Zillow's Price Data Exploration

Chan Yu Yankai Liu Yifei Bi

OverView

1.Program Background
2.Trainset Exploration
3.Property Exploration
4.Geograhic&correlation

Reporter : Yankai Liu

1.Program Background
2.Trainset Exploration
3.Property Exploration
4.Geograhic&correlation

Program background

Question1:What is zillow and Zestimates?
Zillow: Real estate price evaluation company
Zestimates: Using algorithms and data to estimate the price.

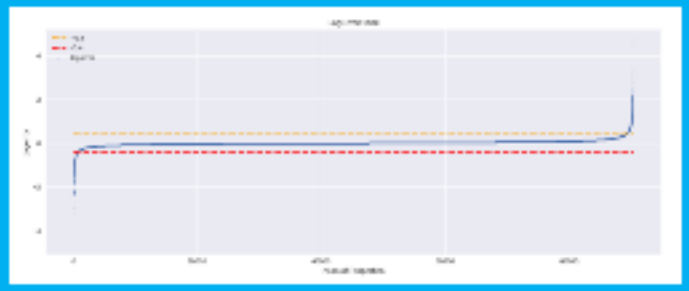
Question2:Why is Zestimates so important?

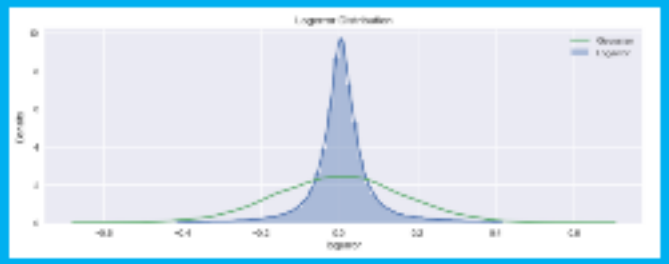
Zestimates can evaluate the price of real estate. But Zestimates is based on statistical and machine learning models. So the efficiency and accuracy of the model are very important.

