

## BÖLÜM 3. İBB'İN ŞEHİR TİYATROLARINDA SAHNELENEN OYUNLAR VERİ SETİNİN GÖRSELLEŞTİRİLMESİ

2017 ve 2020 yılları arasında İstanbul'da Şehir Tiyatrolarında Sahnelenen Oyunlar'ın sahne bilgisi, tiyatro oyun ismi, kategorilerine göre seyirci sayıları veri hazırlama aşamalarından geçirilerek daha anlaşılabilir sonuçlara ulaşmak için yapılan veri görselleştirme çalışmasıdır.

### 3.1. Kullanılacak Olan Veriyi Anlama ve Yükleme Aşaması

İBBaçık veri portalından “Şehir Tiyatrolarında Sahnelenen Oyunlar” veri seti kullanılmıştır.[7]

- PLAY\_DATE (Oyun Tarihi)
- PLAY\_NAME (Oyun İsmi)
- THEATER\_NAME (Sahne İsmi)
- PLAY\_CATEGORY (Oyun Kategorisi)
- IS\_GUEST\_PLAY (Misafir Oyun mu?)
- PLAY\_TYPE (Oyun Tipi)
- LONGITUDE (Boylam)
- LATITUDE (Enlem)
- NUMBER\_OF\_AUDIENCE (Seyirci Sayısı)

```
#veriyi içe aktaralım
data=pd.read_csv('theater_play.csv')
data.head()
```

	PLAY_DATE	PLAY_NAME	THEATER_NAME	PLAY_CATEGORY	IS_GUEST_PLAY	PLAY_TYPE	LONGITUDE	LATITUDE	NUMBER_OF_AUDIENCE
0	2017-12-27	KARAGÖZ'ÜN UYKUSU	Gaziosmanpaşa Feriğ Egemen Çocuk Tiyatrosu Sah...	Çocuk	0	Yerli	28.916482	41.057846	50
1	2019-03-28	ŞAHANE ZÜĞÜRTLER	Kadıköy Haldun Taner Sahnesi	Yetişkin	0	Yabancı	29.023408	40.991770	1
2	2017-11-30	BAK BİZİM ŞARKIMIZI ÇALIYORLAR	Kağıthane Sadabad Sahnesi	Müzikal	0	Yabancı	28.970972	41.079366	84
3	2020-01-24	ÇİN SABAHTA	Gaziosmanpaşa Sahnesi	Yetişkin	0	Yerli	28.916482	41.057846	46
4	2019-01-27	POLLYANNA	Kağıthane Sadabad Sahnesi	Çocuk	0	Yabancı	28.970972	41.079366	150

Şekil 3. 1 Veri Seti İçeriği

head fonksiyonu ile veri setimizin ilk beş satırını görüntüledikten sonra tail fonksiyonu ile de son beş satırını aşağıdaki gibi görüntüledim.tail fonksiyonu ile veri setimizde kaç satır olduğu çıkarımı da yapılmaktadır.

```
data.tail()
```

	PLAY_DATE	PLAY_NAME	THEATER_NAME	PLAY_CATEGORY	IS_GUEST_PLAY	PLAY_TYPE	LONGITUDE	LATITUDE	NUMBER_OF_AUDIENCE
27293	2019-12-22	ELMA KURDU KIRTIK	Gaziosmanpaşa Ferih Egemen Çocuk Tiyatrosu Sah...	Çocuk	0	Yerli	28.916482	41.057846	18
27294	2019-12-03	BİSKÜVİ ADAM	Fatih Reşat Nuri Sahnesi	Çocuk	0	Yabancı	28.956194	41.015900	34
27295	2017-10-15	HİSSE-İ ŞAYIA (BİR EVLİLİK KOMEDİSİ)	Fatih Reşat Nuri Sahnesi	Yetişkin	0	Yerli	28.956194	41.015900	17
27296	2019-12-09	UZLAŞMA	Kadıköy Haldun Taner Sahnesi	Yetişkin	0	Yabancı	29.023408	40.991770	20
27297	2018-03-02	YANGIN YERİNDE ORKİDELER	Kadıköy Haldun Taner Sahnesi	Yetişkin	0	Yerli	29.023408	40.991770	2

Şekil 3. 2 Veri Seti Son Satırlar İçeriği

Veri setimizde ilk beş ve son beş veri seti içeriğini görüntüledikten sonra shape fonksiyonu ile satır ve sütun sayısı bilgisine ulaşılmıştır.27298 satır ve 9 sütun bulunmaktadır.

```
data.shape
```

```
(27298, 9)
```

Şekil 3. 3 Satır ve Sütun Sayısı

```
data.columns
```

```
Index(['PLAY_DATE', 'PLAY_NAME', 'THEATER_NAME', 'PLAY_CATEGORY',  
      'IS_GUEST_PLAY', 'PLAY_TYPE', 'LONGITUDE', 'LATITUDE',  
      'NUMBER_OF_AUDIENCE'],  
      dtype='object')
```

Şekil 3. 4 Veri Setindeki Sütunlar

```
types=data.dtypes
print(types)
```

PLAY_DATE	object
PLAY_NAME	object
THEATER_NAME	object
PLAY_CATEGORY	object
IS_GUEST_PLAY	object
PLAY_TYPE	object
LONGITUDE	float64
LATITUDE	float64
NUMBER_OF_AUDIENCE	int64
dtype:	object

Şekil 3. 5 Veri Setindeki Sütunların Veri Tipleri Bilgisi

Şehir Tiyatrolarında Sahnelenen Oyunlar Veri Setini incelemek için describe fonksiyonu kullanarak, veri setimizin sayısal sütunlarının istatistiksel bilgilerini özet tablosu olarak oluşturuldu.

```
data.describe()
```

