

Waste Time Series ve Location Verilerinin Analizi

(Task2)

1. Projeye Genel Bakış

Bu projede, atık yönetim süreçlerini analiz etmek amacıyla “waste_data_timeseries” veri kümesi kullanılmıştır. Bu veri kümesi, her ziyarette kaydedilen atık kutularının doluluk seviyelerini yüzde olarak içermektedir. Amaç, veri temizleme ve dönüştürme işlemlerinin ardından, veriyi anlamlı görselleştirmelerle analiz etmektir. Power BI kullanılarak, karar vericilere atık toplama verilerinin günün farklı saatlerinde ve çeşitli konumlarda nasıl bir dağılım gösterdiği hakkında bilgi sağlayan bir gösterge paneli oluşturulmuştur.

2. Veri Tanımlaması

2.1 Elimizde “waste_data_timeseries.csv” veri seti bulunmaktadır , 6 sütundan oluşmaktadır.

SPID	Visit Date	# Paper	% Paper	# Plastic	% Plastic
15801	2019-01-01 08:09:50	3	1	2	0.75
16160	2019-01-01 08:22:15	2	0.75	1	0.75
9960	2019-01-01 08:29:48	2	1	0	0
16135	2019-01-01 08:38:08	0	0	1	0.25
16017	2019-01-01 08:42:03	3	0.75	2	0.62
9942	2019-01-01 09:04:28	1	1	0	0

SPID: Konumun benzersiz kimliğini temsil eden UNIX tipinde atanmış kimlik numarası.

VisitDate: Ziyaretin gerçekleştirildiği tarih ve saati gösterir.

Paper: O noktada bulunan kağıt atık kutu sayısını gösterir. Birim, muhtemelen kilogram veya belirli bir hacim birimi olabilir.

% Paper: Kağıt atığının doluluk yüzdesini gösterir. Bu, kağıt kutusunun toplam kapasitesine göre ne kadar dolu olduğunu ifade eder.

Plastic: O noktada bulunan plastik atık kutu sayısını gösterir. Birim, muhtemelen kilogram veya belirli bir hacim birimi olabilir.

% Plastic: Plastik atığının doluluk yüzdesini gösterir. Bu, plastik kutusunun toplam kapasitesine göre doluluk oranını ifade eder.

2.2 Verilen bir diğer veri seti ise “location_data.csv” . Bu veri seti ise 3 sütundan oluşmaktadır.

Y	SPID Y	Latitude Y	Longitude Y
0	10222	39.93	32.86
1	10374	39.92	32.99
2	10375	39.92	32.96
3	10377	39.92	32.98
4	10912	39.99	33.11
5	10914	40.01	33.16

SPID: Konumun benzersiz kimliğini temsil eden UNIX tarzında atanmış kimlik numarası. Bu, atık toplama noktalarının tanımlanması için kullanılır.

Latitude: Atık toplama konumunun enlem (kuzey-güney) koordinatını gösterir.

Longitude: Atık toplama konumunun boylam (doğu-batı) koordinatını gösterir.

3. Verilerin Temizlenmesi ve Yeni Veri Setlerinin Oluşturulması

3.1 waste_data_timeseries.csv incelenmesi temizlenmesi ve yeni veri setinin oluşturulması .Bu işlem python dili kullanılarak yapıldı.

```
import pandas as pd
waste_data = pd.read_csv('waste_data_timeseries.csv')
waste_data['% Paper'] = waste_data['% Paper'].apply(lambda x: min(x, 1))
waste_data['% Plastic'] = waste_data['% Plastic'].apply(lambda x: min(x, 1))
waste_data['total_waste'] = waste_data['# Paper'] + waste_data['# Plastic']
waste_data['total_waste%'] = ((waste_data['# Paper'] * waste_data['% Paper']) + (waste_data['# Plastic'] * waste_data['% Plastic'])) / waste_data['total_waste']
waste_data['VisitDate'] = pd.to_datetime(waste_data['VisitDate'])
waste_data = waste_data.sort_values(by='VisitDate')
data_types = waste_data.dtypes
print(data_types)
waste_data.to_csv("waste_data_update.csv", index=False)
```

Bu kod, "waste_data_timeseries.csv" adlı veri kümesi üzerinde veri temizleme, dönüşüm ve yeni sütunlar ekleme işlemleri yapmaktadır. İlk olarak, dosya pandas ile okunur. Sonrasında % Paper ve % Plastic sütunlarındaki değerler 1'den büyükse 1'e sabitlenerek doluluk oranlarının %100'den fazla olmaması sağlanır. Yeni bir total_waste sütunu oluşturularak kağıt ve plastik atıkların toplam miktarı hesaplanır. Ardından total_waste% adında bir sütun eklenir ve bu sütun, ağırlıklı doluluk oranlarını göstermek için hesaplanır. Bu işlemde, # Paper ve # Plastic değerleri, ilgili doluluk yüzdeleriyle çarpılarak atık toplam doluluğunu temsil eden bir ortalama değeri elde edilir. VisitDate sütunu datetime formatına çevrilip tarih sırasına göre sıralanarak veri zaman dizisi analizi için düzenlenir. Kodun sonunda, güncellenmiş veri yeni bir "waste_data_update.csv" dosyasına kaydedilir ve veri türleri konsola yazdırılır.

SPID	Visit Date	Paper_No	Paper_Percentage	Plastic_No	Plastic_Percentage	Total_waste_No	Total_waste_Percentage
15801	2019-01-01 08:09:50	3	1	2	0.75	5	0.9
16160	2019-01-01 08:22:15	2	0.75	1	0.75	3	0.75
9960	2019-01-01 08:29:48	2	1	0	0	2	1
16135	2019-01-01 08:38:08	0	0	1	0.25	1	0.25
16017	2019-01-01 08:42:03	3	0.75	2	0.62	5	0.7
9942	2019-01-01 09:04:28	1	1	0	0	1	1

Aynı zamanda sütunların adından özel karakterleri manuel biçimde değiştirip , sütun adlarını değiştirildi.

