Zillow's Price Feature Engineering

Chan Yu Yankai Liu Yifei Bi

Procedures Guide

01=**
03=**
04-**

1.Preparation

2.Impute Data 1

3. Creating New Features

4.Impute Data 2

Reporter: Chan Yu

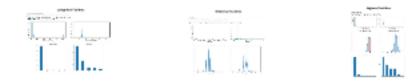


1.Convert Float64 to Float32

	parcelid	airconditioningtypeid	architecturalstyletypeid	basementsqft	bathrooment	bedrooment	buildingclasstypeid
Ö	10754147	1.0	7.0	535.0	0.0	0.0	4.0
1	10759547	1.0	7.0	535.0	0.0	0.0	4.0
2	10843547	1.0	7.0	535.0	0.0	0.0	5.0
3	10859147	1.0	7.0	535.0	0.0	0.0	3.0
4	10879947	1.0	7.0	535.0	0.0	0.0	4.0

5 rows x 54 columns

2.Group Features into Different Groups



3. Prepare Functions For Corresponding Group

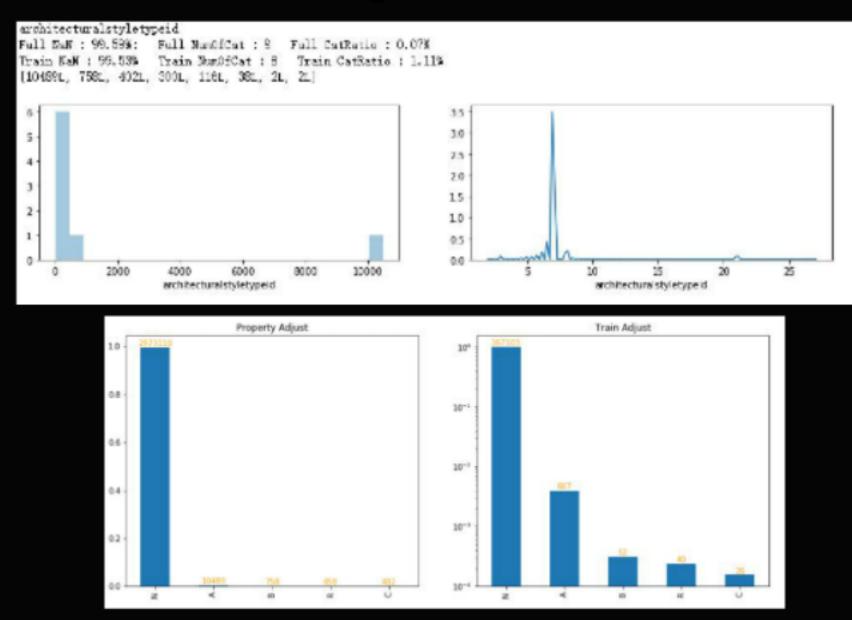
1.Preparation

2.Impute Data 1

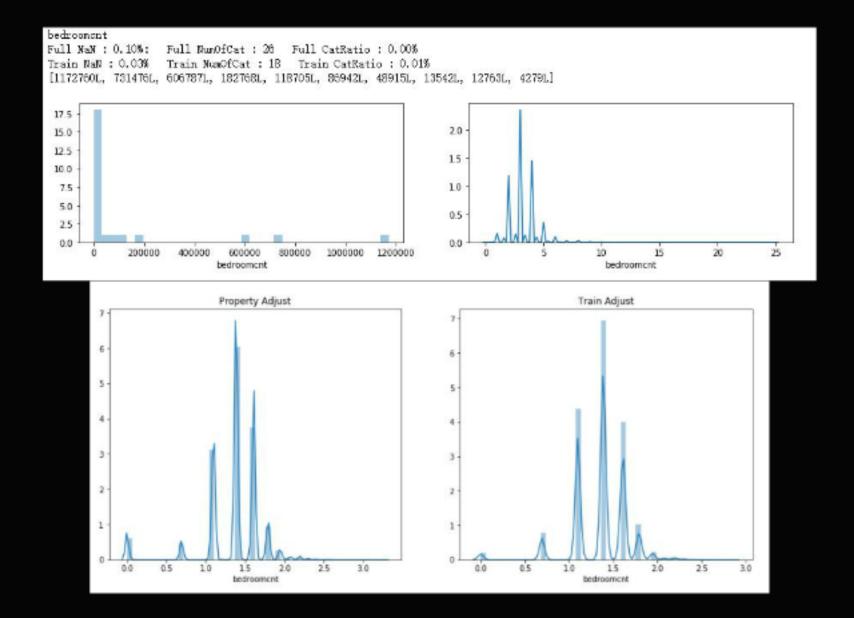
3.Creating New Features

4.Impute Data 2

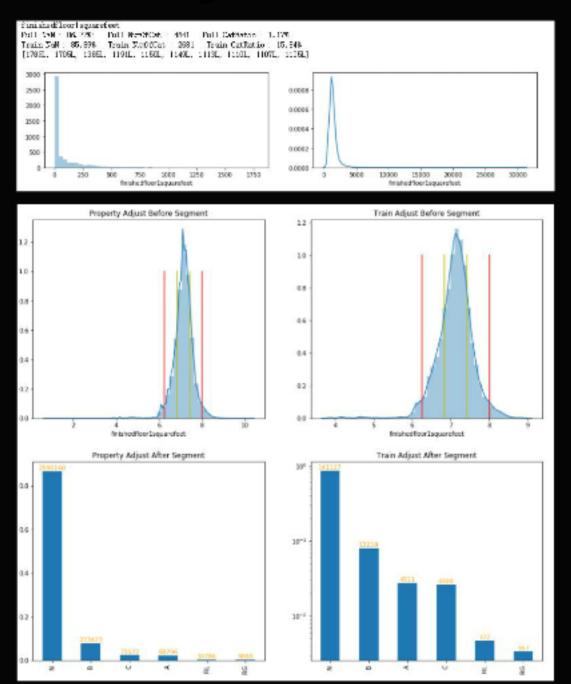
Categorical Features



Numerical Features



