# Istanbul AirBnb Keşifçi Veri Analizi

Name Surname – Student ID

İçindekiler Tablosu

[Istanbul AirBnb Keşifçi Veri Analizi](#_heading=h.gjdgxs) **1**

[**1. Amaç ve Kapsam**](#_heading=h.u9hmdsaolagg) **1**

[**2. Verilerin Temizlenmesi**](#_heading=h.3znysh7) **3**

[**2.1 Nan Değerlerin Temizlenmesi ve Tamamlanması**](#_heading=h.2et92p0) **5**

[**2.2 Temel İstatistik Bilgilerin İncelenmesi**](#_heading=h.tyjcwt) **8**

[**3. Verilerin Analiz Edilmesi**](#_heading=h.3dy6vkm) **11**

[**3.1 En çok hangi semtte AirBnb bulunmaktadır?**](#_heading=h.1t3h5sf) **11**

[**3.2 En çok hangi tür AirBnb'ler bulunmaktadır?**](#_heading=h.4d34og8) **13**

[**3.3 Semtlere göre AirBnB tür dağılımı nasıldır?**](#_heading=h.2s8eyo1) **14**

[**3.4 Gecelik ortalama fiyatlar semtlere göre nasıl değişiklik göstermektedir?**](#_heading=h.17dp8vu) **16**

[3.5 Türlerine göre AirBnb'lerin ortalama gecelik getirisi nasıl değişmektedir?](#_heading=h.3rdcrjn) 17

[**3.6 AirBnb türlerine göre gecelik fiyatlar semtlere göre nasıl değişmektedir?**](#_heading=h.26in1rg) **21**

[**3.7 AirBnb türlerine göre konaklanan minimum gece sayısı nasıl değişmektedir?**](#_heading=h.lnxbz9) **22**

[3.8 Birden fazla Airbnb ilanına sahip hostlar var mı?](#_heading=h.15g4k73572dd) 25

[3.9 İlanlarda en çok hangi kelimeler kullanılıyor?](#_heading=h.ifuawj5j4ycy) 26

[**4. Sonuç ve Öneriler**](#_heading=h.35nkun2) **27**

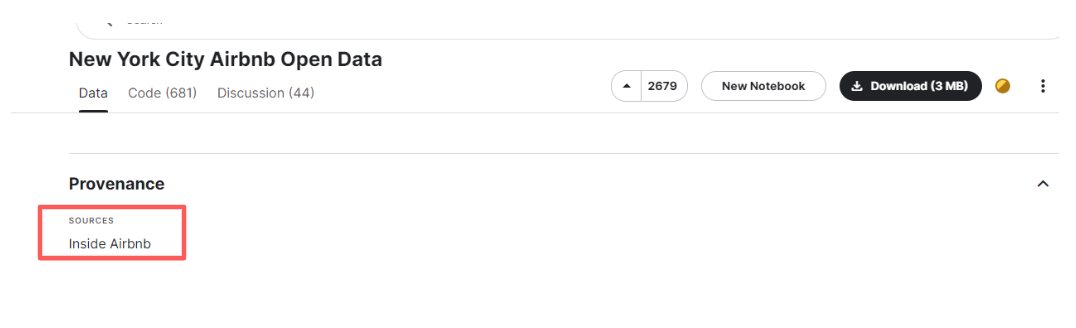
## **1. Amaç ve Kapsam**

Bu projenin amacı İstanbul'da yatırım amaçlı bir gayrimenkul edinmek isteyen biri gözünden İstanbul'daki AirBnb'lere ait verileri incelemektir. Bu kapsamda aşağıdaki sorulara cevap aranacaktır:

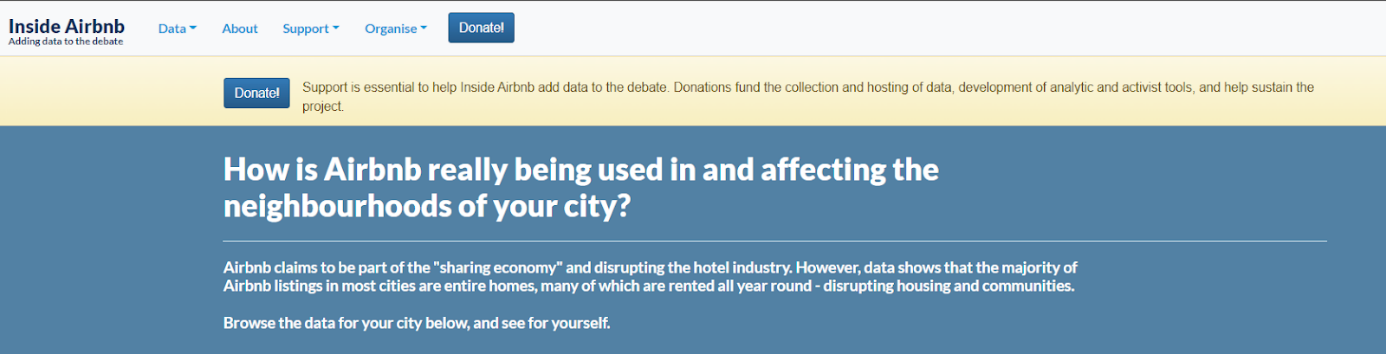
* 1. En çok hangi semtte AirBnb bulunmaktadır?
  2. En çok hangi tür AirBnb'ler bulunmaktadır?
  3. Semtlere göre AirBnB tür dağılımı nasıldır?
  4. Gecelik ortalama fiyatlar semtlere göre nasıl değişiklik göstermektedir?
  5. Gecelik ortalama fiyatlar semtlere göre nasıl değişiklik göstermektedir?
  6. AirBnb türlerine göre gecelik fiyatlar semtlere göre nasıl değişmektedir?
  7. AirBnb türlerine göre konaklanan gece sayısı nasıl değişmektedir?
  8. Birden fazla Airbnb ilanına sahip hostlar var mı?
  9. İlanlarda en çok hangi kelimeler kullanılıyor?

Problemin ve araştırma sorularının belirlenmesinde sınıfta incelenen NYC Airbnb veri seti üzerinde yapılan analizler ve konuşmalardan esinlenilmiştir.

İlk olarak Kaggle sitesindeki NYC Airbnb verinin metadası incelenerek veri setinin ana kaynağı olan "Inside Airbnb" sayfasına ulaşılmıştır.



