 424 yorum yapılmıştır. Son yorum tarihi oldukça günceldir. Bu yorumların olumlu yorum mu, olumsuz yorum mu olduğunun tespiti için proje sisteminin kolayca sunduğu http bağlantısına tıklamak yeterli olmuştur. Konak hakkında yapılan yorumların muazzam derecede iyi olduğu tespit edilmiştir. Ayrıca, 2 yatak odasına sahip geniş ve 4.86 yorum yıldızı almış bir yerdir. Tek problem, neredeyse her gününün rezerve edilmiş olmasıdır. Rezervasyonun çok önceden yapılması şarttır. Proje, yalnızca kısa vadeli seyahat ve rezervasyon planlarının yapılması için değil, aynı zamanda uzun vadeli seyahat ve rezervasyon planlarının yapılmasında da olumlu rol oynamaktadır.

Bölüm 5

Projenin Geliştirilmesi

5.1 Konakların Harita Üzerindeki Gösterimi



Şekil 5.1: En çok yorum alan konakların coğrafi konumları

Konakların harita üzerindeki gösterimi için 'folium' kütüphanesi kullanılmıştır. Folium; Python programlama dilinde kullanılan bir kütüphanedir ve interaktif haritalar

oluşturmak için kullanılır. Bu kütüphane, Leaflet.js adlı JavaScript harita kütüphanesini Python ortamına entegre ederek interaktif haritalar oluşturulmasını sağlar.

Proje kapsamında, veri seti içerisinde bulunan her konağa ait enlem ve boylam bilgileri (koordinatları) kullanılarak harita oluşturulmuş ve işaretçiler eklenerek bu harita bir HTML dosyası içerisine kaydedilmiştir.

'latitude' ve 'longitude' parametreleri, haritanın ortasının koordinatlarını, zoom_start ise haritanın başlangıçtaki yakınlaştırma düzeyini belirler. İşaretçiler folium.Marker ile eklenmiştir. Şekil 5.1'de yer alan mavi renkli işaret, projede incelenen tüm konakların orta noktasını gösterir ve ilgili konak Santorini Adası'na denk gelmektedir.

5.2 TF-IDF Matrisinin Oluşturulması

Projede bir içerik öneri sistemi gerçekleştirilmesi amaçlandığından veri setinde yer alan verilerden özellik seçimi yapılmış ve hem TF-IDF algoritmasından hem de kosinüs benzerliğinden yararlanılmıştır. Projede Python programlama dili kullanılmıştır. Bu programlama dilinin seçilme nedeni kütüphanelerinin açık kaynak kodlu olmasıdır (Kumaş, 2021).

TF-IDF (Term Frequency-Inverse Document Frequency), metin madenciliği ve bilgi algoritmalarında kullanılan popüler bir ölçüdür. Bu ölçü, çoğu zaman bir belgedeki bir terimin önemini belirlemek için kullanılır. Aynı zamanda belgeler arasındaki benzerlikleri değerlendirmek amacıyla da kullanılabilir (Gelemet vd., 2022).

TF-IDF matrisi, bir belge koleksiyonundaki belgelerin terim frekanslarını ve terimin belge koleksiyonu içindeki yaygınlığını içerir. İki ana kısmı vardır:

- TF (Term Frequency - Terim Frekansı)
- IDF (Inverse Document Frequency - Ters Belge Frekansı)

TF (Terim Frekansı); bir belgedeki bir terimin ne kadar sık geçtiğini belirten bir ölçüdür. Bir terimin bir belgedeki frekansı, o terimin belgedeki toplam kelime sayısına bölünerek bulunur.