3.2 Oluşturulan "waste_data_update.csv" veri seti kullanılarak verileri aylık şeklinde yeniden düzenlendi.

```
import pandas as pd

waste_data = pd.read_csv('waste_data_update.csv')

waste_data['VisitDate'] = pd.to_datetime(waste_data['VisitDate'])

waste_data['YearMonth'] = waste_data['VisitDate'].dt.to_period('M')

def monthly_percentage(df, column_count, column_percentage):

    weighted_sum = (df[column_count] * df[column_percentage]).sum()

    total_sum = df[column_count].sum()

    return weighted_sum / total_sum if total_sum > 0 else 0

monthly_results = waste_data.groupby('YearMonth').apply(

    lambda x: pd.Series({

        'Monthly Paper Percentage': monthly_percentage(x, '# Paper', '% Paper'),

        'Monthly Plastic Percentage': monthly_percentage(x, '# Plastic', '% Plastic')

    })

).reset_index()

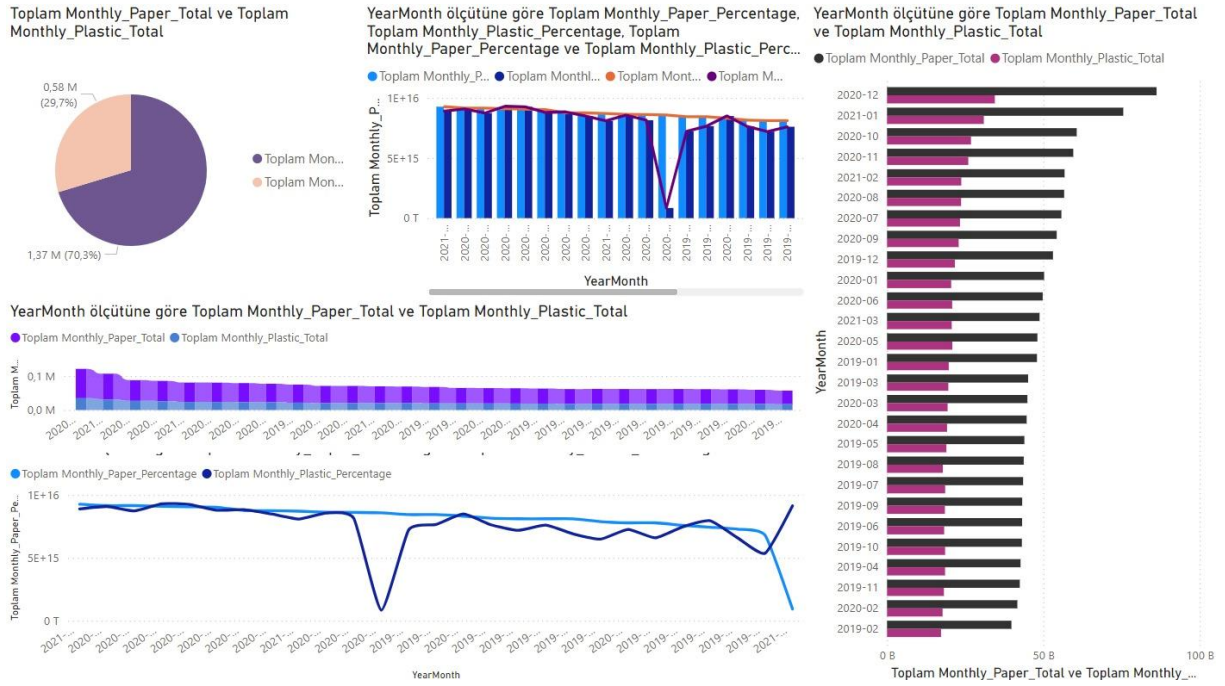
print(monthly_results)

monthly_results.to_csv("monthly_paper_plastic_percentage.csv", index=False)
```

Bu kod, her ay için kağıt ve plastik atık doluluk oranlarının ağırlıklı ortalamalarını hesaplamaktadır. İlk olarak, "waste_data_update.csv" dosyasını pandas ile okuyarak VisitDate sütununu datetime formatına çevirir. Daha sonra YearMonth adında bir sütun ekleyerek her kaydın ait olduğu yıl ve ayı dönüştürür. monthly_percentage adlı bir fonksiyon oluşturulur; bu fonksiyon, belirtilen sayısal sütun ile yüzdelik sütunları çarparak ağırlıklı toplamı hesaplar ve her ay için bu ağırlıklı doluluk oranını bulur. Veri, YearMonth sütununa göre gruplandırılarak kağıt ve plastik için aylık doluluk ortalamaları hesaplanır. Sonuçlar monthly_results adlı veri çerçevesine kaydedilir, ekrana yazdırılır ve ardından "monthly_paper_plastic_total_counts_percentage.csv" dosyasına kaydedilir.

Year Month	Monthly Paper Percentage	Monthly Plastic Percentage	Monthly Total Percentage	Monthly Paper Total	Monthly Plastic Total
2019-01	0.84	0.72	0.81	4794	1973
2019-02	0.81	0.76	0.8	3977	1726
2019-03	0.76	0.75	0.75	4511	1960
2019-04	0.81	0.76	0.8	4269	1855
2019-05	0.81	0.72	0.78	4391	1898
2019-06	0.79	0.65	0.75	4319	1827
2019-07	0.68	0.54	0.64	4351	1859

3.2.1 Veri Setinin görselleştirilmesi.



Şekil1

Şekil1 görselinde, belirli zaman dilimlerinde kağıt ve plastik atık miktarlarının aylık olarak nasıl değiştiği ve aralarındaki yüzdesel dağılım farklı grafiklerle detaylı bir şekilde gösterilmiştir. Sol üstte yer alan pasta grafik, toplam kağıt ve plastik miktarlarının yüzdesel dağılımını gözler önüne seriyor. Grafik üzerinde, toplam kağıt miktarının %70,3 oranında olduğunu ve değeri 1,37 milyon olarak gösterildiğini, plastik miktarının ise %29,7 oranında ve 0,58 milyon değerinde olduğunu görmekteyiz. Bu grafik, toplam atık içinde kağıt ve plastiğin oranlarını karşılaştırmak için kullanılmış.

Sağ üstte bulunan çubuk grafik ise aylık bazda kağıt ve plastik atık miktarlarının zaman içindeki değişimini ayrıntılı bir şekilde yansıtıyor. X eksenini her bir ayı temsil ederken, Y eksenini toplam kağıt ve plastik miktarını gösteriyor. Bu grafikte, mavi çubuklar kağıt atık miktarını, pembe çubuklar ise plastik atık miktarını temsil ediyor. Grafik, kağıt ve plastik atıkların aylara göre nasıl dalgalandığını görmeyi sağlıyor.

Orta altta yer alan yığılmış sütun grafik ise kağıt ve plastik atık miktarlarının bir araya getirilmiş olarak zamanla nasıl değiştiğini gösteriyor. Yığılmış sütun grafikte, her ay için kağıt (açık mavi) ve plastik (mor) atık miktarları üst üste eklenerek toplam atık miktarı bir bütün olarak sunuluyor. Bu grafik, aylık toplam atık miktarlarının zaman içindeki değişimlerini anlamak açısından oldukça kullanışlı.

Alt ortadaki çizgi grafik, kağıt ve plastik atık yüzdesel oranlarının zamanla nasıl değiştiğine odaklanıyor. Bu grafikte mavi çizgi kağıt yüzdesini, açık mavi çizgi ise plastik yüzdesini gösteriyor. Zaman içerisinde yüzdesel oranların nasıl dalgalandığını ve belirli aylarda ani değişiklikler olup olmadığını bu grafik üzerinden inceleyebiliriz.

Sağdaki yatay çubuk grafik ise her ay için toplam kağıt ve plastik miktarlarını karşılaştırmak amacıyla oluşturulmuş. Y eksenini her bir ayı temsil ederken, X eksenini toplam kağıt ve plastik miktarını gösteriyor. Siyah çubuklar kağıt atık miktarını, pembe çubuklar ise plastik atık miktarını ifade ediyor. Her ay için kağıt ve plastik miktarları yatay olarak kıyaslanarak, hangi aylarda kağıt veya plastik atığın daha fazla olduğu kolayca gözlemlenebiliyor.

Bu rapor, kağıt ve plastik atık miktarlarının ve yüzdelere ilişkin aylık olarak nasıl değiştiğini görsel olarak analiz etmek için farklı grafik türlerini bir arada kullanıyor. Her bir grafik, veri analizi açısından farklı bir bakış açısı sunarak, toplam atıkların ve yüzdelere ilişkin zaman içindeki değişimlerini anlamayı kolaylaştırıyor.