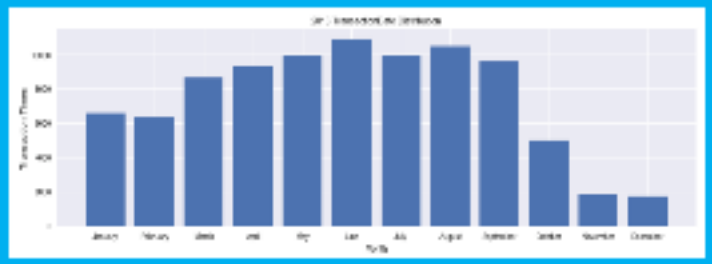
2

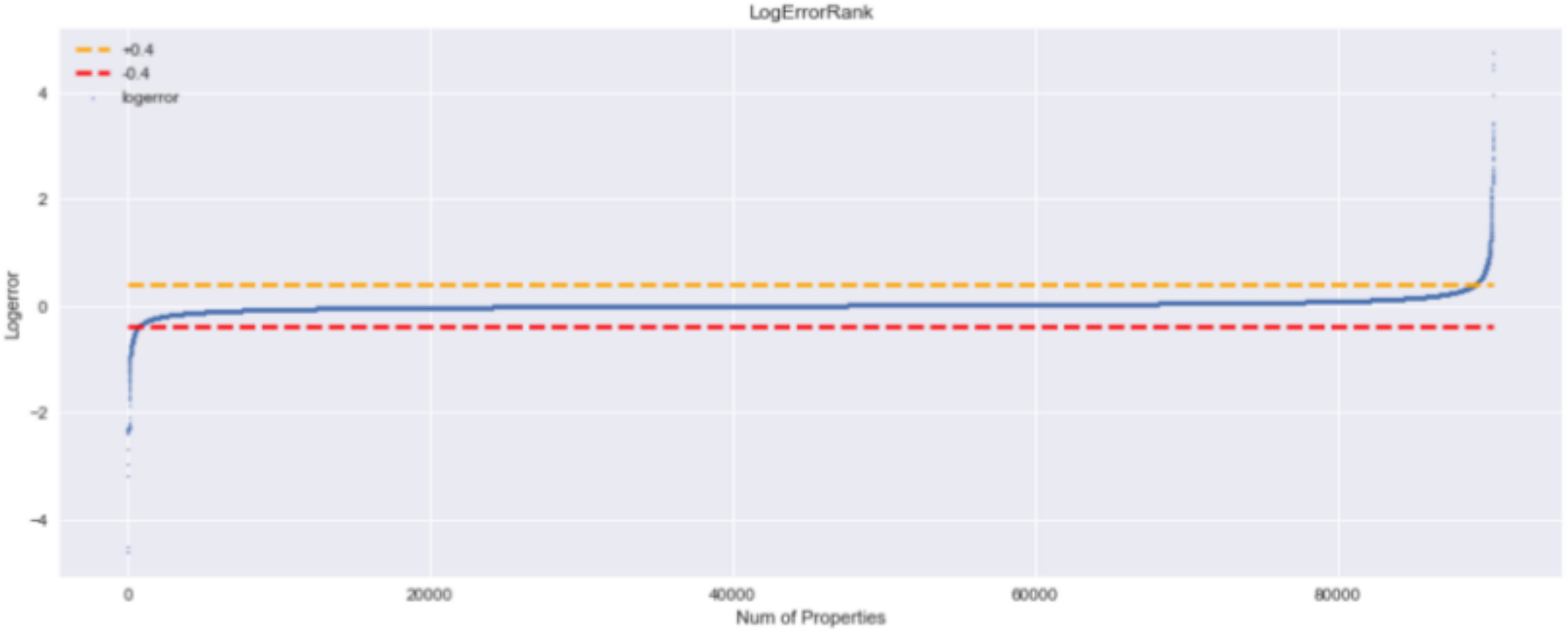
Trainset exploration

1.Program Background
2.Trainset Exploration
3.Property Exploration
4.Geograhic&correlation