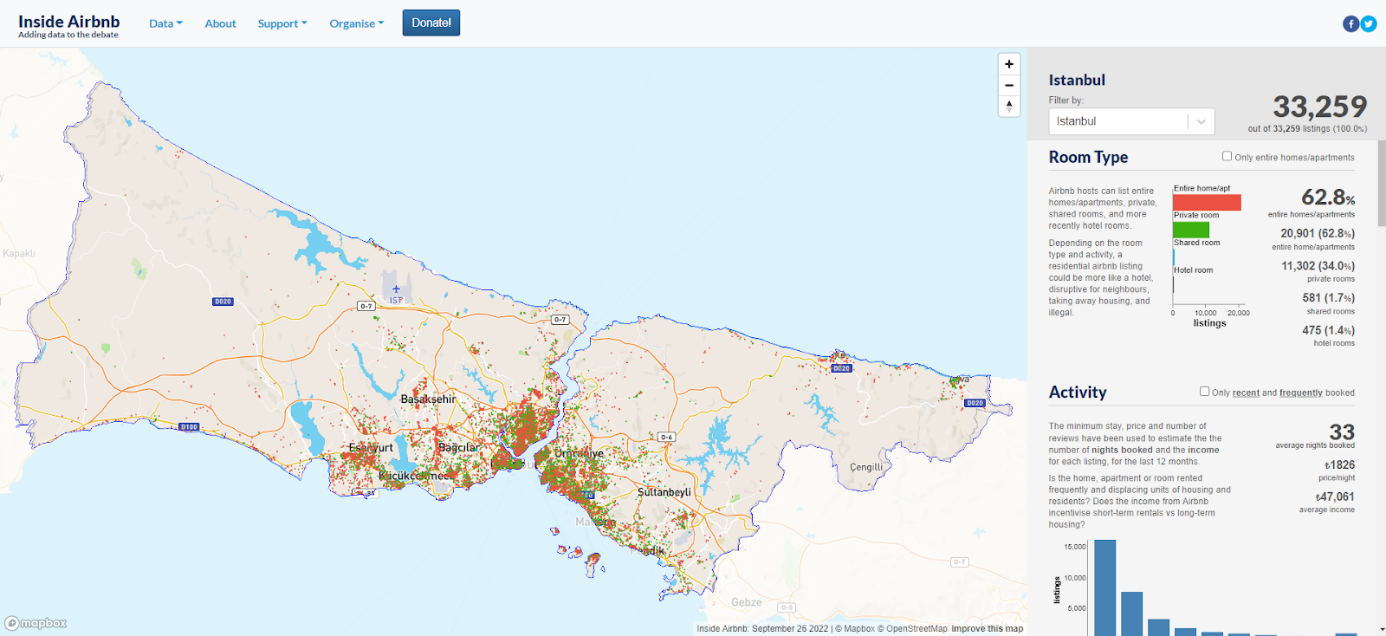
Inside Airbnb, Airbnb'nin konut toplulukları üzerindeki etkisi hakkında veri ve savunuculuk sağlayan misyon odaklı bir projedir. Inside Airbnb'nin amacı paydaşlarının konut kiralamanın rolünü anlamaları, doğru ve etkili karar vermeleri üzerine bir vizyonla veri sağlamayı hedeflemektedir.



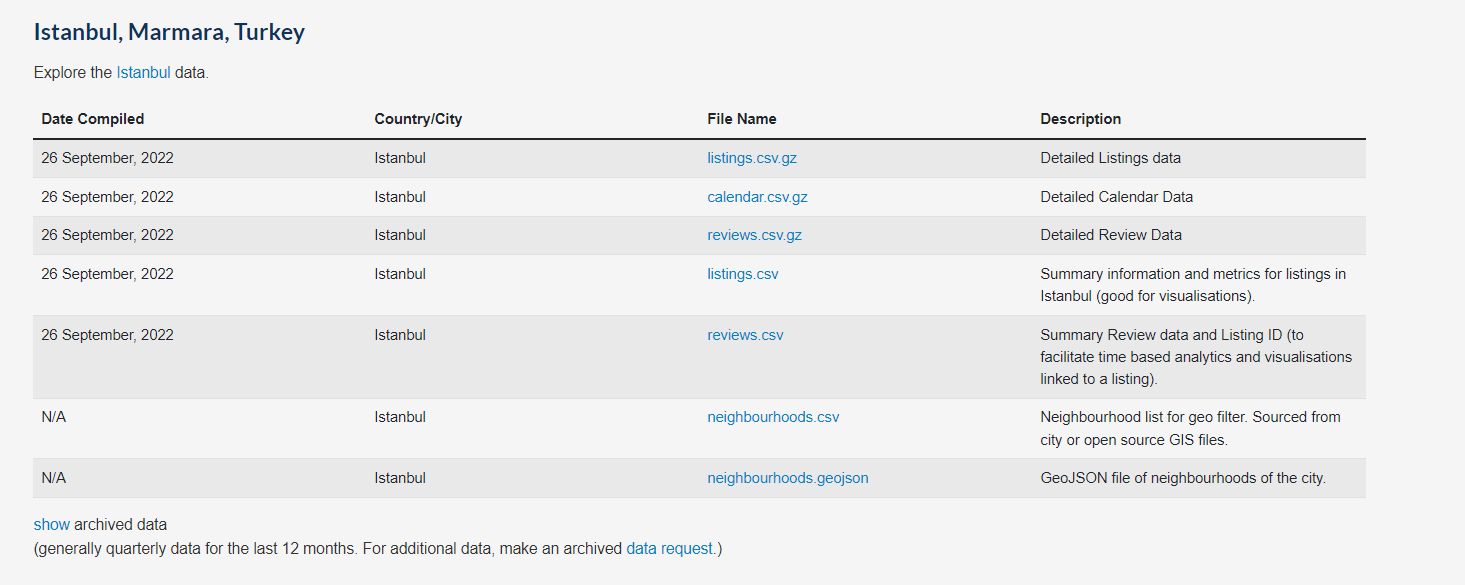
Inside Airbnb sitesinin sunduğu veriler, Airbnb sitesindeki halka açık bilgilerden elde edilir. Veriler, Inside Airbnb tarafından temizlenir, düzenlenir ve kullanıma hazır hale getirilir.

Türkiye'den sadece İstanbul'a ait veriler sunulduğu için Ankara yerine İstanbul'a ait veriler ile ilerlenmiştir.

İstanbul AirBnb veri seti; ev sahipleri, coğrafi müsaitlik durumu, mahalle ve yorumlar hakkında daha fazla bilgi edinmek için bilgiler içerir.



Veri seti 26 Eylül 2022'de güncellenmiştir.



## **2. Verilerin Temizlenmesi**

##gerekli kütüphanelerin import edilmesi

import pandas as pd #veri manipülasyonu için

import numpy as np #veri manipülasyonu için

import seaborn as sns #görselleştirme için

import matplotlib.pyplot as plt

%matplotlib inline

import warnings

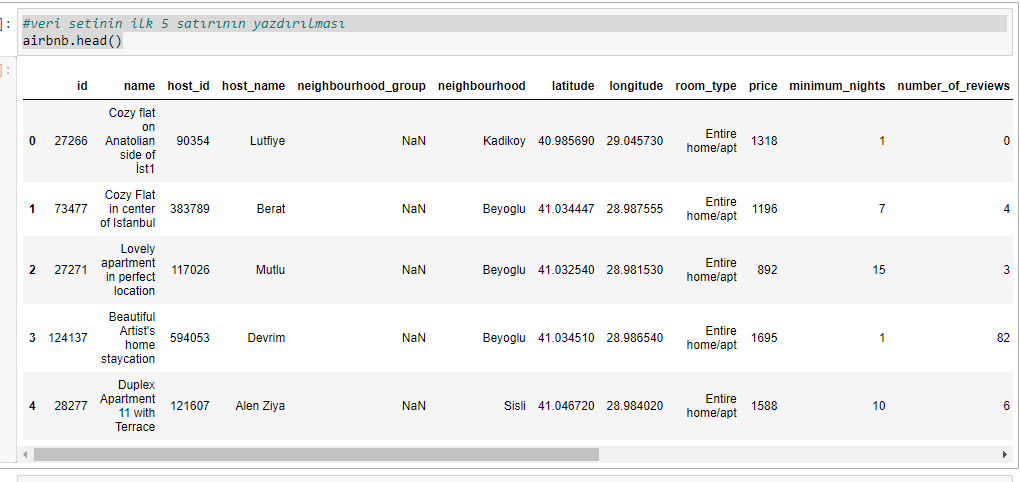
warnings.filterwarnings("ignore")

#veri setinin yüklenmesi

airbnb = pd.read\_csv("listings.csv")

#veri setinin ilk 5 satırının yazdırılması

airbnb.head()

