



WOB INC

Paris Konut Fiyat Tahmini

Hazırlayanlar

Türker Bayazıt

Kübra Karaalp

Zeynep Ecem Güneş

Amaç

Bu projedeki amacımız Paris'te yaşayan kişilerin ev satın alırken dikkat ettiği kriterlere göre analiz yapıp tercih ettikleri kriterlerle ev satış fiyatlarını incelemektir.

Giriş

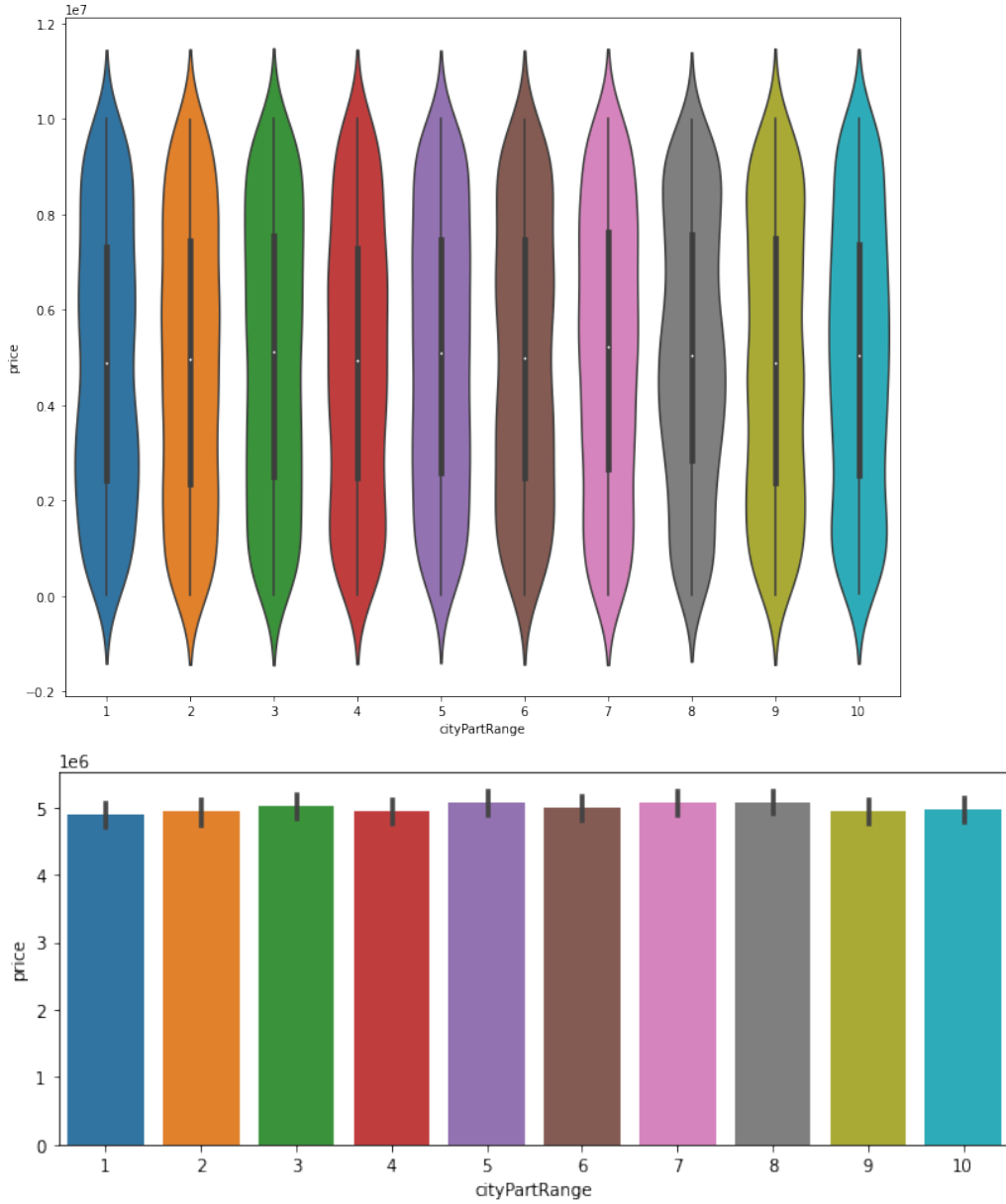
Bu projede Paris'te yaşayan kişilerin hangi evleri daha çok tercih ettiğini, ev alırken aradıkları özellikleri görmek istiyoruz. İnsanların ev satın alırken çeşitli tercihleri olur. Bu kriterler havuzunun olması, çok katlı olması, kaç yılında yapılmış olması olabilirken bazı insanlar için kentten daha uzak bir evde ona göre malzeme kullanılmış olması olabilir.

