

1. 数据分析要求 3.1 数据可视化和摘要

数据摘要 标称属性，给出每个可能取值的频数

数值属性，给出5数概括及缺失值的个数

数据可视化 使用直方图、盒图等检查数据分布及离群点

3.2 数据缺失的处理 观察数据集中缺失数据，分析其缺失的原因。分别使用下列四种策略对缺失值进行处理：

将缺失部分剔除 用最高频率值来填补缺失值 通过属性的相关关系来填补缺失值 通过数据对象之间的相似性来填补缺失值 注意：在处理后，要可视化地对比新旧数据集。

```
In [24]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
#在cell中显示图形
```

去除数据集中完全相同的行

```
In [25]: def delDuplic(data):
    #显示重复行的总个数
    print(data.duplicated().sum())
    #去除完全重复行，并将结果写入df变量，避免占用过多空间
    data.drop_duplicates(inplace=True)
    #确认是否成功删除了重复行
    print(data.duplicated().sum())
```

统计标称型数据的频数

```
In [26]: def nominal_summary(df,nominal_index=None,new_df=None,head_n=50):
    #根据频数绘制直方图
    def bar_describe(data,new_data=None,head_n=50):
        if new_data is None:
            plt.figure(figsize=(24,8))
            plt.title(data.name, fontsize=30)
            plt.bar(data.index[:head_n], data.values[:head_n])
            plt.xticks(rotation=90)
            plt.show()
        else:
            plt.figure(figsize=(24,8))
            plt.subplot(211)
            plt.title(data.name, fontsize=30)
            plt.bar(data.index[:head_n], data.values[:head_n], color='b', label=data)
            plt.xticks(rotation=90)

            plt.figure(figsize=(24,8))
            plt.subplot(212)
            plt.title('new_'+new_data.name, fontsize=30)
            plt.bar(new_data.index[:head_n], new_data.values[:head_n], color='g', label=new_data)
            plt.xticks(rotation=90)
            plt.legend()
            plt.show()
            ,,
            i = 1

        for index in nominal_index:
```

```

freqs1 = dict(df[index].value_counts())
freqs2 = dict(new_df[index].value_counts())

plt.subplot(len(nominal_index), 2, i)
i = i+1
plt.bar(freqs1[0], freqs1[1])
plt.xticks(rotation=90)
plt.title(index, y=0.9, fontdict={'weight': 'normal', 'size': 15})

plt.subplot(len(nominal_index), 2, i)
i = i+1
plt.bar(freqs2[0], freqs2[1], color='green')
plt.xticks(rotation=90)
plt.title(index, y=0.9, fontdict={'weight': 'normal', 'size': 15}),
,,,

# 获取数据中每个取值的频数
frequency = {key: df[key].value_counts() for key in df.columns}
new_frequency = None
if new_df is not None:
    new_frequency = {key: new_df[key].value_counts() for key in new_df.columns}

if nominal_index is None:
    nominal_index=df.columns
if new_frequency is None:
    for key in nominal_index:
        bar_describe(data=frequency[key], new_data=None, head_n=head_n)
else:
    for key in nominal_index:
        bar_describe(data=frequency[key], new_data=new_frequency[key], head_n=head_n)
return None

```

数值属性缺失值个数及五数概括

```

In [27]: def numerical_summary(df, numerical_index=None, new_df=None):
    shape=df.shape # 样本总数
    # 格式化输出五数概括和缺失值个数
    def data_describe(data, new_data=None):
        print('descriptive statistics (%s):' % data.name)
        info=data.describe()
        new_info = None
        if new_data is not None:
            new_info = new_data.describe()

        print("Min: {:.4f}\tQ1(25%): {:.4f}\tQ2(50%): {:.4f}\tQ3(75%): {:.4f}\tMa
              info['min'], info['25%'], info['50%'], info['75%'], info['max']))
        print("Missing: {:d}.".format(int(shape[0] - info['count'])))
        print()

        if new_info is not None:
            print("\033[95m new Min: {:.4f}\tQ1(25%): {:.4f}\tQ2(50%): {:.4f}\tQ3(
                  new_info['min'], new_info['25%'], new_info['50%'], new_info['75%'], new_info
                  print("\033[95m new Missing: {:d} \033[0m".format(int(shape[0] - new_inf
                  print()

        if numerical_index is None:
            print("Please provide the numerical index needed to be describe")
            return None

    # 获取数值属性的5数概况和缺失值个数
    if new_df is None:
        for key in numerical_index:
            data_describe(df[key])

```

```

    else:
        for key in numerical_index:
            data_describe(df[key], new_df[key])

    return None

```

数据可视化，用于绘制数值属性的直方图，盒图及散点图

```

In [28]: def box_plot(df, numerical_index, new_df=None):
    if new_df is None:
        for index in numerical_index:
            plt.boxplot(list(df[index].dropna(axis=0)))
            plt.title(index, fontdict={'weight': 'normal', 'size': 15})
            plt.show()
    else:
        i = 1
        col = len(numerical_index)
        plt.figure(figsize=(15, 5*col))

        for index in numerical_index:
            plt.subplot(col, 2, i)
            i = i+1
            plt.boxplot(list(df[index].dropna(axis=0)))
            plt.title(index, fontdict={'weight': 'normal', 'size': 15})

            plt.subplot(col, 2, i)
            i = i+1
            plt.boxplot(list(new_df[index].dropna(axis=0)))
            plt.title(index, fontdict={'weight': 'normal', 'size': 15})

        plt.tight_layout()

def hist_plot(df, numerical_index, new_df=None):
    if new_df is None:
        for index in numerical_index:
            plt.hist(list(df[index].dropna(axis=0)), bins=20)
            plt.title(index, fontdict={'weight': 'normal', 'size': 15})
            plt.show()
    else:
        i = 1
        col = len(numerical_index)
        plt.figure(figsize=(15, 5*col))

        for index in numerical_index:
            plt.subplot(col, 2, i)
            i = i+1
            plt.hist(list(df[index].dropna(axis=0)), bins=20)
            plt.title(index, fontdict={'weight': 'normal', 'size': 15})

            plt.subplot(col, 2, i)
            i = i+1
            plt.hist(list(new_df[index].dropna(axis=0)), bins=20, color='green')
            plt.title(index, fontdict={'weight': 'normal', 'size': 15})

        plt.tight_layout()

def scatter_plot(df, numerical_index, new_df=None):
    if new_df is None:
        for index in numerical_index:
            plt.scatter(list(df[index].dropna(axis=0)))
            plt.title(index, fontdict={'weight': 'normal', 'size': 15})

```

```

    plt.show()
else:
    i = 1
    col = len(numerical_index)
    plt.figure(figsize=(15, 5*col))

    for index in numerical_index:
        plt.subplot(col, 2, i)
        i = i+1
        plt.scatter(list(df[index].dropna(axis=0)))
        plt.title(index, fontdict={'weight': 'normal', 'size': 15})

        plt.subplot(col, 2, i)
        i = i+1
        plt.scatter(list(new_df[index].dropna(axis=0)), color='green')
        plt.title(index, fontdict={'weight': 'normal', 'size': 15})

```

将缺失部分剔除

```
In [29]: def miss_drop(df):
    return df.dropna(axis=0)
```

用最高频率值来填补缺失值

```
In [30]: def miss_mode_fill(df):
    cpdf = df.copy(deep=True)
    for key in cpdf.columns:
        cpdf[key].fillna(cpdf[key].mode()[0], inplace=True)
    return cpdf
```

通过属性的相关关系来填补缺失值

```
In [31]: from sklearn.preprocessing import OrdinalEncoder
from sklearn.ensemble import RandomForestRegressor

def attr_corr_fill(df, miss_index, complete_index):
    def set_miss_values(df, complete_index):
        enc_label = OrdinalEncoder()
        enc_fea = OrdinalEncoder()
        missing_index = complete_index[0]

        # Take out the existing numerical data (no NaN) and throw them in Random Forest
        train_df = df[complete_index]
        # known & unknow values
        known_values = np.array(train_df[train_df[missing_index].notnull()])
        unknow_values = np.array(train_df[train_df[missing_index].isnull()])

        # y is the know missing_index
        y = known_values[:, 0].reshape(-1, 1)
        enc_label.fit(y)
        y = enc_label.transform(y)

        # X are the features
        X = known_values[:, 1:]
        test_X = unknow_values[:, 1:]
        all_X = np.row_stack((X, test_X))
        enc_fea.fit(all_X)
        X = enc_fea.transform(X)

        # fit
        rfr = RandomForestRegressor(random_state=0, n_estimators=2000, n_jobs=-1)
```

```

rfr.fit(X, y.ravel())
# predict
predicted_values = rfr.predict(enc_fea.transform(unknow_values[:, 1:]))
predicted_values = enc_label.inverse_transform(predicted_values.reshape(-1, 1))
# fill in with predicted values
df.loc[(df[missing_index].isnull()), missing_index] = predicted_values
return df

cpdf = df.copy(deep=True)
for i in range(0, len(miss_index)):
    complete_index.insert(0, miss_index[i])
    cpdf = set_miss_values(cpdf, complete_index)
return cpdf

```

通过数据对象之间的相似性来填补缺失值

```

In [32]: from impute.imputation.cs import fast_knn
import sys

def obj_corr_fill(df, numerical_index, k=30):
    sys.setrecursionlimit(100000) #Increase the recursion limit of the OS
    # start the KNN training
    cpdf = df.copy(deep=True)
    imputed_training=fast_knn(cpdf[numerical_index].values, k=k)
    imputed_training=pd.DataFrame(data=imputed_training, columns=numerical_index)
    cpdf[numerical_index] = imputed_training[numerical_index]
    return cpdf

```

数据集1：Chicago Building Violations

1.进行初步处理

```

In [33]: #导入数据
df_CBV = pd.read_csv('./Chicago Building Violations/building-violations.csv')
#删除重复值
delDupli(df_CBV)
# 移除无用的属性和作用重复的属性
df_CBV = df_CBV.drop(['ID', 'VIOLATION LAST MODIFIED DATE', 'VIOLATION DATE', 'VIOLATION',
                      'VIOLATION INSPECTOR COMMENTS', 'INSPECTOR ID', 'LATITUDE',
                      'INSPECTION WAIVED', 'SSA'], axis=1)

```

6546

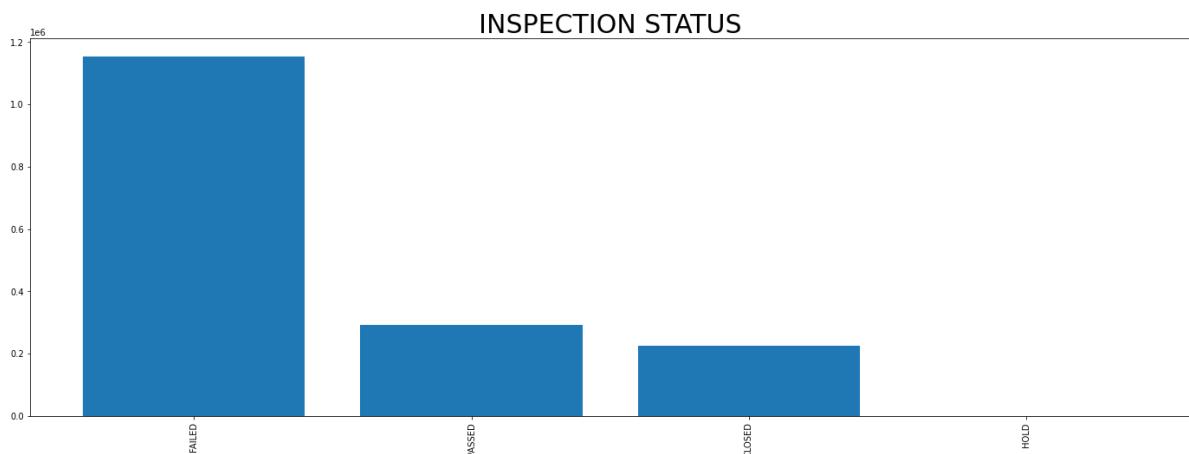
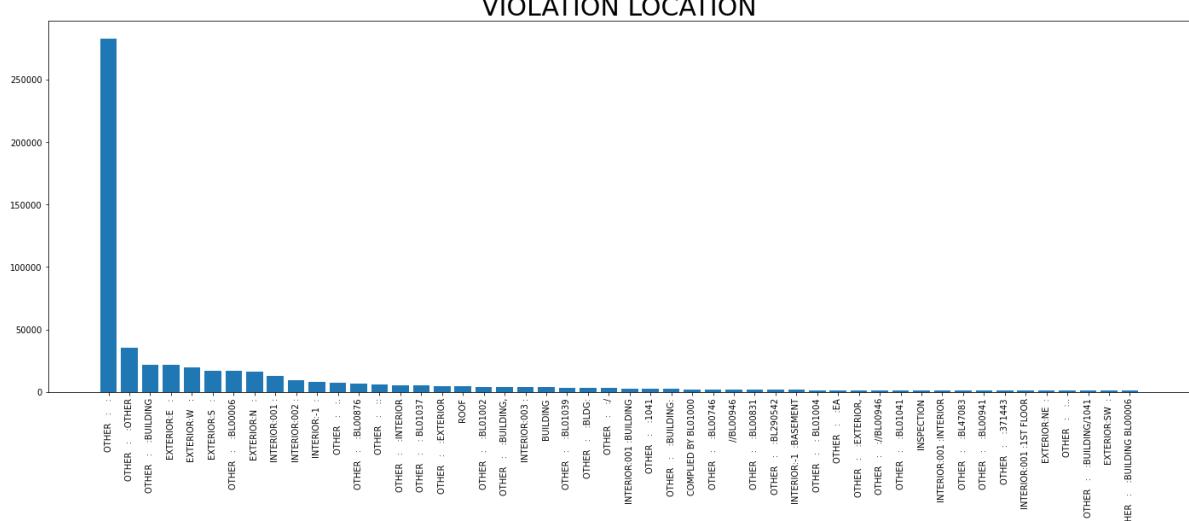
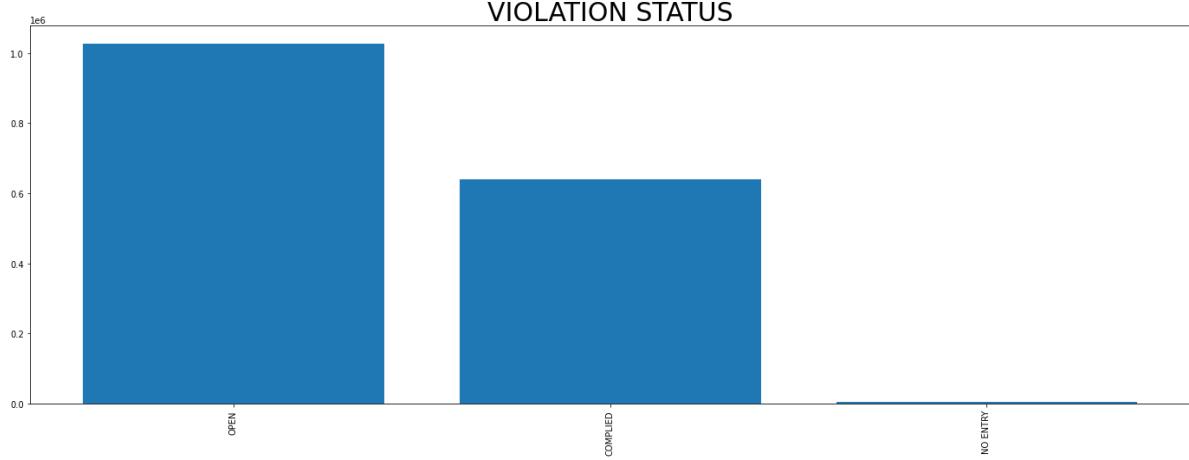
0

2.统计标称型数据的频数

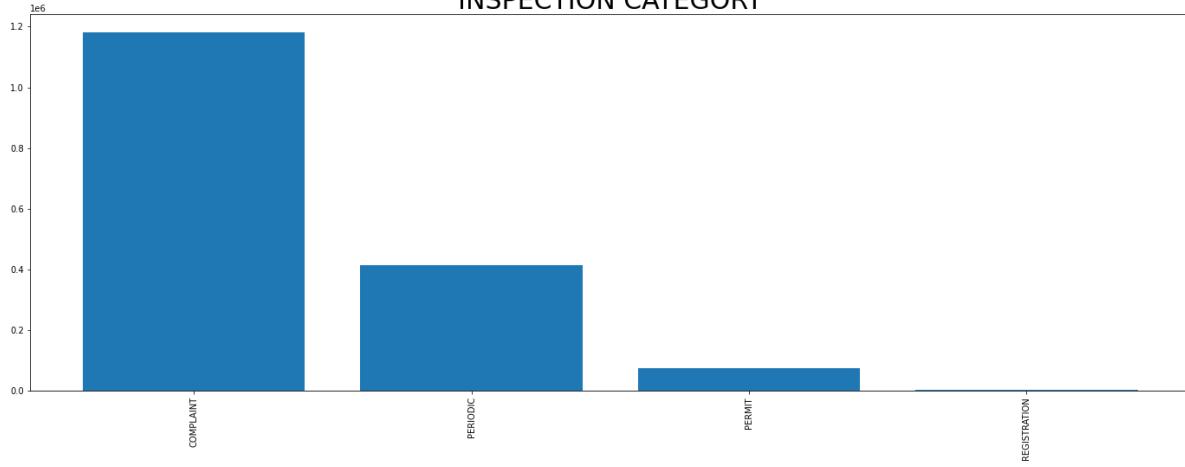
```

In [34]: #获取标称型数据
nominal_index_CBV = list(df_CBV.select_dtypes(['object']))
nominal_summary(df_CBV, nominal_index_CBV)

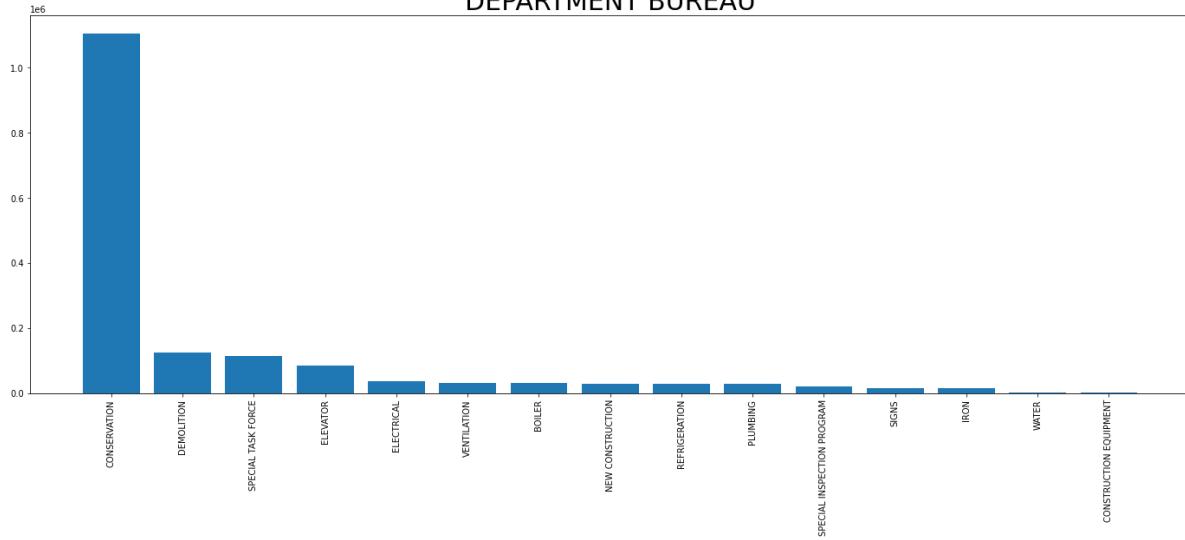
```



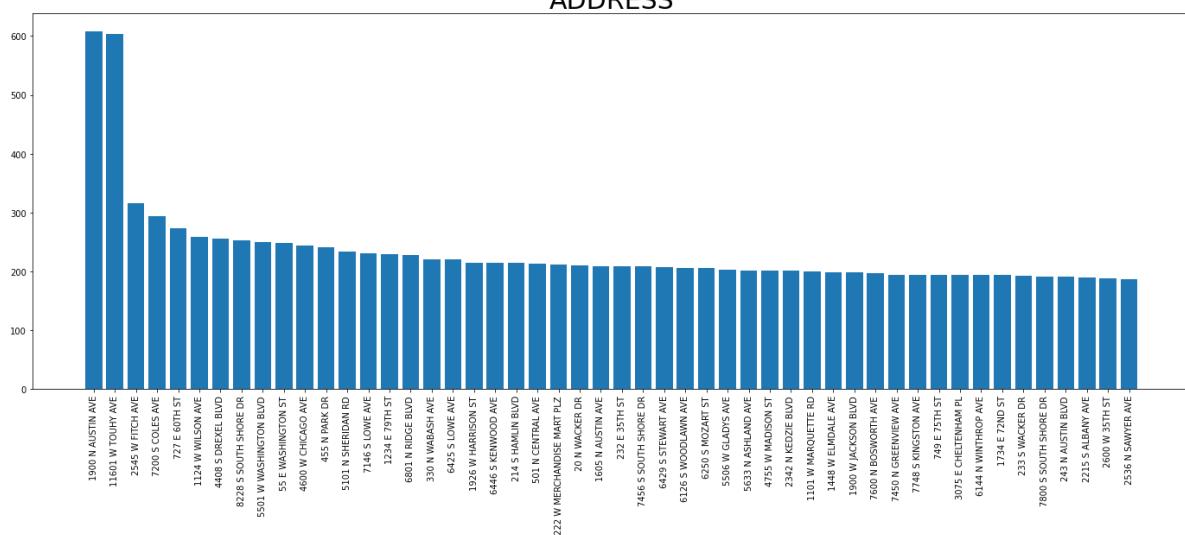
INSPECTION CATEGORY



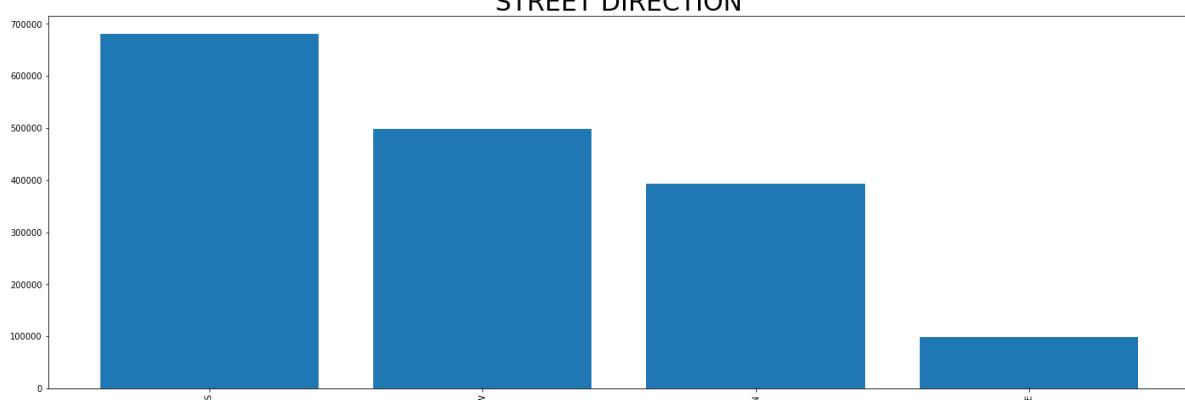
DEPARTMENT BUREAU

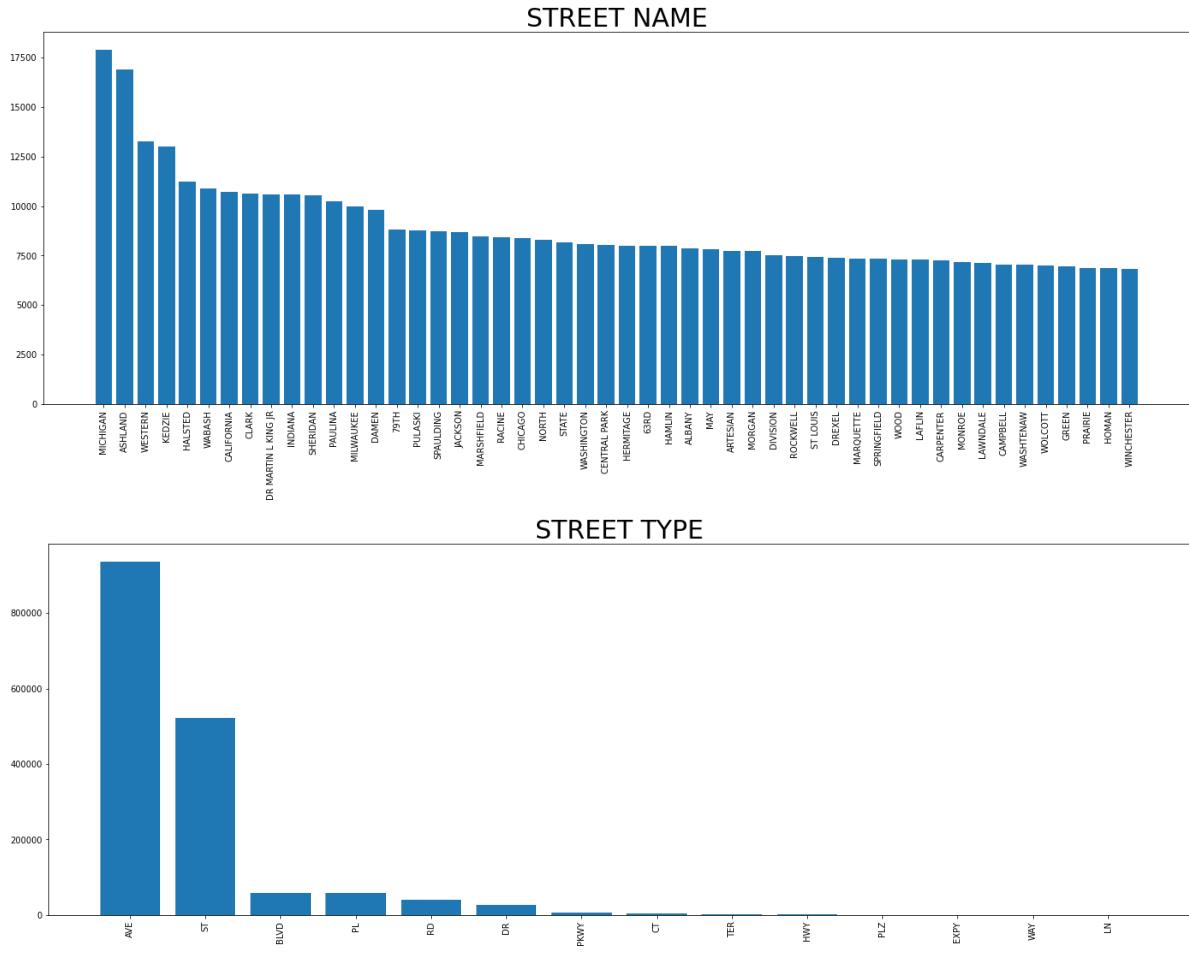


ADDRESS



STREET DIRECTION





3. 数值型数据的五数概括及缺失值数据

```
In [35]: #获得数据集的数值型数据
numerical_index_CBV = list(df_CBV.select_dtypes(['int', 'float']))
numerical_summary(df_CBV, numerical_index_CBV)
```

```

descriptive statistics (INSPECTION NUMBER):
Min: 265575.0000      Q1(25%): 2305167.0000    Q2(50%): 10418949.0000   Q3(75%): 116
87122.0000          Max: 13050915.0000
Missing: 0

descriptive statistics (STREET NUMBER):
Min: 1.0000      Q1(25%): 1648.0000      Q2(50%): 3747.0000      Q3(75%): 6229.0000
Max: 13770.0000
Missing: 0

descriptive statistics (PROPERTY GROUP):
Min: 1000.0000  Q1(25%): 20560.0000      Q2(50%): 154450.0000      Q3(75%): 367118.0000
Max: 677975.0000
Missing: 0

descriptive statistics (Community Areas):
Min: 1.0000      Q1(25%): 24.0000      Q2(50%): 36.0000      Q3(75%): 58.0000
Max: 77.0000
Missing: 2276

descriptive statistics (Zip Codes):
Min: 4299.0000  Q1(25%): 21190.0000      Q2(50%): 21569.0000      Q3(75%): 22248.0000
Max: 22620.0000
Missing: 1509

descriptive statistics (Boundaries - ZIP Codes):
Min: 1.0000      Q1(25%): 15.0000      Q2(50%): 28.0000      Q3(75%): 52.0000
Max: 61.0000
Missing: 2276

descriptive statistics (Census Tracts):
Min: 1.0000      Q1(25%): 179.0000      Q2(50%): 374.0000      Q3(75%): 572.0000
Max: 801.0000
Missing: 1544

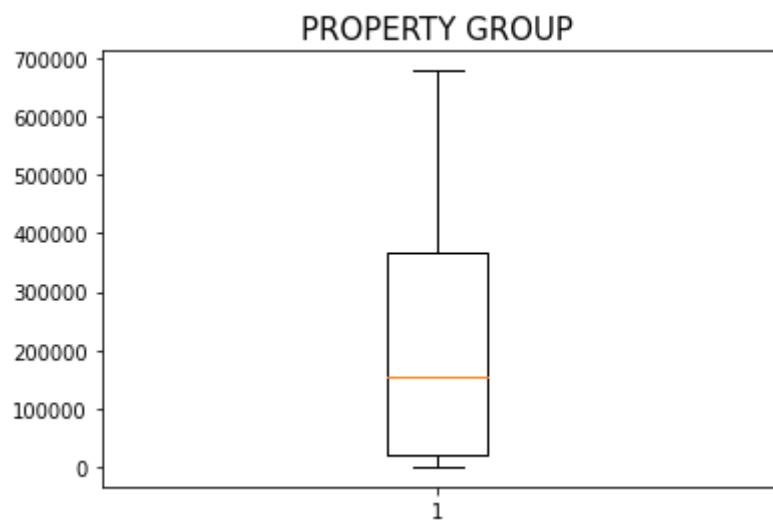
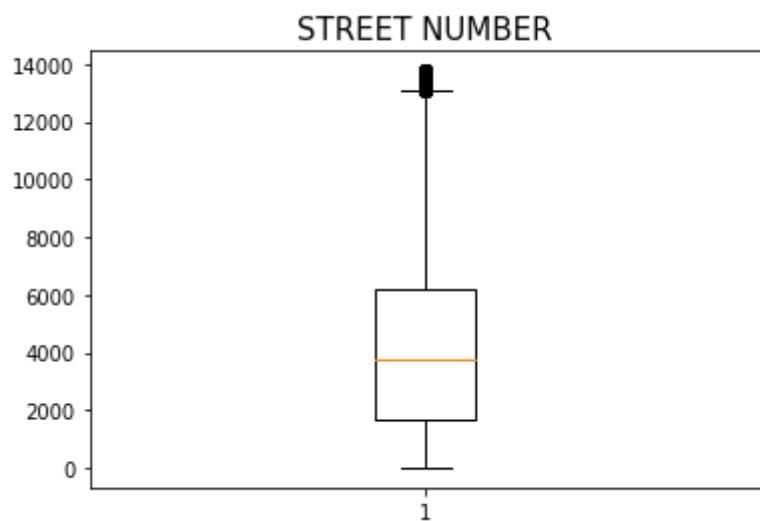
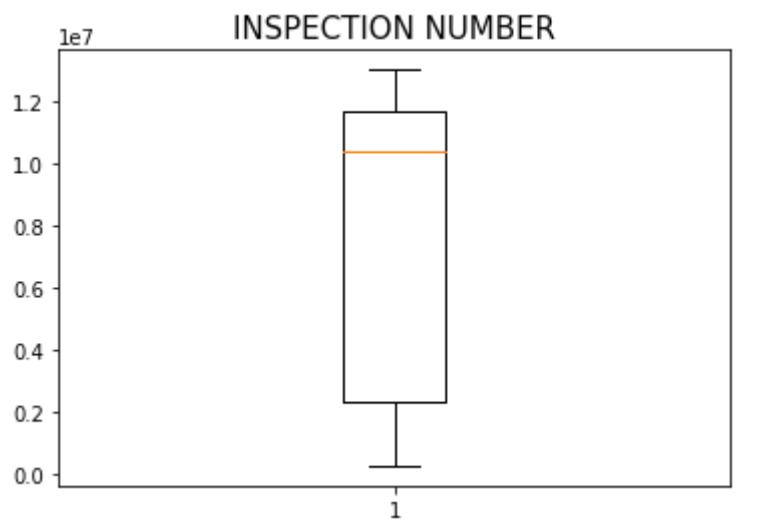
descriptive statistics (Wards):
Min: 1.0000      Q1(25%): 12.0000      Q2(50%): 25.0000      Q3(75%): 37.0000
Max: 50.0000
Missing: 2276

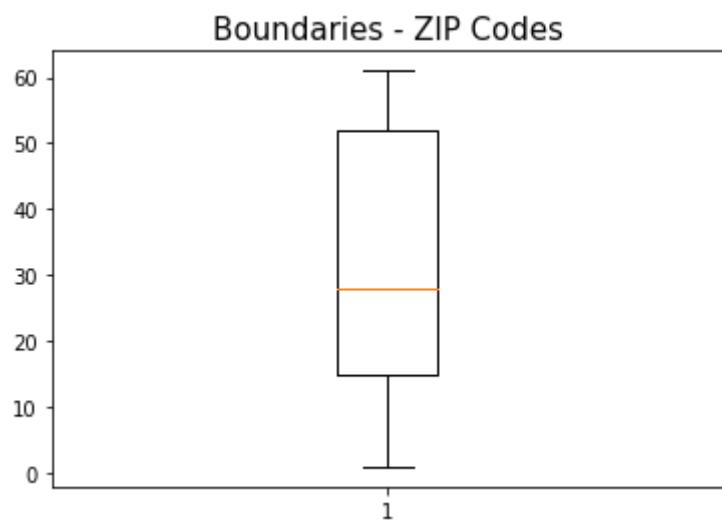
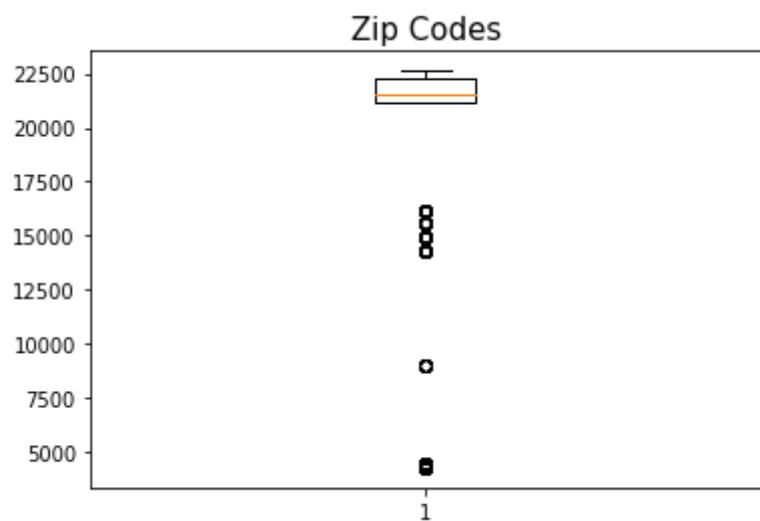
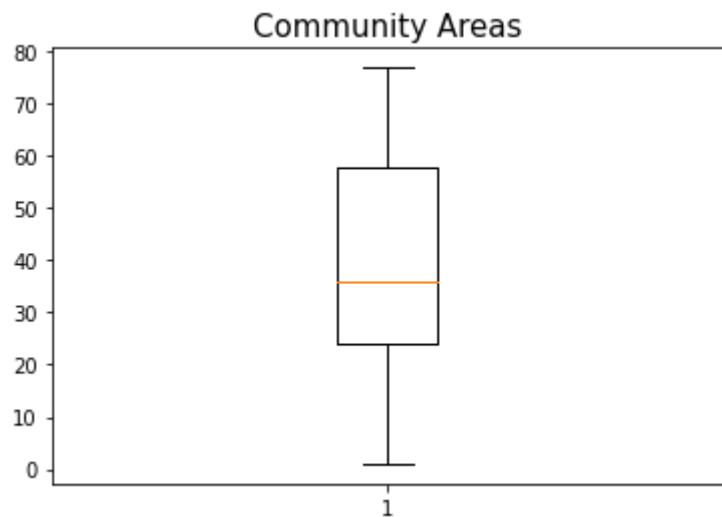
descriptive statistics (Historical Wards 2003–2015):
Min: 1.0000      Q1(25%): 14.0000      Q2(50%): 28.0000      Q3(75%): 41.0000
Max: 53.0000
Missing: 2276

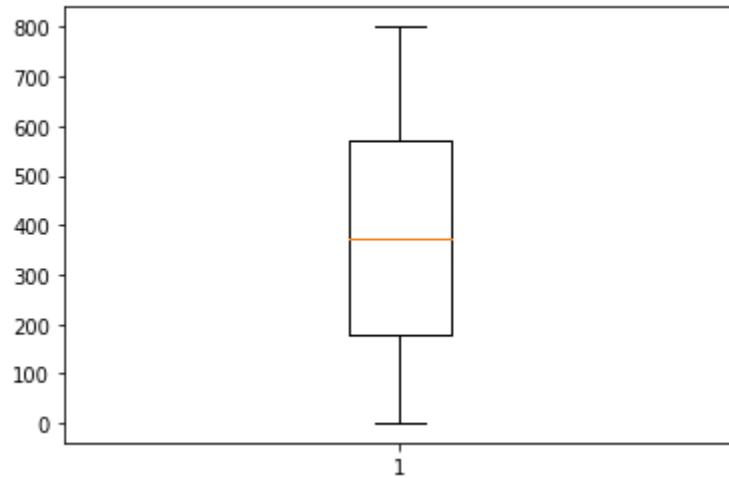
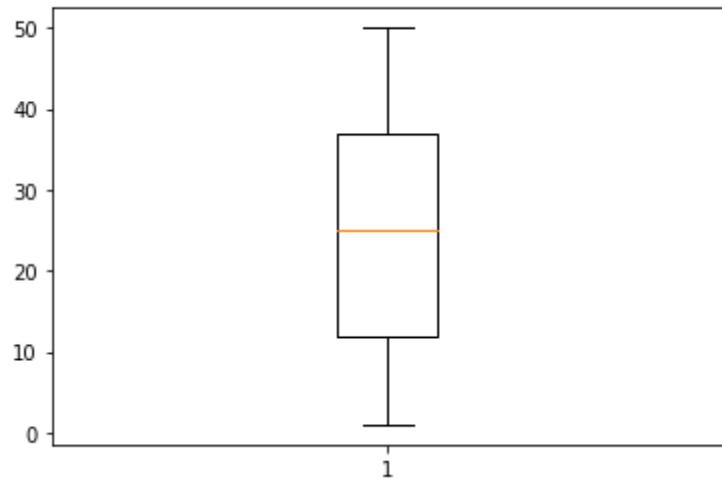
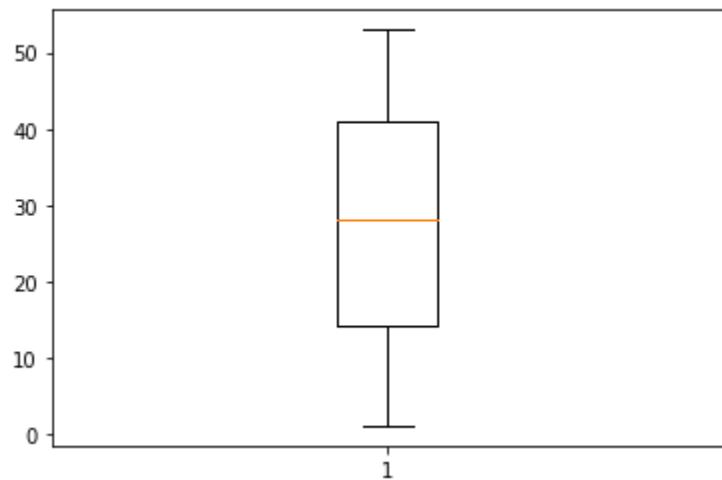
```

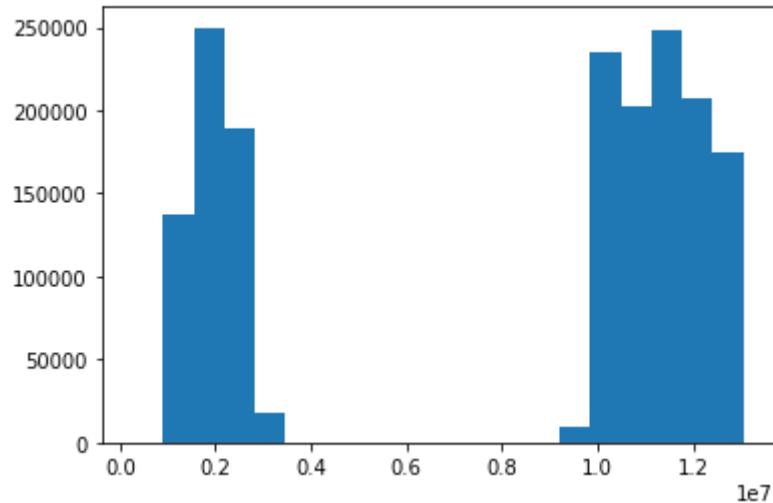
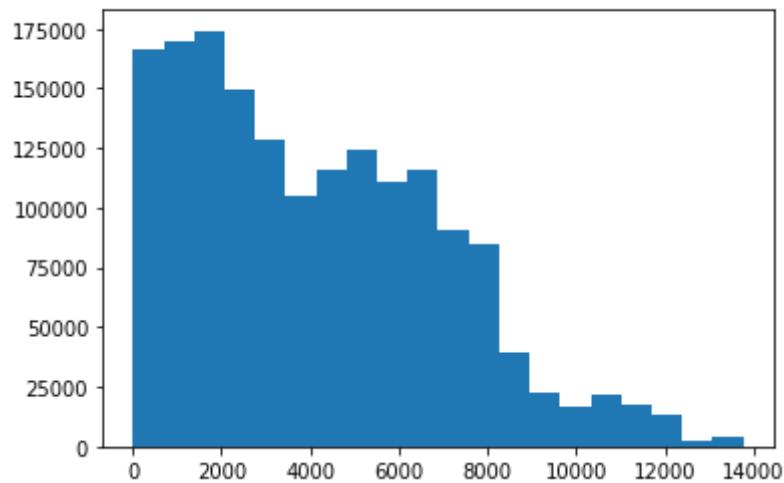
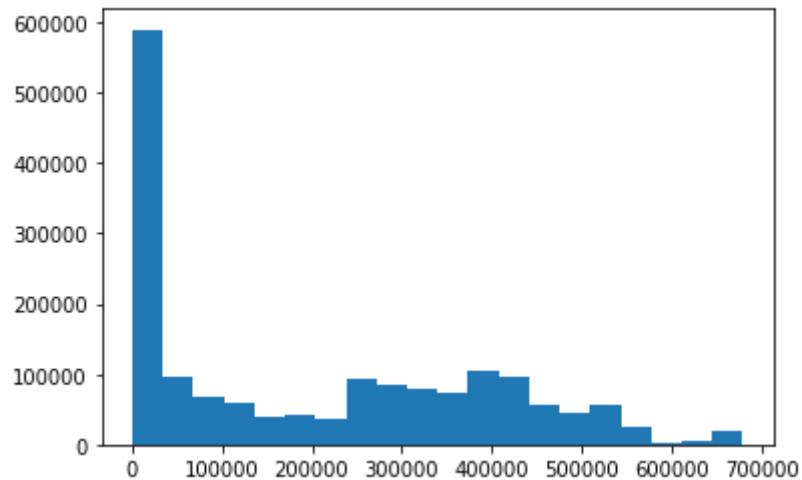
4.数据可视化 使用直方图、盒图等检查数据分布及离群点

```
In [38]: box_plot(df_CBV, numerical_index_CBV)
hist_plot(df_CBV, numerical_index_CBV)
```

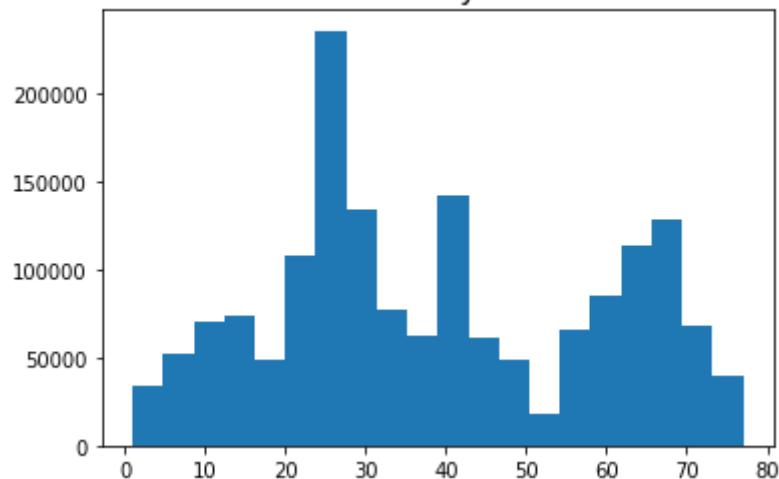




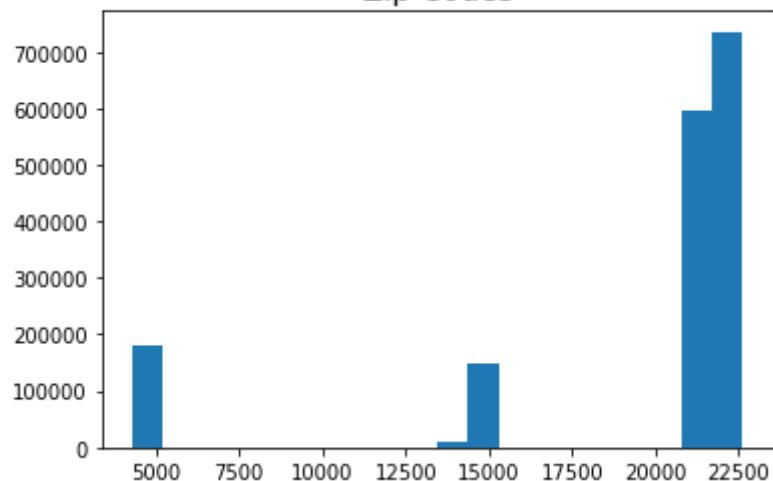
Census Tracts**Wards****Historical Wards 2003-2015**

INSPECTION NUMBER**STREET NUMBER****PROPERTY GROUP**

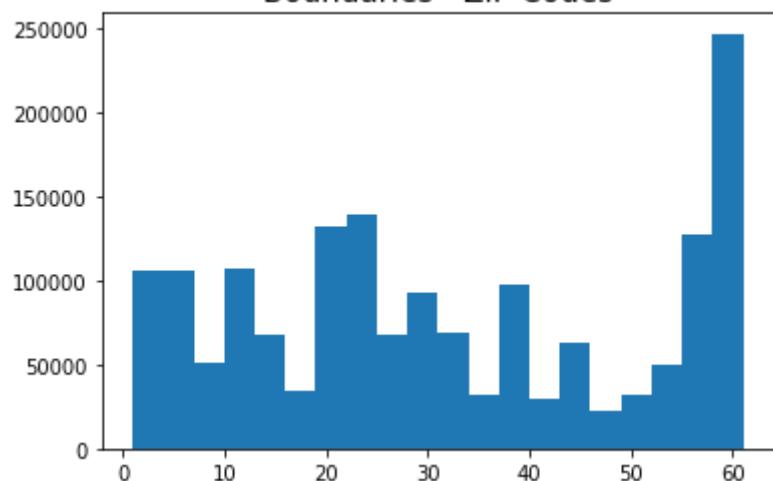
Community Areas

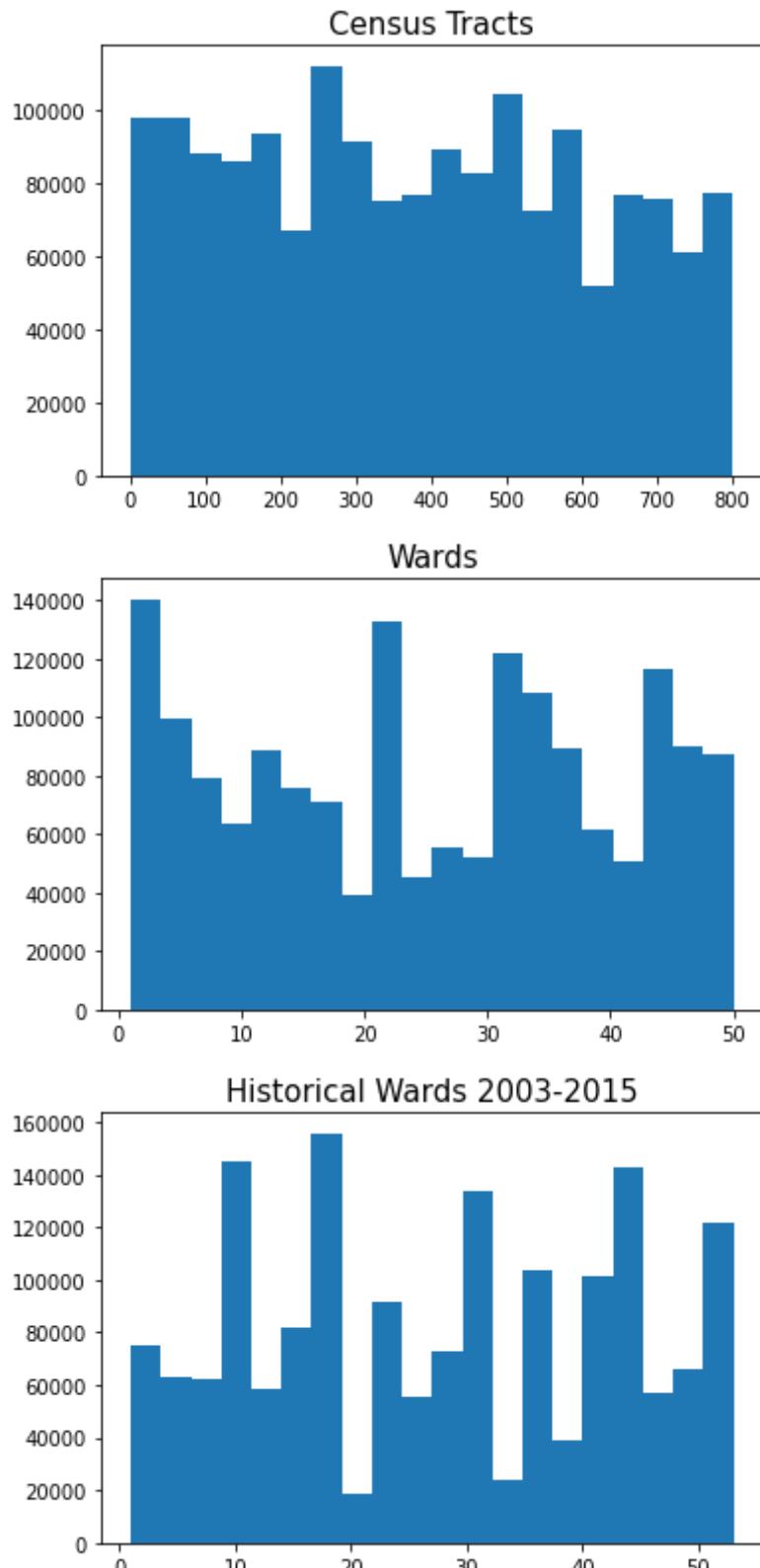


Zip Codes



Boundaries - ZIP Codes



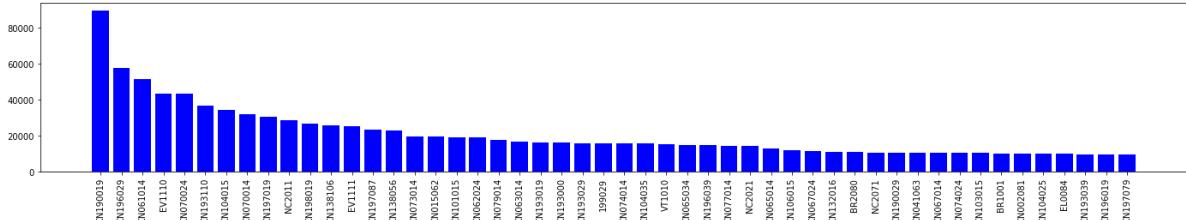


5. 观察数据集中缺失数据，分析其缺失的原因。

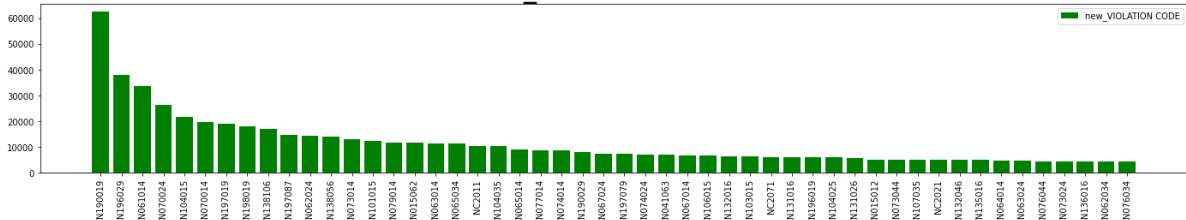
5.1 将缺失部分删除

```
In [39]: new_df_CBV = miss_drop(df_CBV)
nominal_summary(df_CBV, nominal_index_CBV, new_df_CBV)
box_plot(df_CBV, numerical_index_CBV, new_df_CBV)
hist_plot(df_CBV, numerical_index_CBV, new_df_CBV)
```