3.3 Oluşturulan “waste_data_update.csv” veri seti kullanılarak verilerin günlük olarak en yoğun saatlerini gösterecek şekilde yeniden düzenlenecek.

```
import pandas as pd

waste_data = pd.read_csv('waste_data_update.csv')

waste_data['VisitDate'] = pd.to_datetime(waste_data['VisitDate'])

waste_data['Date'] = waste_data['VisitDate'].dt.date

waste_data['Hour'] = waste_data['VisitDate'].dt.hour

daily_hourly_totals = waste_data.groupby(['Date', 'Hour'])['total_waste'].sum().reset_index()

max_hour_per_day =
daily_hourly_totals.loc[daily_hourly_totals.groupby('Date')['total_waste'].idxmax()]

print(max_hour_per_day)

max_hour_per_day.to_csv("daily_max_hour_waste.csv", index=False)
```

Bu kod, her gün için saat bazında atık toplama miktarını analiz ederek en fazla atık toplanan saati bulur. İlk olarak "waste_data_update.csv" dosyasını okuyarak VisitDate sütununu datetime formatına çevirir ve tarih ile saat bilgilerini ayrı sütunlara böler. Ardından, Date ve Hour sütunlarına göre gruplandırma yaparak her gün için saat bazında total_waste değerlerinin toplamını hesaplar ve daily_hourly_totals adlı bir veri çerçevesine kaydeder. Daha sonra, her gün için en fazla atık toplanan saati bulmak amacıyla, total_waste sütununa göre maksimum değeri veren saatleri max_hour_per_day veri çerçevesinde toplar. Son olarak, bu veriyi "daily_max_hour_waste.csv" dosyasına kaydeder ve ekrana yazdırır.

Date	Hour	Total_waste
2019-01-01	12	33
2019-01-02	9	51
2019-01-03	8	41
2019-01-04	10	49
2019-01-05	10	41
2019-01-06	16	43
2019-01-07	9	42
2019-01-08	9	21
2019-01-09	10	40

Oluşan veri setinin bir parçasının görseli bu şekilde.

3.4 Oluşturulan "waste_data_update.csv" veri seti kullanılarak verilerin iklimsel olarak ayrıştırılması.

```
import pandas as pd

df = pd.read_csv('monthly_paper_plastic_total_counts_percentage.csv')
df = df[df['YearMonth'] < '2021-01']
df['YearMonth'] = pd.to_datetime(df['YearMonth'])
df['Year'] = df['YearMonth'].dt.year
df['Month'] = df['YearMonth'].dt.month

seasons = {
    'Spring': [3, 4, 5], 'Summer': [6, 7, 8, 9], 'Fall': [10, 11], 'Winter': [12, 1, 2]
}

seasonal_data = []

for year in [2019, 2020]:
    for season, months in seasons.items():
        season_df = df[(df['Year'] == year) & (df['Month'].isin(months))]
        paper_avg = season_df['Monthly Paper Percentage'].mean()
        plastic_avg = season_df['Monthly Plastic Percentage'].mean()
        total_avg = (paper_avg + plastic_avg) / 2
```

```

paper_total = season_df['Monthly Paper Total'].sum()
plastic_total = season_df['Monthly Plastic Total'].sum()

seasonal_data.append({
    'Year': year,
    'Season': season,
    'Paper_Avg': paper_avg,
    'Plastic_Avg': plastic_avg,
    'Total_Avg': total_avg,
    'Paper_Total': paper_total,
    'Plastic_Total': plastic_total
})

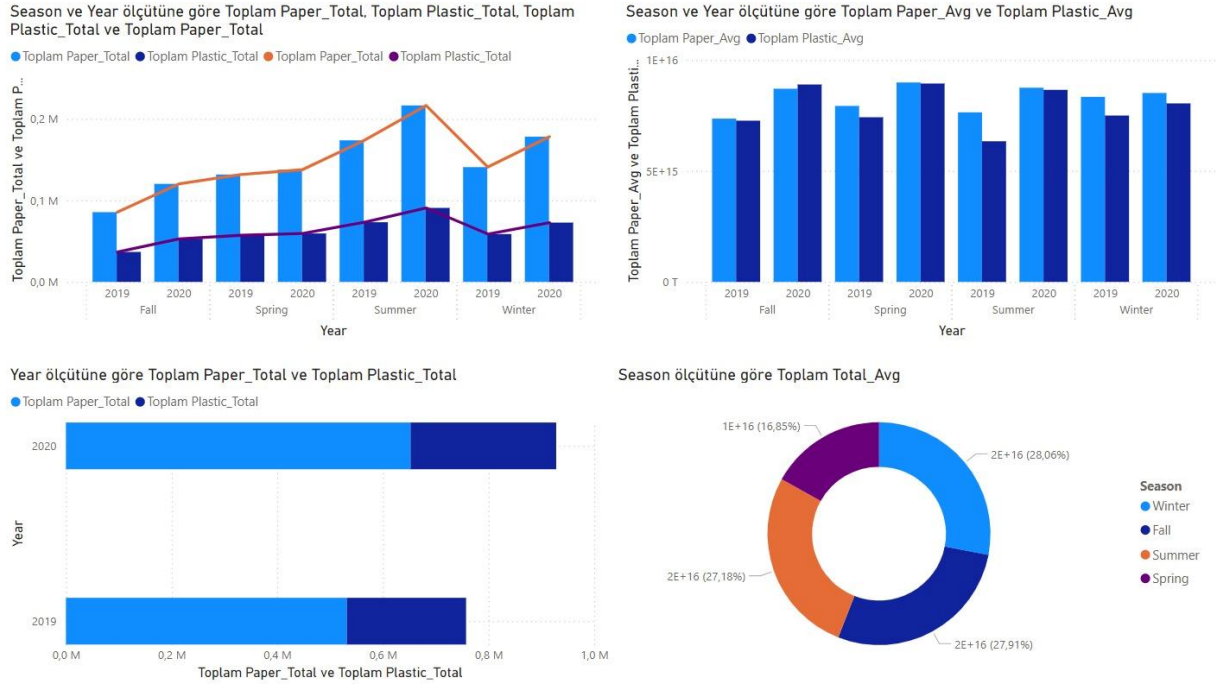
seasonal_df = pd.DataFrame(seasonal_data)
seasonal_df.to_csv('seasonal_results.csv', index=False)
print("Yeni veri seti 'seasonal_results.csv' dosyasına kaydedildi.")

```

Bu kod, 2019 ve 2020 yıllarındaki aylık kağıt ve plastik atık verilerini mevsimlere göre gruplandırarak ortalama ve toplam hesaplamaları yapmaktadır. İlk olarak, "monthly_paper_plastic_total_counts_percentage.csv" dosyasından verileri okuyarak 2021 yılına ait verileri filtreler. Ardından, YearMonth sütununu datetime formatına çevirip yıl ve ay bilgilerini çıkarır. Mevsimler, bir sözlük yapısı kullanılarak tanımlanır. Kod, her yıl ve mevsim için döngüye girerek ilgili verileri filtreler ve mevsime ait kağıt ve plastik yüzdelerinin ortalamalarını, toplam kağıt ve plastik kutu sayılarını hesaplar. Hesaplanan değerler seasonal_data listesine eklenir. Son olarak, bu veriler yeni bir DataFrame'e dönüştürülür ve "seasonal_results.csv" dosyasına kaydedilir. Kod, işlemin başarılı olduğunu bildiren bir mesajla tamamlanır.