Biz de insanların ev satın almalarına hangi kriterlerin etkili olduğunu görüp buna göre bir çalışma yapmak istiyoruz. Bu projede Paris'te yaşayan kişilerin hangi evleri daha çok tercih ettiğini, ev alırken aradıkları özellikleri görmeyi amaçlıyoruz. Verimiz Paris'teki konut fiyatlarında oluşan hayali bir veri setidir. Yapmış olduğumuz veri analizleri sonucunda veri seti için birden fazla regresyon modeli ile eğittiğimiz veri setini test edip en uygun regresyon modelini bulmayı amaçlıyoruz.

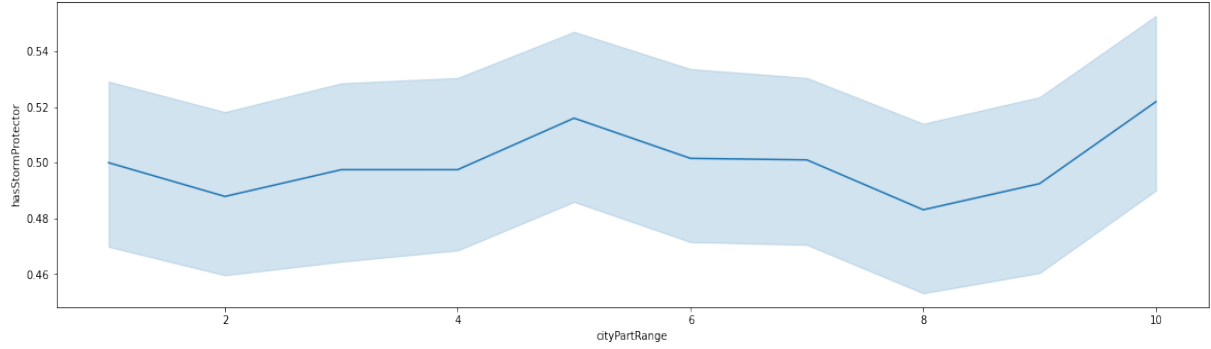
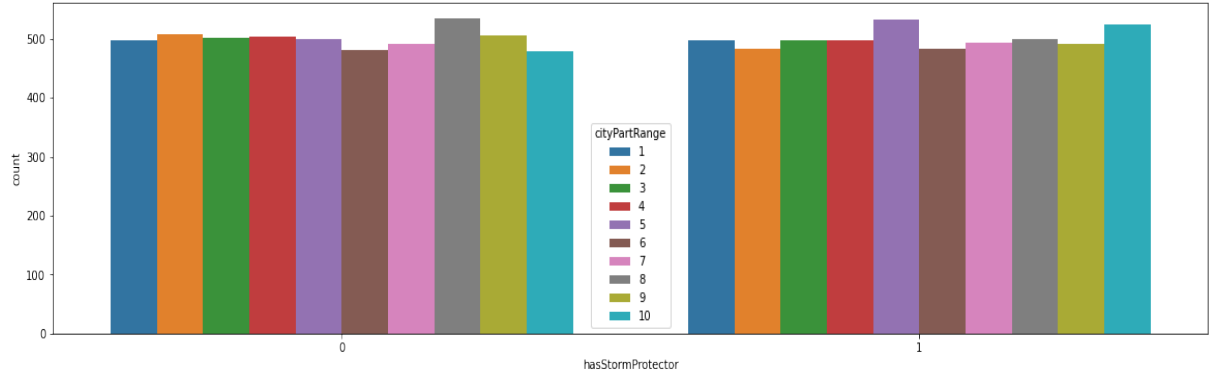
	count	mean	std	min	25%	50%	75%	max
squareMeters	10000.0	4.987013e+04	2.877438e+04	89.0	25098.50	50105.5	74609.75	99999.0
numberOfRooms	10000.0	5.035840e+01	2.881670e+01	1.0	25.00	50.0	75.00	100.0
hasYard	10000.0	5.087000e-01	4.999493e-01	0.0	0.00	1.0	1.00	1.0
hasPool	10000.0	4.968000e-01	5.000148e-01	0.0	0.00	0.0	1.00	1.0
floors	10000.0	5.027630e+01	2.888917e+01	1.0	25.00	50.0	76.00	100.0
cityCode	10000.0	5.022549e+04	2.900668e+04	3.0	24693.75	50693.0	75683.25	99953.0
cityPartRange	10000.0	5.510100e+00	2.872024e+00	1.0	3.00	5.0	8.00	10.0
numPrevOwners	10000.0	5.521700e+00	2.856667e+00	1.0	3.00	5.0	8.00	10.0
made	10000.0	2.005488e+03	9.308090e+00	1990.0	1997.00	2005.5	2014.00	2021.0
isNewBuilt	10000.0	4.991000e-01	5.000242e-01	0.0	0.00	0.0	1.00	1.0
hasStormProtector	10000.0	4.999000e-01	5.000250e-01	0.0	0.00	0.0	1.00	1.0
basement	10000.0	5.033104e+03	2.876730e+03	0.0	2559.75	5092.5	7511.25	10000.0
attic	10000.0	5.028011e+03	2.894332e+03	1.0	2512.00	5045.0	7540.50	10000.0
garage	10000.0	5.531212e+02	2.620502e+02	100.0	327.75	554.0	777.25	1000.0
hasStorageRoom	10000.0	5.030000e-01	5.000160e-01	0.0	0.00	1.0	1.00	1.0
hasGuestRoom	10000.0	4.994600e+00	3.176410e+00	0.0	2.00	5.0	8.00	10.0
price	10000.0	4.993448e+06	2.877424e+06	10313.5	2516401.95	5016180.3	7469092.45	10006771.2