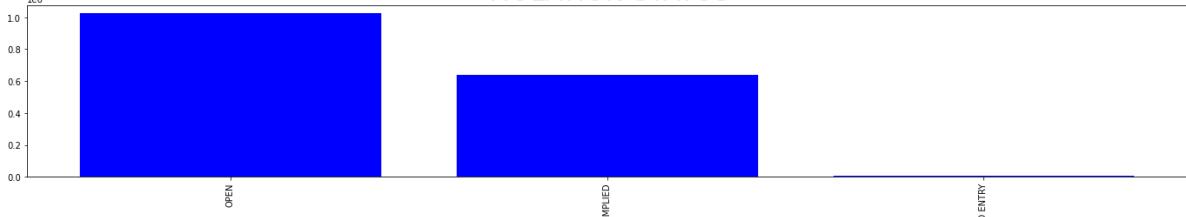
VIOLATION CODE



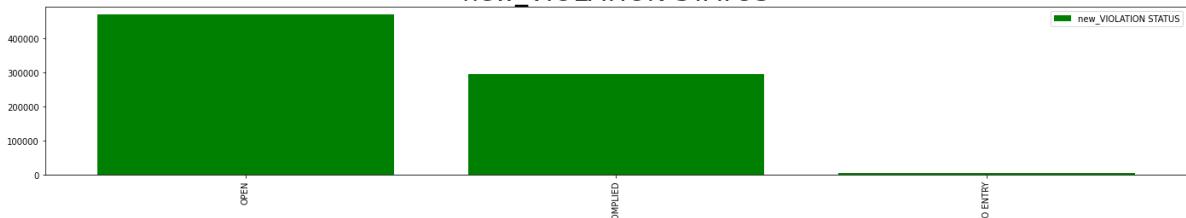
new VIOLATION CODE



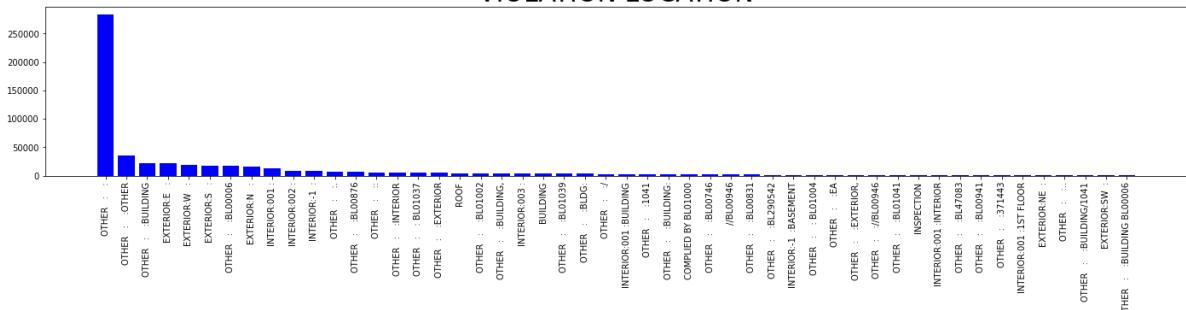
VIOLATION STATUS



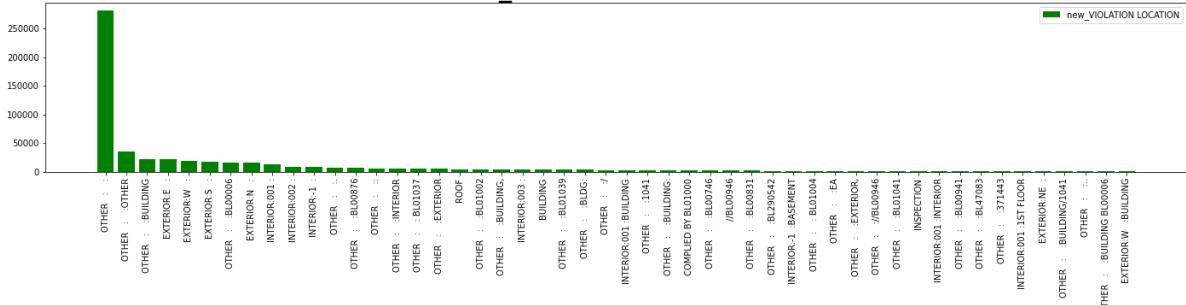
new VIOLATION STATUS



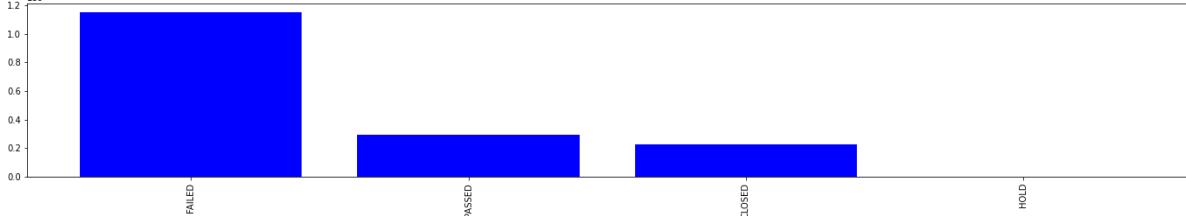
VIOLATION LOCATION



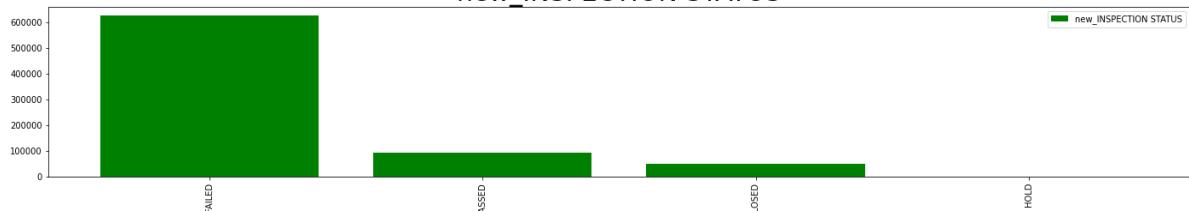
new VIOLATION LOCATION



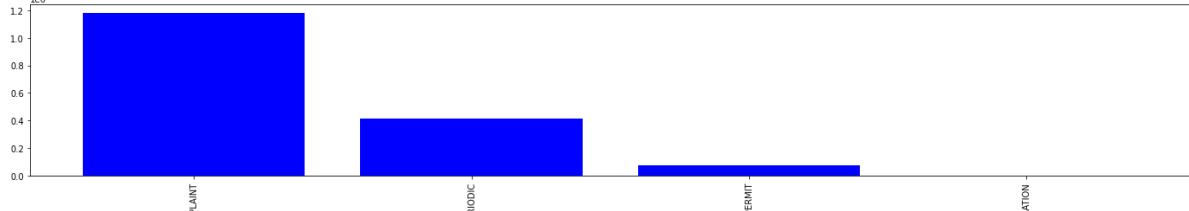
INSPECTION STATUS



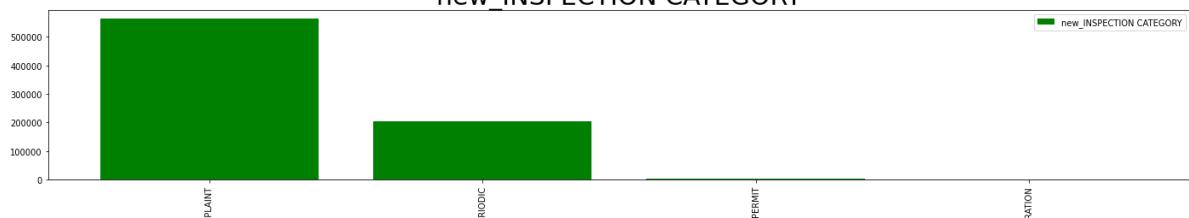
new_INSPECTION STATUS



INSPECTION CATEGORY



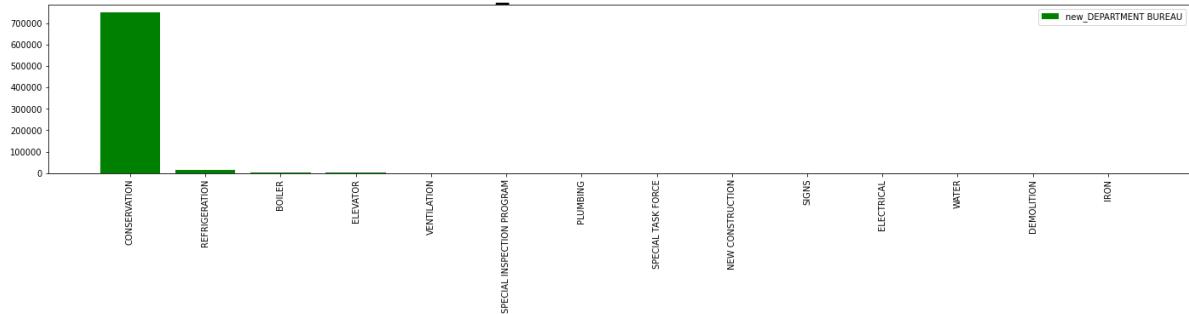
new_INSPECTION CATEGORY



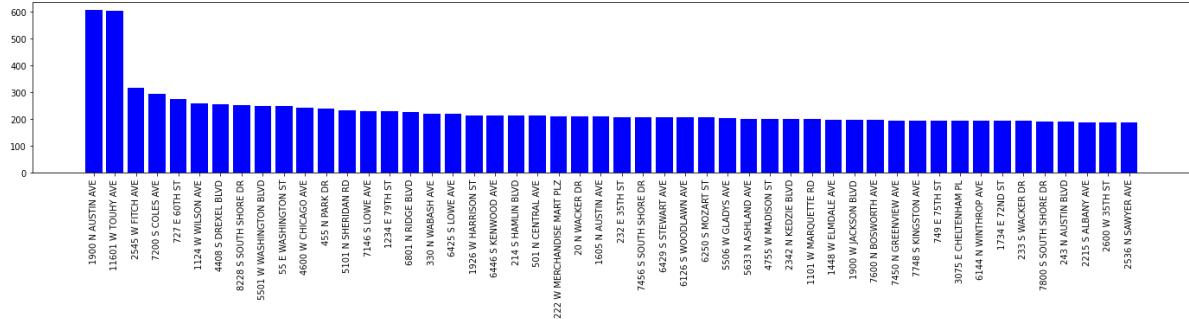
DEPARTMENT BUREAU



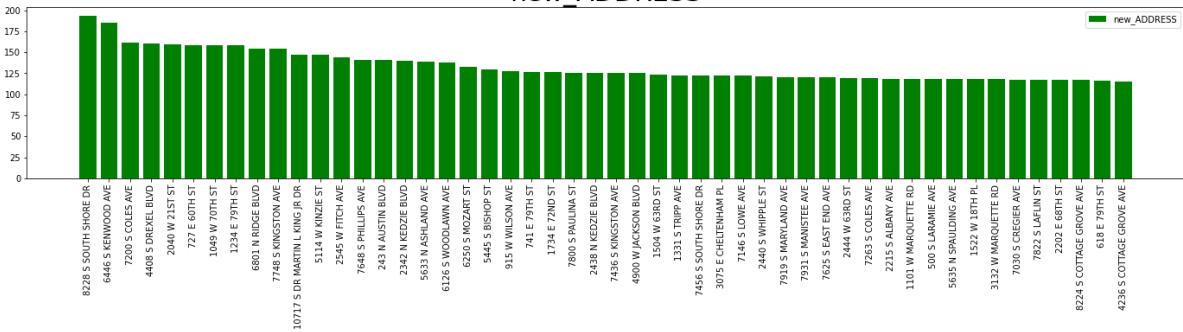
new_DEPARTMENT BUREAU



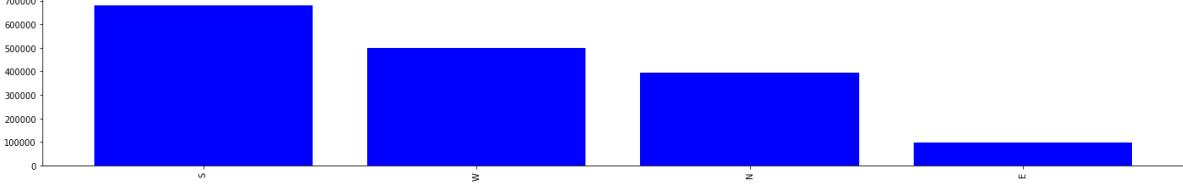
ADDRESS



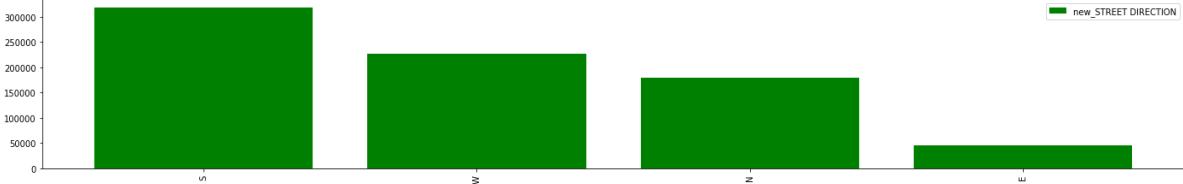
new ADDRESS



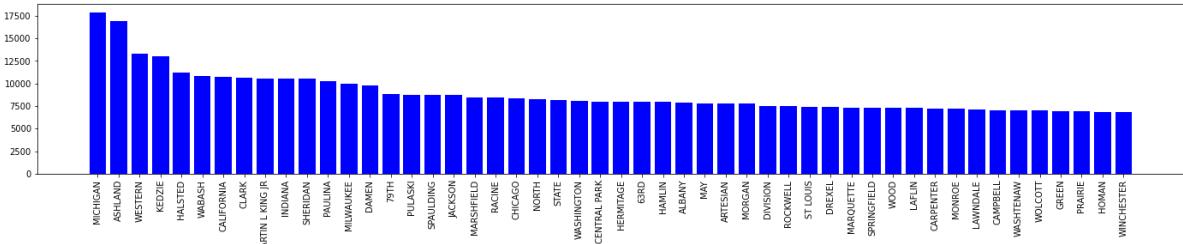
STREET DIRECTION



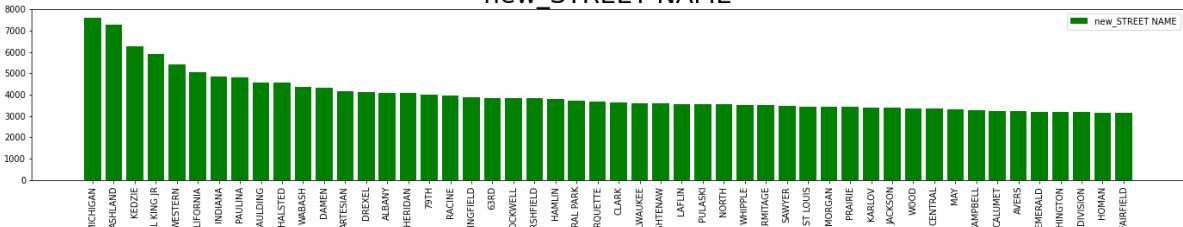
new STREET DIRECTION



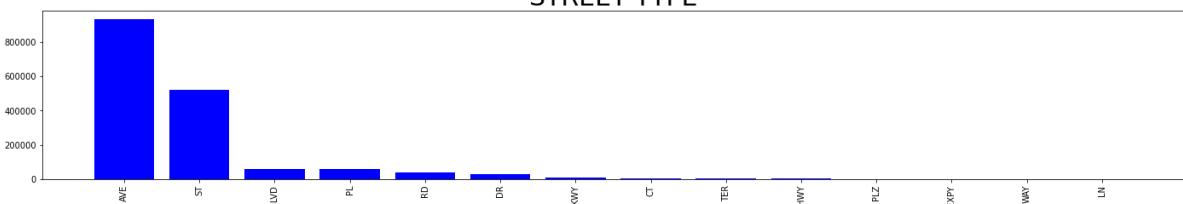
STREET NAME



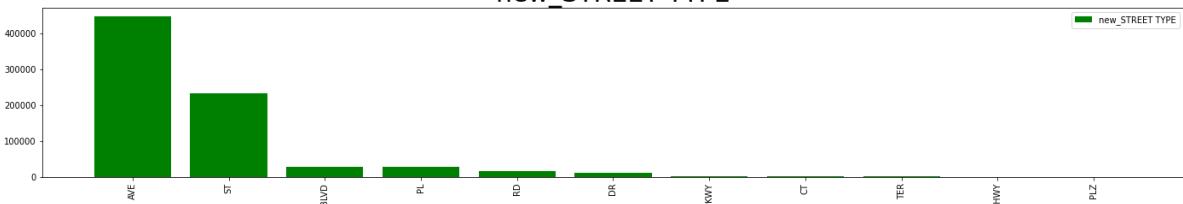
new STREET NAME

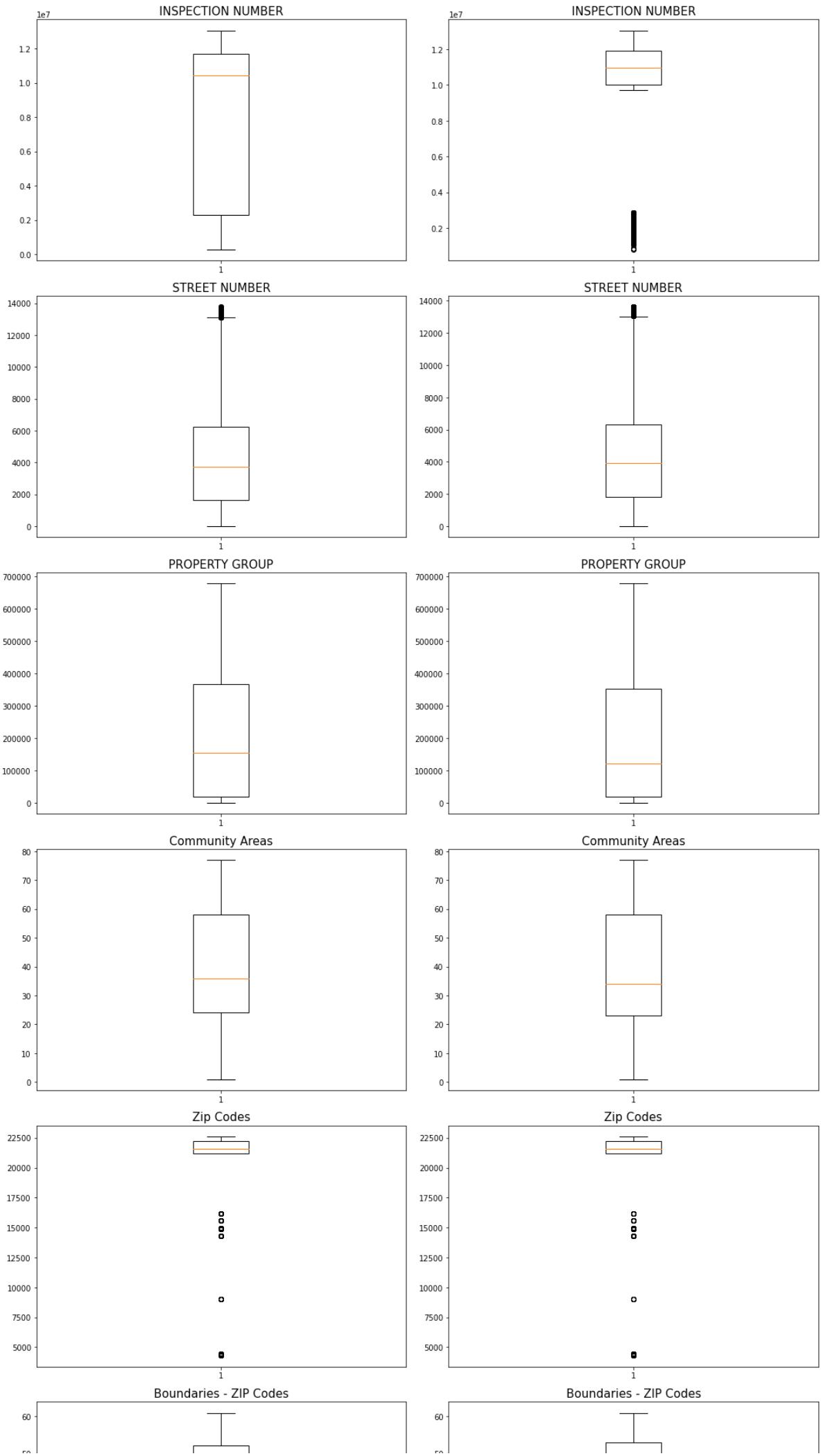


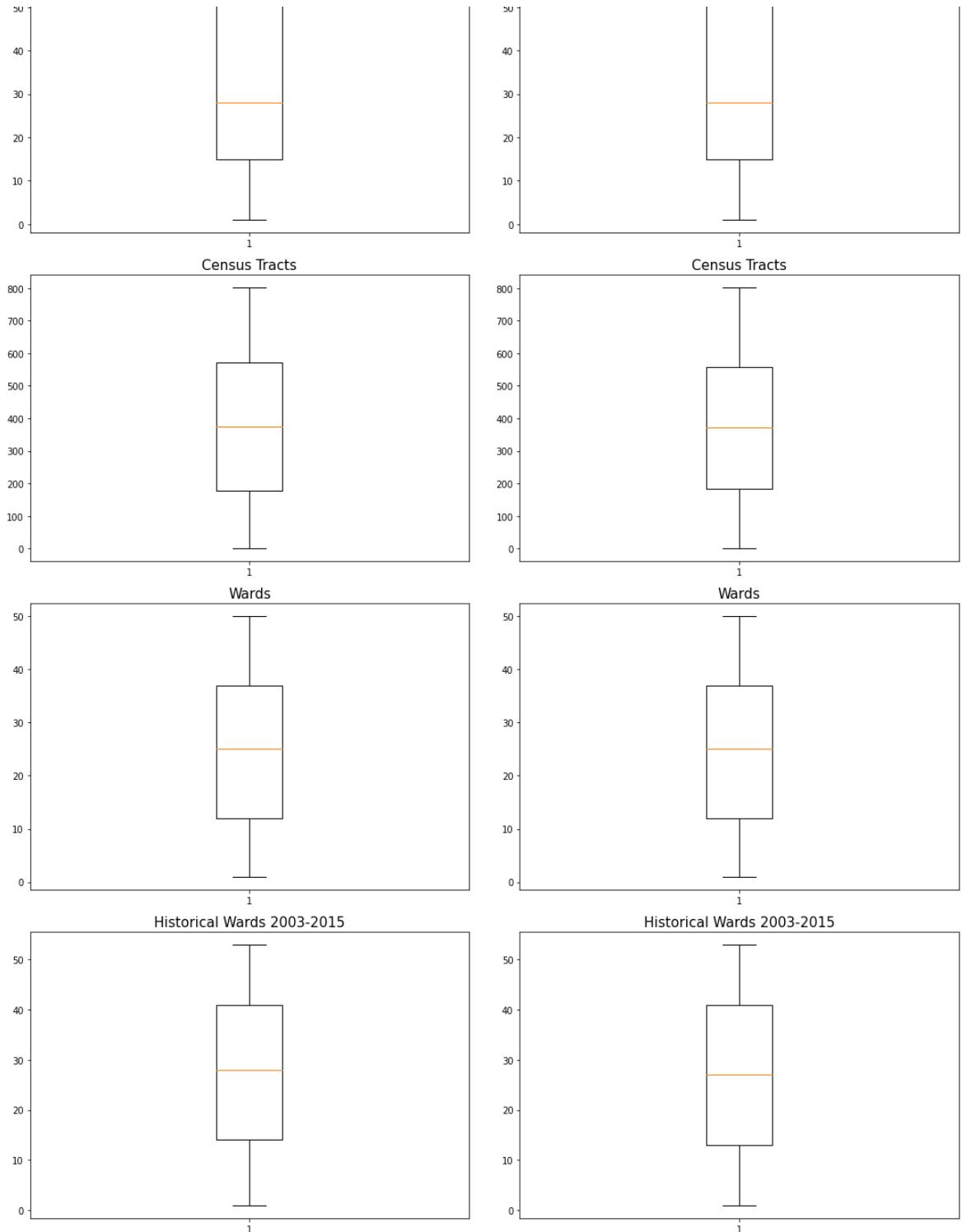
STREET TYPE

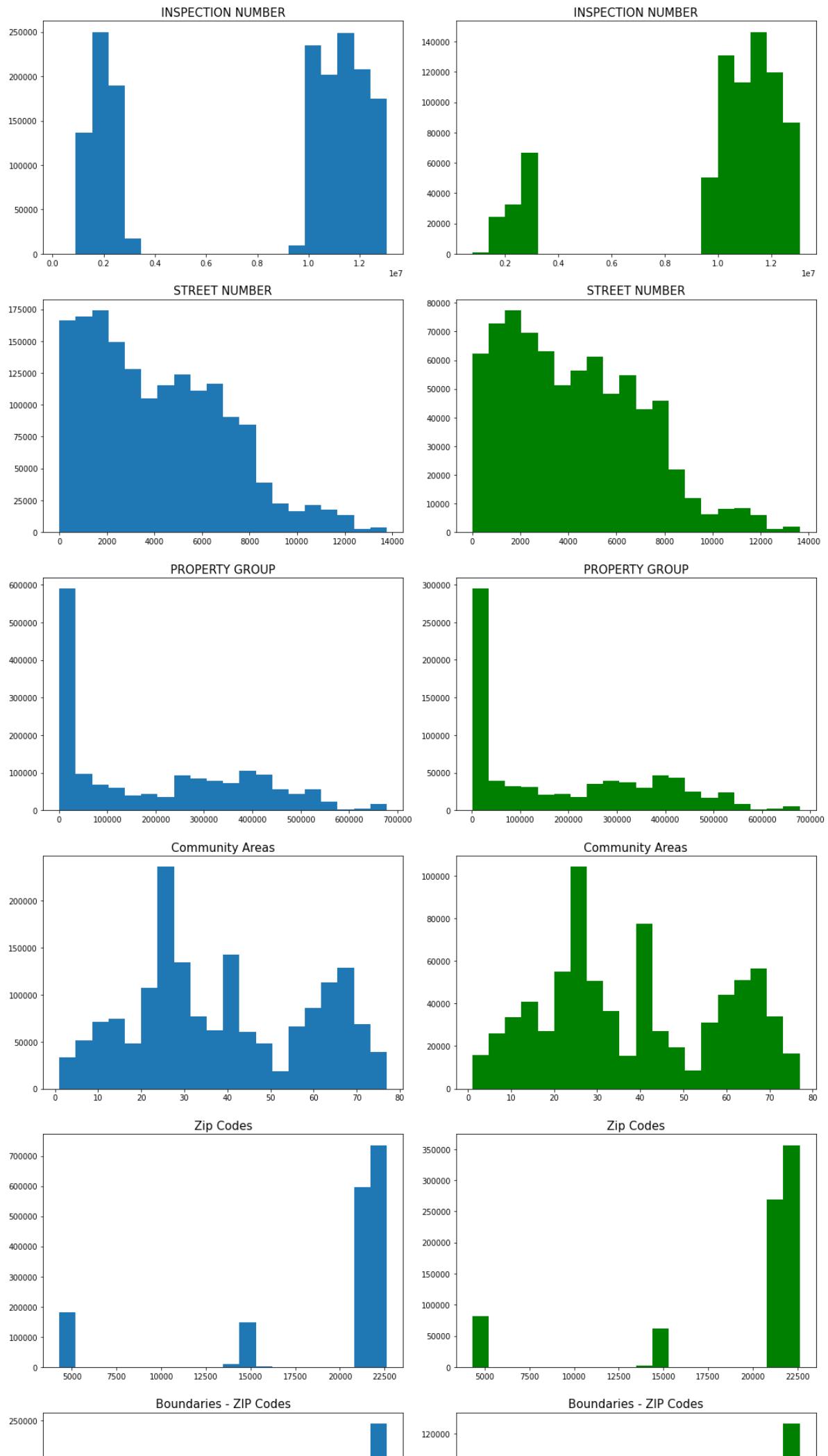


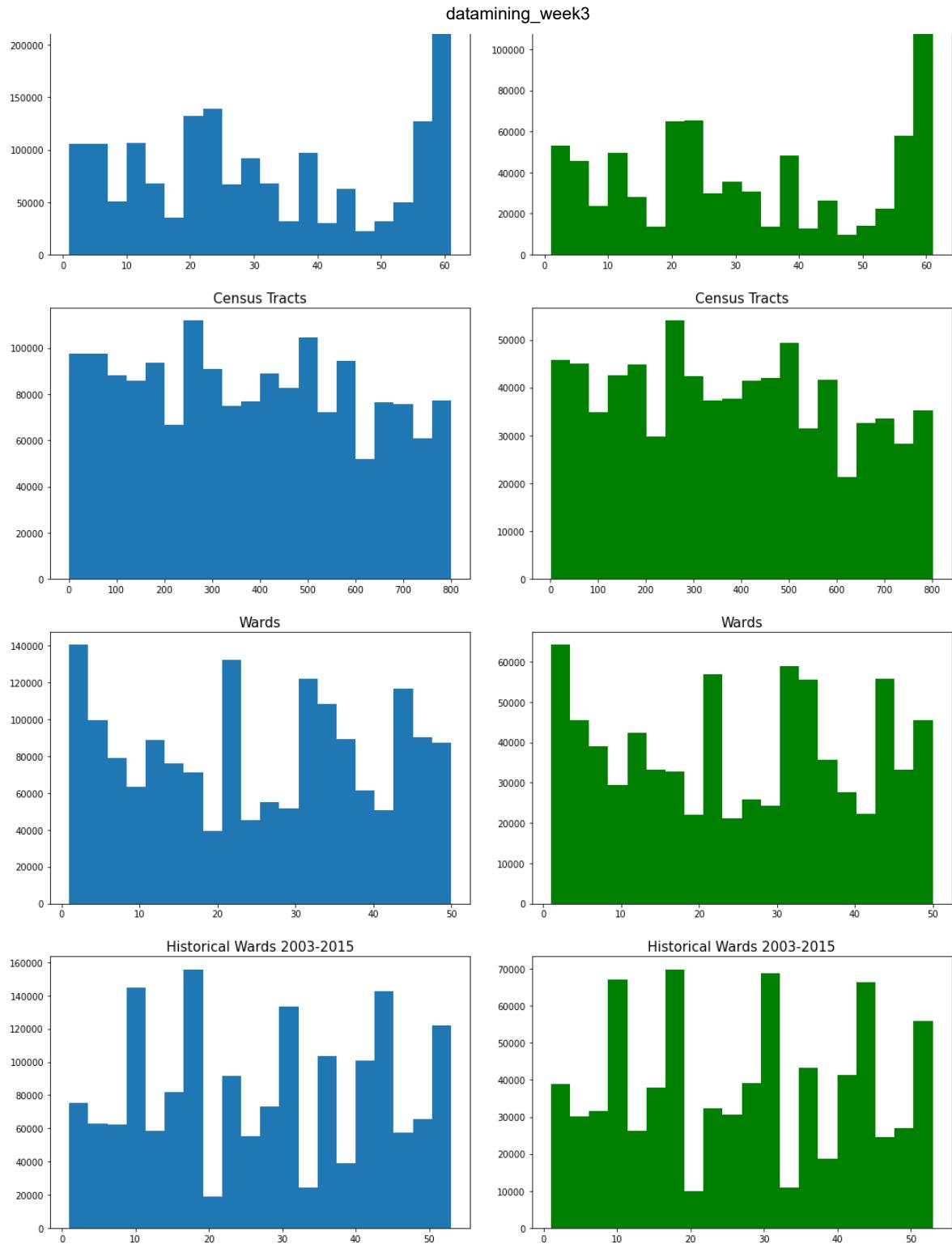
new STREET TYPE







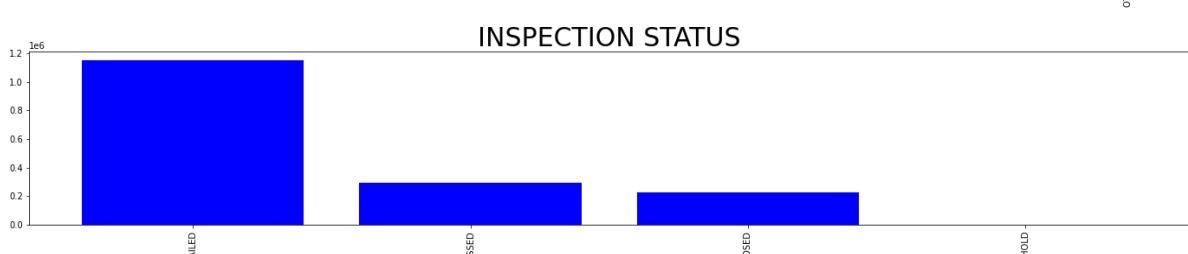
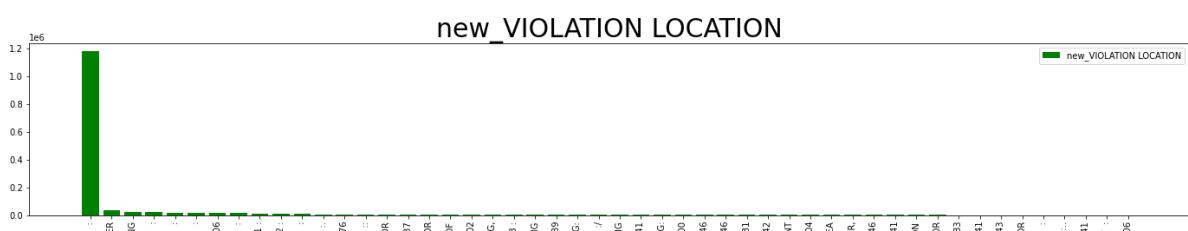
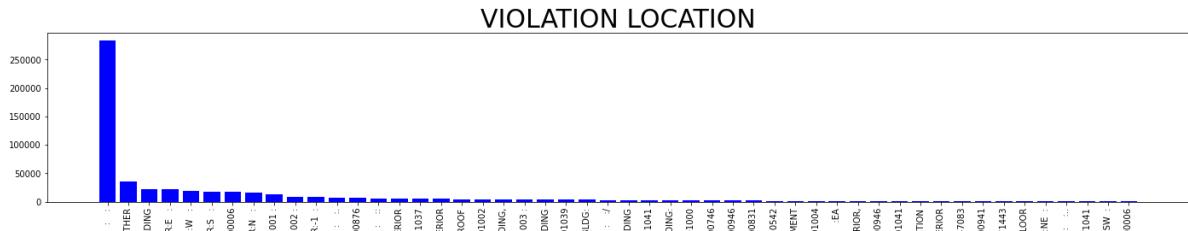
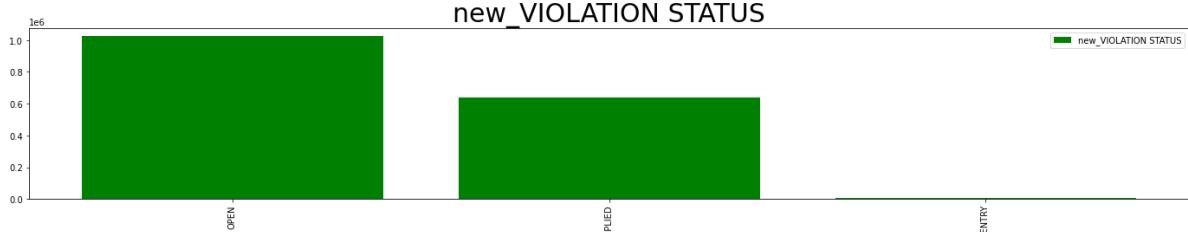
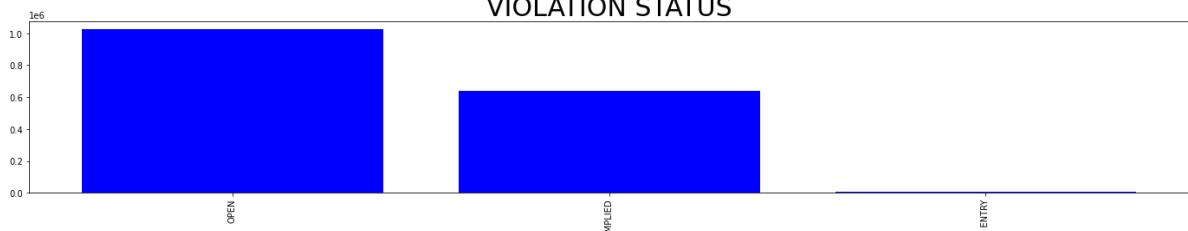
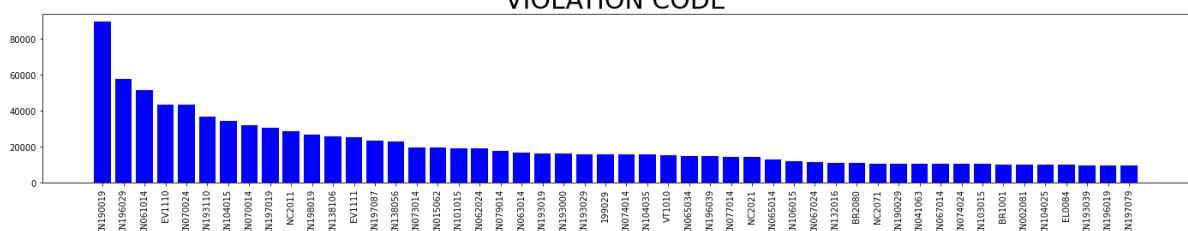




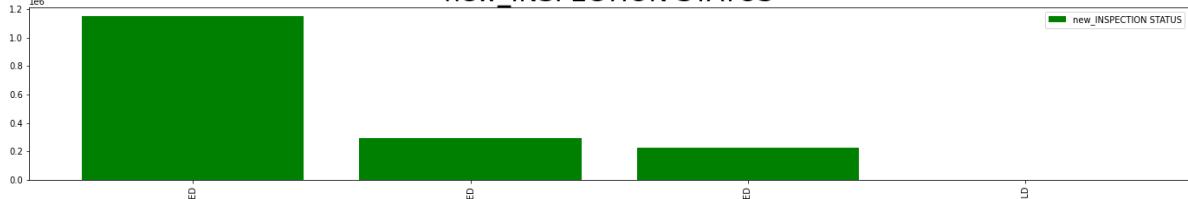
[分析]通过图示可以分析得出数值属性INSPECTION NUMBER与原始数据集相差过大，多个标称属性也出现了数据丢失过多的情况，说明此种方式处理缺失值不可取。

5.2用最高频率值来填补缺失值

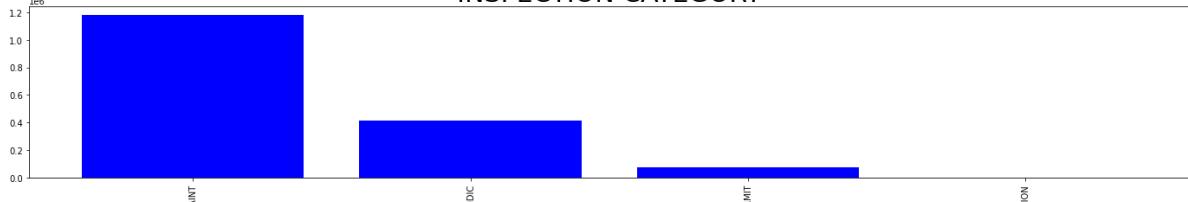
```
In [40]: new_df_CBV = miss_mode_fill(df_CBV)
nominal_summary(df_CBV, nominal_index_CBV, new_df_CBV)
box_plot(df_CBV, numerical_index_CBV, new_df_CBV)
hist_plot(df_CBV, numerical_index_CBV, new_df_CBV)
```



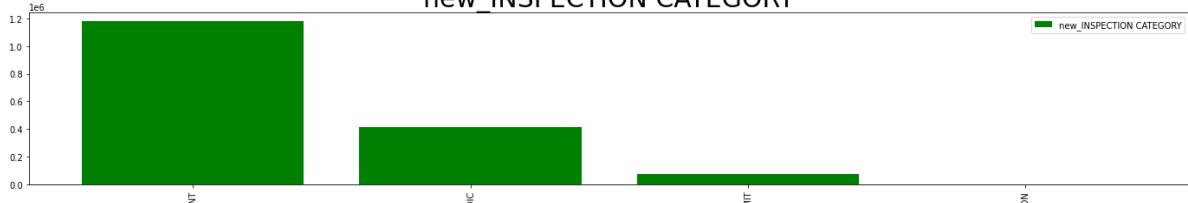
new_INSPECTION STATUS



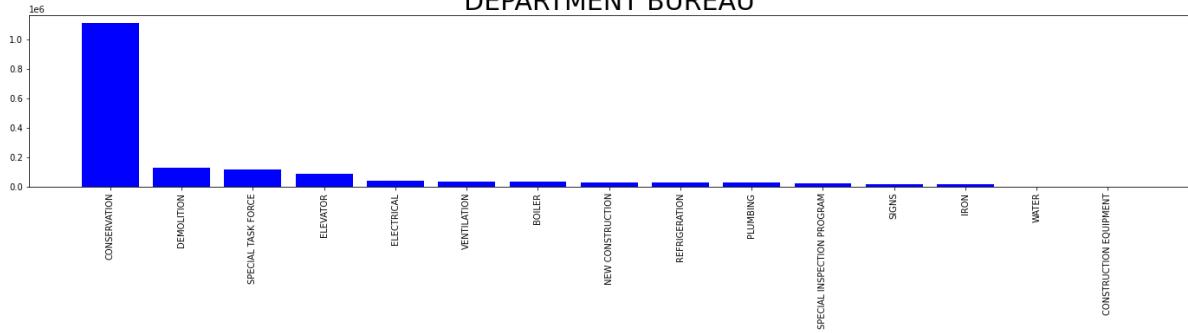
INSPECTION CATEGORY



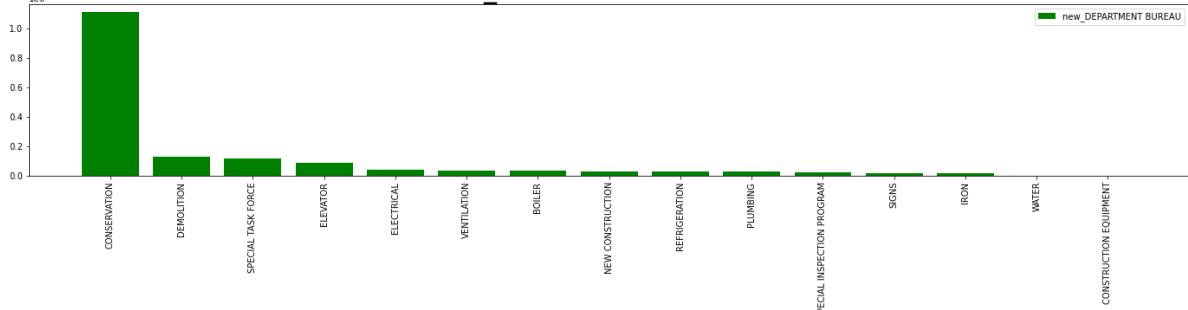
new_INSPECTION CATEGORY



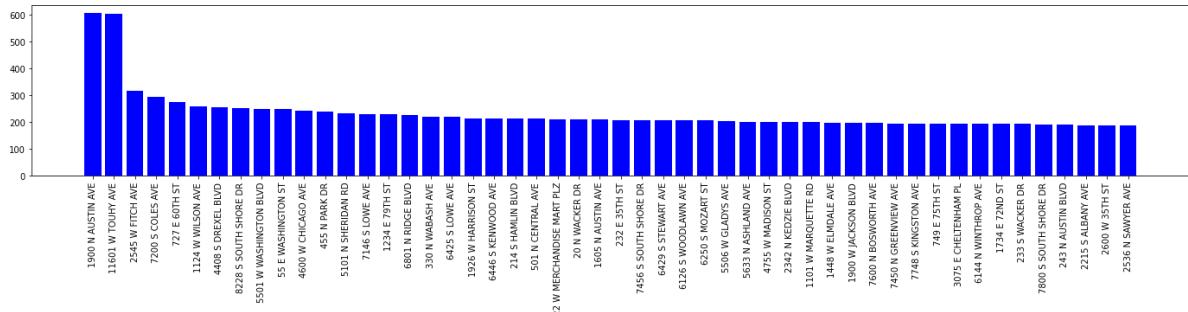
DEPARTMENT BUREAU



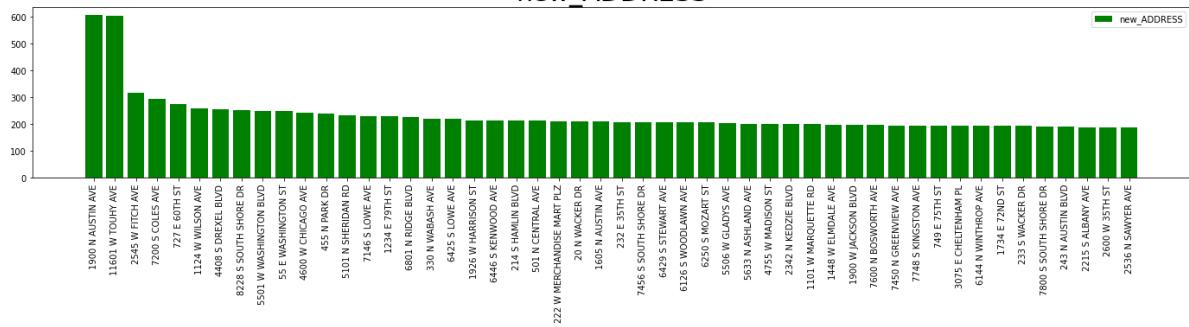
new_DEPARTMENT BUREAU



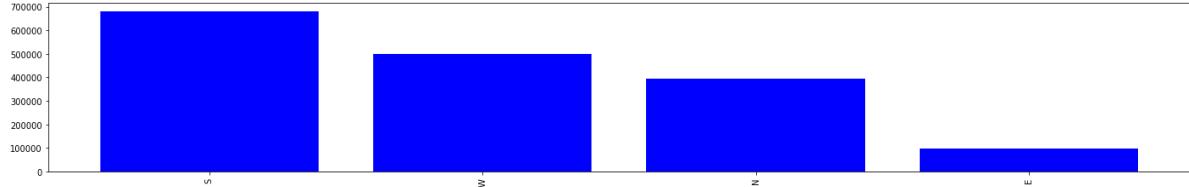
ADDRESS



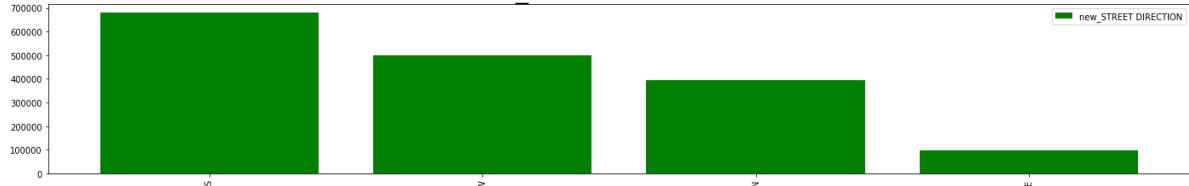
new_ADDRESS



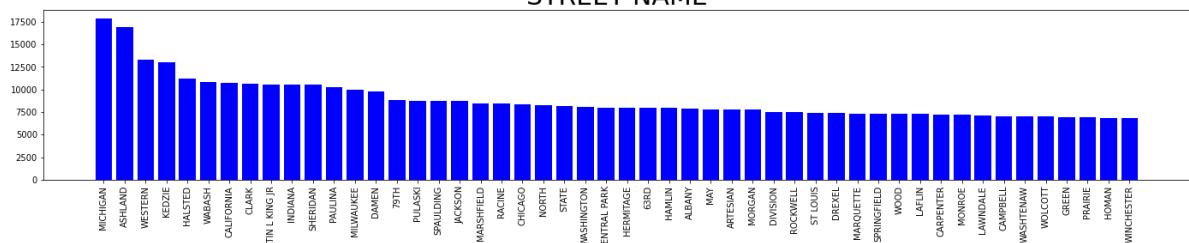
STREET DIRECTION



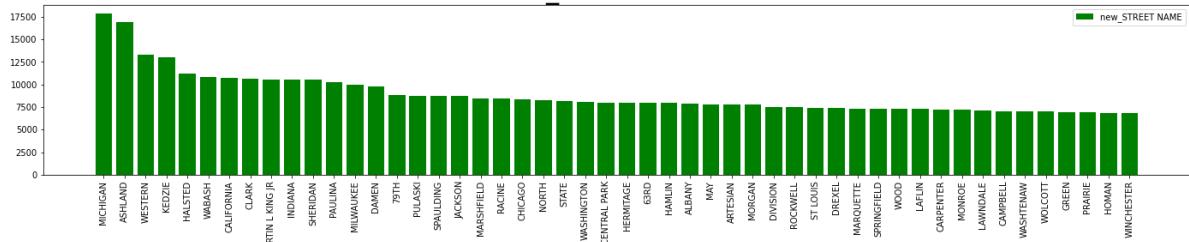
new_STREET DIRECTION



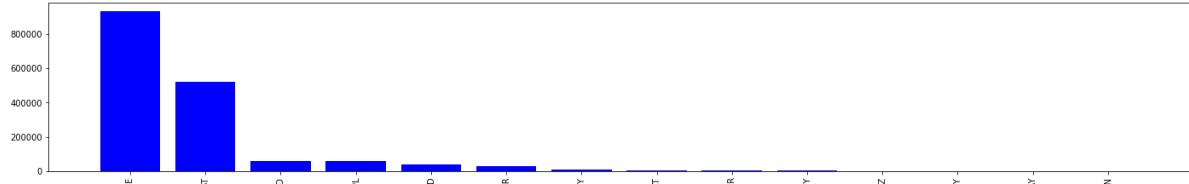
STREET NAME



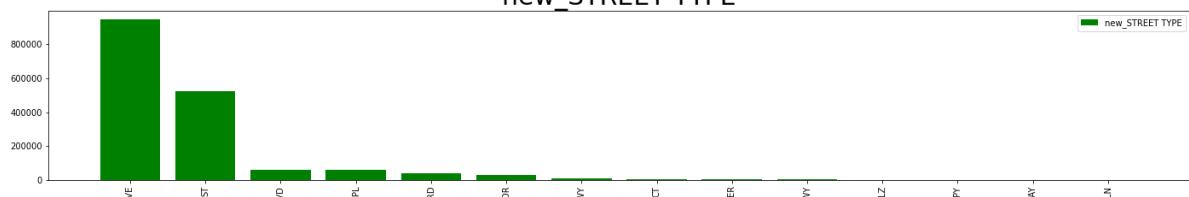
new_STREET NAME

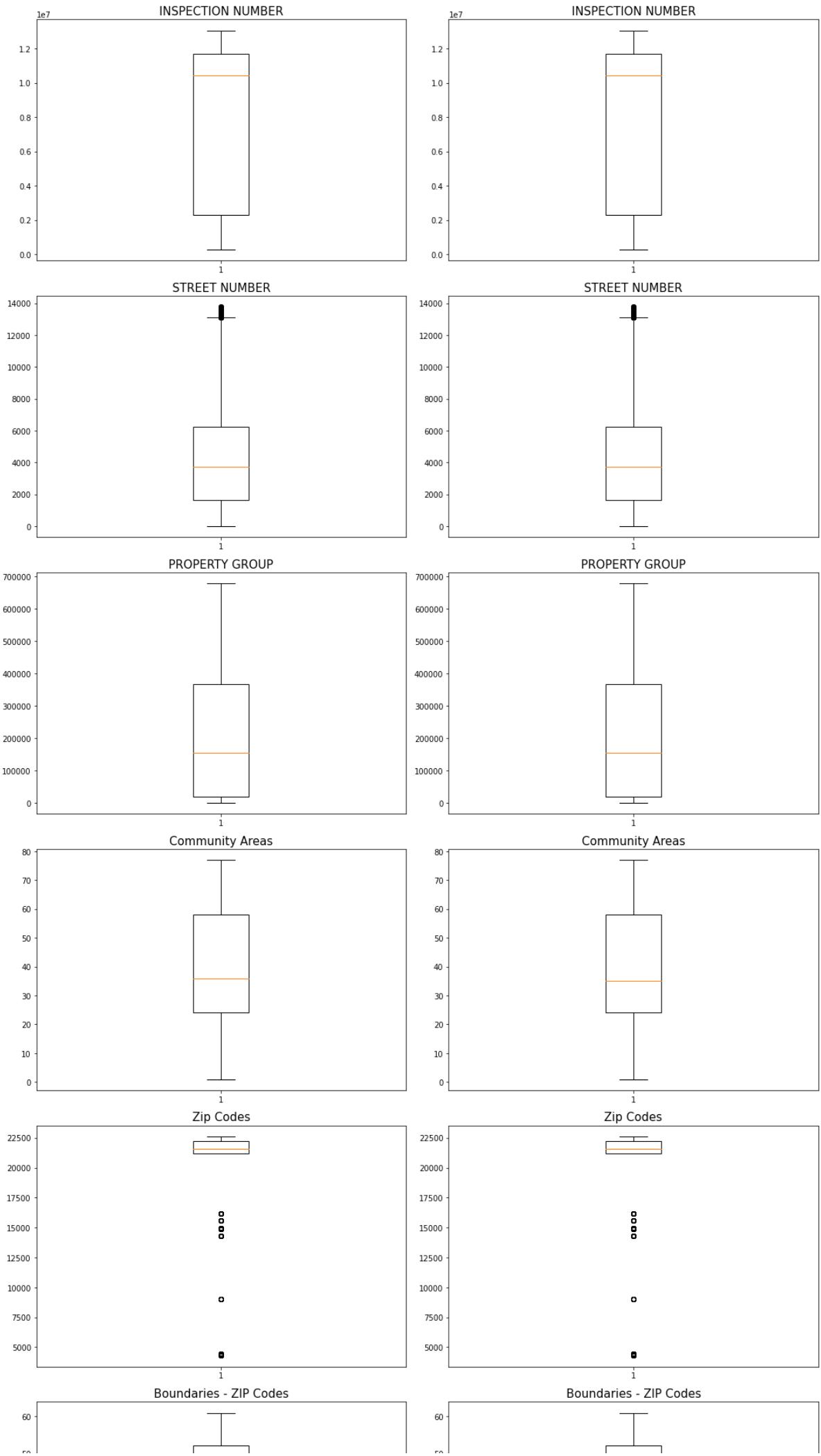


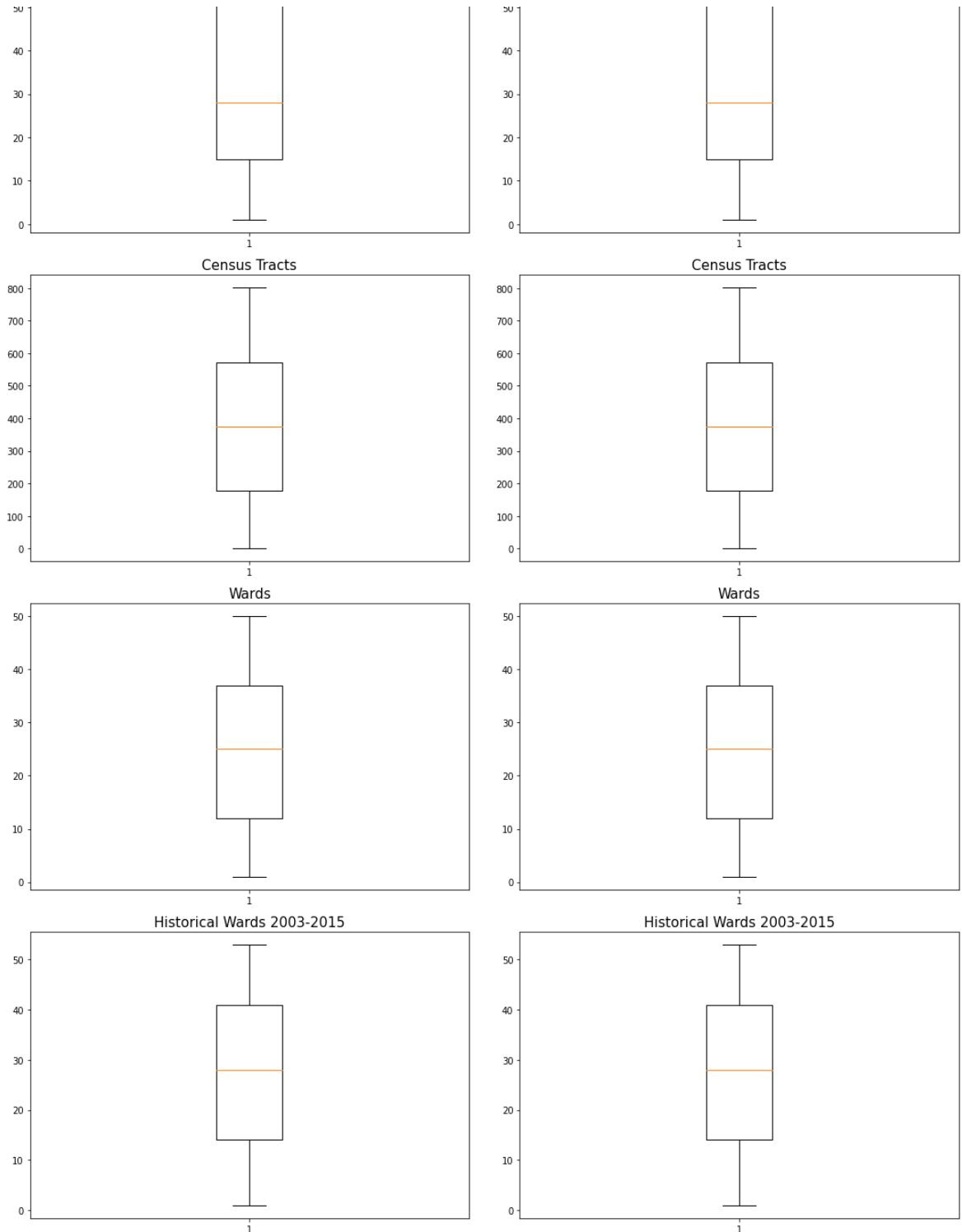
STREET TYPE

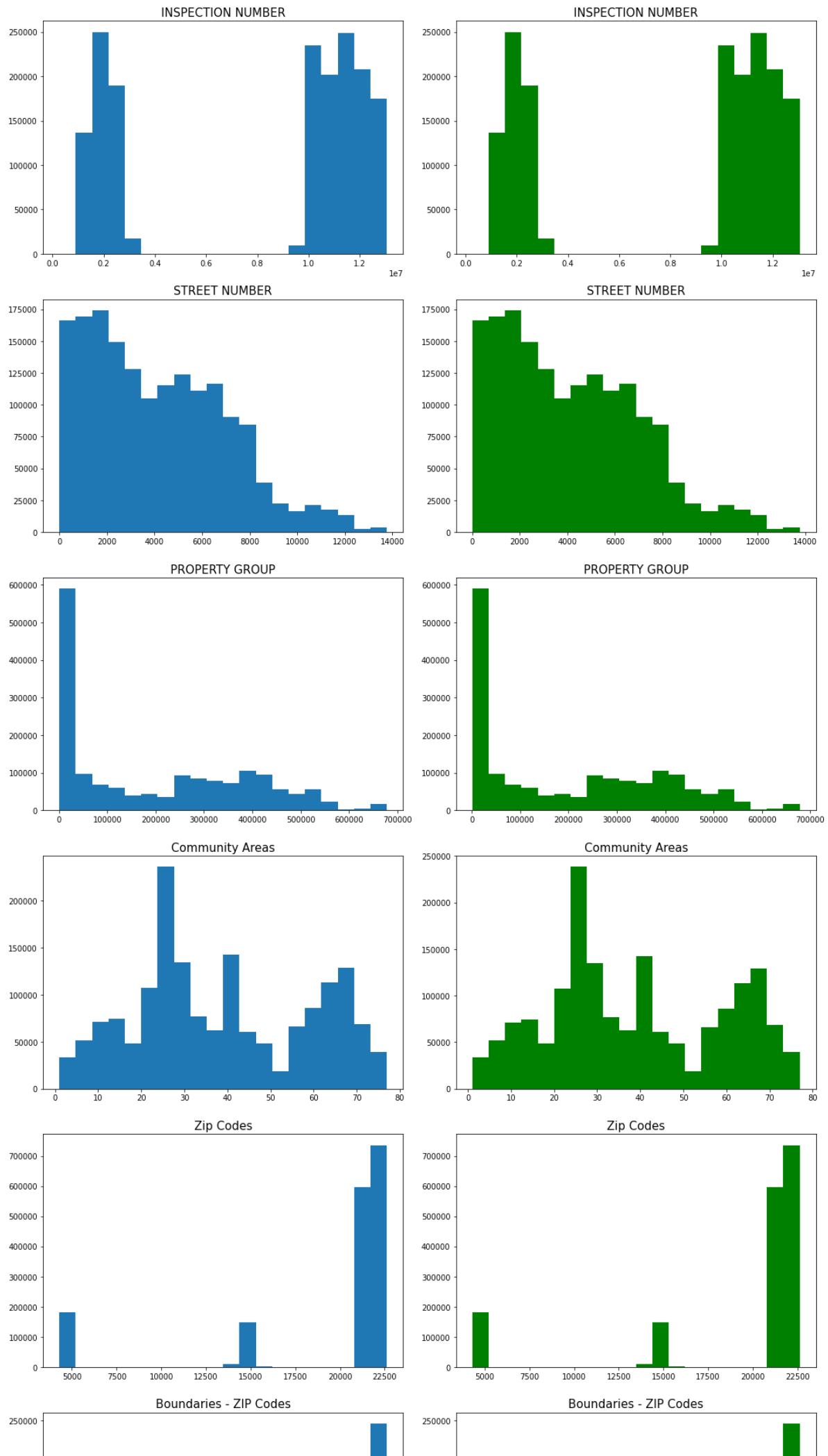


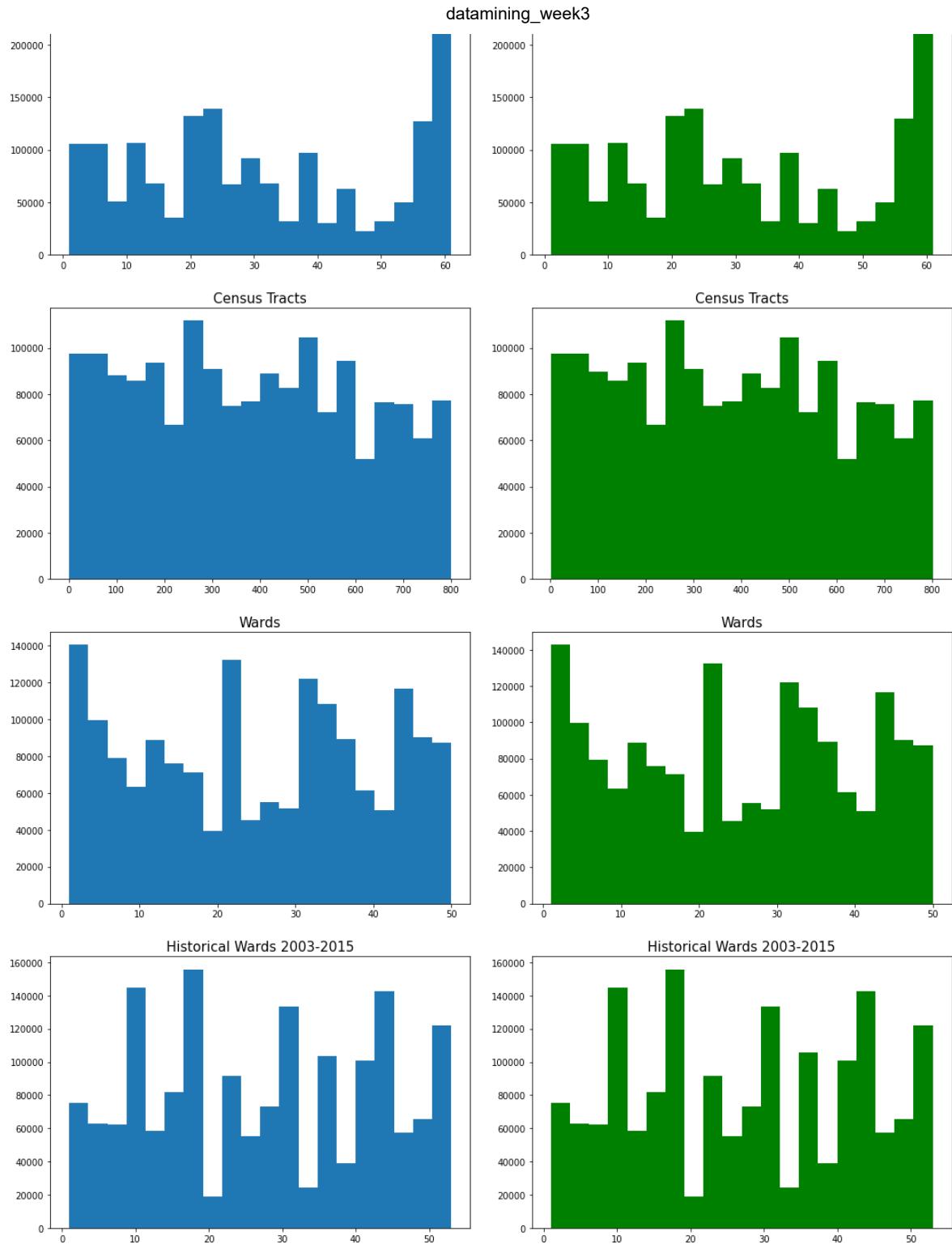
new_STREET TYPE











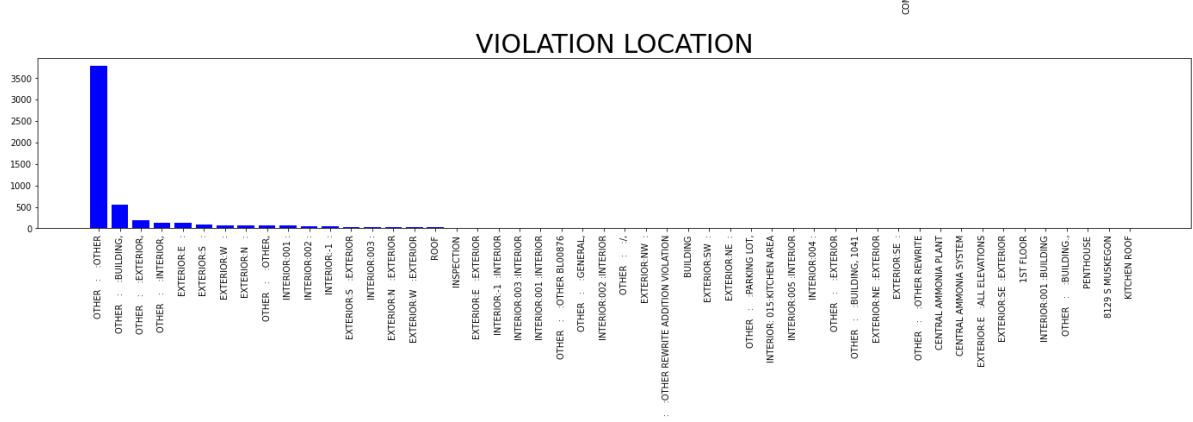
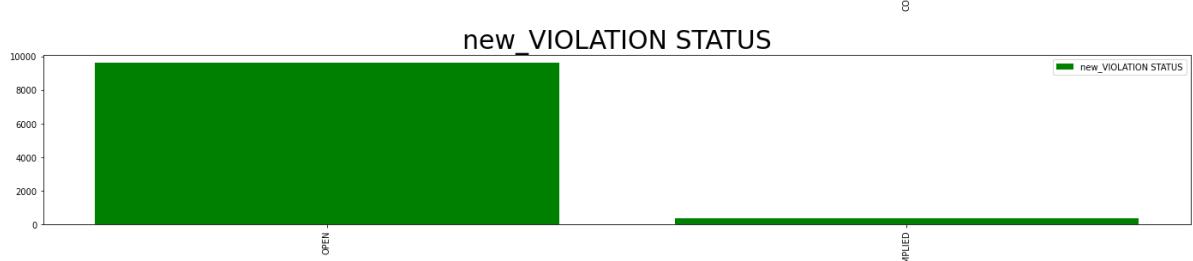
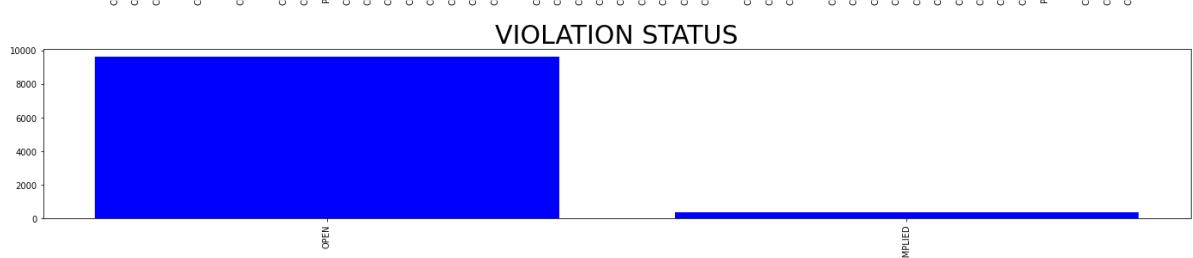
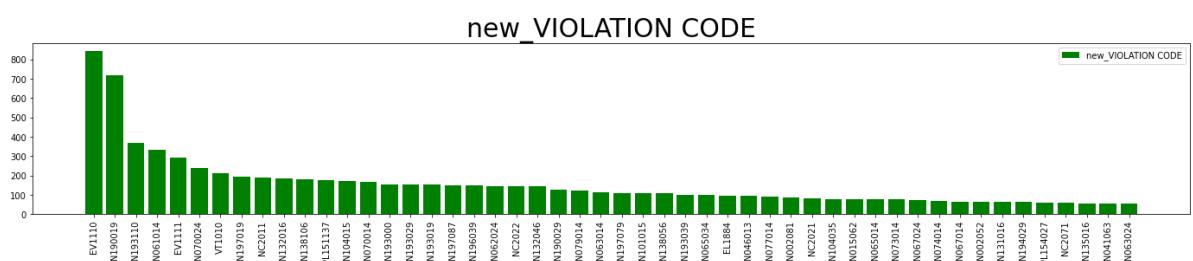
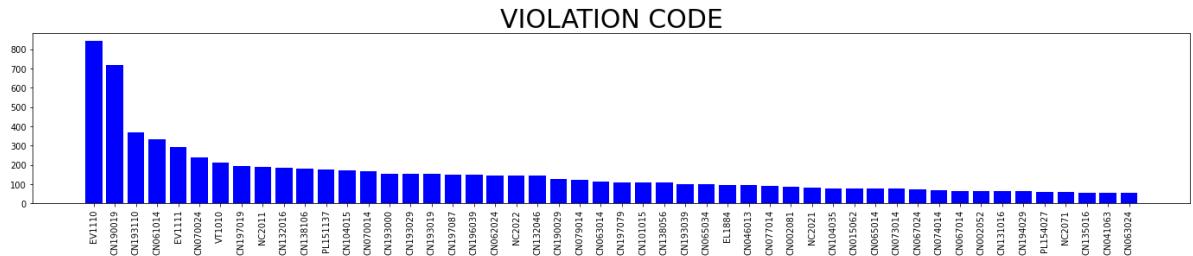
[分析]标称属性VIOLATION LOCATION的频数减少过多，影响预测后的数据分布，故此方式不合理