Year	Season	Paper_Avg	Plastic_Avg	Total_Avg	Paper_Total	Plastic_Total
2019	Spring	0.79	0.74	0.77	13171	5713
2019	Summer	0.76	0.63	0.7	17363	7319
2019	Fall	0.74	0.73	0.73	8557	3674
2019	Winter	0.83	0.75	0.79	14077	5867
2020	Spring	0.9	0.89	0.9	13760	5942
2020	Summer	0.88	0.87	0.87	21642	9074
2020	Fall	0.87	0.89	0.88	12020	5283
2020	Winter	0.85	0.8	0.83	17809	7274

3.4.1 Veri setinin görselleştirilmesi.



Şekil2

Şekil2 olarak belirtilen görselde , belirli yıllar ve mevsimlere göre toplam kağıt ve plastik miktarlarının nasıl değiştiği, ortalama değerlerin karşılaştırması ve yüzdesel dağılımı detaylı bir şekilde sunulmuştur. Grafikler, 2019 ve 2020 yılları arasındaki kağıt ve plastik atık miktarlarının toplam ve ortalama değerlerini göstererek her mevsimdeki değişimleri incelememize olanak tanır.

Sol üstteki ilk grafik, mevsim ve yıl bazında toplam kağıt ve plastik miktarlarının dağılımını gösterir. Çubuk grafikler kağıt (mavi) ve plastik (mor) miktarlarını temsil ederken, çizgiler yıllara göre eğilim değişikliklerini vurgulamaktadır. Örneğin, 2020 yazında kağıt miktarında belirgin bir artış gözlemlenirken, plastik miktarında ise bir düşüş görülmektedir. Bu grafik, kağıt ve plastik atıkların toplam miktarlarının mevsimsel olarak nasıl değiştiğini anlamak için faydalıdır.

Sağ üstteki ikinci grafik, yine mevsim ve yıl bazında kağıt ve plastik atık miktarlarının ortalamalarını kıyaslar. Burada mavi çubuklar kağıt atığının ortalama değerlerini, açık mavi çubuklar ise plastik atığının ortalama değerlerini göstermektedir. Yıl ve mevsim bazında incelendiğinde, 2020 yazında kağıt atığının ortalama miktarının plastikten daha fazla olduğu görülmektedir. Bu grafik, farklı yıllardaki mevsimsel değişiklikleri ortalama bazında analiz etmek isteyenler için değerli bilgiler sunmaktadır.

Sol alttaki üçüncü grafik, yıllara göre toplam kağıt ve plastik atık miktarlarını kıyaslayan yatay çubuk grafik. Y ekseninde yıllar yer alırken, X ekseninde toplam kağıt ve plastik miktarlarını gösterir. Burada, 2020 yılında kağıt miktarının 2019'a göre daha yüksek olduğu net bir şekilde görülmektedir. Bu grafik, yıllık bazda kağıt ve plastik atıkların toplam miktarlarını karşılaştırmak için idealdir.

Sağ alttaki dördüncü grafik ise mevsim bazında kağıt ve plastik atıkların toplam ortalama değerlerinin yüzdesel dağılımını sunar. Pasta grafikte kış (mavi), sonbahar (mor), yaz (turuncu) ve ilkbahar (açık mavi) renkleri ile mevsimler belirtilmiştir. Yaz ve ilkbahar aylarının toplam ortalama değerleri diğer mevsimlere göre daha yüksek bir yüzdesel orana sahipken, kış aylarında en düşük oran gözlemlenmektedir. Bu grafik, mevsimlerin toplam atık miktarlarına katkısını yüzdesel olarak görmemizi sağlar.

3.5 Oluşturulan “waste_data_update.csv” veri seti kullanılarak verilerin özel günlerdeki durumunu göstermek için analiz edildi ve yeni veri seti oluşturuldu.

```
import pandas as pd

waste_data = pd.read_csv('waste_data_update.csv')

waste_data['VisitDate'] = pd.to_datetime(waste_data['VisitDate'])

special_days = [

    # 2019 yılı

    '2019-01-01', # Yılbaşı

    '2019-04-23', # Ulusal Egemenlik ve Çocuk Bayramı

    '2019-05-01', # Emek ve Dayanışma Günü

    '2019-05-19', # Atatürk'ü Anma, Gençlik ve Spor Bayramı

    '2019-06-04', # Ramazan Bayramı Arife Günü

    '2019-06-05', # Ramazan Bayramı 1. Gün

    '2019-06-06', # Ramazan Bayramı 2. Gün

    '2019-06-07', # Ramazan Bayramı 3. Gün

    '2019-07-15', # Demokrasi ve Millî Birlik Günü

    '2019-08-10', # Kurban Bayramı Arife Günü

    '2019-08-11', # Kurban Bayramı 1. Gün

    '2019-08-12', # Kurban Bayramı 2. Gün

    '2019-08-13', # Kurban Bayramı 3. Gün

    '2019-08-14', # Kurban Bayramı 4. Gün

    '2019-08-30', # Zafer Bayramı

    '2019-10-29', # Cumhuriyet Bayramı

    # 2020 yılı

    '2020-01-01', # Yılbaşı

    '2020-04-23', # Ulusal Egemenlik ve Çocuk Bayramı
```

```

'2020-05-01', # Emek ve Dayanışma Günü
'2020-05-19', # Atatürk'ü Anma, Gençlik ve Spor Bayramı
'2020-05-23', # Ramazan Bayramı Arife Günü
'2020-05-24', # Ramazan Bayramı 1. Gün
'2020-05-25', # Ramazan Bayramı 2. Gün
'2020-05-26', # Ramazan Bayramı 3. Gün
'2020-07-15', # Demokrasi ve Millî Birlik Günü
'2020-07-30', # Kurban Bayramı Arife Günü
'2020-07-31', # Kurban Bayramı 1. Gün
'2020-08-01', # Kurban Bayramı 2. Gün
'2020-08-02', # Kurban Bayramı 3. Gün
'2020-08-03', # Kurban Bayramı 4. Gün
'2020-08-30', # Zafer Bayramı
'2020-10-29', # Cumhuriyet Bayramı
# 2021 yılı
'2021-01-01', # Yılbaşı
]

special_days = pd.to_datetime(special_days)

special_day_summary = waste_data[waste_data['VisitDate'].dt.date.isin(special_days.date)]

special_day_summary = (
    special_day_summary
    .groupby(special_day_summary['VisitDate'].dt.date)
    .agg({
        '# Paper': 'sum', '% Paper': 'mean', '# Plastic': 'sum', '% Plastic': 'mean', 'total_waste': 'sum'
    })
    .reset_index()
    .rename(columns={'VisitDate': 'Special Day'})
)

print(special_day_summary)

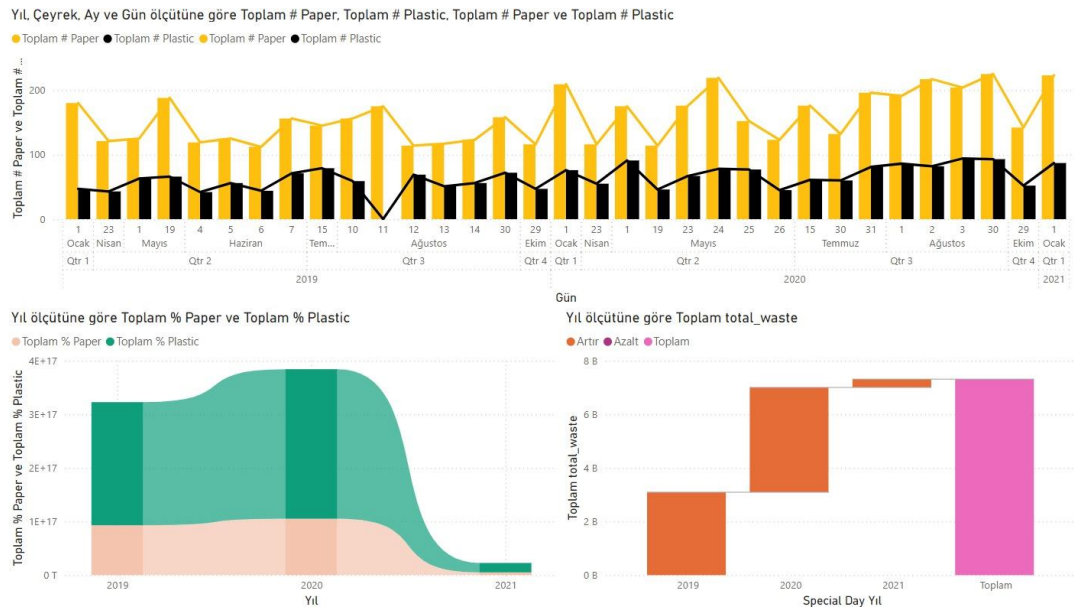
special_day_summary.to_csv("special_day_summary_2019_2021.csv", index=False)