	LONGITUDE	LATITUDE	NUMBER_OF_AUDIENCE
count	27298.000000	27298.000000	27298.000000
mean	28.986390	41.040191	66.653601
std	0.060749	0.030050	167.922210
min	28.876578	40.991770	1.000000
25%	28.956194	41.017064	11.000000
50%	28.988829	41.025025	32.000000
75%	29.023408	41.057846	67.000000
max	29.104798	41.097745	3588.000000

Şekil 3. 6 Veri Seti Sayısal Sütunlar Özeti

nunique fonksiyonu kullanılarak veri seti içeriğinde bulunan her bilginin kaç farklı değer aldığı bilgisine ulaşılmıştır.

```
data.nunique()
PLAY_DATE      770
PLAY_NAME      350
THEATER_NAME    77
PLAY_CATEGORY   3
IS_GUEST_PLAY   2
PLAY_TYPE       2
LONGITUDE      10
LATITUDE        10
NUMBER_OF_AUDIENCE 791
dtype: int64
```

Şekil 3. 7 Veri Setindeki Sütunların Benzersiz Değer Sayıları

Sütunların benzersiz değer sayılarını `nunique` fonksiyonu ile görüntüledikten sonra az değere sahip olan sütunların benzersiz değerleri `unique` fonksiyonu kullanarak görüntülendi.

```
data['PLAY_CATEGORY'].unique()
array(['Çocuk', 'Yetişkin', 'Müzikal'], dtype=object)

data['PLAY_TYPE'].unique()
array(['Yerli', 'Yabancı'], dtype=object)

data['LONGITUDE'].unique()
array([28.916482 , 29.023408 , 28.970972 , 29.023817 , 28.988829 ,
       29.0129   , 28.8765777, 28.956194 , 29.104798 , 28.99044  ])

data['LATITUDE'].unique()
array([41.057846 , 40.99177   , 41.079366 , 41.017064 , 41.046788 ,
       41.019561 , 41.09774468, 41.0159   , 41.025025 , 41.045777  ])
```

Şekil 3. 8 Bazı Sütunların Benzersiz Değerleri

### 3.2. Veri Seti İçeriğinde Değişiklik Yapma Aşaması

Bu aşamada IS\_GUEST\_PLAY sütununun aldığı değerler değiştirilip daha anlaşılır hale getirilecektir.

```
data['IS_GUEST_PLAY'].unique()
array([0, 1], dtype=int64)
```

Şekil 3. 9 IS\_GUEST\_PLAY Sütununun Aldığı Değerler

Bu sütun değerlerinde 1 misafir oyun anlamında kullanılırken 0 misafir oyun değil anlamında kullanılmaktadır. Anlaşılır olması için 1 ve 0 yerine anlamlarını getirerek aşağıdaki şekilde değişiklik yapılmıştır.

```
data=data.replace({'IS_GUEST_PLAY': {1: 'Misafir', 0 : 'Misafir değil'}})
data['IS_GUEST_PLAY'].unique()
array(['Misafir değil', 'Misafir'], dtype=object)
```

Şekil 3. 10 IS\_GUEST\_PLAY Sütununun Değerlerinin Değiştirilmesi

### 3.3. Kullanılacak Verinin Temizleme Aşaması

Veri setimizde boş veri ya da eksik veri kontrolü yapıldıktan sonra eğer boş ya da eksik veri var ise gerekli işlemler yapılacaktır.

```
data.isnull().sum()
PLAY_DATE      0
PLAY_NAME      0
THEATER_NAME    1
PLAY_CATEGORY  0
IS_GUEST_PLAY  0
PLAY_TYPE      0
LONGITUDE      0
LATITUDE       0
NUMBER_OF_AUDIENCE 0
dtype: int64
```

Şekil 3. 11 Veri Setimizin Boş Veri Bilgisi

Yukarıda görüldüğü gibi THEATER\_NAME sütununda 1 adet boş verimiz bulunmaktadır. Boş verimizin hangi satırda olduğuna şu şekilde bakılmıştır:

```
data[data['THEATER_NAME'].isna()]
```

	PLAY_DATE	PLAY_NAME	THEATER_NAME	PLAY_CATEGORY	IS_GUEST_PLAY	PLAY_TYPE	LONGITUDE	LATITUDE	NUMBER_OF_AUDIENCE
3500	2018-05-18	LE TARTUFFE "TARTUF" (ÖZDEMİR SABANCI EMİRGA...	NaN	Yetişkin	Misafir	Yabancı	28.970972	41.079366	4

Şekil 3. 12 Boş Verinin Bulunduğu Satır

Boş veri yerine gelmesi gereken değer ile doldurulacaktır. Daha sonra tekrardan boş veri kontrolü yapılmıştır.

```
data['THEATER_NAME'].fillna('Özdemir Sabancı Emirgan Anadolu Lisesi',inplace=True)
```

```
data.isnull().sum()
```

```
PLAY_DATE      0
PLAY_NAME      0
THEATER_NAME    0
PLAY_CATEGORY  0
IS_GUEST_PLAY  0
PLAY_TYPE      0
LONGITUDE      0
LATITUDE       0
NUMBER_OF_AUDIENCE 0
dtype: int64
```

Şekil 3. 13 Boş Verinin Doldurulması ve Son Boş Veri Bilgisi

### 3.4. Veri Setini Görselleştirme Aşaması

İlk olarak en çok oynanan beş tiyatroyu ve kaç kez oynandıklarını görüntüledikten sonra bu oyunları liste halinde görüntülenmiştir. Daha sonra da bu verilerimizi sütun grafiği haline getirerek veri görselleştirmesi yapıldı.

```
data['PLAY_NAME'].value_counts()[0:5]
```

```
KARINCALAR - BİR SAVAŞ VARDI    1025
NORA - BİR BEBEK EVİ            884
KAHVEDE ŞENLİK VAR              846
ŞAHANE ZÜĞÜRTLER                763
MATRUŞKA                        735
Name: PLAY_NAME, dtype: int64
```

