

KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ



2020-2021 GÜZ DÖNEMİ

BİL 413

ÖRÜNTÜ TANIMA

**Emlak Firması Veriseti Oluşturma, Problem Belirlenmesi ve Veri
Analizi**

Vize Ödevi

Hazırlayan

EBRU ŞİMŞEK

18110131003 (ebrusimsek201@gmail.com)

Ders Öğretim Üyesi

Özet_ Yapay zekanın gelişmesiyle her sektöre uygun farklı çözümler üretilmiş ve üretilmektedir.Örüntü tanıma insan becerileriyle yıllar sürecektir. Örüntülerin yapay zeka ile makineye hızlı bir şekilde bulundurulmasını içerir. Örüntü tanıma makine öğrenmesinin öğrenme biçiminde denilebilir.Gözetimli öğrenme yapay zekada sık kullanılan bir makine öğrenmesi veri analiz biçimidir. Makine tarafından veriler arasındaki örüntüler tanınarak öğrenme dediğimiz eğitim gerçekleştirilir.Eğitim sonucu modeller geliştirilir ve bu modellere yeni verilerde analiz için kullanılır.

Bu çalışmada da bir emlak firmasına ait ilan kayıtlarının bilgisini içeren veriseti oluşturulmuş, kayıtların internet üzerinden günlük görüntülenme sayısının tahmini şeklinde problem belirlenmiş ve regresyon kullanılarak veri analiz edilmiştir.

Anahtar Kelimeler_ yapay zeka, regresyon, örüntü tanıma, makine öğrenmesi

1. GİRİŞ

Yapay Zeka insanoğlunun davranışlarının modellenmesi ve sistemler geliştirmesi ile ilgilenen bir bilgisayar bilimidir. İnsanın düşünme yöntemlerini analiz ederek bunların benzeri yapay yönergeleri geliştirmeye çalışmaktadır. Çevresi ile etkileşime girebilen ve edindiği verileri kullanarak zeka olarak tanımlanabilecek davranışlar sergileyen makineler üretebilmeye çalışır.[1-3]

Makine öğrenmesi yapay zekanın, verilere ulaşp insan desteği olmadan bu verilerden yeni bilgiler edinen bilgisayar programları geliştirmek için kullanılan bir dalıdır. Sistemin öğrenmesini ve gerekli kararları almasını sağlayan algoritmaların geliştirilmesini amaçlar. Algoritma geliştirmeyi veriler arasındaki insanoğlunun zor farkedeyeceği, farketmesine ömrünün dahi yetmeyeceği örüntüleri hızlıca keşfederek örüntü tanıma ile gerçekleştirir. Hem çok büyük boyutlu veri ile çalışabilmesi hem zamandan kazanç sağlaması nedeniyle makine öğrenmesi diğer bir tabirle "yapay öğrenme" her sektörde kullanılmaya başlamıştır. Veri madenciliği işlemleri toplanan veriyi işlemekle yükümlü olurken makine öğrenmesi bu veriyi tekrar tekrar öğrenerek yeni sonuçlar geliştirir. İşlenen verilere yeni veri kaynakları ekleyerek sonuç potansiyeli artırılabilir ve daha efektif veriler elde edebilir. Buna dayanarak eğitim veri seti ile analiz edilerek makineye öğretilip, test verileri de bu öğrenme ile analiz edilir.[1-3]

Temel olarak makine öğrenmesi gözetimli öğrenme, gözetimsiz öğrenme ve pekiştirmeli öğrenme olarak 3 şekilde veriyi analiz edebilir. Verilerin hangi girdi değerinde hangi çıktıyı vereceği biliniyorsa bu şekilde eğitilmiş makine öğrenmesine gözetimli öğrenme denir. Çıkış değeri bilinmediğinde eğitim veri seti üzerinden benzer örüntülerden faydalanarak eğitilen makine öğrenmesine gözetimsiz öğrenme denir. Pekiştirmeli öğrenme ise bunlardan farklı olarak sistemin eğitim verisinden ziyade denediği olası durumlardan doğru veya yanlış tüm durumların hatırlanmasıyla oluşan eğitilmiş makine öğrenmesidir. .[1-3]

Bu çalışmada da örnek verilerden öğrenmeye dayanan gözetimli öğrenmenin bir yöntemi olan regresyon analizi ile tahmin çalışması yapılmıştır. Bir emlak firmasına ait ilan kayıtlarının bilgisini içeren veriseti oluşturulmuş, kayıtların internet üzerinden görüntülenme sayısının tahmini şeklinde problem belirlenmiş ve regresyon kullanılarak veri analiz edilmiştir.

Bu çalışmanın müşterilerin istek, bütçe ve ihtiyacından doğan ilgilerinin ne yönde olduğunun belirlenmesi, yeni ilanların muhtemel ilgi görme durumunun tahmin edilmesi açısından önemlidir. Firma müşteri ilgilerine göre gerekli eklemeleri ev sahipleri ile görüşerek yaptırabilir yada gözden kaçabilecek ilanları tavsiye sistemleri ile görülebilir yapılabilir.

2. BÖLÜM 1

PROBLEM BELİRLEME

Emlak piyasasında, her bir emlak satılır ya da kiralanır. Günümüzde insanların kiralık ya da satılık emlak ararken başvurdukları ilk kaynak e-ticaret siteleridir. Kiralanan emlak göz önüne alınırsa, emlak sahibi ve kiracı piyasadaki arz-talep ilişkisine göre belirlenen kira değeri ile ilgilenir. [4-7]

Emlak sektörü yapay zeka kullanımında oldukça yenidir. Genellikle emlak veriseti üzerinden kira tahmini yapılmaktadır. Bu sebeple farklı bir yaklaşım sunabilmek adına bu çalışmada ilanın günlük görüntülenme sayısına odaklanılmıştır. Arz-talep ilişkisini güçlendirebilmenin bir yolu olarak görüntülenme sayısı yani siteyi ziyaret eden müşterilerin ilgisi takip edilerek yönlendirilebilir. Müşteri ilgisinin yoğunlaştığı ilan özellikleri takip edilerek ev sahibi ile konuşulup ev özelliklerini güncelleştirmesi sağlanabilir. Yada site ziyaretçisi olan muhtemel müşterilere az görüntülenen ilanlar öneri/tavsiye sistemi ile sunulabilir.

VERİSETİ OLUŞTURMA

Veri kümesi

Kullanılan veriseti bir emlak firmasına ait ilanların özelliklerini içermektedir ve yeni oluşturulmuştur hazır alınmamıştır.

Veri kümesi 105 kayıt ve 13 öznitelik içermektedir. Dosya formatı csv olup 'data.csv' olarak saklanmıştır. 105 satır ve 13 sütundan oluşmaktadır. Matrix olarak düşünüldüğünde 105x13 boyutlarındadır. Probleme daha iyi çözüm üretmek için öznitelikler seçilerek 'data2.csv' ve 'data3.csv' oluşturulmuş ve denenmiştir.

Veri Kümesi Öznitelik Bilgileri

Çalışmada kullanılan öznitelikler filtreleme seçeneklerinden yola çıkılarak ve makinenin algılaması için kategorik değerler sayısallaştırılarak şöyle belirlenmiştir:

- Metrekare
- Oda
 - 0 -> 1+1
 - 1 -> 2+1
 - 2 -> 3+1
 - 3 -> 4+1
 - 4 -> 5+1
 - 5 -> 6+1
 - 6 -> 7+1
- Bina yaşı
- Bulunduğu kat
- Site içinde mi
 - 0 -> hayır
 - 1 -> evet
- Isınma biçimi
 - 0 -> soba
 - 1 -> doğalgaz sobası
 - 2 -> merkezi
 - 3 -> doğalgaz kombi
 - 4 -> klima
 - 5 -> şömine
- Konut tipi
 - 0 -> daire
 - 1 -> müstakil
 - 2 -> rezidans
 - 3 -> villa
- Eşyalımı
 - 0 -> Hayır
 - 1 -> Evet
- Lokasyon
 - 0 -> Mahalle 1
 - 1 -> Mahalle 2
 - 2 -> Mahalle 3

- 3 ->Mahalle 4
- 4 ->Mahalle 5
- 5 ->Mahalle 6 (elit)
- 6 ->Mahalle 7 (en elit)

Veriseti Örneği

	metrekare	oda	binayasi	bulundugukat	site	kira	isi	balkon	banyo	tip	esya	lokasyon	see
0	295	2	14	1	0	2650	5	1	1	3	0	6	20
1	125	2	11	5	0	950	1	2	1	0	0	0	65
2	255	4	24	7	1	2350	3	1	2	0	0	3	45
3	155	3	1	13	1	1150	2	2	1	0	0	1	30
4	290	5	15	15	1	2550	3	3	2	0	0	6	25
5	220	2	4	16	1	1850	3	1	2	0	1	3	25
6	250	2	1	14	0	2250	3	1	2	0	0	4	10
7	150	1	2	1	1	1050	3	0	1	0	0	5	60
8	280	3	15	3	0	2450	3	2	2	2	1	6	10
9	215	5	4	1	1	1750	4	3	1	1	0	3	25
10	240	5	1	2	1	2200	3	4	2	0	0	3	25