IDF (Ters Belge Frekansı) ise; bir terimin belge koleksiyonundaki genel yaygınlığını ölçer. Nadir terimlere daha yüksek ağırlık verilmesini sağlar. IDF hesaplama kısmında "+1" terim t'nin hiçbir belgede geçmediği durumda sıfıra bölme hatası olmaması için eklenmiştir. Son olarak, TF ve IDF birleştirilerek TF-IDF skoru elde edilir.

$$TF(t, d) = \frac{\text{(Number of occurrences of term } t \text{ in document } d)}{\text{(Total number of terms in the document } d)}$$
$$IDF(t, D) = \log_e \frac{\text{(Total number of documents in the corpus)}}{\text{(Number of documents with term } t \text{ in them)}}$$
$$TF-IDF(t, d, D) = TF(t, d) \times IDF(t, D)$$

Şekil 5.2: TF-IDF Matrisi

TF-IDF matrisi, belge koleksiyonundaki her bir belge ve içerdikleri terimler arasındaki ölçümleri içerir. Bu matris, genellikle metin madenciliği uygulamalarında özellik matrisi olarak kullanılır ve metin belgelerini vektör uzayında temsil eder. Bu sayede belgeler arasındaki benzerlik ölçülebilir. Tüm bunlar için 'TfidfVectorizer' kullanılır.

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import linear_kernel

birlesik['http'] = 'https://tr.airbnb.com/rooms/' + birlesik['id'].astype(str)

def recommend_listing(user_input, birlesik, tfidf_vectorizer, tfidf_matrix):

    user_tfidf = tfidf_vectorizer.transform([user_input])
```

Şekil 5.3: İhtiyaca en uygun 5 konak

5.3 Cosine Similarity Hesaplama

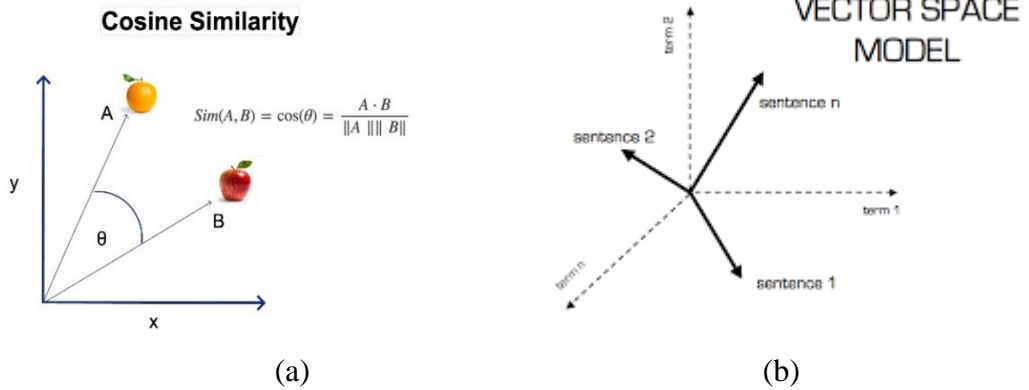
```
cosine_sim = linear_kernel(user_tfidf, tfidf_matrix)

similar_indices = cosine_sim[0].argsort()[::-6:-1]
recommendations = birlesik.iloc[similar_indices]

return recommendations[['name', 'neighbourhood_left', 'price', 'latitude', 'longitude', 'id', 'http']]
```

Şekil 5.4: Cosine similarity kullanımı

Cosine Similarity (Kosinüs Benzerliği), özellikle metin madenciliği ve bilgi algoritmalarında kullanılan iki vektörün birbirine olan benzerliğini ölçen bir metrik veya uzaklık ölçüsüdür (Amin ve Garg, 2019; Carrasco ve Rosillo, 2021). Kosinüs benzerliği metinler arasındaki vektörel uzaklığı ölçerek benzerliğin bulunmasını temel alır (Tata ve Patel, 2007; Liao vd., 2021).



Şekil 5.5: Cosine similarity ve uzay vektör modeli

İçerikler birer vektör olarak düşünüldüğünde kosinüs benzerliği; iki vektörün çarpımının iki vektörün boylarının çarpımına oranı olarak hesaplanır (Xia vd., 2015).

İki vektör arasındaki kosinüs benzerliği hesaplamak için şu formül kullanılır:

$$\text{Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \quad (5.6)$$

Bu formülde;

- $A \cdot B$, iki vektörün iç çarpımını ifade eder.
- $\|A\|$ ve $\|B\|$, sırasıyla A ve B vektörlerinin Euclidean normlarını ifade eder.

Cosine Similarity değeri -1 ile 1 arasında değişir. Değer ne kadar büyükse, vektörler o kadar benzerdir; değer ne kadar küçükse, vektörler o kadar farklıdır. Eğer iki vektör birbirine tamamen dik ise, cosine similarity 0 olur.

Cosine Similarity'nin metin madenciliğinde sıklıkla kullanılmasının nedeni, metinleri vektör olarak temsil edip ardından bu metrikle benzerlik ölçümü yapılabilmesidir. Örneğin, belgeler arasındaki benzerliği ölçmek, metin sınıflandırma, öneri sistemleri ve bilgi çıkarma gibi birçok uygulamada kullanılabilir.

5.4 Önerilerin Alınması ve Benzerlik Sıralama