Şekil 3. 14 En Çok Oynanan Beş Tiyatro Bilgisi

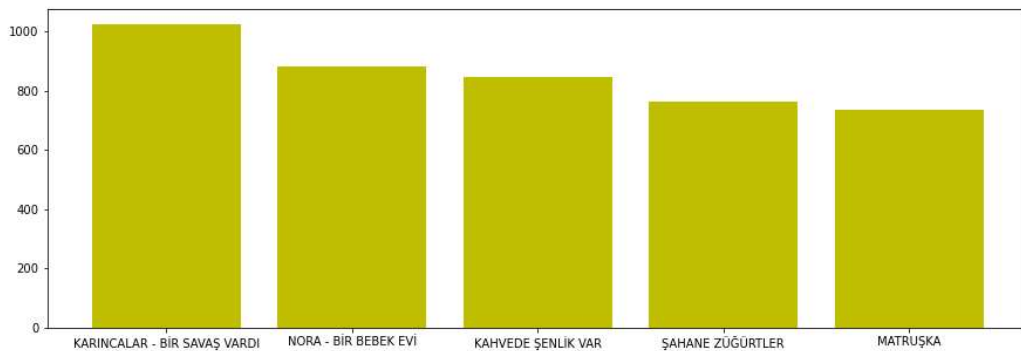
```
list(data['PLAY_NAME'].value_counts()[0:5].keys())
```

```
['KARINCALAR - BİR SAVAŞ VARDI',
 'NORA - BİR BEBEK EVİ',
 'KAHVEDE ŞENLİK VAR',
 'ŞAHANE ZÜĞÜRTLER',
 'MATRUŞKA']
```

Şekil 3. 15 En Çok Oynanan Beş Tiyatro Listesi

```
from matplotlib import pyplot as plt
```

```
plt.figure(figsize=(15,5))
plt.bar(list(data['PLAY_NAME'].value_counts()[0:5].keys()), list(data['PLAY_NAME'].value_counts()[0:5]), color="y")
plt.show()
```



Şekil 3. 16 En Çok Oynanan İlk Beş Tiyatro Sütun Grafiği



En çok tiyatronun sahnelendiği beş tiyatro sahnesini ve bu sahnelerde kaç kez tiyatro sahnelendiğini Şekil 3.17 'de görüntüledikten sonra Şekil 3.18'de bu sahneleri liste halinde görüntülenmiştir. Daha sonra da bu verilerimizi Şekil 3.19'da sütun grafiği haline getirerek veri görselleştirmesi yapıldı.

```
data['THEATER_NAME'].value_counts()[0:5]
```

```
Kağıthane Sadabad Sahnesi      3553
Fatih Reşat Nuri Sahnesi       3364
Gaziosmanpaşa Sahnesi         3347
Ümraniye Sahnesi              3260
Harbiye Muhsin Ertuğrul Sahnesi 3200
Name: THEATER_NAME, dtype: int64
```

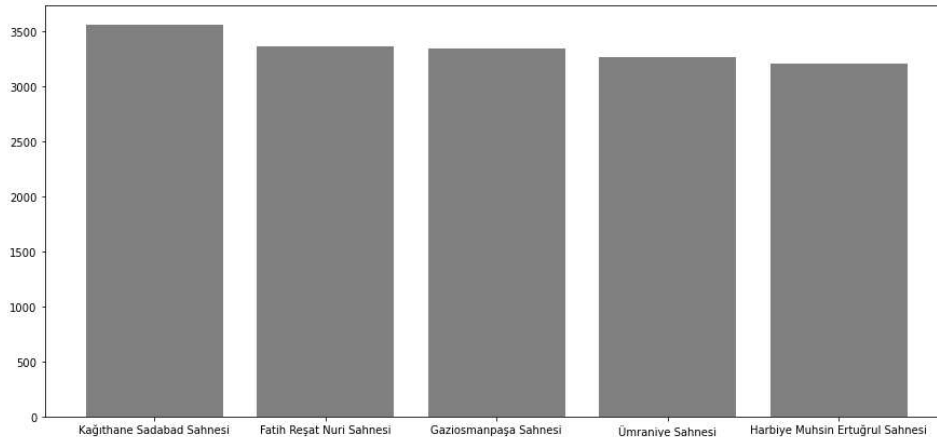
Şekil 3. 17 Tiyatronun En Fazla Oynandığı Beş Sahne

```
list(data['THEATER_NAME'].value_counts()[0:5].keys())
```

```
['Kağıthane Sadabad Sahnesi',
 'Fatih Reşat Nuri Sahnesi',
 'Gaziosmanpaşa Sahnesi',
 'Ümraniye Sahnesi',
 'Harbiye Muhsin Ertuğrul Sahnesi']
```