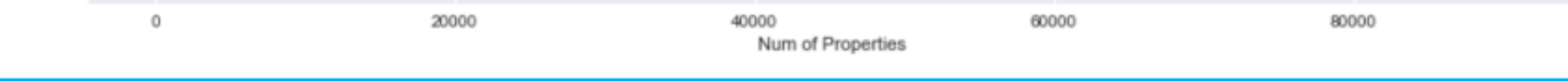


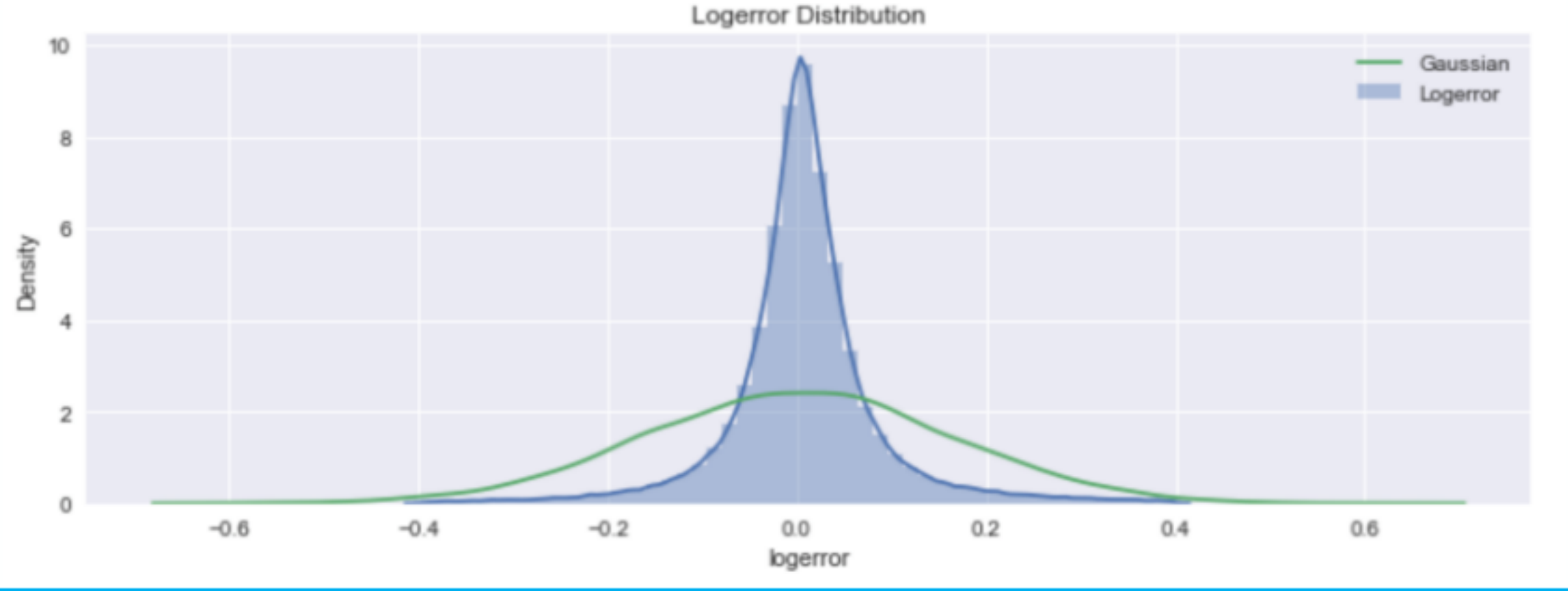


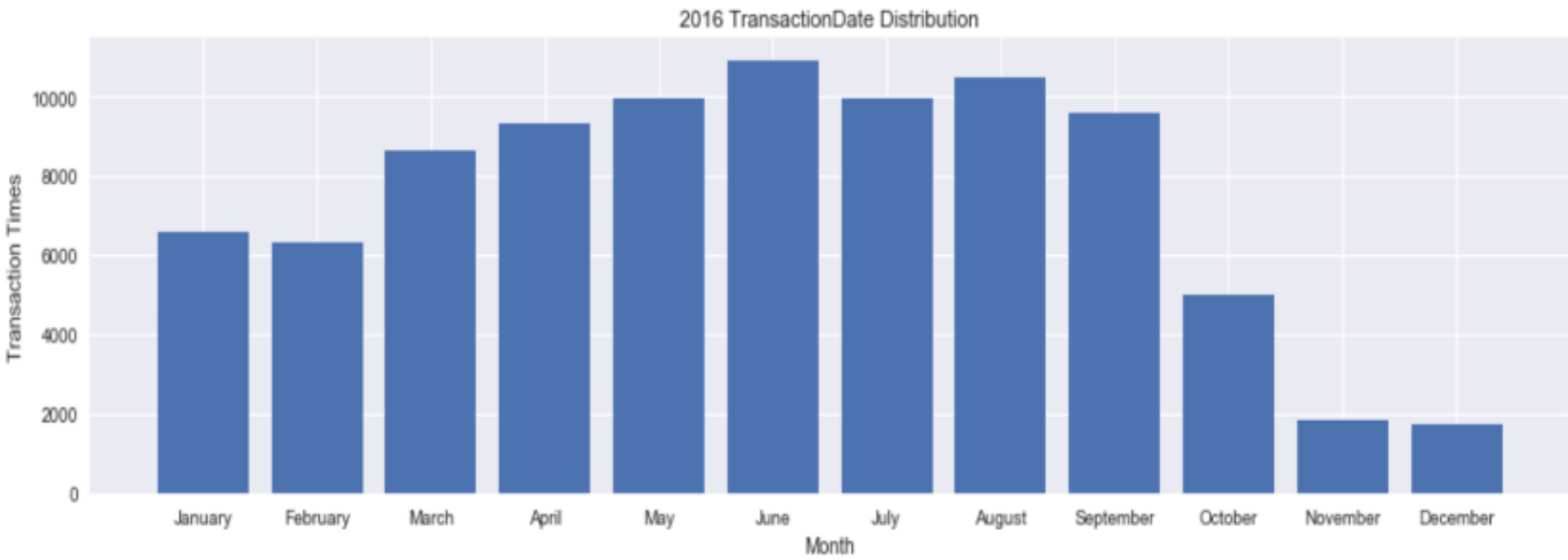



LogErrorRank
+0.4
-0.4
logerror
4
2
0
−2
−4
Logerror
0
20000
40000
60000
80000
Num of Properties

Logerror Distribution
10


0
20000
40000
60000
80000
Num of Properties

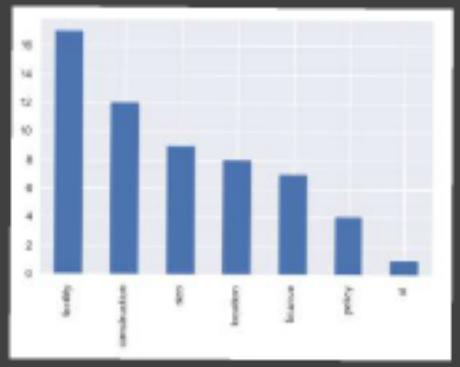

Logerror Distribution
Gaussian
Logerror
Density
10
8
6
4
2
0
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
logerror