5.3通过属性的相关关系来填补缺失值（由于数据集过大，个人计算机难以处理，故自此步起仅选取前一万行进行处理）

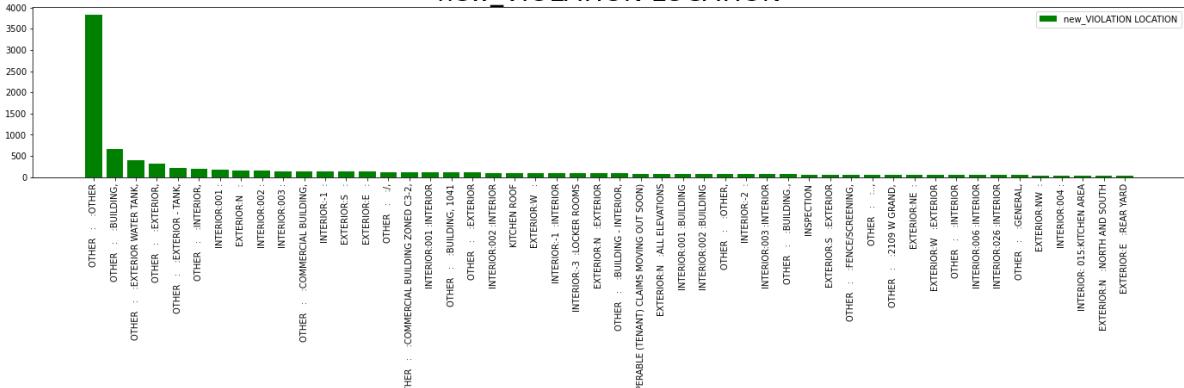
```
In [42]: now_df_CBV = df_CBV[:10000]
miss_index = list()
complete_index = list()
col = now_df_CBV.columns.values
i = 0
for index in list(now_df_CBV.isnull().any()): #index==True时，对应的列中具有缺失值
    if(index):
        miss_index.append(col[i])
    else:
```

```
    complete_index.append(col[i])
i = i + 1

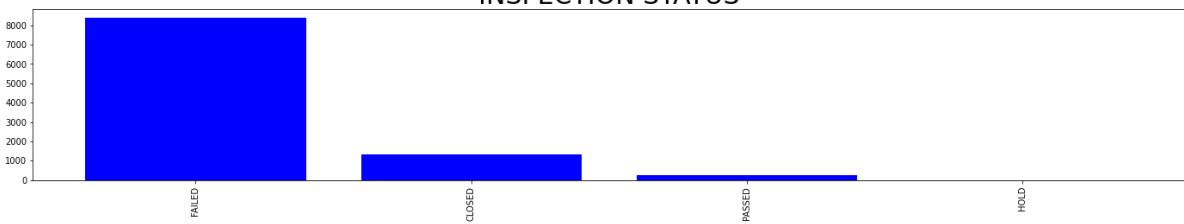
new_df_CBV = attr_corr_fill(now_df_CBV, miss_index, complete_index)
nominal_summary(df=now_df_CBV, nominal_index=nominal_index_CBV, new_df=new_df_CBV)
box_plot(now_df_CBV, numerical_index_CBV, new_df_CBV)
hist_plot(now_df_CBV, numerical_index_CBV, new_df_CBV)
```



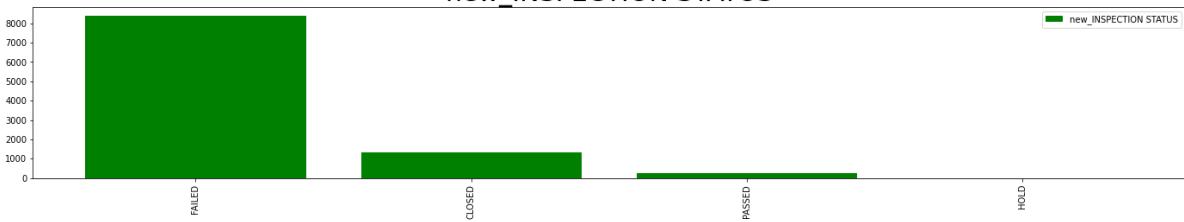
new_VIOLATION LOCATION



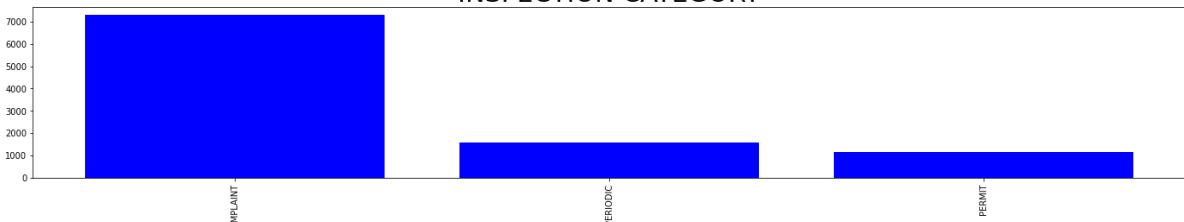
INSPECTION STATUS



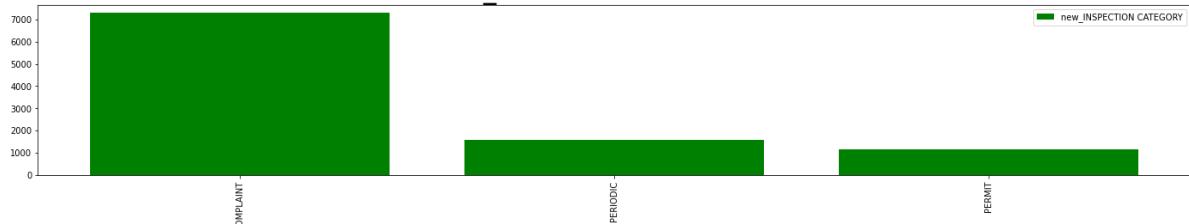
new_INSPECTION STATUS



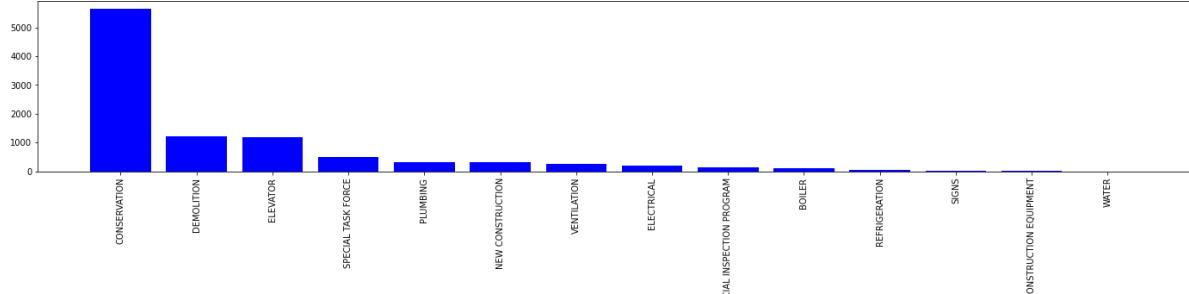
INSPECTION CATEGORY



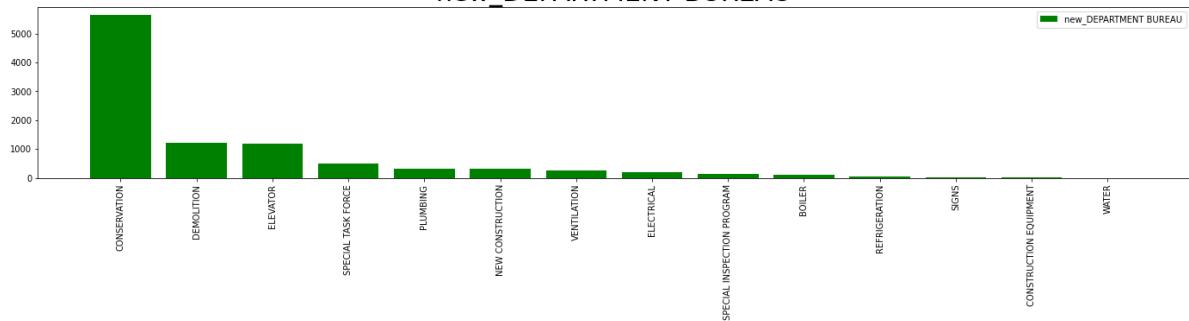
new_INSPECTION_CATEGORY



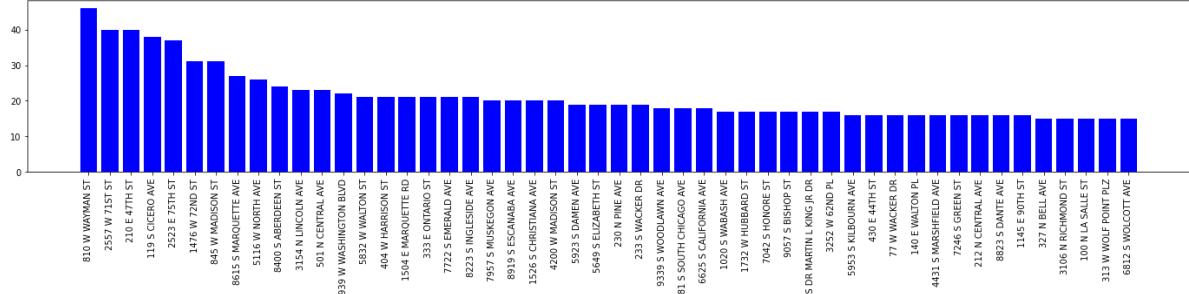
DEPARTMENT BUREAU



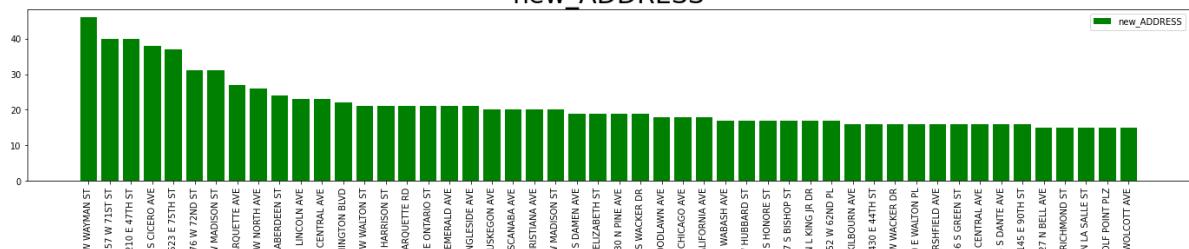
new_DEPARTMENT_BUREAU



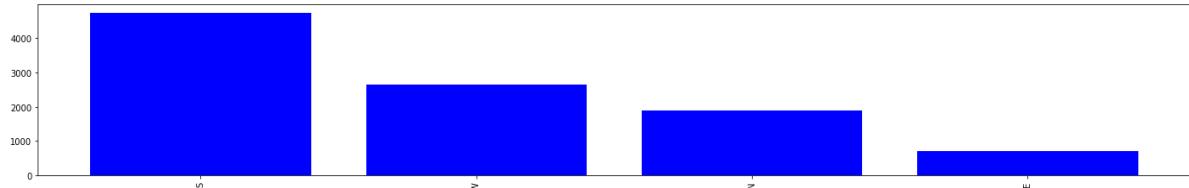
ADDRESS



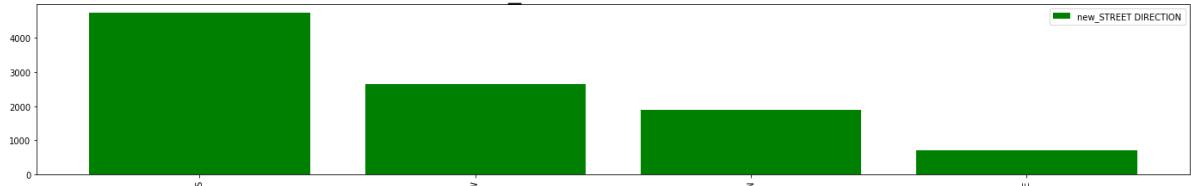
new_ADDRESS



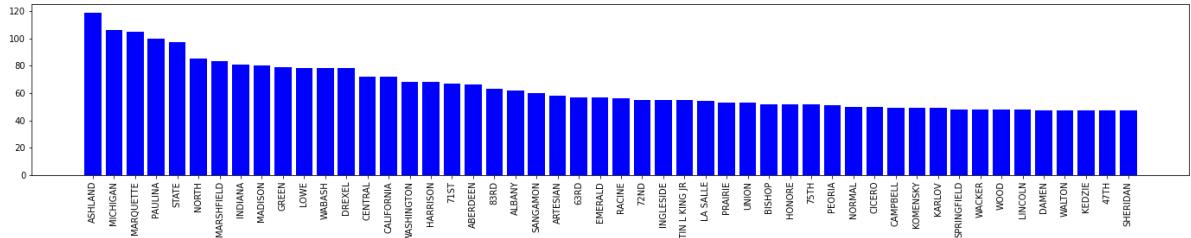
STREET DIRECTION



new_STREET DIRECTION



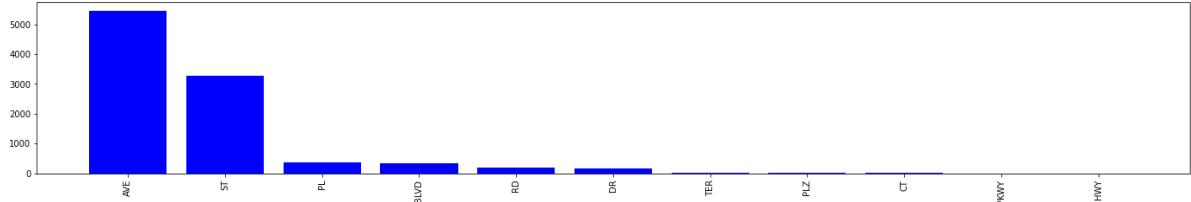
STREET NAME



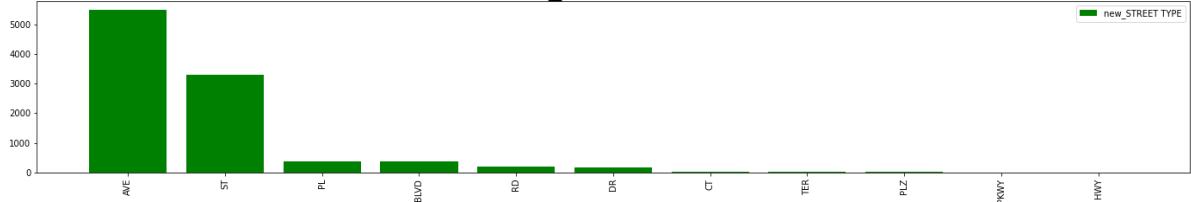
new_STREET NAME

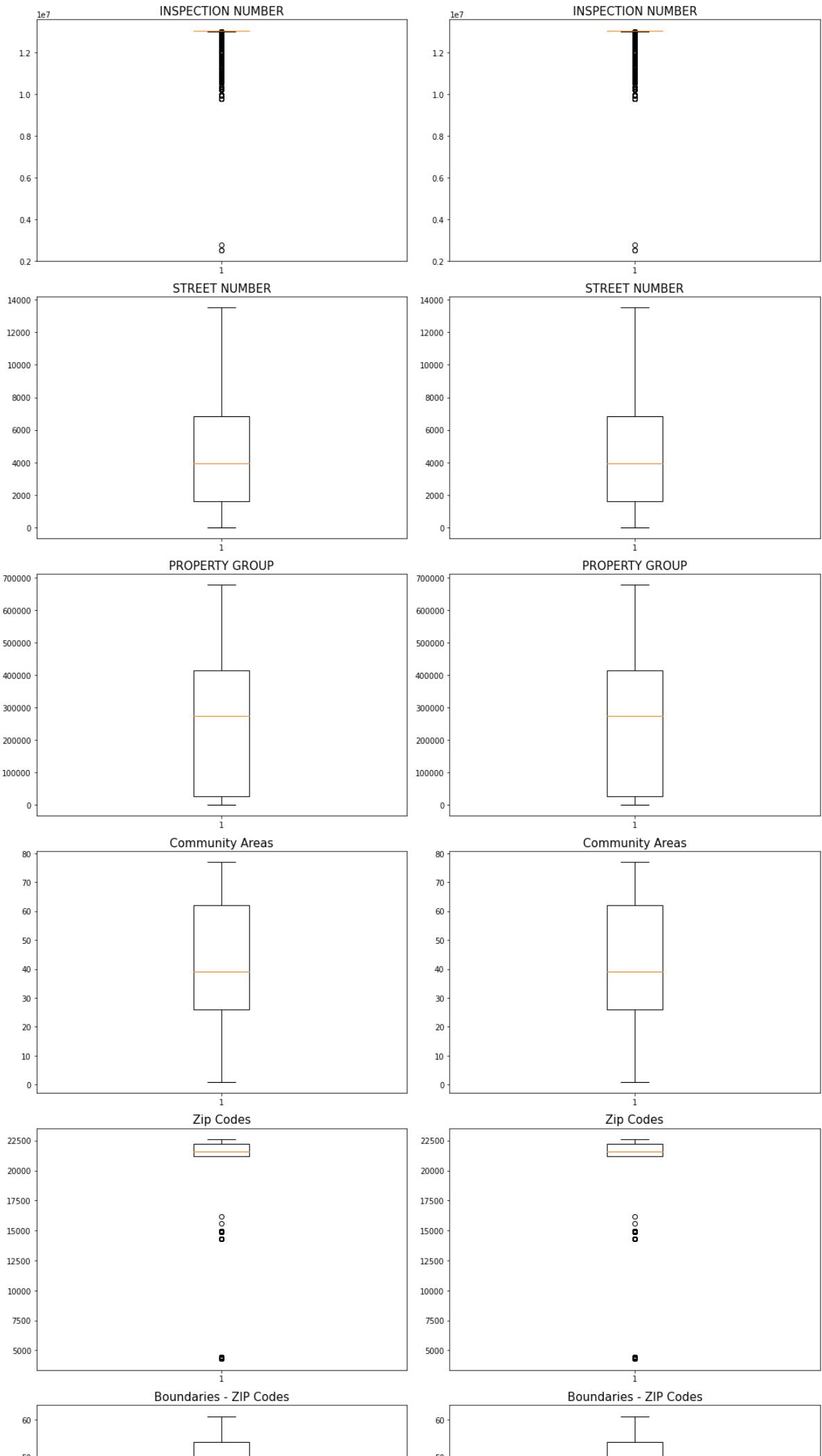


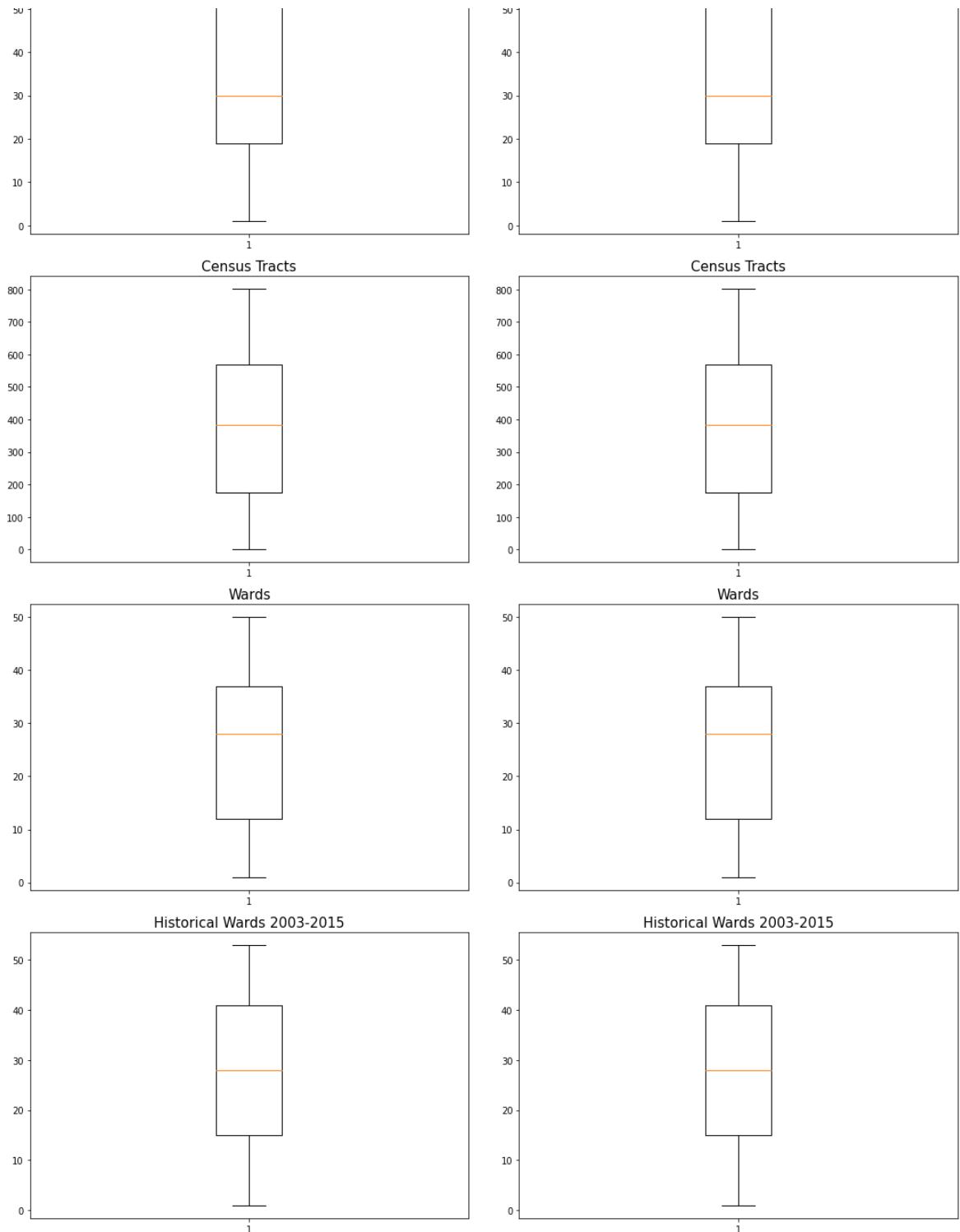
STREET TYPE

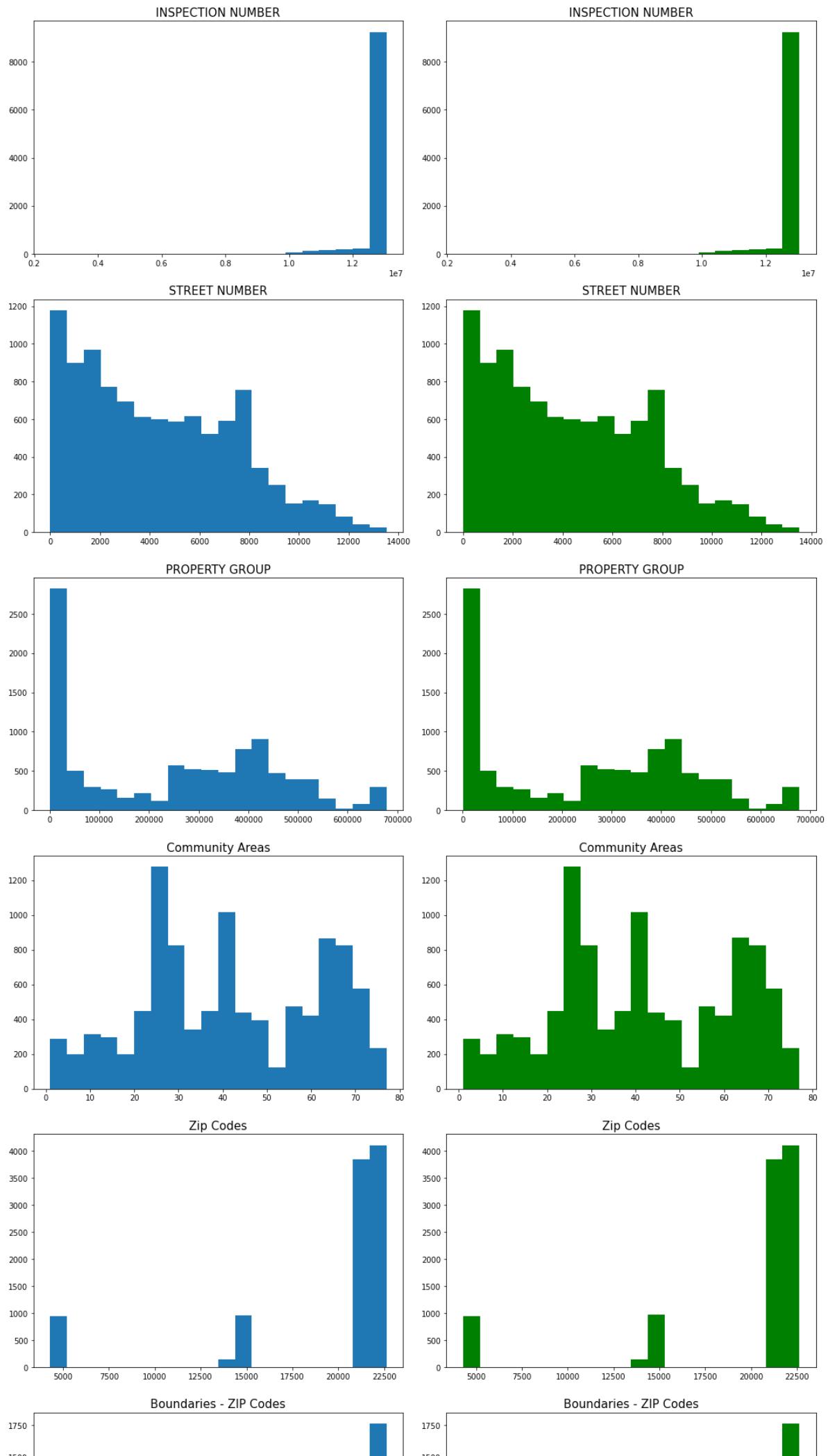


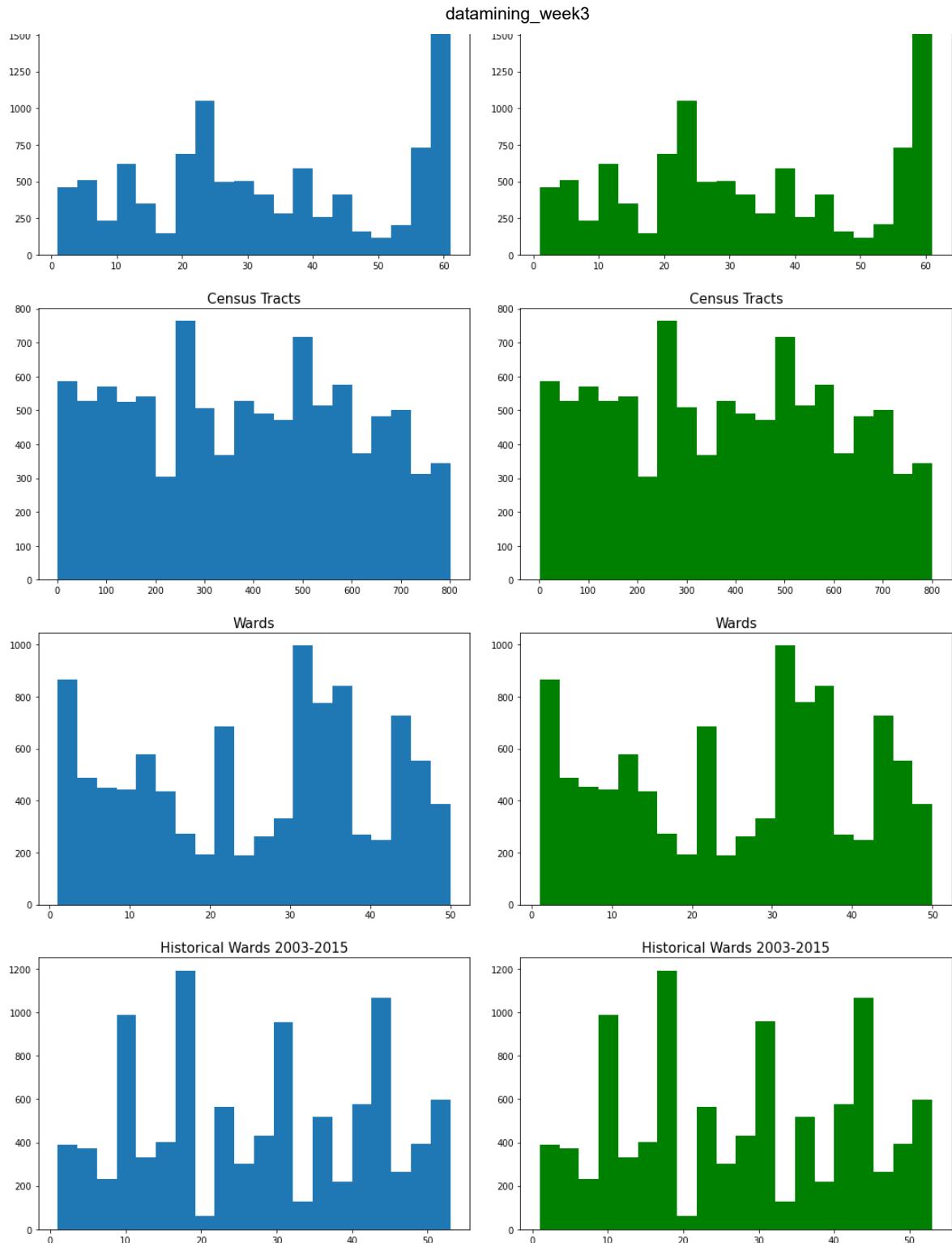
new_STREET TYPE







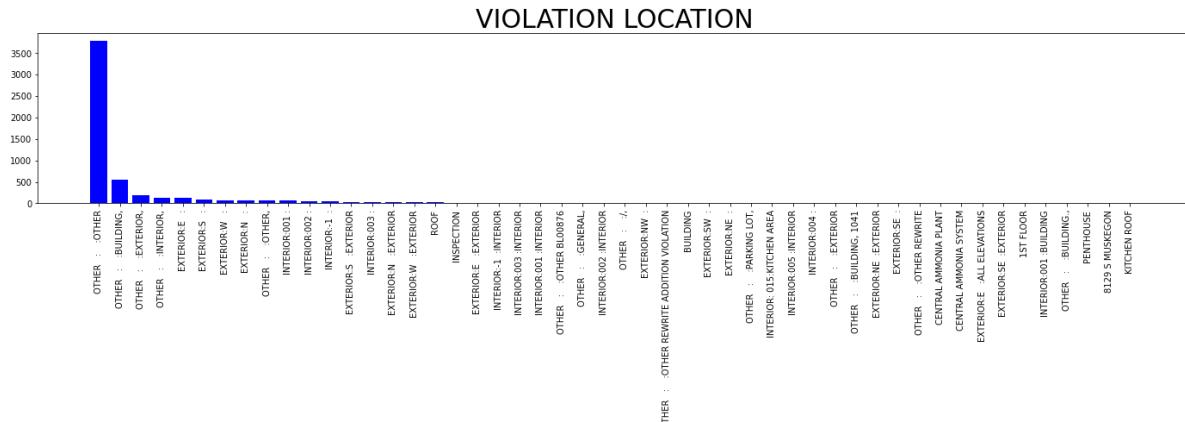
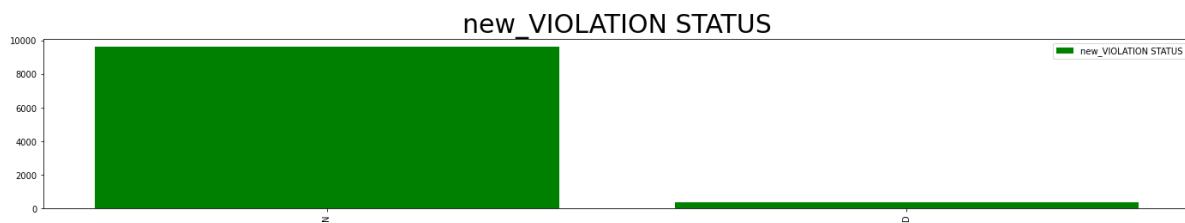
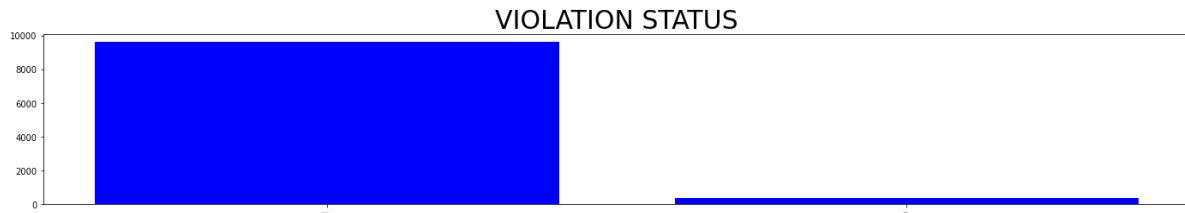
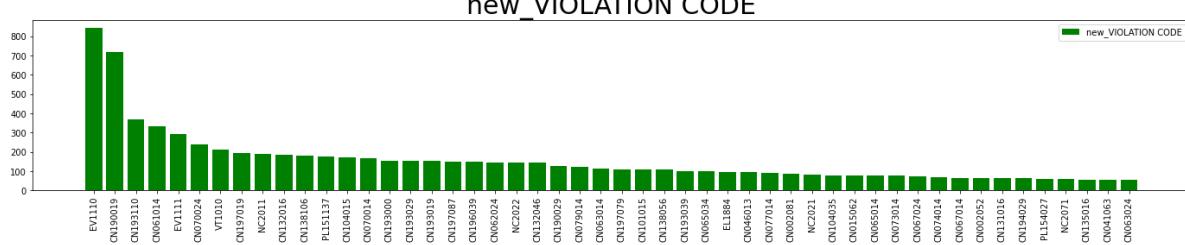
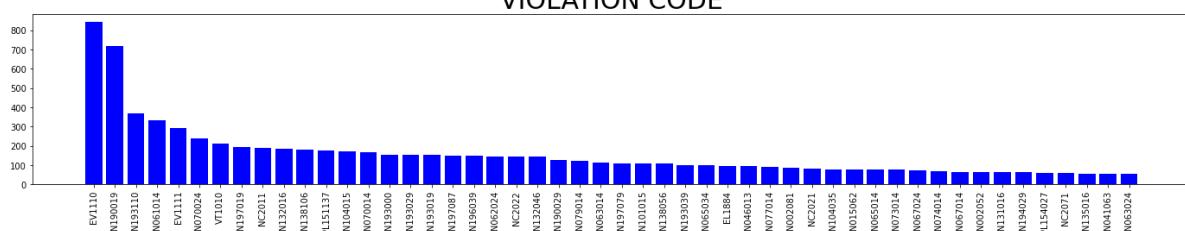




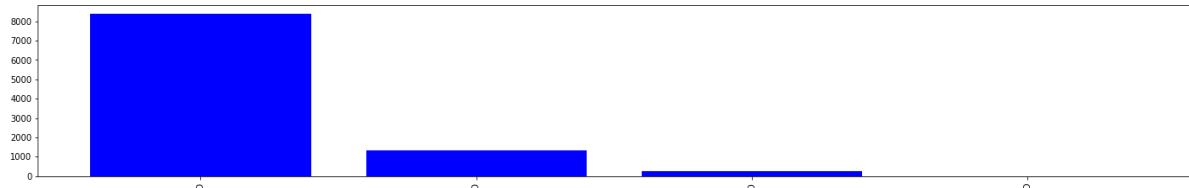
[分析]预测后的数据分布与原始数据差距不大，由此可知本方法预测合理

5.4 通过数据对象之间的相似性来填补缺失值

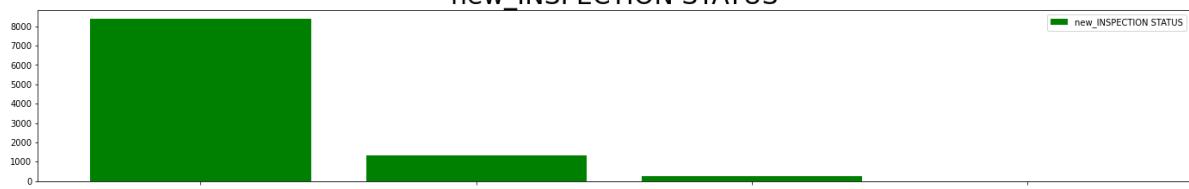
```
In [43]: new_df_CBV = obj_corr_fill(df=now_df_CBV, numerical_index=numerical_index_CBV, k=len(r)
nominal_summary(now_df_CBV, nominal_index_CBV, new_df_CBV)
box_plot(now_df_CBV, numerical_index_CBV, new_df_CBV)
hist_plot(now_df_CBV, numerical_index_CBV, new_df_CBV)
```



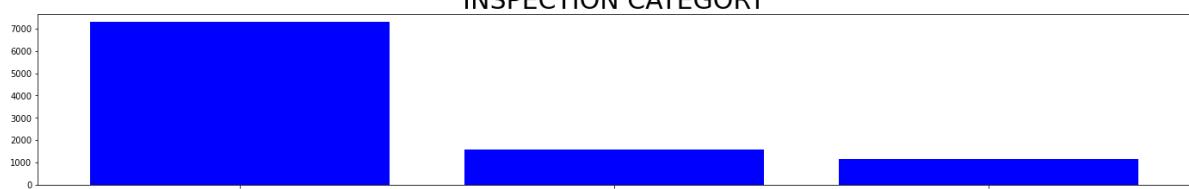
INSPECTION STATUS



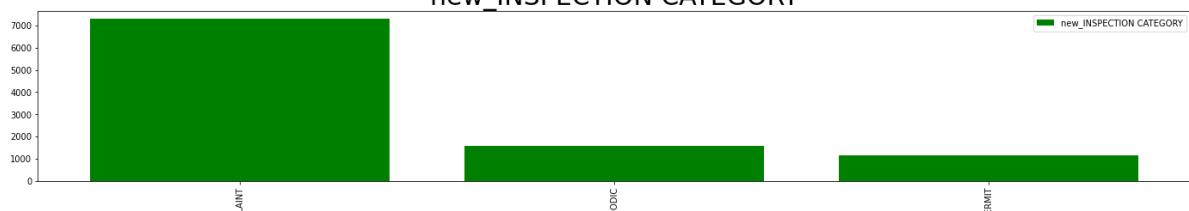
new_INSPECTION STATUS



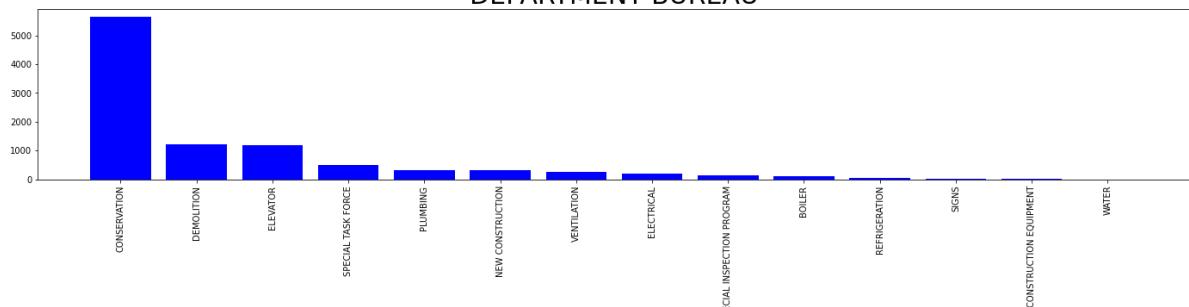
INSPECTION CATEGORY



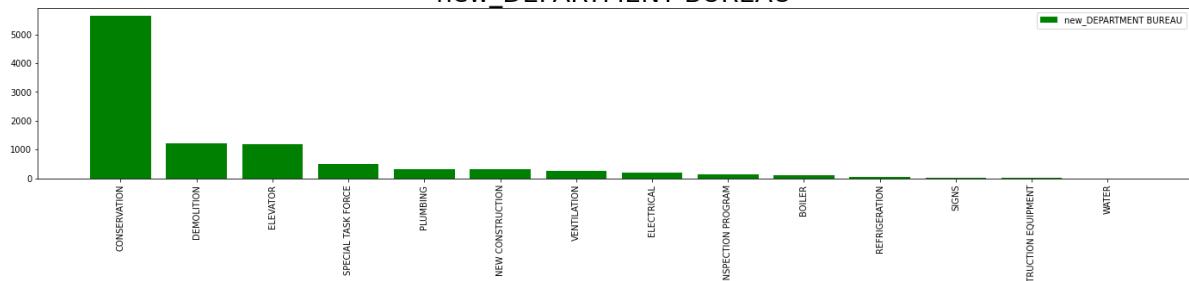
new_INSPECTION CATEGORY



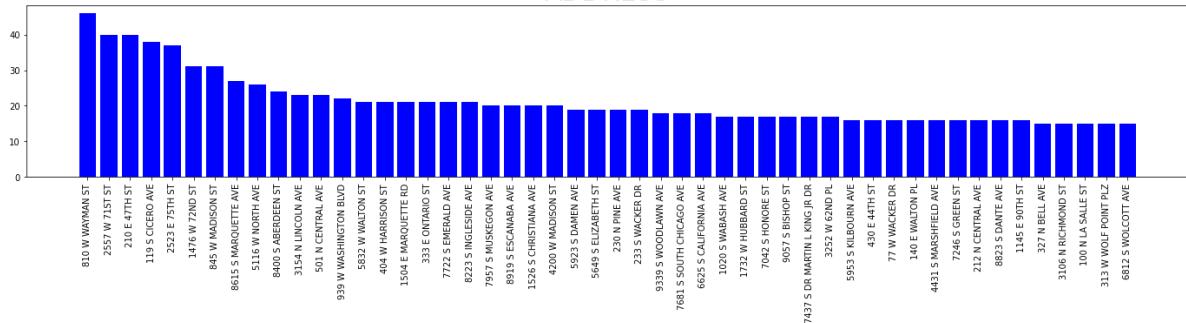
DEPARTMENT BUREAU



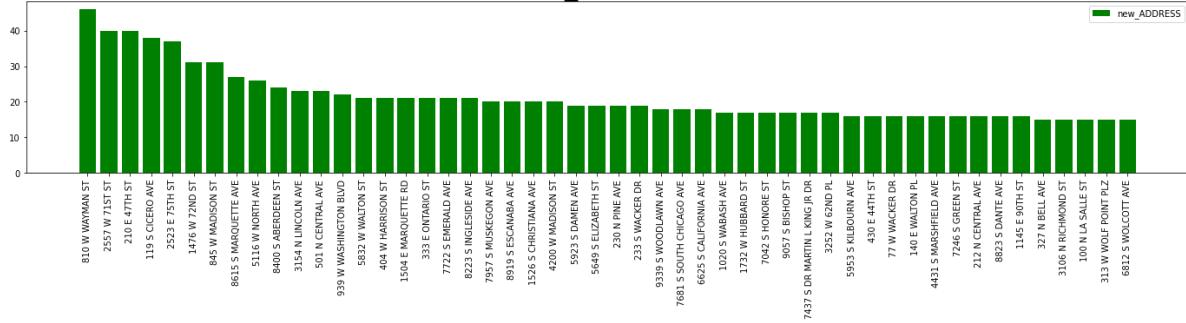
new_DEPARTMENT BUREAU



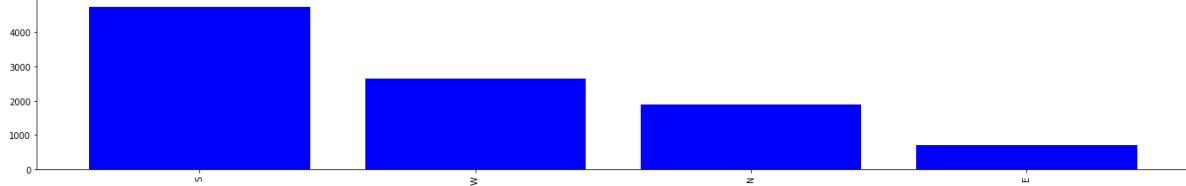
ADDRESS



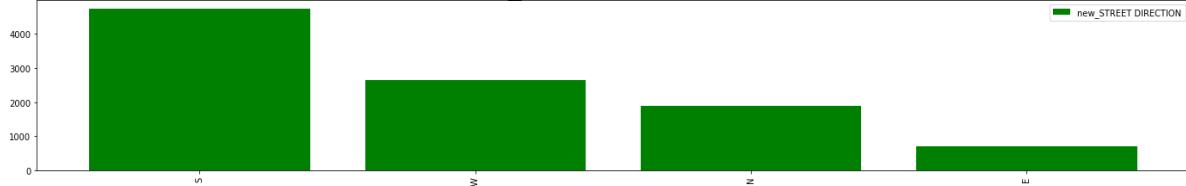
new_ADDRESS



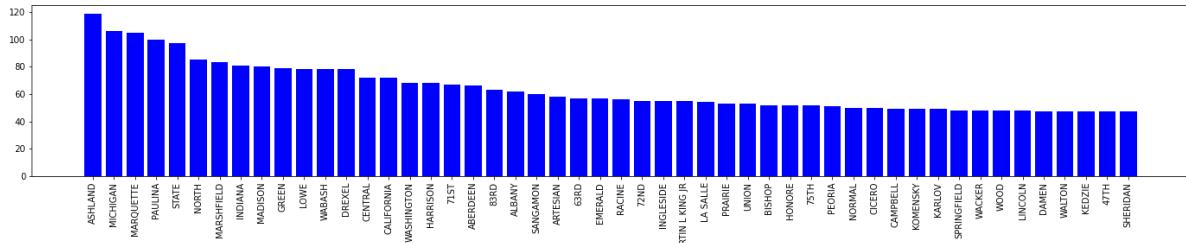
STREET DIRECTION



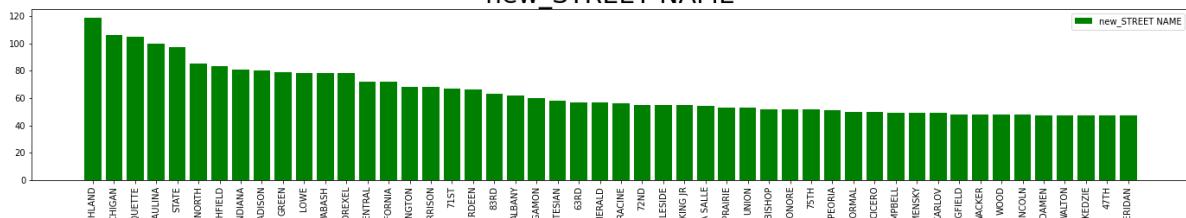
new_STREET DIRECTION



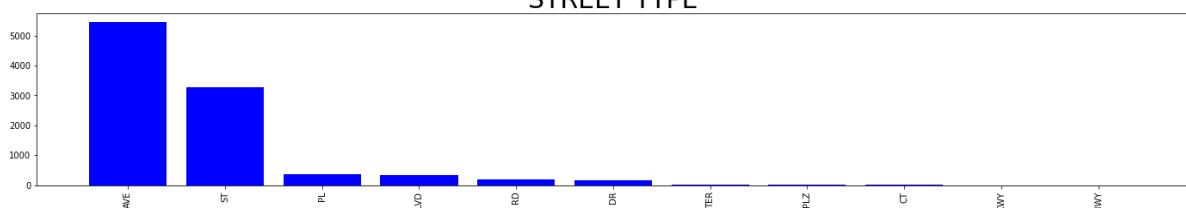
STREET NAME

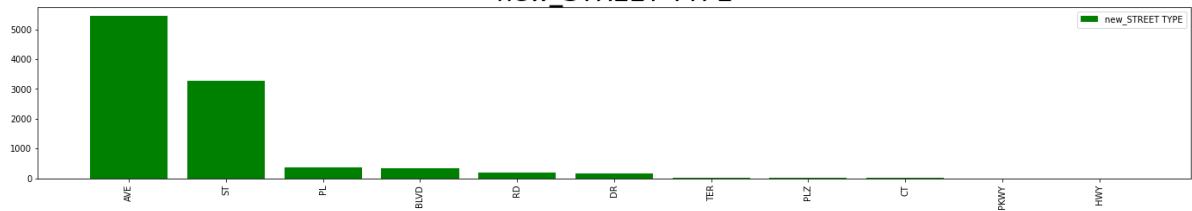


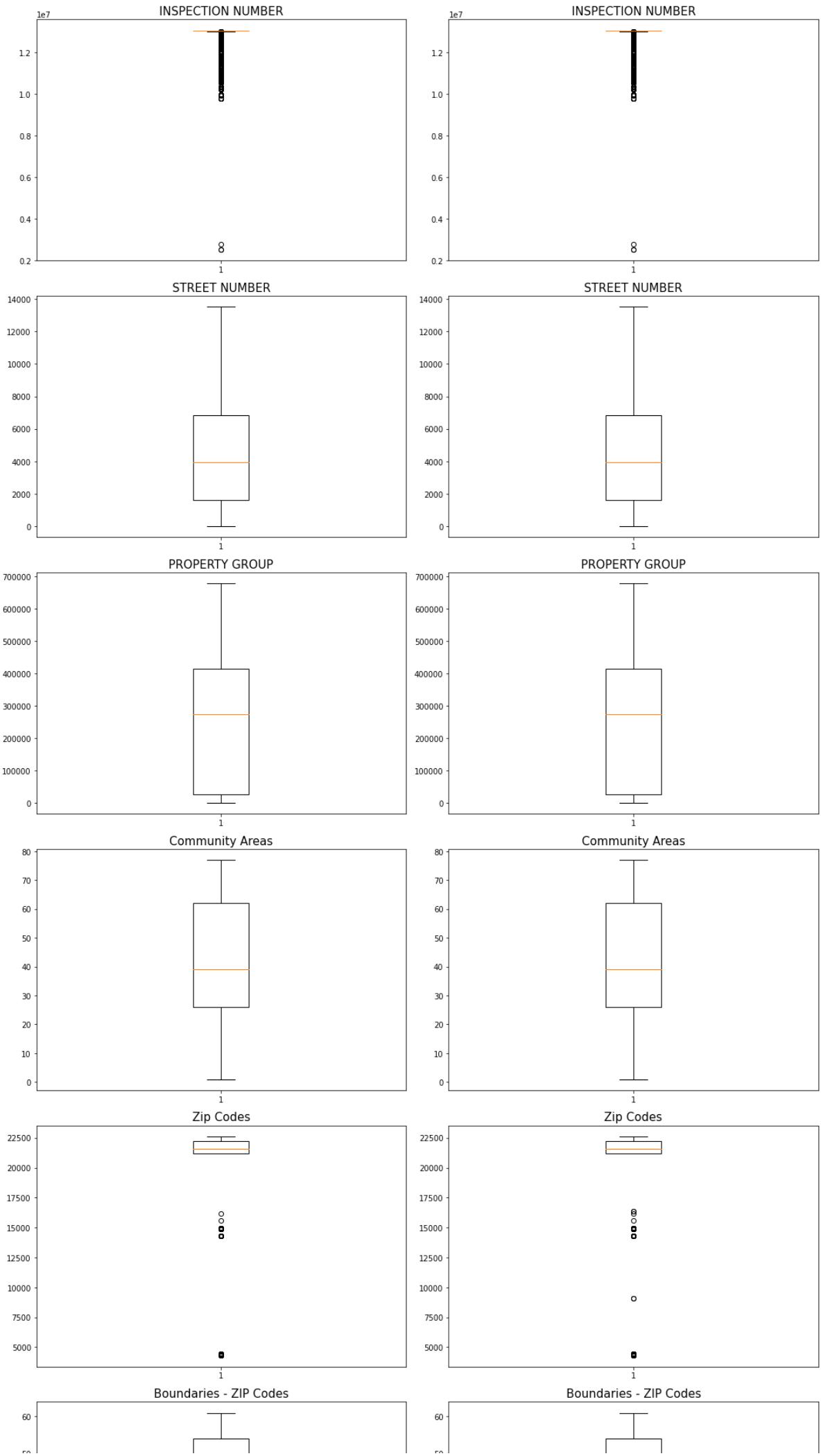
new_STREET NAME

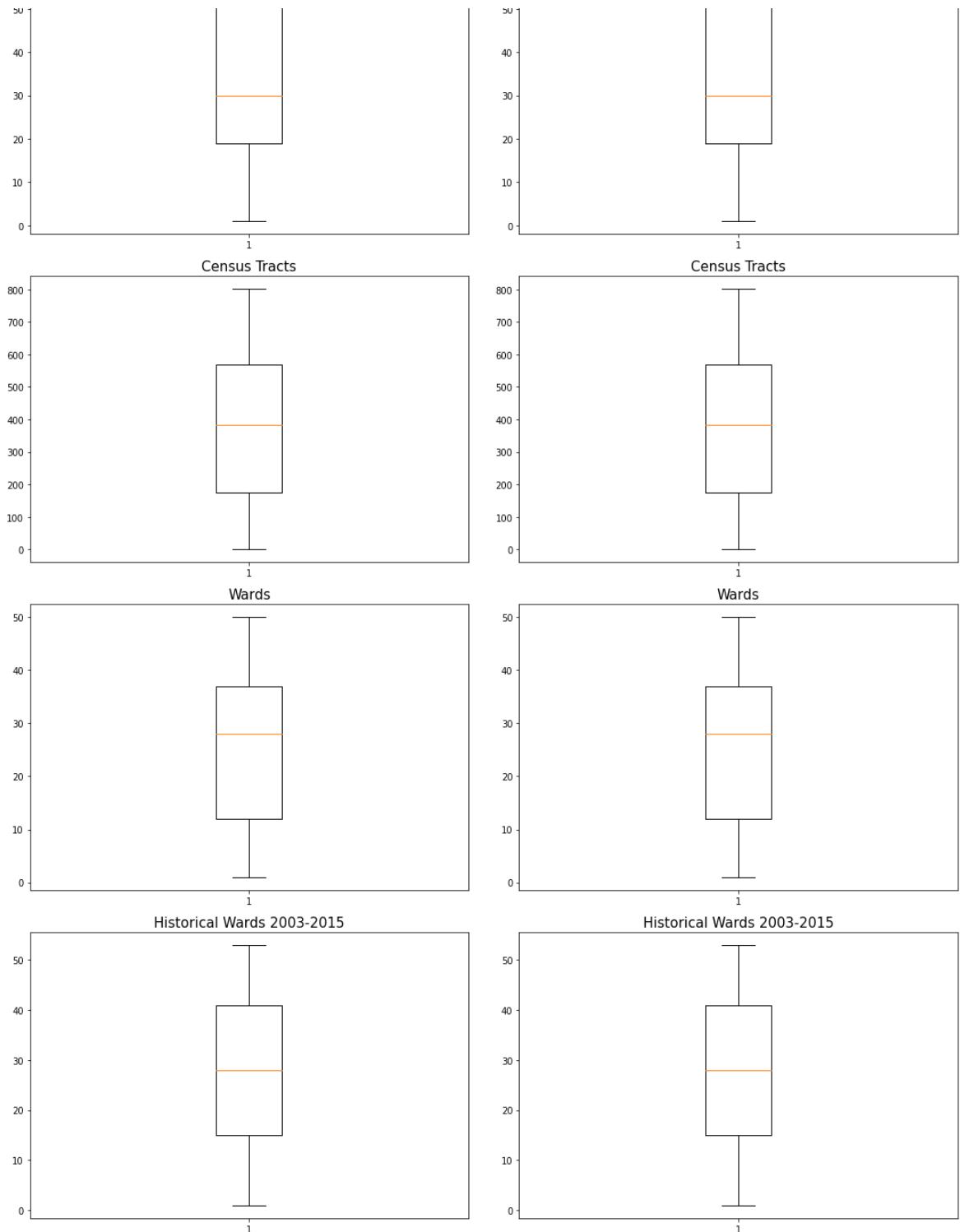


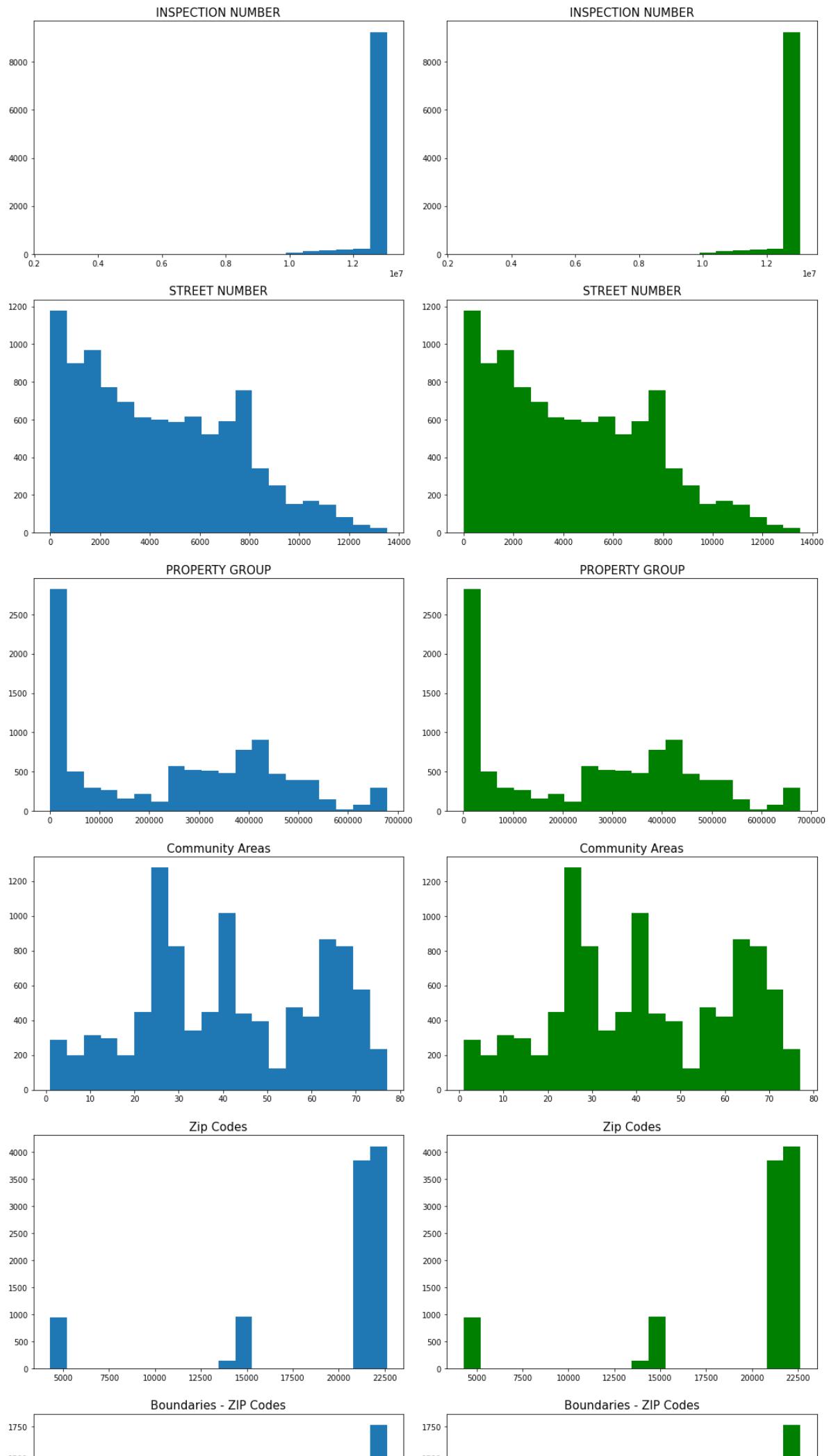
STREET TYPE

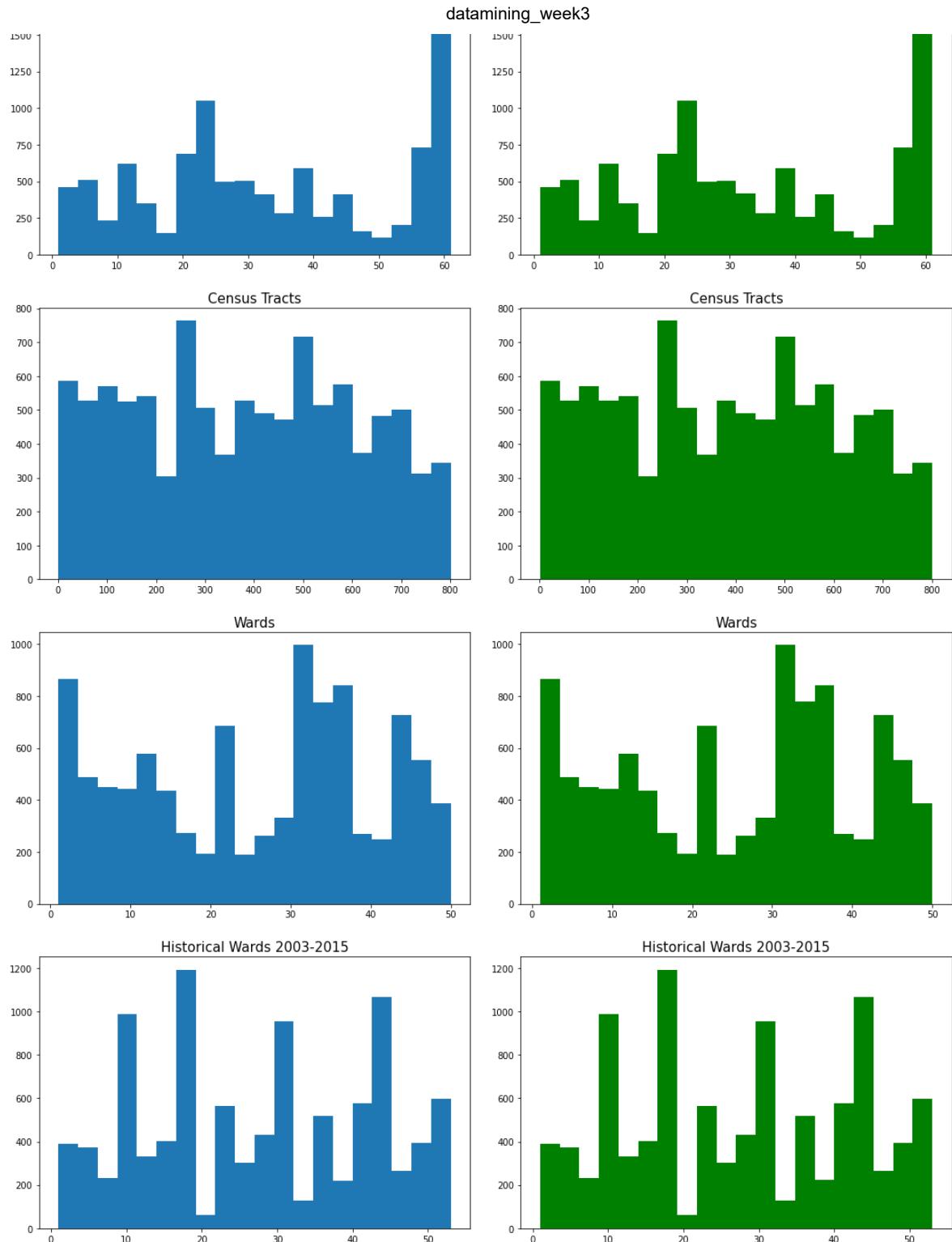


new_STREET TYPE









[分析]预测后的数据分布与原始数据差距不大，由此可知本方法预测合理

数据集2：Wine Reviews

1. 进行初步处理

```
In [54]: #导入数据
df_wine1 = pd.read_csv('./Wine Reviews/winemag-data_first150k.csv', index_col=0)
df_wine2 = pd.read_csv('./Wine Reviews/winemag-data-130k-v2.csv', index_col=0)
#删除重复值
delDupli(df_wine1)
delDupli(df_wine2)
```