Resim 1- Veriseti (data.csv)

	metrekare	kira	banyo	tip	see
0	295	2650	1	3	20
1	125	950	1	0	65
2	255	2350	2	0	45
3	155	1150	1	0	30
4	290	2550	2	0	25
5	220	1850	2	0	25
6	250	2250	2	0	10
7	150	1050	1	0	60
8	280	2450	2	2	10
9	215	1750	1	1	25

Resim 2- Veriseti 2 (data2.csv)

	oda	bulundugukat	kira	isi	balkon	esya	lokasyon	see
0	2	1	2650	5	1	0	6	20
1	2	5	950	1	2	0	0	65
2	4	7	2350	3	1	0	3	45
3	3	13	1150	2	2	0	1	30
4	5	15	2550	3	3	0	6	25
5	2	16	1850	3	1	1	3	25
6	2	14	2250	3	1	0	4	10
7	1	1	1050	3	0	0	5	60
8	3	3	2450	3	2	2	6	10
9	5	1	1750	4	3	1	3	25
10	5	2	2200	3	4	2	3	25

Resim 3-Veriseti 3 (data3.csv)

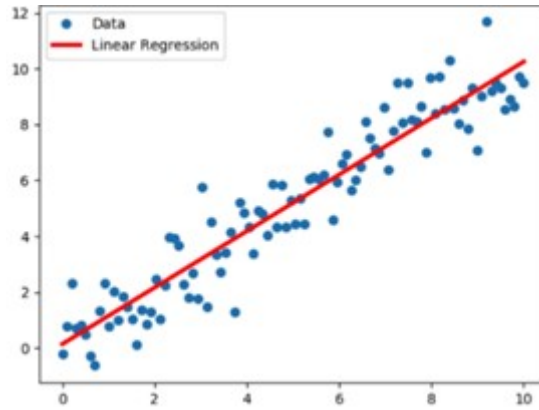
3. BÖLÜM 2

ÇÖZÜME UYGUN YAKLAŞIM BELİRLEME

DOĞRUSAL REGRESYON

Doğrusal regresyon, iki veya daha fazla değişken arasındaki ilişkiyi tanımlamak için kullanılan doğrusal bir model yaklaşımıdır. Basit doğrusal regresyonda bağımlı değişken ve bağımsız değişken olmak üzere 2 değişken vardır. İki tür doğrusal regresyon modeli vardır. Bunlar basit regresyon ve çoklu regresyondur.

- Basit doğrusal regresyon, bağımlı bir değişkeni tahmin etmek için bir bağımsız değişken kullanıldığında gerçekleşir.
- Birden fazla bağımsız değişken mevcut olduğunda, işleme çoklu doğrusal regresyon adı verilmektedir.



Resim 4- Doğrusal regresyon

Bağımsız değişkenlerin dağılımına bağlı olarak y bağımlı değişkeni etkilenecektir. Bu ilişkisel durum bir doğru grafiğine oturtulursa

$$y=ax+b$$

şeklindeki fonksiyon ile doğrusal ilişki açıklanıp çözüme erişilir. Burada;

- x:bağımsız değişken
- y: bağımlı değişken (hedef, sonuç)
- a:eğim
- b:doğrunun ne kadar kaydırılacağı

Problem çözümünde kullanılmak üzere doğrusal regresyon seçilmiştir. Bu analiz biçimi, bağımlı değişkenin değerini en iyi öngören bir ya da daha fazla bağımsız değişkeni kullanarak doğrusal denklemin katsayılarını tahmin eder. Doğrusal regresyon modelleri nispeten basittir ve tahminler üretebilen, yorumlanması kolay bir matematiksel formül sağlar. Doğrusal regresyon, iş dünyasında ve akademik çalışmalarda çeşitli alanlara sıklıkla uygulanmaktadır. Ayrıca problemin çözümünde sürekli bir değer beklenildiği içinde uygun olduğu düşünülmüştür.

EN KÜÇÜK KARELER YÖNTEMİ

Tahmin edilen katsayıların ve değişkenlerin hesaplanması için geliştirilen ve gerçek katsayılara en yakın sonuçları veren yöntemlerden bir tanesi En Küçük Kareler (EKK) yöntemidir. Bir diğer ifade ile regresyon modelindeki hata terimini minimum yapan yöntem EKK yöntemidir.

$y = a + bx$ doğrusu üzerindeki (x_i, y) noktaları ile verilen (x_i, y_i) serpiye noktaları arasındaki uzaklıkların kareleri toplamını minimum yapan a ve b katsayılarını bulma işleminden ibarettir. Bu katsayılar bulununca, $y = a + bx$ doğrusu (regresyon doğrusu) bulunmuş olur.

a ve b katsayıları şu formüllerle bulunur:

$$\hat{a} = \bar{Y} - \hat{b}\bar{X}$$

•

$$\hat{b} = \frac{\sum_{i=1}^n (Y_i - \bar{Y}) \cdot (X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

PYTHON İLE VERİ ANALİZİ

Veri Analizi

Bu çalışmada 2:53 GHz Intel Core i5-M460 işlemci, 3GB RAM, 500GB SSD bellek ve xubuntu işletim sistemine sahip sistem ve Anaconda dağıtımı üzerinden jupyter nootebook kullanılarak Python programlama dili kullanılmıştır.

Veriseti Python ile analiz edileceği için veriseti pandas ile okutularak dataframe şeklinde programa eklenip üzerinde işlemler gerçekleştirilmiştir.

Önemli istatistikler

	metrekare	oda	binayasi	bulundugukat	site	kira	ısı	balkon	banyo	tip	esya	lokasyon	see
count	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000
mean	197.857143	3.428571	11.323810	7.000000	0.523810	1902.857143	3.038095	2.076190	1.371429	0.361905	0.228571	3.342857	28.152381
std	63.763600	1.343482	7.445307	5.550468	0.501828	706.012250	0.634335	0.957714	0.523618	0.810078	0.421927	1.920680	18.018036
min	80.000000	1.000000	1.000000	1.000000	0.000000	850.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	4.000000
25%	150.000000	2.000000	4.000000	2.000000	0.000000	1350.000000	3.000000	2.000000	1.000000	0.000000	0.000000	2.000000	20.000000
50%	200.000000	4.000000	11.000000	5.000000	1.000000	1800.000000	3.000000	2.000000	1.000000	0.000000	0.000000	3.000000	22.000000
75%	240.000000	4.000000	16.000000	12.000000	1.000000	2400.000000	3.000000	3.000000	2.000000	0.000000	0.000000	5.000000	30.000000
max	400.000000	6.000000	30.000000	19.000000	1.000000	4100.000000	5.000000	5.000000	3.000000	3.000000	1.000000	6.000000	100.000000

Resim 5- istatistikler

Eğitim ve test ayrımı

Veri kümesinde 105 örnek olduğu için %80 eğitim, %20 test için ayrılmıştır. 84 örnek eğitim için kullanılmış, 21 örnek test için kullanılmıştır.

Model Geliştirilmesi

Eğitim veriseti kullanılarak doğrusal regresyon ile model geliştirilmiştir. 0.47 başarı oranı elde edilmiştir.

Tahmin

Test verisinden geliştirilen model kullanılarak tahmin yapılmıştır.

```
y_test [[25],[10],[45],[10],[20],[20],[17],[20],[20],[20],[20],[20],  
[22],[40],[60],[25],[30],[20],[20],[10],[25]]
```

```
Tahmin [[20.92092217],[41.03184783],[33.09037289],[16.66438646],  
[31.90140059],[29.3673973],[31.82020325],[12.75046328],[14.54410829],  
[30.6728571],[26.07109307],[23.81403749],[33.55480376],[39.49093009],  
[42.88874868],[12.03389079],[52.45473292],[37.23694734],[39.2886704],  
[20.48690707],[28.3278486]]
```

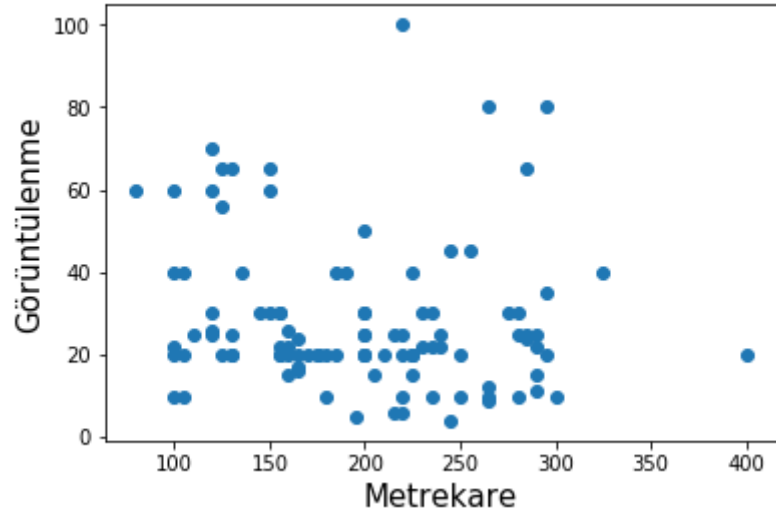
y_test	Tahmin
25	21
10	41
45	33
10	17
20	32

