

1) HOUSE PRICES - ADVANCED REGRESSION TECHNIQUES

TRAIN DATA SET



VERİ ANALİZİ

GÖZLEMLER: 1460

DEĞİŞKENLER 81

KATEGORİK DEĞİŞKENLER: 43

KATEGORİK OLUP SINIF SAYISI 20'DEN FAZLA OLAN DEĞİŞKENLER:1

NUMERİK DEĞİŞKENLER: 38

NUMERİK OLUP SINIF SAYISI 10'DAN AZ OLAN DEĞİŞKENLER:11



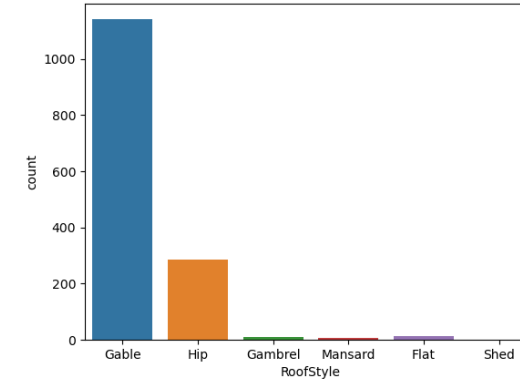


KATEGORİK DEĞİŞKEN ANALİZİ

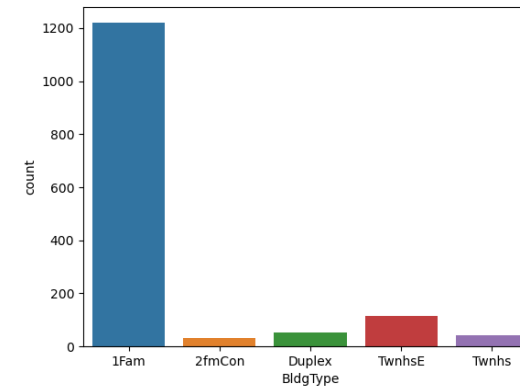
- Kategorik değişkenler(sınıf sayısı 20'den fazla olanlar bu gruba dahil değildir.) ve numerik olup sınıf sayısı 10'dan az olanlar final_cat_cols olarak tanımlanmıştır.
- Final_cat_cols 53 adet değişken içermektedir.
- Final_cat_cols içerisinde Alley(Type of alley access to property) değişkeni 1369 eksik gözleme sahiptir. PoolQC değişkeni :1453 ,Fence:1179 değişkeni , MiscFeature değişkeni: 1406 eksik gözleme sahiptir.
- En fazla eksik gözlem içeren: PoolQC havuz kalitesini, Alley: mülke erişim türünü, Fence: çit kalitesini ve MiscFeature değişkeni diğer değişkenlerde belirtilmeyen bazı özellikleri içermektedir. Bu bağlamda bu verilerin eksik olması noktasında tüm evlerde havuz ve çit olmadığı bilgisine erişilebilir.
- MiscFeature değişkeni incelendiğinde alt kırılımları: Asansör, tenis kortu, garaj gibi bilgileri içermektedir. Buradan da yine bu özelliklerin her ev için bulunmadığı anlaşılabılır.

KATEGORİK DEĞİŞKEN ANALİZİ

- Kategorik değişken analizinde veri setinde satışa çıkarılan evlerden çatı türlerinden üçgen çatı olanlar en fazladır. Bunun sebebi hem ısı hem de su sızıntıları açısından büyük avantajlar bu tip çatıları kişilerin daha fazla tercih etmesi olabilir.

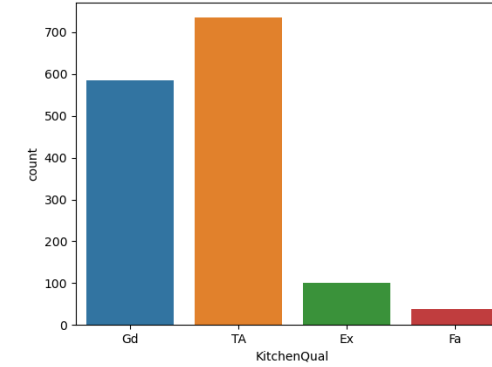


- Kategorik değişken analizinde satışa çıkarılan evlerden türü 1 ailenin yaşayacağı türde konutlar daha fazladır.