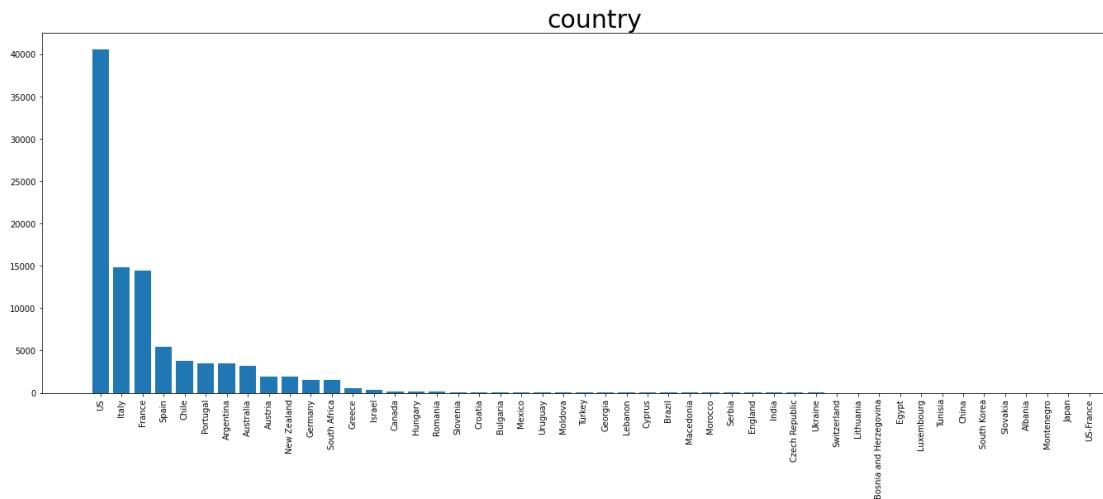
53079
0
9983
0

2.统计标称型数据的频数

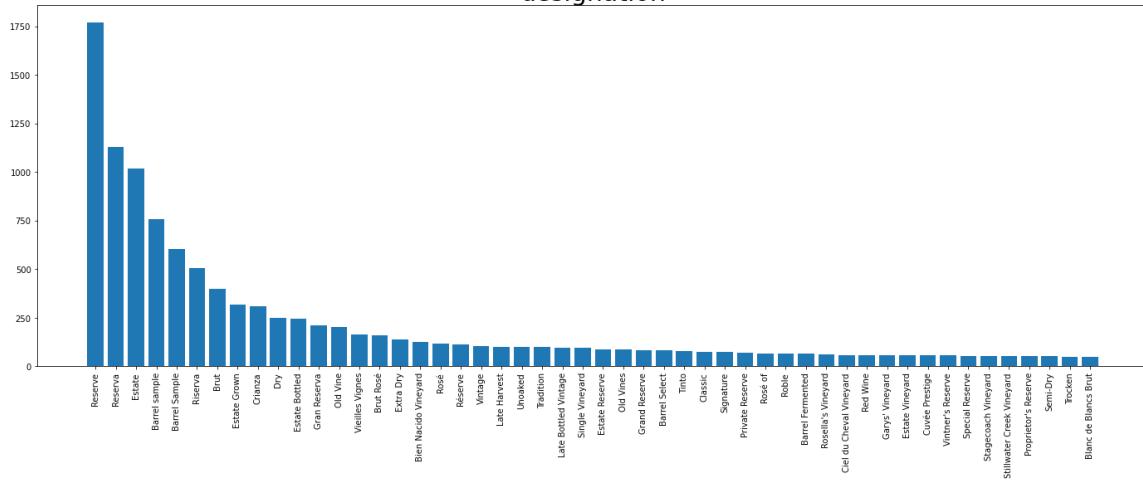
2.1针对数据集winemag-data_first150k.csv

In [55]:

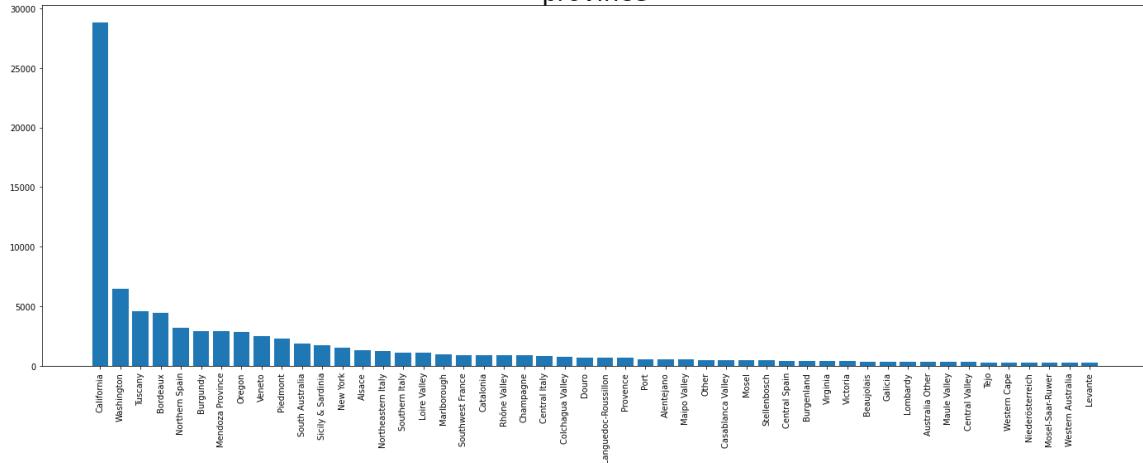
```
nominal_index_wine1 = list(df_wine1.select_dtypes(['object']))  
nominal_summary(df_wine1,nominal_index_wine1)
```



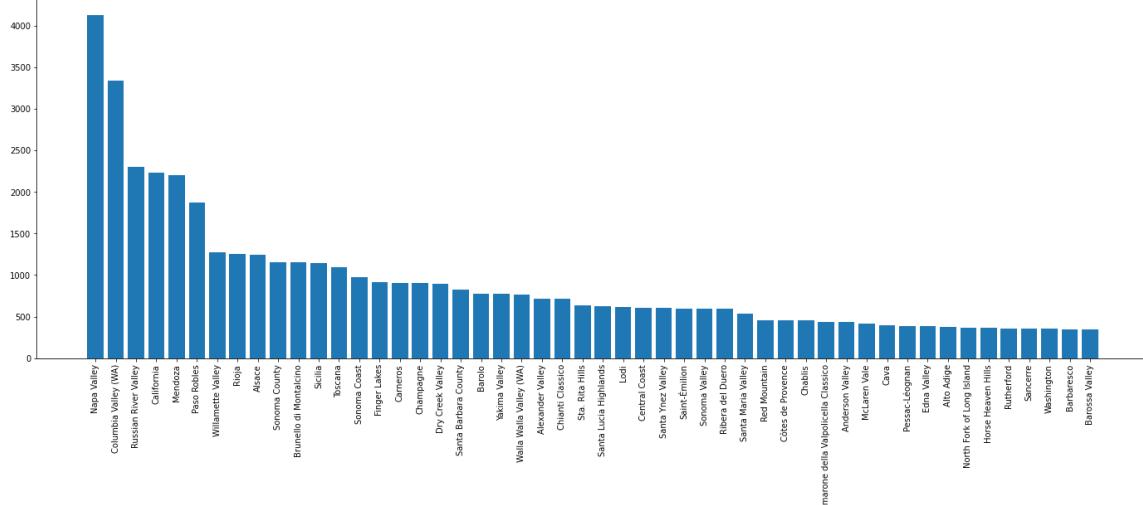
designation



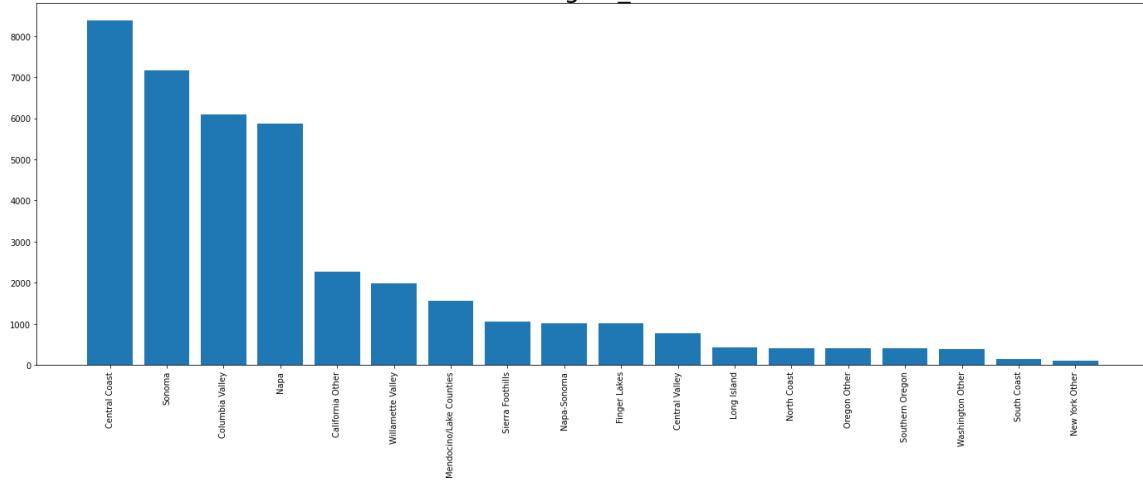
province



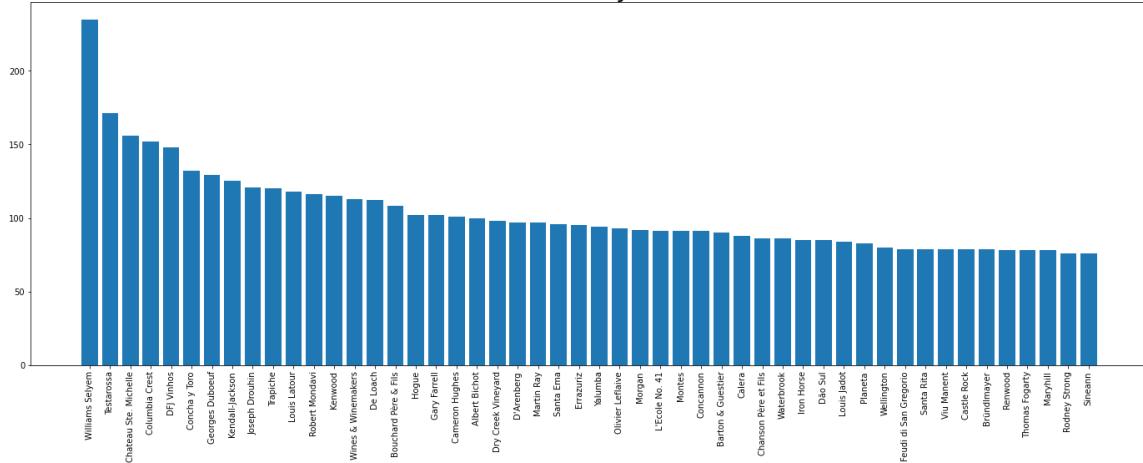
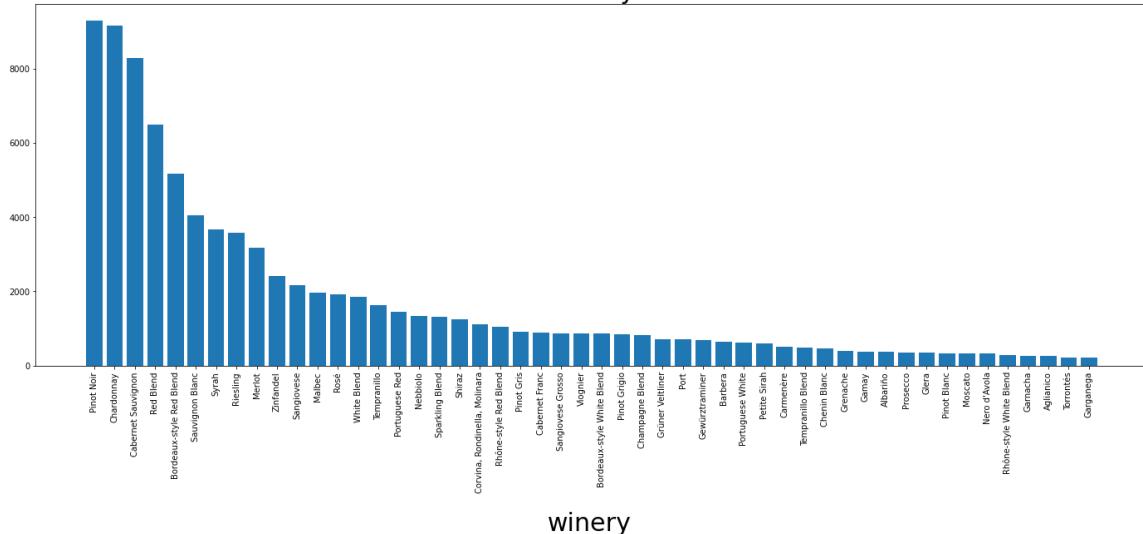
region_1



region_2

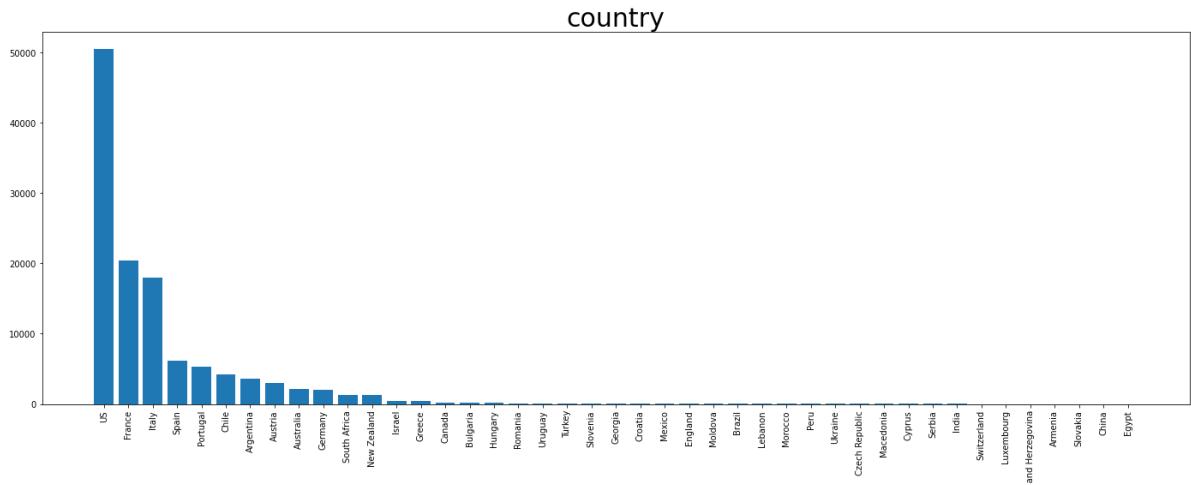


variety

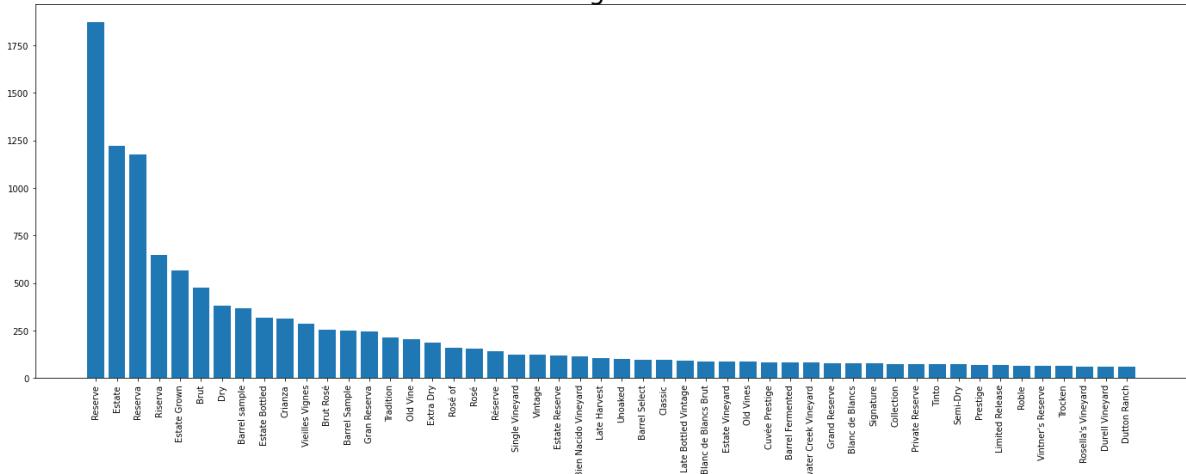


2.2 针对数据集winemag-data-130k-v2.csv

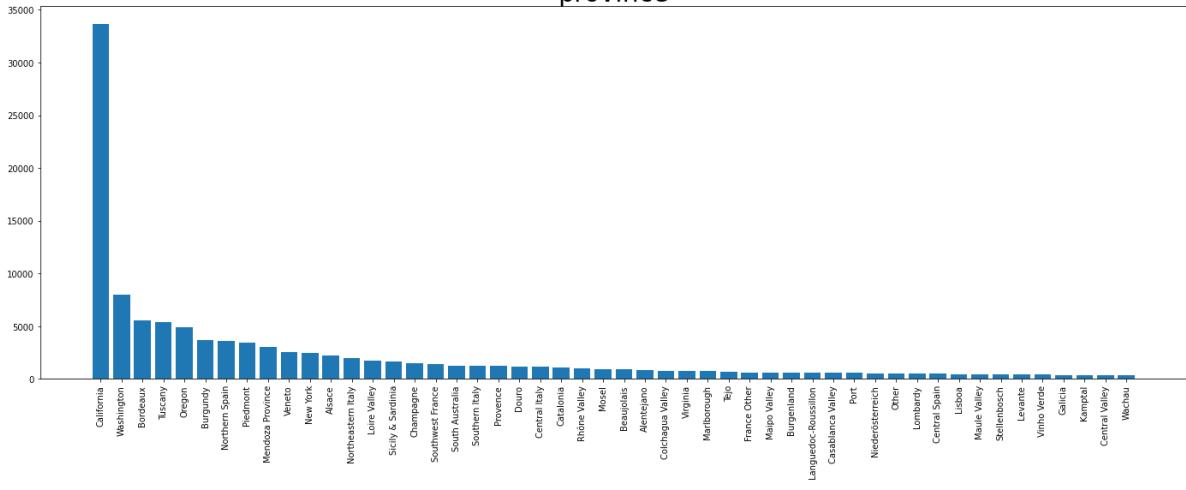
```
In [56]: nominal_index_wine2 = list(df_wine2.select_dtypes(['object']))  
nominal_summary(df_wine2,nominal_index_wine2)
```



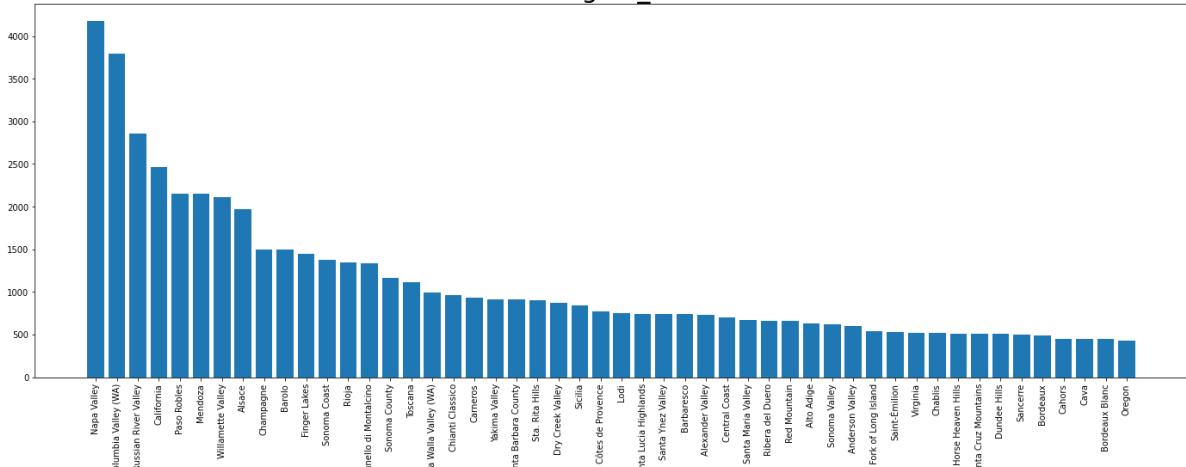
designation



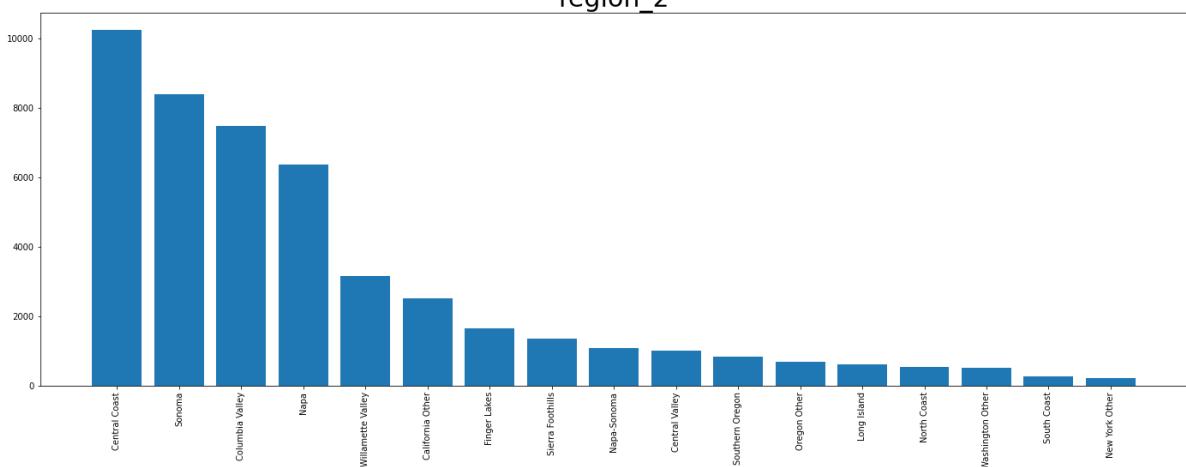
province



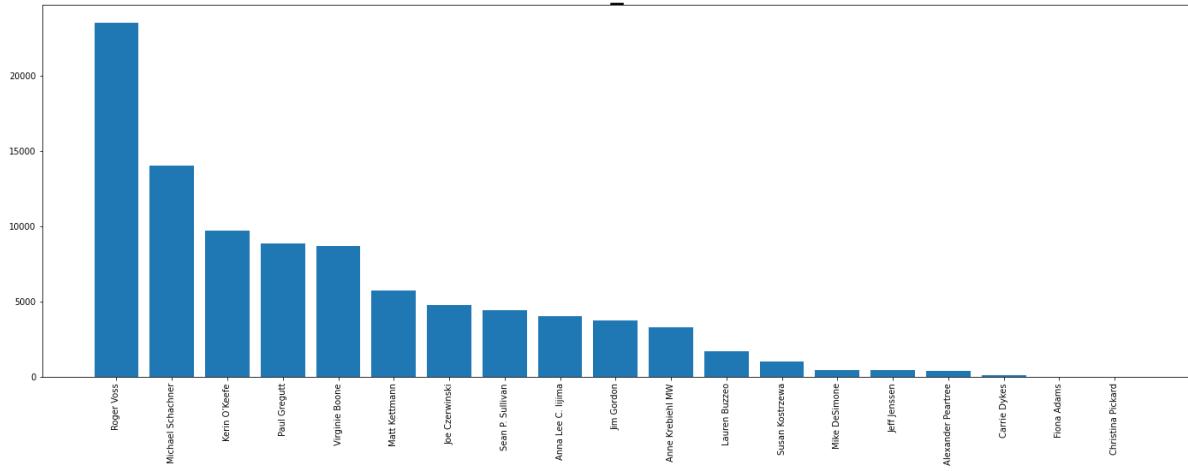
region_1



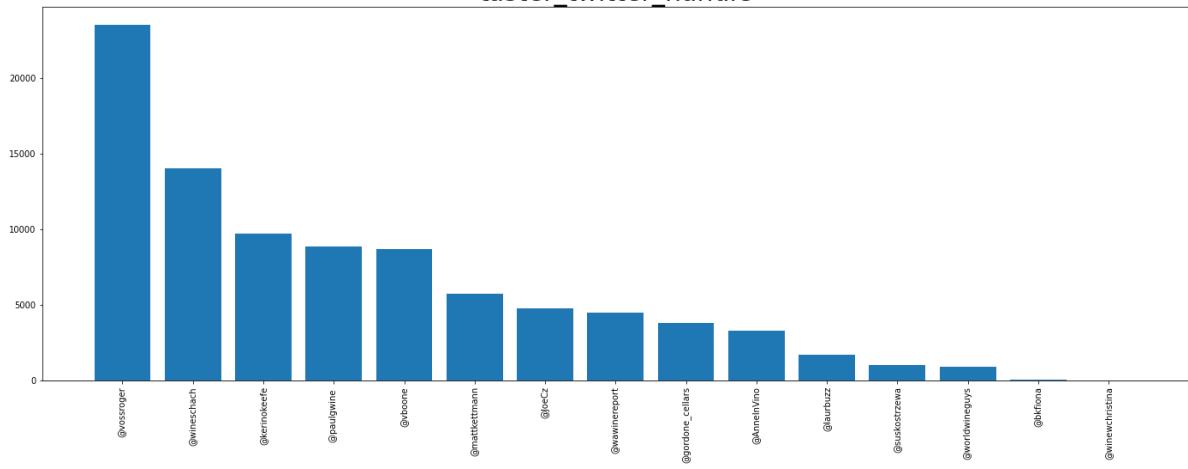
region_2



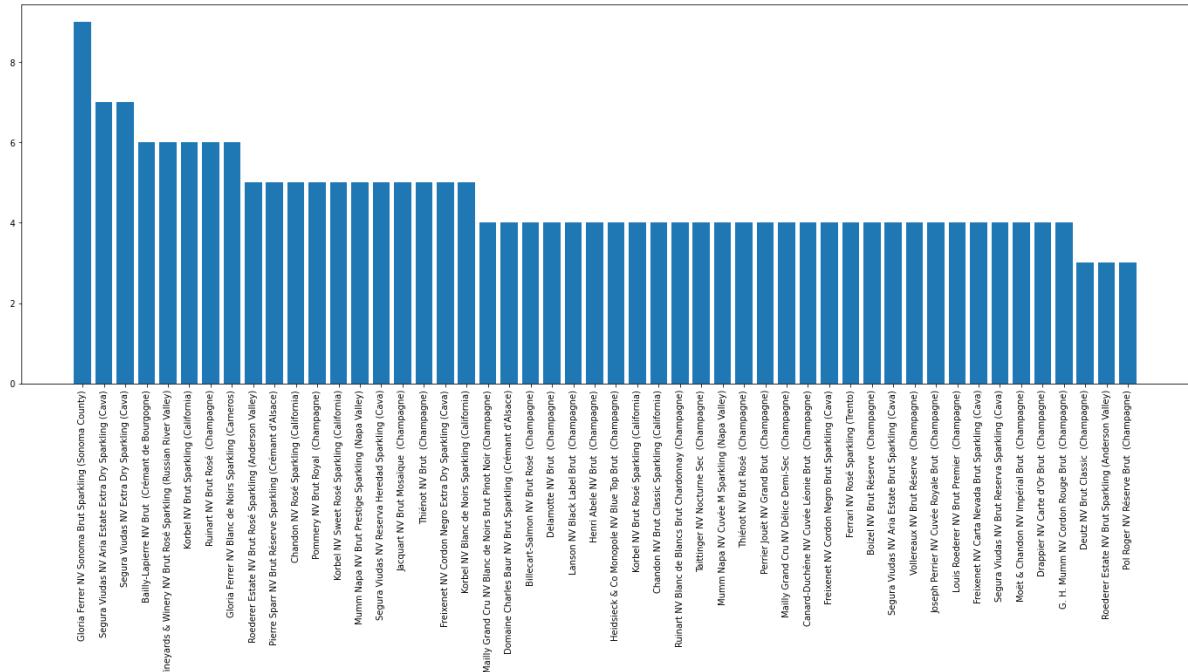
taster_name

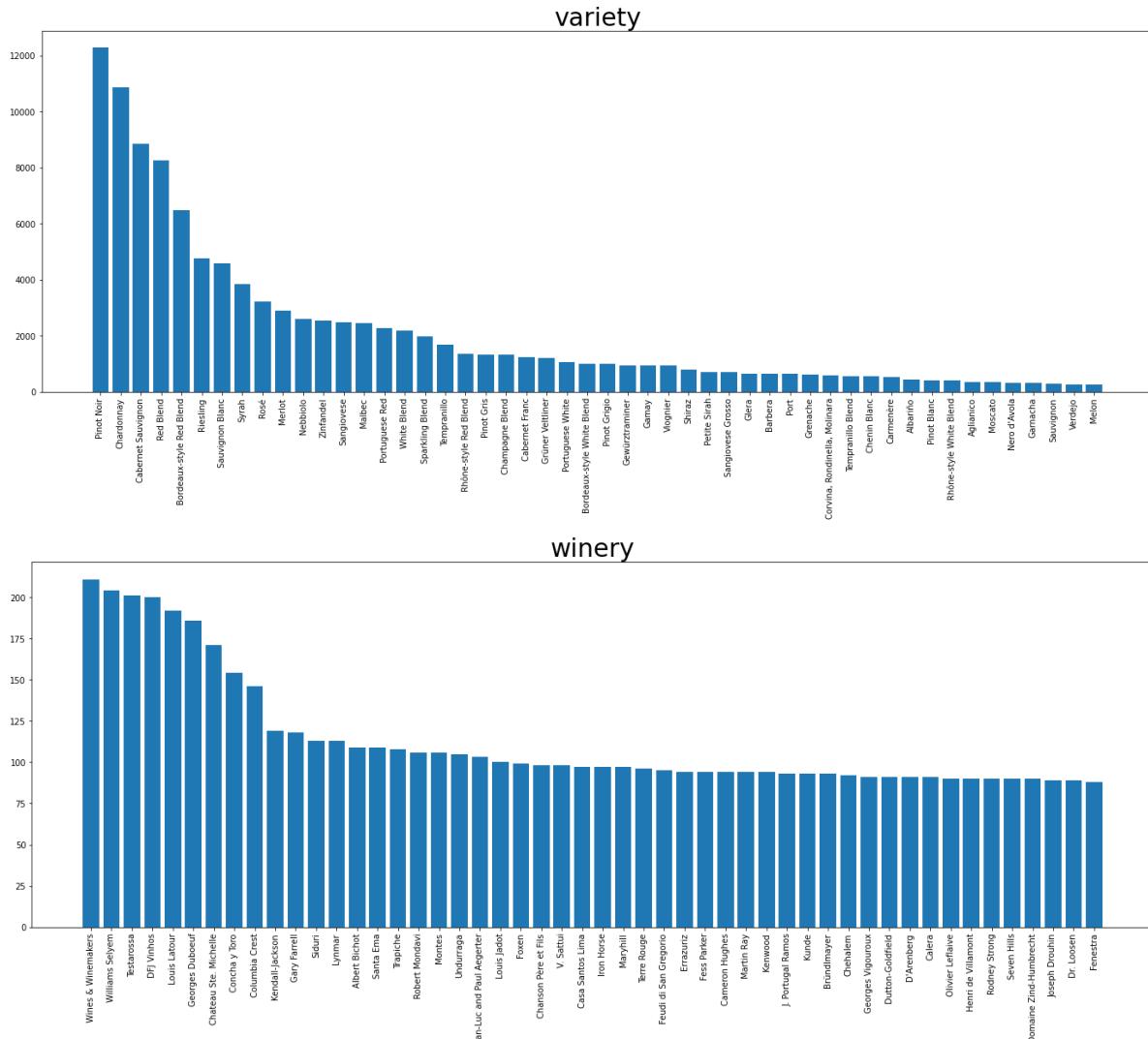


taster_twitter_handle



title





3.数值型数据的五数概括及缺失值数据

3.1针对数据集winemag-data_first150k.csv

In [57]:

```
#获得数据集的数值型数据
numerical_index_wine1 = list(df_wine1.select_dtypes(['int', 'float']))
numerical_summary(df_wine1, numerical_index_wine1)
```

```
descriptive statistics (points):
Min: 80.0000    Q1(25%): 86.0000          Q2(50%): 88.0000          Q3(75%): 90.0000
Max: 100.0000
Missing: 0

descriptive statistics (price):
Min: 4.0000    Q1(25%): 16.0000          Q2(50%): 25.0000          Q3(75%): 40.0000
Max: 2300.0000
Missing: 8720
```

3.2针对数据集winemag-data-130k-v2.csv

In [58]:

```
#获得数据集的数值型数据
numerical_index_wine2 = list(df_wine2.select_dtypes(['int', 'float']))
numerical_summary(df_wine2, numerical_index_wine2)
```

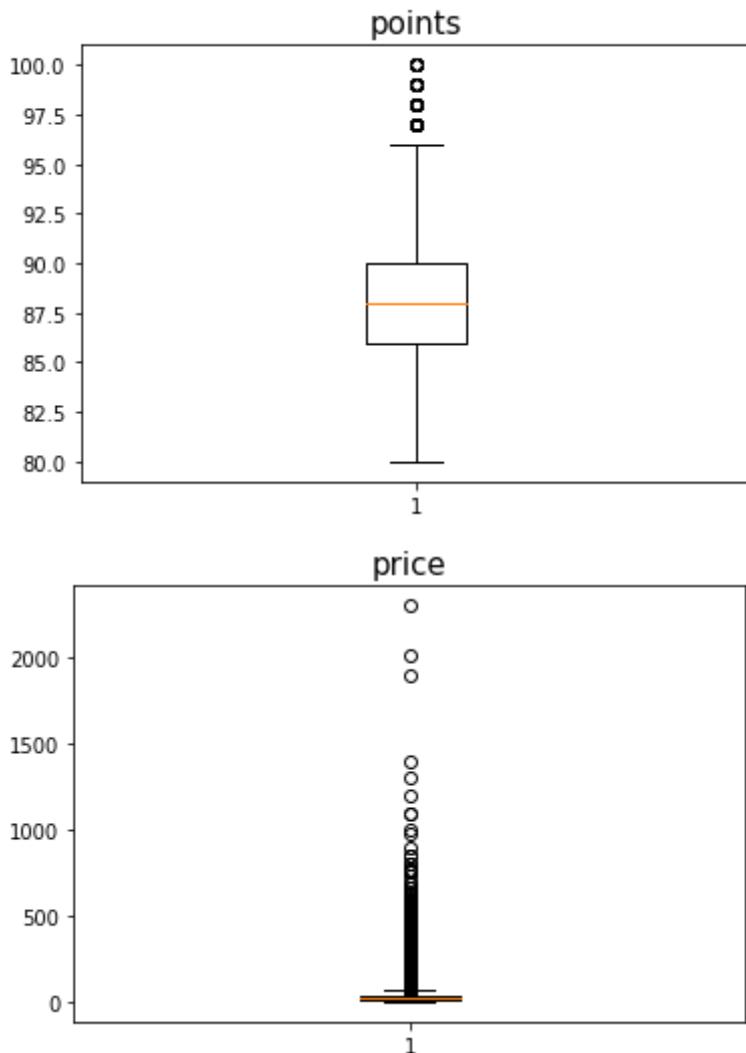
```
descriptive statistics (points):
Min: 80.0000    Q1 (25%): 86.0000        Q2 (50%): 88.0000        Q3 (75%): 91.0000
Max: 100.0000
Missing: 0

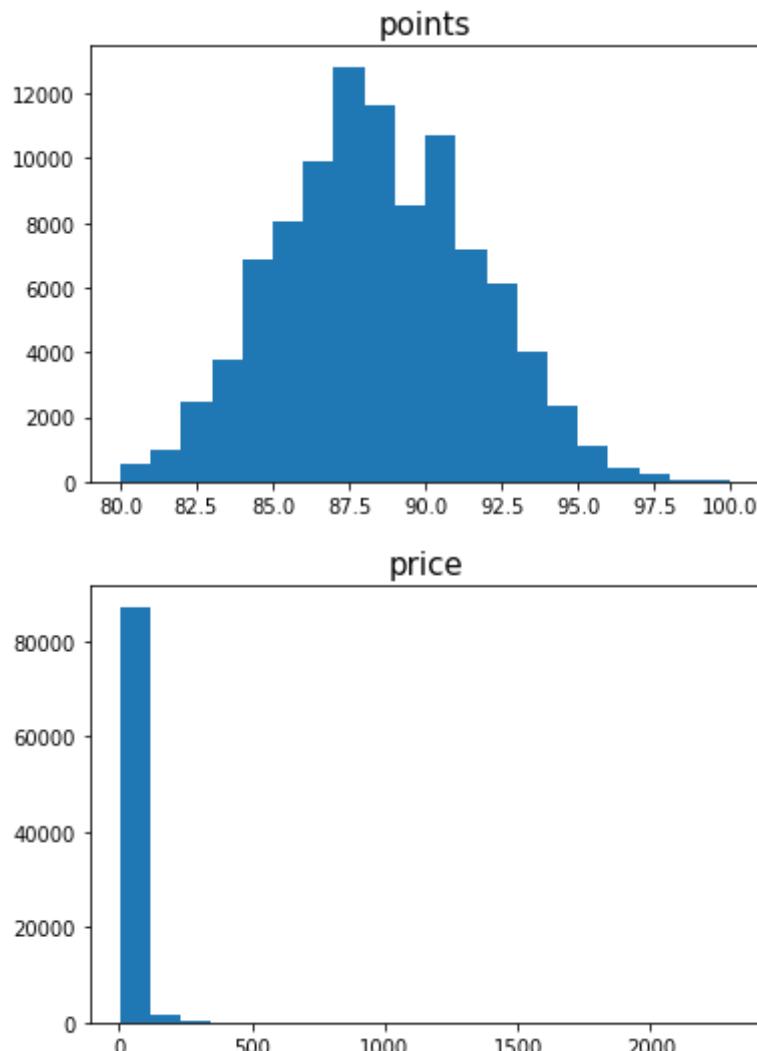
descriptive statistics (price):
Min: 4.0000    Q1 (25%): 17.0000        Q2 (50%): 25.0000        Q3 (75%): 42.0000
Max: 3300.0000
Missing: 8395
```

4.数据可视化 使用直方图、盒图等检查数据分布及离群

4.1针对数据集winemag-data_first150k.csv

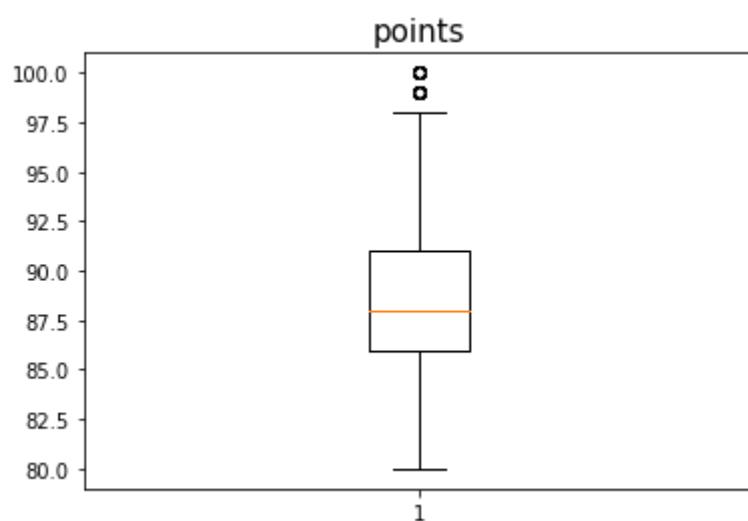
```
In [59]: box_plot(df_wine1, numerical_index_wine1)
hist_plot(df_wine1, numerical_index_wine1)
```

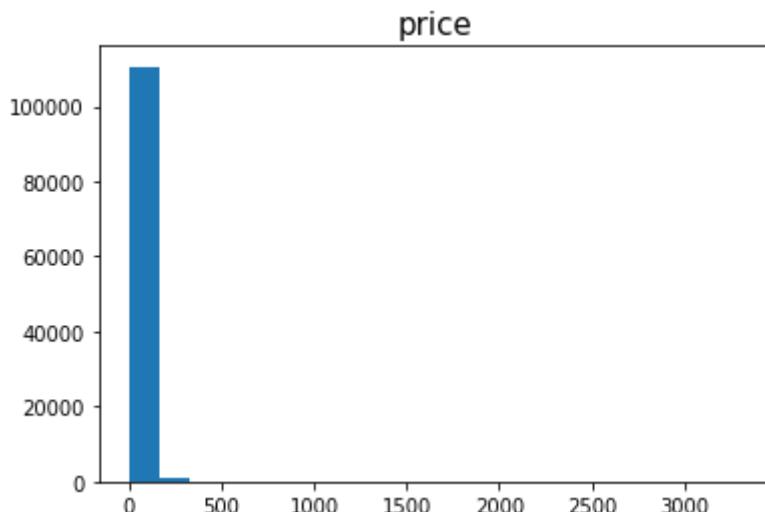
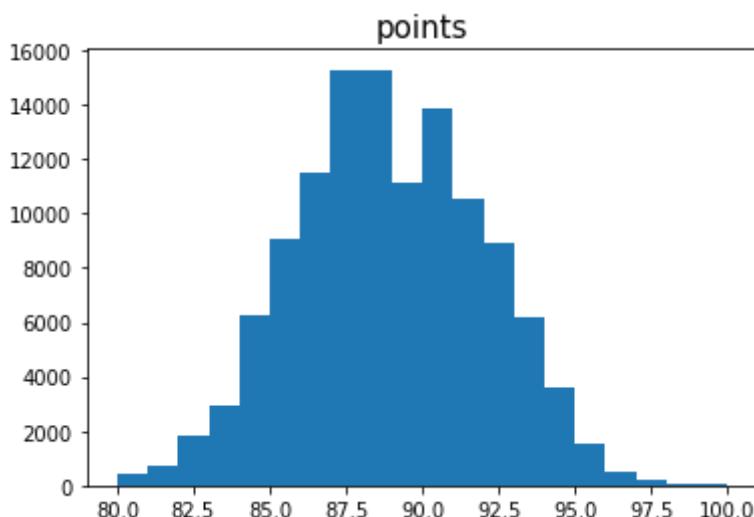
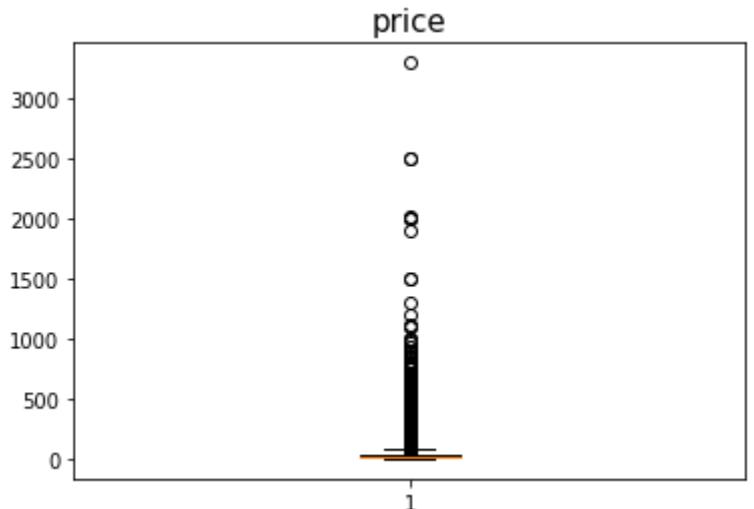




4.2 针对数据集winemag-data-130k-v2.csv

```
In [60]: box_plot(df_wine2, numerical_index_wine2)  
hist_plot(df_wine2, numerical_index_wine2)
```





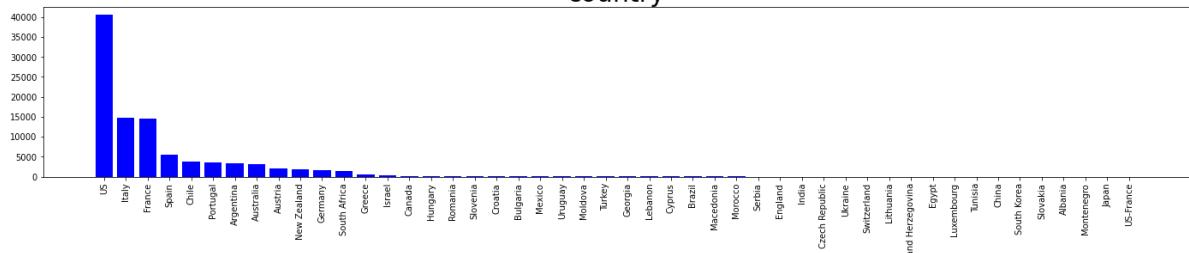
5. 观察数据集中缺失数据，分析其缺失的原因。

5.1 针对数据集 winemag-data_first150k.csv

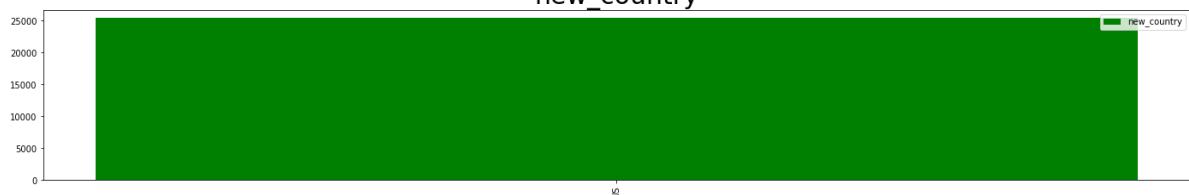
5.1.1 将缺失部分删除

```
In [61]: new_df_wine1 = miss_drop(df_wine1)
nominal_summary(df_wine1, nominal_index_wine1, new_df_wine1)
box_plot(df_wine1, numerical_index_wine1, new_df_wine1)
hist_plot(df_wine1, numerical_index_wine1, new_df_wine1)
```

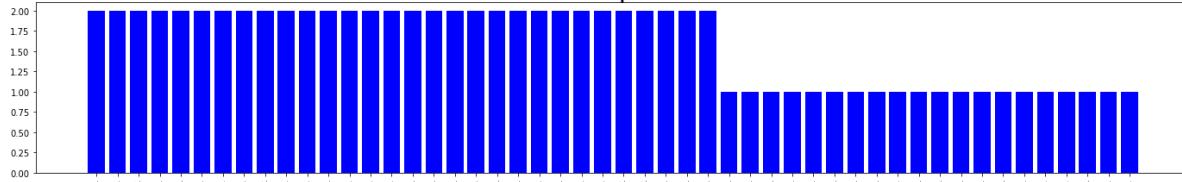
country



new_country



description



Spiry, with raspberry and spearment aromas. Shows cranberry and plum skin on the palate, which is lean and lacking in the texture category. Gets better with airing, but still shows some bitterness on the finish and not much in the middle.

A little bit funky and unsettled when you pop the screwcap, but soon it finds its floral, blueberry base. Remains superficial and sweet in the mouth, with candied flavors, vanilla and mid oak. Highly regular, could use more concentration and density. Glibinoro, a pure expression of Nero d'Avola, sees some oak aging to shape aromas of toasted walnut, nutella and wood spice. In fact, not much of the natural fruit is left, and what is there feels jammy and ripe.

In 2009 this single vineyard offering includes 11% Semillon. It has wonderful texture, and a mix of light herbal flavors that run through a garnet from cedar to lemongrass and up into tart citrus fruit. The acidity is generous but never overpowers. A quintessential food wine.

Show all the hallmarks of an important young Napa Cabernet. Bone dry and quite hard in tannins, it's immensely concentrated in blackberries, black currants, cassis and fine, smoky new oak. Possesses that elusive quality of dramatic terror, a real sense of place. Very fine young wine that needs time. Cellar until at least 2015.

Powerful, in Zinny character, this blend of Dry Creek and Russian River grapes bursts with brawny flavors of wild berries, chocolate-covered mint, tobacco and pepper, wrapped into sturdy tannins. Shows real class and character. A lean, cool, initially austere wine, well balanced with a touch of honeyed apricot sweetness. Has a story, unapologetically bold, with berry fruits and a poised elegant structure. The acidity gives great freshness to go with the structure, making a great food wine.

This wine shows some of the less appealing aspects of American oak: overt spice and vanilla flavors; and a sick, oily mouthfeel. Asparagus and green herb aromas run smoothly into grapefruit and sweet green pea flavors. More citrus fruits emerge on the tart, cleansing finish.

92.94 Barrel sample. A rounded wine, its tannins submerged into the ripe fruits. It feels soft, and there's just a late of alcohol. The structure is soft, gentle, opulent.

Steer clear of this one if you like your Sauvignons done in stainless. This one is 100% barrel fermented, yielding some toasty, smoky aromas that blend well with the forceful tropical fruit flavors. There's guava and pineapple, tangerine and menthol on the long finish.

Fresh, tropical, and welcoming. Sweet and lush in the mouth with full blast of mango and melon flavors. Medium-bodied with good acidity. Finishes correct and everything seems to be united and tall together. Above-average value Chardonnay from Argentina.

Simple, and candied, with blackberry jam and cola flavors. The tannins are rugged, and the wine finishes with a touch of acidity. Seems at its best now.

This solid Bordeaux achieves a nice balance between ripe fruit and a more austere earthiness. There's a solid core of blackberry flavor and firm, broad tannins that suggest it will develop more depth with nearer maturing.

92.92 Barrel sample. This is a solid and dense wine with hints of bitter chocolate along with superripe blackberry fruits. The wine has concentration, richness and a fine aging potential.

92.94 Barrel sample. A rounded wine, its tannins submerged into the ripe fruits. It feels soft, and there's just a late of alcohol. The structure is soft, gentle, opulent.

Almondic notions of peach precede a fresh, light and easy-drinking palate with fresh acidity. Very tannic, rough.

Pleasant enough and Chard-like, with aromas of tangerine, peach and a kiss of honeyed oak. It's got enough residual sugar to almost fall into the off-dry category. Well made and certainly affordable.

Dark, earthy and rubbery aromas go along with a reasonably fruity bouquet, and as the wine is good. The palate is full, a bit rough and rugged, with ripe, sweet black-cherry and cassis flavors. Concentrated and a touch syrupy, but good to what amounts to basic cabernet.

From Minkler Family estate vineyards, this intense and thrilling Riesling has a Germanic 8.5% alcohol and 5.5% residual sugar balance, yet with acids so vivid and juicy that its sweetness is completely counterbalanced. Loaded with spicy pear and tart tangarine fruit flavors, it's a penetrating, dense and highly refreshing wine that explores a style that is both complex and user friendly.

Deep, briny, wild-berry and leather aromas in the mouth. It's ripe and sweetly fruity, with soft tannins and Pepper-spice notes. It's bone dry, with expressive acidity and unpolished heabity.

92.94 Barrel sample. With so much spacy fruit, intensely ripe blackberry flavors and solid tannins, all the elements of excellence are here, and powered by a firm structure. This wine has weight and a fine Cabernet concentration.

If you've ever wondered what makes this wine so special, the answer is simple elegance. This shows polished fruit and spice aromas that are intense, balanced and beautifully integrated. The blend is 52% Cabernet Sauvignon, with the balance made up of Merlot, Cabernet Franc and Petit Verdot. The palate is dense, rich, smooth and supple. Drink this after 2018.

Ripe, plump, game, truffle, leather and menthol are some of the aromas you'll find on this earthy wine. The slightly round palate offers dried black cherry, chopped sage, mint and roasted tannins alongside racy tannins that leave a slighty numbing finish.

The fruity, perfumed characters show instant red berry aromas, but the old wood flavor doesn't jutchime with the fruit, giving a slightly muddled feel to the wine.

Dusty cherry and plum aromas are nothing special. The palate is similarly just OK, with stately flavors of citrus and fruits. This is lightly oxidized and pretty tasting on the finish.

Strongly aromatic of black pepper, this blend of Syrah, Zinfandel, Cabernet Sauvignon, Merlot and Cab Franc is herbal like. Cab and slightly sour/sweet on the finish.

The bouquet on this blend of chardonnay, Sauvignon Blanc and Pinot Noir is stuck in neutral, while flavors of citrus and lemon never towards sour. The finish is tannic and tangy.

Mild lemon and apple aromas are compromised by a harsh, stony element. The mouthfeel is basic and spritzzy, while generic apple and green melon flavors in tropical sweetness.

Following up on a bouquet that's lightly herbal and lactic, the palate feels hard and tannic, with oak-based resin sticking out. Flavors of black pepper, licorice and sage pitch away the wine's black-fruit component, leaving a minty finish that's sake dominant and mildly artifical tasting.

This is herbal and tonikoly on the nose, then acry in feel, with little midlevel stuffing on the palate. Flavors of raspberry are acute and offset by mild buttery oak notes. This wine shows some elements of classic Spain but weaknesses as well.

Maderized in color, this wine has a yeasty, creamy nose with baked white-fruit aromas and caramelly. It's OK in feel, with pickled, milky briny flavors of apple and apricot. The finish is showing some oxidation, leading to a chunky, fizzy feel.

This Chardonnay is slightly metallic and briny on the bouquet. It feels plump and chunky, with citrus and pineapple flavors coming in front of a quiet finish.

Mint aromas speak of the wood used in aging this wine. With some burnt tannin character, it's perhaps not a wine that will show great fruit.

A sweetly nose to taste, then this turns into a plummy, tannic, astrigent red with a hint of barrel toast. The nose on this Viura is flat, with walnut oil and dusty apple notes. The mouthfeel is solid, while the flavor's suggest dry apricot, apple and nutty notes. This is dry and modest on the finish.

This is a pricky, slightly racy wine with a tangy, garby feel and briny, pickled flavors of herald red berry fruits. It's ready tasting on the finish, with a snaply feel and lasting herbal flavors.

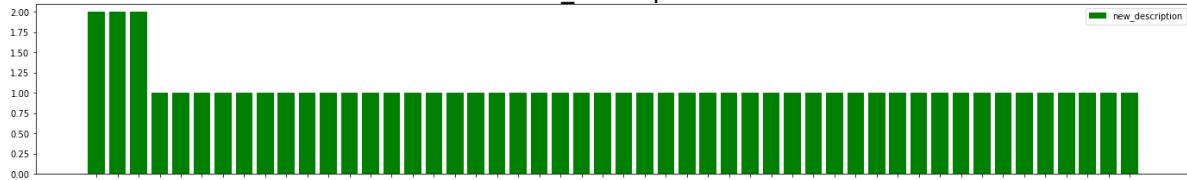
This Cava has a funny nose that's unusual. The palate's a bit flabby, with sweet banana and peaches flavor. There's not much to the finish and overall this comes across as overly simple and strained.

Showcasing some of Sardinia's most widely planted indigenous grapes—Carignano, Cannonau and Bonarda—Saracu is bold, round and ewe, with a long, smooth, velvety texture. On the nose, the wine presents black-cherry, spice and mid-to-coconuts.

Dripping with notes of juicy black fruit, sweet tobacco and clove, this is layered and full-bodied, with soft tannins and an intensity that should grow more intriguing with age. From the prestigious Calveras County Dalton Vineyard, this is blends with 21% Petter Saracu, also from Dalton.

This is a big, fruit-driven style of Shiraz, with great aromas and flavors of camphor, vanilla, dark toast and cassis. There's just enough structure to give it focus, although it lacks a bit of textural richness on the midpalate. Drink this supple wine over the next few years.

new_description



As weighty as a Chardonnay, this is a Sauvignon Blanc whose nuances will be killed if you over-chill it. You could even decant it for a while and let it approach room temperature. Dry and rich, it shows savory, complex tiers of pineapples, limes, green apples, sweet golden tobacco, vanilla, white pepper and even a touch of thyme.

Savory on the nose with modest apple and pear aromas, accented by a curiously smoked ham note, this smoky off-dry Riesling with a touch of honey crisp apple sweetness has a briny, umami-laden complexity that would make it an interesting companion to prosciutto wrapped around morsels of ham and fig.

This tremendous 100% varietal wine hails from Oakville and was aged over three years in oak. Juicy, tart-cherry fruit, and a compelling hint of caramel greet the palate, framed by elegant, fine tannins and a subtle minty tone in the background. Balanced and rewarding from start to finish, it has years ahead of it to develop further nuance. Enjoy 2022-2030.

Deep, briny, wild-berry and pepper aromas. In the mouth, it's ripe and sweetly fruity, with soft tannins and pepper-spice notes. It's bone dry, with adequate acidity. Classic Sonoma Zin in its expressive exuberance and unpolished readability.

A useful red wine, medium-bodied and fruity, with black cherry, blueberry and spice flavors are savory, easy to like with everything from burgers to steaks and summer fries. It's a Rhône-style blend with 15% Tempranillo.

Made using fruit that was sourced from 30- and 40-year-old plantings, this has captivating flavors of black cherry and cassis, wrapped in smoky hints of herb. It shows plenty of intense power, with many more layers surfacing with breathing time. In fact, it's just as good on its second day.

Sommeliers will love this Chardonnay for its team, chablis-like elegance. The lemon and passion fruit flavors stand alongside stony minerality and a tarty, sourdough bread tartness, while the brisk acidity makes the mouth water. Elegant, rather than opulent, this gather momentum as it warms in the glass.

A Southern Rhône-style blend of 62% Grenache, 31% Cinsault and 7% Syrah, this is loaded with flavors of wild herb and bright berry. From the Circuitait comes a hint of muscatine, while the Syrah lends muscle and peppery fruit. This has excellent concentration, complex flavors, and appealing length.

Acidity dominates this wine, giving it a tartness—like a squeeze of lime juice, just that is an enormous rich, dry, vibrant wine, explosive in raspberries, cherries and ripe, sweet persimmons. New French oak, with its caramel-like toastiness, is perfectly in keeping with the wine's volume. It's a bit disjointed now, but give it 3-4 years in the cellar, and it should come around.

This young wine emulates the style of Northeast Italy in its tart, grassy flavors. The fruit treks toward light melon and green apple. It's crisp and refreshing. This wine shows great Chardonnay character, from the creamy texture to the notes of buttered toast, vanilla, lemon and lime.

In comparison to the winery's regular 2021 Pinot Noir, the reserve seems more acidic, and the fruit not quite as ripe. Russet rather than raspberry is the dominant note, with a whiff of chocolate or the finish. This wine shows great Chardonnay character, from the creamy texture to the notes of buttered toast, vanilla, lemon and lime.

Winemaker Greg Griswold travels into uncharted territory in this three "Pinot" blend—45% Pinotage, 42% Pinot Meunier and 13% Pinot Noir. Bursting in soft layers of Rainier cherry, it delves into a bed of mineral on the finish and will pair magnificently with smoked salmon. This is 95% Syrah, unlike past vintages where it was inoculated after a Côte du Rhône-style blend. It opens with enticing wafts of soft leather and toasted nut, before bringing in a smooth mix of black and purple fruits, along with chocolate and anise highlights. It's supple and proportionate, with polished, coffee-flavored tannins.

Made using fruit that was sourced from Burgundy for its team, chablis-like elegance. The lemon and passion fruit flavors stand alongside stony minerality and a tarty, sourdough bread tartness, while the brisk acidity makes the mouth water. Elegant, rather than opulent, this gather momentum as it warms in the glass.

Sommeliers will love this Chardonnay for its team, chablis-like elegance. The lemon and passion fruit flavors stand alongside stony minerality and a tarty, sourdough bread tartness, while the brisk acidity makes the mouth water. Elegant, rather than opulent, this gather momentum as it warms in the glass.

A fine effort in a cool year, this reserve combines the best barrels from the winery's three estate vineyards, which occupy three different sub-AVAs. Hence the Willamette Valley appellation. With excellent depth and power, it showcases a core of cherry, berry and chocolate, accented with sweet spices and a hint of fresh herb.

This is a new cuvee for Traestum, from a vineyard in the Red Hills of Dundee. It shows some old vine character in its brambly, Burgundian elegance. Light flavors of raspberry and cherry float above a bed of tannin; yet there's enough muscle for the wine to extend gracefully through a lingering finish, with a nice of chocolate. Enjoy it over the next six years, yes. This is a new cuvee for Traestum, from a vineyard in the Red Hills of Dundee. It shows some old vine character in its brambly, Burgundian elegance. Light flavors of raspberry and cherry float above a bed of tannin; yet there's enough muscle for the wine to extend gracefully through a lingering finish, with a nice of chocolate. Enjoy it over the next six years, yes.

Rustic and ripe, but easy to like, with red licorice, mocha, licorice, toasty oak and dusty tannins. By 2016 it should start to find its harmonious sweet spot. This is a new cuvee for Traestum, from a vineyard in the Red Hills of Dundee. It shows some old vine character in its brambly, Burgundian elegance. Light flavors of raspberry and cherry float above a bed of tannin; yet there's enough muscle for the wine to extend gracefully through a lingering finish, with a nice of chocolate. Enjoy it over the next six years, yes.

Fine acidity and a bracing minerality, like a tickle of cold steel on a winter day, mark this Chardonnay. It has oak notes, but they're wisely kept in the background, allowing the citrus, green apple and peach flavors to stand. This beautiful wine continues the winery's string of great reserve Chardonnays, and is their best since the 2006 vintage. Thickly moussed, Roederer's Brut, 60% Chardonnay blended with 40% Pinot Noir, aged two years on yeast is affordable and luxurious. The flavors are serious, suggesting raspberries and cherries with earthy notes of mushrooms, clove and henna. It would not be surprising if this wine was still gaining traction in 2020, but you can decant it and drink it now.