Yöntem

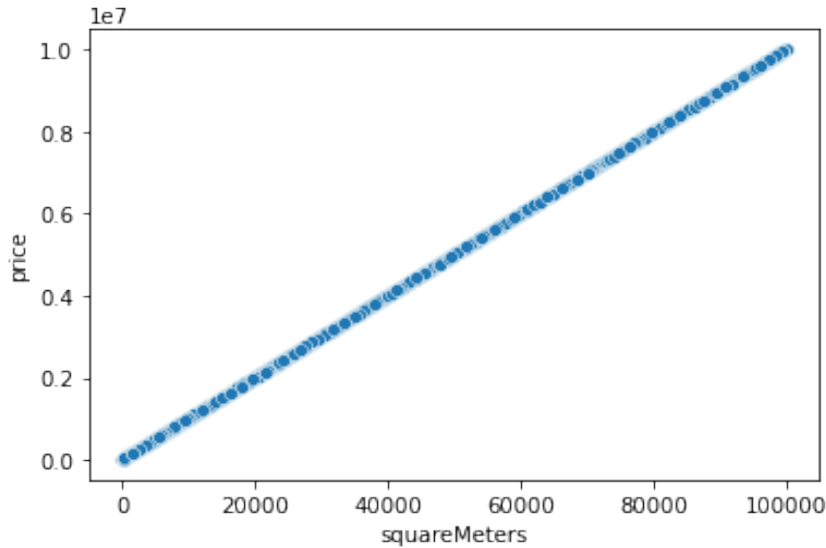
Projede ilk olarak verimizi tanıdıktan sonra analize başlandı. Verideki değişkenlerin kendi aralarında ilişkilerini inceledik bunları yaparken çeşitli görselleştirme seçenekleri için birçok farklı grafik kullandık. Veriyi daha iyi anlamak adına box plot, scatter plot, join plot, histogram, bar plot, line plot gibi birden fazla grafik kullanıldı.