20	29
17	32
20	13
20	15
20	31
20	26
20	24
22	34
40	39
60	43
25	12
30	52
20	37
20	39
10	20
25	28

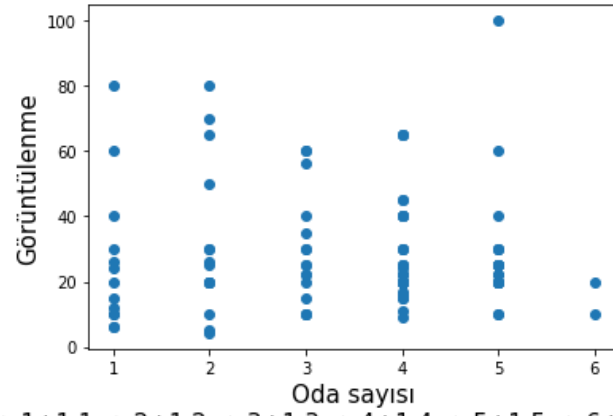
Grafikler

Dağılım Grafikleri

Bağımsız değişkenlerin bağımlı değişkene olan etkilerini gözlemlemeyi sağlar.

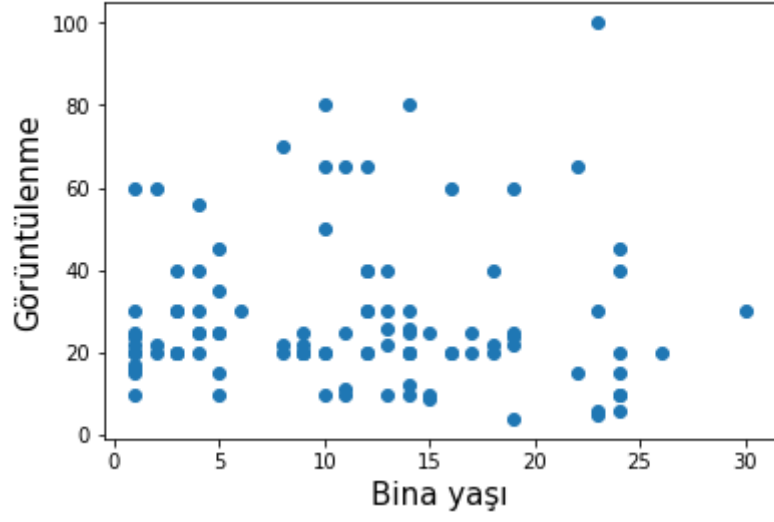


Resim 6- Metrekare-görüntülenme ilişkisi

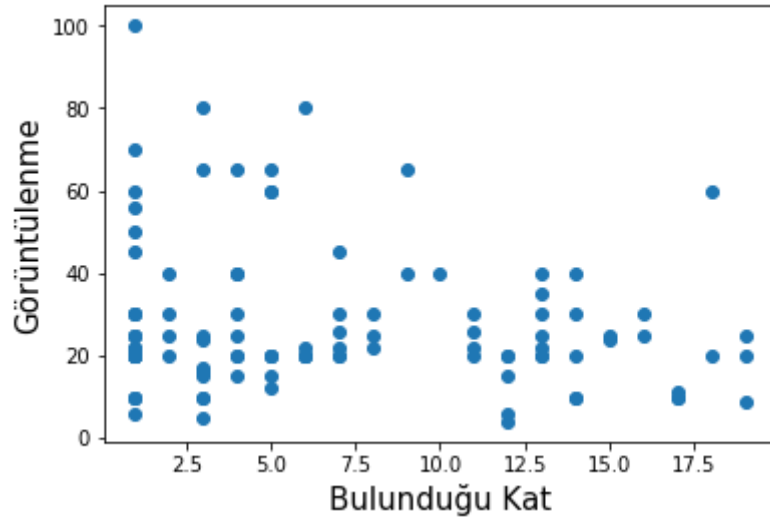


0=>1+1 1=>2+1 2=>3+1 3=>4+1 4=>5+1 5=>6+1 6=>7+1

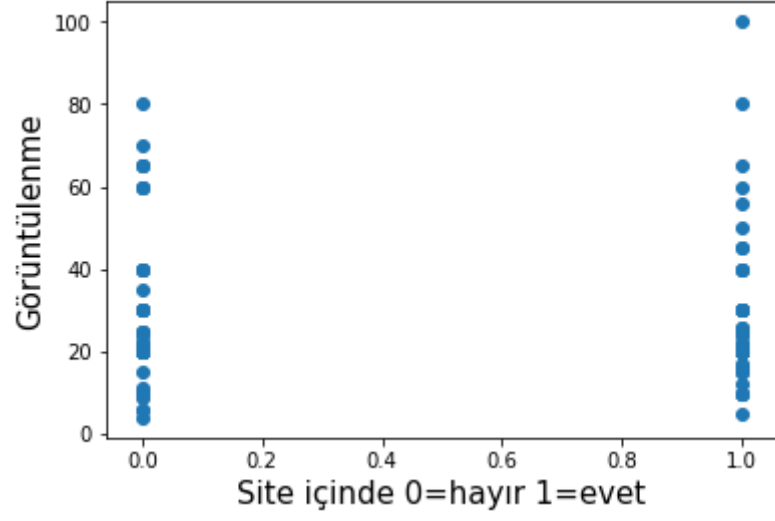
Resim 7-Oda sayısı görüntülenme ilişkisi



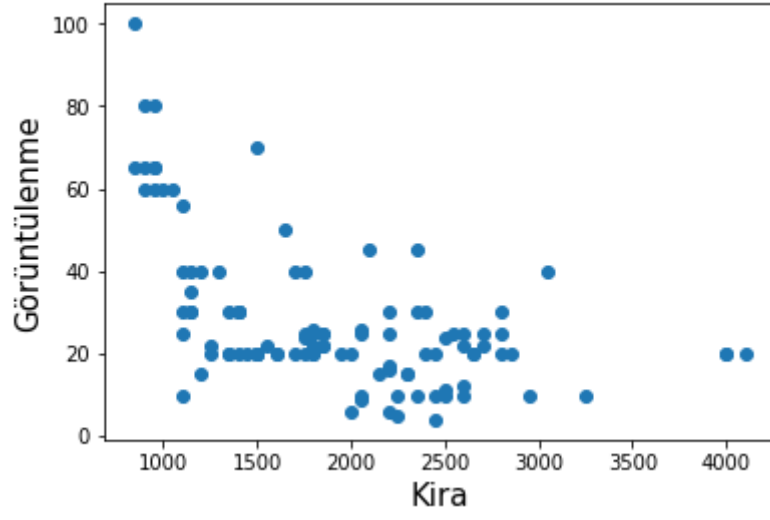
Resim 8- Bina yaşı görüntülenme ilişkisi



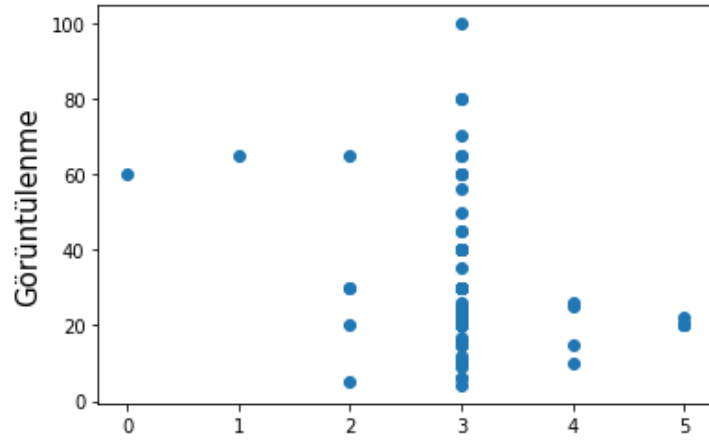
Resim 9- Bulunduğu kat görüntülenme ilişkisi



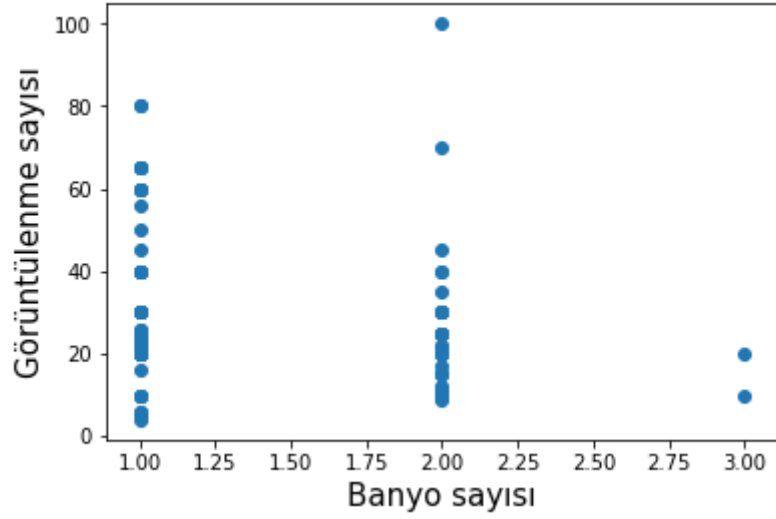
Resim 10- Site durumu görüntülenme ilişkisi