Şekil 3. 18 Tiyatronun En Fazla Oynandığı Beş Sahne Listesi

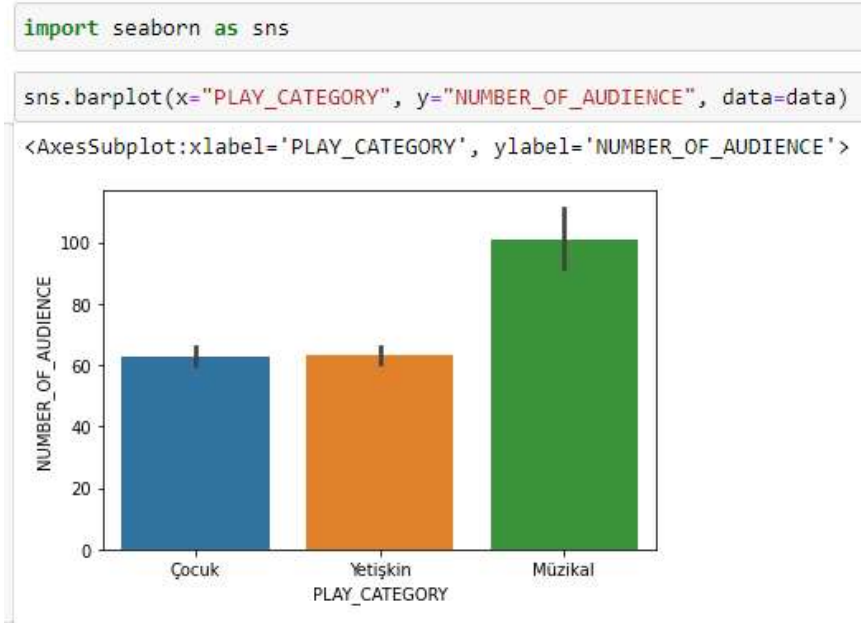
```
plt.figure(figsize=(15,7))
plt.bar(list(data['THEATER_NAME'].value_counts()[0:5].keys()), list(data['THEATER_NAME'].value_counts()[0:5]), color="gray")
plt.show()
```



Şekil 3. 19 Tiyatronun En Fazla Oynandığı Beş Sahne Sütun Grafiği

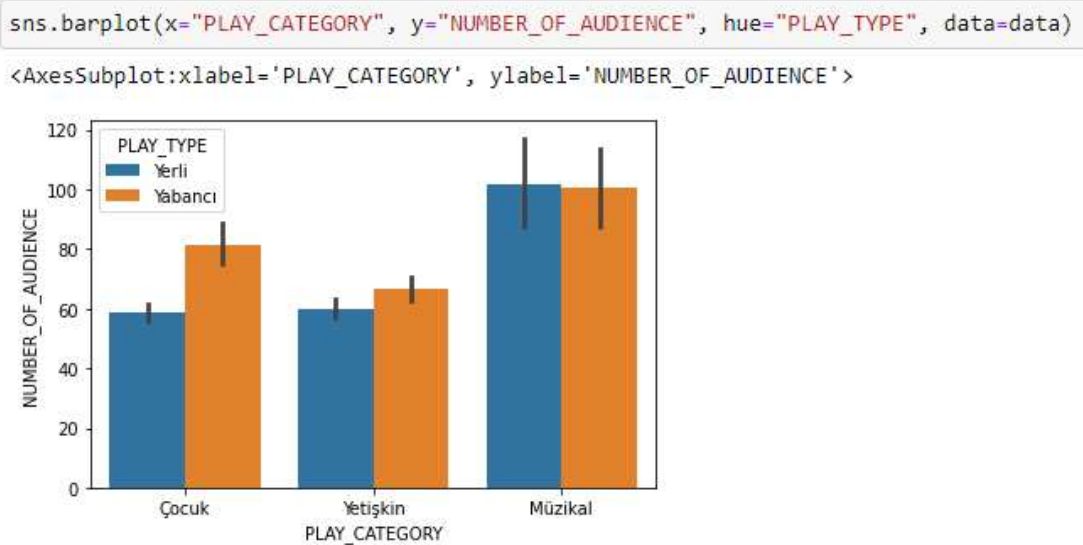


Oyun kategorilerinde seyirci sayıları için Şekil 3.20’de veri görselleştirilmesi yapıp sütun grafiği oluşturulmuştur. Oyun kategorilerinde en çok seyirciye sahip kategorinin müzikal olduğu görülmektedir.



Şekil 3. 20 Oyun Kategorisi Sütun Grafiği

Oyun kategorileri sütun grafiğinde oyun tipinde belirtilmiş hali:

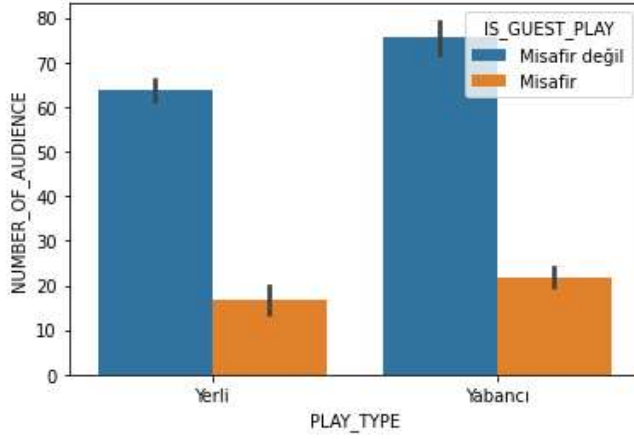


Şekil 3. 21 Oyun Kategorisi Sütun Grafiği

Oyun tipinde seyirci sayıları için misafir oyun ayrıntısıyla Şekil 3.22’de veri görselleştirmesi yapıp sütun grafiği oluşturulmuştur.

```
sns.barplot(x="PLAY_TYPE", y="NUMBER_OF_AUDIENCE", hue="IS_GUEST_PLAY", data=data)
```

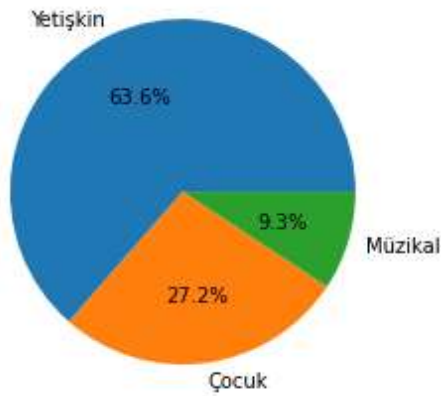
```
<AxesSubplot:xlabel='PLAY_TYPE', ylabel='NUMBER_OF_AUDIENCE'>
```



Şekil 3. 22 Oyun Tipi Sütun Grafiği

Oyun kategorilerinin veri görselleştirilmesi yapılarak Şekil 3.23’de pasta grafiği oluşturulmuştur. Oyun kategorilerinden en fazla olan kategorinin yetişkin olduğu görülmektedir.