Veri seti analizimizin bir parçası olan bu grafikler violin plot ve bar plot. Grafiklere baktığımızda şehrin uzaklığının ayrıcalıklı olması fiyatta bir değişikliğe neden olmamıştır. Görmekteyiz ki şehir uzaklığı insanların ev alırken fiyatında bir değişiklik olması için önemli bir kriter değildir.



Bu grafiklerde şehrin uzaklığına bağlı olarak evlerin fırtına koruyucu sahip olup olmamasına bakıyoruz. Şehir uzaklığının evlerde fırtına koruyucu olmasına ya da olmamasına bir etkisi olmadığını görüyoruz. Fırtına koruyucu olan evler ve olmayan evlerin sayısı arasında çok ciddi bir fark görülmemekte.



Grafikte ev satış fiyatı evin metrekaresi arasındaki ilişki görüyoruz. Ev satış fiyatı ve evin metrekaresi arasında doğrusal bir ilişki olduğunu söyleyebiliriz. Yani evin metrekaresi arttıkça evin satış fiyatı da aynı şekilde bir artış gösteriyor.

Sonuç

Uygulamış olduğumuz olduğumuz regresyon modellerinin sonuçlarına bakacak olursak;

Linear Regression

```
Mean Squared Error : 3704307.671084843
Root Mean Squared Error : 1924.6578062307187
R-Square : 0.9999995521735113
```

R-Square değerini baktığımızda modelimizin 1'e çok yakın yani çok iyi bir model olduğunu görüyoruz.

Ridge Regression

```
Mean Squared Error : 30155434.41168997
Root Mean Squared Error : 5491.39639906736
R-Square : 0.9999963544058689
```

Ridge regresyon modelinin de iyi bir model olduğunu gözlemliyoruz.

Lasso Regression

```
Mean Squared Error : 3702061.1219909205
Root Mean Squared Error : 1924.074094724764
R-Square : 0.9999995524451044
```

Lasso regresyon modelimiz iyi bir modeldir.

Elastic Net Regression

```
Mean Squared Error : 6085755507993.838
Root Mean Squared Error : 2466932.4084769404
R-Square : 0.26427209567500975
```

Bu regresyon modelimiz değerlerimize baktığımızda bize iyi sonuçlar veren bir model değildir. Hata kareler ortalaması da çok yüksektir.

KNeighbors Regressor

```
Mean Squared Error : 1604736347133.637
Root Mean Squared Error : 1266781.886172058
R-Square : 0.8059979064029191
```

Modelimiz iyi bir modeldir diyebiliriz.

Support Vector Regressor

Mean Squared Error : 8272133225288.896
Root Mean Squared Error : 2876131.6425520056
R-Square : -4.662266578270291e-05

Model kötü bir modeldir. Kesinlikle kullanılamaz.

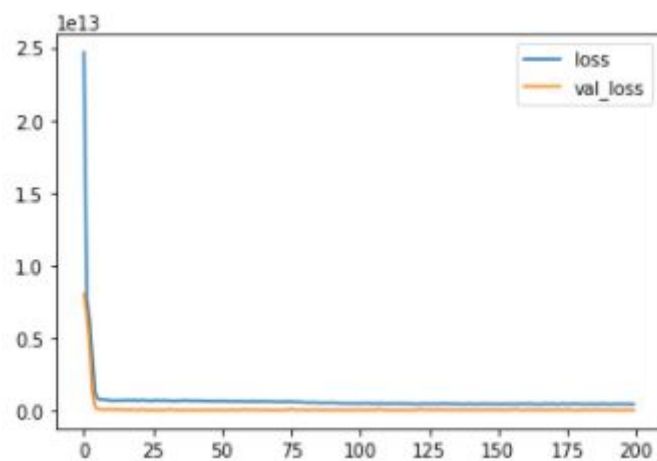
Random Forest Regressor

Mean Squared Error : 15288451.916178092
Root Mean Squared Error : 3910.0450018098377
R-Square : 0.9999981517264909

Modelimiz iyi bir modeldir.