```

Bu Python kodu, atık verilerini içeren bir CSV dosyasını okuyarak belirli özel günlerdeki atık miktarlarını analiz eder. İlk olarak, pandas kütüphanesi kullanılarak veri seti yüklenir ve VisitDate sütunu datetime formatına dönüştürülür. Ardından, Türkiye'nin 2019, 2020 ve 2021 yıllarındaki özel günler bir liste olarak tanımlanır ve bu tarihler datetime formatına çevrilir. Kod, daha sonra yalnızca bu özel günlerdeki verileri filtreleyerek, her bir gün için toplam kağıt ve plastik kutu sayısını, bu atık türlerinin yüzdelerinin ortalamasını ve toplam atık miktarını hesaplar. Elde edilen sonuçlar, özel günlerin tarihlerini içeren bir veri çerçevesinde toplanır ve special_day_summary_2019_2021.csv adlı yeni bir dosyaya kaydedilir. Bu süreç, atık yönetimi ve geri dönüşüm uygulamalarının etkinliğini değerlendirmek için önemli bir adım sağlar.

Special Day	# Paper	% Paper	# Plastic	% Plastic	Total_waste
2019-01-01	180	0.72	47	0.27	227
2019-04-23	121	0.77	43	0.42	164
2019-05-01	125	0.54	63	0.44	188
2019-05-19	188	0.63	66	0.25	254
2019-06-04	119	0.78	42	0.42	161
2019-06-05	125	0.51	56	0.36	181
2019-06-06	112	0.61	44	0.27	156
2019-06-07	156	0.61	71	0.38	227
2019-07-15	145	0.5	79	0.39	224
2019-08-10	156	0.36	59	0.12	215

3.5.1 Veri setinin görselleştirilmesi.



Şekil3

Bu görsel, 2019, 2020 ve 2021 yıllarında belirlenen özel günler ve tatil dönemleri verisine dayanarak kağıt ve plastik atıkların miktarını inceleyen grafiklerden oluşmaktadır. Veriler, yalnızca bu özel günlere ait bilgilerden seçilmiş ve yıl, çeyrek, ay bazında toplam atık miktarlarının karşılaştırması yapılmıştır. Bu sayede, özel günlerdeki kağıt ve plastik atık üretimindeki değişimlerin analizi gerçekleştirilmiştir.

Üstteki Grafik: İlk grafik, yıllık çeyrek, ay ve gün ölçütüne göre kağıt ve plastik atıkların toplam miktarlarını göstermektedir. Sarı çubuklar kağıt, siyah çubuklar ise plastik atık miktarlarını ifade ederken, trend çizgileri yıl içerisindeki değişimleri vurgulamaktadır. 2020 yılı ikinci çeyreğinde kağıt atık miktarında belirgin bir artış görülmektedir, bu da muhtemelen o dönemki tatil günlerinde kağıt tüketiminin artmış olabileceğine işaret eder.

Sol Alttaki Grafik: Bu grafik, yıllık bazda kağıt ve plastik atıkların yüzdesel dağılımını sergilemektedir. Turuncu renk kağıt, yeşil renk ise plastik atığı temsil etmektedir. 2020 yılında plastik atık yüzdesinde bir artış görülmekte, bu da o yılki özel günlerde plastik tüketiminin daha yoğun olabileceğini göstermektedir. 2021 yılına gelindiğinde ise plastik kullanımının oranında düşüş yaşandığı fark edilmektedir.

Sağ Alttaki Grafik: Son grafik, yıllık ölçekte toplam atık miktarındaki artış ve azalışı göstermek için tasarlanmıştır. Turuncu çubuklar artış, pembe çubuklar toplam atık miktarını belirtirken, bu grafik yıllık toplam atık eğilimini anlamak için kullanılır. Örneğin, 2021 yılı için genel atık miktarının önceki yıllara göre daha az olduğu görülmektedir, bu da çevresel farkındalığın arttığı veya tüketimin azaldığı bir dönem olarak değerlendirilebilir.

3.6 Oluşturulan “waste_data_update.csv” veri seti kullanılarak verilerin hafta içi ve hafta sonu olarak ayrımı yapılmıştır , veri seti oluşturulmuştur .

```
import pandas as pd

waste_data = pd.read_csv('waste_data_update.csv')

waste_data['VisitDate'] = pd.to_datetime(waste_data['VisitDate'])

waste_data['Year'] = waste_data['VisitDate'].dt.year

waste_data['DayOfWeek'] = waste_data['VisitDate'].dt.weekday

waste_data = waste_data[(waste_data['Year'] < 2021) | ((waste_data['Year'] == 2021) &
(waste_data['VisitDate'] <= '2021-03-24'))]

summary_data = []
```

```
for year in [2019, 2020, 2021]:
```

```
    yearly_data = waste_data[waste_data['Year'] == year]
    weekday_data = yearly_data[yearly_data['DayOfWeek'] < 5]
    weekday_paper_count = weekday_data['# Paper'].sum()
    weekday_paper_percent = (weekday_data['# Paper'] * weekday_data['% Paper']).sum() /
    weekday_paper_count if weekday_paper_count > 0 else 0
    weekday_plastic_count = weekday_data['# Plastic'].sum()
    weekday_plastic_percent = (weekday_data['# Plastic'] * weekday_data['% Plastic']).sum() /
    weekday_plastic_count if weekday_plastic_count > 0 else 0
    weekday_total = weekday_data['total_waste'].sum()
```

```
    weekend_data = yearly_data[yearly_data['DayOfWeek'] >= 5]
    weekend_paper_count = weekend_data['# Paper'].sum()
    weekend_paper_percent = (weekend_data['# Paper'] * weekend_data['% Paper']).sum() /
    weekend_paper_count if weekend_paper_count > 0 else 0
    weekend_plastic_count = weekend_data['# Plastic'].sum()
    weekend_plastic_percent = (weekend_data['# Plastic'] * weekend_data['% Plastic']).sum() /
    weekend_plastic_count if weekend_plastic_count > 0 else 0
    weekend_total = weekend_data['total_waste'].sum()
```

```
    summary_data.append({
        'Year': year,
        'Weekday # Paper': weekday_paper_count, 'Weekday % Paper': weekday_paper_percent,
        'Weekday # Plastic': weekday_plastic_count,
        'Weekday % Plastic': weekday_plastic_percent, 'Weekday Total': weekday_total,
        'Weekend # Paper': weekend_paper_count,
        'Weekend % Paper': weekend_paper_percent,
        'Weekend # Plastic': weekend_plastic_count,
        'Weekend % Plastic': weekend_plastic_percent, 'Weekend Total': weekend_total
    })
```

```
summary_df = pd.DataFrame(summary_data)
```