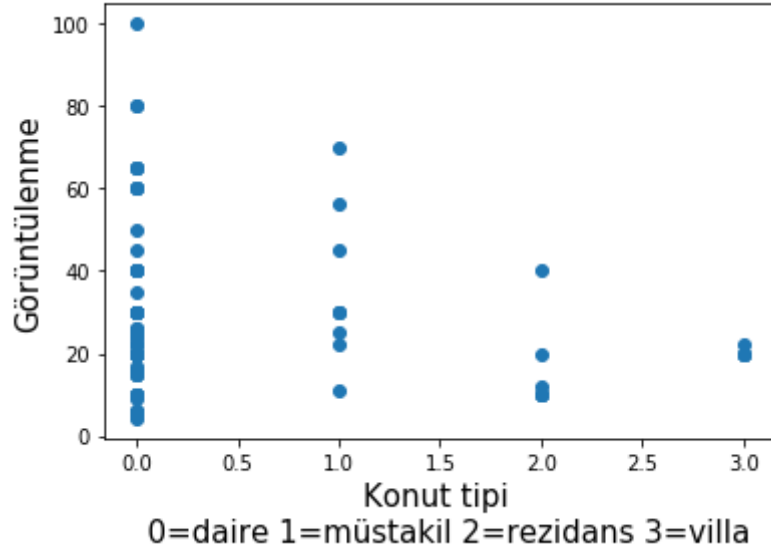
Resim 11-Kira görüntülenme ilişkisi



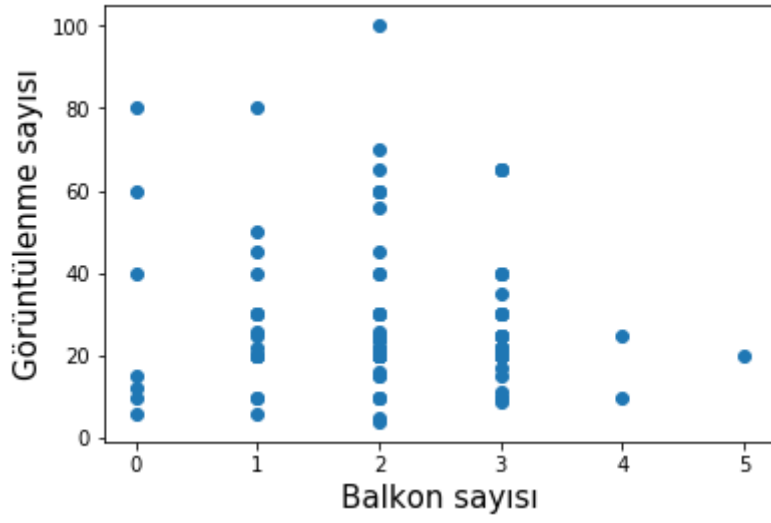
0=soba 1=doğalgaz sobası 2=merkezi 3=doğalgaz kombi 4=klima 5=şömine
Resim 12-Isınma durumu görüntülenme ilişkisi



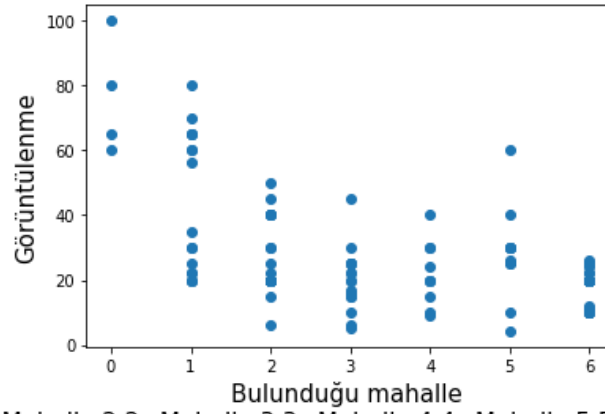
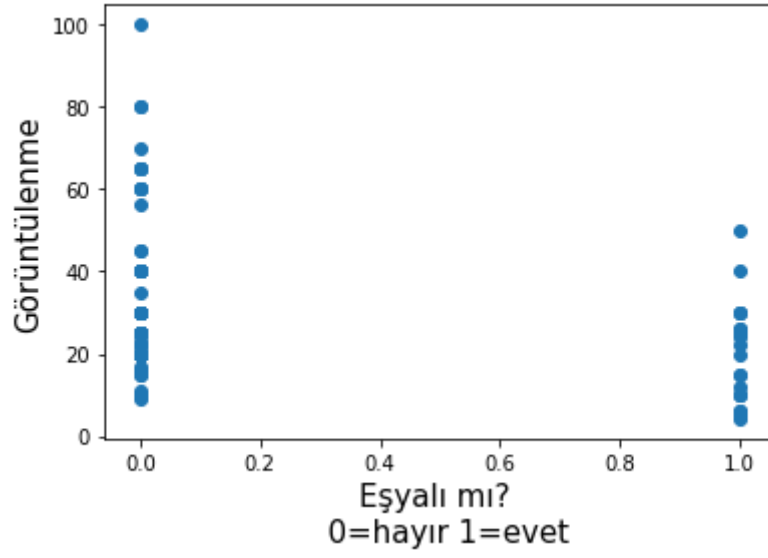
Resim 13-Banyo sayısı görüntülenme ilişkisi



Resim 14-Konut tipi görüntülenme ilişkisi

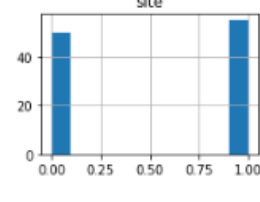
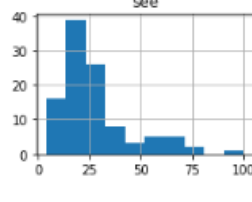
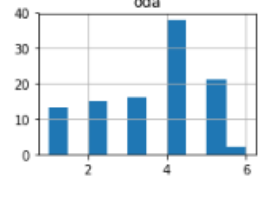
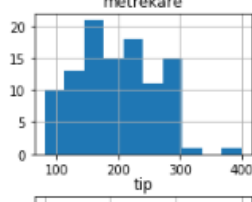
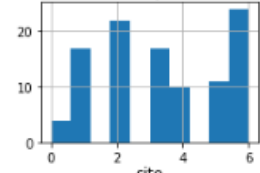
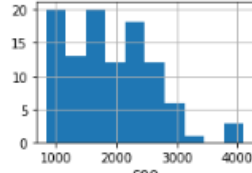
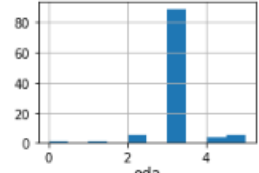
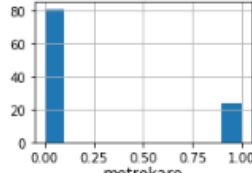
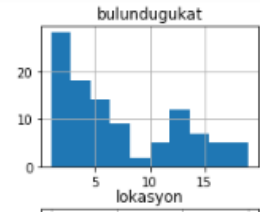
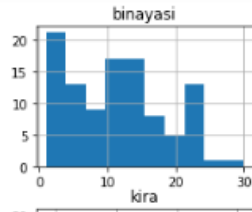
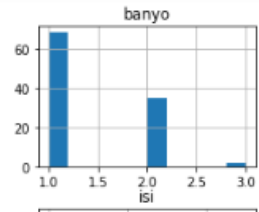
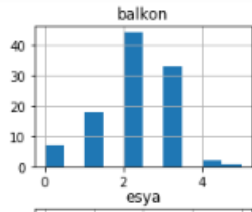