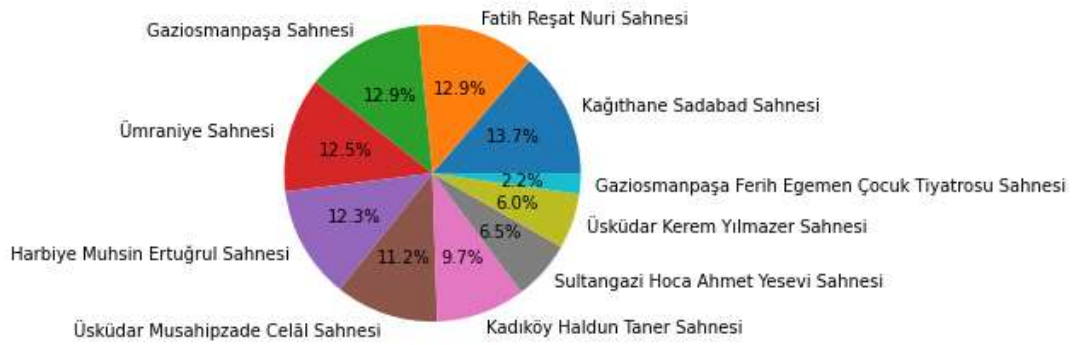
```
data.PLAY_CATEGORY.value_counts().plot(kind="pie", autopct="%0.1f%%")  
plt.ylabel(" ");
```



Şekil 3. 23 Oyun Kategorisi Pasta Grafiği

En çok tiyatronun oynandığı 10 sahnenin veri görselleştirmesi yapılarak Şekil 3.24’de pasta grafiği oluşturulmuştur.En çok tiyatronun oynandığı sahnenin Kağıthane Sadabad Sahnesi olduğu görülmektedir.

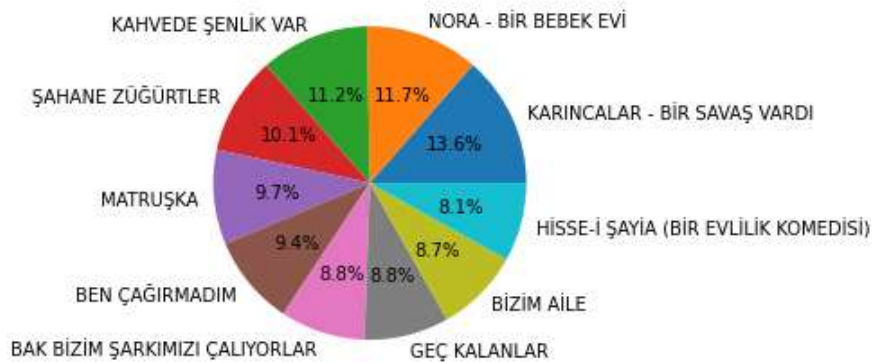
```
data.THEATER_NAME.value_counts()[0:10].plot(kind="pie", autopct="%0.1f%%")
plt.ylabel(" ");
```



Şekil 3. 24 Tiyatro Sahnelerinin Pasta Grafiği

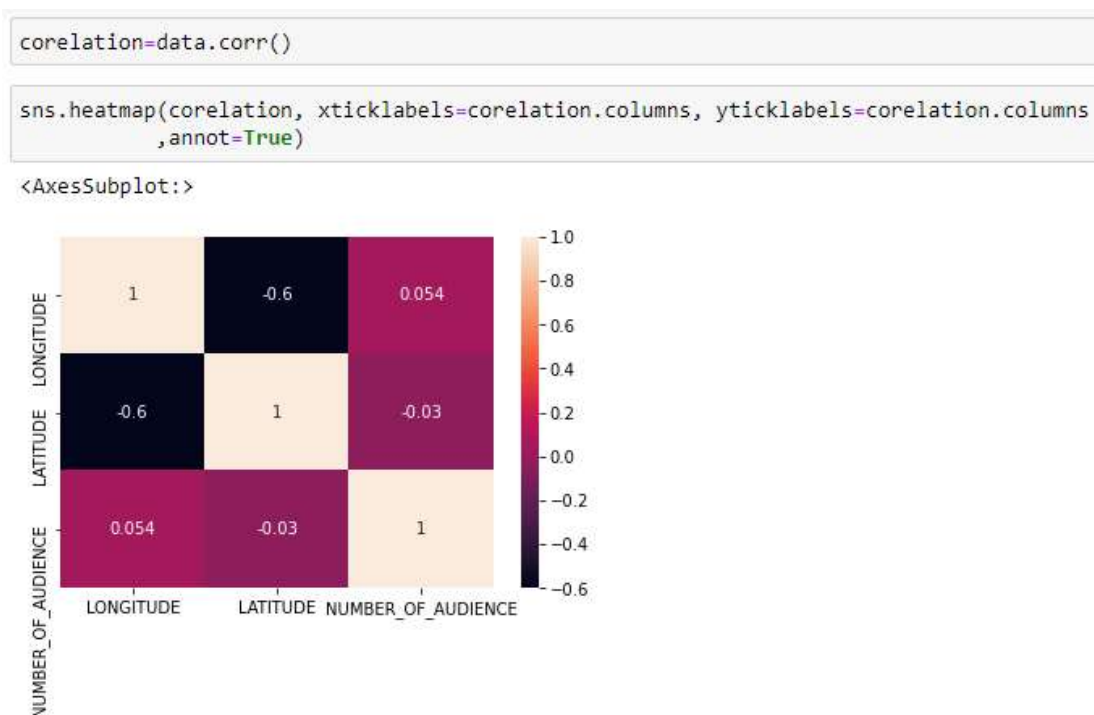
En çok izlenen 10 tiyatronun veri görselleştirilmesi yapılarak Şekil 3.25’de pasta grafiği oluşturulmuştur.En çok izlenen tiyatronun Karıncalar – Bir Savaş Vardı olduğu görülmektedir.

```
data.PLAY_NAME.value_counts()[0:10].plot(kind="pie", autopct="%0.1f%%")
plt.ylabel(" ");
```