```
user_input = input("What are you interested in? ")

recommendations = recommend_listing(user_input, birlesik, tfidf_vectorizer, tfidf_matrix)
recommendations
```

What are you interested in? New Villa 5.0 in Santorini with breakfast

	name	neighbourhood_left	price	latitude	longitude	id	http
920	Villa in Santorini - ★5.0 · 1 bedroom · 2 beds...	Θήραç (Santorini)	460.0	36.426598	25.428297	2752382	https://tr.airbnb.com/rooms/2752382
4516	Villa in Santorini · 1 bedroom · 3 beds · 1 bath	Θήραç (Santorini)	260.0	36.399440	25.452770	15534062	https://tr.airbnb.com/rooms/15534062
4923	Villa in Santorini - ★5.0 · 1 bedroom · 3 beds...	Θήραç (Santorini)	265.0	36.400540	25.451960	17350145	https://tr.airbnb.com/rooms/17350145
6779	Villa in Santorini - ★5.0 · 1 bedroom · 3 beds...	Θήραç (Santorini)	1000.0	36.465970	25.370660	21751897	https://tr.airbnb.com/rooms/21751897
22778	Villa in Santorini - ★5.0 · 1 bedroom · 2 beds...	Θήραç (Santorini)	220.0	36.352530	25.447259	639530165910224872	https://tr.airbnb.com/rooms/639530165910224872

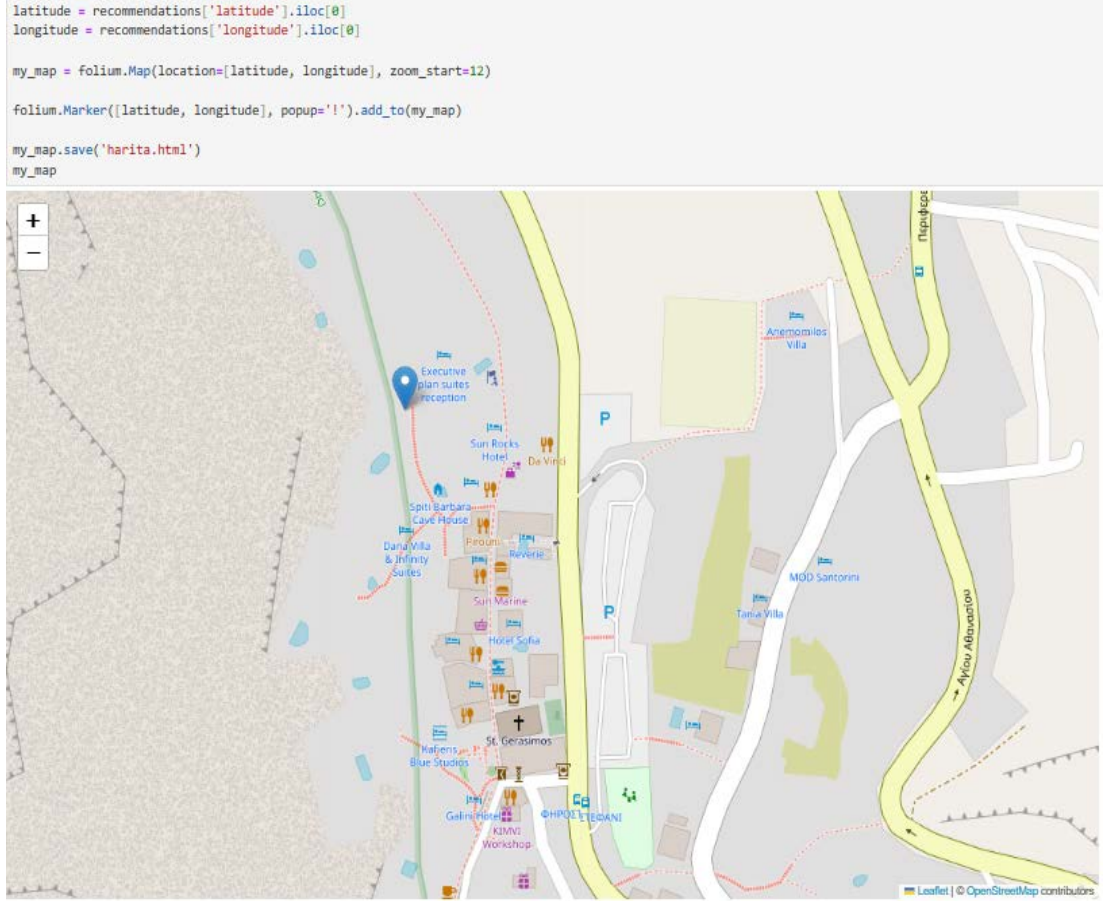
Şekil 5.7: Anahtar kelimelere göre önerilerin alınması

Kullanıcıdan alınan girişler ışığında, Cosine Similarity Hesaplama başlığı altında elde edilen benzerliklere göre listeleme girişleri sıralanır ve en yüksek benzerlik skorları en üstte yer almak üzere sıralanır.

Örneğin; Santorini’de, 5 yıldız yorum almış villa tipi bir konakta, kahvaltı dahil konaklamak isteyen bir kişiye öneri sunabilmek için; “*New Villa 5.0 in Santorini with breakfast*” gibi bir giriş yapmak mümkündür. Böylece, bu bilgilere en benzer nitelikteki sonuçlar birçok ilan arasından sıyrılıp saniyeler içerisinde alt alta sıralanır. Her farklı girişte farklı sonuçlar elde edilecektir.

Sonuçlara kısa sürede erişebilmek için http uzantılarının yer aldığı ek bir sütun da oluşturulmuştur.

5.5 Sonuçların Görüntülenmesi



Şekil 5.8: Önerilen konakların coğrafi konumları

Sıralanan sonuçların harita üzerinde görüntülenmesi ile ulaşım ve çevreye ilişkin bilgi sahibi olabilmek kolaylaşmıştır. Bunun için yine Konakların Harita Üzerindeki Gösterimi başlığı altında açıklanmış olan ‘folium’ kütüphanesi kullanılmıştır.

Bölüm 6

Sonuç

Bu projede, ihtiyaca göre konaklama öneri sistemi geliştirilmesine odaklanılmıştır. Metin verileri arasındaki benzerlikleri ölçmek amacıyla TF-IDF vektörlere ve Cosine Similarity gibi yöntemler kullanılarak bir öneri sistemi oluşturulmuştur.