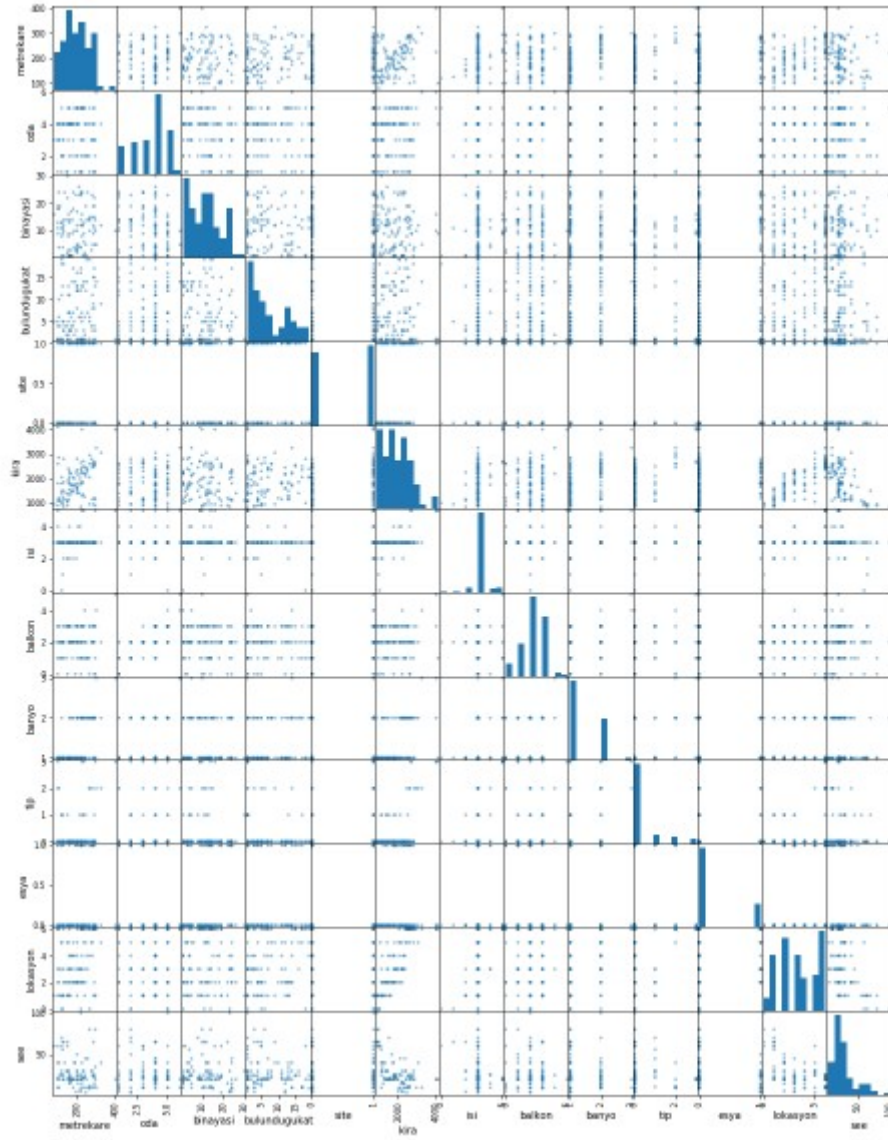
Resim 15-Balkon sayısı görüntülenme ilişkisi



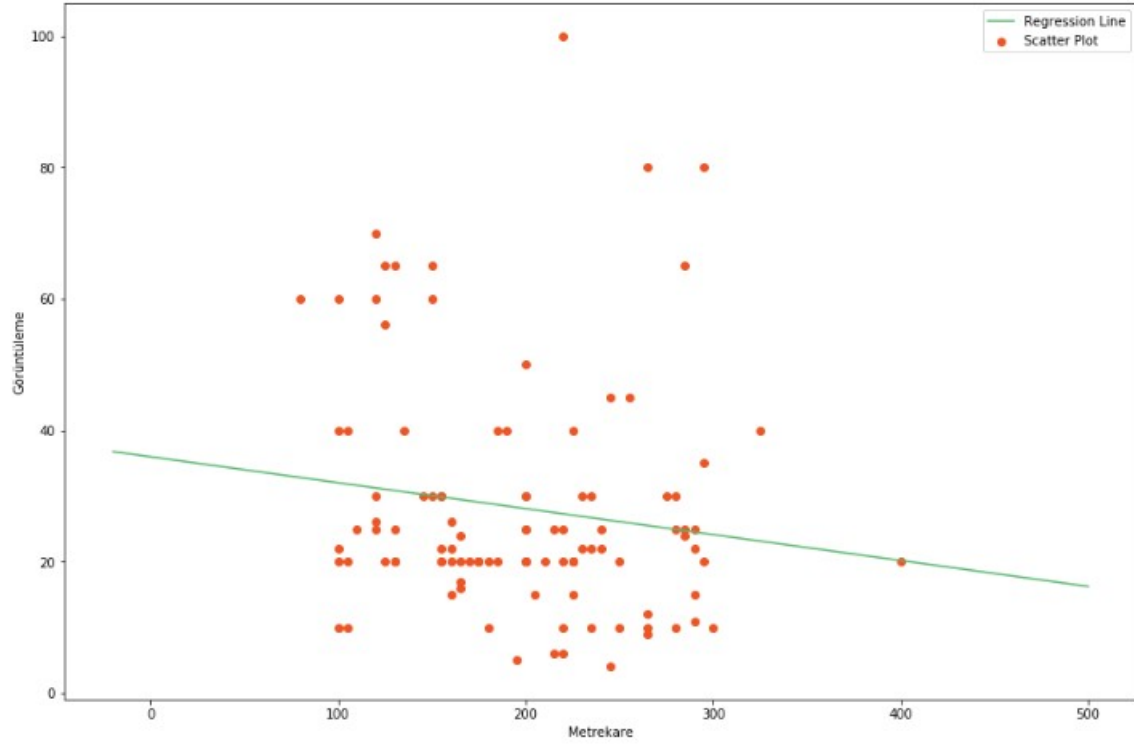
0=Mahalle 1 1=Mahalle 2 2=Mahalle 3 3=Mahalle 4 4=Mahalle 5 5=Mahalle 6 6=Mahalle 7