Şekil 3. 25 Tiyatroların Pasta Grafiği

Isı haritaları değişkenleri satırlara ve sütunlara yerleştirip, tablodaki hücreleri renklendirerek çok değişkenli verileri çapraz incelemek amacıyla kullanılmaktadır. Isı haritaları, birden çok değişken arasında varyans göstermek, herhangi bir değişkenin birbirine benzer olup olmadığını göstermek ve aralarında herhangi bir korelasyon olup olmadığını tespit etmek amacıyla da kullanılabilir. Şekil 3.26'da verilerimizin ısı grafiği bulunmaktadır.



Şekil 3. 26 Isı Haritası

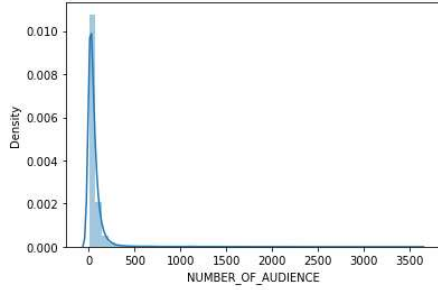
Distplot, verilerin tek değişkenli dağılımını, yani bir değişkenin yoğunluk dağılımına karşı veri dağılımını temsil eder.

Şekil 3.27'de seyirci sayısının histogram grafiği oluşturulmuştur. 0-500 seyirci sayısı aralığında bir yoğunluk görülmektedir.

```
sns.distplot(data['NUMBER_OF_AUDIENCE'])
```

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).  
warnings.warn(msg, FutureWarning)

```
<AxesSubplot:xlabel='NUMBER_OF_AUDIENCE', ylabel='Density'>
```

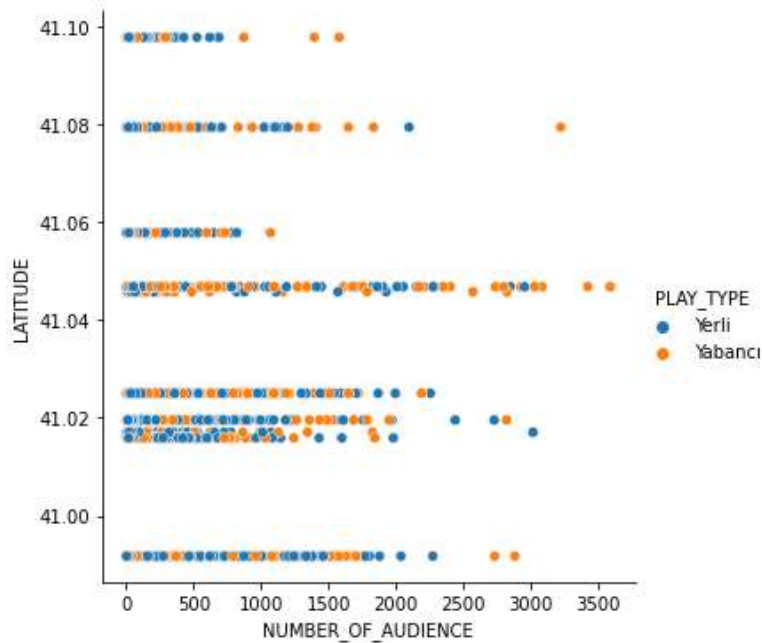


Şekil 3. 27 Seyirci Sayısı Histogram Grafiği

Tiyatro oyun tipi enlem ve seyirci sayısı değişkenleriyle Şekil 3.28’de dağılım grafiği oluşturulmuştur.

```
sns.relplot(x='NUMBER_OF_AUDIENCE', y='LATITUDE', hue='PLAY_TYPE', data=data)
```

```
<seaborn.axisgrid.FacetGrid at 0x11cec4f4580>
```