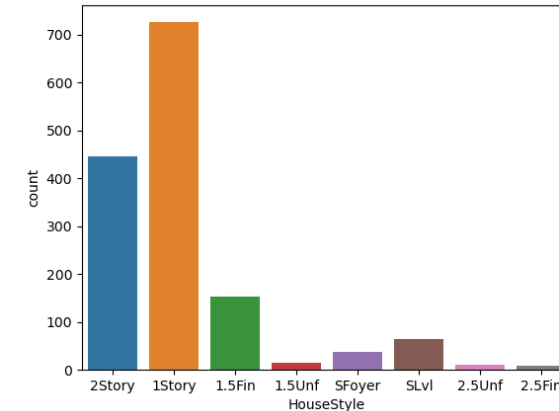


KATEGORİK DEĞİŞKEN ANALİZİ

- Kategorik değişken analizinde veri setinde satışı çıkarılan evlerden değişken tipi mutfak kalitesi iyi olanlar daha fazladır. Bu da insanların ev alırken mutfak kalitesine özen gösterdiğinin bir göstergesidir.

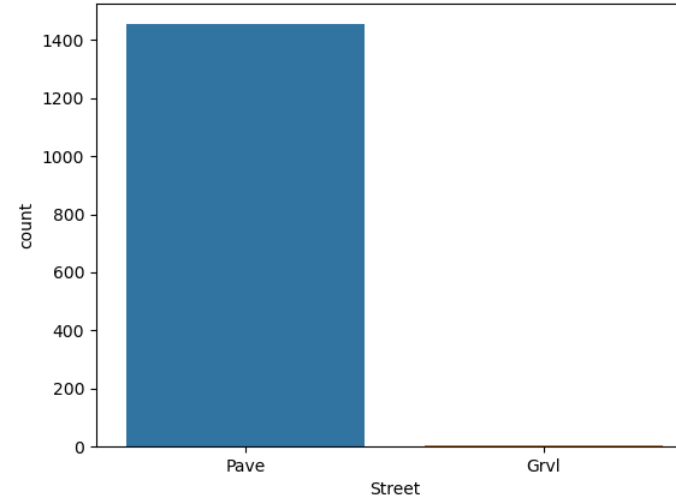


- Kategorik değişken analizinde veri setinde satışı çıkarılan evlerden ev tipi tek katlı olanlardır.



KATEGORİK DEĞİŞKEN ANALİZİ

- Veri setinde pave ve gravel olarak iki sokak bulunmaktadır. Satışa çıkartılan en fazla ev pave sokağı içerisinde bulunanlardır. Buradan Pave ve Gravel sokakların yapısı karşılaştırılabilir, pave sokağındaki evlerin daha çok rağbet görüp görmediği tartışılabilir.



- Numerik değişken olarak görünüp sınıf sayısı 10'dan az olan değişkenler:

OverallCond',
'BsmtFullBath',
'BsmtHalfBath',
'FullBath',
'HalfBath',
'BedroomAbvGr',
'KitchenAbvGr',
'Fireplaces',
'GarageCars',
'PoolArea',
'YrSold']

- Sınıf sayısı 10'dan az olan değişkenler çıkartıldığında Numerik olan 27 değişken ise:

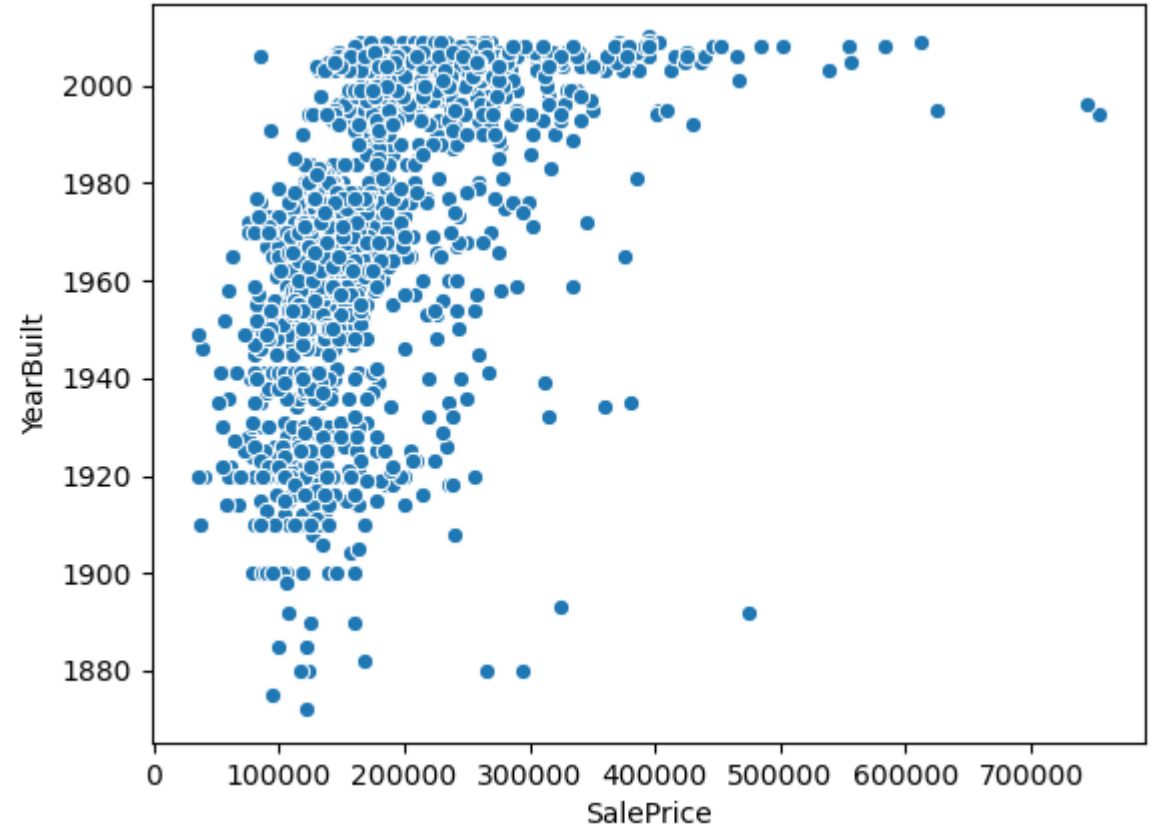
« 'Id', 'MSSubClass', 'LotFrontage', 'LotArea', 'OverallQual', 'YearBuilt', 'YearRemodAdd', 'MasVnrArea', 'BsmtFinSF1', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', '1stFlrSF', '2ndFlrSF', 'LowQualFinSF', 'GrLivArea', 'TotRmsAbvGrd', 'GarageYrBlt', 'GarageArea', 'WoodDeckSF', 'OpenPorchSF', 'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'MiscVal', 'MoSold', 'SalePrice' »