TF-IDF vektörlere uygulanmasının ardındaki amaç, metin verilerini sayısal bir forma dönüştürmekte etkili bir yöntem oluşudur. Her bir konaklama yerinin belirli özelliklerini temsil eden vektörler oluşturularak, kullanıcının girişiyle benzer sonuçlar çıkarılmıştır.

Cosine Similarity kullanılarak elde edilen bu sonuçlar, başarılı bir şekilde kullanıcının taleplerine uygun olarak konaklama önerilerini sıralamış, elde edilen benzerlik skorlarına göre yapılan sıralama sonucunda, kullanıcıya en uygun konaklama yerlerini sunmak mümkün olmuştur. Bu öneri sistemi, kullanıcının terim girişlerine en yakın özelliklere sahip konaklama yerlerini belirleyerek, kullanıcı deneyimini geliştirmeyi hedeflemiştir.

Ayrıca, önerilen konakları harita üzerinde görmeyi, http uzantılarının yer aldığı ek bir sütun ile, ilgili ilanlara çok kısa bir sürede erişebilmeyi, konakları fiiliyatta daha kolay bulmayı ve harita üzerinden civar olanakları keşfetmeyi mümkün kılmıştır.

Geliştirilen öneri sistemi, yeni verilerle güncellenebilir ve genişletilebilir bir yapıya sahiptir. Bu sayede, sistemin performansı zamanla artırılabilir ve kullanıcıların değişen taleplerine, tercihlerine daha uygun öneriler sunabilir.

Kaynaklar

- Ahmed, H., Traore, I., & Saad, S. (2017). Detection of online fake news using n-gram analysis and machine learning techniques. *In International conference on intelligent, secure, and dependable systems in distributed and cloud environments*, 127-138.
- Aggarwal, C. C. (2016). *Recommender systems*. Springer.
- Akça, O. (2024). Öneri sistemi ile ürün tavsiyesi, *YBS Ansiklopedi*, 12, 1.
- Amin, D. M., & Garg, A. (2019). Performance Analysis of Data Mining Algorithms. *Journal of Computational and Theoretical Nanoscience*, 16 (9), 3849-3853.
- Budak, H., & Gümüştaş, E. (2022). Kişiselleştirilmiş ürün öneri sistemi için kullanıcı bazlı işbirlikçi filtreleme ve kümeleme kullanan hibrit bir yaklaşım, 1. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, 21-43, 253-268.
- Carrasco, S. S., & Rosillo, R. C. (2021). Word embeddings, cosine similarity and deep learning for identification of professions & occupations in health-related social media. *In Proceedings of the Sixth Social Media Mining for Health (#SMM4H) Workshop and Shared Task*, 74-76.
- Cingiz, M.Ö., & Marangoz, K. (2021). Kullanıcı tabanlı ve öge tabanlı işbirlikçi filtreleme ile kümeleme algoritmalarının değerlendirilmesi. *Avrupa Bilim ve Teknoloji Dergisi*, 28, 453-458. doi:10.31590/ejosat.1005391