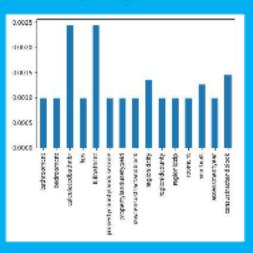
Segment Features



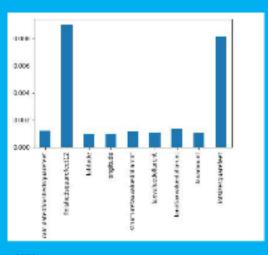


1.Preparation2.Impute Data 13.Creating New Features4.Impute Data 2

Impute By Median



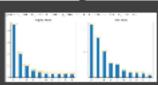
Impute By Mode



02

Missing ratio < 10%





Cluster Geo Info

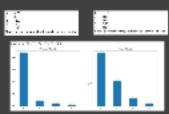
г	-	-0.00	 	- Marie	-	_
٠	100	100	 -	1	75	
٠	-		 _			
		-	 	100		
	200					
٠	-		 -			

7 algrithms and each for 5 classes

elitarian de la company	
11.000	
200	
75m	
9,000	

1.Preparation 2.Impute Data 1 3.Creating New Features 4.Impute Data 2

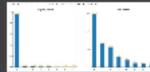




Date Features

-	analysis y	400	parties 2	-	100
-	18		1610	1111	4
-	- 6		1610	00.00	
	18		1010	22.24	
7	***		10.12	20.00	
	**		1472	25, 64	

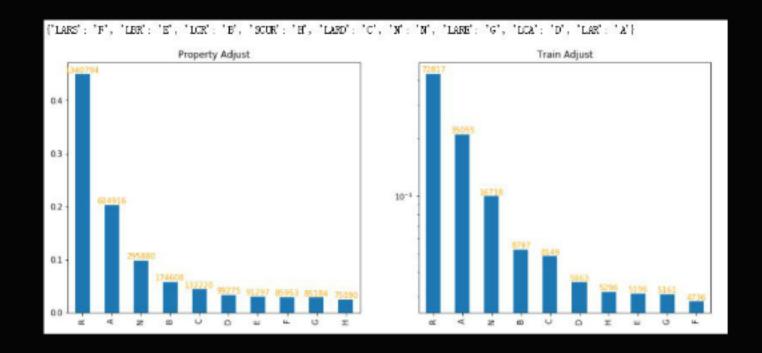
Can desinquencyyear



propertyzoningdesc

```
0 NaN
1 LCA11*
2 LAC2
3 LAC2
4 LAM1
Name: propertyzoningdesc, dtype: object
```