Şekil 3. 28 Dağılım Grafiği

Şekil 3.30’da bulunan 3 boyutlu görselleştirme Şekil 29’da bulunan kod ile yapılmıştır.

```
import plotly

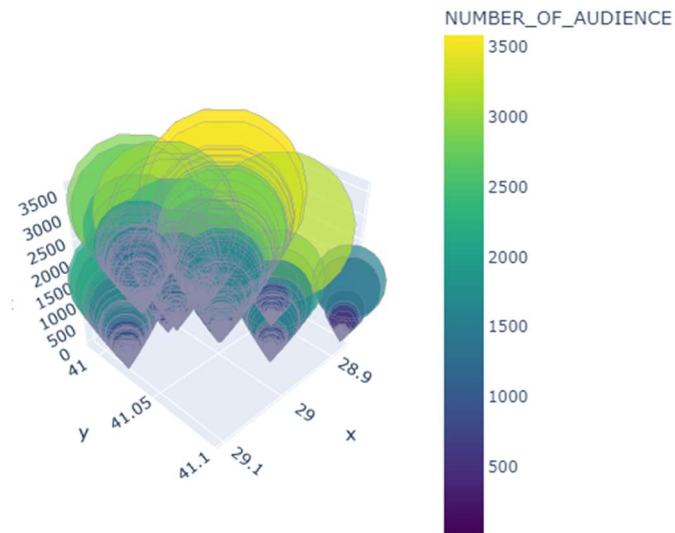
import plotly.graph_objects as go

fig = go.Figure(data=go.Scatter3d(
    x=data['LONGITUDE'],
    y=data['LATITUDE'],
    z=data['NUMBER_OF_AUDIENCE'],
    text=data['THEATER_NAME'],
    mode='markers',
    marker=dict(
        sizemode='diameter',
        sizeref=25,
        size=data['NUMBER_OF_AUDIENCE'],
        color = data['NUMBER_OF_AUDIENCE'],
        colorscale = 'Viridis',
        colorbar_title = 'NUMBER_OF_AUDIENCE',
        line_color='rgb(140,140, 170)'
    )
))

fig.update_layout(height=600, width=600,
                  title=' ')

fig.show()
```

Şekil 3. 29 3 Boyutlu Görselleştirme Kodu



Şekil 3. 30 3 Boyutlu Görselleştirme



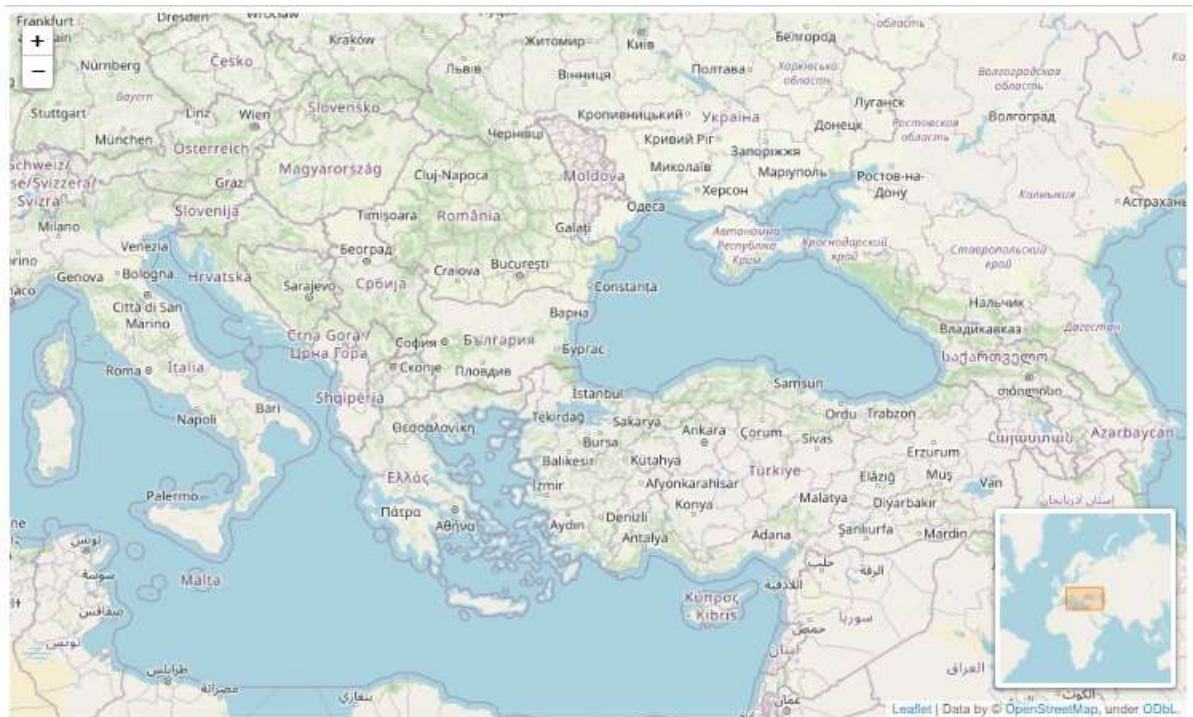
Minimap ile haritalama yapmak için MiniMap() foksiyonu kullanılarak Şekil 3.31 'de yapılan kodlama ile Şekil 3.32'de görülen harita oluşturulmuştur.

```
import folium

from folium import plugins
m= folium.Map(location=(42, 29), zoom_start=5)

minimap = plugins.Minimap()
m.add_child(minimap)
m
```