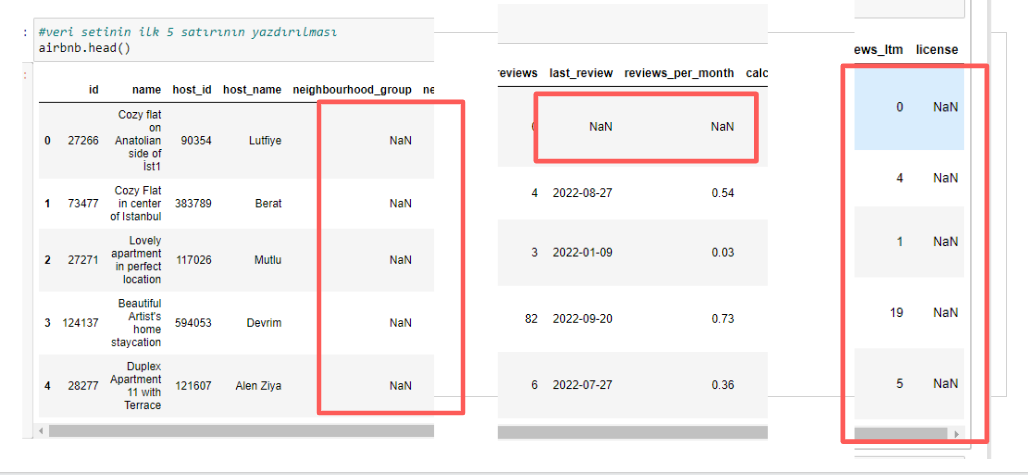


#veri setinin son 5 satırının yazdırılması

airbnb.tail()

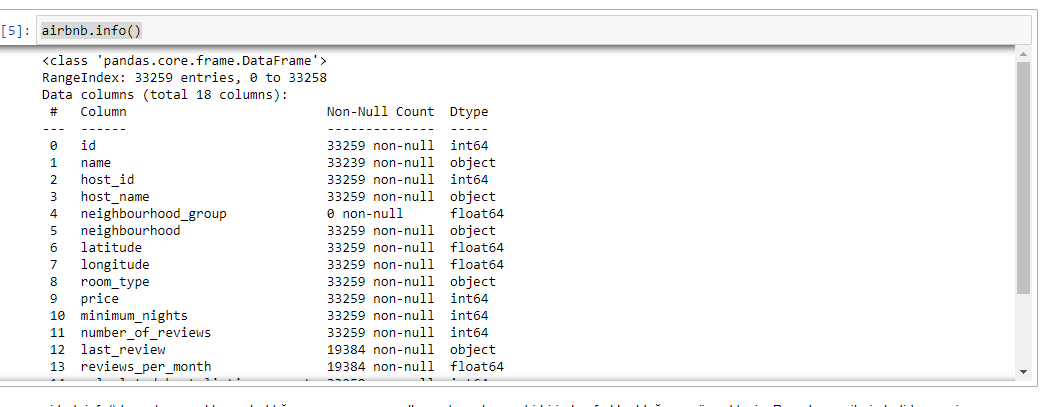
### **2.1 Nan Değerlerin Temizlenmesi ve Tamamlanması**

Veri setine ilk bakışta çok fazla NaN olduğu, neighbourhood\_group ve license sütunlarının tamamen NaNdan oluştuğu görülmektedir.



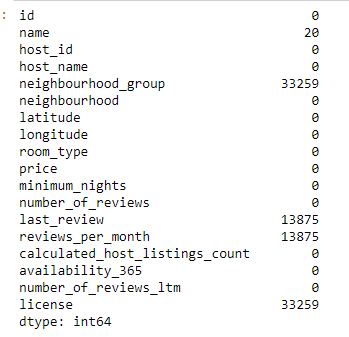
Veri setiyle ilgili daha detaylı bilgi almak için:

airbnb.info()



airbnb.info() komutunun çıktısına baktığımız zaman non-null count sayılarının birbirinden farklı olduğunu görmekteyiz. Boş olan verilerin belirlenmesi ve temizlenmesi için:

airbnb.isnull().sum()



neighbourhood\_group ve license sütunlarının tamamen NaNdan oluştuğu gözlemimiz bu komutun çıktılarıyla doğrulanmıştır. Bu verilerin doldurulması mümkün olmadığından veri setinden çıkarılması gerekmektedir.

airbnb.drop(['neighbourhood\_group','license'], axis=1, inplace=True)

airbnb.info()

airbnb.info() komutunun çıktısından aynı zamanda last\_review sütunun veri türünün object olduğu görülmektedir. Veri setine bakıldığında bu verilerin aslında yıl-ay-gün formatında tarih verileri olduğu görülmektedir. Bu verilerin türlerinin düzeltilmesi gerekmektedir:

airbnb["last\_review"] = pd.to\_datetime(airbnb.last\_review)

Verilerin türleri düzelmekle beraber name, last\_review ve reviews\_per\_month verilerinde eksiklikler olduğu göze çarpmaktadır.

1. Name alanı cevap aranan sorularla alakalı olmadığı için boş bırakılabilir.
2. "last\_review" tarih değerleridir; listeleme için herhangi bir değerlendirme yoksa tarih değeri mevcut olmayacaktır. Bizim sorularımız için bu sütun önemsizdir, bu nedenle bu değerlerin eklenmesine gerek yoktur.
3. "review\_per\_month" sütununa eksik değerler için basitçe 0.0 ekleyebiliriz; toplam 0 inceleme ile bu mantığı takip ederek ayda 0,0 inceleme oranı olacaktır.

airbnb.reviews\_per\_month.isnull().sum()

airbnb.fillna({'reviews\_per\_month':0}, inplace=True)

#"reviews\_per\_month" içindeki tüm NaN değerlerinin 0 ile değiştirilmesi

airbnb.isnull().sum()

##Son olarak birden fazla kez listelenen ilanların da temizlenmesi için:

airbnb.duplicated().sum()

airbnb.drop\_duplicates(inplace=True)

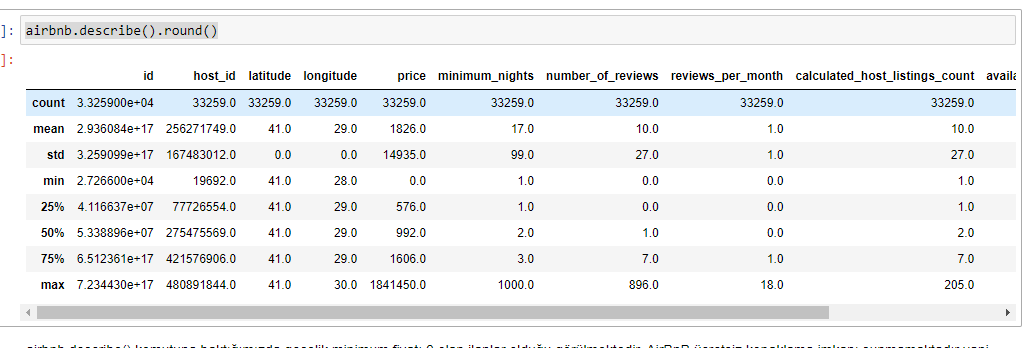
airbnb.info()

#veri setinin son haline bakıldığında boş verilerin temizlediği ve

#birden fazla kez girilmiş ilan olmadığı görülmektedir.

### **2.2 Temel İstatistik Bilgilerin İncelenmesi**

airbnb.describe().round()



airbnb.describe() komutuna baktığımızda gecelik minimum fiyatı 0 olan ilanlar olduğu görülmektedir. AirBnB ücretsiz konaklama imkanı sunmamaktadır yani price=0 olan verilerin hatalı olduğu söylenebilir.