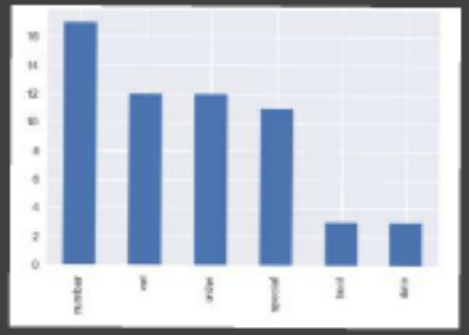

2016 TransactionDate Distribution
Transaction Times
0
2000
4000
6000
8000
10000
January
February
March
April
May
June
July
August
September
October
November
December
Month

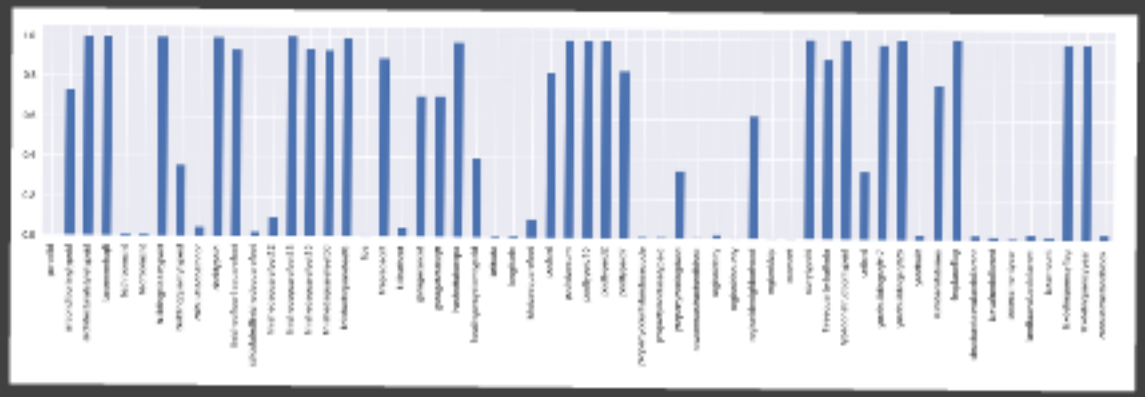
3

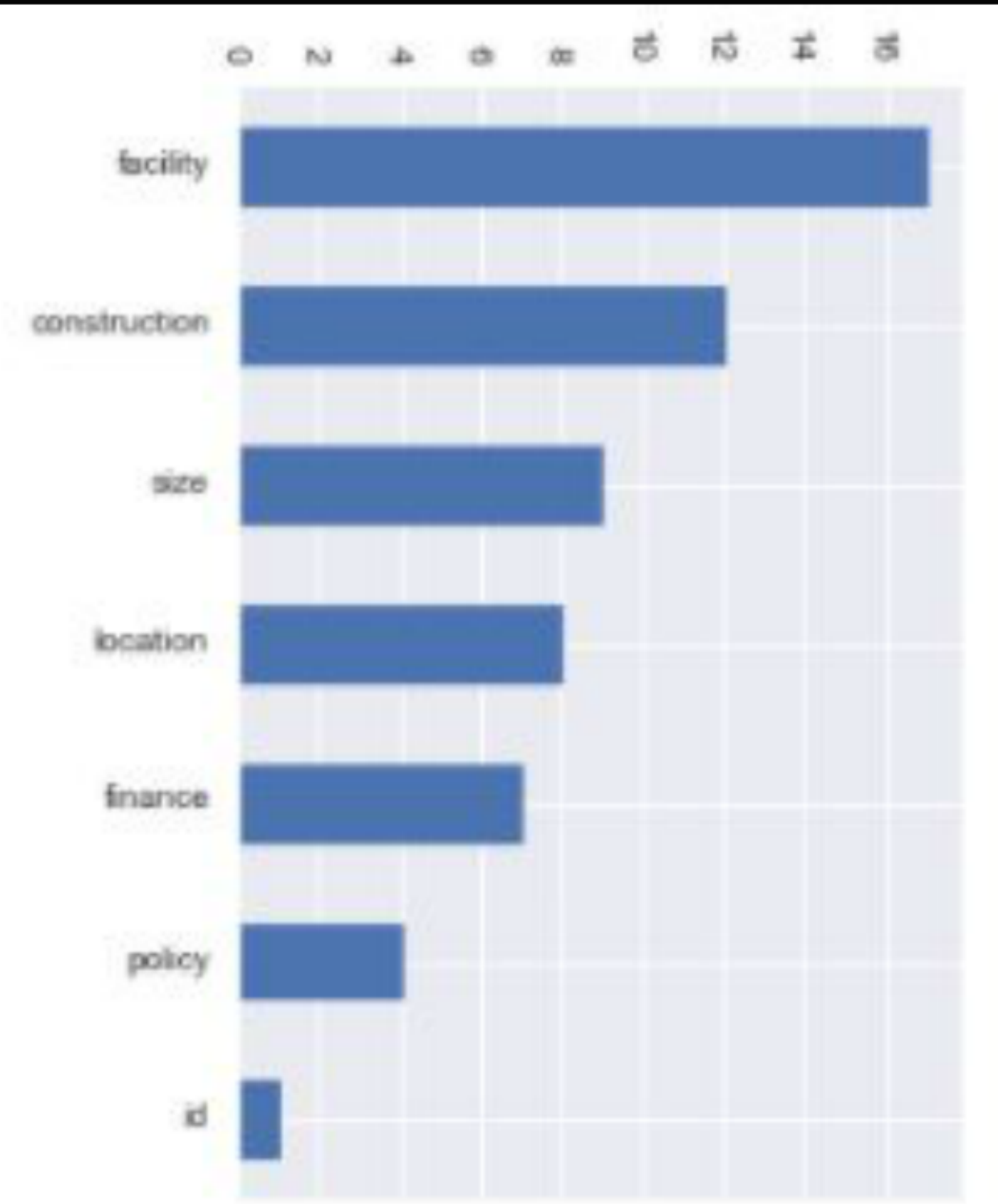
Properties exploration

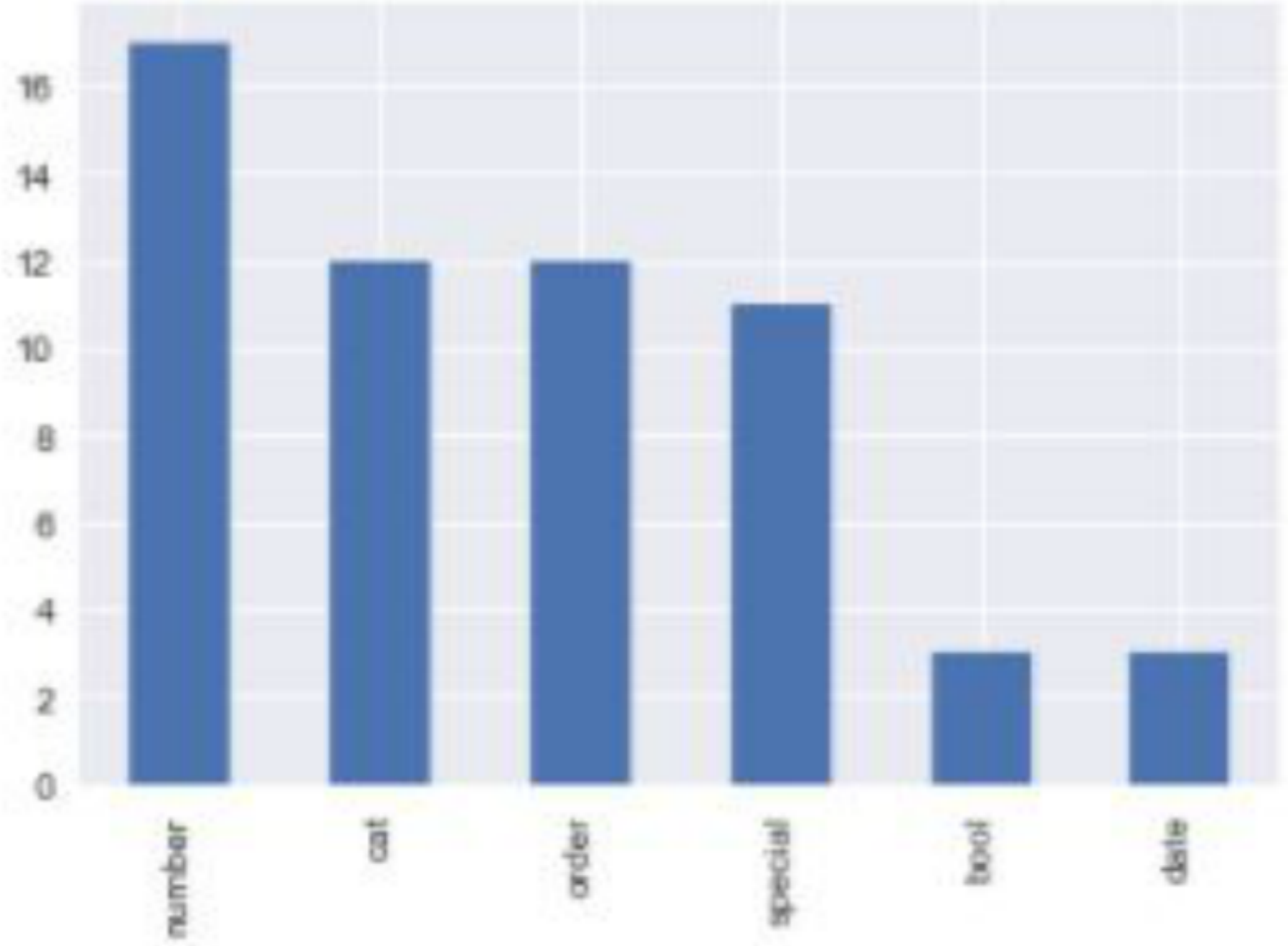
1.Program Background
2.Trainset Exploration
3.Property Exploration
4.Geograhic&correlation