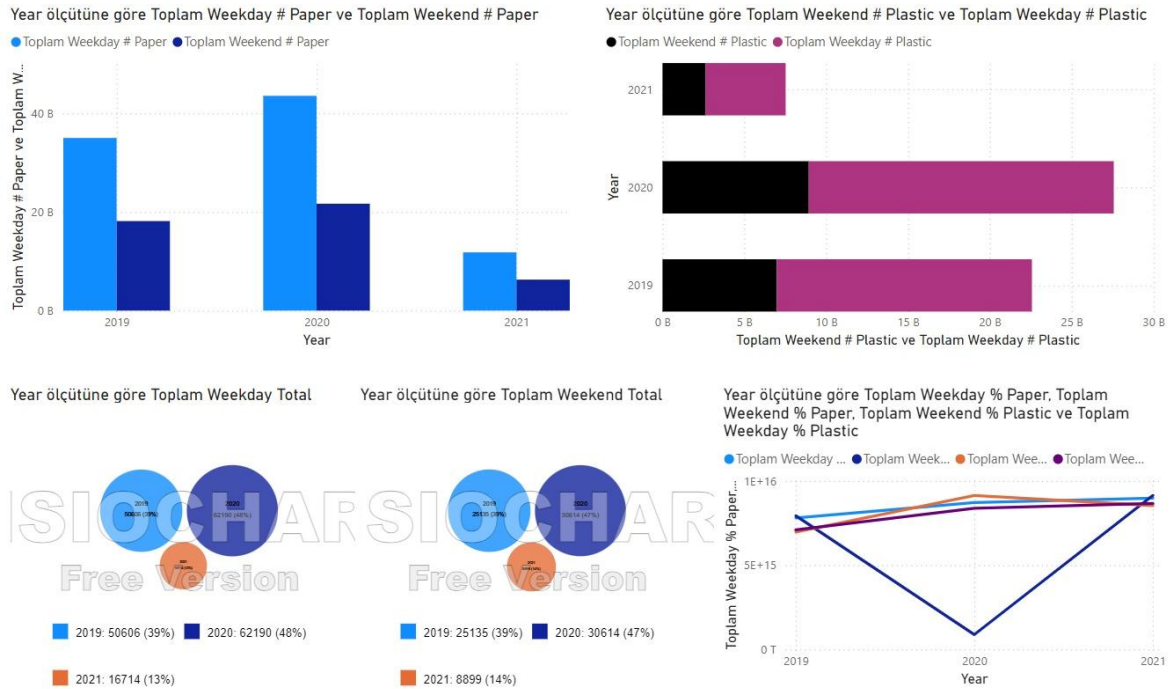
```
print(summary_df)
```

```
summary_df.to_csv("weekday_weekend_summary_2019_2021.csv", index=False)
```

Bu Python kodu, atık verilerini içeren bir CSV dosyasını okuyarak haftanın günlerine göre atık miktarlarını analiz eder. İlk olarak, veri seti yüklenir ve VisitDate sütunu datetime formatına dönüştürülerek yıl ve haftanın günü bilgileri eklenir. 2021 yılı verileri yalnızca 2021-03-24 tarihine kadar sınırlandırılır. Daha sonra, her yıl için hafta içi ve hafta sonu verileri filtrelenir. Haftanın günleri için toplam kağıt ve plastik kutu sayıları ile yüzdeleri hesaplanır. Sonuçlar, yıllara göre organize edilerek bir listeye eklenir ve ardından bu liste bir DataFrame'e dönüştürülür. Sonuçlar, weekday_weekend_summary_2019_2021.csv adlı yeni bir dosyaya kaydedilir. Bu analiz, atık yönetimi uygulamalarının etkinliğini değerlendirirken hafta içi ve hafta sonu atık toplama farklarını anlamaya yardımcı olur.

Year	Weekday # Paper	Weekday % Paper	Weekday # Plastic	Weekday % Plastic	Weekday Total	Weekend # Paper	Weekend % Paper	Weekend # Plastic	Weekend % Plastic	Weekend Total
2019	35010	0.78	15596	0.71	50606	18158	0.79	6977	0.7	25135
2020	43552	0.87	18638	0.84	62190	21679	0.88	8935	0.91	30614
2021	11807	0.9	4907	0.87	16714	6282	0.91	2617	0.85	8899

3.6.1 Veri setinin görselleştirilmesi.



Şekil4

Bu veri setinde elde edilen Şekil4 'deki panelde, 2019'dan 2021'e kadar olan yıllar için hafta içi ve hafta sonu olmak üzere iki farklı zaman diliminde kağıt ve plastik materyallerin miktarları ve yüzdeleri incelenmiştir. Veri seti, hem kağıt hem de plastik materyallerin yıl bazında hafta içi ve hafta sonu toplamalarını içermekte ve bu materyallerin oranlarını yüzdelik olarak göstermektedir. Görselleştirmeler, yıllık bazda hafta içi ve hafta sonu toplam kağıt ve plastik miktarlarının dağılımlarını daha iyi anlamamızı sağlamaktadır.

İlk grafik, 2019'dan 2021'e kadar olan yıllarda hafta içi ve hafta sonu kağıt materyal miktarlarının karşılaştırmasını sunmaktadır. 2019 ve 2020 yıllarında hafta içi günlerinde kağıt tüketiminin hafta sonuna göre oldukça yüksek olduğu görülmektedir. 2021 yılına gelindiğinde, her iki zaman diliminde de kağıt kullanımında belirgin bir düşüş yaşanmıştır. Bu azalma, özellikle 2021'deki düşük hafta içi kağıt kullanımı ile dikkat çekmektedir.

İkinci grafik, hafta içi ve hafta sonu plastik materyal miktarlarını karşılaştırmakta ve yıllara göre dağılımlarını göstermektedir. 2019 ve 2020 yıllarında hafta içi plastik kullanımı hafta sonuna göre daha fazlayken, 2021 yılında plastik kullanımında genel bir azalma yaşandığı gözlenmiştir. Ancak, hafta içi plastik tüketimi hafta sonuna kıyasla daha az bir düşüş göstermiştir. Bu da 2021'deki plastik tüketiminde belirgin bir değişim yaşandığını işaret etmektedir.

Üçüncü grafik, yıllara göre hafta içi ve hafta sonu toplam miktarların kağıt ve plastik olarak ayrıldığı bir dağılım sunmaktadır. Bu dağılım, plastik kullanımının yıllar içerisinde kağıt kullanımına kıyasla daha homojen bir şekilde azaldığını göstermektedir. Ayrıca, 2021 yılında hafta sonu plastik tüketiminin hafta içi plastik tüketiminden biraz daha azaldığı dikkat çekmektedir.

Dördüncü grafik, hafta içi toplamların yıllara göre dağılımını daire grafiği ile sunmaktadır. Bu grafikte, her bir yıl için toplam hafta içi kullanım yüzdeleri verilmektedir. 2020 yılında hafta içi toplam kullanım en yüksek seviyedeysen, 2021 yılında bu miktarda ciddi bir düşüş yaşandığı görülmektedir. Bu düşüş, hafta içi materyal tüketiminin azaldığını ve genel tüketim trendinin yıllar içinde değişim gösterdiğini ifade etmektedir.