Ayrıca son 365 gün içindeki uygunluğu 0 olan konaklama seçenekleri olduğu da görülmektedir. Insider AirBnB sitesinde bu durumun aktif olmayan ya da uzun süreli tutulmuş konaklamaları temsil ettiği belirtilmektedir.

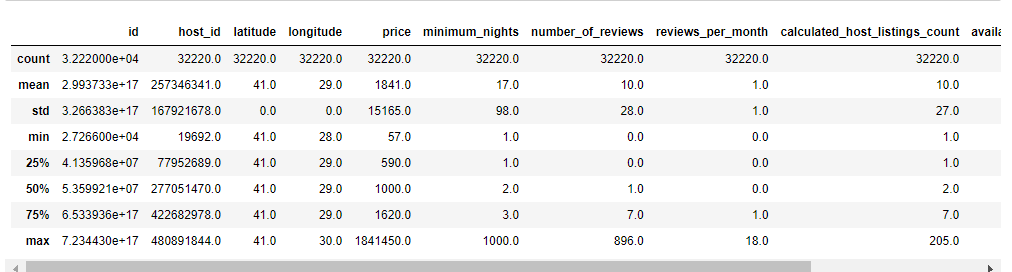
price=0 ve availability\_365=0 olan veriler analizi olumsuz etkileyeceğinden temizlenmelidir:

#price = 0 vey availability\_365 = 0 olan satırların silinmesi

indexNames = airbnb[ (airbnb['price'] == 0) | (airbnb['availability\_365'] == 0) ].index

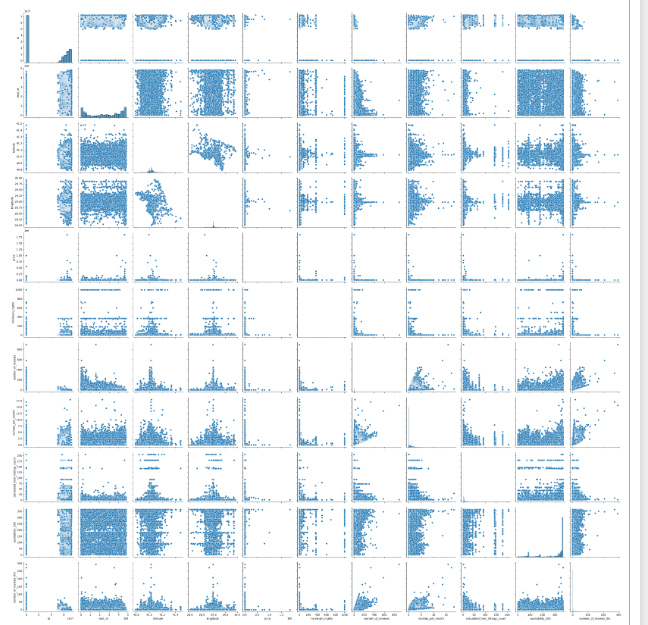
airbnb.drop(indexNames , inplace=True)

airbnb.describe().round()



plt.figure(figsize=(30, 30))

sns.pairplot(airbnb, height=3, diag\_kind="hist")



plt.figure(figsize=(15,12))

palette = sns.diverging\_palette(20, 220, n=256)

corr=airbnb.corr(method='pearson')

sns.heatmap(corr, annot=True, fmt=".2f", cmap=palette, vmax=.3, center=0,

square=True, linewidths=.5, cbar\_kws={"shrink": .5}).set(ylim=(11, 0))

plt.title("Correlation Matrix",size=15, weight='bold')



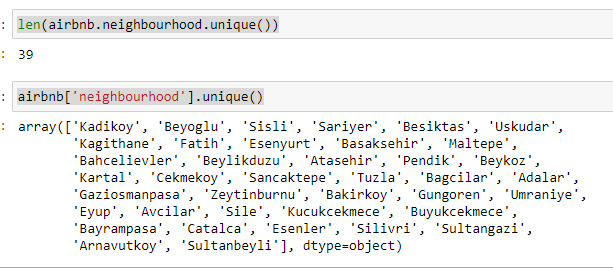
## **3. Verilerin Analiz Edilmesi**

### **3.1 En çok hangi semtte AirBnb bulunmaktadır?**

Semtlere göre konaklamaların dağılımının yapılabilmesi için ilk olarak hangi semtlerde AirBnB evleri olduğuna bakılması gerekmektedir.

len(airbnb.neighbourhood.unique())

airbnb['neighbourhood'].unique()



Buna göre İstanbul'un tüm semtlerinde AirBnb bulunmaktadır. Semt sayısı bilgisi Wikipedi'den edinilmiştir:

f,ax = plt.subplots(figsize=(40,15))

ax = sns.countplot(airbnb.neighbourhood,palette="muted")

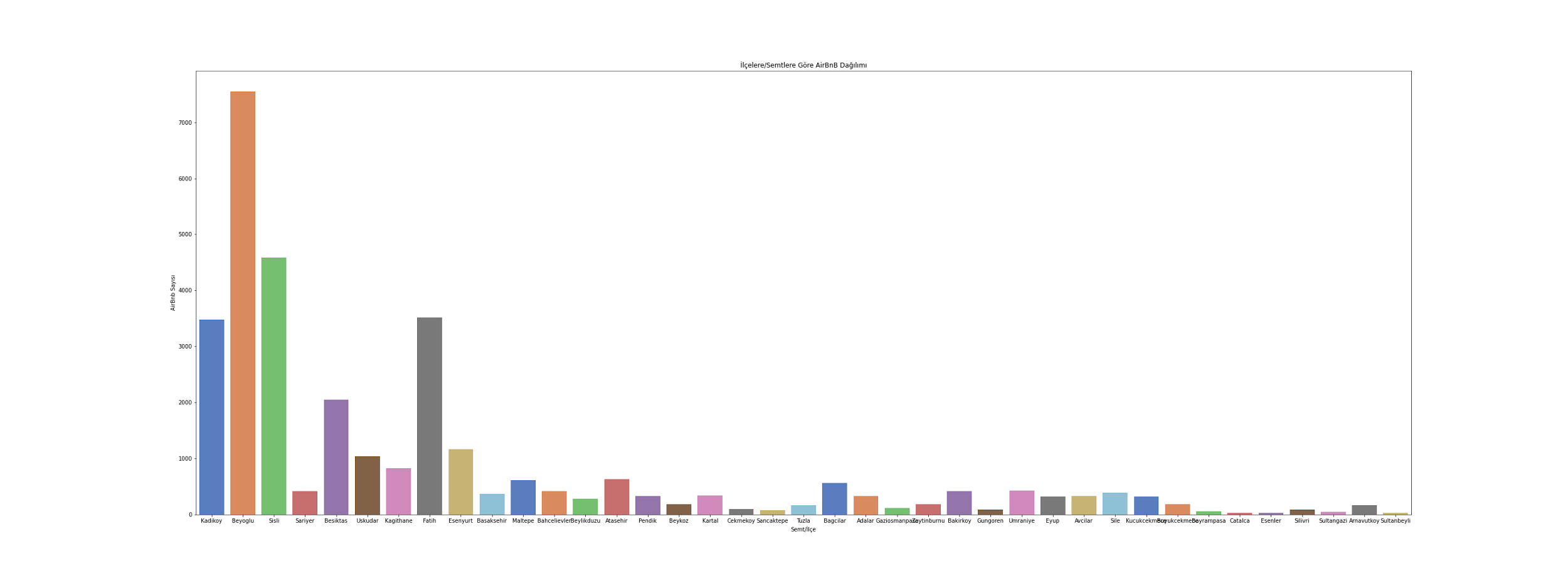
plt.title('İlçelere/Semtlere Göre AirBnB Dağılımı')

plt.xlabel('Semt/İlçe')

plt.ylabel('AirBnb Sayısı')