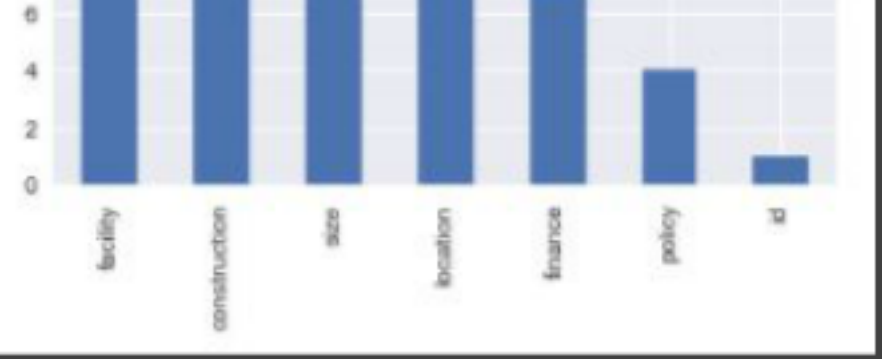


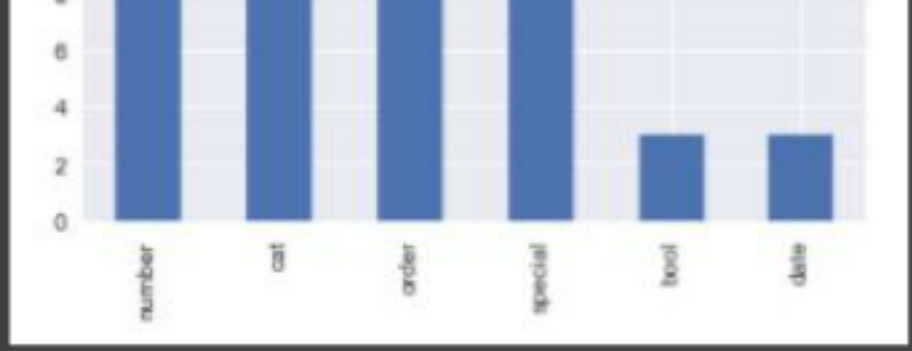


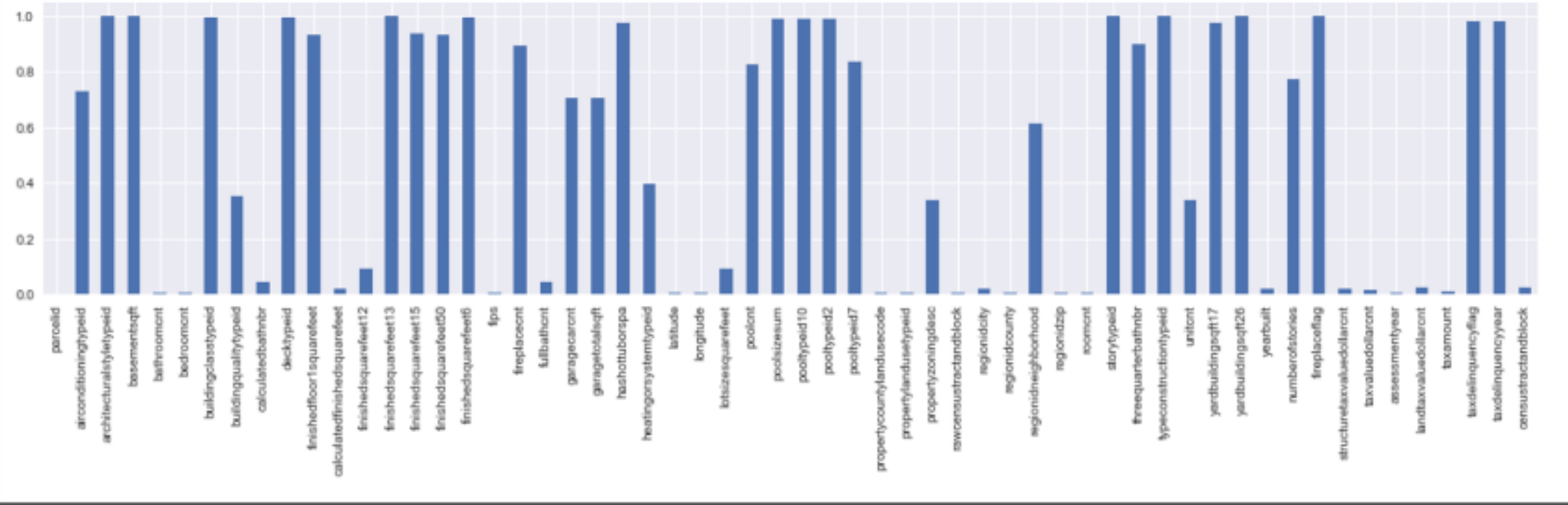


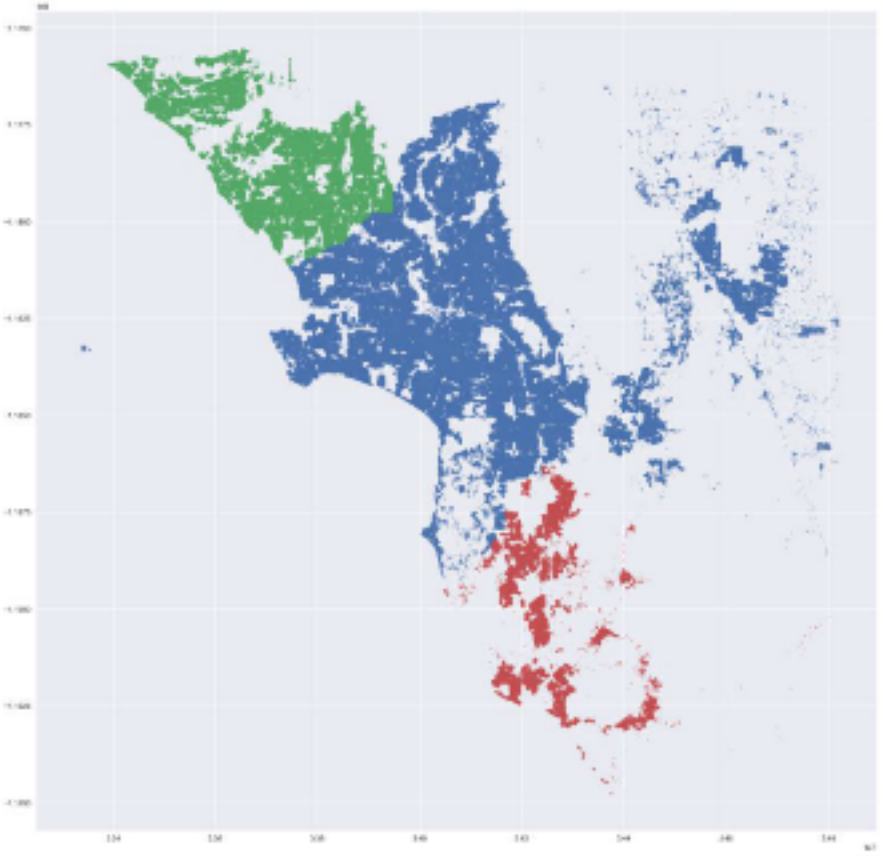
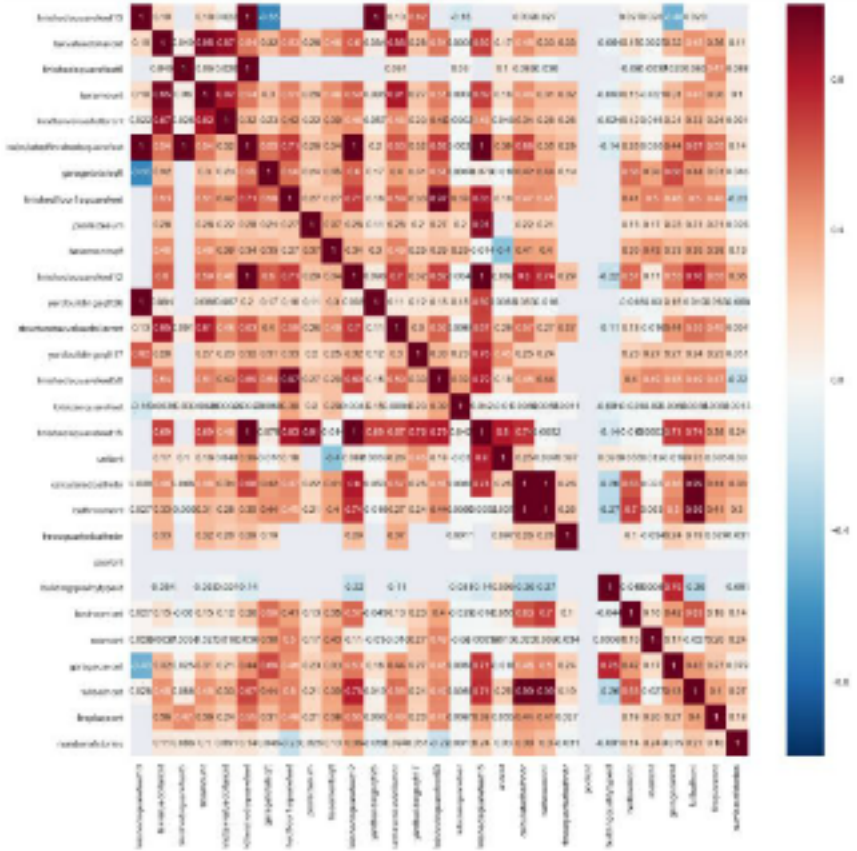

0
2
4
6
8
10
12
14
16
facility
construction
size
location
finance
policy
id


16
14
12
10
8
6
4
2
0
number
cat
order
special
bool
date


0
2
4
6
facility
construction
size
location
finance
policy
id


0
2
4
6
number
cat
order
special
bool
date


0.0
0.2
0.4
0.6
0.8
1.0
parcelid
airconditioningtypeid
architecturalstyletypeid
basementsqft
bathroomcnt
bedroomcnt
buildingclasstypeid
buildingqualitytypeid
calculatedbathnbr
decktypeid
finishedfloor1squarefeet
calculatedfinishedsquarefeet
finishedsquarefeet12
finishedsquarefeet13
finishedsquarefeet15
finishedsquarefeet50
finishedsquarefeet6
fips
fireplacecnt
fullbathcnt
garagecarcnt
garagetotalsqft
hashottuborspa
heatingorsystemtypeid
latitude
longitude
lotsizesquarefeet
poolcnt
poolsizesum
pooltypeid10
pooltypeid2
pooltypeid7
propertycountylandusecode
propertylandusetypeid
propertyzoningdesc
rawcensustractandblock
regionidcity
regionidcounty
regionidneighborhood
regionidzip
roomcnt
storytypeid
threequarterbathnbr
typeconstructiontypeid
unitcnt
yardbuildingsqft17
yardbuildingsqft26
yearbuilt
numberofstories
fireplaceflag
structuretaxvaluedollarcnt
taxvaluedollarcnt
assessmentyear
landtaxvaluedollarcnt
taxamount
taxdelinquencyflag
taxdelinquencyyear
censustractandblock

Geograhic&correlation

1.Program Background
2.Trainset Exploration
3.Property Exploration
4.Geograhic&correlation