Histogram Grafikleri

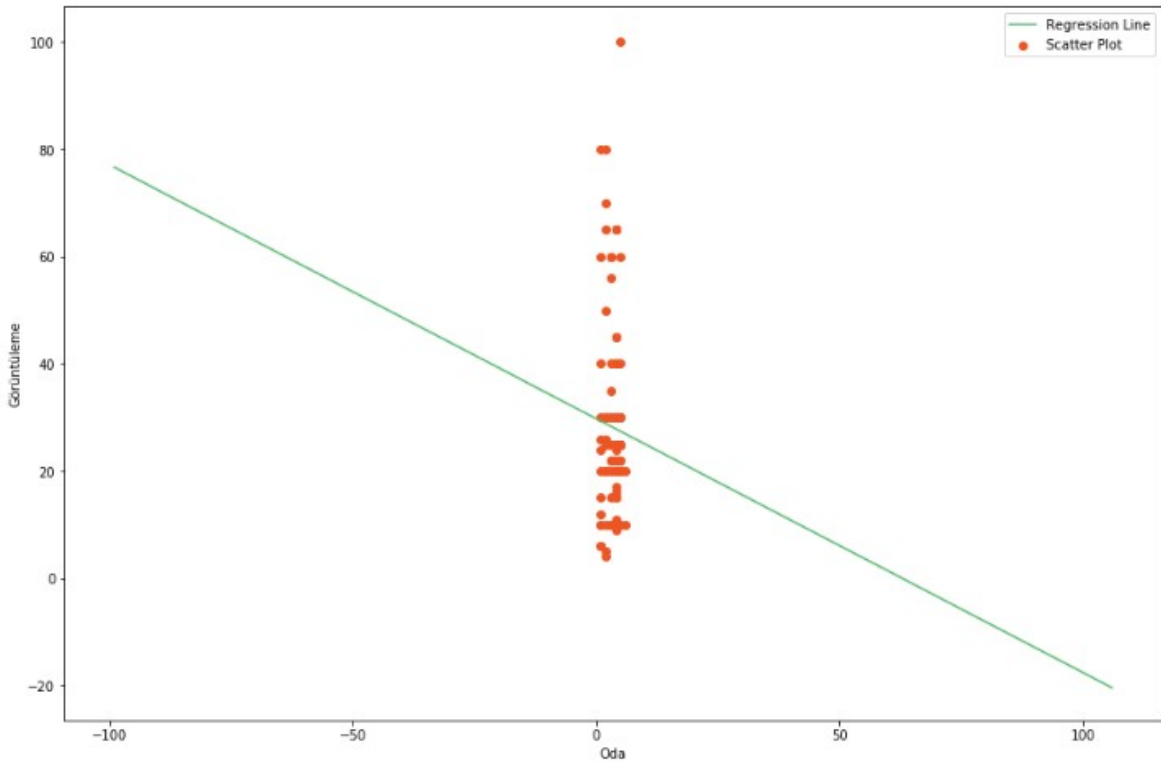




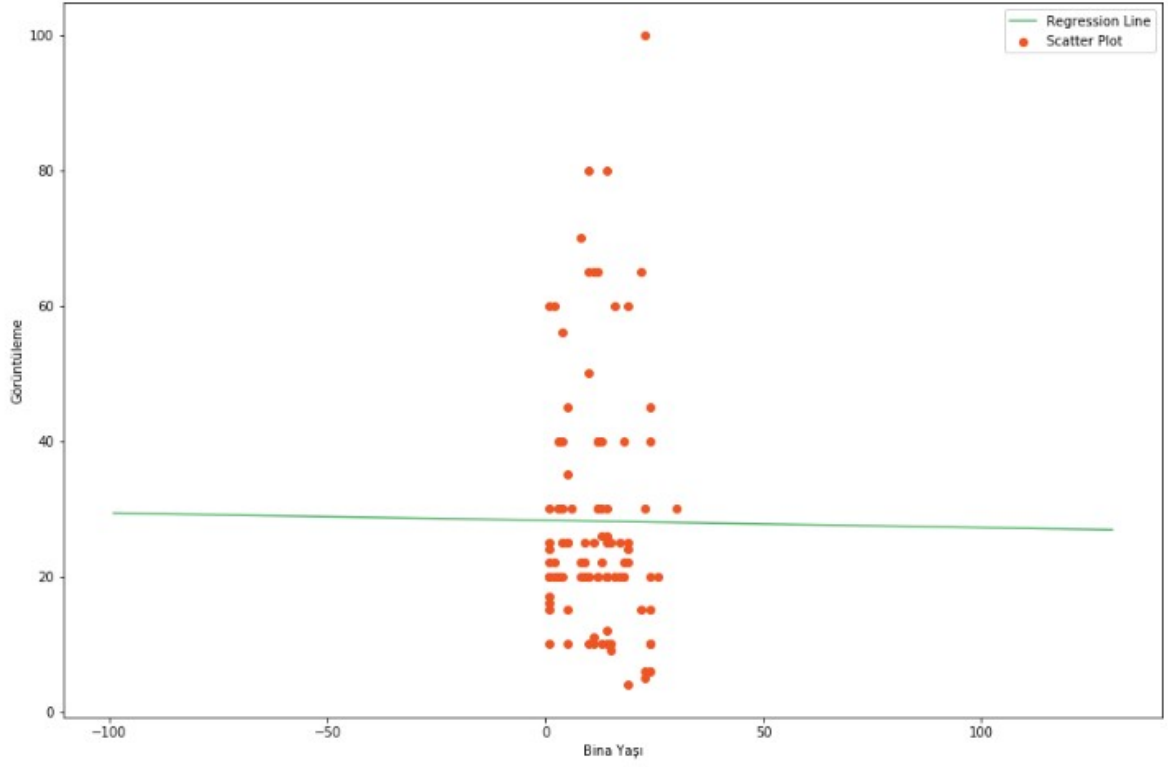
EKK grafikleri



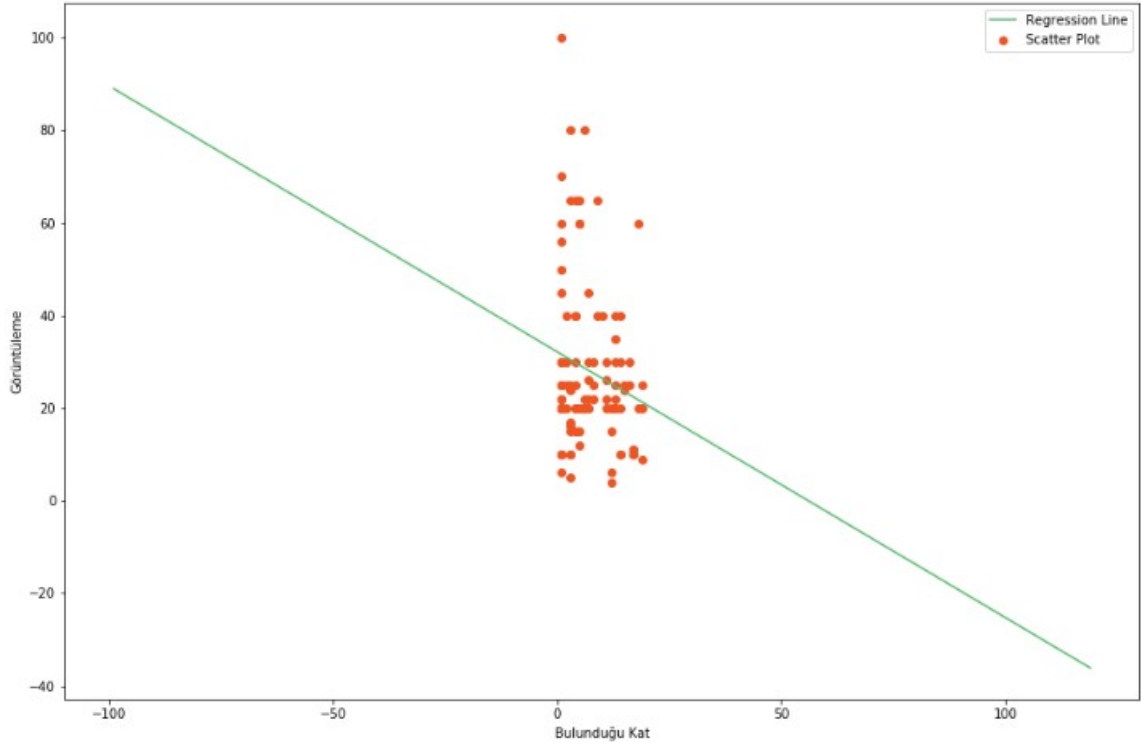
Metrekare ile görüntülenme negatif ilişkidir.



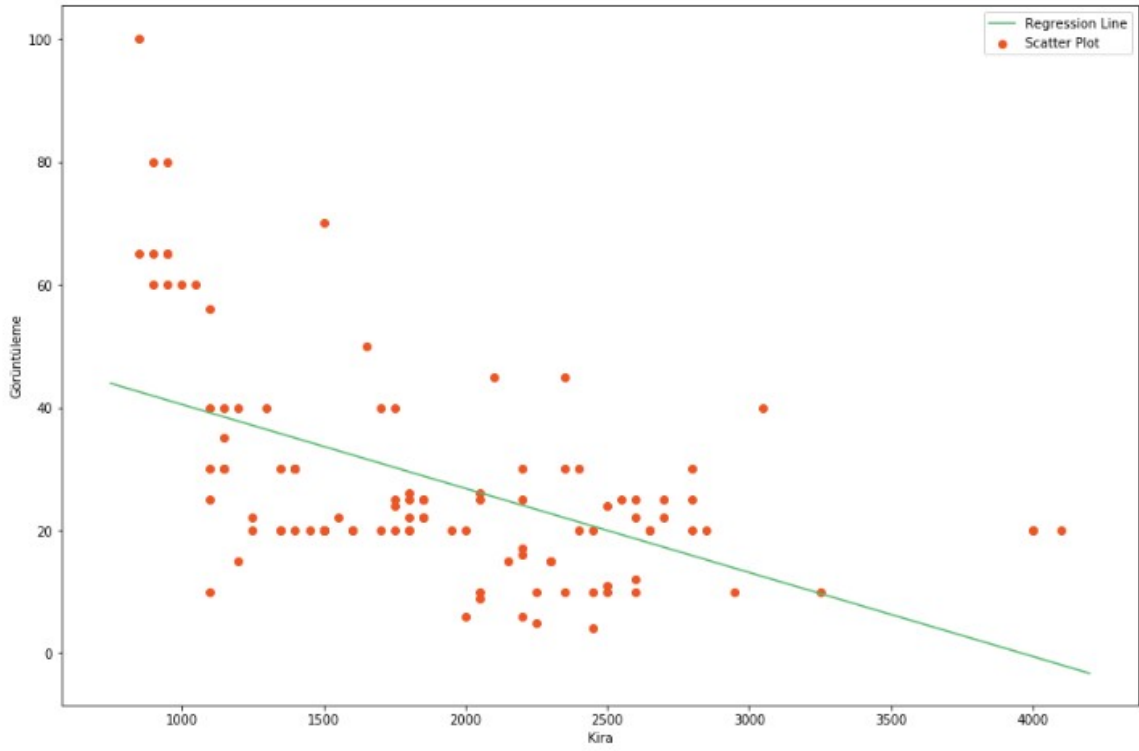
Oda sayısı ile görüntülenme negatif ilişkidir.