Artificial Neural Networks

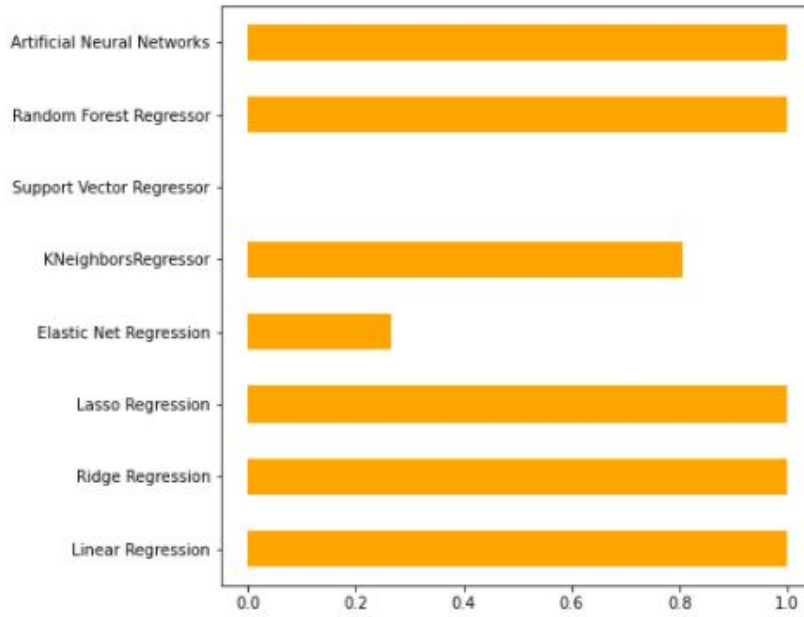
Mean Squared Error : 11956612800.738155
Root Mean Squared Error : 109346.2976087355
R-Square : 0.9985545239753458



Kullanılabilir iyi bir modeldir.

Sonuç olarak bütün modellerimizin tahmin performanslarına baktığımızda bunlar R^2 değerleri oluyor, Şu sonuçları elde ediyoruz.

	R-Square	Root Mean Squared Error
Linear Regression	1.000000	1.924658e+03
Ridge Regression	0.999996	5.491396e+03
Lasso Regression	1.000000	1.924074e+03
Elastic Net Regression	0.264272	2.466932e+06
KNeighborsRegressor	0.805998	1.266782e+06
Support Vector Regressor	-0.000047	2.876132e+06
Random Forest Regressor	0.999998	3.910045e+03
Artificial Neural Networks	0.998555	1.093463e+05



Grafiklere baktığımızda modellerimizde en yüksek performansı veren ve bununla beraber düşük hataları olan modellerimizi tercih ediyoruz. Performans olarak Linear Regression, Ridge Regression, Lasso Regression, Random Forest Regressor ve Artificial Neural Networks modellerinin en iyi olduğunu görüyoruz. Bununla beraber KNeighbors Regressor de kullanılabilir bir modeldir fakat ilk tercihimiz olmayacaktır. Bütün modellerimizden en düşük hatayı veren modellerimiz ise Linear, Lasso Regression ve Artificial Neural Networks modellerimizdir.

KAYNAKÇA

<https://www.kaggle.com/ruchi798/housing-prices-eda-and-prediction>

<https://towardsdatascience.com/create-a-model-to-predict-house-prices-using-python-d34fe8fad88f>

<https://www.pluralsight.com/guides/linear-lasso-ridge-regression-scikit-learn>

<https://www.kaggle.com/faressayah/linear-regression-house-price-prediction>

<https://www.kaggle.com/ashydv/housing-price-prediction-linear-regression>

<https://machinelearningmastery.com/spot-check-regression-machine-learning-algorithms-python-scikit-learn/>