#plt.show()

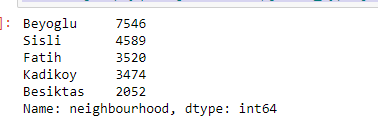
plt.savefig('Semtlere\_Gore\_AirBnB\_Dagilimi.png')



Bu dağılıma göre sırasıyla en çok Beyoğlu, Şişli, Fatih, Kadıköy ve Beşiktaş'ta AirBnB olduğu görülmektedir. Bu gözlemin doğrulaması için:

airbnb.neighbourhood.value\_counts().head(5)

#airbnb.groupby('neighbourhood')['room\_type'].count()

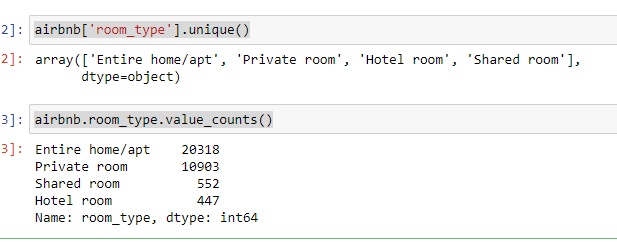


### **3.2 En çok hangi tür AirBnb'ler bulunmaktadır?**

İlk olarak AirBnb türlerine bakılması gerekmektedir:

airbnb['room\_type'].unique()

airbnb.room\_type.value\_counts()



f,ax = plt.subplots(figsize=(12,5))

ax = sns.countplot(airbnb.room\_type,palette="muted")

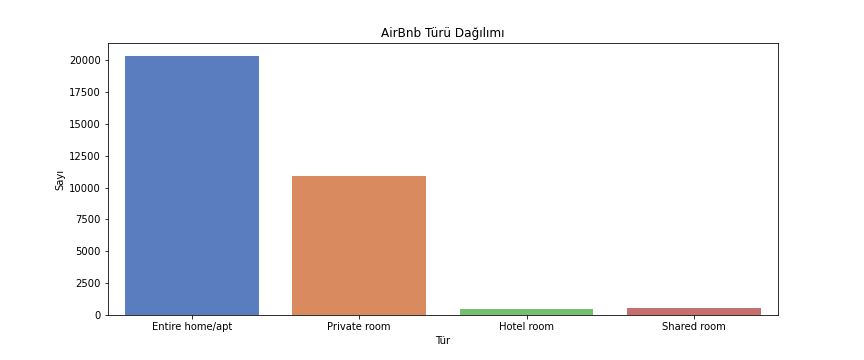
plt.title('AirBnb Türü Dağılımı')

plt.xlabel('Tür')

plt.ylabel('Sayı')

#plt.show()

plt.savefig('AirBnb\_Turu\_Dagilimi.png')



Tüm ev seçeneğinin diğer AirBnb türlerinin toplamının neredeyse 2 katı olduğunu görebiliriz.

### **3.3 Semtlere göre AirBnB tür dağılımı nasıldır?**

fig, ax = plt.subplots(figsize=(8, 30))

sns.countplot(data=airbnb, y='neighbourhood', hue='room\_type', ax=ax)

plt.title('Semstlere Göre AirBnb Tür Dağılımı', size=15)

plt.xlabel('Sayı', size=15)

plt.ylabel('Semtler', size=15)

plt.show()

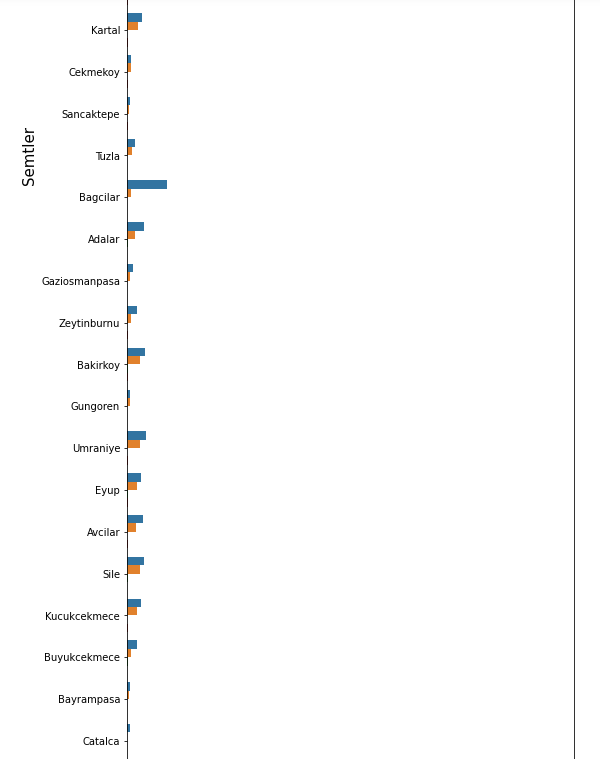
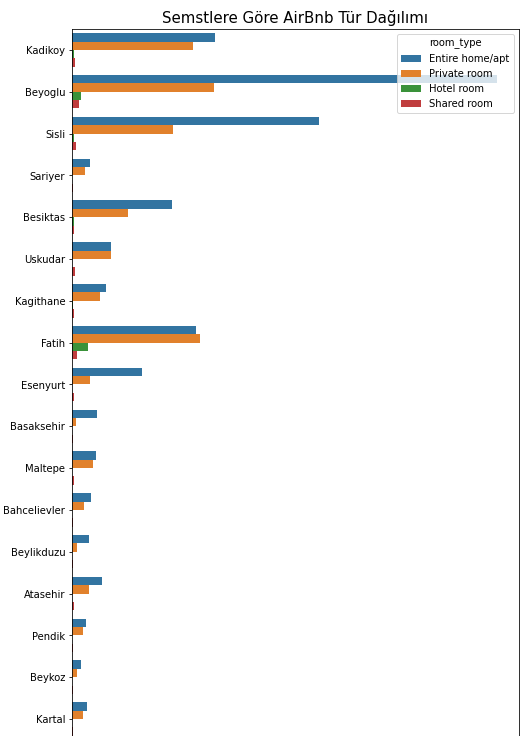


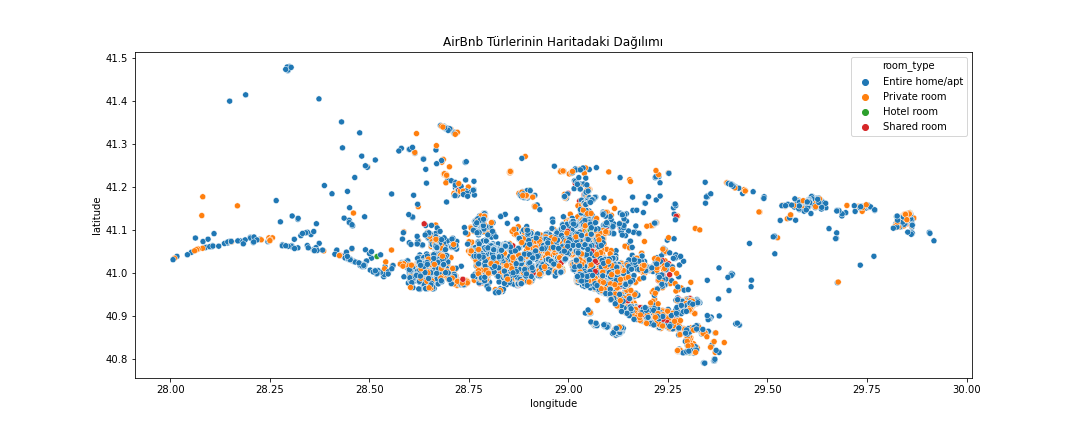
fig = plt.figure(figsize=(15,6))

sns.scatterplot(airbnb['longitude'], airbnb['latitude'], hue=airbnb['room\_type'])

plt.title('AirBnb Türlerinin Haritadaki Dağılımı')

