

MAKİNE ÖĞRENMESİ

DÖNEM PROJESİ

RAPOR

Berke YALÇIN 200701057

Hamza DEMİR 190701156

Öncelikle kiralanmak istenilen evin özelliklerini aldık:

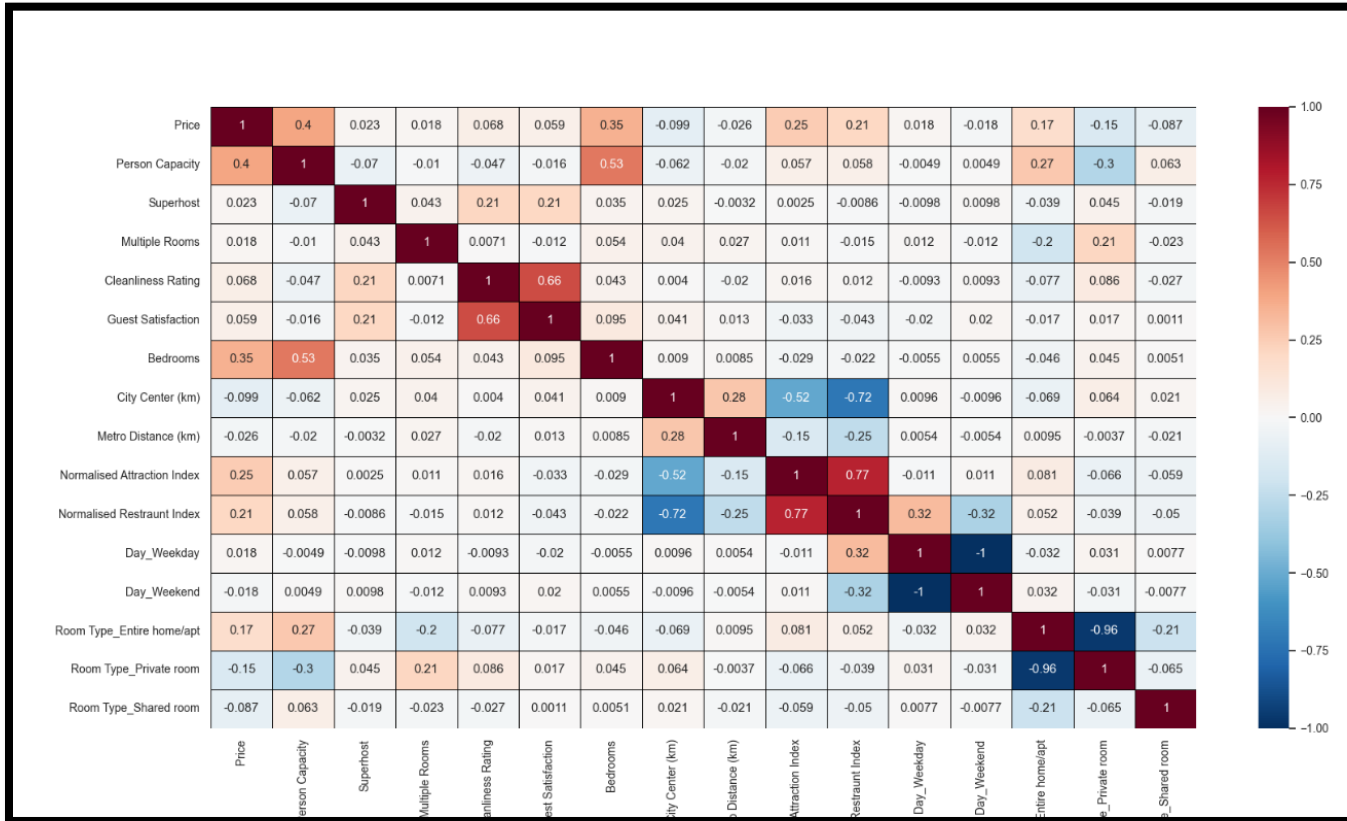
```
Kiralama yapmak istediğiniz şehri giriniz: Paris
Person Capacity özelliğinin değerini girin: 2
Superhost özelliğinin değerini girin: 0
Multiple Rooms özelliğinin değerini girin: 0
Bedrooms özelliğinin değerini girin: 1
City Center (km) özelliğinin değerini girin: 2.4
Metro Distance (km) özelliğinin değerini girin: 1.7
Normalised Attraction Index özelliğinin değerini girin: 24.78
Normalised Restraunt Index özelliğinin değerini girin: 43.22
Day_Weekday özelliğinin değerini girin: 0
Day_Weekend özelliğinin değerini girin: 1
Room Type_Entire home/apt özelliğinin değerini girin: 0
Room Type_Private room özelliğinin değerini girin: 1
Room Type_Shared room özelliğinin değerini girin: 0
```

Diğer şehirleri veritemizden kaldırdık.

```
#Diğer şehirleri verisetimizden kaldırabiliriz
data = data[data['City'] == target_city]

data = data.drop("City", axis=1)
```