This is the medium-dry Riesling from this estate vineyard, a notch lower in alcohol than the dry bottling, and overall a bit rounder, carries the same excellent texture, length and mix of ripe fruits, with pears and peaches taking center stage.

A fine effort in a cool year, this reserve combines the best barrels from the winery's three estate vineyards, which occupy three different sub-AVAs. Hence the Willamette Valley appellation. With excellent depth and power, it showcases a core of cherry, berry and chocolate, accented with sweet spices and a hint of fresh herb.

This is a new cuvee for Traestum, from a vineyard in the Red Hills of Dundee. It shows some old vine character in its brambly, Burgundian elegance. Light flavors of raspberry and cherry float above a bed of tannin; yet there's enough muscle for the wine to extend gracefully through a lingering finish, with a nice of chocolate. Enjoy it over the next six years, yes.

Give this wine 3-5 years in the cellar to let the parts come together. Right now, it's an unexpected melding of sweet cherry pie, mocha, licorice, toasty oak and dusty tannins. By 2016 it should start to find its harmonious sweet spot. This is a new cuvee for Traestum, from a vineyard in the Red Hills of Dundee. It shows some old vine character in its brambly, Burgundian elegance. Light flavors of raspberry and cherry float above a bed of tannin; yet there's enough muscle for the wine to extend gracefully through a lingering finish, with a nice of chocolate. Enjoy it over the next six years, yes.

Chardonnays this rich need crisp acidity, minerality and a dry finish to balance them. That is exactly what you'll find in this beautiful example. For all the lightness of mouthfeel, the flavors are serious, suggesting raspberries and cherries with earthy notes of mushrooms, clove and henna. It would not be surprising if this wine was still gaining traction in 2020, but you can decant it and drink it now.

This is a good wine, but it relies too much on oak, overwhelming the fruit. The result is a rich but unbalanced wine, with underlying orange, pineapple and peach flavors.

Bottled with a platter of Petit Verdot and Merlot, this Cabernet is rich, soft and smooth. The blackberry, currant and chocolate flavors are at their full, primary maturity, and promise a long, evolving aging process in the years ahead. It's a shame that just 100 cases of this wine were made, as this bright, fresh, exciting Zinfandel hails from the best of California, not just Washington. The fresh-picked strawberry and raspberry flavors are concentrated and vibrating with juicy acidity. They leap into a supple, lengthy midpalate, with an underpinning of rock.

Rustic and ripe, but easy to like, with red licorice, mocha and blackberry jam flavors. Made from an assortment of grape varieties, including Syrah, Merlot and Barbera. The vineyard is a fine one, the fruit source for highly regarded photos from Tandem and Ballwick's own Silver Pines label. The wine is dry and silky, with strong acidity and tannins that currently render it tough to appreciate. Give it 5-6 years to let the raspberry and clove flavors integrate with the oak.

With mouthwatering acidity, this wine captures each of those flavors, deviating slightly towards tart cherries. The finish enters Syrah territory, with earthy leathery and spice. A non vintage blend of Zinfandel Syrah and Petite Sirah, this wine captures each of those flavors, deviating slightly towards tart cherries. The finish enters Syrah territory, with earthy leathery and spice.

This is a sound, everyday Merlot with some special qualities that give it an extra point or two. It's dry and smooth in tannins, with layers of cherries, blackberries, currants and licorice. Not likely to age well, so give it a little time to breathe and drink now.

A very distinctive Sauvignon Blanc with feisty notes of biscuit, spice and mustard at the fore along with earthy tannins and a sense of smoky acidity that runs from head to tail. This is a sound, everyday Merlot with some special qualities that give it an extra point or two. It's dry and smooth in tannins, with layers of cherries, blackberries, currants and licorice. Not likely to age well, so give it a little time to breathe and drink now.

This is a solid, well-structured Cabernet Franc from Blackrock Vineyards in the Yalumba Valley. The notes are light and high toned with still vanilla, fresh and juicy tannins. The 40% new American oak gives much of the flavor as well as the tannins, which come off as a bit too aggressive. While the nose is ripe and plush with sweet smelling red cherry and strawberry aromas, this is a surprisingly brisk, light-tinted fruit and drying slightly astrigent tannins on the finish.

A solid, well-structured Cabernet, dry tannins and balanced. Shows blackberry and black currant flavors that are tinged with smoky oak. Drink now.

This is a good Syrah that's dry and spicy and ready to drink now with a bright berry flavor. The blackberry flavors have earthy, tobacco and grilled meat notes. It's a bit sharp. Drink up.

Not as dry as the name suggests, this offering seems fairly sweet, with sugary lime, orange and vanilla flavors. You'll find a rich streak of mouthy yeast and acidity driving it that typical Chardonnay note.

This Pinot is ripe in jammy cherry fruit, and it has a smoky, buttery richness from oak. But it's a little soft and heavy for the variety, which is perhaps to be expected given its Alexander Valley origin. There's lots of ripe jammy blackberry and cherry fruits in this Petit Verdot-based blend. It's dry, robust, spicy and a little rustic. Drink it now with pizza or a burger.

With a sour, scratchy mouthfeel, this wine lacks acidity. But it is rich and layered in strawberries and red raspberries, and is easy to drink.

A big but not too-tannic Petit Verdot, this is juicy and structured with fed fruit and a light layer of spice. Tending toward black licorice and dark chocolate, I could use a splash of something to brighten it up. But it's otherwise a bottle worth trying if you're curious to try a vinyl Petit Verdot.

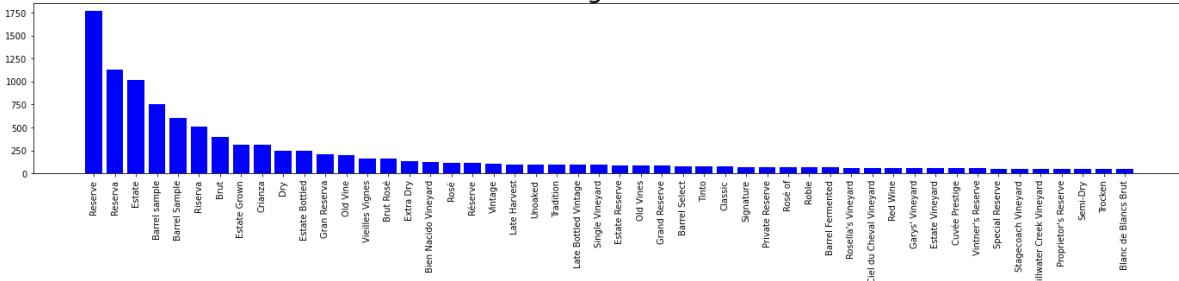
This wine is ripe, forward flavors of pineapples, golden mangoes, honey and spicy peaches. You'll find a rich streak of mouthy yeast and acidity driving it that typical Chardonnay note.

Net fully ripe, this has a flavor of green tomato, earthy aldehydes, notes of cherry and cola. It shows the delicate texture and crisp acidity you expect from this coastal vineyard, but don't try to age this wine. Drink it now.

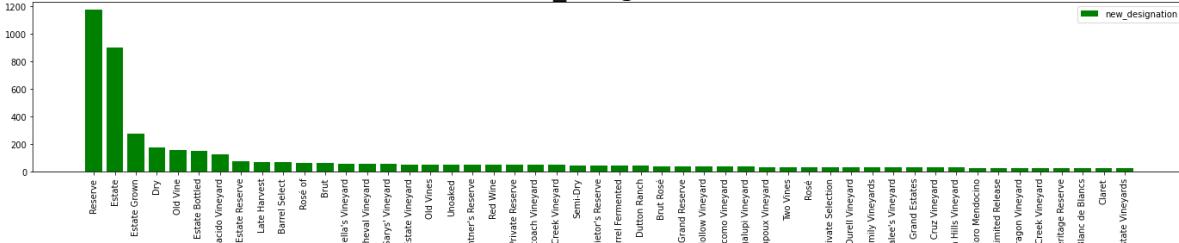
Based on Merlot, this Bordeaux-style blend is very ripe and forward in blueberry, cherry and raspberry jam, and it finishes in a swirl of dusty baking spices. It's a good wine, if a bit rustic, with some earthy tannins and high alcohol.

On its own, this is dry and earthy, with a taste of Medieval Eastern Hammurabi, spiced oats and oranges. It's certainly offbeat, but it can certainly be paired with food. How about hummus with tortilla chips? This is crisp in acidity, tasting slightly sweet in citrus and apricot jam, with notes of green pepper and gooseberry. It has the versatility to pair with a wide range of foods, including Asian-style fare and roasted chicken. There's a soft voluptuousness to this Syrah that makes it instantly drinkable. It's intense in blackberry, cassis, mocha and grilled prosciutto notes, with lots of new French-oak influence adding toast, vanilla and caramel flavors. As sweet as it tastes, it finishes dry, long and spicy.

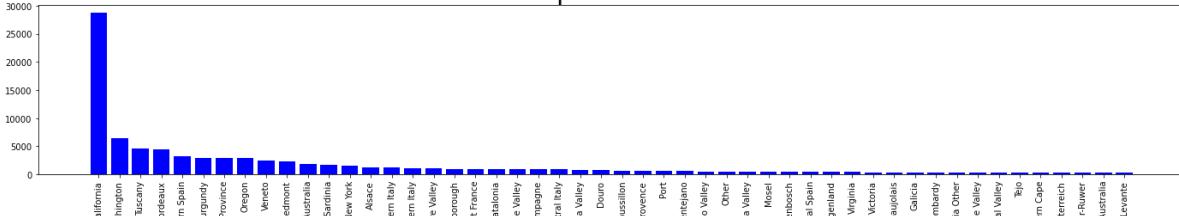
designation



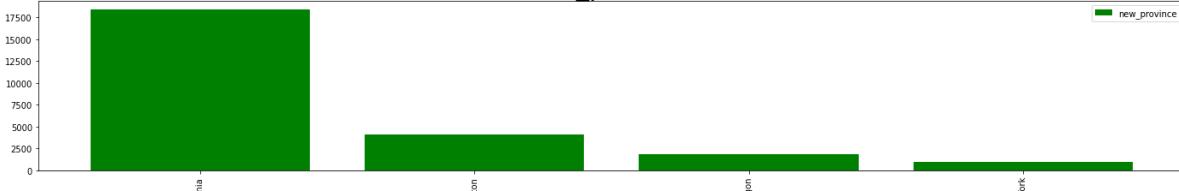
new_designation



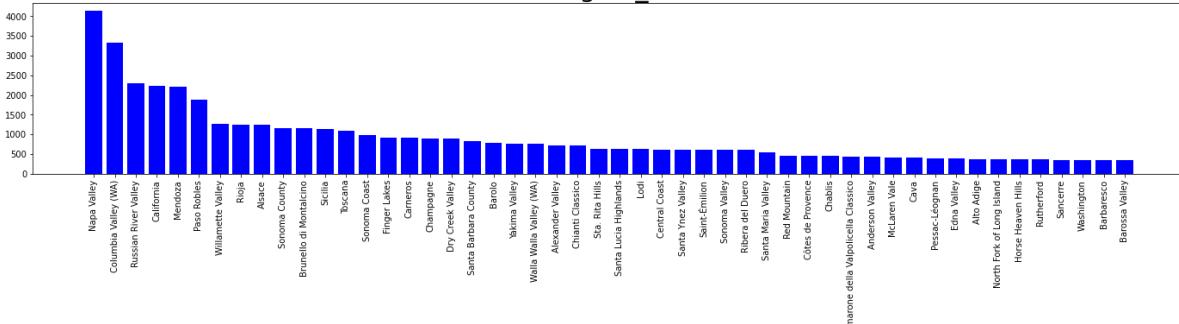
province



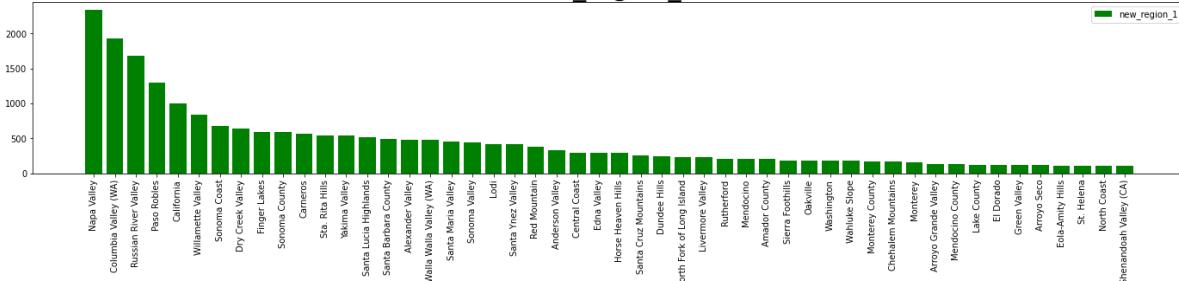
new_province



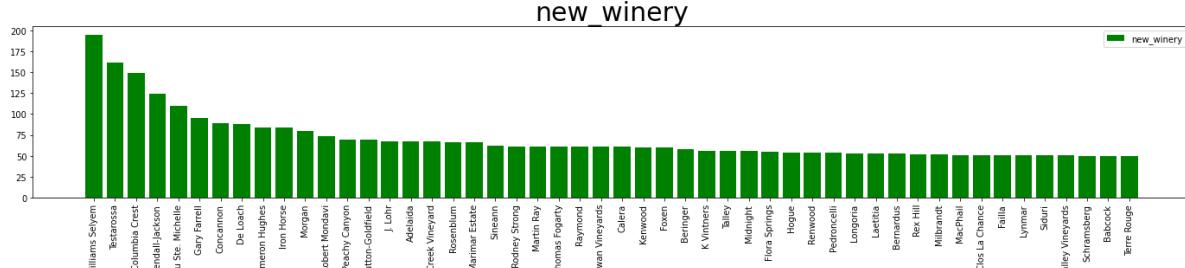
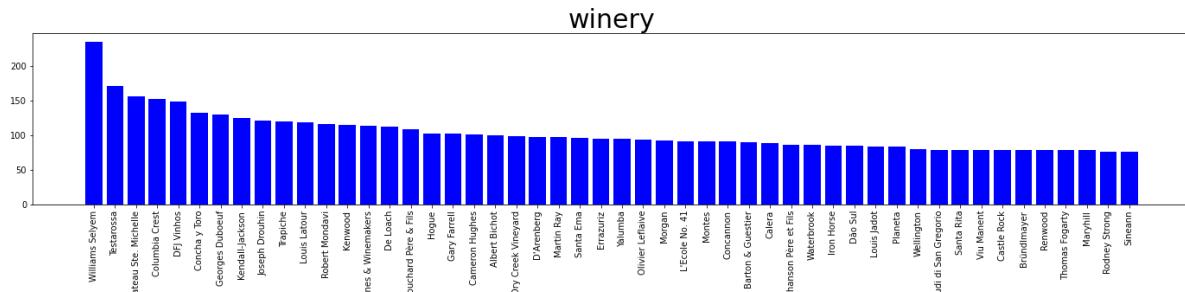
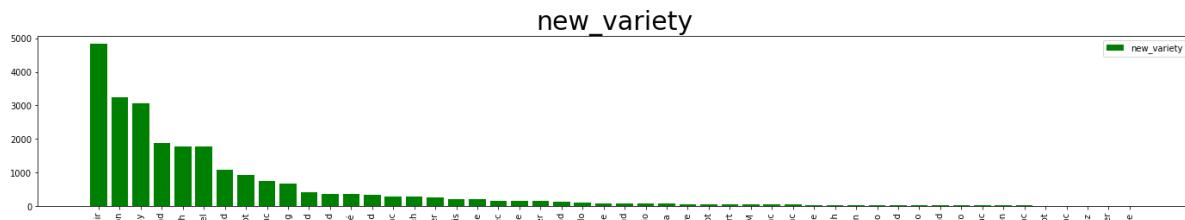
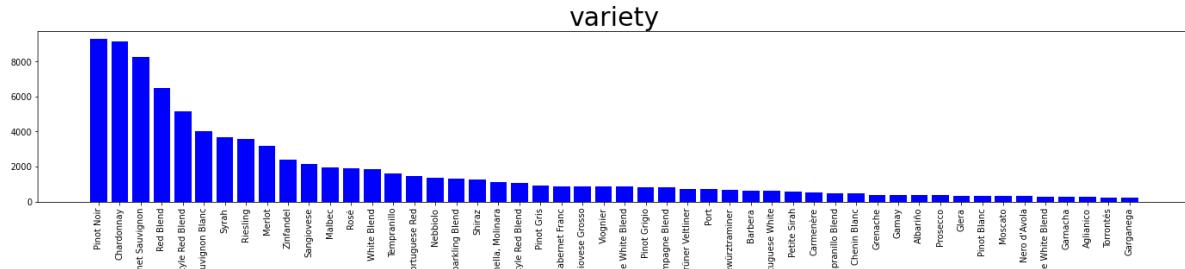
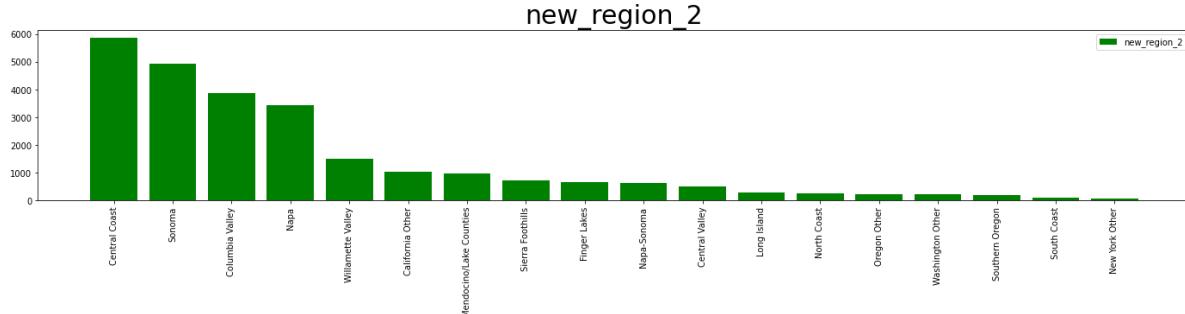
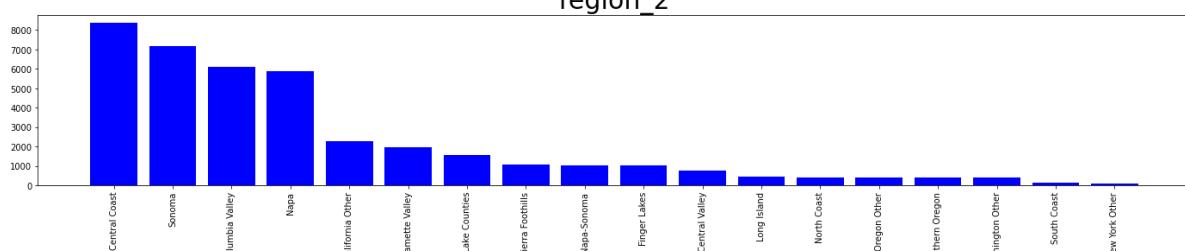
region_1

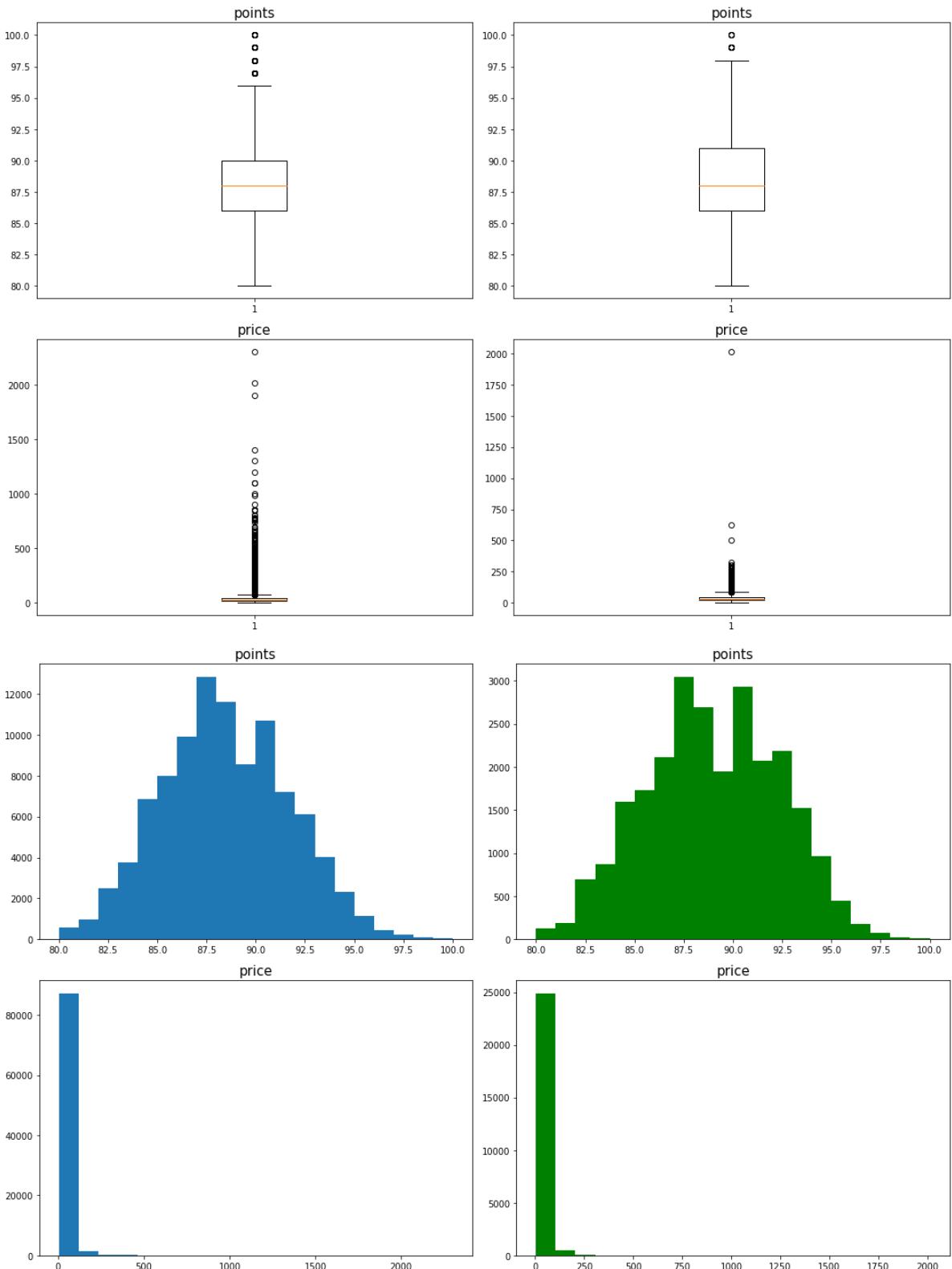


new_region_1



datamining week3



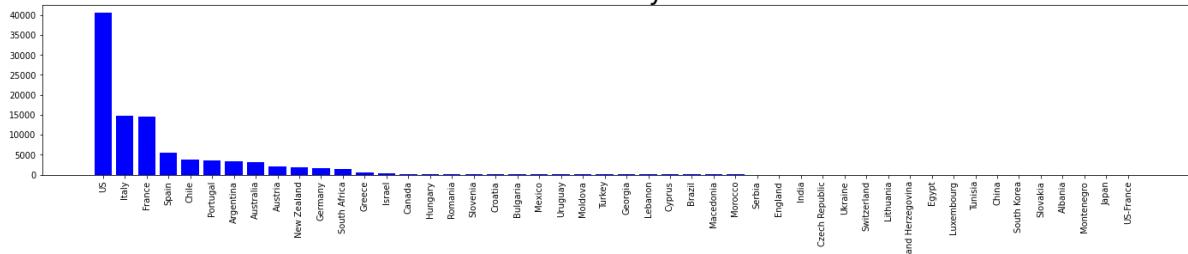


[分析]从图中分析得出本方法会导致country减少至一个，description和designation出现较大变化，province减少至三个，造成明显的数据分布错误，故此方法不可取

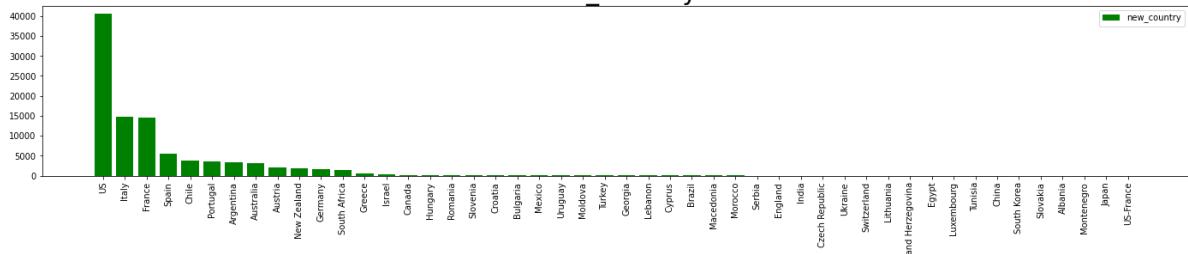
5.1.2用最高频率值来填补缺失值

```
In [62]: new_df_wine1 = miss_mode_fill(df_wine1)
nominal_summary(df_wine1, nominal_index_wine1, new_df_wine1)
box_plot(df_wine1, numerical_index_wine1, new_df_wine1)
hist_plot(df_wine1, numerical_index_wine1, new_df_wine1)
```

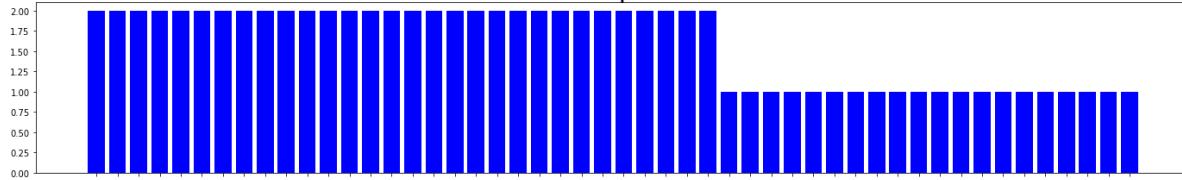
country



new_country



description



Spiry, with raspberry and spearment aromas. Shows cranberry and plum skin on the palate, which is lean and lacking in the texture category. Gets better with airing, but still shows some bitterness on the finish and not much in the middle.

A little bit funky and unsettled when you pop the screwcap, but soon it finds its floral, blueberry base. Remains superficial and sweet in the mouth, with candied flavors, vanilla and mid oak. Highly regular, could use more concentration and density. Glibinoro, a pure expression of Nero d'Avola, sees some oak aging to shape aromas of toasted walnut, nutella and wood spice. In fact, not much of the natural fruit is left, and what is there feels jammy and ripe.

In 2009 this single vineyard offering includes 11% Semillon. It has wonderful texture, and a mix of light herbal flavors that run through a garnet from citrus fruit. The acidity is generous but never overpowering. A quintessential food wine. Shows all the hallmarks of an important young Napa Cabernet. Bone dry and quite hard in tannins, it's immensely concentrated in blackberries, black currants, cassis and fine, smoky new oak. Possesses that elusive quality of dramatic terror, a real sense of place. Very fine young wine that needs time. Cellar until at least 2015.

A seductive nose of pineapple, peat, vanilla and buttered toast leads, followed by nicely knit, well-fruitful, but not at all sweet palate. Here, heat, medium, pineapple and tangerine is offset by light toast and vanilla accents. Balanced and smoothly rounded. It shows graceful restraint that suggests it will improve over the next year or two.

Powful in Zinny character, this blend of Dry Creek and Russian River grapes bursts with brawny flavors of wild berries, chocolate-covered mint, tobacco and pepper, wrapped into sturdy tannins. Shows real class and character. Sauvignon on the nose with modest apple and pear aromas accented by a curious smoked ham note. This smoky, unbalanced complexity, that could make an interesting companion to prosciutto wrapped around medallions of melon and fig.

A lean, cool, initially austere wine, well balanced with red berry fruits and a poised elegant structure. The acidity gives great freshness to go with the structure, making a great food wine.

This wine shows some of the less appealing aspects of American oak: overt spice and vanilla flavors; and a sick, oily mouthfeel.

Very ripe and forward in baked fruit or pie-filling flavors of cherries and raspberries, spiced with cinnamon, anise, and nutmeg. The wine is big, simple and is ready to drink now.

Sharp, simple and candied, with blackberry jam and cola flavors. The tannins are rugged, and the wine finishes with a touch of acidity. Seems at its best now.

This solid Bordeaux achieves a nice balance between ripe fruit and a more austere earthiness. There's a solid core of blackberry flavor and firm, broad tannins that suggest it will develop more depth with nearer maturing.

Asparagus and green herb aromas run smoothly into grapefruit and sweet green-pea flavors. More citrus fruits emerge on the tart, cleansing finish.

92-94 Barrel sample. This is a cold and dense wine with hints of bitter chocolate along with superripe blackberry fruits. The wine has concentration, richness and a fine aging potential.

92-94 Barrel sample. A rounded wine, its tannins submerged into the ripe fruits. It feels soft, and there's just a late of alcohol. The structure is soft, gentle, opulent.

Steer clear of this one if you like your Sauvignons done in stainless. This one is 100% barrel fermented, yielding some toasty, smoky aromas that blend well with the forceful tropical fruit flavors. There's guava and pineapple, tangelo and menthol on the long finish.

Pleasant enough and Chard-like, with aromas of tangerine, peach and a kiss of honeyed oak. It's got enough residual sugar to almost fall into the off-dry category. Well made and certainly affordable.

Dark, earthy and rubbery aromas go along with a reasonably fruity bouquet, and as the wine is good, the acids so vivid and juicy that its sweetness is completely counterbalanced. Loaded with spicy pear and tart tangerine fruit flavors, it's a penetrating, dense and highly refined wine that explores a style that is both complex and user friendly.

From Minkler Family estate vineyards, this intense and thrilling Riesling has a Germanic 8.5% alcohol and 5.5% residual sugar balance, yet with acids so vivid and juicy that its sweetness is completely counterbalanced. Loaded with spicy pear and tart tangerine fruit flavors, it's a penetrating, dense and highly refined wine that explores a style that is both complex and user friendly.

Dark, earthy and rubbery aromas in the mouth, it's ripe and sweetly fruity, with soft tannins and Pepper-spice notes. It's bone dry, with slightly acidic notes.

Dusty cherry, wild-berry and leather aromas go along with a reasonably fruity bouquet, and as the wine is good, the acids so vivid and juicy that its sweetness is completely counterbalanced. Loaded with spicy pear and tart tangerine fruit flavors, it's a penetrating, dense and highly refined wine that explores a style that is both complex and user friendly.

Stony aromatics of black pepper, this blend of Syrah, Zinfandel, Cabernet Sauvignon, Merlot and Cab Franc is herby like. Cab and slightly sour-sweet on the finish.

The bouquet on this blend of Chardonnay, Sauvignon Blanc and Pinot Noir is stuck in neutral, while flavors of citrus and lemon never towards sour. The finish is tannic and tangy.

Mild lemon and apricot aromas are compromised by a harsh, stony element. The mouthfeel is basic and spritzzy, while generic apple and green melon flavors and in tropical sweetness.

Following up on a bouquet that's lightly herbal and lactic, the palate feels hard and tannic, with oak-based resin sticking out. Flavors of black pepper, licorice and sage pitch away the wine's black-fruit component, leaving a minty finish that's oak-dominated and mildly artifical tasting.

This is herbal and tonikoly on the nose, then acry in feel, with little midlevel stuffing on the palate. Flavors of raspberry are acute and offset by mild berry oak notes. This wine shows some elements of classic Spain but weaknesses as well.

Maderized in color, this wine has a yeasty, creamy nose with baked white-fruit aromas and caramel. It's OK in feel, with picked, milky briny flavors of apple and apricot. The finish is showing some oxidation, leading to a chunky, fizzy feel.

This Chardonnay is slightly metallic and briny on the bouquet. It feels plump and chunky, with citrus and pineapple flavors coming in front of a quiet finish.

Mint aromas speak of the rind used in aging this wine. With some burnt tannin character, it's perhaps not a wine that will show great fruit.

A sweetly nose to start, then this turns into a plummy, tannic, astrigent red with a hint of barrel toast. The nose on this Viura is flat, with walnut oil and dusty apple notes. The mouthfeel is solid, while the flavor's suggest dry apricot, apple and nutty notes. This is dry and moderate on the finish.

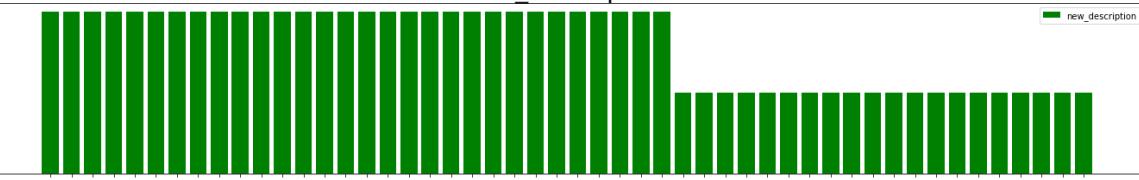
This is a pricky, slightly racy wine with a tangy, garby feel and briny, pickled flavors of herald, red berry fruits. It's ready tasting on the finish, with a snap-y feel and lasting herbal flavors.

This Cava has a funny nose that's unusual. The palate's a bit flabby, with sweet banana and peachy flavors. There's not much to the finish and overall this comes across as overly simple and strained.

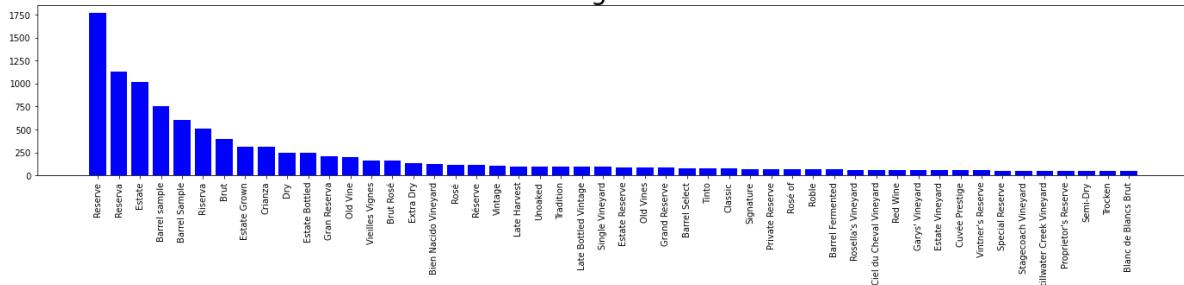
Dripping with notes of juicy black fruit, sweet tobacco and clove, this is layered and full-bodied, with soft tannins and an intensity that should grow more intriguing with age. From the prestigious Calveras County Dalton Vineyard, this is blends with 21% Peter Sraha, also from Dalton.

This is a big, fruit-driven style of Shiraz, with great aromas and flavors of camphor, vanilla, dark toast and cassis. There's just enough structure to give it focus, although it lacks a bit of textural richness on the midpalate. Drink this supple wine over the next few years.

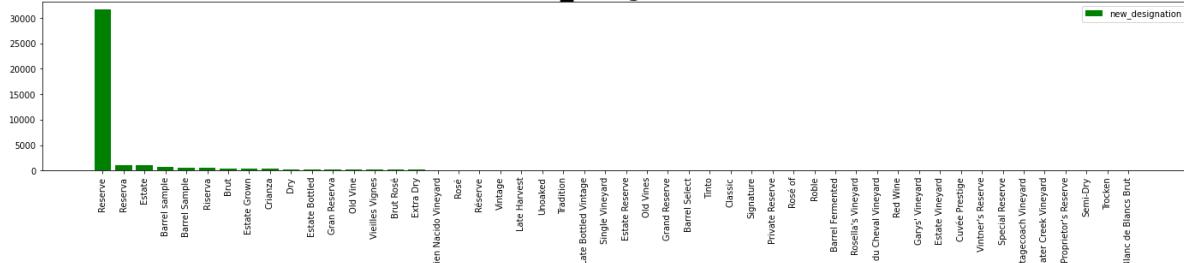
new_description



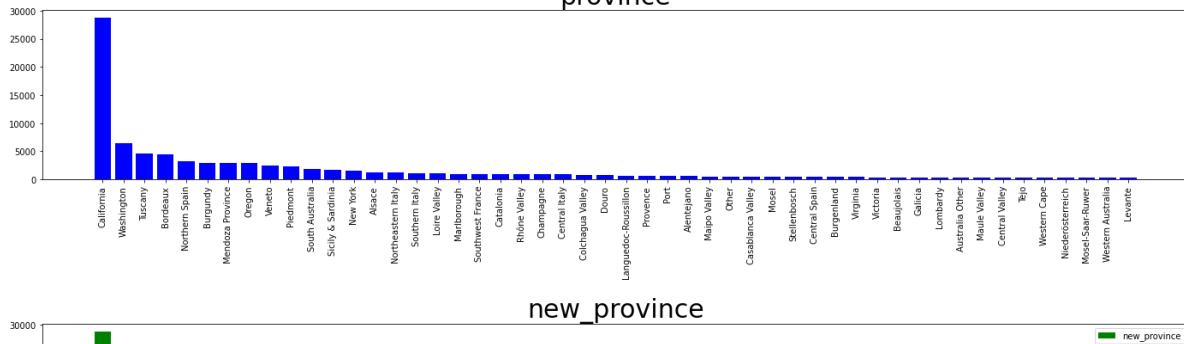
designation



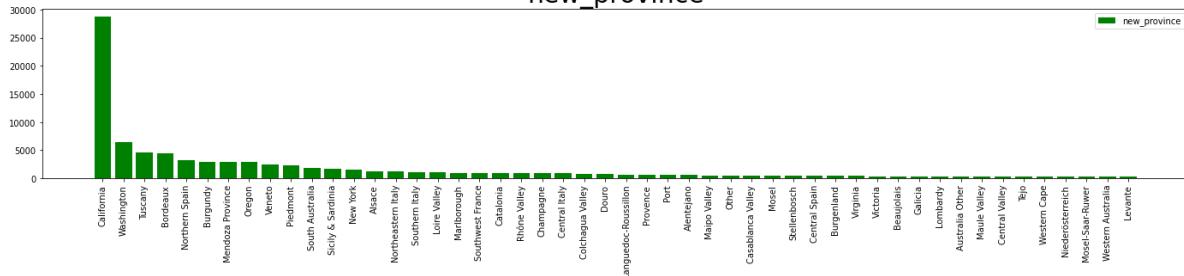
new_designation



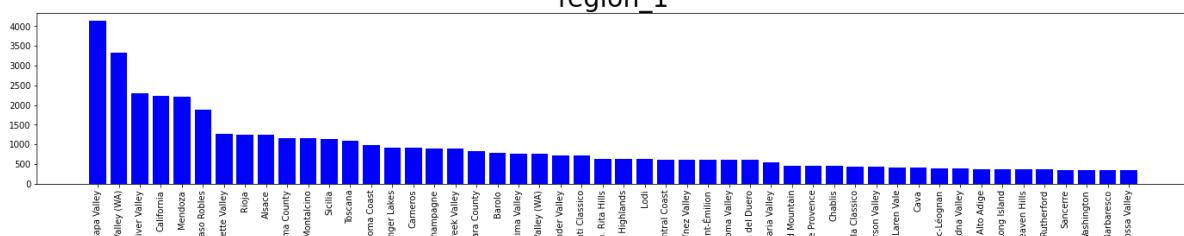
province



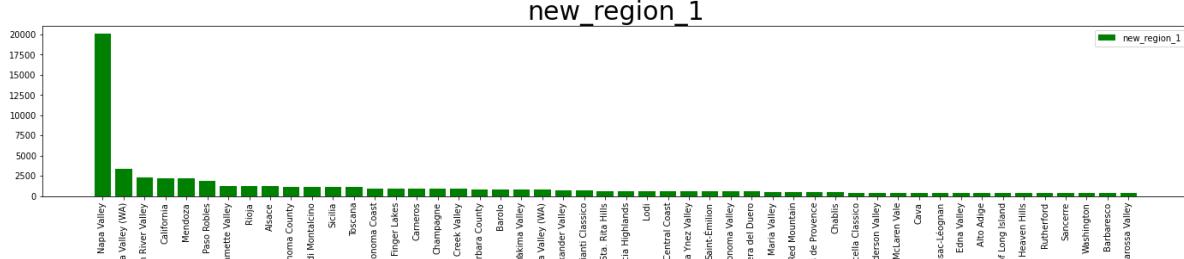
new_province



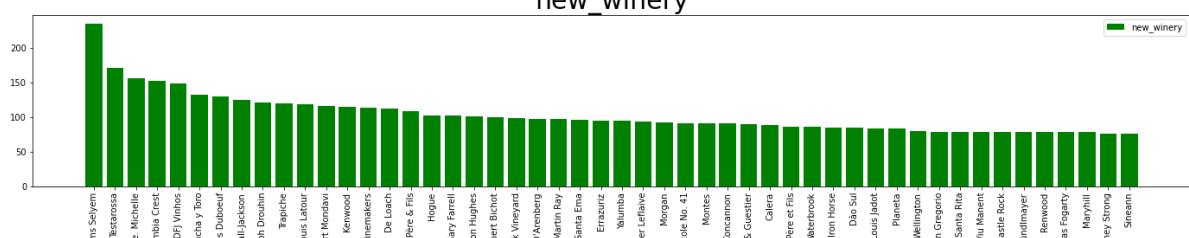
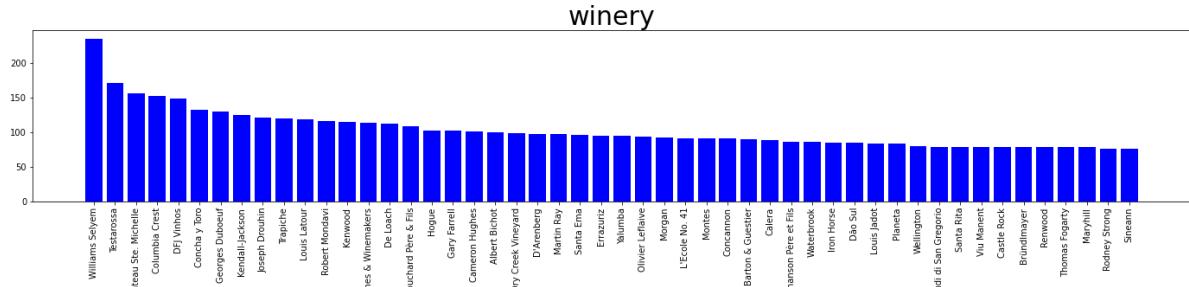
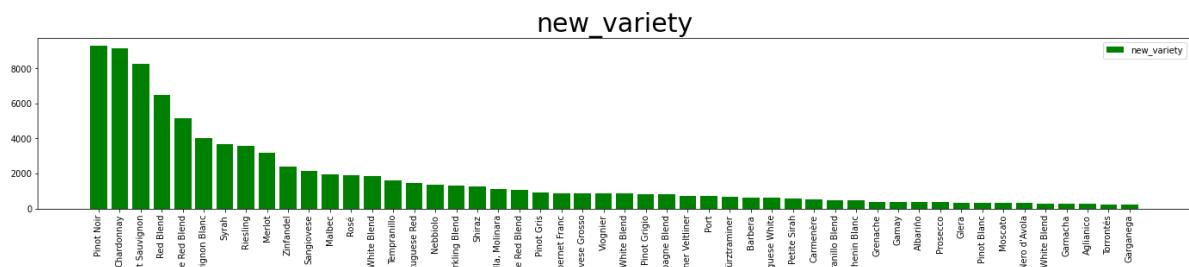
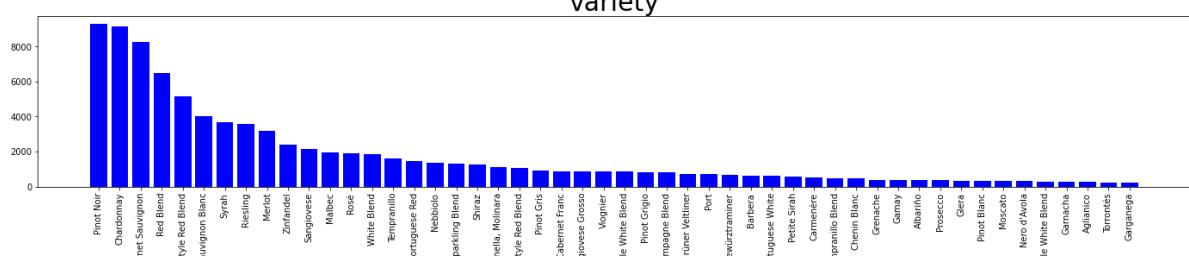
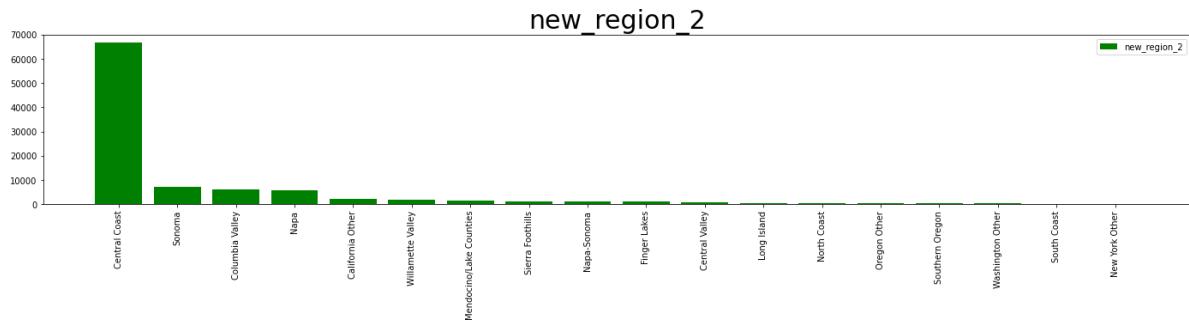
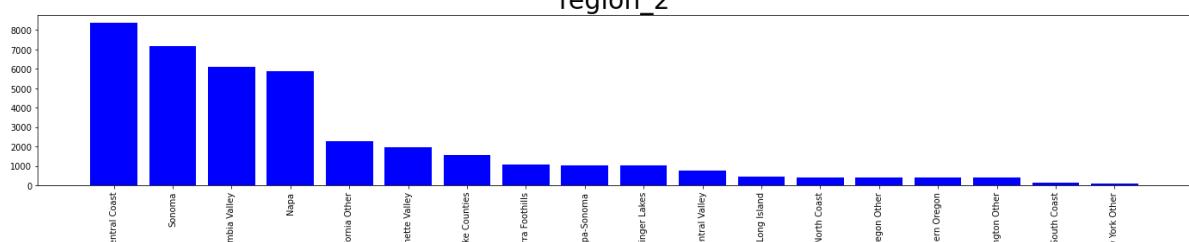
region_1

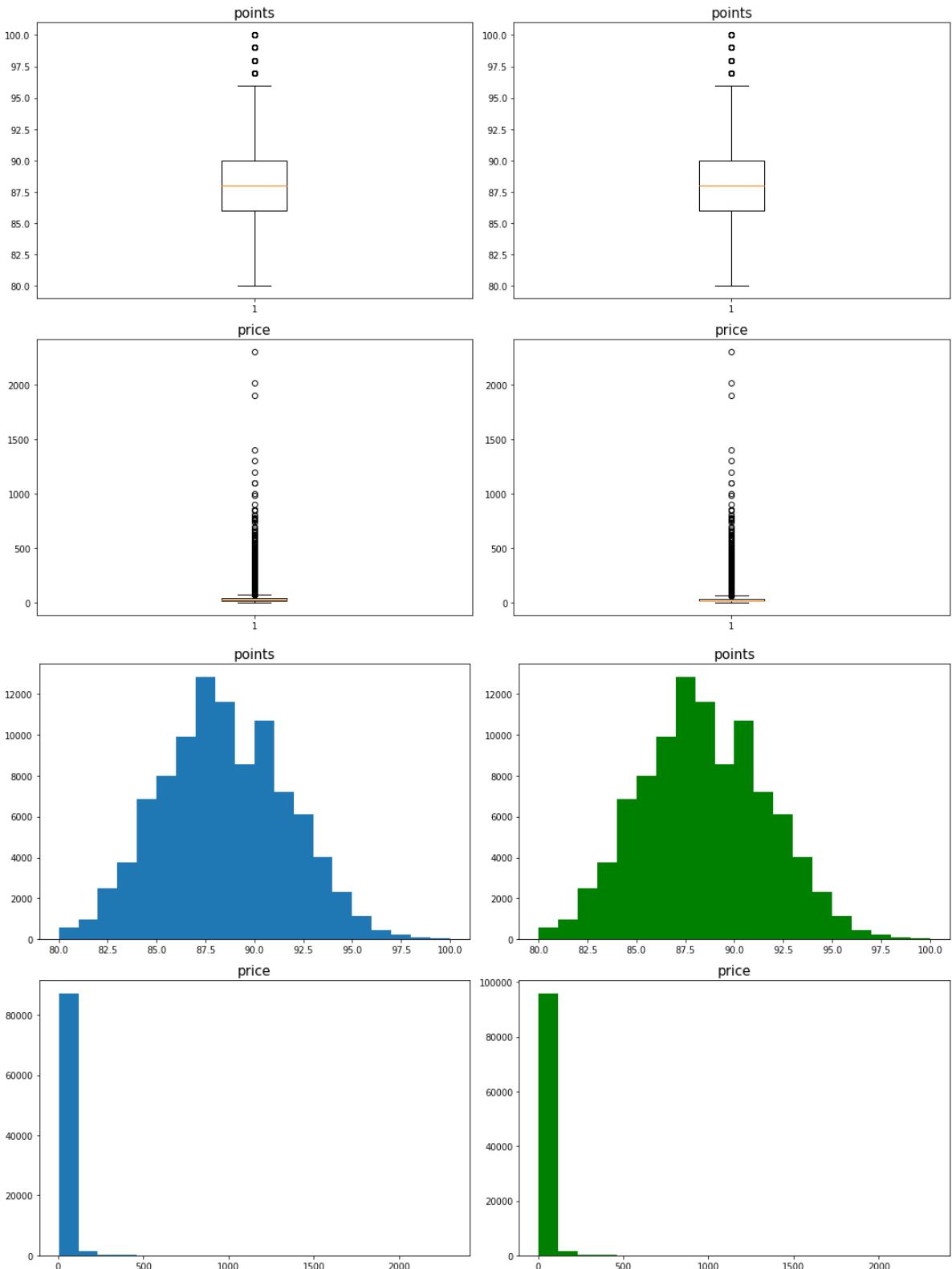


new_region_1



datamining week3





[分析]从图中分析得出本方法会导致country减少至一个，description、designation、regin_1\regin_2出现较大变化，造成明显的数据分布错误，故此方法不可取

5.1.3通过属性的相关关系来填补缺失值

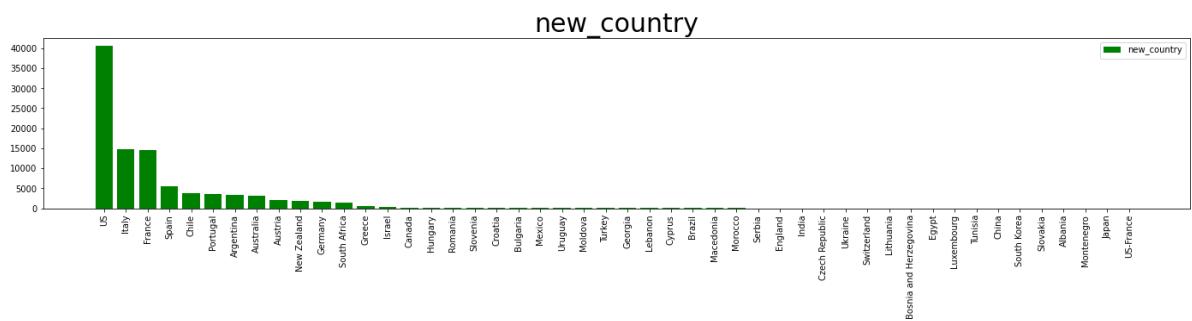
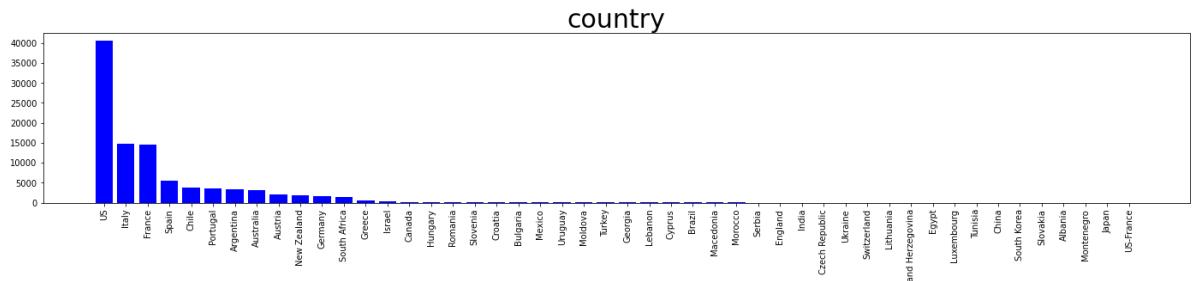
```
In [63]: miss_index = list()
complete_index = list()
col = df_wine1.columns.values
i = 0
for index in list(df_wine1.isnull().any()): #index==True时，对应的列中具有缺失值
    if(index):
        miss_index.append(col[i])
    else:
```

```

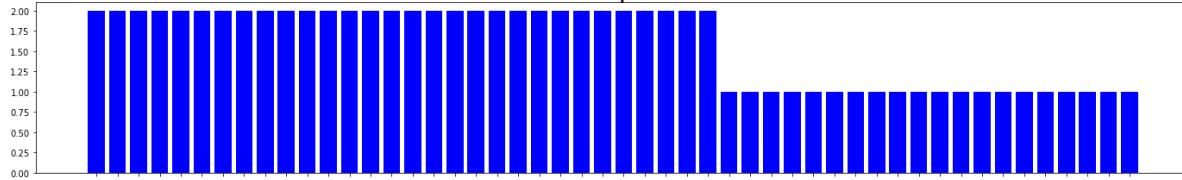
complete_index.append(col[i])
i = i + 1

new_df_wine1 = attr_corr_fill(df_wine1, miss_index, complete_index)
nominal_summary(df=df_wine1, nominal_index=nominal_index_wine1, new_df=new_df_wine1)
box_plot(df_wine1, numerical_index_wine1, new_df_wine1)
hist_plot(df_wine1, numerical_index_wine1, new_df_wine1)

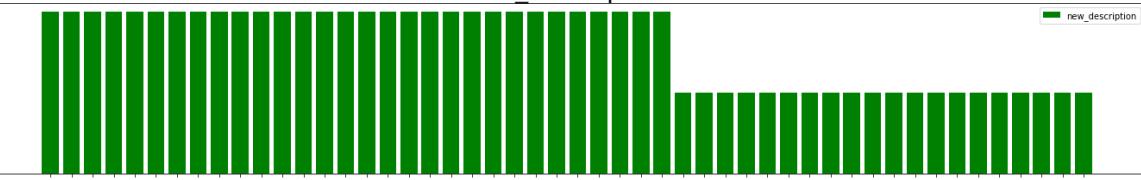
```



description



new_description



Spry, with raspberry and spearmint aromas. Shows cranberry and plum skin on the palate, which is lean and lacking in the texture category. Gets better with airing, but still shows some bitterness on the finish and not much in the middle.

92-94 Barrel sample. A big, concentrated, powerful wine, yet also restrained by elegant fresh fruits. There is plenty of style here, a complex mix of fruits, acidity and balanced tannins.

A little bit tannic and unsettled when you pop the screwcap. But soon it finds its floral, blueberry base. Remains superficial and sweet in the mouth, with candied flavors, vanilla and mid oak. Highly regular, could use more concentration and density.

Giblino, a pure expression of Nero d'Avola, sees some oak aging to shape aromas of toasted walnut, nutella and wood spice. In fact, not much of the natural fruit is left, and what is there feels jammy and ripe.

In 2009 this single vineyard offering includes 11% Semillon. It has wonderful texture, and a mix of light herbal flavors that run through a gamut from citrus fruit to lemongrass and up to tart citrus fruit. The acidity is generous but not searing. The acidity is generous but not searing. A quintessential food wine.

Shows all the hallmarks of an important young Napa Cabernet. Bone dry and quite hard in tannins, it's immensely concentrated in blackberries, black currants, cassis and fine, smoky new oak. Possesses that elusive quality of dramatic terror. A real sense of place. Very fine young wine that needs time. Cellar until at least 2015.

A seductive nose of pineapple, pear, vanilla and buttered toast leaves, followed by a nicely knit, well-fruitful, but not-all-sweet palate. Here, near meadow, pineapple and tangerine is offset by light toast and vanilla accents.

Balanced and smoothly finished. It shows graceful restraint that suggests it will improve over the next year or two.

Powerful, in Zinny character, this blend of Dry Creek and Russian River grapes bursts with briny flavors of onion berries, chocolate-covered mint, tobacco and mint. Shows ripe, wrapped into sturdy tannins. Shows real class and character.

A lean, cool, initially austere wine, well balanced with red berry fruits and a poised elegant structure. The acidity is good, the finish is clean and crisp, with a touch of minerality.

This wine shows some of the less appealing aspects of American oak: sweet spice and vanilla flavors; oily, mouthfeel.

Very ripe and forward in baked fruit or pie-filling flavors of cherries and raspberries, spiced with cinnamon, anise and nutmeg. The wine is a simple and is ready to drink now.

Medium-bodied with good acidity. Finishes correct and everything seems to be united and put together. Above-average value Chardonnay from Argentina.

Simple and candied, with blackberry jam and clove flavors. The tannins are rugged, and the wine finishes with a scowl of acidity. Seems at its best now.

This solid Bordeaux achieves a nice balance between ripe fruit and a more austere earthiness. There's a solid core of blackberry flavor and firm, broad tannins that suggest it will develop more depth with nearer cellar aging.

Asparagus and green herb aromas turn smoothly into grapefruit, passionfruit and sweet green-pepper flavors. More citrus fruit emerges on the tart, cleansing finish.

92-94 Barrel sample. This is a solid and dense wine with hints of bitter chocolate along with superripe blackberry fruits. The wine has concentration, richness and a fine ageing potential.

92-94 Barrel sample. A rounded wine, its tannins submerged into the ripe fruits. It feels soft, and there is just a hint of alcohol. The structure is soft, generous, opulent.

Steer clear of this one if you like your Sauvignons done in stainless. This one is 100% barrel fermented, yielding some toasty, smoky aromas that bend well with the forceful tropical fruit flavors. There's guava and pineapple, smoke and menthol on the long finish.

Very tannic, rough. Aromatics notions of peach, lime and easy-drinking palate with very acidity.

Pleasant enough and Chard-like, with aromas of tangerine, peach and a kiss of honeyed oak. It's got enough residual sugar to almost fall into the off-dry category. Well made and certainly affordable.

This solid Bordeaux achieves a nice balance between ripe fruit and a more austere earthiness. There's a solid core of blackberry flavor and firm, broad tannins that suggest it will develop more depth with nearer cellar aging.

Asparagus and green herb aromas turn smoothly into grapefruit, passionfruit and sweet green-pepper flavors. More citrus fruit emerges on the tart, cleansing finish.

92-94 Barrel sample. This is a solid and dense wine with hints of bitter chocolate along with superripe blackberry fruits. The wine has concentration, richness and a fine ageing potential.

92-94 Barrel sample. A rounded wine, its tannins submerged into the ripe fruits. It feels soft, and there is just a hint of alcohol. The structure is soft, generous, opulent.

The fruit, wild-berry and deeper aromas in the mouth, it's ripe and sweetly fruity, with soft tannins and pepper-spice notes. It's bone dry, with added acidity. Classic Sonoma Zin in its expressive exuberance and unpolished readability.

92-94 Barrel sample. With so much spiciness, that, intensely ripe blackberry flavors and solid tannins, all the elements of excellence are here, and powered by a firm structure. This wine has weight and a fine Cabernet concentration.

If you've ever wondered what makes this wine so special the answer is simple: elegance. This shows polished fruit and spice aromas that are intense, balanced and beautifully integrated. The blend is 52% Cabernet Sauvignon, with the balance made up of Merlot, Cabernet Franc and Petit Verdot. The palate is dense, rich, smooth and supple. Drink this after 2018.

Ripe plum, game, truffle, leather and menthol are some of the aromas you'll find on this ripe wine. The slightly round palate offers dried blackberry, chopped sage, mint and roasted banana alongside racy tannins that leave a mouth-drying finish.

The fruity, perfumed characters show instant fruit, berry aromas, but the old wood flavor doesn't quite chime with the fruit living a slightly muddled fest to the wine.

Dusty cherry and plum aromas are nothing special. The palate is similarly soft OAK, with stony flavors of citrus and red fruits. This is lightly oxidized and pithy tasting on the finish.

Strongly aromatic of black pepper, this blend of Syrah, Zinfandel, Cabernet Sauvignon, Merlot and Cab Franc is herbal like Cab and slightly sour/sweet on the finish.

The bouquet on this blend of Chardonnay, Sauvignon Blanc and Pinot Noir is stuck in neutral, while flavors of citrus and lemon never towards the finish is tingly and tangy.

Mild Lemon and Apple aromas are compromised by a harsh, starchy element. The mouthfeel is basic and spiczy, while generic apple and green melon flavors end in tropical sweeteness.

Following up on a bouquet that's lightly herby and lactic, the palate feels hard and tannic, with oak-based resin striking out. Flavors of black pepper, licorice and sage push away the wine's black-fruit component, leaving a minty finish that's oak dominant and mostly artificial tasting.

This is herbal and tonikatory on the nose, then dry in feel, with little mid-sweetness on the palate. Flavors of raspberry are pale and offset by mild berry oak notes. This wine shows some elements of classic Spain but weakness as well.

Maderized in color, this wine has a yeasty, creamy nose with baked white-fruit aromas, and caramel. It's Ok in feel, with picked, mildly briny flavors of apple and apricot. The finish is showing some oxidation, leading to a chunky, fizzy feel.

This Chardonnay is slightly metallic and briny on the bouquet. It feels plump and chunky, with citrus and pineapple flavors coming in front of a quiet finish.

Mint aromas speak of the wood used in aging this wine. Minty tannins are nothing special. The palate is similarly soft OAK, with stony flavors of citrus and red fruits. This is lightly oxidized and pithy tasting on the finish.