Şekil 3. 31 Haritalama Kodu



Şekil 3. 32 Minimap ile Haritalama

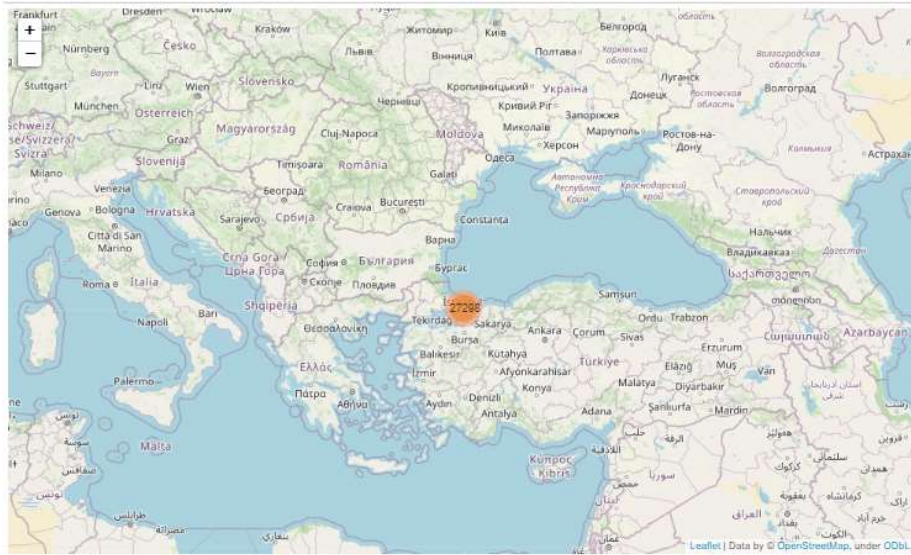


Plugins içerisinde bulunan MarkerCluster fonksiyonu kullanılarak parametre olarak konum bilgileri verilmiştir.Şekil 3.33’de MarkerCluster fonksiyonu kullanılarak yapılan kodlama ve oluşan harita çıktısı görülmektedir.

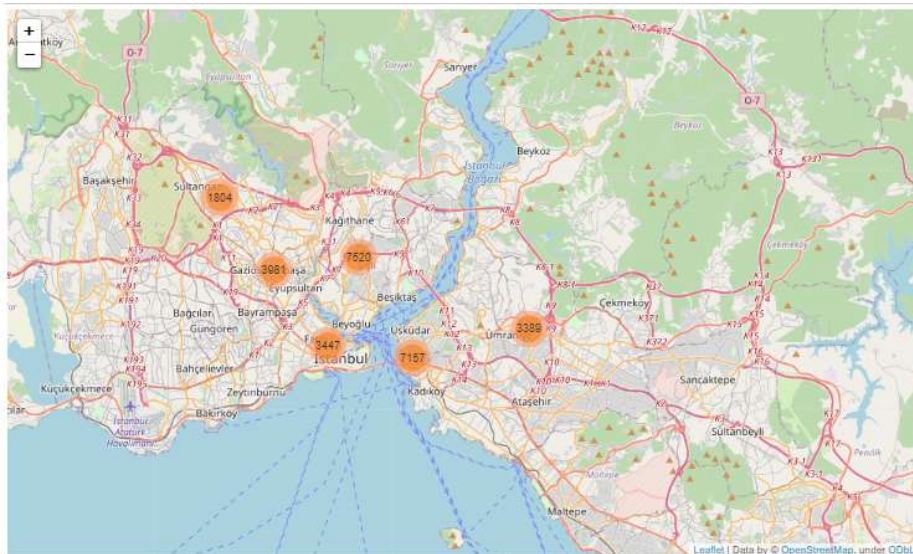
```
m = folium.Map([42, 29], zoom_start=5, width="%100", height="%100")
location=data[["LATITUDE", "LONGITUDE"]]

plugins.MarkerCluster(location).add_to(m)

m
```



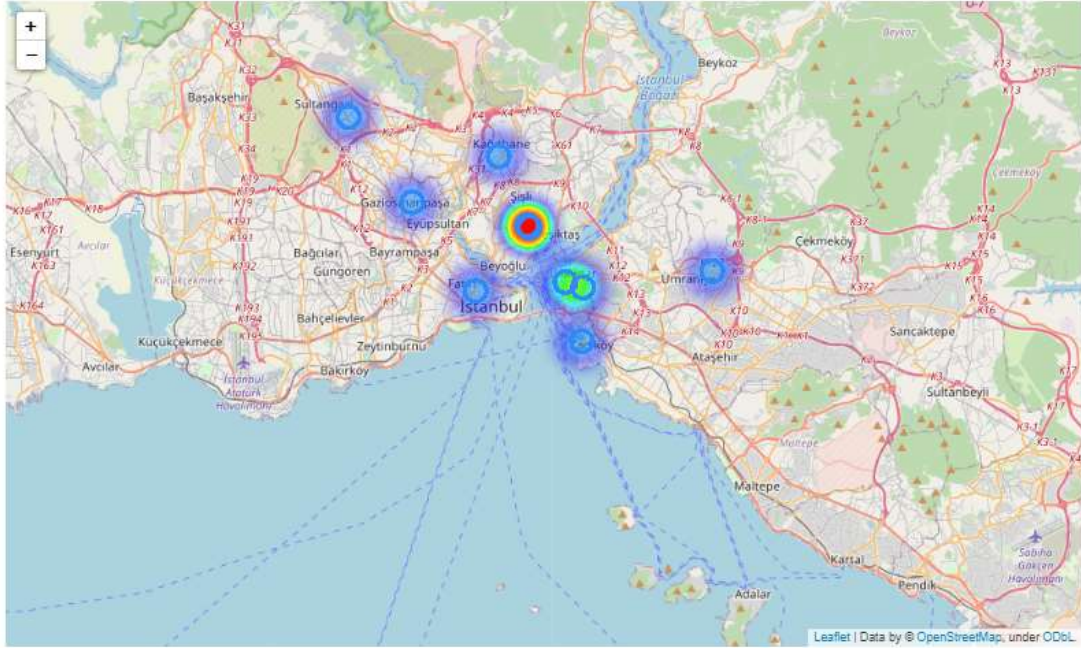
Şekil 3.33 MarkerCluster İle Görseleştirme



Şekil 3.34 MarkerCluster Zoom Yapılmış Hali



Şekil 3.37’de ısı haritası oluşturulmuştur.



Şekil 3. 37 Isı Haritası

```
heat_data = data.groupby(["LATITUDE", "LONGITUDE"])['NUMBER_OF_AUDIENCE'].mean().reset_index().values.tolist()
folium.plugins.HeatMap(heat_data).add_to(m)
m
```