NUMERİK DEĞİŞKEN ANALİZİ

NUMERİK DEĞİŞKEN ANALİZİ

- Binların kaç yıllık olduğunun fiyat üzerindeki etkisini incelemek istersek:
- İki değişken arasında pozitif yönlü korelasyon vardır.

Görsele bakarak en fazla fiyata sahip olan binaların 1980-2000 yılları arasında yapıldığını söylemek mümkündür.



TARGET ANALİZİ

TARGET=SALEPRICE

- Kategorik değişken olan elektrik sisteminin, evlerin fiyatı üzerinde etkisini incelemek istersek. Elektrik sistemi Standard Circuit Breakers & Romex olan herhangi bir yangın gibi tehlikeli durumlarda otomatik kapatma özelliğine sahip olan elektrik sistemleri bulunan evlerin fiyatlarının daha yüksek olduğu sonucuna varılabilir.

	TARGET_MEAN
Electrical	
FuseA	122196.893617
FuseF	107675.444444
FuseP	97333.333333
Mix	67000.000000
SBrkr	186825.113193

TARGET ANALİZİ

TARGET=SALEPRICE

- Kategorik değişken olan ısıtma sisteminin, evlerin fiyatı üzerinde etkisini incelemek istersek. Isıtma sistemi yerden ısıtma olan evlerin fiyatının diğer türlere göre daha düşük olduğu sonucuna varılabilir.

	TARGET_MEAN
Heating	
Floor	72500.000000
GasA	182021.195378
GasW	166632.166667
Grav	75271.428571
OthW	125750.000000
Wall	92100.000000

TARGET ANALİZİ

TARGER=SALEPRICE

- Kategorik değişken olan dış malzeme kalitesinin, evlerin fiyatı üzerinde etkisini incelemek istersek. Evlerin dış yapısının görünüşünün satış fiyatları ve rağbet üzerinde etkili olduğunu görürüz.
- Aşağıda bulunan görsele göre dış malzemesi orta-iyi ve çok iyi olan evlerin fiyatları arasındaki artış ile Fair(kötü) dış malzemeye bakıldığında fiyatın neredeyse yarısı kadar düştüğünü gözlemleyebiliriz.

	TARGET_MEAN
ExterQual	
Ex	367360.961538
Fa	87985.214286
Gd	231633.510246
TA	144341.313466

TARGET ANALİZİ

TARGET=SALEPRICE

- Numerik değişken olan bir evin büyüklüğü ile fiyat arasındaki ilişkiyi incelemek istersek:
- Aralarında yüksek bir korelasyon bulunmasa da evlerin büyüklüğünün arttıkça fiyatının da arttığını söyleyebiliriz.

	LotArea
SalePrice	
34900	7879.0
35311	9000.0
37900	5925.0
39300	5000.0
40000	8500.0
	...
582933	13891.0
611657	12919.0
625000	35760.0
745000	15623.0
755000	21535.0

TEŞEKKÜRLER