A sweet note to start, then this turns into a plummy, tannic, aromatic red with a hint of barrel roast.

The nose on this Viura is full, with white flowers suggesting dry grape, apple and nutty notes. This is dry and modest on the finish.

This is a funny nose that's unusual. The palate is a bit flabby, with sweet banana and peaches flavors. There's not much to the finish and overall it's come across as overly simple and strained.

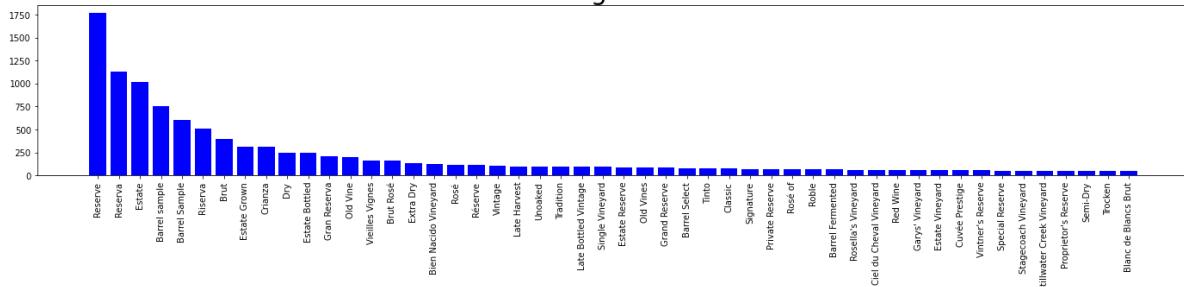
This Chardonnay comes of Sardinia's most widely planted indigenous grapes—Carrasino, Cargnolo and Bonaiu. Syrah is bold, round and wine with a long, smooth, velvety texture. On the nose, the wine presents black cherry, spice and mid-toasted oak notes.

Dripping with notes of juicy black fruit, sweet tobacco and clove, this is layered and full-bodied, with soft tannins and an intensity that should grow more intriguing with age. From the gorgeous Culveras County Dalton Vineyard, this is blended with 11% Petite Sirah, also from Dalton.

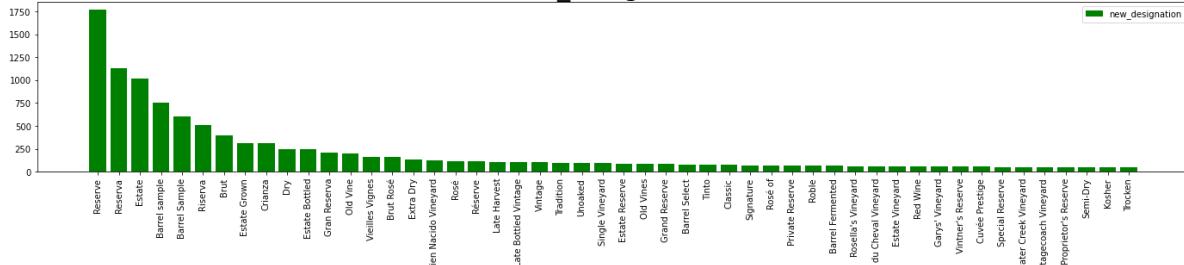
From beautiful vineyards on the Vulcano volcano in Southern Italy, Durini opens with savory aromas of leather and cured meat, followed by dried berries and prune. Polished tannins and tones of lead pencil and smoke indicate the wine is aging nicely.

This is a big, fruit-driven style of Shiraz, with great aromas and flavors of camphor, vanilla, dark toast and cassis. There's just enough structure to give it focus, although it lacks a bit of textural richness on the midpalate. Drink this supple wine over the next few years.

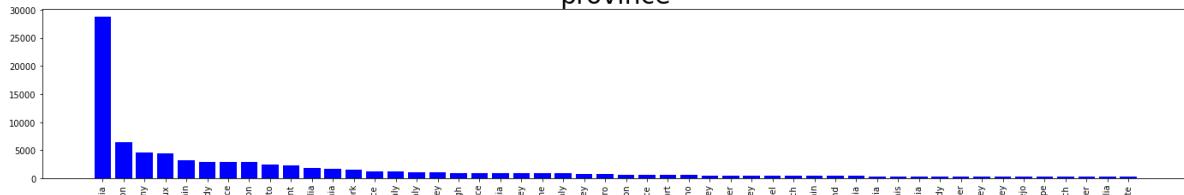
designation



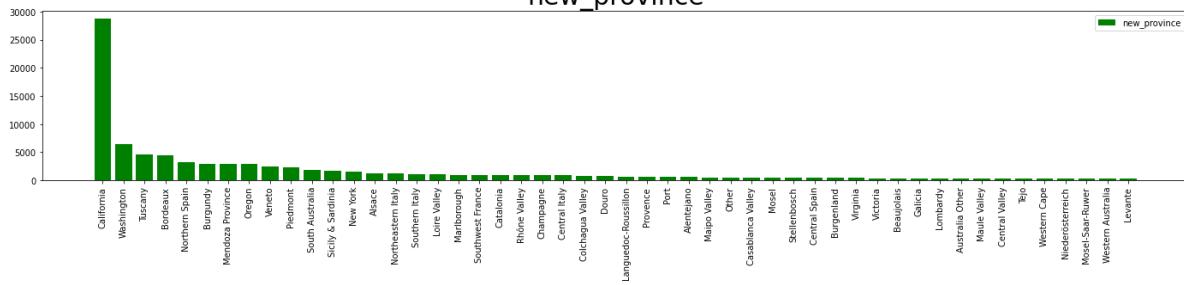
new_designation



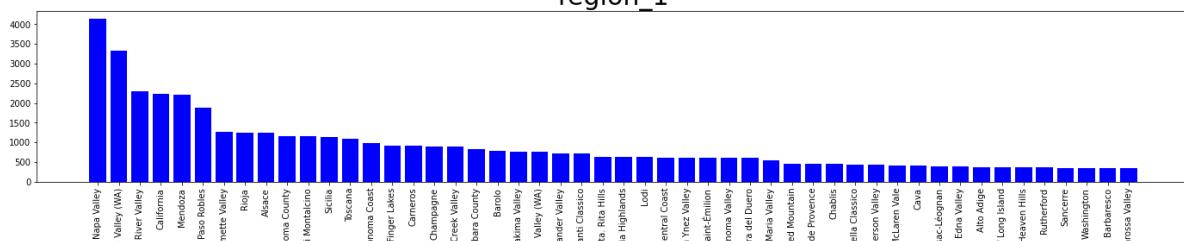
province



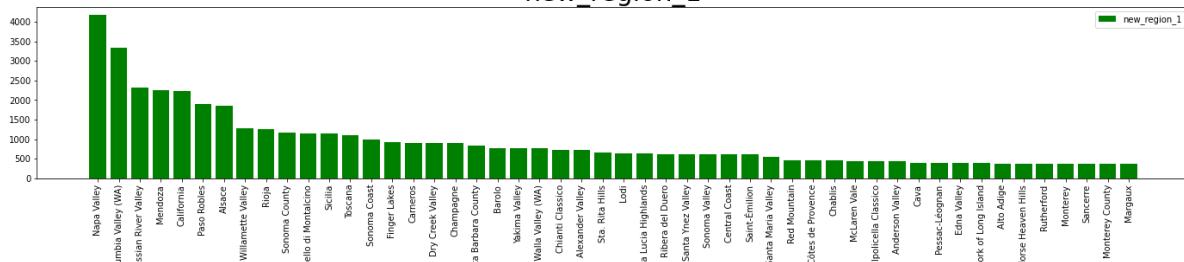
new_province



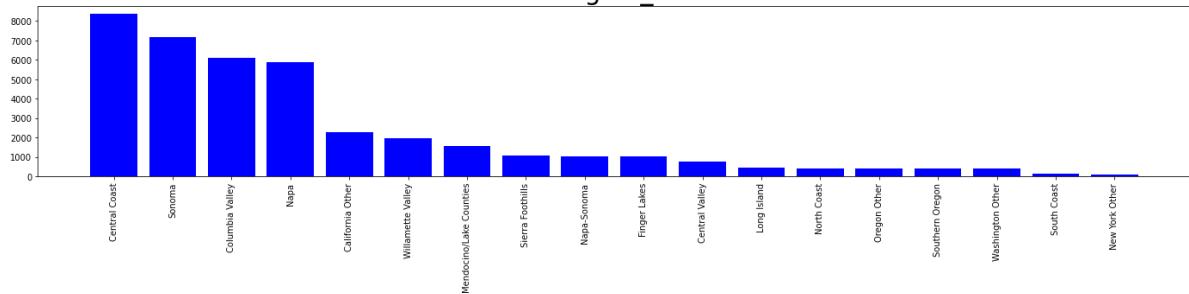
region_1



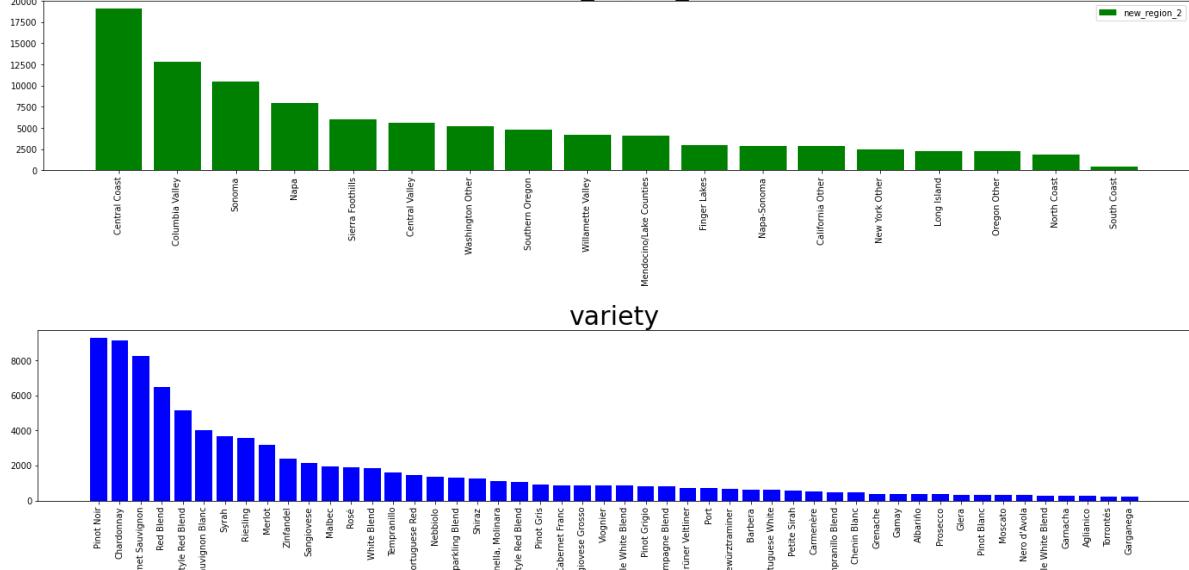
new_region_1



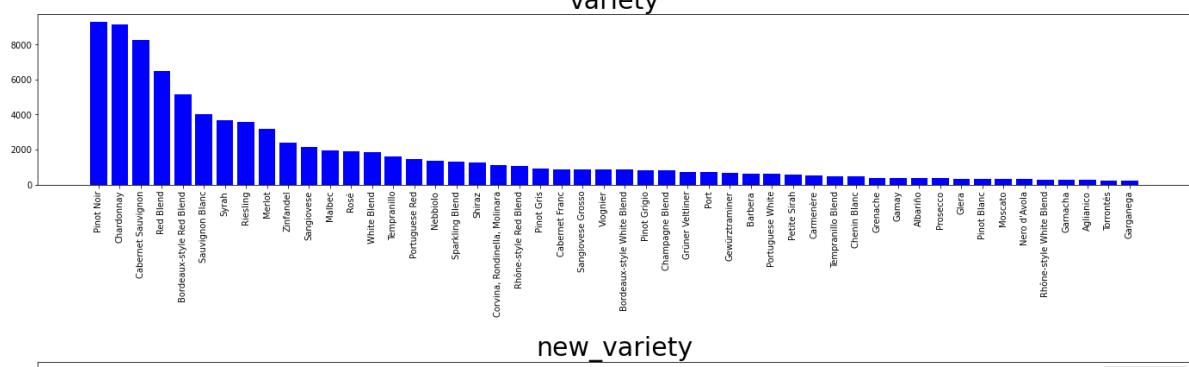
region_2



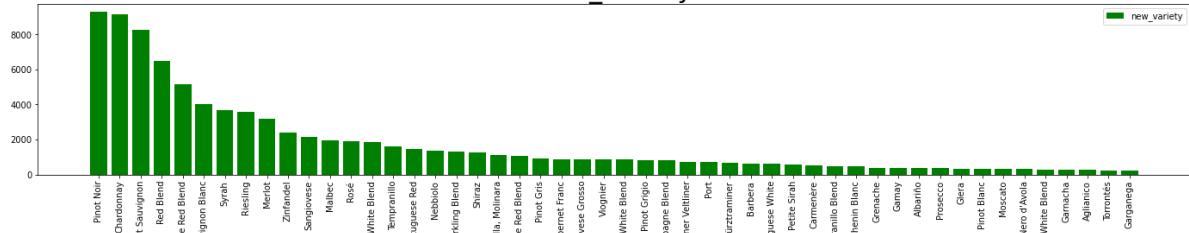
new_region_2



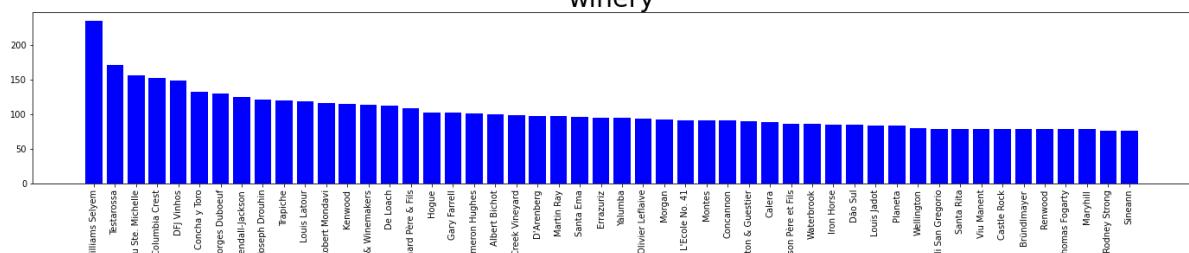
variety



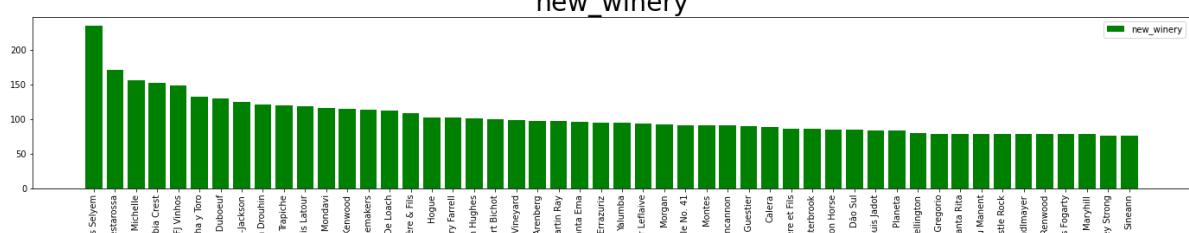
new_variety

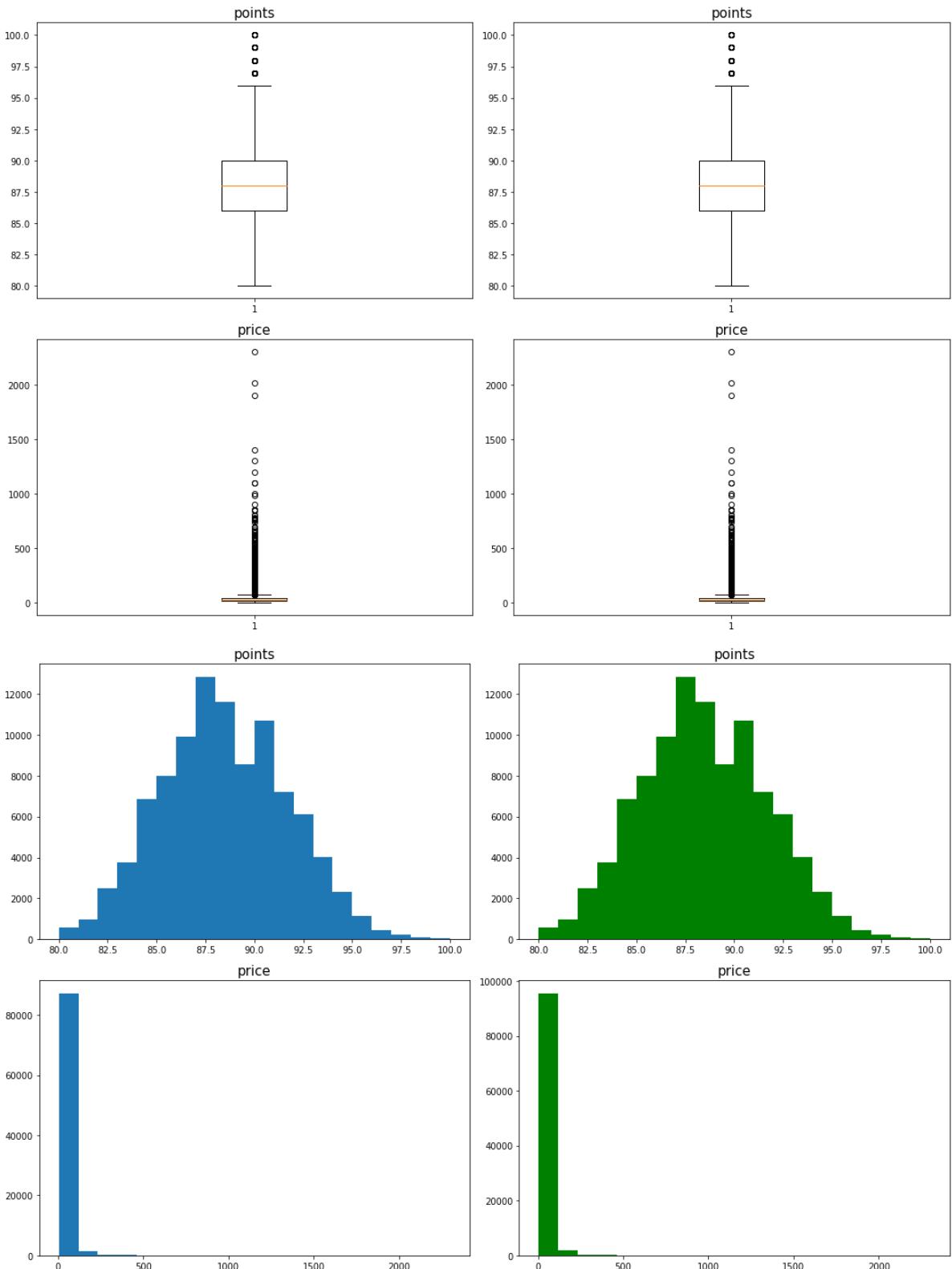


winery



new_winery





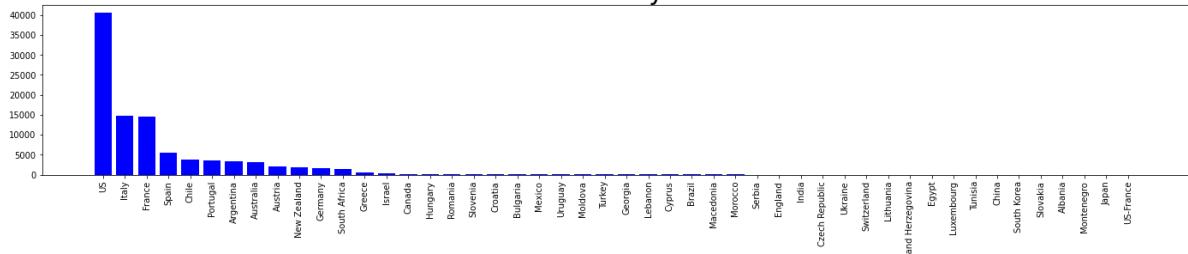
[分析]预测后的数据分布与原始数据差距不大，由此可知本方法预测合理

5.1.4 通过数据对象之间的相似性来填补缺失值

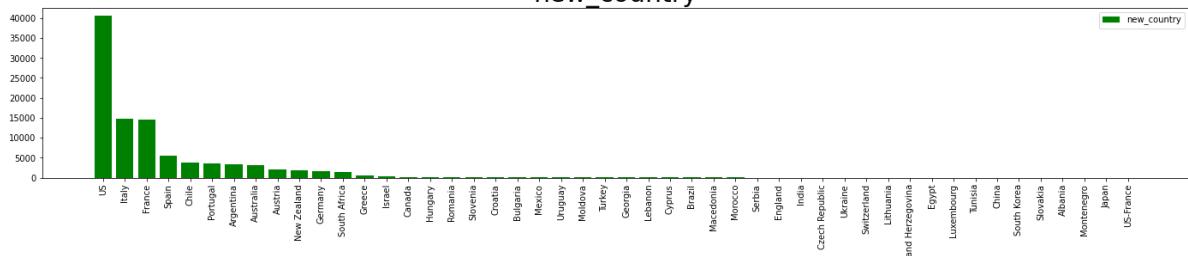
```
In [65]: new_df_wine1 = obj_corr_fill(df=df_wine1, numerical_index=numerical_index_wine1, k=1)
nominal_summary(df_wine1, nominal_index_wine1, new_df_wine1)
box_plot(df_wine1, numerical_index_wine1, new_df_wine1)
hist_plot(df_wine1, numerical_index_wine1, new_df_wine1)
```

```
E:\app\miniconda3\lib\site-packages\impute\imputation\cs\fast_knn.py:113: RuntimeWarning: invalid value encountered in true_divide
    weights = distances/np.sum(distances)
```

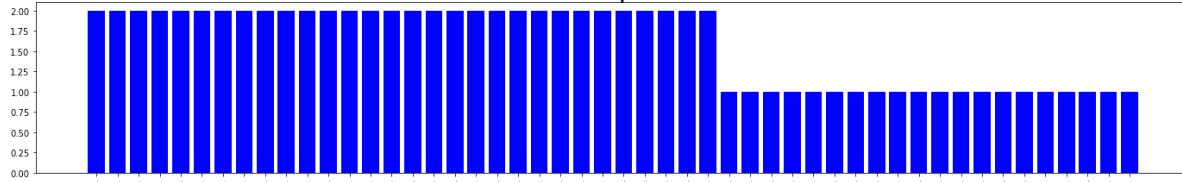
country



new_country



description



Spiry, with raspberry and spearment aromas. Shows cranberry and plum skin on the palate, which is lean and lacking in the texture category. Gets better with airing, but still shows some bitterness on the finish and not much in the middle.

A little bit funky and unsettled when you pop the screwcap, but soon it finds its floral, blueberry base. Remains superficial and sweet in the mouth, with candied flavors, vanilla and mid oak. Highly regular, could use more concentration and density.

Gabinoro, a pure expression of Nero d'Avola, sees some oak aging to shape aromas of toasted walnut, nutella and wood spice. In fact, not much of the natural fruit is left, and what is there feels jammy and ripe.

In 2009 this single vineyard offering includes 11% Semillon. It has wonderful texture, and a mix of light herbal flavors that run through a gamut from citrus to lemongrass and up into tart citrus fruit. The acidity is generous but not searing, it penetrates but never overpowers. A quintessential food wine.

Show all the hallmarks of an important young Napa Cabernet. Bone dry and quite hard in tannins, it's immensely concentrated in blackberries, black currants, cassis and fine, smoky new oak. Possesses that elusive quality of dramatic terror, a real sense of place. Very fine young wine that needs time. Cellar until at least 2015.

A seductive nose of pineapple, peat, vanilla and buttered toast leads, followed by nicely knit, well-fruitful, but not-all-sweet palate. Here, heat, medium, pineapple and tannins is offset by light toast and vanilla accents. Balanced and smoothly rounded. It shows graceful restraint that suggests it will improve over the next year or two.

Powful in Zinny character, this blend of Dry Creek and Russian River grapes bursts with briny flavors of wild berries, chocolate-covered mint, tobacco and pepper, wrapped into sturdy tannins. Shows real class and character.

Sauvignon on the nose with modest apple and pear aromas accented by a curious smoked ham note. This smoky, unbalanced complexity, that could make a interesting companion to prosciutto wrapped around medallions of melon, and fragrant, citrusy melon.

A lean, cool, initially austere wine, well balanced with red berry fruits and a poised elegant structure. The acidity gives great freshness to go with the structure, making a great food wine.

This wine shows some of the less appealing aspects of American oak: overt spice and vanilla flavors; and a sick, oily mouthfeel.

Very ripe and forward in baked fruit or pie-filling flavors of cherries and raspberries, spiced with cinnamon, anise, and nutmeg. The wine is bit simple and is ready to drink now.

Sharp, simple and candied, with blackberry jam and cola flavors. The tannins are rugged, and the wine finishes with a scowl of acidity. Seems at its best now.

This solid Bordeaux achieves a nice balance between ripe fruit and a more austere earthiness. There's a solid core of blackberry flavor and firm, broad tannins that suggest it will develop more depth with nearer maturing.

Asparagus and green herb aromas run smoothly into grapefruit and sweet green pea flavors. More citrus fruits emerge on the tart, cleansing finish.

92.90 Barrel sample. This is a solid and dense wine with hints of bitter chocolate along with superripe blackberry fruits. The wine has concentration, richness and a fine aging potential.

92.90 Barrel sample. A rounded wine, its tannins submerged into the ripe fruits. It feels soft, and there's just a late of alcohol. The structure is soft, gentle, opulent.

Steer clear of this one if you like your Sauvignons done in stainless. This one is 100% barrel fermented, yielding some toasty, smoky aromas that blend well with the forceful tropical fruit flavors. There's guava and pineapple, tangelo and menthol on the long finish.

Very tannic, rough. Atomatic notions of peach precede a fresh, light and easy-drinking palate with tarty acidity.

Pleasant enough and Chard-like, with aromas of tangerine, peach and a kiss of honeyed oak. It's got enough residual sugar to almost fall into the off-dry category. Well made and certainly affordable.

This solid Bordeaux achieves a nice balance between ripe fruit and a more austere earthiness. There's a solid core of blackberry flavor and firm, broad tannins that suggest it will develop more depth with nearer maturing.

Asparagus and green herb aromas run smoothly into grapefruit and sweet green pea flavors. More citrus fruits emerge on the tart, cleansing finish.

92.90 Barrel sample. This is a solid and dense wine with hints of bitter chocolate along with superripe blackberry fruits. The wine has concentration, richness and a fine aging potential.

92.90 Barrel sample. A rounded wine, its tannins submerged into the ripe fruits. It feels soft, and there's just a late of alcohol. The structure is soft, gentle, opulent.

The fruity, perfumed character shows instant red berry aromas, but the old wood flavor doesn't jive with the fruit, giving a slightly muddled feel to the wine.

Dusty cherry and plum aromas are nothing special. The palate is similarly just OK, with stony flavors of citrus and fruits. This is lightly oxidized and pretty tasting on the finish.

Strongly aromatic of black pepper, this blend of Syrah, Zinfandel, Cabernet Sauvignon, Merlot and Cab Franc is herby like. Cab and slightly sour/sweet on the finish.

The bouquet on this blend of Chardonnay, Sauvignon Blanc and Pinot Noir is stuck in neutral, while flavors of citrus and lemon never towards sour. The finish is tannic and tangy.

Mild lemon and apricot aromas are compromised by a harsh, stony element. The mouthfeel is basic and spritz, while generic apple and green melon flavors end in tropical sweetness.

Mild lemon and apricot aromas are compromised by a harsh, stony element. The mouthfeel is basic and spritz, while generic apple and green melon flavors end in tropical sweetness.

Following up on a bouquet that's lightly herby and lactic, the palate feels hard and tannic, with oak-based resin sticking out. Flavors of black pepper, licorice and sage pack away the wine's black-fruit component, leaving a minty finish that's oak-dominated and mildly artifical tasting.

This is herby and tonikoly on the nose, then acy in feel, with little midlevel stuffing on the palate. Flavors of raspberry are acute and offset by mild buttery oak notes. This wine shows some elements of classic Spain but weaknesses as well.

Maderized in color, this wine has a yeasty, creamy nose with baked white-fruit aromas and caramel. It's OK in feel, with pickled, milky briny flavors of apple and apricot. The finish is showing some oxidation, leading to a chunky, fizzy feel.

This Chardonnay is slightly metallic and briny on the bouquet. It feels plump and chunky, with citrus and pineapple flavors coming in front of a quiet finish.

Mint aromas speak of the wood used in aging this wine. With some burnt tannin character, it's perhaps not a wine that will show great fruit.

A sweetly nose to start, then this turns into a plummy, tannic, astrigent red with a hint of barrel toast.

The nose on this Viura is flat, with walnut oil and dusty apple notes. The mouthfeel is solid, while the flavor's suggest dry apricot, apple, and nutty notes. This is dry and moderate on the finish.

This is a pricky, slightly racy wine with a tangy, briny feel and briny pickled flavors of herald red berry fruits. It's ready tasting on the finish, with a snappy feel and lasting herbal flavors.

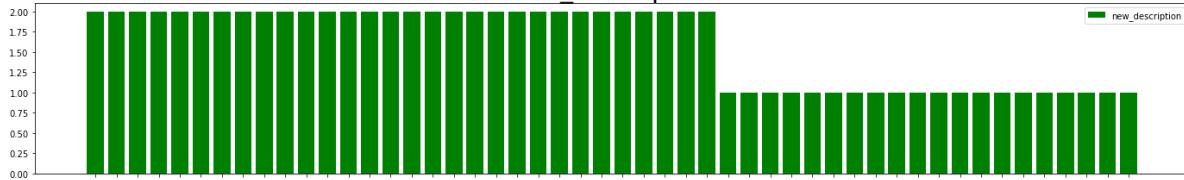
This Cava has a funky nose that's unusual. The palate's a bit flabby, with sweet banana and peaches flavor. There's not much to the finish and overall this comes across as overly simple and strained.

Dripping with notes of juicy black fruit, sweet tobacco and clove, this is layered and full-bodied, with soft tannins and an intensity that should grow more intriguing with age. From the prestigious Calveras County Dalton Vineyard, this is blends with 21% Peter Sraha, also from Dalton.

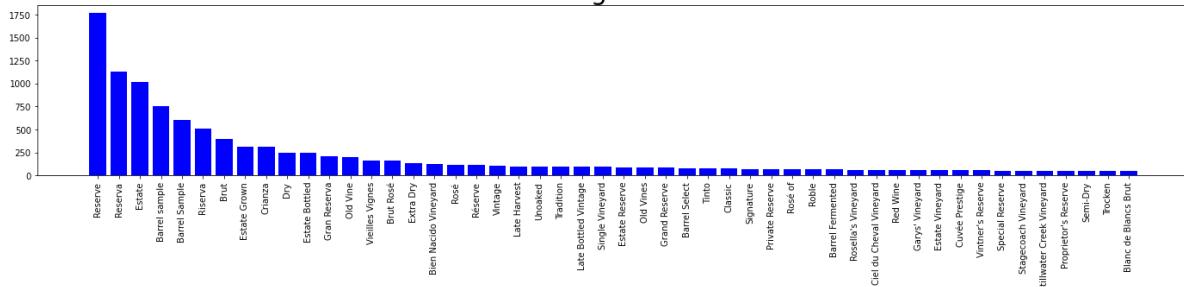
From beautiful vineyards on the Vulcano volcano in Southern Italy, Dianus opens with savory aromas of feather and cured meat, followed by dried berries and prune. Polished tannins and tones of lead pencil and smoke indicate the wine is aging nicely.

This is a big, fruit-driven style of Shiraz, with great aromas and flavors of camphor, vanilla, dark toast and cassis. There's just enough structure to give it focus, although it lacks a bit of textural richness on the midpalate. Drink this supple wine over the next few years.

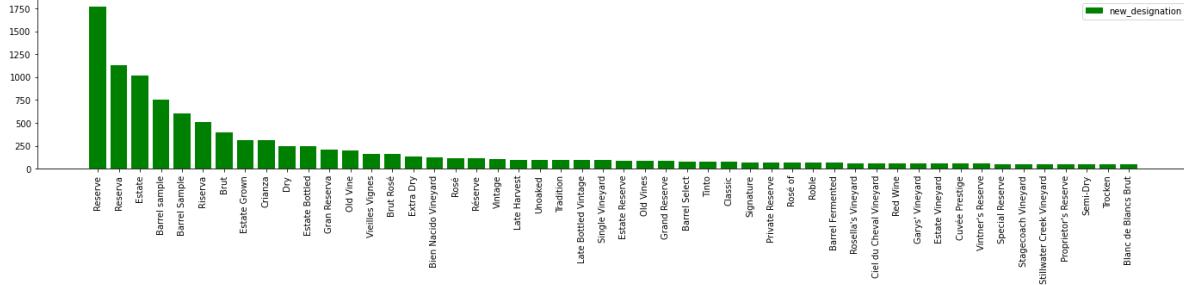
new_description



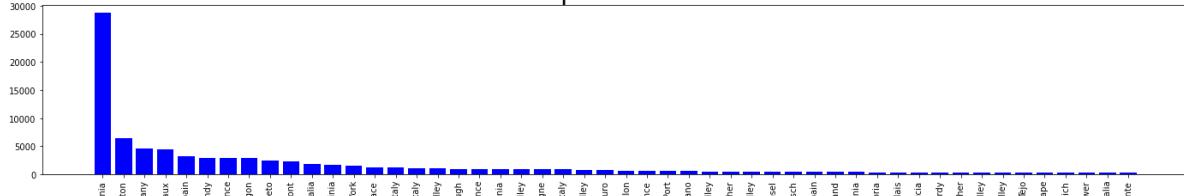
designation



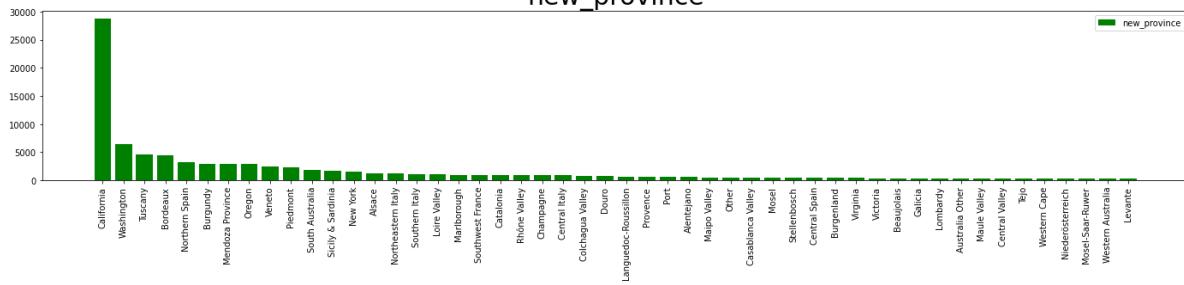
new_designation



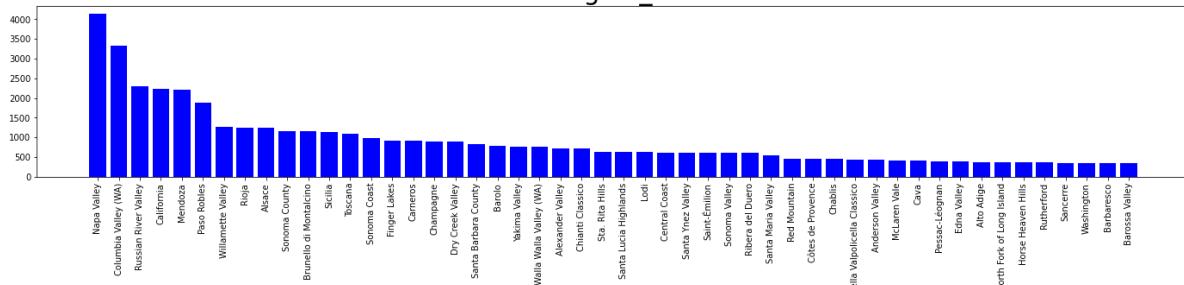
province



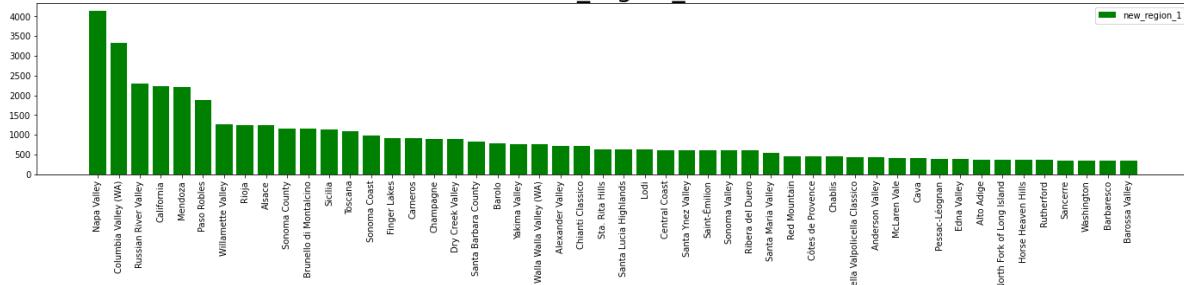
new_province



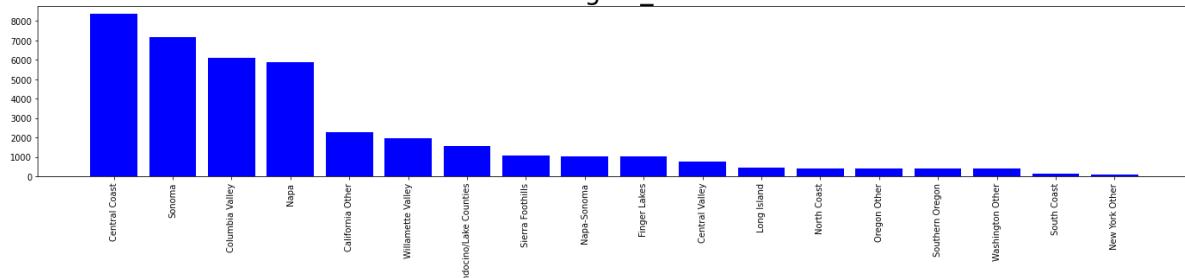
region_1



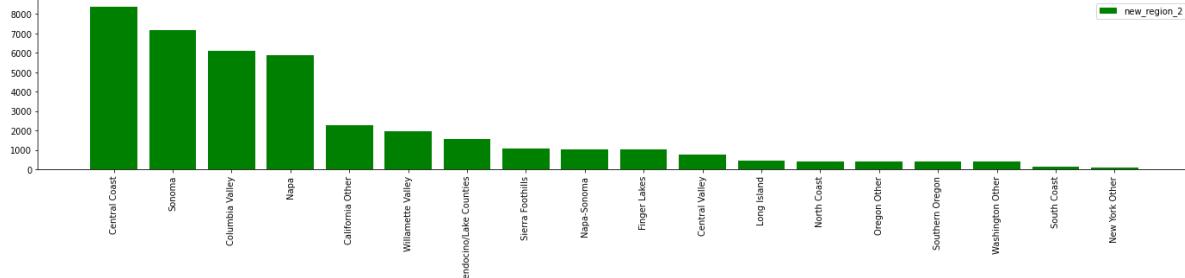
new_region_1



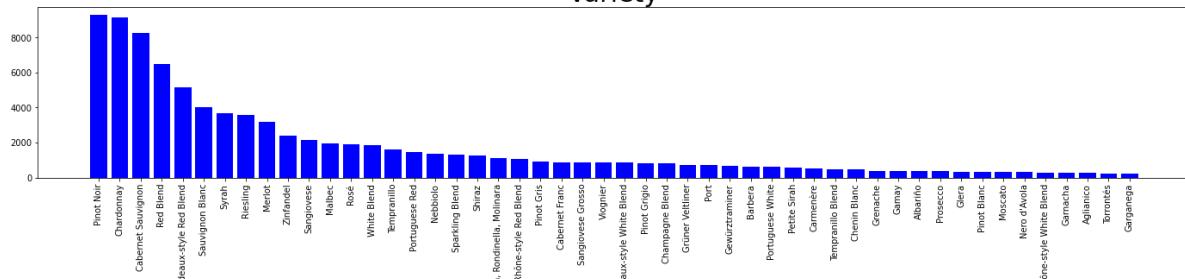
region_2



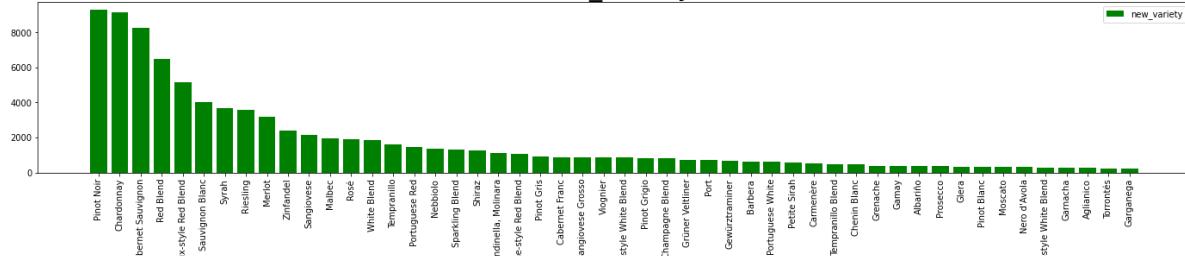
new_region_2



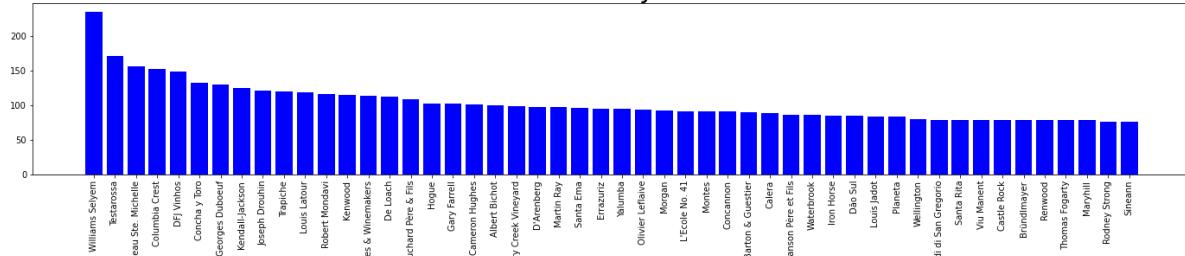
variety



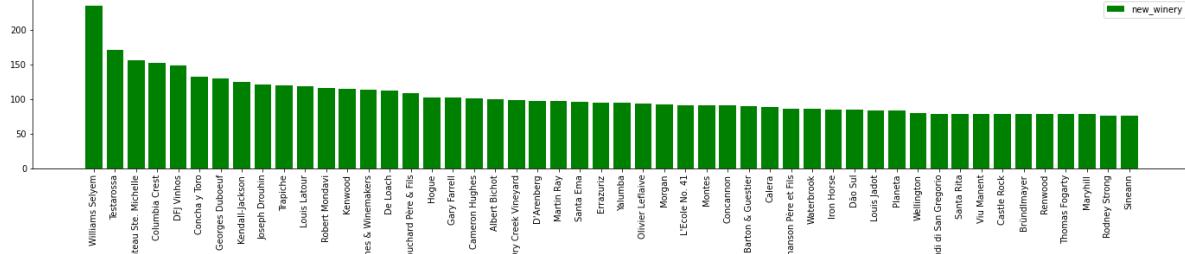
new_variety

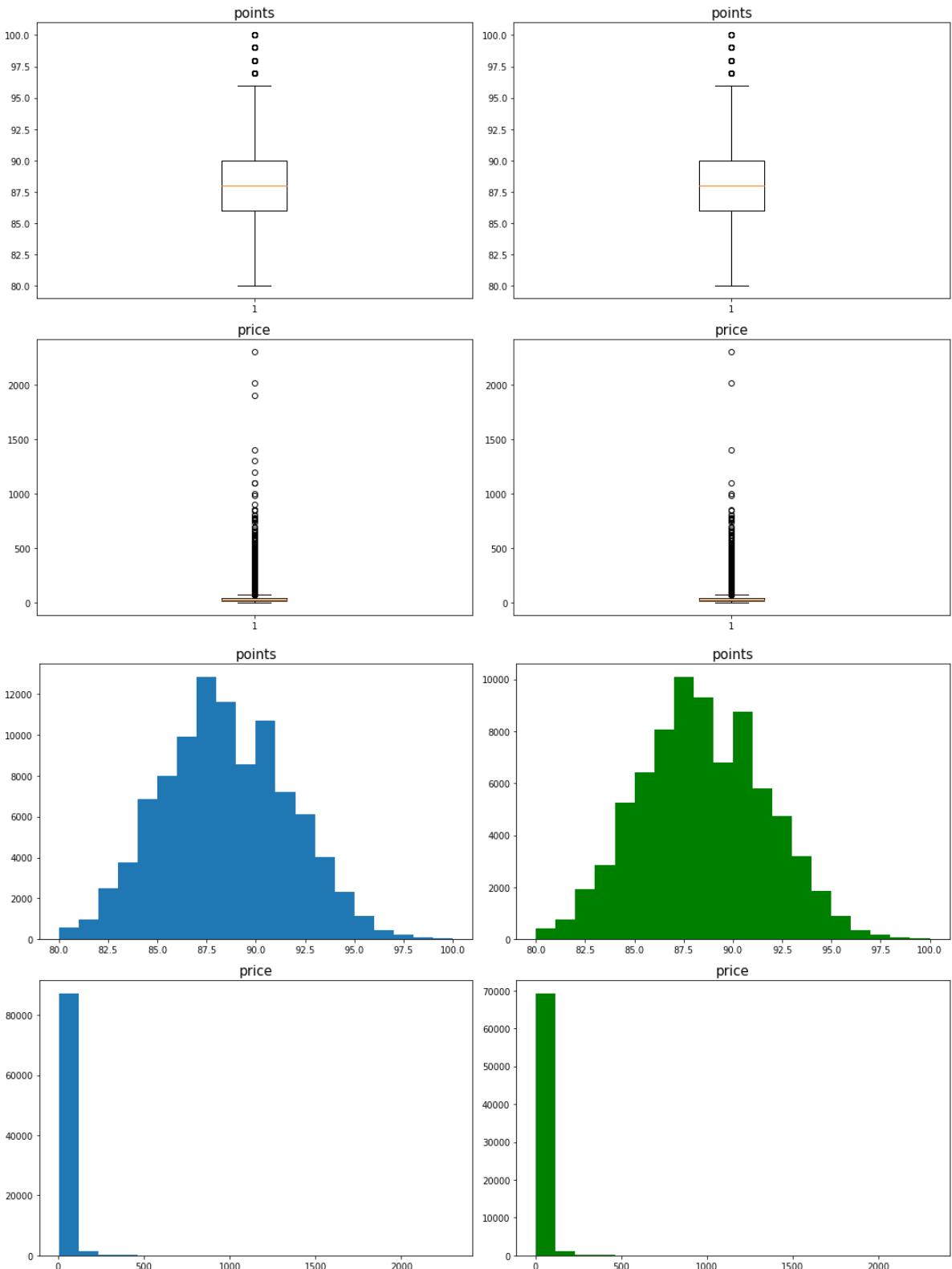


winery



new_winery





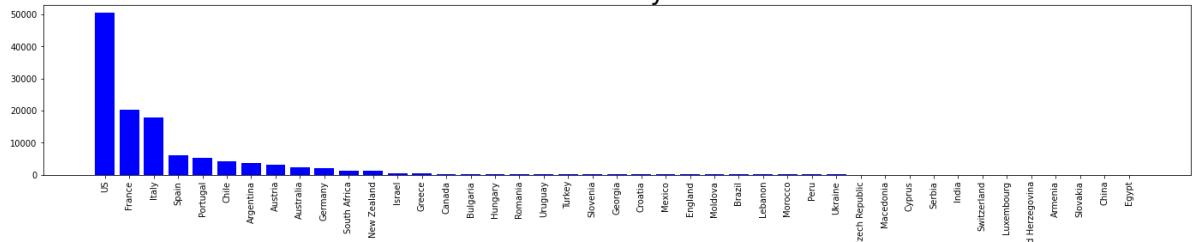
[分析]预测后的数据分布与原始数据差距不大，由此可知本方法预测合理

5.2 针对数据集winemag-data-130k-v2.csv

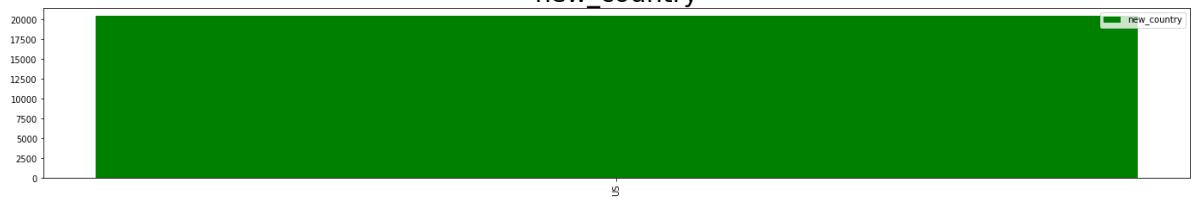
5.2.1 将缺失部分删除

```
In [66]: new_df_wine2 = miss_drop(df_wine2)
nominal_summary(df_wine2,nominal_index_wine2,new_df_wine2)
box_plot(df_wine2,numerical_index_wine1,new_df_wine2)
hist_plot(df_wine2,numerical_index_wine1,new_df_wine2)
```

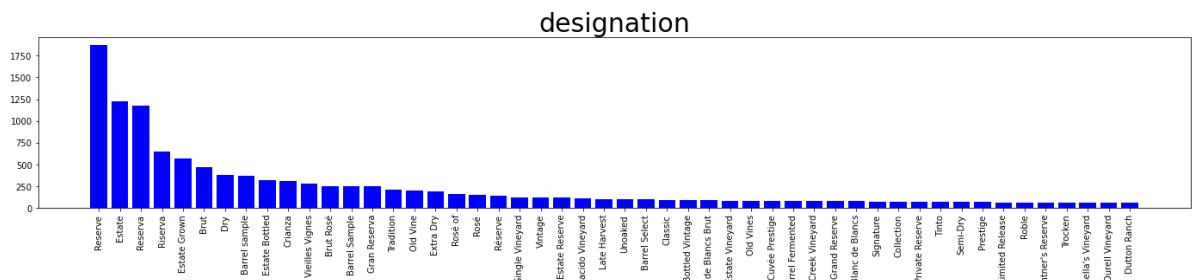
country



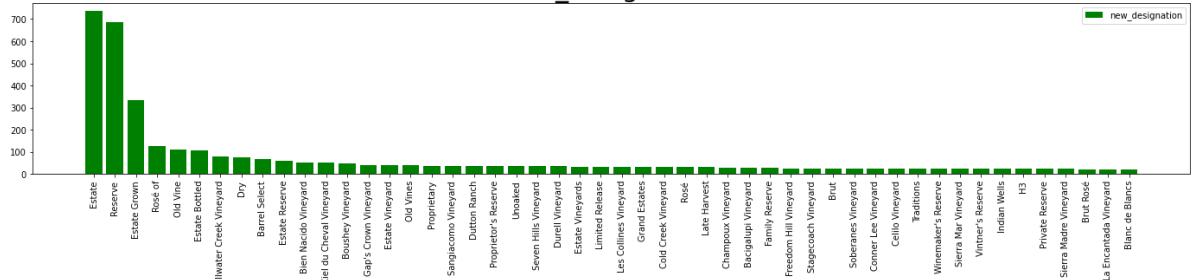
new_country



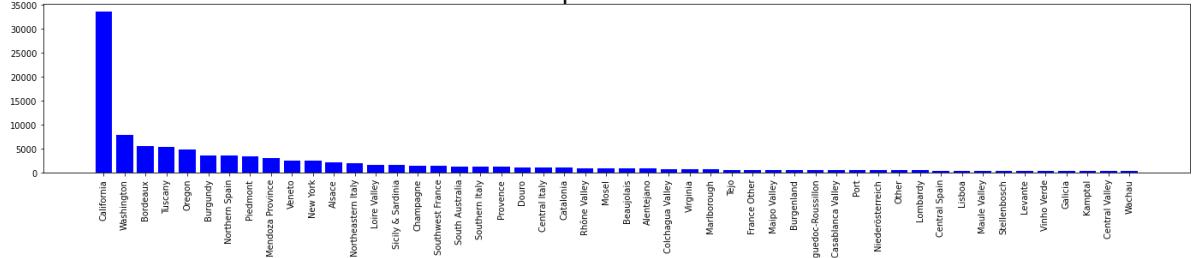
This is just this side of uncouth, a decadent tip and rich late harvest Riesling. The residual sugar is at 23%, the alcohol a moderate 12.5%.