Ardından seçtiğimiz şehir için bir heatmap oluşturduk, böylece özelliklerin diğer özellikler üzerinde etkisini analiz edebildik:

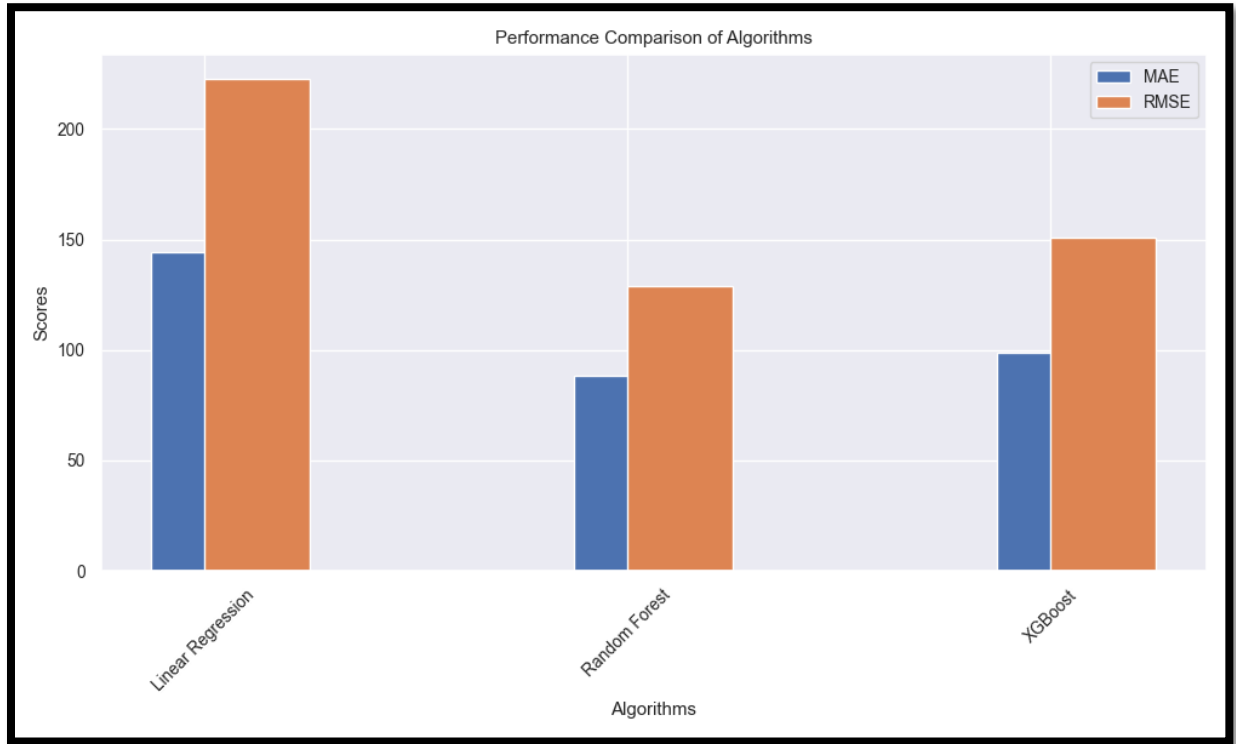


Fonksiyonlarımız ile verisetimizi temizledik:

```
+def data_preparation(data):...  
+def encode_categorical_features(data):...
```

“Business”, “Private Room”, “Shared Room”, “Attraction Index”, “Restaunt Index” sütunlarımızı kaldırdık, “Day” ve “Room Type” sütunlarımızı numerik olmadıkları için encode ettik.

Lineer Regresyon, Random Forest ve XGBoost algoritmalarıyla girdiğimiz özellikteki evin fiyatını tahmin ettik ve bu arada algoritmaların performanslarını karşılaştırarak en doğru tahmini aldık.