Son grafik, hafta içi ve hafta sonu kağıt ve plastik yüzdelerinin yıllık bazda değişimlerini çizgi grafiği ile göstermektedir. 2019 ve 2020 yıllarında hem kağıt hem de plastik yüzdelerinde artış gözlemlenirken, 2021 yılında her iki materyalde de düşüş yaşandığı görülmektedir. Bu grafik, tüketim oranlarının zaman içinde nasıl değiştiğini ve hangi yıllarda daha fazla veya daha az kullanıldığını anlamamıza olanak tanımaktadır.

3.7 Oluşturulan “waste_data_update.csv” veri seti kullanılarak verilerin SPID değerlerindeki toplam atık yüzdesine göre kullanıcıya durum belirtme .

```
import pandas as pd

waste_data = pd.read_csv('waste_data_update.csv')

waste_data['VisitDate'] = pd.to_datetime(waste_data['VisitDate'])

spid_summary = waste_data.groupby('SPID').agg({

    'total_waste%': 'mean'

}).reset_index()
```

```
def classify_waste_level(average):

    if average >= 0.90:

        return "trash can should be added"

    elif average < 0.50:

        return "Trash bin should be reduced"

    else:

        return "Ideal"

spid_summary['Case'] = spid_summary['total_waste%'].apply(classify_waste_level)

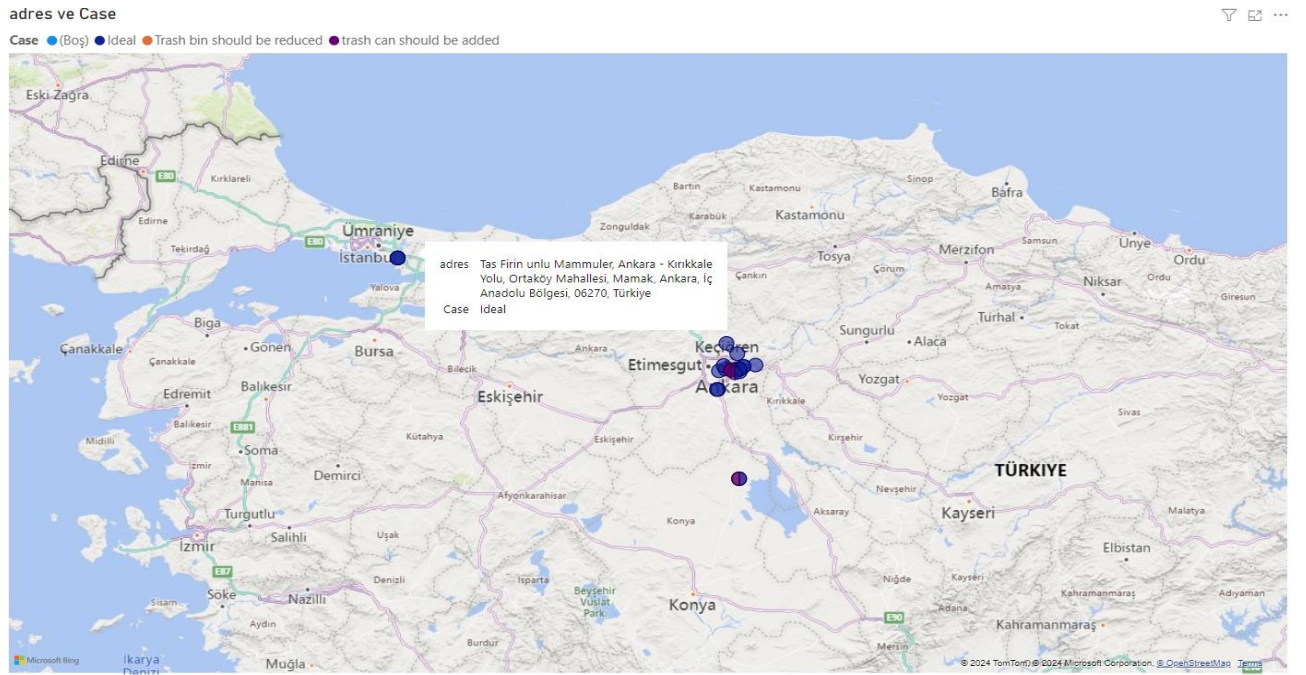
print(spid_summary)

spid_summary.to_csv("waste_bin_recommendations.csv", index=False)
```

Bu Python kodu, atık verilerini içeren bir CSV dosyasını okuyarak, her bir SPID (atık toplama noktası) için toplam atık yüzdesinin ortalamasını hesaplar. İlk olarak, veri seti yüklenir ve VisitDate sütunu datetime formatına dönüştürülür. Daha sonra, SPID bazında total_waste% sütunundaki değerler gruplandırılarak her SPID için ortalama değerler hesaplanır. Sonrasında, atık seviyesini sınıflandırmak için bir fonksiyon tanımlanır; bu fonksiyon, ortalama değere göre atık kutusu ekleme, azaltma veya ideal durumu belirtir. Sınıflandırma sonuçları yeni bir sütun olarak spid_summary DataFrame'ine eklenir. Sonuçlar konsola yazdırılır ve waste_bin_recommendations.csv adlı bir dosyaya kaydedilir. Bu analiz, atık yönetimi için önerilerde bulunarak, her bir atık toplama noktasındaki durumun değerlendirilmesine yardımcı olur.

SPID	Total_waste%	Case
7444	0.82	Ideal
9935	0.83	Ideal
9939	0.86	Ideal
9942	0.89	Ideal
9959	0.87	Ideal
9960	0.87	Ideal
10374	0.86	Ideal
10375	0.87	Ideal
10377	0.92	trash can should be added
10912	0.81	Ideal
10914	0.79	Ideal
12462	0.8	Ideal
12854	0.84	Ideal
14419	0.78	Ideal
15184	0.75	Ideal
15372	0.91	trash can should be added
15373	0.72	Ideal
15377	0.9	Ideal
15479	0.9	Ideal
15750	0.84	Ideal
15769	0.3	Trash bin should be reduced

3.7.1 Veri setinin görselleştirilmesi.



Şekil5

Bu harita, Türkiye'deki bazı bölgelerde çöp kutusu konumlarının durumunu analiz etmek amacıyla hazırlanmıştır. Harita üzerinde gösterilen noktalar, farklı konumlarda yer alan adresleri temsil ederken, her bir adres için belirlenmiş bir durum (case) bilgisi bulunmaktadır. Durumlar, ideal konumlar, çöp kutusunun azaltılması gereken yerler veya ek çöp kutusuna ihtiyaç duyulan alanlar olarak üç kategoriye ayrılmıştır. Harita üzerinde renklerle belirtilen bu durumlar, atık yönetimi ve kaynak optimizasyonu için önemli bilgiler sağlamaktadır.

Haritada yer alan her bir düğüm, belirli bir adres ve durum bilgisi içerir. Mavi renkli düğümler ideal konumları ifade ederken, turuncu renkli düğümler çöp kutusu sayısının azaltılması gereken yerleri göstermektedir. Kırmızı renkteki düğümler ise çöp kutusu eklenmesi gereken alanları belirtmektedir. Haritaya detaylı bir bakış, Ankara ve çevresinde yoğunlaşan konumların bulunduğunu gösteriyor. İstanbul ve çevresinde de belirli bölgelerde ideal çöp kutusu konumları bulunmaktadır. Bu görselleştirme, atık yönetimi ile ilgili kararlar alınırken hangi bölgelerde müdahaleye ihtiyaç duyulduğunu belirlemeye yardımcı olur.