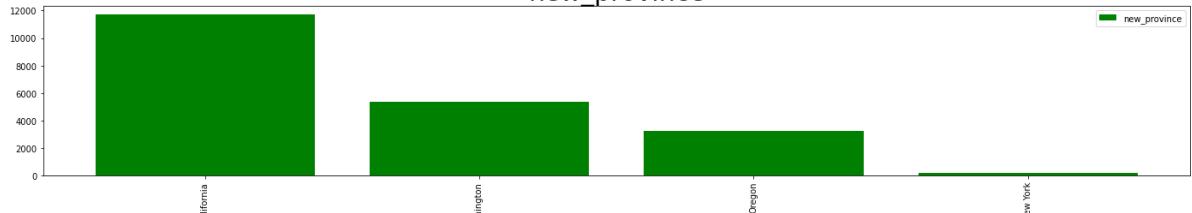
new_designation



province

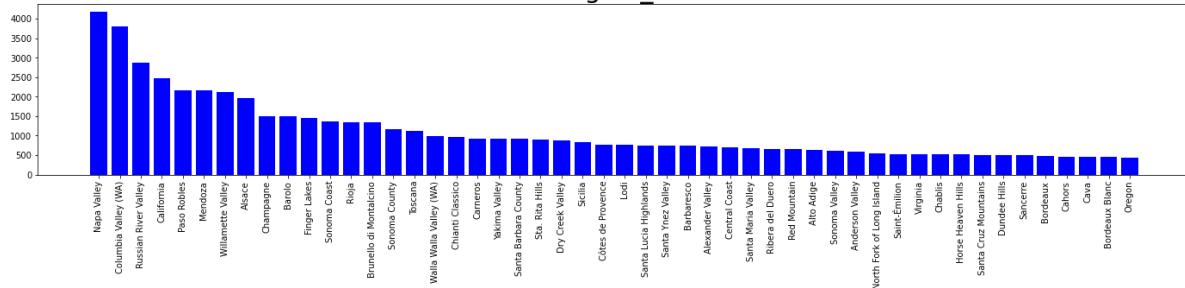


new_province

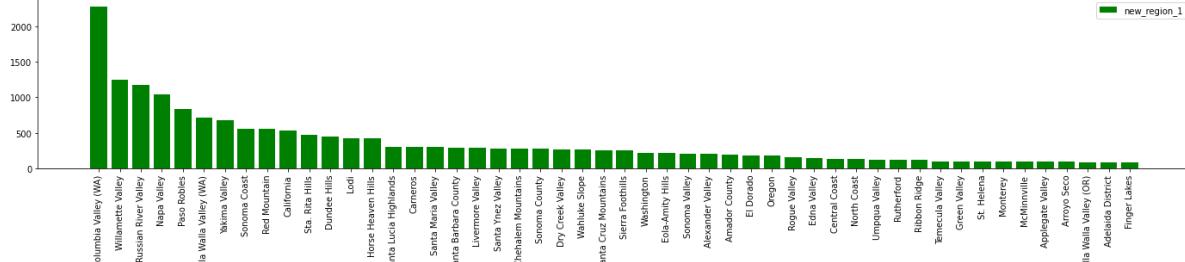


datamining week3

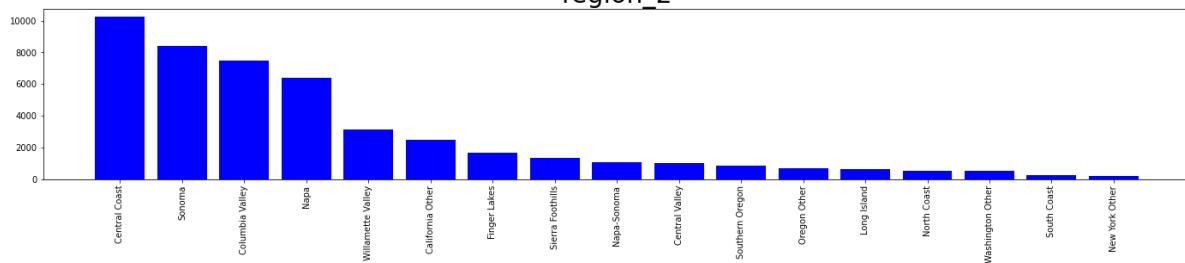
region_1



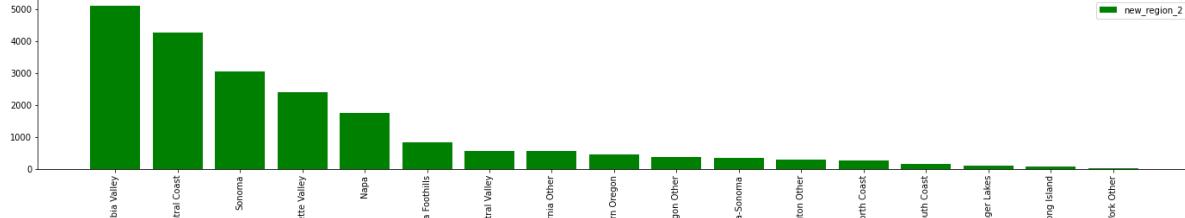
new_region_1



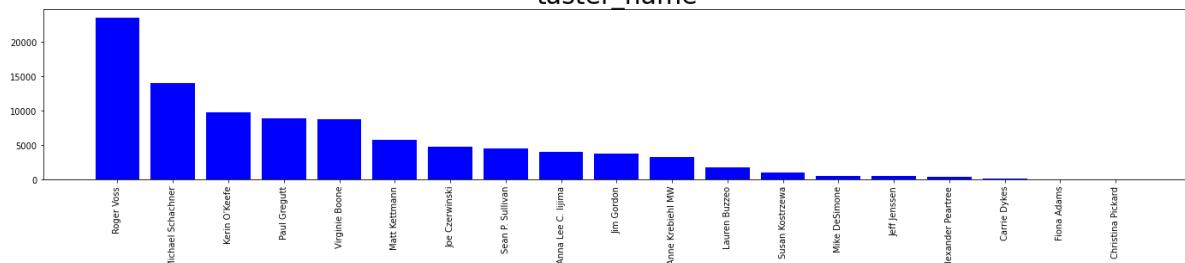
region 2



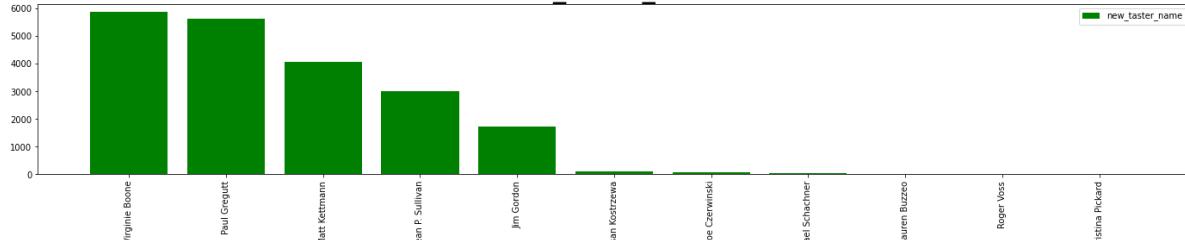
new_region_2



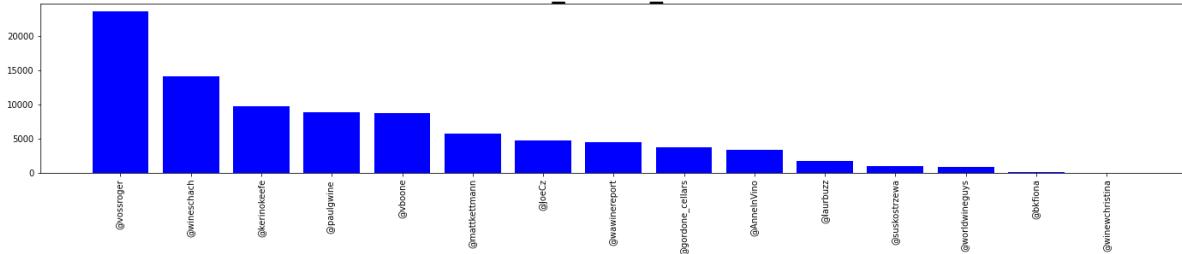
taster name



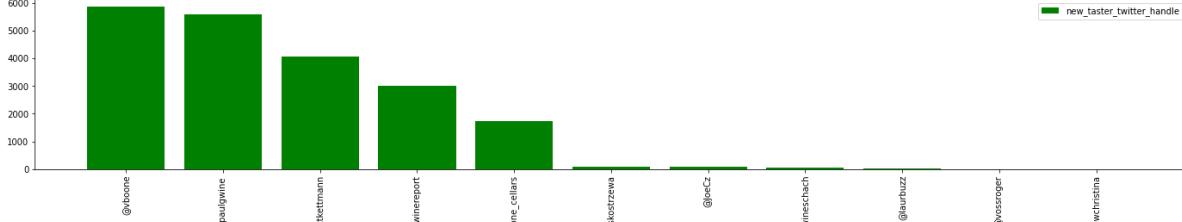
new_taster_name



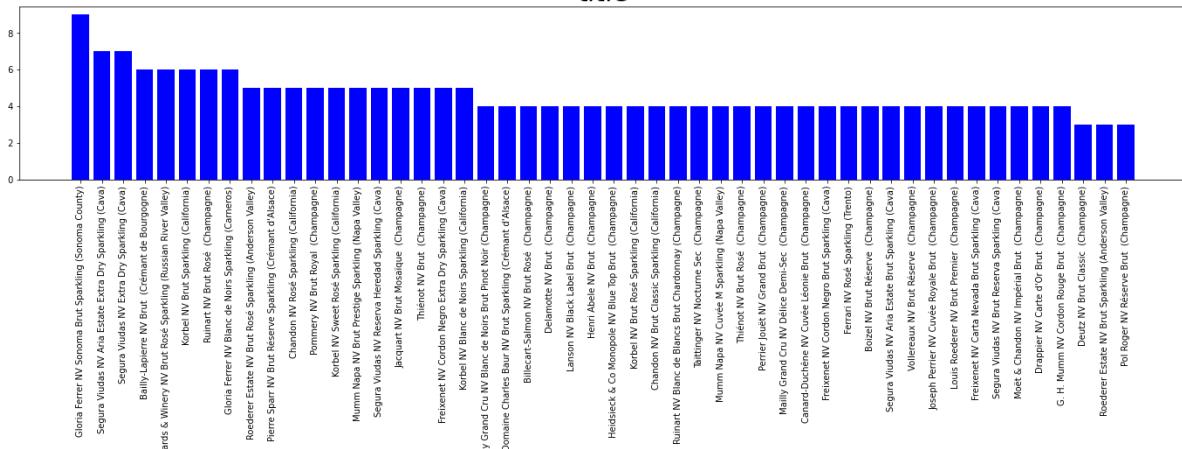
taster_twitter_handle



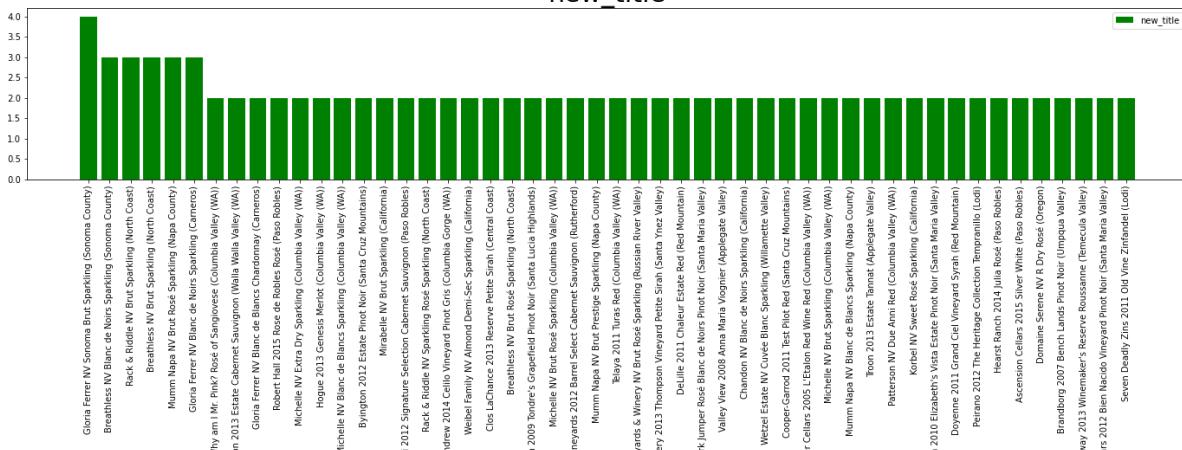
new_taster_twitter_handle

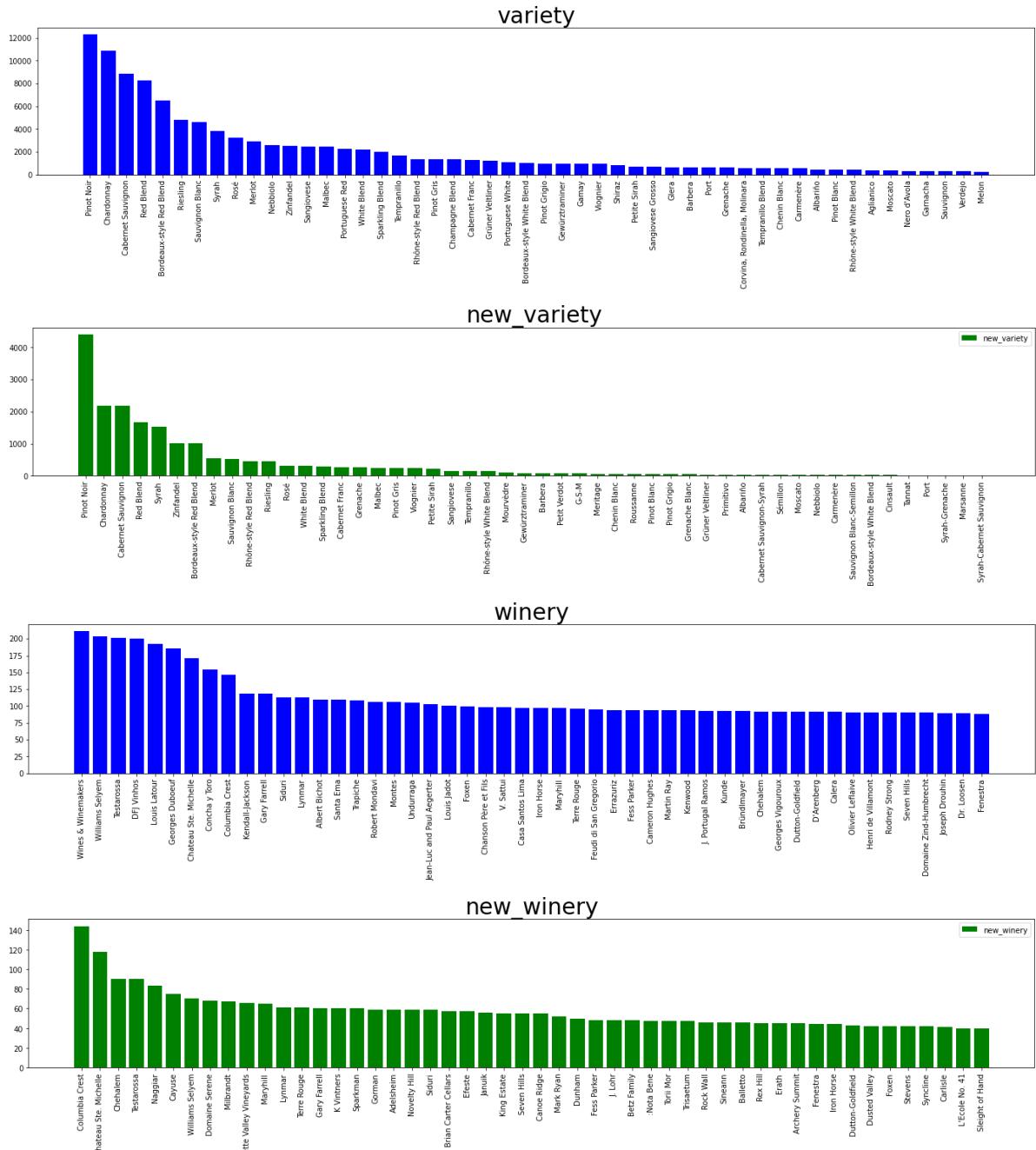


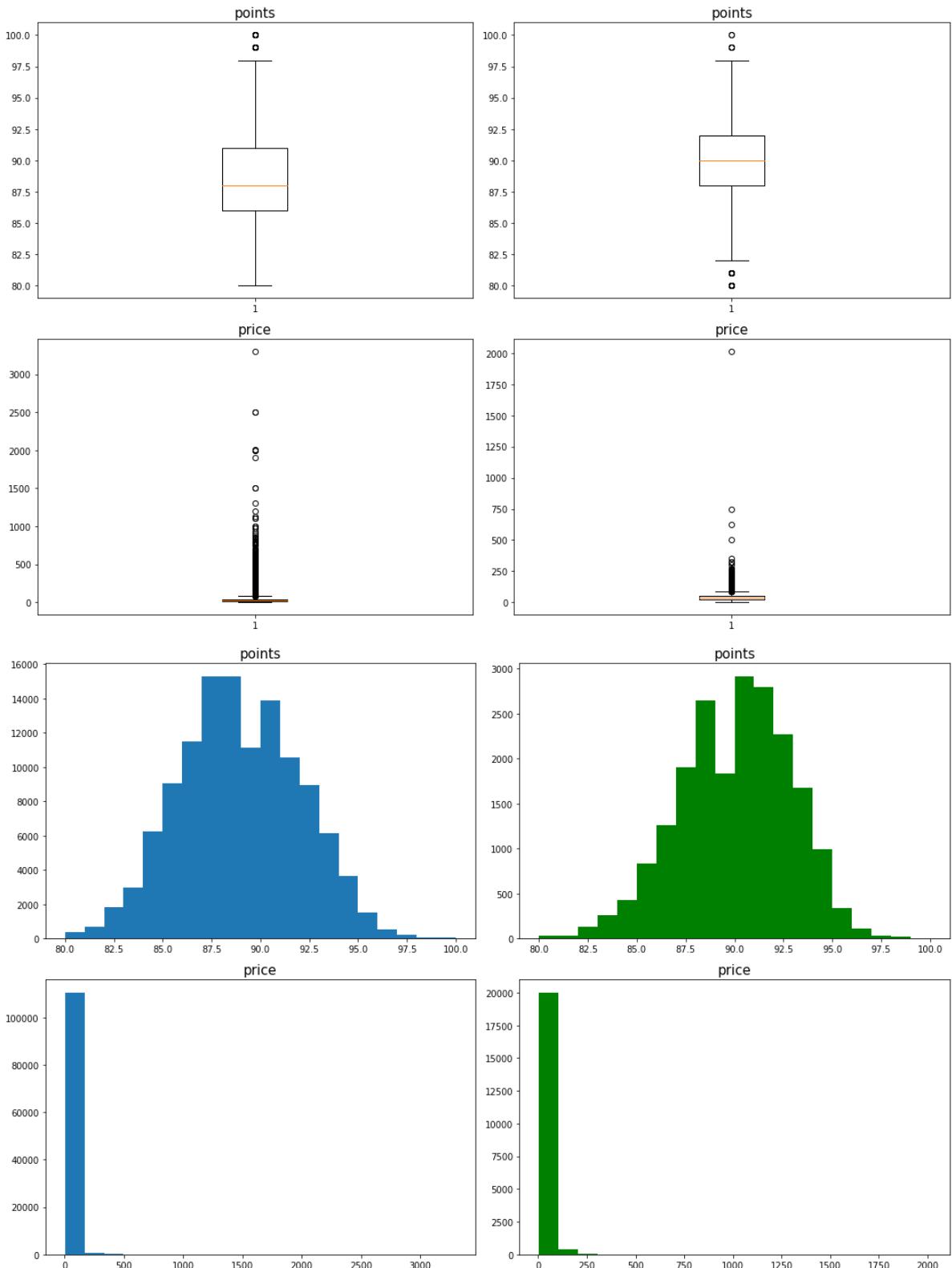
title



new_title





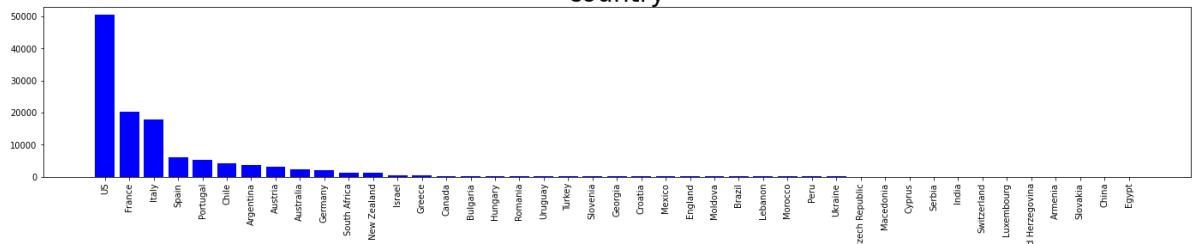


[分析]从图中分析得出本方法会导致country减少至一个，description、region_1、varity出现较大变化，province减少至4个，造成明显的数据分布错误，故此方法不可取

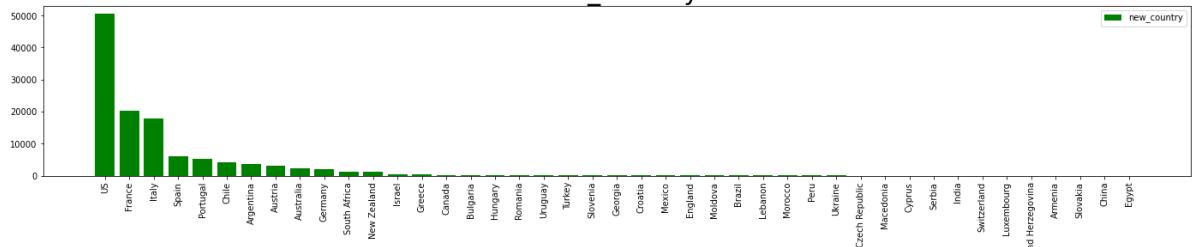
5.2.2用最高频率值来填补缺失值

```
In [67]: new_df_wine2 = miss_mode_fill(df_wine2)
nominal_summary(df_wine2,nominal_index_wine2,new_df_wine2)
box_plot(df_wine2,numerical_index_wine2,new_df_wine2)
hist_plot(df_wine2,numerical_index_wine2,new_df_wine2)
```

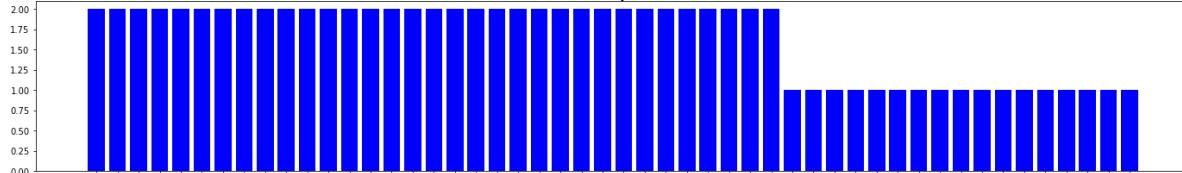
country



new_country



description



lifted pear-drop aromas are accompanied by the pearly green of watercress and anguria. The concentrated palate is grounded by earthy, savory pear notes and framed by fresh, ripe fruit. There is a dollop of firier fruit at the core, swathed with a hint of baking spice and a note of bitter sage leave an astrigent finish.

The subdued nose eventually reveals black cherry, anise, and a whiff of chopped hemlock. Ekyogami and enjyaku, this has aromas of mature black-skinned fruit, violet and an earthy whiff of truffle. The lively simple palate offers black cherry and a hint of baking spice alongside subtle tannins.

Aromas of scorched soil, toasted hazelnut, dried cherries, and an earthy whiff of game carry over to the straightforward, informal palate. The vibrant saffron, pomegranate and a hint of anise bring zesty acidity and defined tannins.

Cigar box, cafe au lait, and dried tobacco aromas are followed by coffee and cherry flavors, with barrel spices lingering on the finish. The wood gets a bit out front but it still delivers enjoyment.

This zesty red has pretty aromas that suggest small red berries, blue flower and a whiff of moist soil. The pale, bright flavors of lemonlime and white peach are balanced with a tannin minerality. A beautiful example of an ancient indigenous variety.

Aromas of citrus, melon and orange blossom start this crisp, elegant white from Cephalonia. On the palate, bright flavors of limoncello and white peach are balanced with a subtle minerality. A beautiful example of an ancient indigenous variety honeysuckle and lavender aromas show on the relatively奔放的 nose of this bottling from a large vineyard on Highway 46 east of Faso Bello. There is a compellingly grippy texture to the sip, with ripe flavors of a soft and honeyed melon. A soft, acidity lingers to the next level.

92-94 Barbet sample. With so much spicy fruit, intensely ripe blackberry flavors and bold tannins, all the elements of excellence are here, and powered by a firm structure. This wine has weight and a fine Cabernet concentration. Hugely delicious, just a joy to drink. So soft and velvety, so rich in blackberry jam, cassis and chocolate flavors, with the tannins so smooth and round. Really wonderful over the next few years, although the softness and high alcohol suggest it's not an age.

Ripe plum, game, truffle, leather and menthol are some of the aromas you'll find on this earthy wine. The tightly wound palate offers dried black cherry, chopped sage, mint and roasted coffee bean alongside raspy tannins that leave a mouth-drying finish.

100% Syrah. Aromas of blackberry jam and cassis. The tannins are rugged, and the wine finishes with a sour note of acidity. Seems at its best now.

Sharp, simple and candied, with blackberry jam and citrus flavors. The tannins are rugged, and the wine finishes with a sour note of acidity. A quixotic wine, ready to lemoning us up into something else. The acidity is generous but not sharp, and the palate is well balanced. All never overpowers.

Subtle aromas of dark berry, tilled earth and baking spice lead the nose. The ripe, juicy palate delivers mature plum, cassis, anise and black pepper alongside round, solid tannins. Drink through 2018.

Classically structured, this still needs to open up but already offers aromas of dark berry, methol, toast, dried lens and a whiff of cocoa. The firm palate delivers red cherry, raspberry, mocha, tobacco and licorice accompanied by bracing tannins and fresh acidity.

Dark berry, toasted hazelnut and metallic aromas (it's out of the glass). The lightly wounded palate is still rather closed off and the lean side in terms of fruit richness, but sowly reveals sour cherry, licorice and note of ground pepper. A tannic backbone provides structure.

Vibrant and spicy, this saucy red delivers aromas and flavors of ripe blackberry, black cherry, ground pepper and chopped Mediterranean herbs. Firm, fine-grained tannins support the juicier flavors. An anise note backs up the finish.

Seductively tart (in lemon pith, cranberry and pomegranate), this effervescent light-bodied quaff is infinitely enjoyable, both on its own or at the table. It continues to expand on the palate into an increasing array of fresh flavors, finishing in cherry and orange.

Sleek by Amazonas standards, this opens with aromas of ripe black cherry, ground paprika, raisin, and a whiff of Black Flower. On the palate, finely woven tannins, black pepper, coca, and a cinnamon note. Very ripe and forward in baked fruit (peeling flavors of cherries and raspberries, spiced with cinnamon, anise and nutmeg). The wine is big and is ready to drink now.

Stately aromas of traditional Italian prates and international varieties such as Cabernet Sauvignon and Merlot. This is a bold, big wine with a thick, chewy mouthfeel that should be paired with aged cheese or succulent meat. Aromas of white flower, green pear and faint mineral/mingle in the glass. The bright palate offers creamy Granny Smith apples, nectarine, sage and a light minty note. Bright acidity gives it a clean finish.

Stately aromas suggest hay and green herbs, with raspberries in the background. It has a bold, short, 1-temperament mouthfeel. The wine is dry and tart, with a hint of hazelnut along the finish. Subtle aromas of Acacia flower and a whiff of almond lead the way. The round, easycare palate offers pear, citrus zest and a hint of hazelnut along the finish.

Aromas of weedy underbrush, toasty, methyl and tannin matted in spirits emerge in the glass. The palate shows raw sour cherry and a hint of licorice alongside raspy tannins that dry out the finish. Rose, red berry, mint and anise aromas unfurl in the glass. The firm palate offers wild cherry, crushed raspberry, dove, aniseed and mocha note framed in bracing tannins and fresh acidity. Drink 2020-2021. Delicate aromas of honeysuckle and pear tail shape in the glass. The palate isn't very expressive, but eventually displays light notes of mature yellow apple and dried lemon zest. Drink soon to capture the remaining fruit.

This wine shows some of the less appealing aspects of American oak: over sweet and vanilla flavors and a slick, oily mouthfeel.

90-92 Barolo sample. This is a solid and dense wine with hints of bitter chocolate along with superripe blackberry fruits. The wine has concentration, richness and a fine aging potential.

Aromas of raspberry and black pepper lay the groundwork for flavors of raspberry, eucalyptus, black cherry and thyme. Soft, but ever present tannins go on long after the bright cherry finish.

This wine, created by Bé Sariá, is a reinvigorated gift from the past. The palate offers bits of baking spices and touches of dried herbs, but is based primarily on rose petals and ripe blueberries. The nose on this amber-colored wine is initially shy but eventually reveals the stone fruit, crushed rock, spice and a raisin-like note. The nature palate offers tangy-dried apricots, anise, vanilla and mineral notes with bright acidity and depth. A briny note of saline and acidic provide an intrigue to the entry of this wine, which is otherwise integrated in oak and medium to full-bodied. Smooth and linear, it sings on the palate, revealing baked apple and pear with bright acidity and depth. A blend of sauvignon and cabernet sauvignon adds to an overbearing sense of ripeness. In the mouth, this is heavy, with weighty tannins. Flavors of prune and P.K. Sherri pie/crème, chocolate and fig flavors on the finish.

Produced, Syrupy number from Cafayate in Salta, Additional Scents of florale, herbs and ready berry only add to the nose on the palate, and tingers on the finish, giving this well-priced wine focus and balance. It has plenty of acidity, fine tannins and good depth of flavor.

A tiny amount of Cabernet Sauvignon and Rapsberry juice from the nose on the palate, and tingers on the finish, giving this well-priced wine focus and balance. It has plenty of acidity, fine tannins and good depth of flavor.

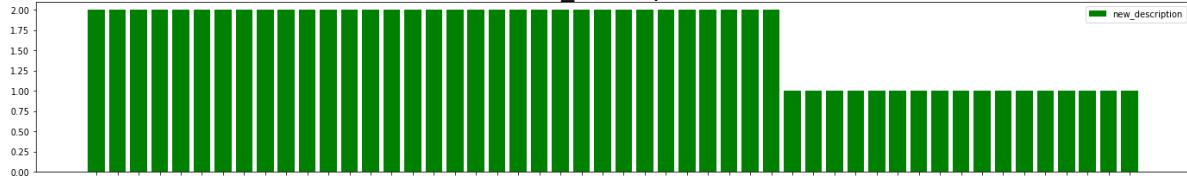
Stale ananas & grapefruit, citrus zest and lime float out of the glass. The radiant palate is more expressive, doing out juicy peach, honeydew melon, nectarine and sage. A mineral note backs up the finish.

Aromas of blueberry, cherry and lime yield on the palate to flavors of cherry, blackberry, cassis/croissants and bacon paper. Smooth tannins, acidity only on entry and follow through, a pleasant finish that is underscored with bright and spicy notes.

This structured blend of Chardonnay, Sauvignon, Picoult and Moscato Giallo offers aromas from toasted oak, ginger and orchard fruit. The elegant palate combines golden Delicious apple, mature apricot, oak-driven spice and candied nectarine testes against taut acidity. Dark and earthy on the nose, this bottling shows sage, menthol, black tea and street-floor elements that decorate the dark hibiscus and pomegranate fruited palate. The palate is much like a fruit-faced black tea, with tart cherry, grippy plump skin and black-peppercorn flavors. Produced from organic grapes, this is a smooth, lightly toasty wine with its lie, aplie and a crisp fruit finish.

Leave dry in the glass this sunny hamlet wine has a nose of black cherry and aniseed. Lightly syrupy and soft palate. Aromas include citrus and floral notes.

new_description



This opens with subtle aromas of wild flower and a whiff of exotic fruit. The bright fruity palate offers white peach, apple and a hint of honeydew melon. It's straightforward and refreshing.

Lifted pear-drop aromatics are accompanied by the peppery green of watercress and angelica. The concentrated palate is grounded by earthy, savory yeast notes and framed by fresh lime peel. There is a dollop of ripe pear fruit at the core, shining and shimmering with added charm.

The subdued nose eventually reveals black cherry, anise and a whiff of chopped herb that carry over to the straightforward, informal palate. Dusty, dry tannins and a note of bitter sage leave an astrigent finish.

Everything and enjoyable, this has aromas of mature black-skinned fruit, violet and an earthy whiff of truffle. The juicy simple palate offers blackberry and a hint of baking spice alongside supple tannins. Drink soon.

This zesty red has pretty aromas that suggest small red berry, blue flower and a whiff of moist soil. The vibrant palate offers sour cherry, pomegranate and a hint of anise alongside zesty acidity and effeted tannins. Aromas of scented soil, roasted hazelnut, dried black cherry and an earthy whiff of game carry over to the palate along with a note of bitter sage. Firm tannins provide support.

Aromas of citrus, melon and orange blossom start crisp, elegant white from Cephalonia. On the palate, bright flavors of lemon/lime and white peach are balanced with a stony minerality. A beautiful example of an ancient indigenous variety.

Gravenstein apple, honeysuckle and jasmine aromas show on the relatively bittersweet nose of this bottling from a large vineyard on Highway 46 east of Poco Bluffs. There is compellingly grippy texture to the spin with ripe flavors of peaches and honeydew melon. A salty acidity takes to the next level.

92-94 Barre sample. Wine so much spicy fruit, intensely ripe blackberry flavors and solid tannins, all the elements of excellence are here, and powered by a firm structure. This wine has weight and a fine Cabernet concentration. Dark berry, toasted hazelnut, and methyl aromas lift off the glass. The slightly round palate is still rather closed and lacks the real side in terms of fruit richness, but, slowly reveals sugar cherries, licorice and a note of ground pepper. A tannic backbone provides structure.

The tannins are so fierce on this baby Cabernet that they effectively lock it down and make it tough to drink. So soft, and velvety, so rich in blackberry jam, cassis and chocolate flavors, with the tannins so smooth and round. Really wonderful over the next few years, although the softness and high alcohol suggest it's not an age-winner.

Ripe plum, grape, truffle, leather and menthol are some of the aromas you'll find on this earthy wine. The tightly wound palate offers dried blackberry, chopped sage, mint and toasted coffee bean alongside raspy tannins that leave a mouth-drying finish. Sharp, simple and candied, with blackberry jam and cola flavors.

Subtle aromas of dark berry, lime, earth and baking spice lead the nose. The fine, juicy yellow notes (x) matured plum, cassis, anise and black pepper alongside round, solid tannins. Drink through 2018.

Classically structured, this still needs to open up but already offers aromas of dark berry, menthol, rose, dried herbs and white pepper. The firm palate delivers red, raspberry, mocha, tobacco and clove accompanied by bracing tannins and fresh acidity.

Dark berry, toasted hazelnut, and methyl aromas lift off the glass. The slightly round palate is still rather closed and lacks the real side in terms of fruit richness, but, slowly reveals sugar cherries, licorice and a note of ground pepper. A tannic backbone provides structure. The wine has weight and a fine Cabernet concentration. Everything about it suggests ageability. Give it a month-drying finish.

Vibrant and spicy, this savoury red delivers aromas and flavors of ripe blackberry, black cherry, ground pepper and chopped Mediterranean herb. Firm, fine-grained tannins support the juicy flavors. An anise note backs up the finish. Seductively tart in lemon pith, cranberry and pomegranate, this refreshing, light-bodied quaff is infinitely enjoyable, both on its own or at the table. It continues to expand on the palate into an increasing array of fresh flavors, finishing in terrytory orange.

Very ripe and forward in baked fruit and pinning flavors of cherries and raspberries, spiced with cinnamon, anise and nutmeg. The wine is full, simple and ready to drink now.

Subtle aromas with a note of black cherry, ground pepper, raisin and a whiff of blue flower. On the palate, finely woven tannins support the fruit. The palate shows raw sour cherry and a hint of rose alongside raspy tannins that dry out the finish.

Smooth, round and ripe, this Savignon and Merlot. This is a bold, big wine with a thick, chewy mouthfeel that should be paired with aged cheese or succulent meat.

Aromas of lime flower, green pea and minty minty mélange in the glass. The right palate offers creamy Granny Smith apple, incense, sage and a light mineral note. Bright acidity gives it a clean finish.

Subtle aromas of aceda flower and a whiff of minted herbs leading the way and Henry fruit running behind. Dry red fruit and herbal notes dominate the finish.

Aromas of weedy underbrush, toast, menthol and cherry marinated in spirits emerge in the glass. The palate shows raw sour cherry, crushed raspberry, clove, anise and a mouth note named in tracking raspberry jam, black pepper, lime and mineral. A blast of blood orange backs up the finish.

Ripe, red berry, mint and anise aromas unite in the glass. The firm palate offers wild cherry, crushed raspberry, clove, anise and a mouth note named in tracking raspberry jam, black pepper, lime and mineral. Dried lemon zest, dried lemon zest, drink soon to capture the remaining fruit.

Stalky aromas suggest hay and green herbs, with raspberry in the background. It's hot and short in moments of mouthfeel with hemal flavors leading the way and Henry fruit running behind. Dry red fruit and herbal notes dominate the finish.

This wine shows some of the less appealing aspects of American oak, overt spice and vanilla flavors and a slick, oily mouthfeel.

90-92 Barrel sample. This is a solid and dense wine with hints of bitter chocolate along with Superfine blackberry fruits. The wine has concentration, richness and a fine ageing potential.

Aromas of raspberry and black pepper lay the groundwork for flavors of blackberry, eucalyptus, black cherry and thyme. Soft but over present tannins grow on long after the bright cherry finish.

This wine, boosted by 8% Syrah, is relatively light in the glass, showing vanilla and plums and red cherries on the nose. The nature palate offers tangine zest, dried apricot, anise, vanilla and mineral together with bright acidity, a balanced complexity and depth.

A briny note of saline and pickle provides an intrigue to the entry of this wine, which is otherwise integrated in oak and medium to full-bodied. Smooth and floral, it hangs on the palate, revealing baked apple and pear around a hint of wild vanilla and herb.

Raisin and molasses aromas and flavors are a big part of this full-bodied, syrupy number from Caléyave in Salta. Additional scents of horseradish, herbs and earthy berry add to an overbearing sense of ripeness in the mouth, this is heavy, with weighty tannins. Flavors of prune and Kirsch precede coffee, Golden Delicious apple, mature apricot, oak-driven spice and candied Kielbasa, zests alongside tangy acidity.

Produced from organic grapes, this is a smooth, lightly toasty wine. It has ripe apple and a crisp, tight, mineral texture that adds to the interesting character of the wine. The aftertaste is a little fresh fruit.

A briny vein of cranberry and raspberry runs from the nose to the palate, flavors of ripe blackberry, eucalyptus and black pepper, smooth tannins appear early on and endure into a pleasing finish that is interspersed with bright and spicy notes.

Stale aromas of honeysuckle and peat take shape in the glass. The palate is very expressive, but eventually displays light notes of mature yellow apple and dried lemon zest, drink soon to capture the remaining fruit.

The freshness of seared Meyer lemon and squeezed lime peel meets with the warmth of vanilla and anise in the nose of this solid bottling. The palate offers white peach and smoked apples laid across a very tense and pithy texture. It finishes on a Juicy Fruit bubblegum element that's intriguing and nice overdone.

This structured blend of Chardonnay, Sauvignon, Roulle and Moscato Giallo offers aromas of toasted oak, ginger and orchard fruit. The elegant palate combines Golden Delicious apple, mature apricot, oak-driven spice and candied Kielbasa, zests alongside tangy acidity.

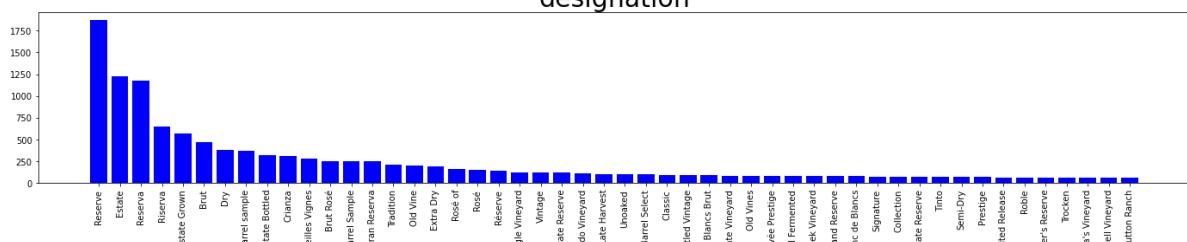
Dark and earthy on the nose, this bottling shows sage, menthol, black tea and forest-floor elements that decorate the dark berries and pomegranate fruits. The palate is much like a fruit-faced black tea.

Everything about this wine is big and heavy, including the name, the pyrrolidinedione, the massive bottle and its size. It tastes great. The color is deep and dark, aroma is full of baby spices of nutmeg and vanilla, and the massive bottle is full of that mass, it tastes great.

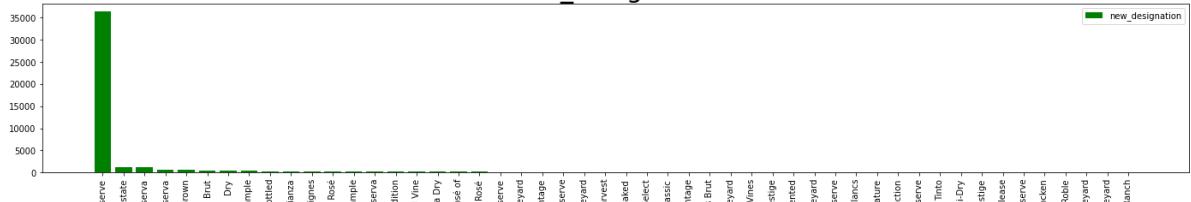
Deep ruby in the glass, this aptly-named wine has a note of black cherry and aniseed. Slightly grippy tannins and strong acidity stand at the core of blueberry, raspberry, aniseete and peppermint flavors. Bunches of mint and orange zest make an impression on the taste buds welling the enduring finish.

Abacela makes a fortified wine from Portuguese grapes, but this is a dry, varietal expression of one of the most obscure. Bone-dry, it's got tongue-scrapping tannins that taste of black tea. It offers dried raspberries, vanilla and cocoa elements, and the overall complexity is impressive. This seems built for a thick steak.

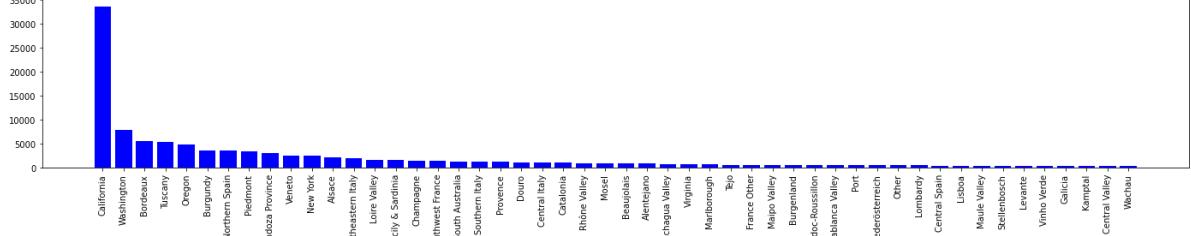
designation



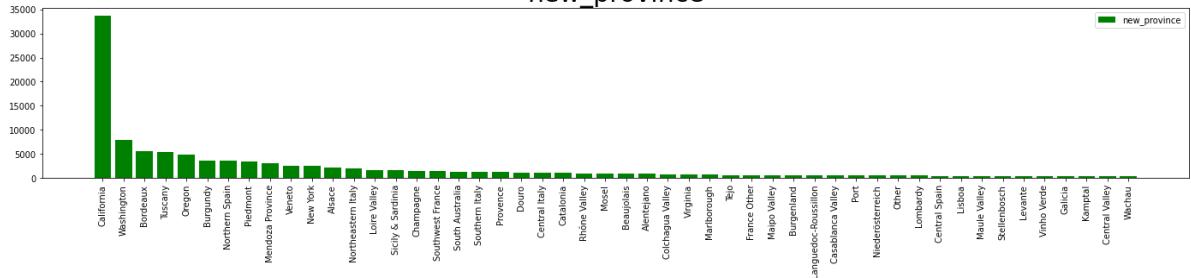
new_designation



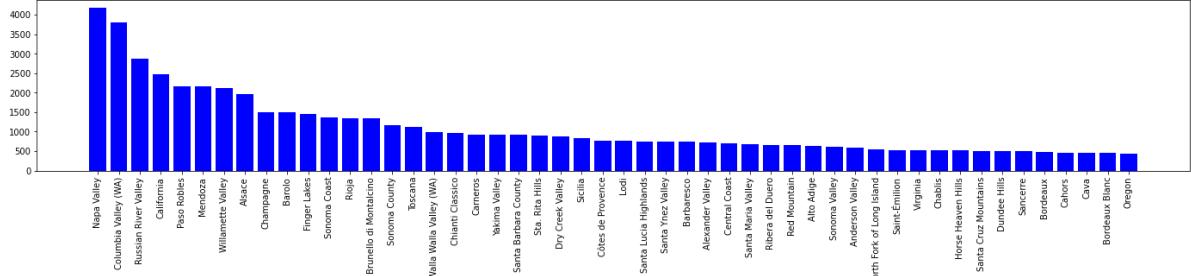
province



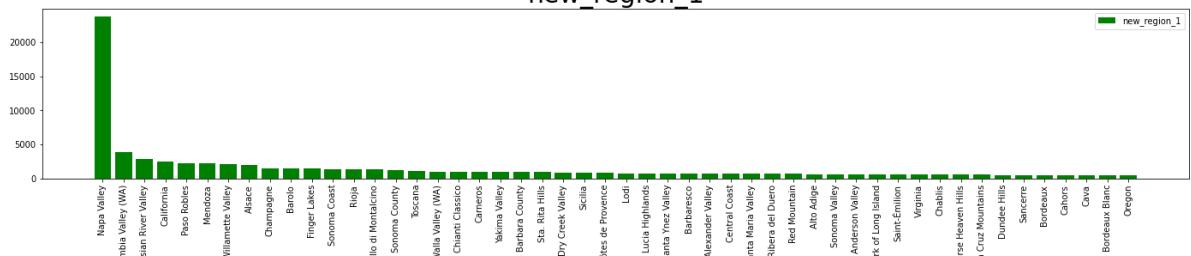
new_province



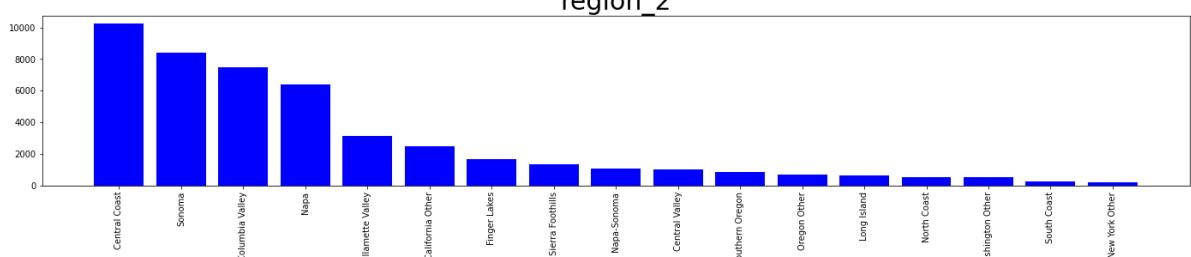
region_1



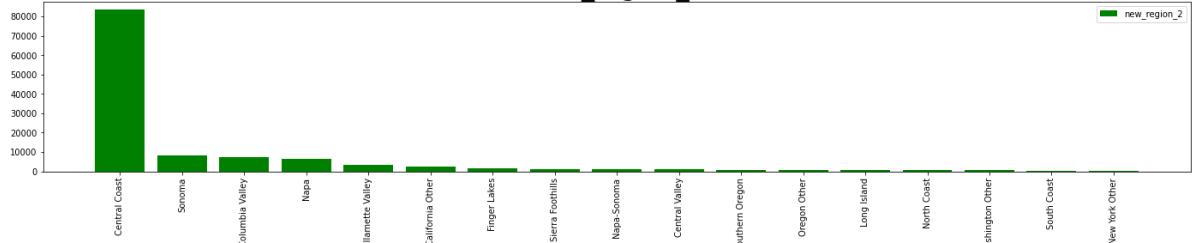
new_region_1



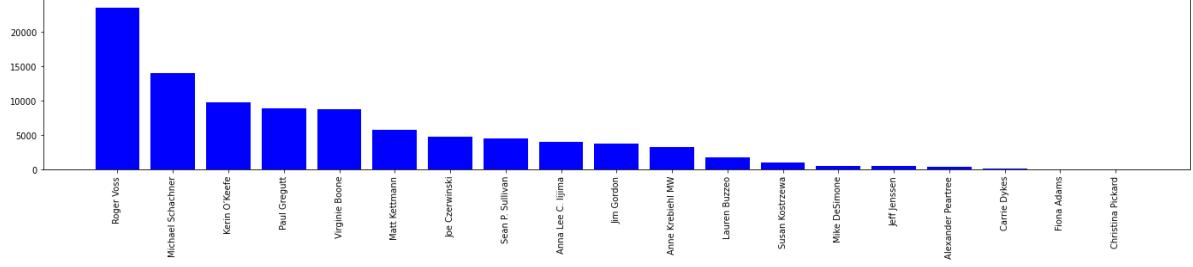
region_2



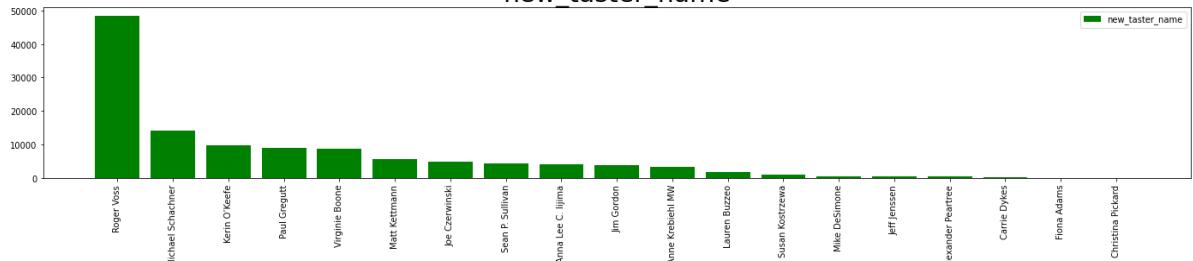
new_region_2



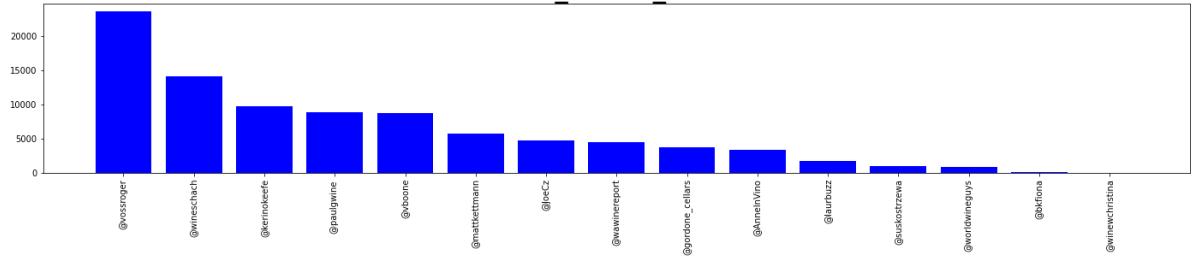
taster_name



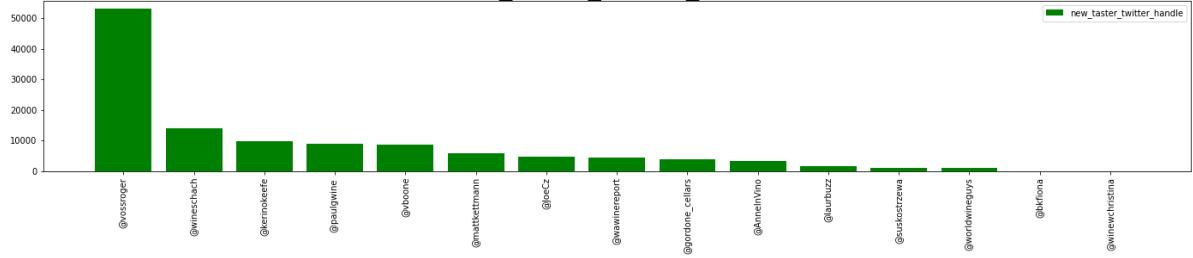
new_taster_name



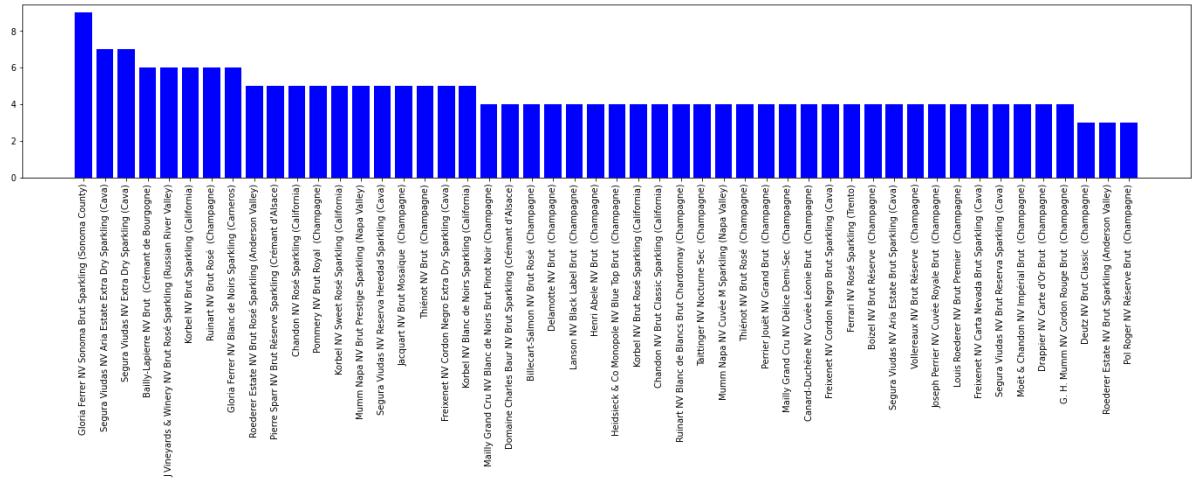
taster_twitter_handle



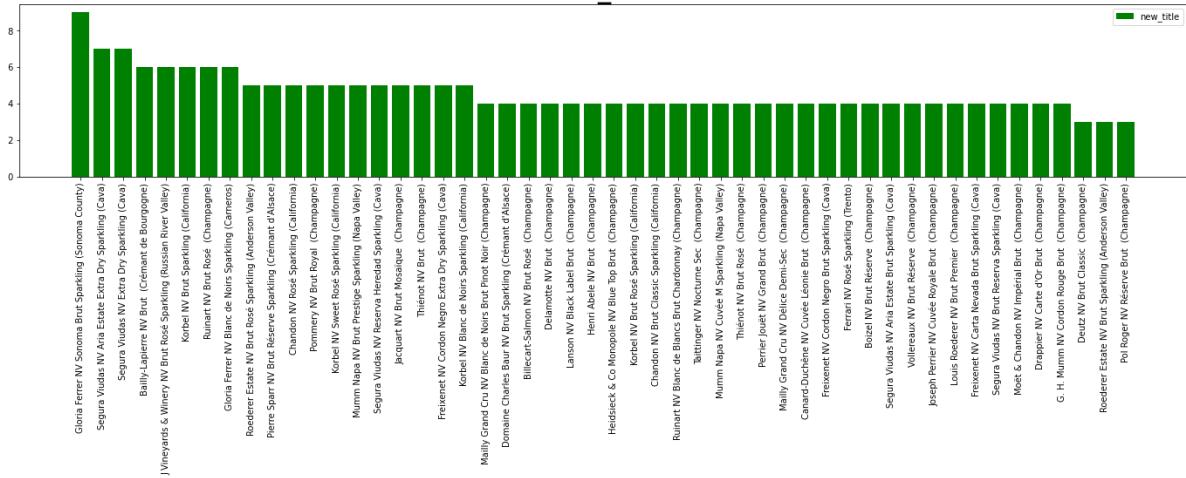
new_taster_twitter_handle



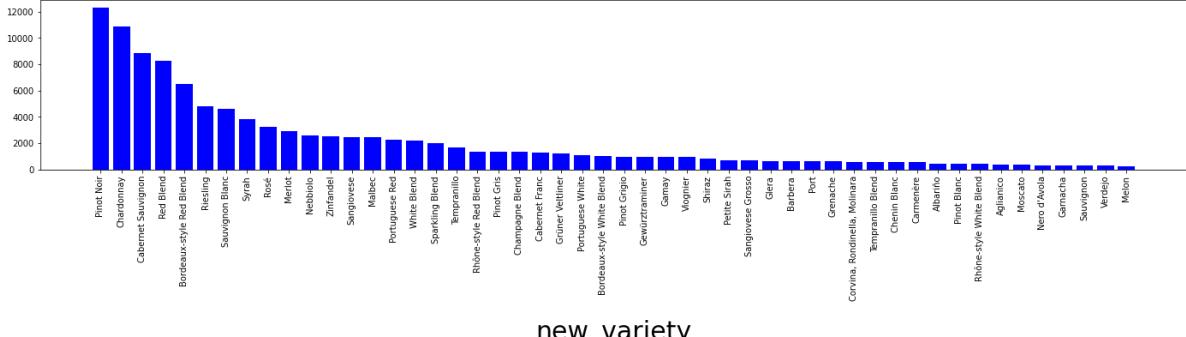
title



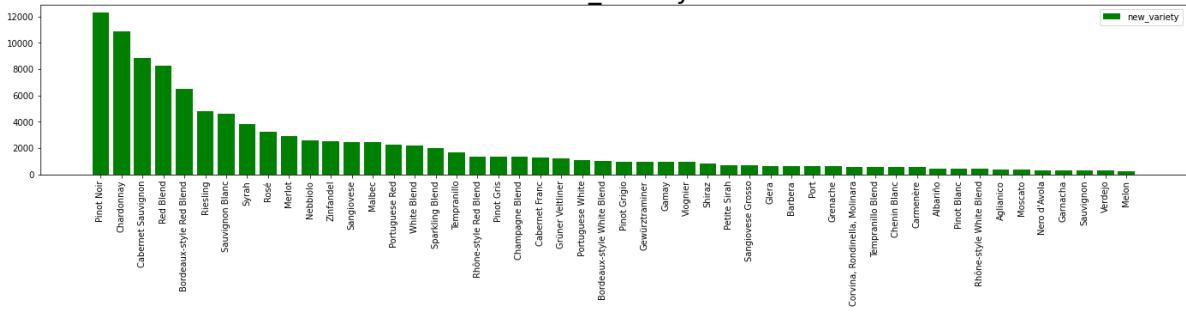
new_title



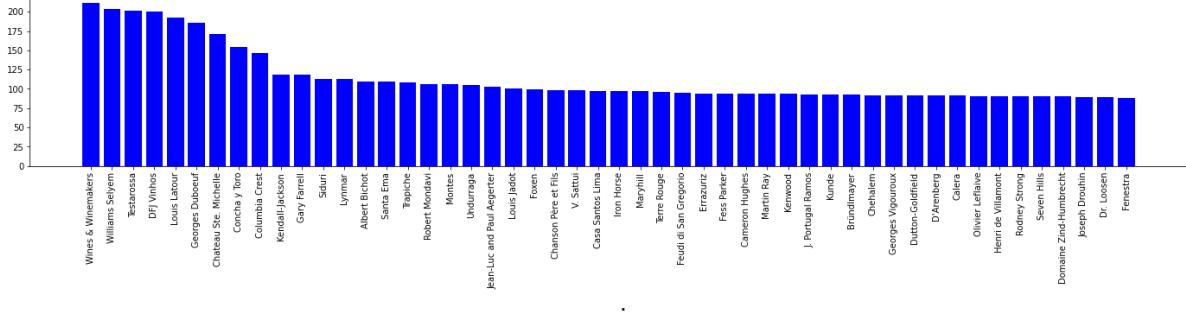
variety



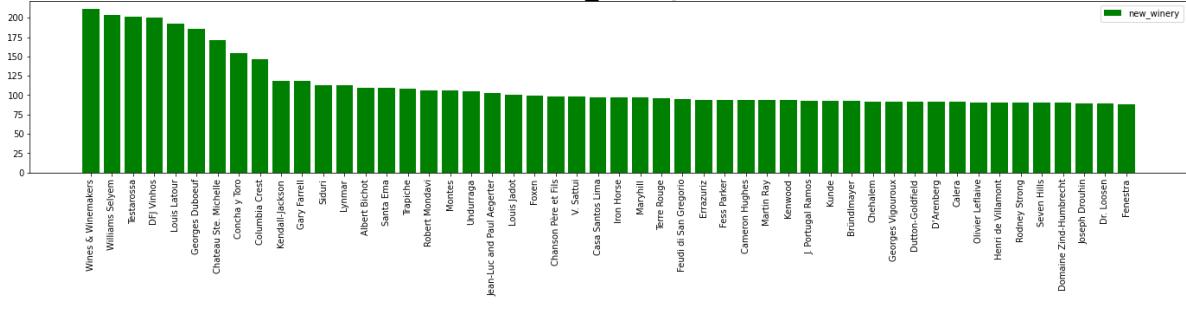
new_variety

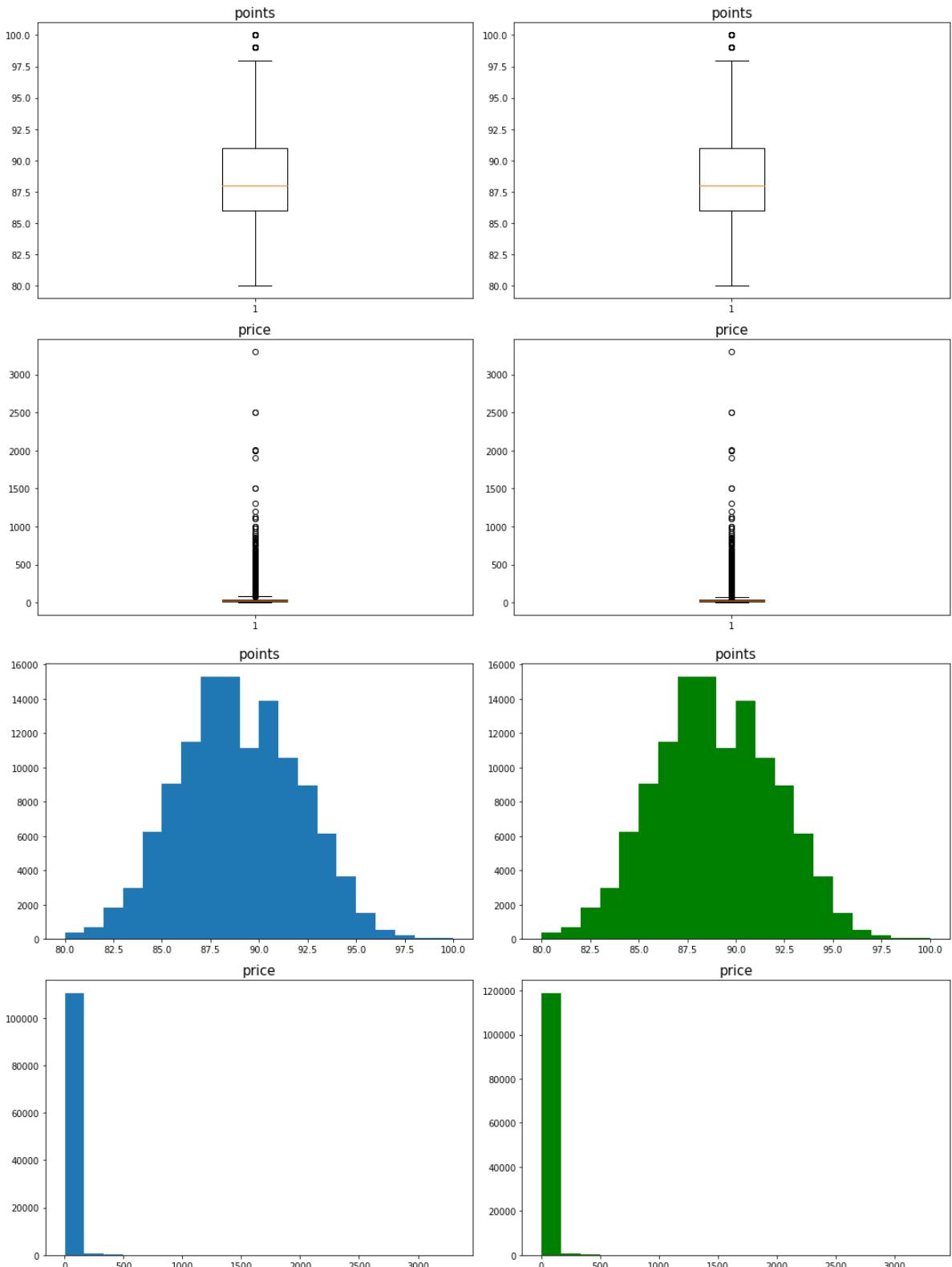


winery



new_winery





[分析]从图中分析得出本方法会导致designation、region_1、region_2等出现较大变化，造成明显的数据分布错误，故此方法不可取

5.2.3通过属性的相关关系来填补缺失值

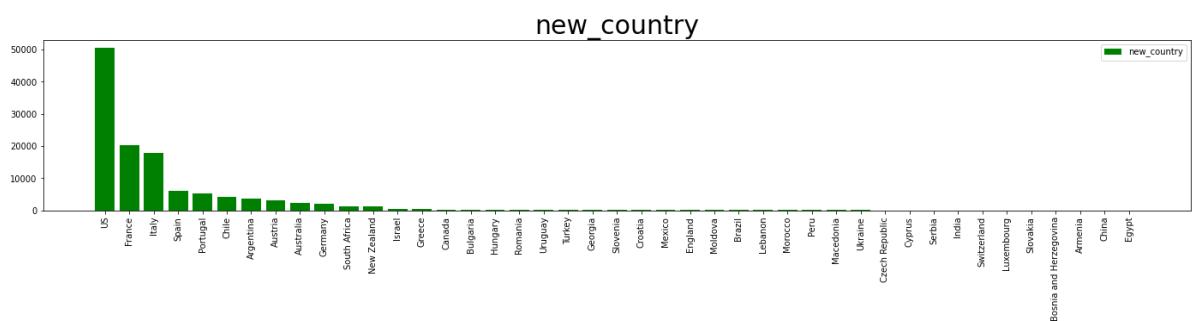
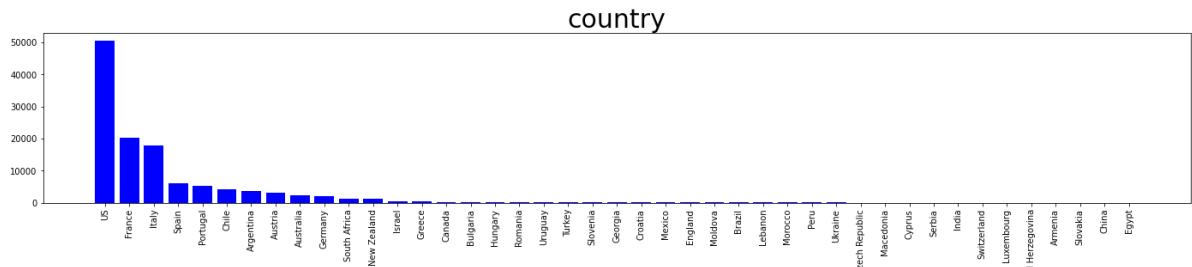
```
In [71]: miss_index = list()
complete_index = list()
col = df_wine2.columns.values
i = 0
for index in list(df_wine2.isnull().any()): #index==True时，对应的列中具有缺失值
    if(index):
        miss_index.append(col[i])
    else:
```

```

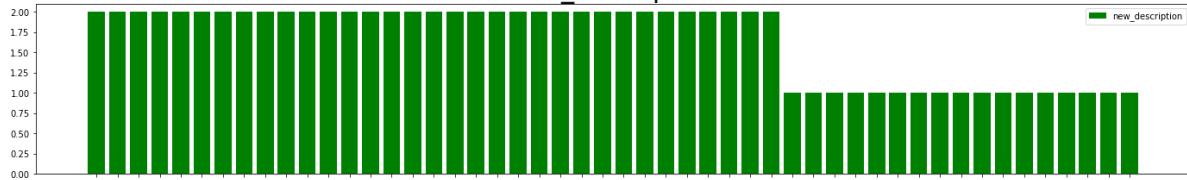
    complete_index.append(col[i])
    i = i + 1

new_df_wine2 = attr_corr_fill(df_wine2, miss_index, complete_index)
nominal_summary(df=df_wine2, nominal_index=nominal_index_wine2, new_df=new_df_wine2)
box_plot(df_wine2, numerical_index_wine2, new_df_wine2)
hist_plot(df_wine2, numerical_index_wine2, new_df_wine2)

```



new_description



This opens with subtle aromas of wild flower and a whiff of exotic fruit. The bright fruity palate offers white peach, apple and a hint of honeydew melon. It's straightforward and refreshing. Lifted pear-drop aromatics are accompanied by the peppery green of watercress and arugula. The concentrated palate is grounded by earthy, savory yeast notes and framed by fresh lime peel. There is a dollop of ripe pear fruit at the core, shining and shimmering with added charm. The subdued nose eventually reveals black cherry, anise and a whiff of chopped herb that carry over to the straightforward, informal palate. Dusty, dry tannins and a note of bitter sage leave an astrigent finish.

Everything and enjoyable, this has aromas of mature black-skinned fruit, violet and an earthy whiff of truffle. The juicier sample palate offers blackberry and a hint of baking alongside supple tannins. Drink soon.

This zesty red has pretty aromas that suggest small red berry, blue flower and a whiff of moist soil. The vibrant palate offers sour cherry, pomegranate and a hint of anise alongside zesty acidity and effeted tannins. Aromas of scorched soil, roasted hazelnut, dried black cherries and an earthy whiff of game carry over to the palate along with a note of bitter sage. Firm tannins provide support.

Aromas of citrus, melon and orange blossom start this crisp, elegant white from Cephelema. On the palate, bright flavors of lemon/lime and honeydew melon. A salty acidity takes it to the next level.

Cigar box, café au lait, and dried tobacco aromas are followed by coffee and cherries, with harelip spices lingering on the fresh. The wood gets a bit out front but it still delivers enjoyment.

Grenache apple, honeysuckle and jasmine aromas show on the relatively modest nose of this bottling from a large vineyard on Highway 46 east of Paso Robles. There is compellingly grippy texture to the spin with ripe flavors of peaches and honeydew melon. A salty acidity takes it to the next level.

92-94 Barre sample. Wine so much spicy fruit, intensely ripe blackberry flavors and solid tannins; all the elements of excellence are here, and powered by a firm structure. This wine has weight and a fine Cabernet concentration. Dark berry, toasted hazelnut, and methyl aromas lift off the glass. The slightly round palate is still rather closed and on the nose side in terms of fruit richness, but slowly reveals sugar cherry, licorice and a note of ground pepper. Aromatic backbone provides structure. The tannins are so fierce on this baby Cabernet that they effectively lock it down and make it tough to drink. So soft, and velvety, so rich in blackberry jam, cassis, and chocolate flavors, with the tannins so smooth and round. Really wonderful over the next few years, although the softness and high alcohol suggest it's not an age-winer.

Ripe plum, grape, truffle, leather and menthol are some of the aromas you'll find on this earthy wine. The tightly wound palate offers dried blackberry, chopped sage, mint, and toasted coffee bean alongside raspy tannins that leave a mouth-drying finish.