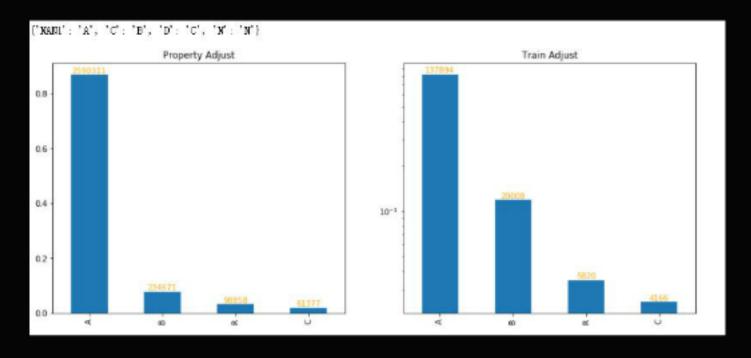
```
0 NaN
1 LCA
2 LAC
3 LAC
4 LAM
Name: propertyzoningdesc, dtype: object
```



propertycountylandusecode

```
0 010D
1 0109
2 1200
3 1200
4 1210
Name: propertycountylandusecode, dtype: object
```

```
0 D
1 NAN1
2 NAN1
3 NAN1
4 NAN1
Name: propertycountylandusecode, dtype: object
```



Cluster Geo Info

	regionidcity	regionidzip	rawcensustractandblock	censustractandblock	latitude	longitude	fips
0	37688.0	96337.0	60378004.0	6.059032e+13	34144440.0	-118654080.0	A
1	37688.0	96337.0	60378000.0	6.059032e+13	34140432.0	-118825360.0	Α
2	51617.0	96095.0	60377032.0	6.059032e+13	33989360.0	-118394632.0	Α
3	12447.0	98424.0	60371412.0	6.059032e+13	34148864.0	-118437208.0	Α
4	12447.0	96450.0	60371232.0	6.059032e+13	34194168.0	-118385816.0	Α

7 algrithms and each for 8 classes

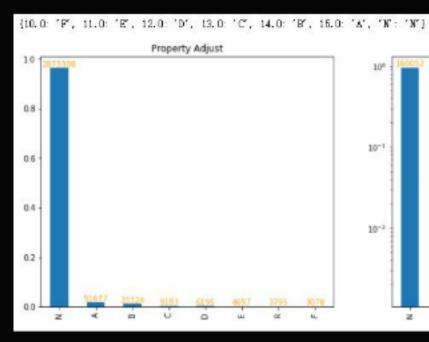
```
charteches compagned as T records.
                                                                                                                                                     reconvertexcondition, reconstructed body, 'basitudy, 'bogitudy, first', long
clusterlocations (clusterlocations clusterlocation, near) ( /clusterlocation, etc.)
nadel-mistacre. Sous classifications in component cvil)
 table fi ladate astind
properties of alluments returned by Practical product telephonetrical
 indefectioner. Please (surflumensel)
 (midset, saled at Lake)
projective ("clume classication, be" benefit, product (classication)
prints; ... !
undelectures. Spectrol Clustering by clumersel)
 cold, fir fellet ed section).
projective ("clusteringstronger" (warde), product (clusteringstron)
nada Leakurten, Edikiak ()
 madel, fit is hast of acetime.
 name time ["above doest me, the", each 's prefint (cinetechnic time).
print(" ... ')
Challenia ayradaylaban
nadel, fit leftet will scatime
 (criterial stackets be of the or tendered party con-
prant? ....
 and trades on Assume retrievel/horizones in position of
 nadel, fit leftart will scuttime
 Cuitsof stedicardence is net be not been been being lesi term
prant.".
  indefectures. Roofs to of satement.
  odd tiskhetelestind
properties of clume classical by by deligation (classical costical)
properties[f] charter location_mi_f, 'cluster location_im', 'cluster location_cm', 'cluster
```

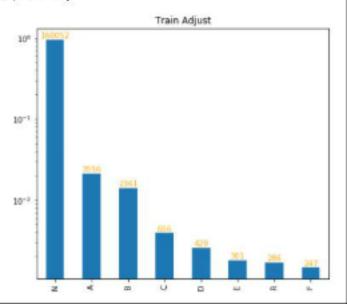
Date Features

	assessmentyear	yearbuilt
0	2016.0	1963.0
1	2015.0	1963.0
2	2016.0	1959.0
3	2016.0	1948.0
4	2016.0	1947.0

	assessmentyear	yearbuilt
0	1.0	62.0
1	2.0	62.0
2	1.0	58.0
3	1.0	69.0
4	1.0	70.0

taxdelinquencyyear







1.Preparation 2.Impute Data 1 3.Creating New Features

4.Impute Data 2