#plt.show()

plt.savefig('AirBnb\_Turlerinin\_Haritadaki\_Dagilimi.png')

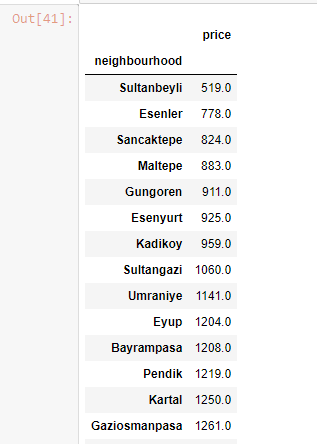


### **3.4 Gecelik ortalama fiyatlar semtlere göre nasıl değişiklik göstermektedir?**

airbnb.groupby(['neighbourhood'])['price'].mean().to\_frame().sort\_values(by='price').round()

#semtegore = airbnb.groupby("neighbourhood")['price'].agg(['min', 'max', 'mean', 'count']).reset\_index()

#semtegore.sort\_values(by="mean", ascending=False)



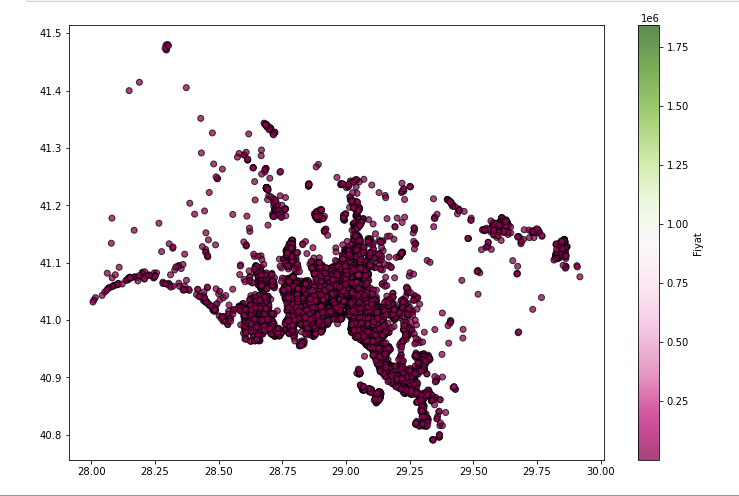
plt.figure(figsize=(12, 8))

plt.scatter(airbnb.longitude, airbnb.latitude, c=airbnb.price, cmap='PiYG', edgecolor='black', linewidth=1, alpha=0.75)

cbar = plt.colorbar()

cbar.set\_label('Fiyat')

**Not: Gecelik fiyatlar birbirine çok yakın olduğundan renk farkları net olarak görülememektedir**



### **3.5 Türlerine göre AirBnb'lerin ortalama gecelik getirisi nasıl değişmektedir?**

airbnb['room\_type'].unique()

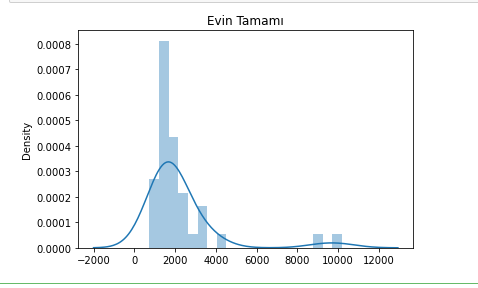
df1 = airbnb[airbnb.room\_type == "Entire home/apt"][["neighbourhood","price"]]

d = df1.groupby("neighbourhood").mean()

plt.title('Evin Tamamı')

sns.distplot(d)

plt.show()



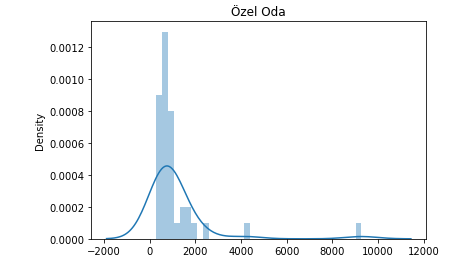
df2 = airbnb[airbnb.room\_type == "Private room"][["neighbourhood","price"]]

d = df2.groupby("neighbourhood").mean()

plt.title('Özel Oda')

sns.distplot(d)

plt.show()



df3 = airbnb[airbnb.room\_type == "Hotel room"][["neighbourhood","price"]]

d = df3.groupby("neighbourhood").mean()

plt.title('Otel Odası')

sns.distplot(d)

plt.show()

df4 = airbnb[airbnb.room\_type == "Shared room"][["neighbourhood","price"]]

d = df3.groupby("neighbourhood").mean()

plt.title('Paylaşımlı Oda')

sns.distplot(d)

plt.show()

plt.figure(figsize=(18,18))

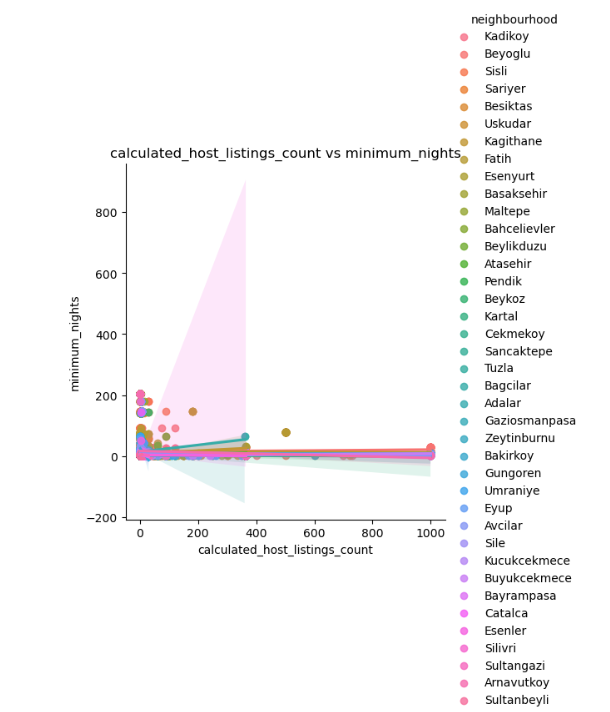
sns.lmplot(x='minimum\_nights',y='calculated\_host\_listings\_count',hue="neighbourhood",data=airbnb)

plt.xlabel('calculated\_host\_listings\_count')

plt.ylabel('minimum\_nights')

plt.title('calculated\_host\_listings\_count vs minimum\_nights')

plt.show()



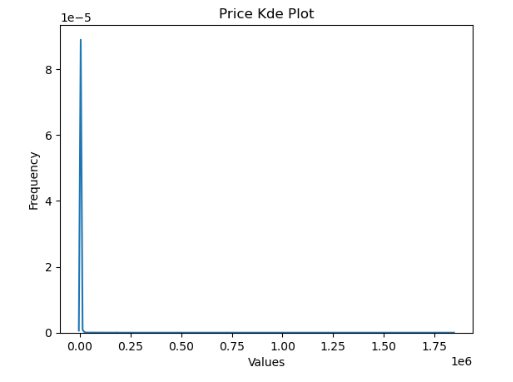
sns.kdeplot(airbnb['price'])

plt.xlabel('Values')

plt.ylabel('Frequency')

plt.title('Price Kde Plot')

plt.show()