- Felfernig, A., & Tran, T. N. T., & Le., V. M. (2021). Recommender systems beyond e-commerce: presence and future. *In Consumer Happiness: Multiple Perspectives* (203-230), Springer, Singapore.
- Gasparetti, F., Sansonetti, G., & Micarelli, A. (2021). Community detection in social recommender systems: a survey. *Applied Intelligence*, 51(6), 3975-3995.

- Gelemet, Ö., & Aydın, H., & Çetinkaya, A. (2022), Netflix verileri üzerinde Tf-idf algoritması ve kosinüs benzerliği ile bir içerik öneri sistemi uygulaması. *AJIT-E: Academic Journal of Information Technology*, 13 (48), 31-52. <https://doi.org/10.5824/ajite.2022.01.002.x>
- Ghazanfar, M. A., & Prügel-Bennett, A. & Szedmak, S. (2012). Kernel-mapping recommender system algorithms. *Information Sciences*, 208, 81-104.
- Ghose, A. Ipeirotis, P.G., & Li, B. (2010). Designing ranking systems for hotels on travel search engines to enhance user experience, *ISIC*.
- Karaman, D., & Gözüaçık, N., & Alagöz, M. O., & İlhan, H., & Çağal, U., & Yavuz, O. (2015). Managing 6LoWPAN Sensors with CoAP on Internet. *23st Signal Processing and Communications Applications Conference (SIU)*, Malatya, Türkiye. doi:10.1109/SIU.2015.7130101
- Karypis, G. (2001). Evaluation of item-based top-n recommendation algorithms. *Proceedings of the 10th International Conference on Information and Knowledge Management*, New York, USA: ACM, 247-254.
- Korkmaz, T., & Çetinkaya, A., & Aydın, H., & Barışkan, M. A. (2021). Analysis of whether news on the Internet is real or fake by using deep learning methods and the TF-IDF algorithm. *International Advanced Researches and Engineering Journal*, 5(1), 31-41.
- Kumaş, E. (2021). Türkçe Twitter verilerinden duygu analizi yapılırken sınıflandırıcıların karşılaştırılması. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, 2(2), 1-5.
- Liao, J., & Huang, Y., & Wang, H., & Li, M. (2021). Matching Ontologies with Word2Vec Model Based on Cosine Similarity. *In The International Conference on Artificial Intelligence and Computer Vision*, 367-374.
- Mahmood, T., & Ricci, F. (2009). Improving recommender systems with adaptive conversational strategies. *In Proceedings of the 20th ACM conference on Hypertext and hypermedia (HT '09)*. Association for Computing Machinery, New York, NY, USA, 73–82. <https://doi.org/10.1145/1557914.1557930>