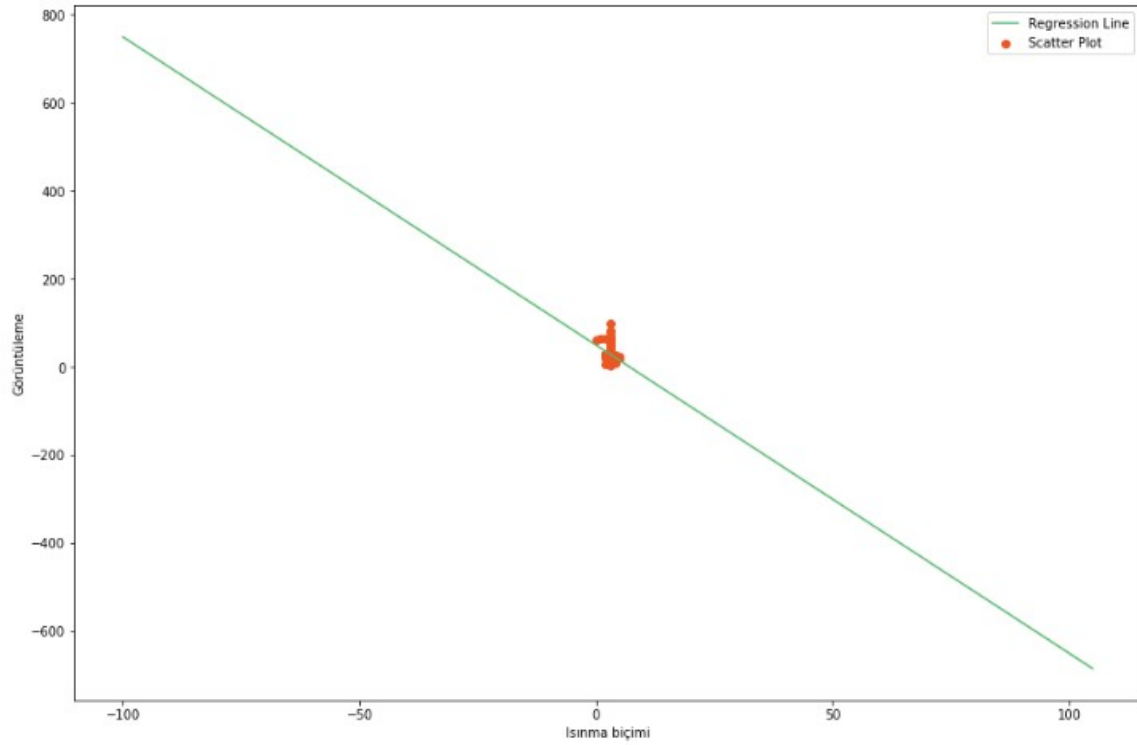
Bina yaşı ile görüntülenme hemen hemen etkisizdir



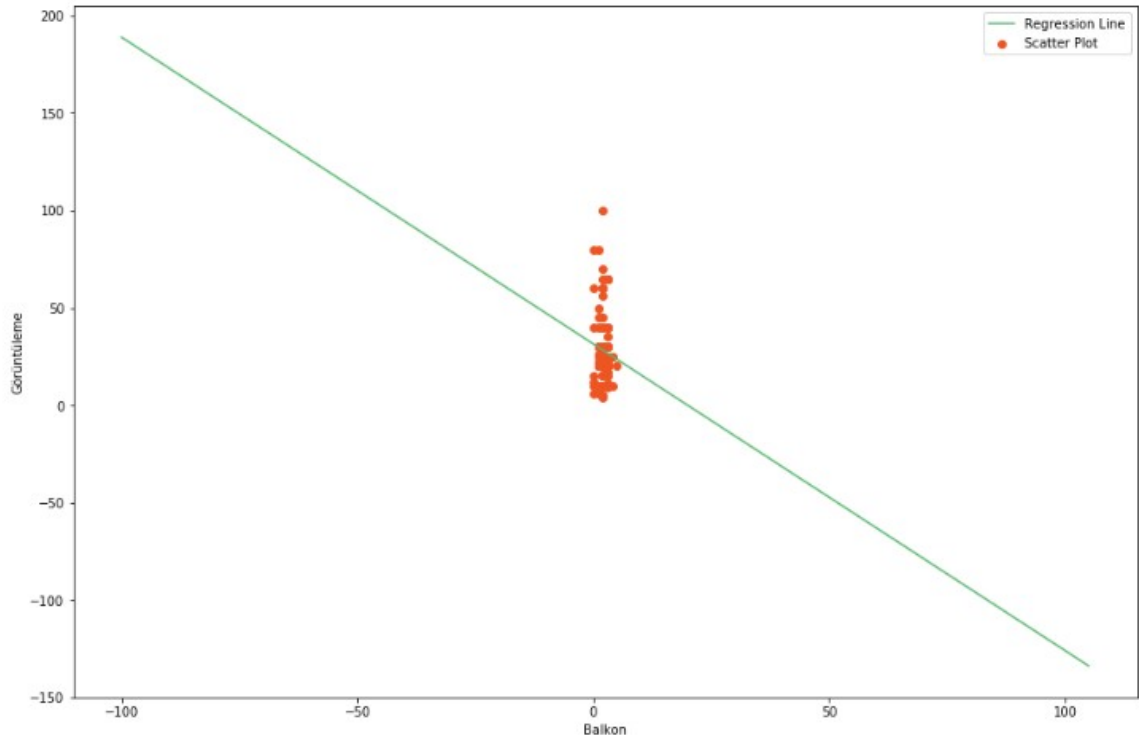
Bulunduğu kat ile görüntülenme negatif ilişkidir.



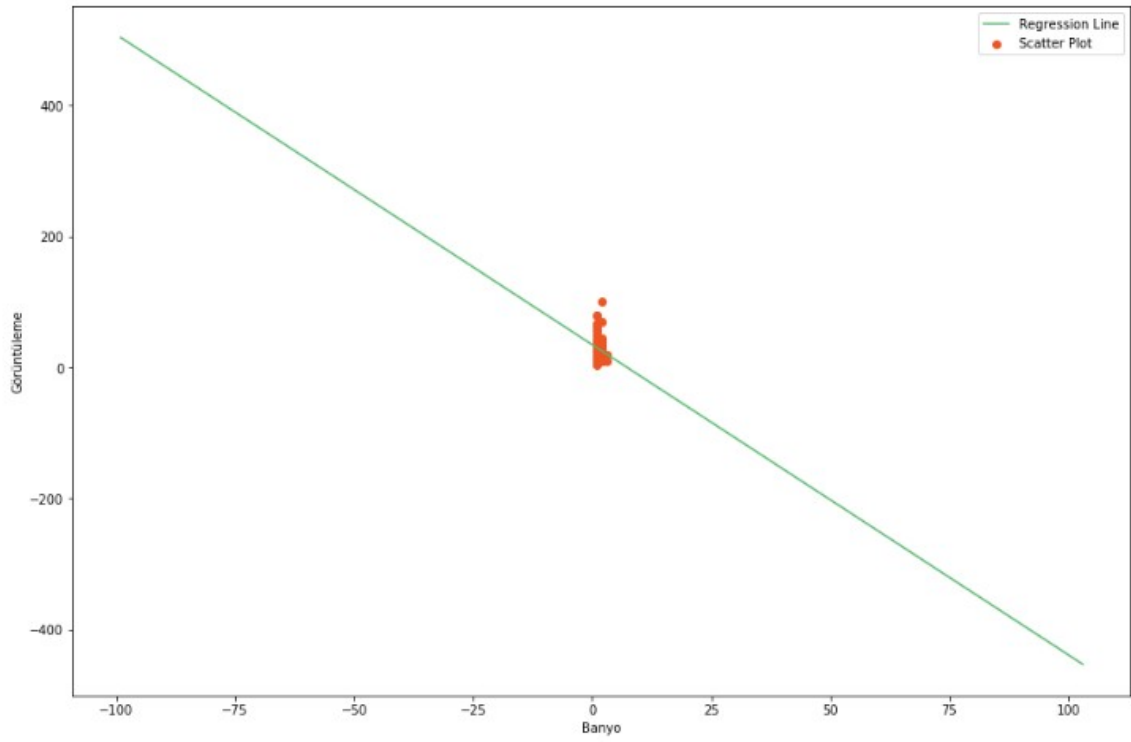
Kira ile görüntülenme negatif ilişkilidir.

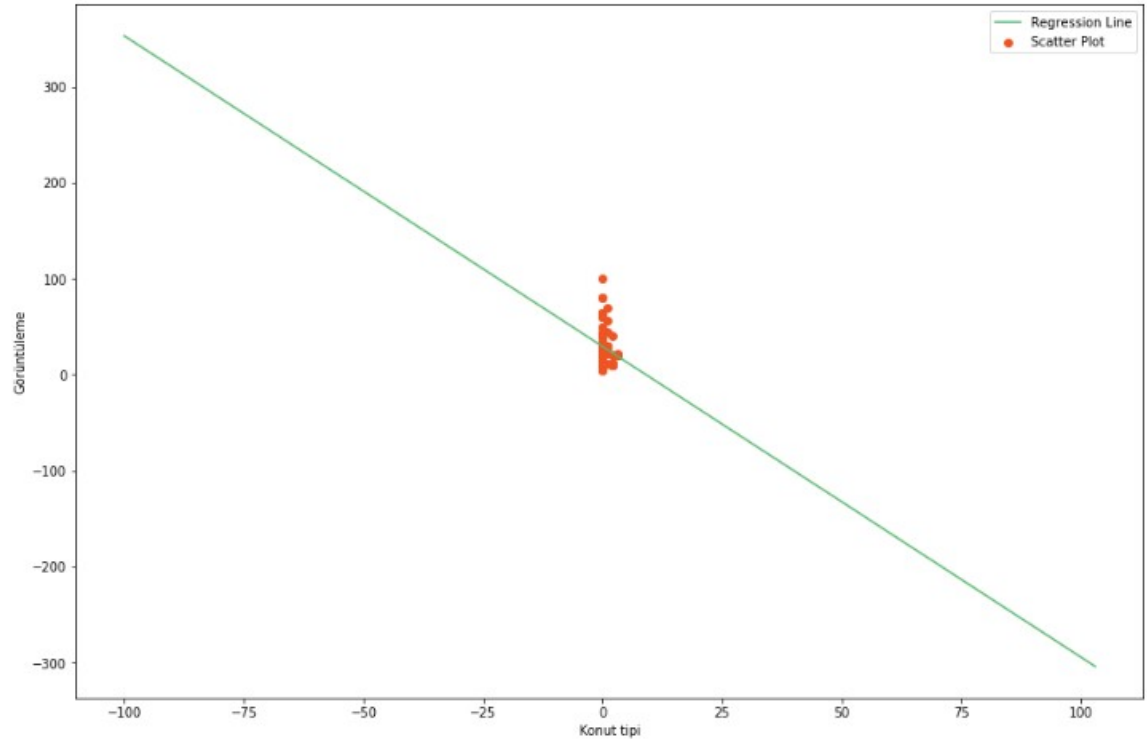


Isınma biçimi ile görüntülenme negatif ilişkilidir.