### **3.6 AirBnb türlerine göre gecelik fiyatlar semtlere göre nasıl değişmektedir?**

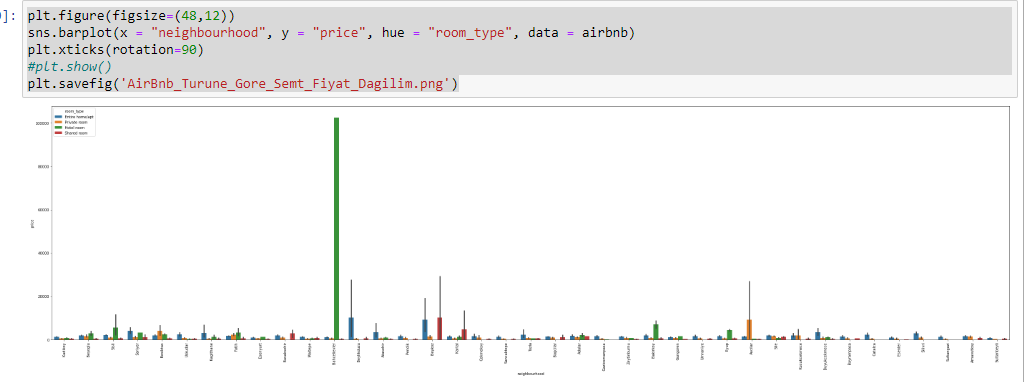
plt.figure(figsize=(48,12))

sns.barplot(x = "neighbourhood", y = "price", hue = "room\_type", data = airbnb)

plt.xticks(rotation=90)

#plt.show()

plt.savefig('AirBnb\_Turune\_Gore\_Semt\_Fiyat\_Dagilim.png')



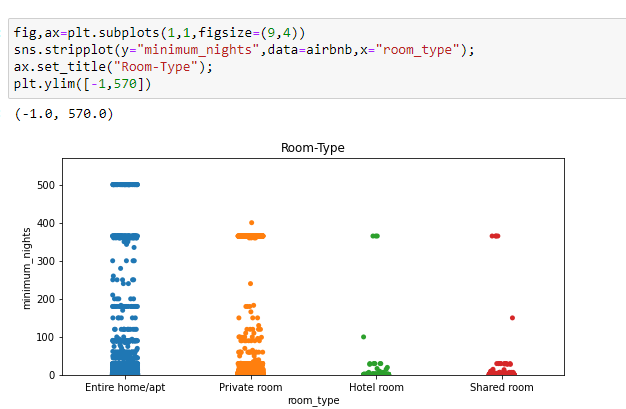
### **3.7 AirBnb türlerine göre konaklanan minimum gece sayısı nasıl değişmektedir?**

fig,ax=plt.subplots(1,1,figsize=(9,4))

sns.stripplot(y="minimum\_nights",data=airbnb,x="room\_type");

ax.set\_title("Room-Type");

plt.ylim([-1,570])



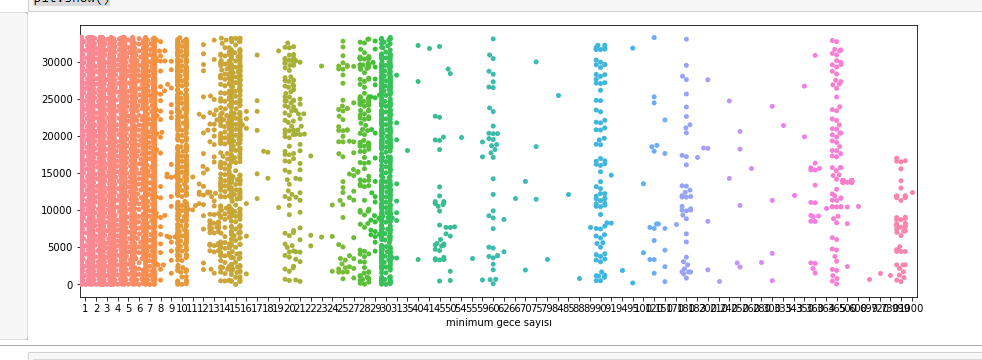
minnight = airbnb[airbnb.room\_type=="Entire home/apt"]['minimum\_nights']

f,ax = plt.subplots(figsize=(15,5))

ax = sns.swarmplot(y= minnight.index,x= minnight.values)

plt.xlabel("minimum gece sayısı")

plt.show()



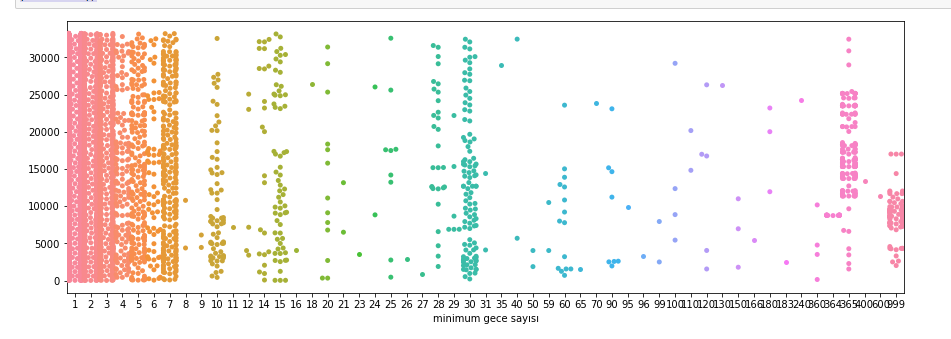
minnight = airbnb[airbnb.room\_type=="Private room"]['minimum\_nights']

f,ax = plt.subplots(figsize=(15,5))

ax = sns.swarmplot(y= minnight.index,x= minnight.values)

plt.xlabel("minimum gece sayısı")

plt.show()



İnsanlar, genellikle aileleri veya arkadaşları ile seyahat ettiklerinde daha rahat olması açısından için tüm evde veya özel odalarda kalmayı tercih edebilirler. Ya da nomad denilen gezginler için de bu konaklamalar uygun olabilir. Özel evlerde 90 güne kadar, özel odalarda ise 30 güne kadar bir yoğunluk olduğu görülmektedir. Yine de en çok 1 veya 2 hafta kalanların çoğunlukta olduğu görülmektedir. Hosting için harcanan eforun azaltılması için bu tür bir konaklama oluşturulabilir

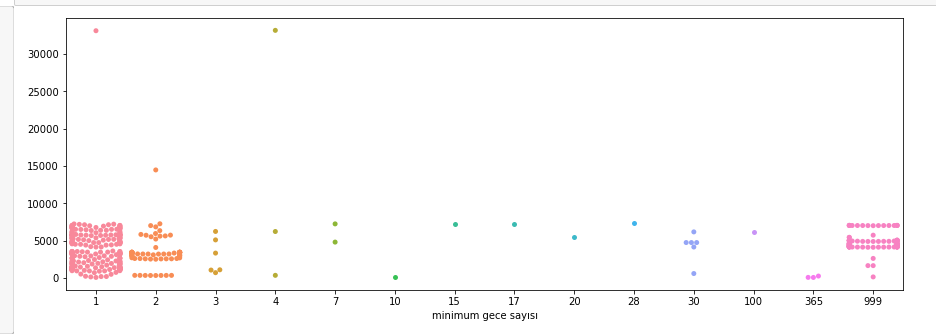
minnight = airbnb[airbnb.room\_type=="Hotel room"]['minimum\_nights']

f,ax = plt.subplots(figsize=(15,5))

ax = sns.swarmplot(y= minnight.index,x= minnight.values)

plt.xlabel("minimum gece sayısı")

plt.show()