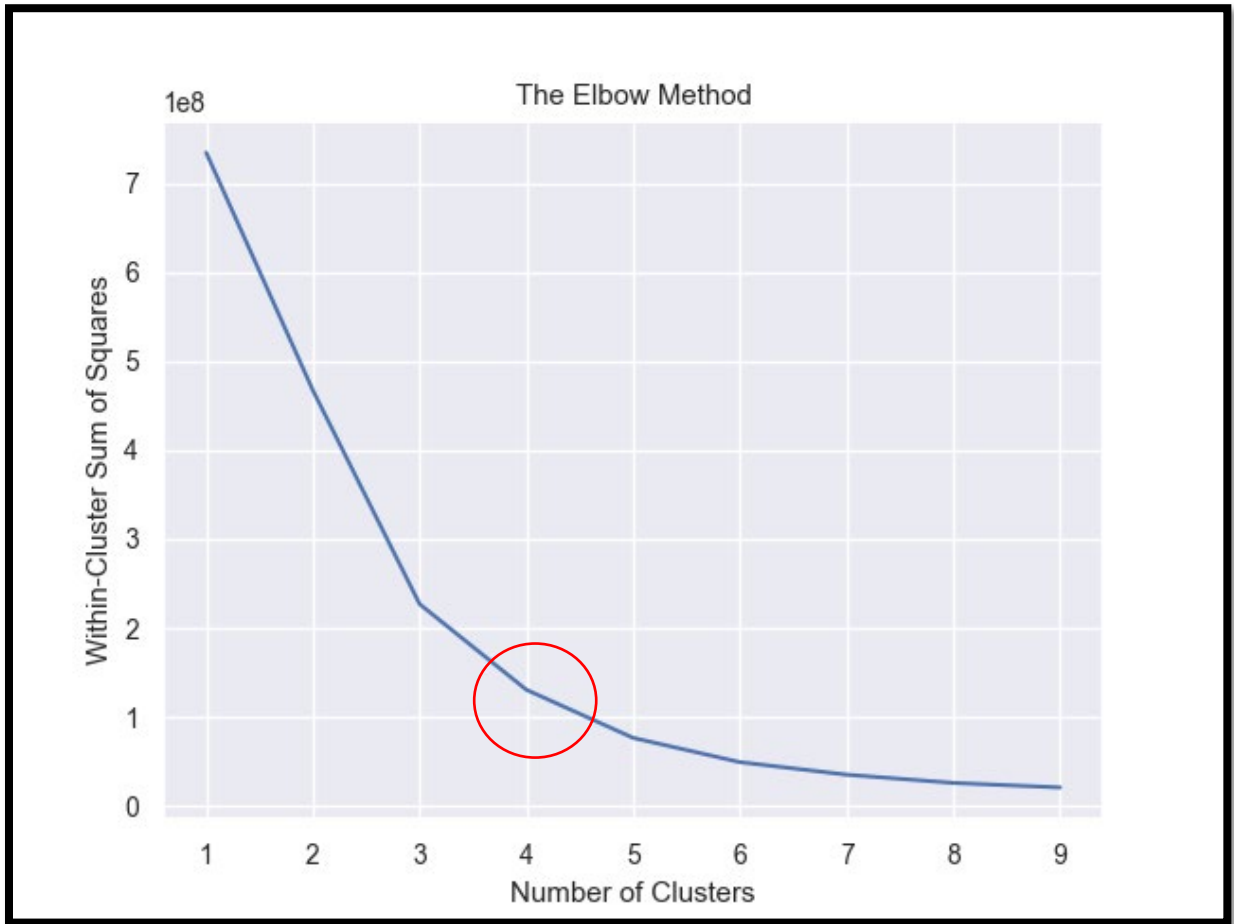
	Algorithm	MAE	MSE	RMSE	R2 Square
0	Linear Regression	144.137911	49635.199703	222.789586	0.249289
1	Random Forest	88.356950	16650.327359	129.036147	0.748171
2	XGBoost	98.768843	22799.171386	150.993945	0.655172
En Verimli Algoritma Random Forest					
Fiyat Tahmini: [324.54795414]					

Evimize benzer özellikteki evlerin fiyatlarını öğrenmek istedik:

K-Means uygulamaya karar verdik ve ilk olarak verisetimize evimizi de ekledik ;

```
# Verisetimize ekle  
merged_data = pd.concat([clean_data, ev_ozellikleri], ignore_index=True)
```

Sonrasında Elbow methoduyla küme sayısını bulmaya çalıştık;



En uygun kırımın 4 kümede gerçekleştiğine karar verdik ve K-Means'ı 4 küme ile uyguladık.

Kümeleme işlemi tamamlandıktan sonra, evimizin hangi kümeye ait olduğunu tespit ettik,

Kendi içlerinde fiyata göre sıralayıp evimize en yakın fiyatlı 6 (3 düşük ve 3 fazla) evi tüm özellikleriyle gösterdik.

Böylece satıcı kendisine önerdiğimiz fiyat aralığındaki diğer evlerin de özelliklerini görüp piyasa hakkında daha fazla bilgiye sahip olabilir:

Price	Person Capacity	Superhost	Multiple Rooms	Cleanliness Rating	Guest Satisfaction	Bedrooms
324.354553	2.0	1.0	0.0	10.0	99.0	1.0
324.354553	2.0	0.0	1.0	9.0	92.0	1.0
324.354553	3.0	0.0	0.0	9.0	92.0	1.0
324.587566	4.0	0.0	0.0	10.0	100.0	1.0
324.587566	3.0	0.0	0.0	10.0	100.0	1.0
324.587566	2.0	0.0	0.0	10.0	94.0	0.0

City Center (km)	Metro Distance (km)	Normalised Attraction Index	Normalised Restaunt Index	Day	Room Type
1.195675	0.042113	19.435618	61.977186	Weekday	Entire home/apt
1.044463	0.214093	19.639506	47.187678	Weekend	Entire home/apt
1.166382	0.181069	26.662266	78.992601	Weekday	Entire home/apt
1.581219	0.245072	18.499240	46.795034	Weekend	Entire home/apt
3.291262	0.243539	10.063800	23.325955	Weekend	Entire home/apt
4.174759	0.394068	14.756150	30.360402	Weekend	Entire home/apt