- Martineau, J. C., & Finin, T. (2009). Delta Tf-idf: An improved feature space for sentiment analysis. *In Third international AAAI conference on weblogs and social media.*
- Meltem, I., & Çamurcu, A. Y. (2011). Web belgeleri kümelemede benzerlik ve uzaklık ölçütleri başarılarının karşılaştırılması. *Marmara Fen Bilimleri Dergisi*, 20(1), 35-49.
- Monti, D., & Rizzo, G., & Morisio, M. (2021). A systematic literature review of multicriteria recommender systems. *Artificial Intelligence Review*, 54, 427-468.
- Sayar, A., & Turdaliev, N. (2018). Makine Öğrenmesi ile Adaptif Otel Öneri Sistemi. *12. Ulusal Yazılım Mühendisliği Sempozyumu*, İstanbul, Türkiye.
- Sjarif, N. N. A., & Azmi, N. F. M., & Chuprat, S., & Sarkan, H. M., & Yahya, Y., & Sam, S. M. (2019). SMS spam message detection using term frequency-inverse document frequency and random forest algorithm. *Procedia Computer Science*, 161, 509-515.
- Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009(2009), 19.
- Tata, S., & Patel, J. M. (2007). Estimating the selectivity of tf-idf based cosine similarity predicates. *ACM Sigmod Record*, 36(2), 7-12.
- Thomas, B., & John, A. K. (2021). Machine learning techniques for recommender systems—a comparative case analysis. *In IOP Conference Series: Materials Science and Engineering* (Vol. 1085, No. 1, p. 012011). IOP Publishing.
- Türker, B. B., & Tugay, R., & Kızıl, I., & Öğüdücü, S. (2019). Hotel recommendation system based on user profiles and collaborative filtering. *2019 4th International Conference on Computer Science and Engineering (UBMK)*, 601-606, Samsun, Türkiye.

- Walek, B., & Farana, R., & Hosek, O., & Lomnica T. (2016). Proposal of expert system for hotel booking system. *IEEE, Carpathian Control Conference (ICCC)*.
- Xia, P., & Zhang, L., & Li, F. (2015). Learning similarity with cosine similarity ensemble. *Information Sciences*, 307, 39-52.
- Yalçın, E. (2016). Naive Bayes sınıflandırıcı tabanlı ikili-veri çoklu-ölçütlü öneri sistemleri (Yayın No. 438122) [Yüksek lisans tezi, Eskişehir Anadolu Üniversitesi]. YÖK Ulusal Merkezi. <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Yücebaş, S. C. (2019). MovieANN: A hybrid approach to movie recommender systems using multilayer artificial neural networks. *Çanakkale Onsekiz Mart Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 5(2), 214-232.
- Zhang, Q., & Lu, J., & Jin, Y. (2021). Artificial intelligence in recommender systems. *Complex & Intelligent Systems*, 7(1), 439-457.