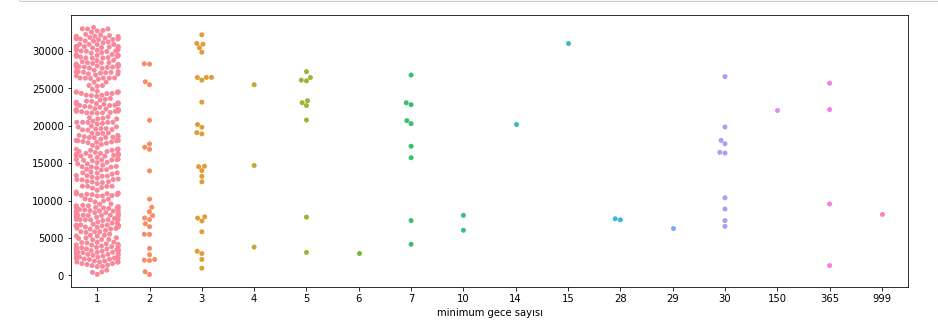
minnight = airbnb[airbnb.room\_type=="Shared room"]['minimum\_nights']

f,ax = plt.subplots(figsize=(15,5))

ax = sns.swarmplot(y= minnight.index,x= minnight.values)

plt.xlabel("minimum gece sayısı")

plt.show()

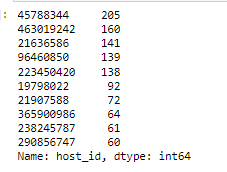


Çoğunlukla gezginler, sırt çantalı gezginler ve düşük bütçeli kişiler ortak odalarda kalmayı sever. Bununla beraber diğer turistler de şehirler arası gezmektedir. Otel odası veya paylaşımlı oda formatında 1-2 gün konaklamanın en fazla olduğu görülmektedir. Bu hızlı dönüşüm nevresim yıkama, temizlik vb. masrafları artırıcı bir etkendir.

### **3.8 Birden fazla Airbnb ilanına sahip hostlar var mı?**

top\_host=airbnb.host\_id.value\_counts().head(10)

top\_host



#'calculated\_host\_listings\_count' sütununu kullanarak 205 ilan sayısının doğrulanması

top\_host\_check=airbnb.calculated\_host\_listings\_count.max()

top\_host\_check

sns.set(rc={'figure.figsize':(10,8)})

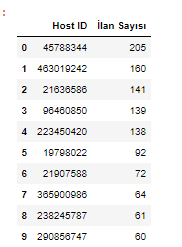
sns.set\_style('white')

top\_host\_df=pd.DataFrame(top\_host)

top\_host\_df.reset\_index(inplace=True)

top\_host\_df.rename(columns={'index':'Host ID', 'host\_id':'İlan Sayısı'}, inplace=True)

top\_host\_df



hosts=sns.barplot(x="Host ID", y="İlan Sayısı", data=top\_host\_df,

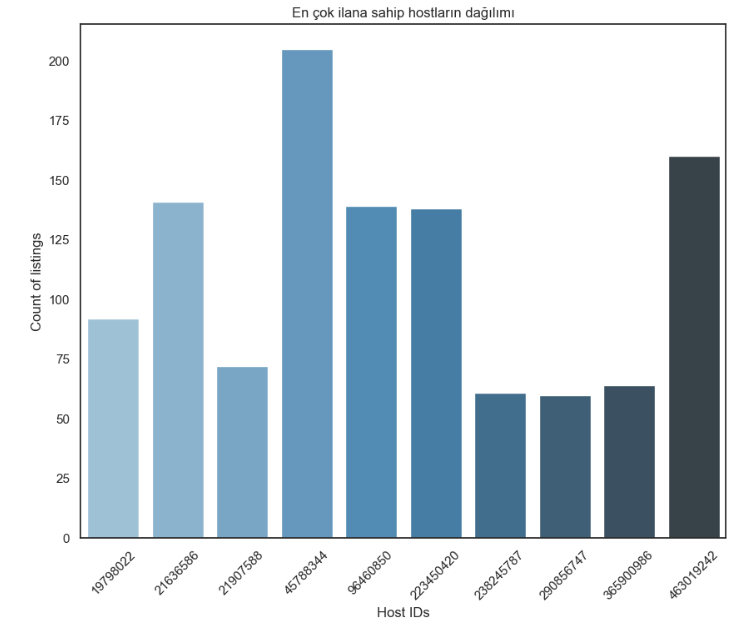
palette='Blues\_d')

hosts.set\_title('En çok ilana sahip hostların dağılımı')

hosts.set\_ylabel('Count of listings')

hosts.set\_xlabel('Host IDs')

hosts.set\_xticklabels(viz\_1.get\_xticklabels(), rotation=45)



### **3.9 İlanlarda en çok hangi kelimeler kullanılıyor?**

from wordcloud import WordCloud, ImageColorGenerator

text = " ".join(str(each) for each in airbnb.name)

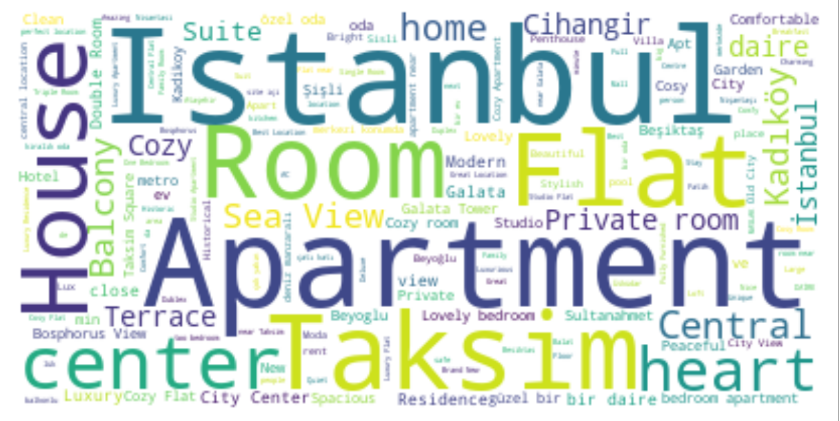
wordcloud = WordCloud(max\_words=200, background\_color="white").generate(text)

plt.figure(figsize=(15,10))

plt.imshow(wordcloud, interpolation='bilinear')

plt.axis("off")

plt.show()



## **4. Sonuç ve Öneriler**

* İstanbul’da en çok tüm ev formatında konaklamalar bulunmaktadır.
* En çok Beyoğlu, Şişli, Fatih, Kadıköy ve Beşiktaş'ta AirBnB bulunmaktadır.
* Konaklama ve semtten bağımsız ortalama bir gecelik konaklama maliyeti 1841 TL’dir
* Özel ev konaklamalar diğer türlerin toplamının neredeyse iki katıdır.
* Müşteri dönüşümünün getireceği masraflardan kaçınmak için tüm ev veya özel oda tarzı konaklamalar oluşturulmalıdır.
* Id, host-id ve name alanlarının kategorik değerler olarak değiştirilmesi
* Enlem ve boylam üzerinden oluşan grafiklerin arkasına harita görseli eklenmesi
* Tür dağılımlarının subplotlarla bir arada verilmesi
* Semtteki airbnb sayısı ve gecelik fiyat arasındaki ilişkinin korelasyonunun hesaplanması
* Konaklamaların review sayılarından yaklaşık müşteri sayıları elde edilerek konaklama değişkenlerine bağlı tercihler analiz edilebilir.
* Veri setinde her evin kaç kez tutulduğu ve ortalama olarak kaç gecelik tutulduğu bilgileri olsaydı daha detaylı analizler yapılabilirdi.
* Bölgelerdeki ortalama ev fiyatlarına ilişkin veriler ve bu veri seti  kullanılarak yaklaşık ROI hesabı yapılabilir.