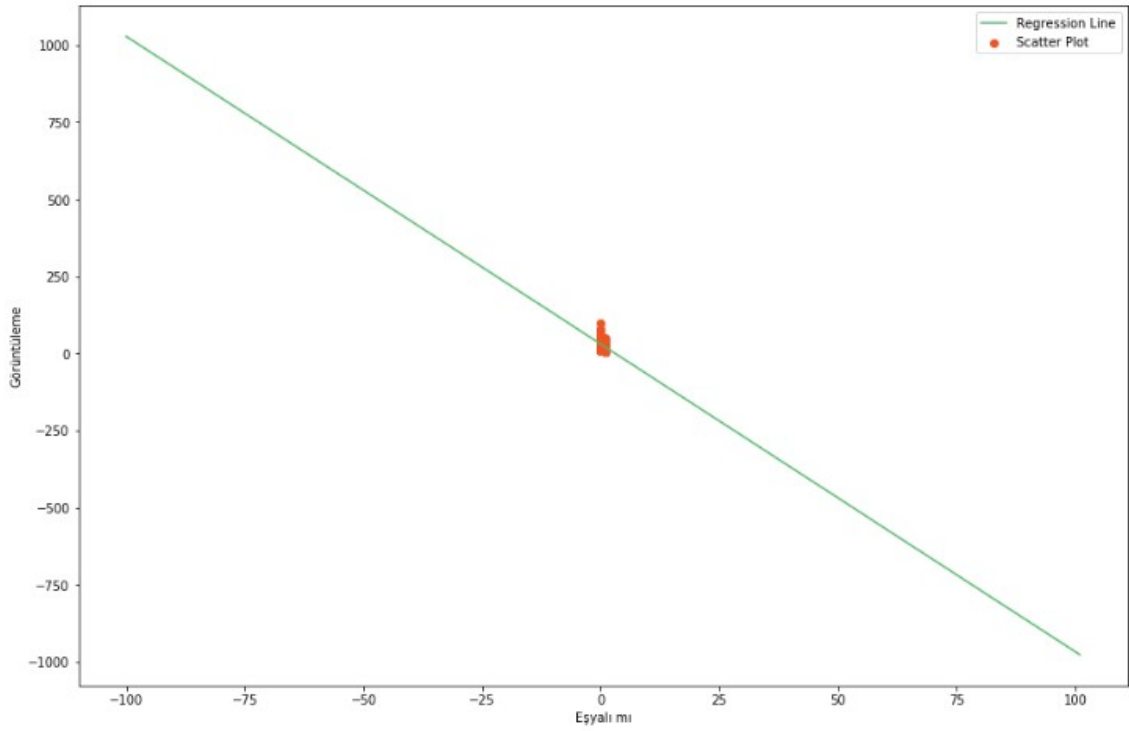


Balkon ile görüntülenme negatif ilişkilidir.

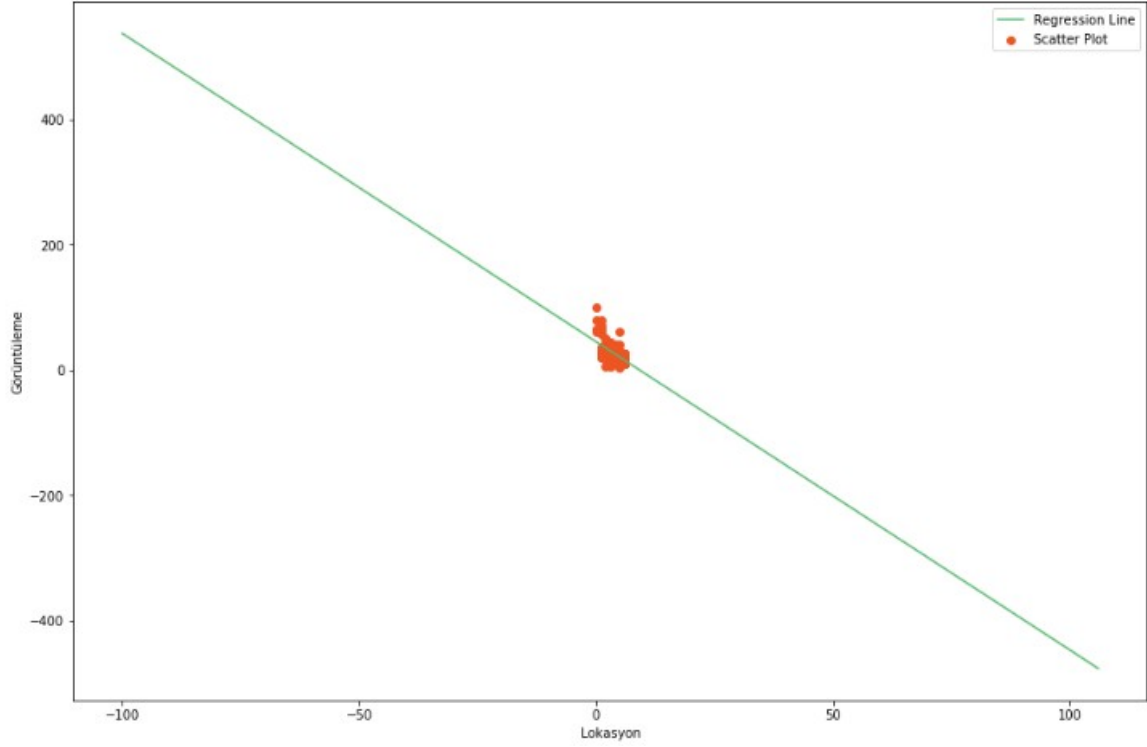




Konut tipi ile görüntülenme negatif ilişkilidir.



Banyo sayısı ile görüntülenme negatif ilişkilidir.



Lokasyon ile görüntülenme negatif ilişkilidir.

SONUÇ

Bu çalışmada emlak ilanlarından görüntülenme sayısını tahmin etme problemi için veriseti oluşturularak doğrusal regresyon ile çözüm geliştirilmiştir. Model başarımları % 47 bulunmuştur. Başarımları ortalama bulunmuştur. Verinin tamamıyla gerçek hayattan alınamaması buna sebep olmuş olabilir. Ayrıca bazı özellikler başarımları olumsuz etkilemiş olabilir. Gerçek veri ile çalışma ve özellik seçimi ile başarımları artırılabilir. Müşterinin ilgisinin belirlenmesi, ilgisine yönelik konutlarda güncellemelerin ev sahibine yaptırılması, ön plana çıkamayan ilanların öneri sistemleri ile müşteriye sunulması arz-talep ilişkisini güçlendirecektir. Bu konuda çalışmaya devam edilmesi firmanın oldukça faydasına olacaktır.

KAYNAKLAR

- [1] Data Mining Concepts and Techniques ,Third Edition, Jiawei Han University of Illinois at Urbana–Champaign Micheline Kamber Jian PeiSimon Fraser University, 2012 by Elsevier Inc (data mining)
- [2] Russell , S. J., & Norvig, P. (2010). Artificial Intelligence A Modern Approach. Upper Saddle River, New Jersey 07458: Pearson Education, Inc.
- [3] Tom Mitchell , E. Fredkin University Professor <https://www.cs.princeton.edu/courses/archive/spr07/cos424/papers/mitchell-dectrees.pdf>
- [4] Broker, E. (n.d.). *Makine Öğrenimi Emlak Sektörüne mi Yöneliyor? | Emlak Broker*. Retrieved December 19, 2020, from <https://www.emlakbroker.com/haberler/makine-ogrenimi-emlak-sektorune-mi-yonel.html>
- [5] *Gayrimenkul Danışmanları Yapay Zekâdan Yararlanabilir mi?* (n.d.). Retrieved December 19, 2020, from <https://tresmo.com/blog/368/gayrimenkul-danismanlari-yapay-zekadan-yararlanabilir-mi>
- [6] Selim, S., & Demirbilek, A. (2009). Türkiye’deki Konutların Kira Değeri nin Analizi: Hedonik Model ve Yapay Sinir Ağları Yaklaşımı. Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 1(1), 73-90.
- [7] GEDİK, H., & ETLİOĞLU, M. (2017). İNTERNETTE EMLAK PAZARLAMASI. Electronic Turkish Studies, 12(32).