3.8 Elimizde bulunan “location_data.csv” veri setindeki bilgiler (latitude ve longitude) kullanılarak adres bilgileri eklenen yeni veri seti oluşturuldu.

```
import pandas as pd

from geopy.geocoders import Nominatim

from geopy.extra.rate_limiter import RateLimiter

file_path = 'location_data.csv'

location_data = pd.read_csv(file_path)

geolocator = Nominatim(user_agent="geoapiExercises")

geocode = RateLimiter(geolocator.reverse, min_delay_seconds=1)

def get_address(lat, lon):

    try:

        location = geocode((lat, lon), language='tr')

        return location.address if location else "Adres bulunamadı"

    except:

        return "Hata"

location_data['adres'] = location_data.apply(lambda row: get_address(row['Latitude'],
row['Longitude']), axis=1)

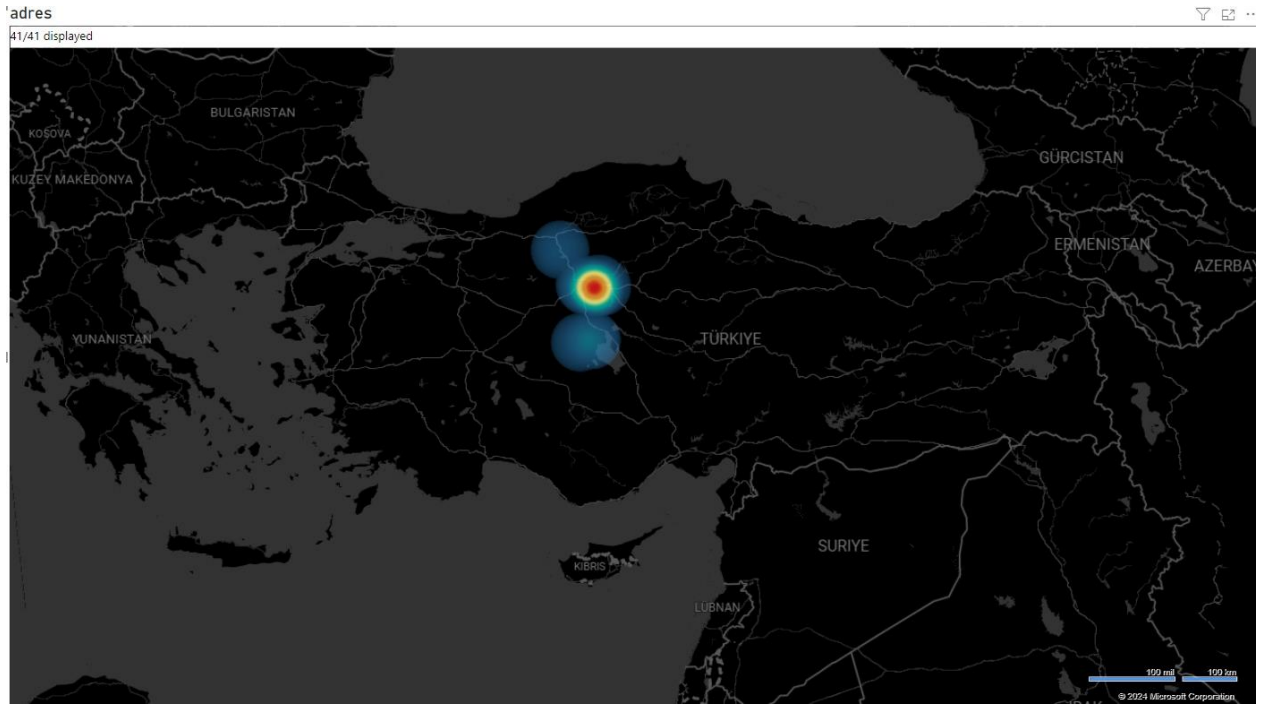
location_data.to_csv('location_data_with_address.csv', index=False)

print("Adres sütunu eklendi ve yeni dosya 'location_data_with_address.csv' olarak kaydedildi.")
```

Bu Python kodu, konum verilerini içeren bir CSV dosyasını okuyarak, her bir konumun coğrafi koordinatlarına (enlem ve boylam) dayanarak adres bilgilerini eklemeyi amaçlar. İlk olarak, dosya pandas kütüphanesi kullanılarak okunur. Ardından, geopy kütüphanesinin Nominatim servisi ile adresleri almak için bir geolocator tanımlanır ve hız sınırlaması uygulanarak aşırı istek gönderimi önlenir. Adres bilgilerini almak için get_address adında bir fonksiyon oluşturulur; bu fonksiyon, verilen enlem ve boylam bilgilerini kullanarak adresi bulur. Bulunan adres, location_data DataFrame'ine yeni bir sütun olarak eklenir. Son olarak, güncellenmiş veri seti location_data_with_address.csv dosyasına kaydedilir ve işlemin tamamlandığına dair bir mesaj konsola yazdırılır. Bu işlem, coğrafi veri setlerini daha anlamlı hale getirerek, analiz ve raporlama süreçlerinde kolaylık sağlar.

SPID	Latitude	Longitude	Adres
10222	39.93	32.86	Talatpaşa Blv., Ulucanlar, Hacettepe Mahallesi, Altındağ, Ankara, İç Anadolu Bölgesi, 06250, Türkiye
10374	39.92	32.99	Kayaş Mahallesi, Mamak, Ankara, İç Anadolu Bölgesi, 06270, Türkiye
10375	39.92	32.96	Tepecik Mahallesi, Mamak, Ankara, İç Anadolu Bölgesi, 06270, Türkiye
10377	39.92	32.98	Kayaş Caddesi, Kayaş, Kayaş Mahallesi, Mamak, Ankara, İç Anadolu Bölgesi, 06270, Türkiye
10912	39.99	33.11	TEM Tesisleri Misafir Otoparkı, Ankara - Kırıkkale yolu, Hasanoğlu Bahçelievler Mahallesi, Hasanoğlu, Kemalpaşa, Lalahan Mahallesi, Mamak, Ankara, İç Anadolu Bölgesi, 06852, Türkiye
10914	40.01	33.16	Şehitlik Mahallesi, Hasanoğlu, Çardakbağı Mahallesi, Elmadağ, Ankara, İç Anadolu Bölgesi, 06850, Türkiye
12462	39.99	33.09	Kavaklı Mahallesi, Altındağ, Ankara, İç Anadolu Bölgesi, Türkiye
12854	39.92	33.01	Taş Firin unlu Mammul, Ankara - Kırıkkale Yolu, Ortaköy Mahallesi, Mamak, Ankara, İç Anadolu Bölgesi, 06270, Türkiye

3.9 Oluşturulan location_data_with_address.csv veri setini kullanarak SPID id numaralarına göre ısı haritasında görselleştirdik.



Şekil6

Bu harita, atık kutularının adres bilgileri kullanılarak oluşturulmuş bir ısı haritasıdır. Haritada, Türkiye'nin belirli noktalarda yoğunlaşmış ısı alanları gözlemlenmektedir. Bu yoğunluk, atık kutularının bulunduğu konumları temsil etmekte olup, hangi bölgelerde daha fazla atık kutusunun yer aldığını ve hangi bölgelerde atık toplama gereksiniminin daha fazla olduğunu görsel olarak ortaya koymaktadır. Yoğun ısı noktaları, atık kutularının en fazla bulunduğu yerleri işaret ederken, daha az yoğun bölgeler, atık kutularının daha seyrek olduğu alanları göstermektedir. Bu tür bir analiz, atık yönetimi için etkili bir planlama aracı olarak kullanılabilir.