Graves-style apple, honeysuckle and jasmine aromas that run through a gamut of fruit flavors. A quintessential food wine Sharp, simple and candied, with blackberry jam and cola flavors. The tannins are rugged, and the wine finishes with a scoop of acidity. Seems at its best now.

Subtle aromas of dark berry, lime, earth and baking spice lead the nose. The fine, juicy yellow notes (mustard, cassis, anise and black pepper) alongside round, solid tannins. Drink through 2018.

Hugely delicious, just a joy to drink. So soft, and velvety, so rich in blackberry jam, cassis, and chocolate flavors, with the tannins so smooth and round. Really wonderful over the next few years, although the softness and high alcohol suggest it's not an age-winer.

Ripe plum, grape, truffle, leather and menthol are some of the aromas you'll find on this earthy wine. The tightly wound palate offers dried blackberry, chopped sage, mint, and toasted coffee bean alongside raspy tannins that leave a mouth-drying finish.

Grenache apple, honeysuckle and jasmine aromas that run through a gamut of fruit flavors. A quintessential food wine Seductively tart in lemon pit, cranberry and pomegranate, this refreshing, light-bodied quaff is infinitely enjoyable, both on its own or at the table. It continues to expand on the palate into an increasing array of fresh flavors, finishing in terrytory orange.

Very ripe and forward in baked fruit and pinning flavors of cherries and raspberries, spiced with cinnamon, anise and nutmeg. The wine is simple and ready to drink now.

Subtle aromas of a mix of black cherry, ground pepper, raisin and a whiff of black cherry flower. On the palate, finely woven tannins support raspberry jam, black pepper, cocoa and a cinnamon note.

Vibrant and spicy, this savory red delivers aromas and flavors of ripe blackberry, black cherry, ground pepper and chopped Mediterranean herb. Firm, fine-grained tannins support the juicy flavors. An anise note backs up the finish.

Subtle aromas of dark berry, lime, earth and baking spice lead the nose. The fine, juicy yellow notes (mustard, cassis, anise and black pepper) alongside round, solid tannins. Drink through 2018.

Dark berry, toasted hazelnut, and methyl aromas lift off the glass. The slightly round palate is still rather closed and on the nose side in terms of fruit richness, but slowly reveals sugar cherry, licorice and a note of ground pepper. Aromatic backbone provides structure. The tannins are so fierce on this baby Cabernet that they effectively lock it down and make it tough to drink. So soft, and velvety, so rich in blackberry jam, cassis, and chocolate flavors, with the tannins so smooth and round. Really wonderful over the next few years, although the softness and high alcohol suggest it's not an age-winer.

Ripe, red berry, mint and anise aromas unite in the glass. The firm palate offers wild cherry, crushed raspberry, clove, aniseote, and a mouth note named in tracking raspberry jam, black pepper, lime, and mineral. A blast of blood orange backs up the finish.

Subtly aromatic, hay and green herbs, with raspberry in the background. It's hot and short in moments of mouthfeel with herby flavors leading the way and fiery fruit running behind. Dry red fruit and herbal notes dominate the finish.

Subtle aromas of a mix of black cherry, ground pepper, raisin and a whiff of black cherry flower. On the palate, finely woven tannins support raspberry jam, black pepper, cocoa and a cinnamon note.

Rose, red berry, mint and anise aromas unite in the glass. The firm palate offers wild cherry, crushed raspberry, clove, aniseote, and a mouth note named in tracking raspberry jam, black pepper, lime, and mineral. A blast of blood orange backs up the finish.

Delicate aromas of honeysuckle and peach take shape in the glass. The pale yellow is expressive, but eventually displays light notes of mature yellow apple and dried lemon zest. Drink soon to capture the remaining fruit.

Stalky aromas suggest hay and green herbs, with raspberry in the background. It's hot and short in moments of mouthfeel with herby flavors leading the way and fiery fruit running behind. Dry red fruit and herbal notes dominate the finish.

Subtly aromatic, hay and green herbs, with raspberry in the background. It's hot and short in moments of mouthfeel with herby flavors leading the way and fiery fruit running behind. Dry red fruit and herbal notes dominate the finish.

This wine shows some of the less appealing aspects of American oak, overt spice, and vanilla flavors and a slick, oily mouthfeel.

90-92 Barrel sample. This is a solid and dense wine with hints of bitter chocolate along with supple blackberry fruits. The wine has concentration, richness and a fine ageing potential.

Atoms as raspberry and black pepper lay the groundwork for flavors of blackberry, eucalyptus, black cherry and thyme. Soft but over present tannins grow on long after the bright cherry finish.

Aromas of weedy underbrush, toast, methyl and cherry marmalade in spirits emerge in the glass. The palate shows raw sour cherry and a hint of rose, alongside raspy tannins that dry out the finish.

Rose, red berry, mint and anise aromas unite in the glass. The firm palate offers wild cherry, crushed raspberry, clove, aniseote, and a mouth note named in tracking raspberry jam, black pepper, lime, and mineral. A blast of blood orange backs up the finish.

This wine, boosted by 8% Syrah, is relatively light on the nose, showing vanilla and plums and red cherries on the nose. The nature palate offers tangine zest, citrus, apricot, anise, vanilla and mineral together with bright acidity, a balanced complexity and depth.

A briny note of saline and pickle provides an intrigue to the entry of this wine, which is otherwise integrated in oak and medium to full-bodied. Smooth and floral, it hangs on the palate, revealing baked apple and pear around a hint of wild vanilla and herb.

The nose on this amber-colored wine is initially shy but eventually reveals ripe stone fruit, crushed rock, spice and a citrusy note. The nature palate offers tangine zest, citrus, apricot, anise, vanilla and mineral together with bright acidity, a balanced complexity and depth.

Raisin and molasses aromas and flavors are a big part of this full-bodied, syrupy number from Cayelaine in Salta. Additional scents of horseradish, herbs and earthy berry add to an overbearing sense of ripeness in the mouth, this is heavy, with weighty tannins. Flavors of prune and Kirsch precede coffee, Golden Delicious apples, mature apricot, oak-driven spice and candied Kharchine zest alongside tangy acidity.

Dark and earthy on the nose, this bottling shows sage, menthol, black tea and forest-floor elements that decorate the dark berries and pomegranate fruits. The palate is much like a fruit-faced black tea. A briny vein of cranberry and raspberry runs from the nose to the palate, and lingers on the finish, giving this wine a crisp, tight, mineral texture that adds to the interesting character of the wine. The aftertaste is a bit fresh fruit.

Produced from organic grapes, this is a smooth, lightly toasty wine. It has ripe apple and a crisp, tight, mineral texture that adds to the interesting character of the wine. The aftertaste is a bit fresh fruit.

Atoms of blueberry, cherry and anise on the palate to flavors of cherry, blackberry, eucalyptus and black pepper. Smooth tannins appear early on and endure into a pleasing finish that intersperses with bright and spicy notes.

Stale aromas of grapefruit, citrus zest and kiwi float out of the glass. The radiant palate is more expressive, dolling out juicy yellow peach, honeydew melon, nectarine and sage. A mineral note backs up the finish.

The freshness of seared Meyer lemon and squeezed lime peel meets with the warmth of vanilla and anise in the nose of this solid bottling. The palate offers white peach and smoked apples laid across a very tense and pithy texture. It finishes on a Juicy Fruit bubblegum element that's intriguing and nice overripe.

This structured blend of Chardonnay, Sauvignon, Roulle and Moscato Giallo offers aromas of toasted oak, ginger and orchard fruit. The elegant palate combines Golden Delicious apples, mature apricot, oak-driven spice and candied Kharchine zest alongside tangy acidity.

Dark and earthy on the nose, this bottling shows sage, menthol, black tea and forest-floor elements that decorate the dark berries and pomegranate fruits. The palate is much like a fruit-faced black tea. A briny vein of cranberry and raspberry runs from the nose to the palate, and lingers on the finish, giving this wine a crisp, tight, mineral texture that adds to the interesting character of the wine. The aftertaste is a bit fresh fruit.

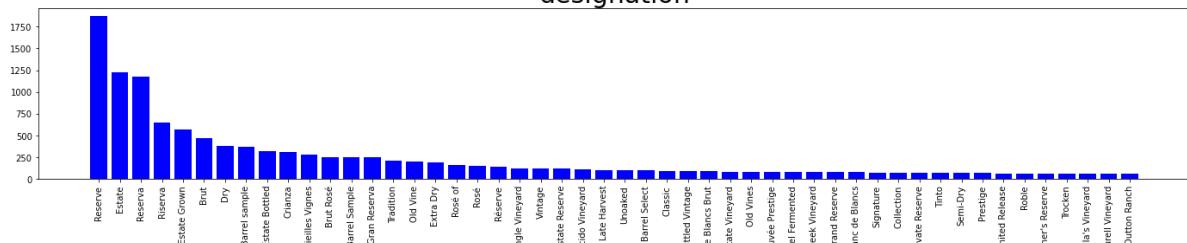
Atoms of black cherry, clove, black pepper and chocolate lay the groundwork for flavors of ripe black cherry, strawberry, chocolate, tangerine and lime. The pyrrolidine base and the massive bottle of all that mass, it tastes great. The color is deep and dark, aroma is full of baby spices of nutmeg and vanilla, and the massive bottle of all that mass, it tastes great.

Everything about this wine is big and heavy, including the name, the pyrrolidine base and the massive bottle of all that mass, it tastes great. The color is deep and dark, aroma is full of baby spices of nutmeg and vanilla, and the massive bottle of all that mass, it tastes great.

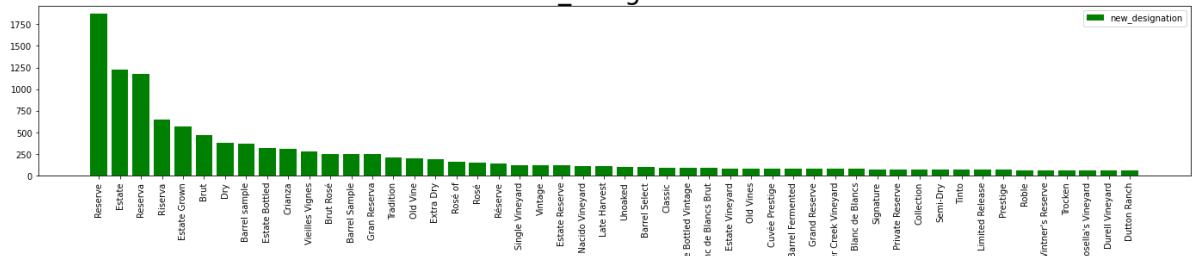
Deep ruby in the glass, this aptly-named wine has a nose of black cherry and aniseed. Slightly grippy tannins and strong acidity stand at the core of blueberry, raspberry, aniseete and peppermint flavors. Bunches of mint and orange zest make an impression on the taste buds welling the enduring finish.

Abacela makes a fortified wine from Portuguese grapes, but this is a dry, varietal expression of one of the most obscure. Bone-dry, it's got tongue-scraping tannins that taste of black tea. It offers dried raspberries, vanilla and cocoa elements, and the overall complexity is impressive. This seems built for a thick steak.

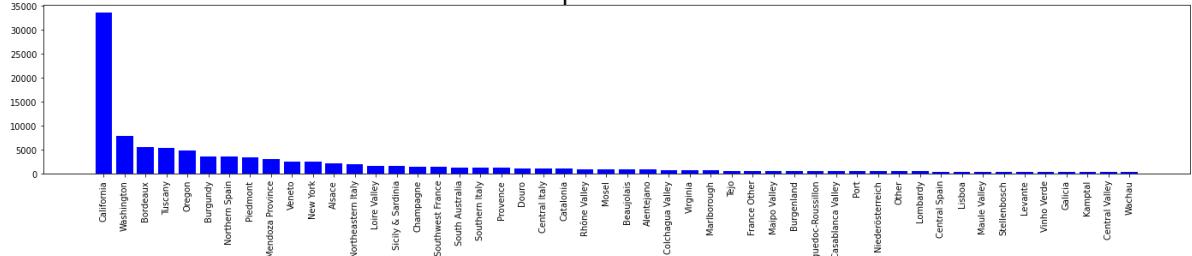
designation



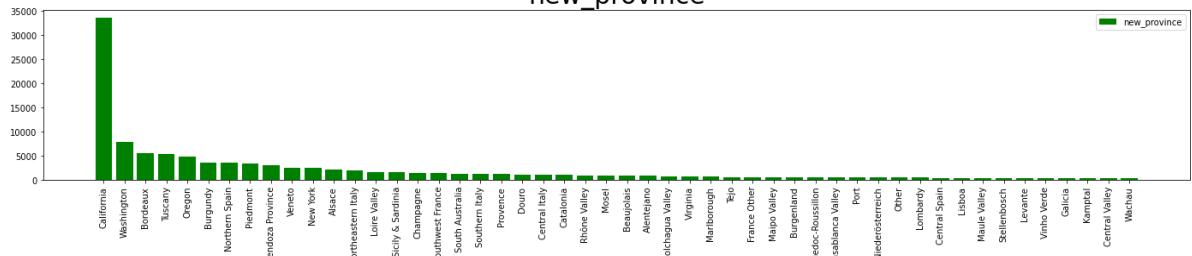
new_designation



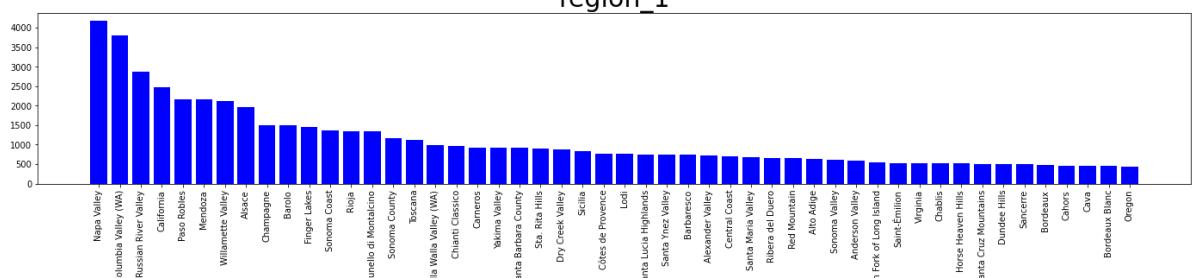
Province



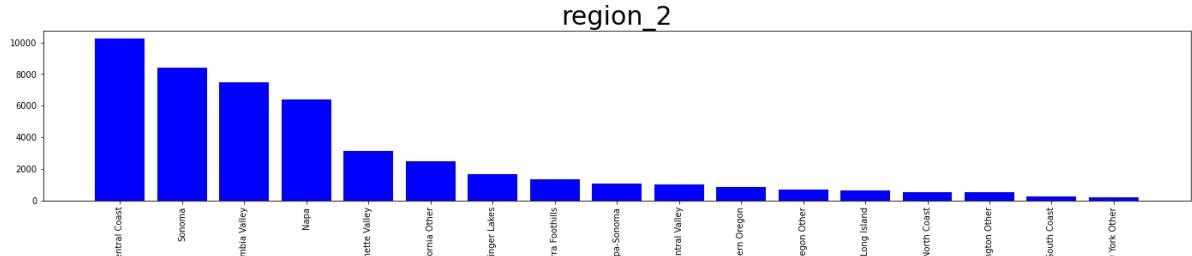
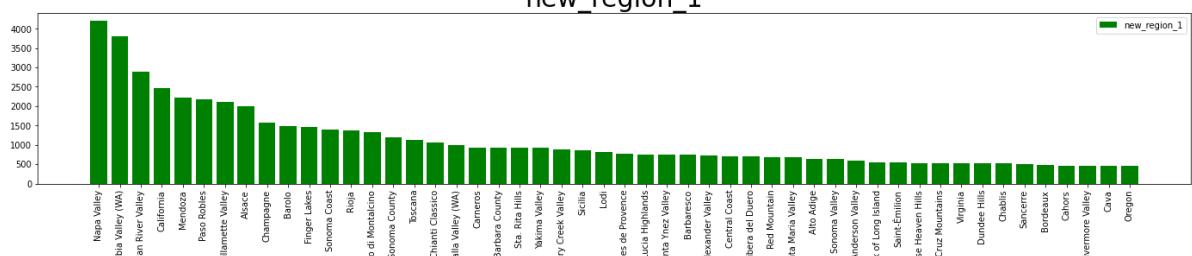
new province

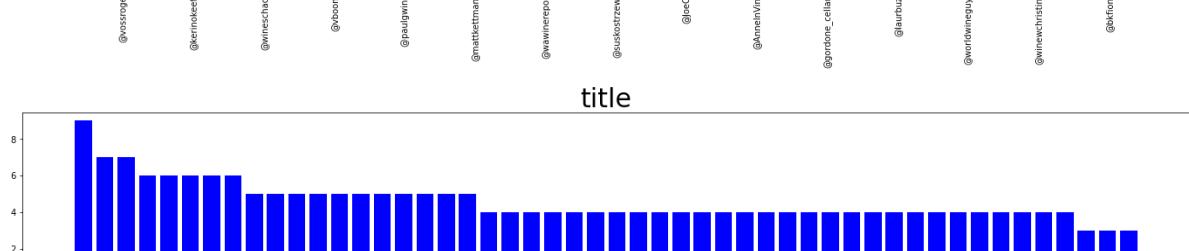
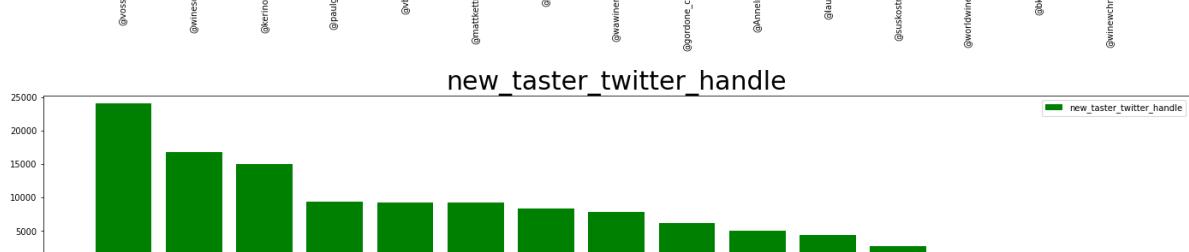
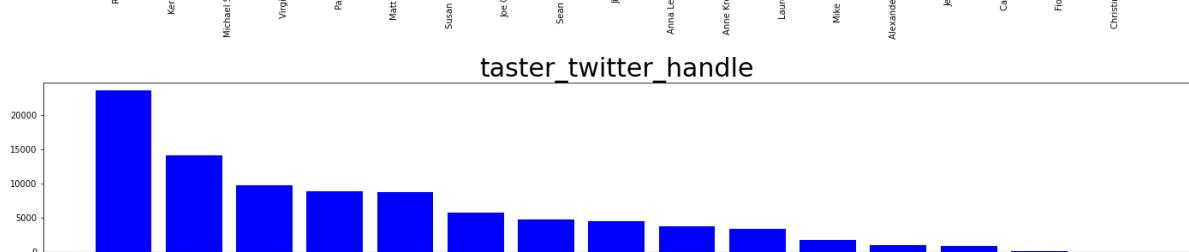
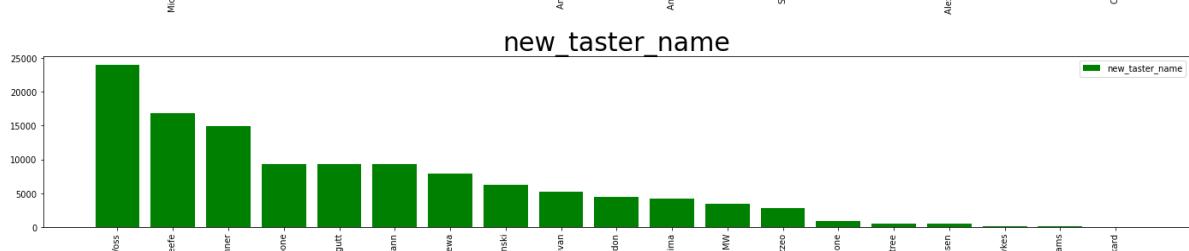
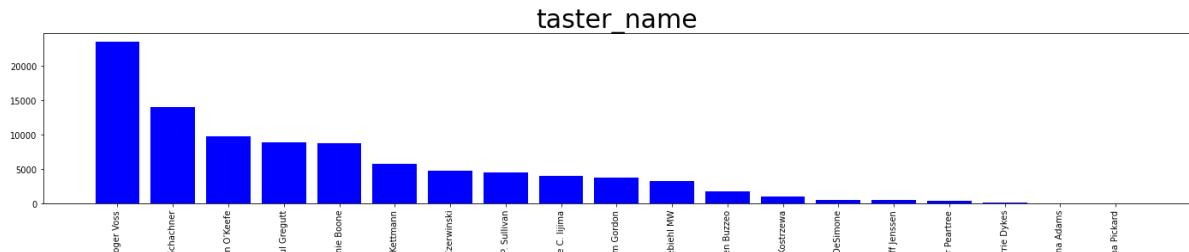
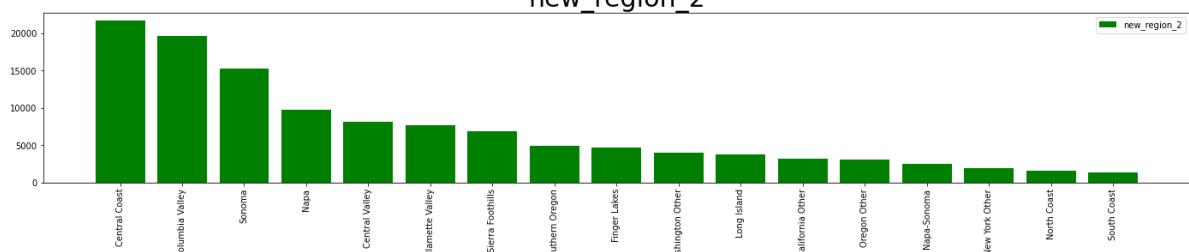


region 1

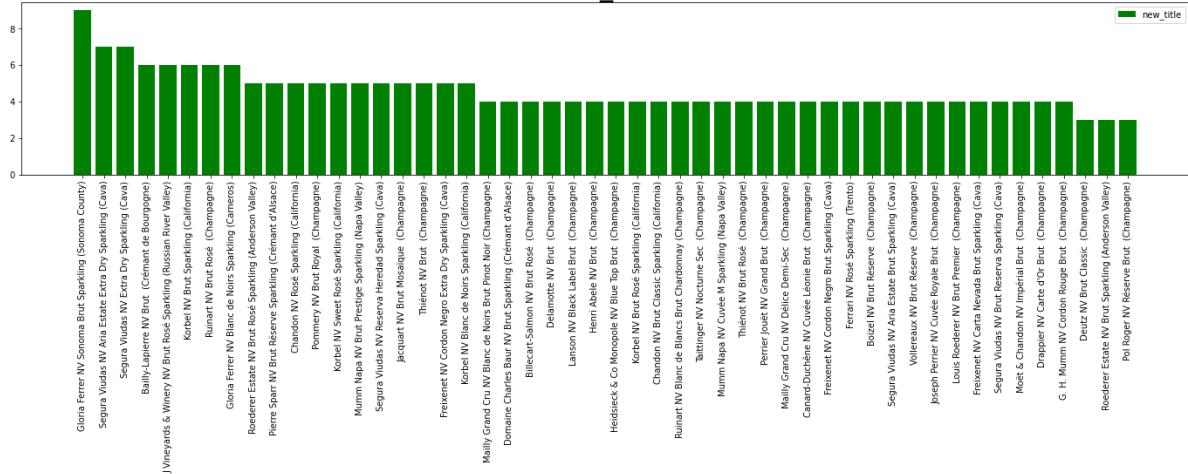


new region 1

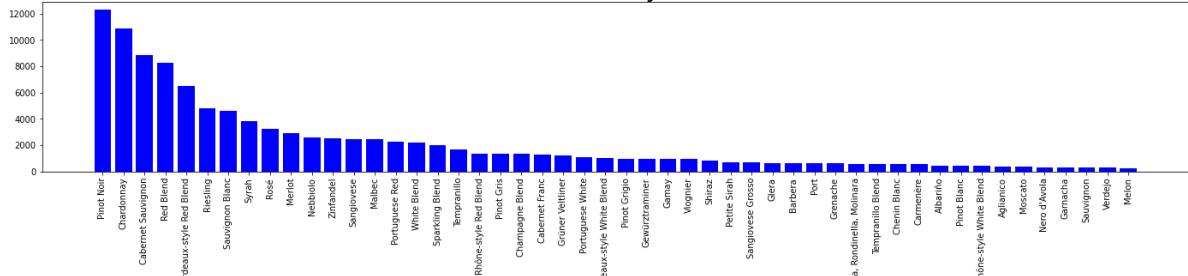




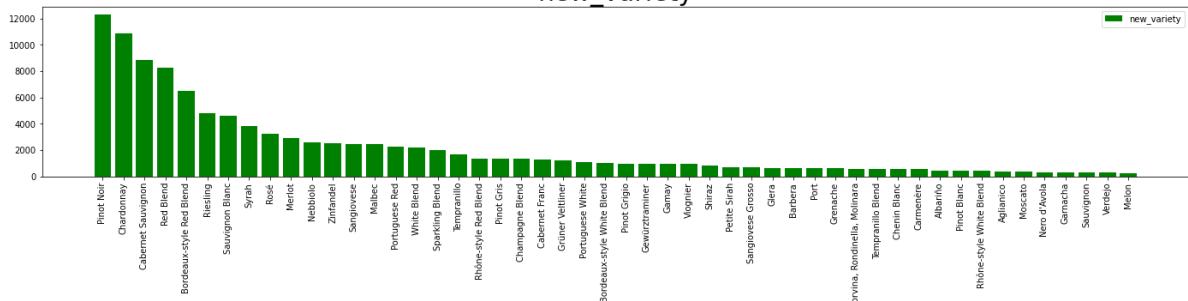
new_title



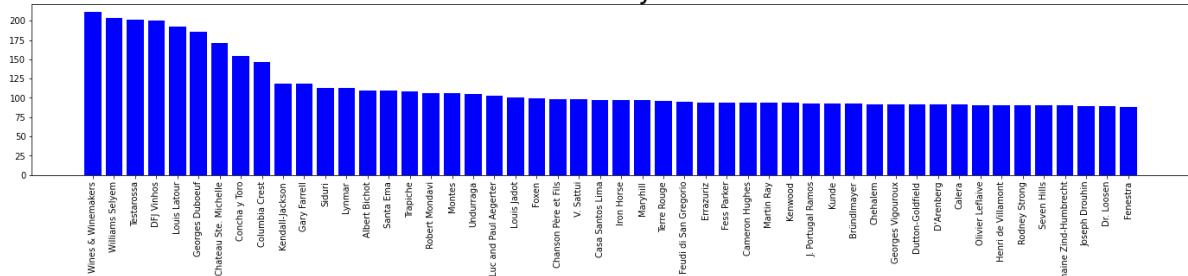
variety



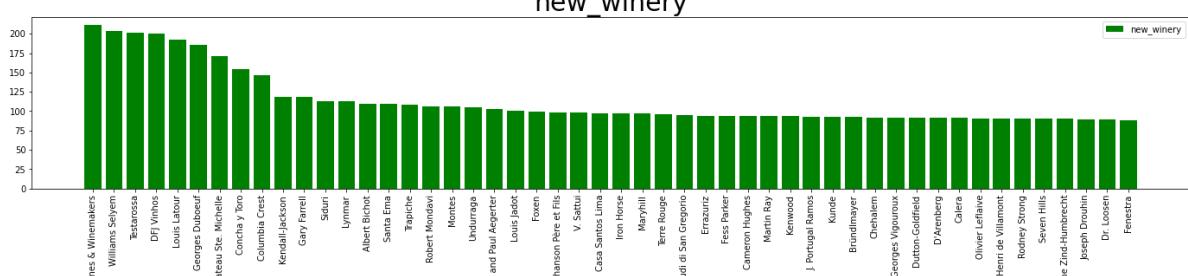
new_variety

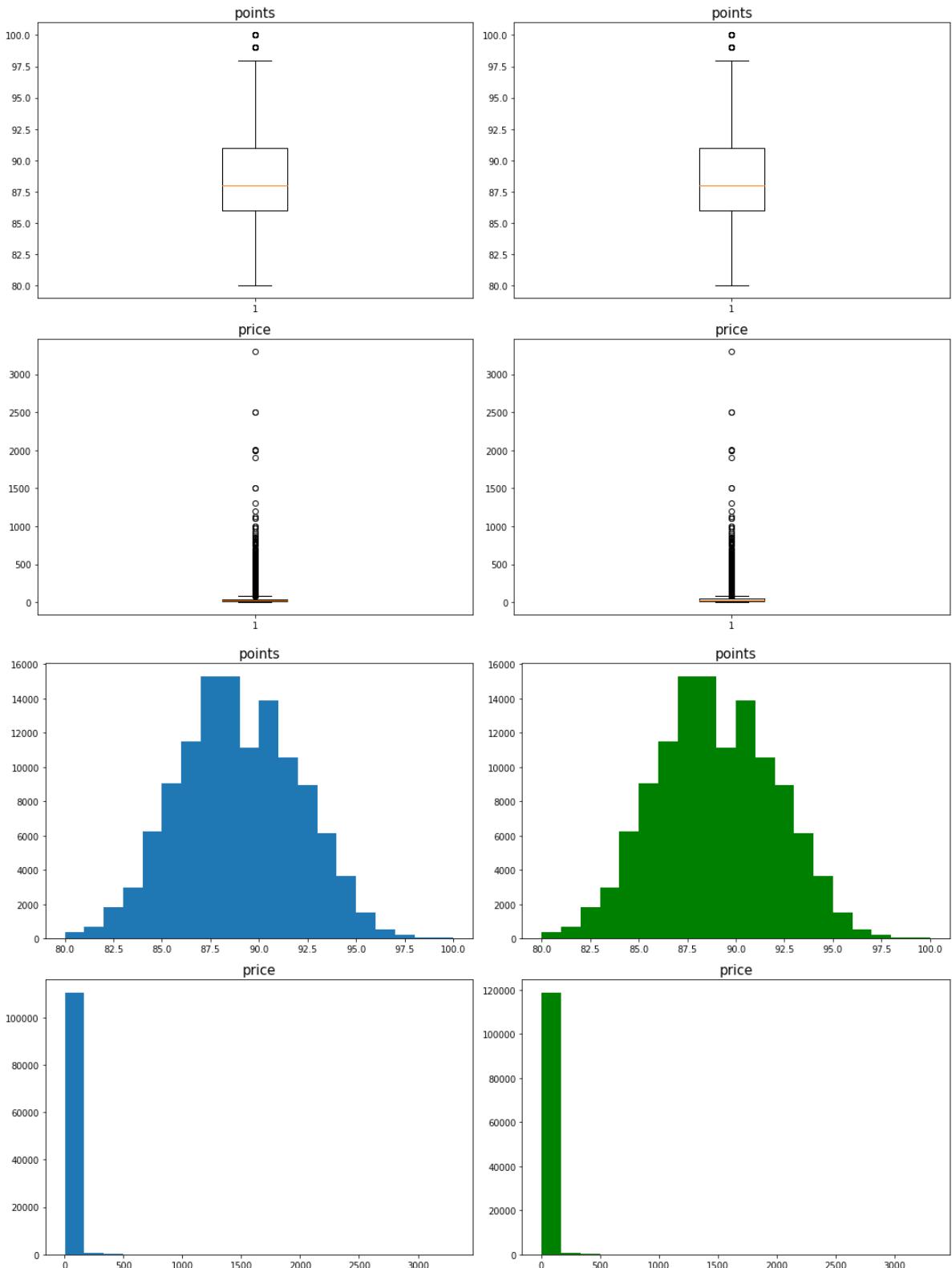


winery



new_winery





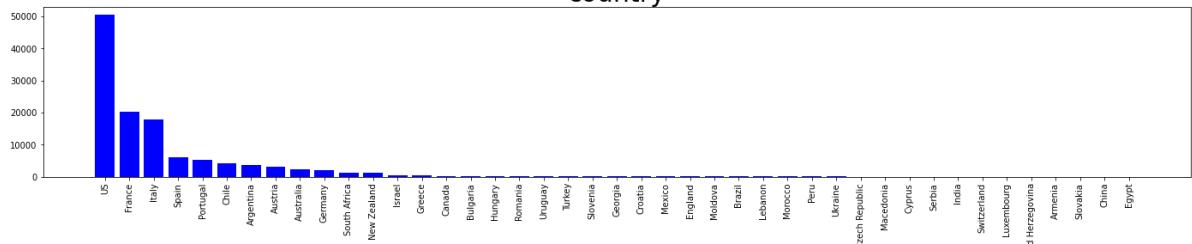
[分析]预测后的数据分布与原始数据差距不大，由此可知本方法预测合理

5.2.4 通过数据对象之间的相似性来填补缺失值

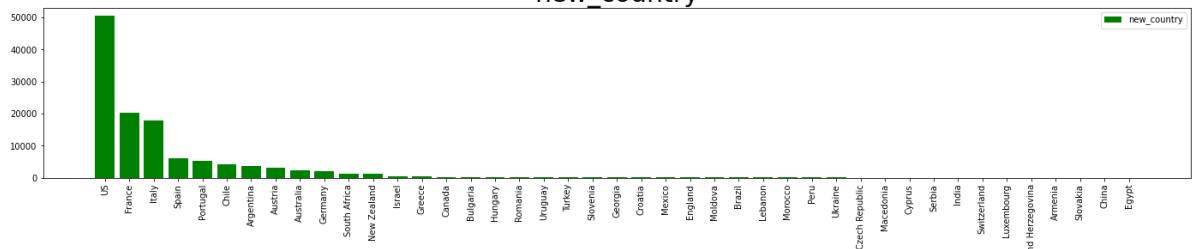
```
In [72]: new_df_wine2 = obj_corr_fill(df=df_wine2, numerical_index=numerical_index_wine2, k=1)
nominal_summary(df_wine2, nominal_index_wine2, new_df_wine2)
box_plot(df_wine2, numerical_index_wine2, new_df_wine2)
hist_plot(df_wine2, numerical_index_wine2, new_df_wine2)
```

```
E:\app\miniconda3\lib\site-packages\impyute\imputation\cs\fast_knn.py:113: RuntimeWarning: invalid value encountered in true_divide
    weights = distances/np.sum(distances)
```

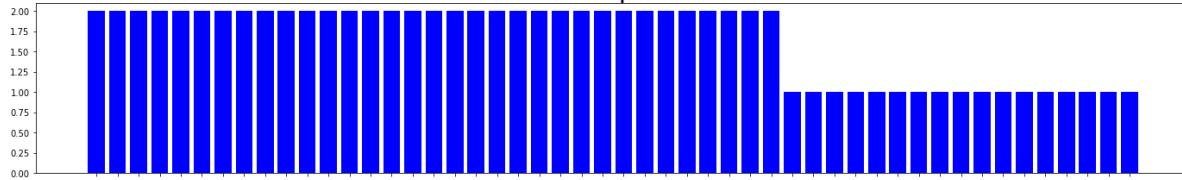
country



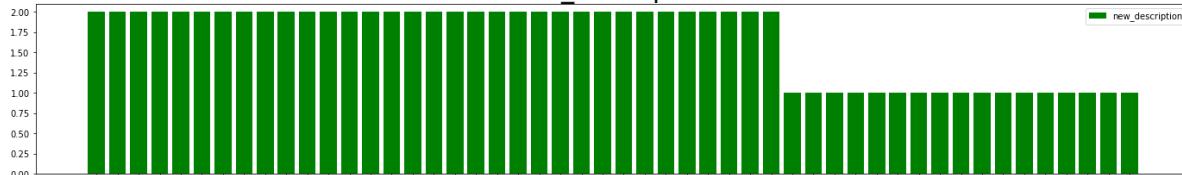
new_country



description



new_description



This opens with subtle aromas of wild flower and a whiff of exotic fruit. The bright fruity palate offers white peach, apple and a hint of honeydew melon. It's straightforward and refreshing. Lifted pear-drop aromatics are accompanied by the peppery green of watercress and arugula. The concentrated palate is grounded by earthy, savory yeast notes and framed by fresh lime peel. There is a dollop of ripe pear fruit at the core, shining and shimmering with added charm. The subdued nose eventually reveals black cherry, anise and a whiff of chopped herb that carry over to the straightforward, informal palate. Dusty, dry tannins and a note of bitter sage leave an astrigent finish.

Everything and enjoyable, this has aromas of mature black-skinned fruit, violet and an earthy whiff of truffle. The juicier sample offers blackberry fruit and a hint of baking alongside supple tannins. Drink soon.

This zesty red has pretty aromas that suggest small red berry, blue flower and a whiff of moist soil. The vibrant palate offers sour cherry, pomegranate and a hint of anise alongside zesty acidity and effeted tannins. Aromas of scorched soil, roasted hazelnut, dried black cherry and an earthy whiff of game carry over to the palate along with a note of bitter sage. Firm tannins provide support.

Aromas of citrus, melon and orange blossom start this crisp, elegant white from Cephelema. On the palate, bright flavors of lemon/lime and white peach are balanced with a stony minerality. A beautiful example of an ancient indigenous variety.

Cigar box, café au lait, and dried tobacco aromas are followed by coffee and cherries, with harelip spices lingering on the fresh. The wood gets a bit out front but it still delivers enjoyment.

Grenache apple, honeysuckle and jasmine aromas show on the relatively robust nose of this bottling from a large vineyard on Highway 46 east of Paso Robles. There is compellingly grippy texture to the spin with ripe flavors of pear and honeydew melon. A salty acidity takes to the next level.

Hugely delicious, just a joy to drink. So soft, and velvety, so rich in blackberry jam, cassis and chocolate flavors, with the tannins so smooth and round. Really wonderful over the next few years, although the softness and high alcohol suggest it's not an age-winer.

Ripe plum, grape, truffle, leather and menthol are some of the aromas you'll find on this earthy wine. The tightly wound palate offers dried blackberry, chopped sage, mint and toasted coffee bean alongside raspy tannins that leave a mouth-drying finish.

Sharp, simple and candied, with blackberry jam and cola flavors. The tannins are rugged, and the wine finishes with a scoop of acidity. Seems at its best now.

Subtle aromas of dark berry, lime, earth and baking spice lead the nose. The fine, juicy yellow notes (mustard, cassis, anise and black pepper alongside roud, solid tannins and black pepper) are here, and powered by a firm structure. This wine has weight and a fine Cabernet concentration.

Dark berry, toasted hazelnut, and methyl amyl ketone lift off the glass. The slightly round palate is still rather closed and the nose is earthy in terms of fruit richness, but slowly reveals sugar cherry, licorice and a note of ground pepper. A tannic backbone provides structure. The tannins are so fierce on this baby Cabernet that they effectively lock it down and make it tough to drink. I might effectively lock it down and make it tough to drink now, yet here's a fabulous core of blackberries that are rich, ripe and intense. Everything about it suggests ageability. Give it at least eight years in a proper cellar, and it could age even longer.

Vibrant and spicy, this savoury red delivers aromas and flavors of ripe blackberry, black cherry, ground pepper and chopped Mediterranean herb. Firm, fine-grained tannins support the juicy flavors. An anise note backs up the finish.

Seductively tart in lemon pit, cranberry and pomegranate, this refreshing, light-bodied quaff is infinitely enjoyable, both on its own or at the table. It continues to expand on the palate into an increasing array of fresh flavors, finishing in terrytory orange.

Very ripe and forward in baked fruit and pie-filling flavors of cherries and raspberries, spiced with cinnamon, anise and nutmeg. The wine is simple and ready to drink now.

Subtle aromas of ripe black cherry, ground pepper, raisin and a whiff of black flower. On the palate, finely woven tannins support the raspberry jam, black pepper, coca and a cinnamon note.

Subtle aromas of acacia flower and hints of mouthfeel with hemal flowers leading the way and hemal fruit running behind. Dry red fruit and herbal notes dominate the finish.

Aromas of weedy underbrush, toast, methol and cherry marmalade in spirits emerge in the glass. The palate shows raw sour cherry and a hint of rosemary alongside raspy tannins that dry out the finish.

Rose, red berry, mint and anise aromas unite in the glass. The firm palate offers wild cherry, crushed raspberry, clove, aniseote, and a mouth note named in tracking raspberry jam, fresh acidity. Drink 2020-2031.

Delicate aromas of honeysuckle and peat take shape in the glass. The pale yellow expressive, but eventually displays light notes of mature yellow apple and dried lemon zest. Drink soon to capture the semipung fruit.

Stalky aromas suggest hay and green herbs, with raspberry in the background. It's hot and short in moments of mouthfeel with hemal flowers leading the way and hemal fruit running behind. Dry red fruit and herbal notes dominate the finish.

This wine shows some of the less appealing aspects of American oak, overt spice, vanilla flavor and a slick, oily mouthfeel.

Atoms as raspberry and black pepper lay the groundwork for flavors of blackberry, eucalyptus, black cherry and thyme. Soft but ever present tannins germinate long after the bright cherry finish.

Aromas of lots of baking spices and touches of dried petals and ripe olallibeerries.

This wine, boosted by 8% Syrah, is relatively light in the glass, showing vanilla and plash red cherries on the nose. The nature palate offers tangine zest, citrus, apricot, anise, vanilla and mint together with bright acidity, a balanced complexity and depth.

A briny note of saline and pickle provide an intrigue to the entry of this wine, which is otherwise integrated in oak and medium to full-bodied. Smooth and floral, it hangs on the palate, revealing baked apple and pear around a hint of wild vanilla and herb.

The nose on this amber-colored wine is initially shy but eventually reveals ripe stone fruit, crushed rock, spice and a minty note. The nature palate offers tangine zest, citrus, apricot, anise, vanilla and mint together with bright acidity, a balanced complexity and depth.

Raisin and molasses aromas and flavors are a big part of this full-bodied, syrupy number from Cayelaine in Salta. Additional scents of horseradish, herbs and earthy berry add to an overbearing sense of ripeness in the mouth, this is heavy, with weighty tannins. Flavors of prune and K. Hersey precede coffee, Golden Delicious apples, mature apricot, oak-driven spice and candied K. Hersey alongside tangy acidity.

A briny vein of cranberry and raspberry runs from the nose to the palate, and lingers on the finish, giving this wine a punchy focus and vibrance. It has plenty of fresh acidity, fine tannins and good depth of flavor.

Aromas of blueberry, cherry and minty on the palate to flavors of cherry, blackberry, eucalyptus and black pepper. Smooth tannins appear early on and endure into a pleasing finish that is interspersed with bright and spicy notes.

The freshness of seared Meyer lemon and squeezed lime peel meets with the warmth of vanilla and anise in the nose of this solid bottling. The palate offers white peach and smoked apples laid across a very tense and pithy texture. It finishes on a Juicy Fruit bubblegum element that's intriguing and nice overdone.

This structured blend of Chardonnay, Sauvignon, Roulle and Moscato Giallo offers aromas of toasted oak, ginger and orchard fruit. The elegant palate combines Golden Delicious apples, mature apricot, oak-driven spice and candied K. Hersey alongside tangy acidity.

Produced from organic grapes, this is a smooth, lightly toasty wine. It has ripe apple and a crisp, tight, minty texture that adds to the interesting character of the wine. The aftertaste is a little fresh fruit.

A briny vein of cranberry and raspberry elements that decorate the dark fruits and pomegranate fruits. The palate is much like a fruit-faced black tea.

Stale aromas of grapefruit, citrus zest and kiwi float out of the glass. The radiant palate is more expressive, delving out juicy yellow peach, honeydew melon, nectarine and sage. A mineral note backs up the finish.

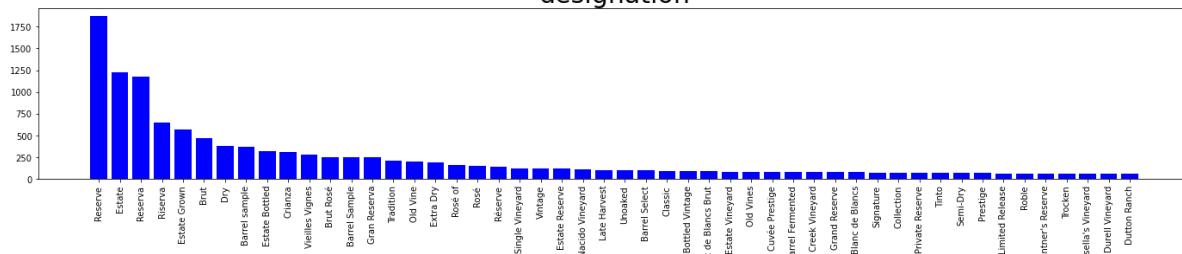
Dark and earthy on the nose, this bottling shows sage, menthol, black tea and forest-floor elements that distract the palate. A hint of coquille arrives toward the finish in a burst of eucalyptus, but restrained tannins and floral notes of bright fruit makes for a satisfying experience.

Everything about this wine is big and heavy, including the name, the pyriformized lid and the massive bottle with wax capsule, built-in at that mass. It tastes great. The color is deep and dark, aroma is full of baby spices of nutmeg and vanilla, and citrus zest make an impression on the taste buds welling the enduring finish.

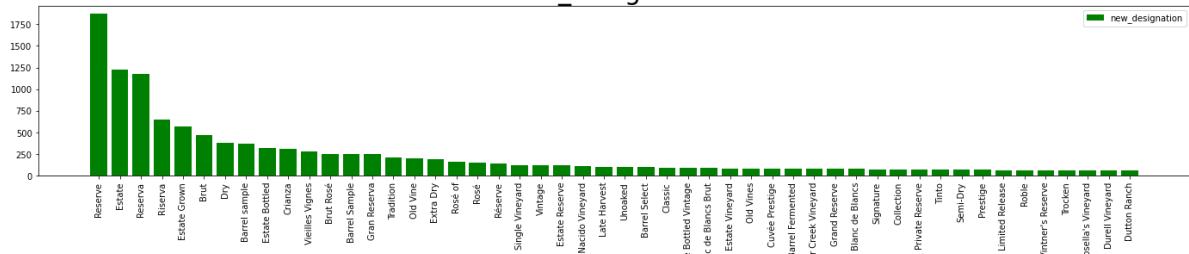
Deep ruby in the glass, this aptly-named wine has a nose of black cherry and aniseed. Slightly grippy tannins and strong acidity stand at the core of blueberry, raspberry, aniseete and peppermint flavors. Bunches of mint and orange zest make an impression on the taste buds welling the enduring finish.

Abacela makes a fortified wine from Portuguese grapes, but this is a dry, varietal expression of one of the most obscure. Bone-dry, it's got tongue-scraping tannins that taste of black tea. It offers dried raspberries, vanilla and cocoa elements, and the overall complexity is impressive. This seems built for a thick steak.

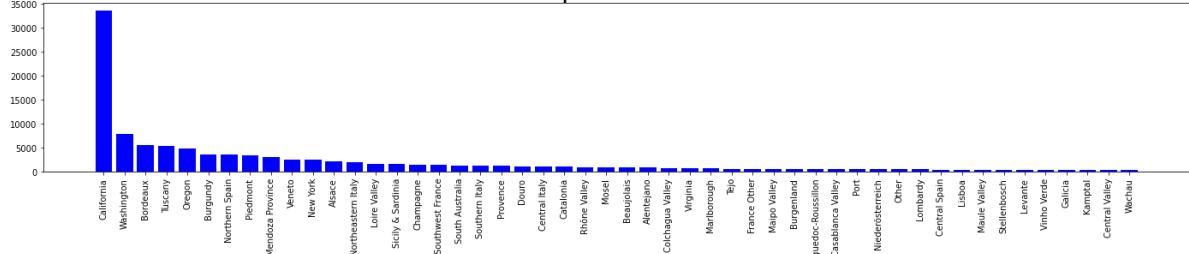
designation



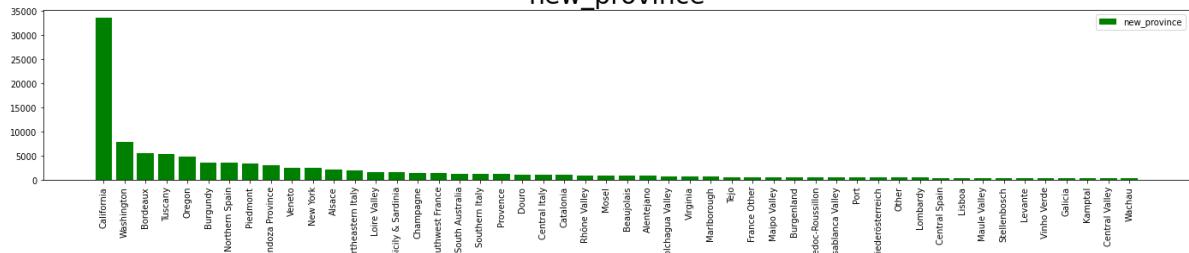
new_designation



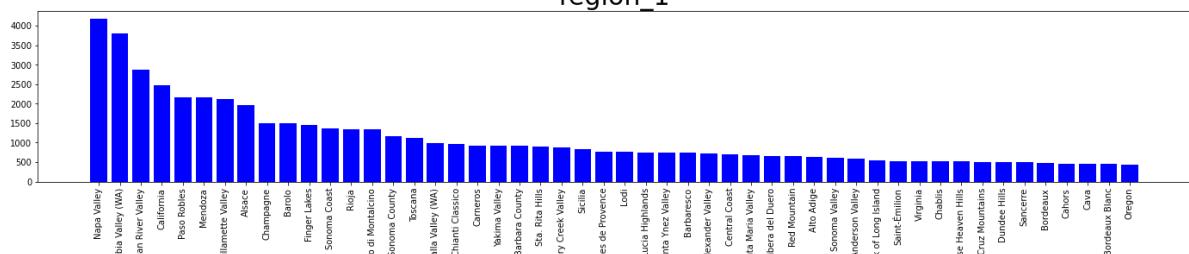
province



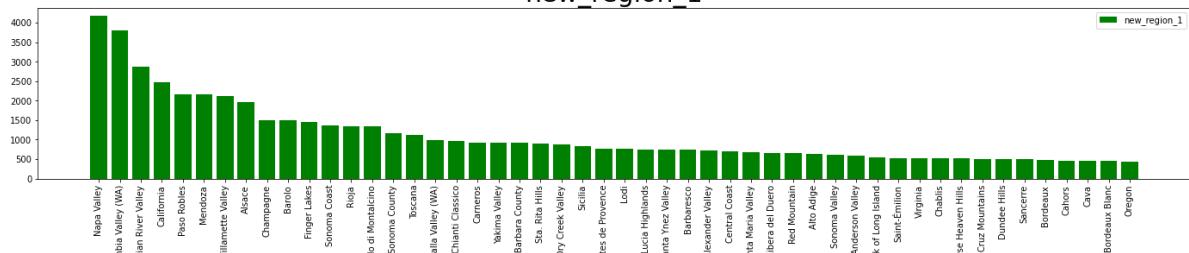
new_province



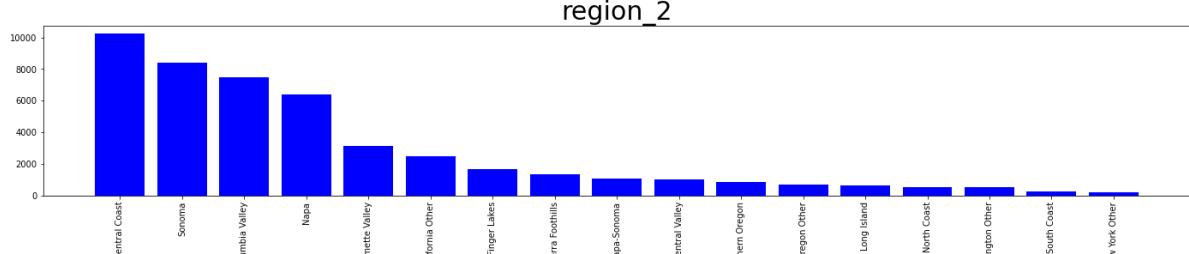
region_1



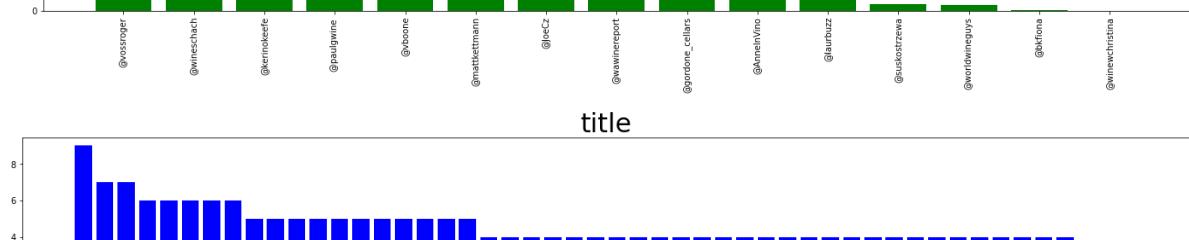
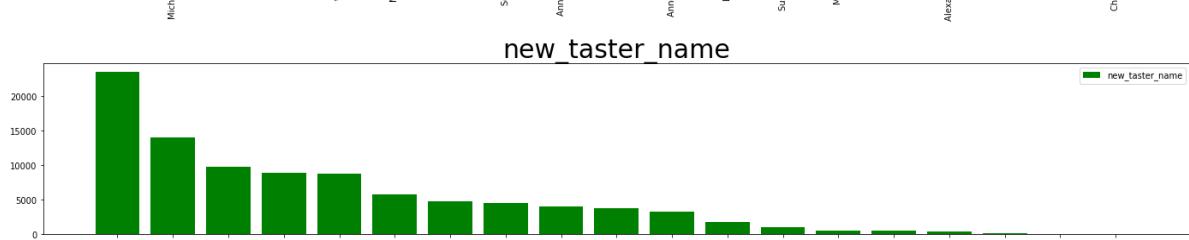
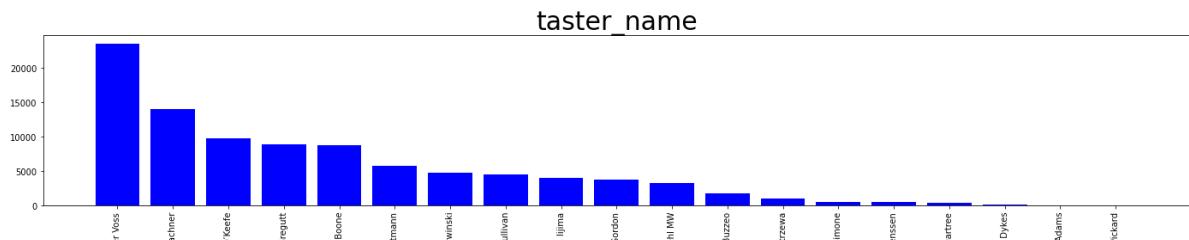
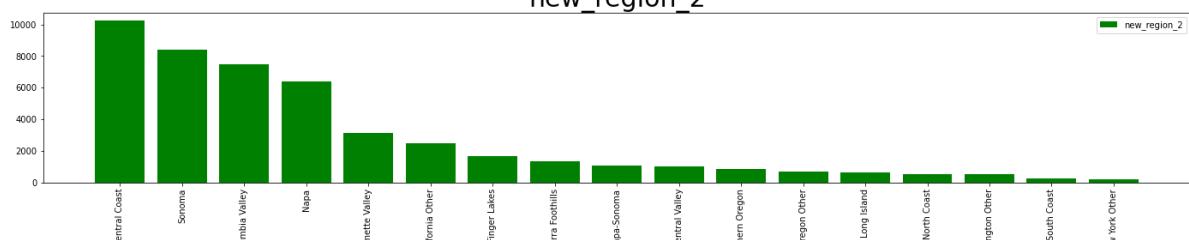
new_region_1



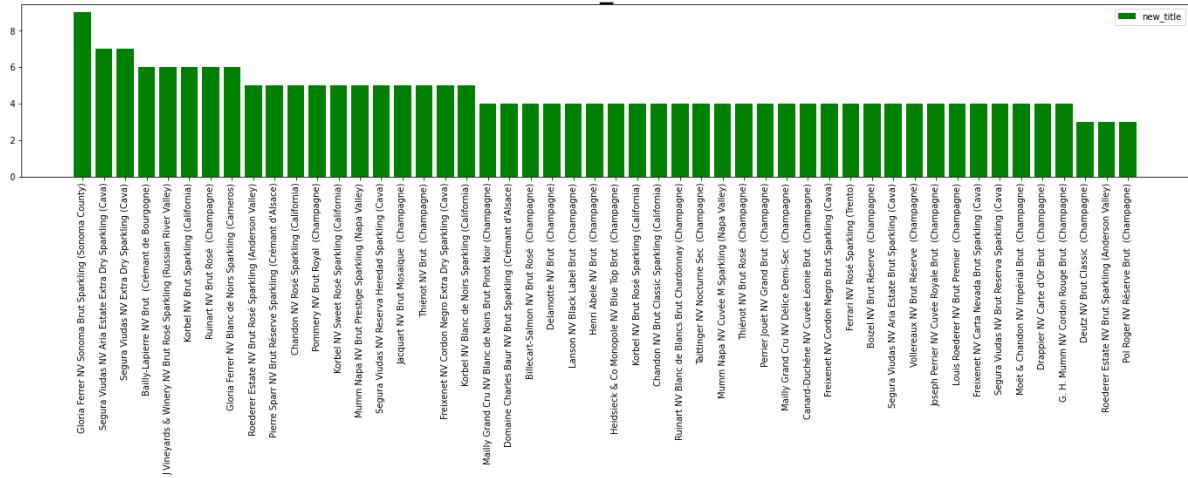
region_2



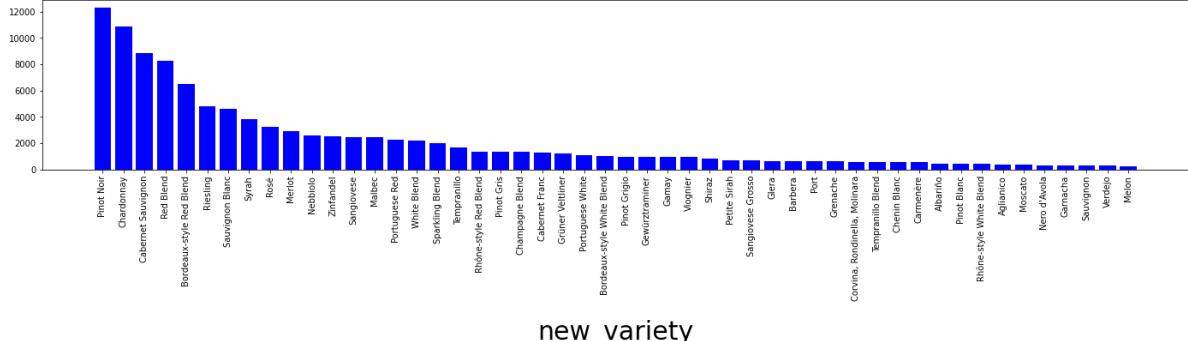
datamining week3



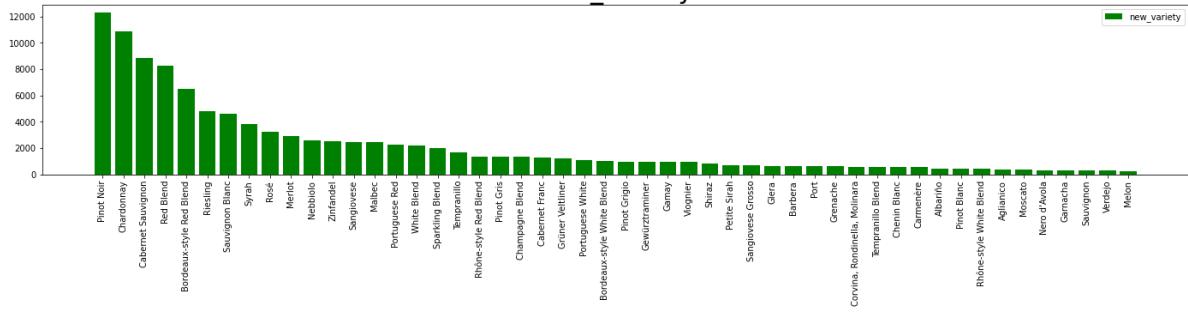
new_title



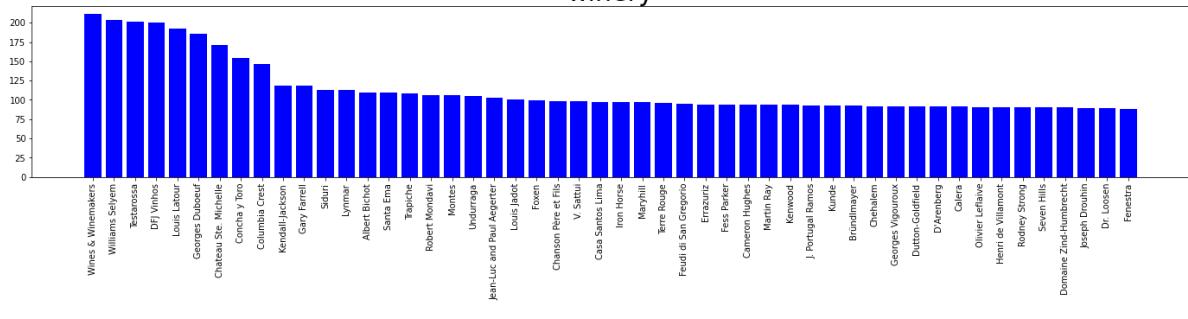
variety



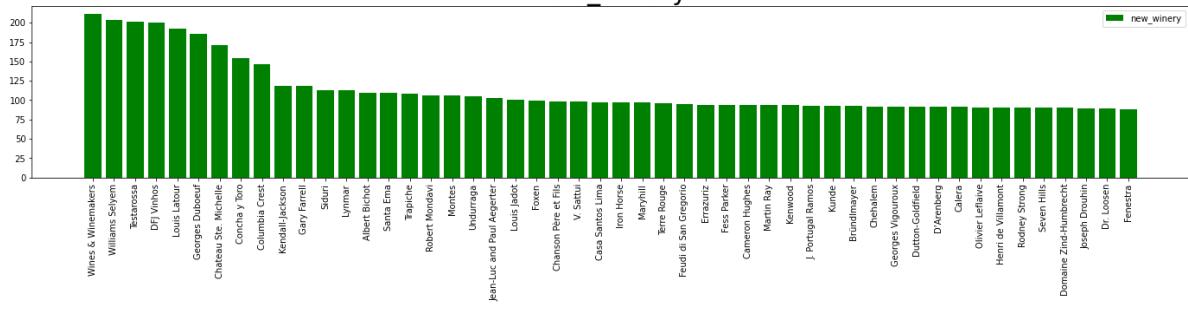
new_variety

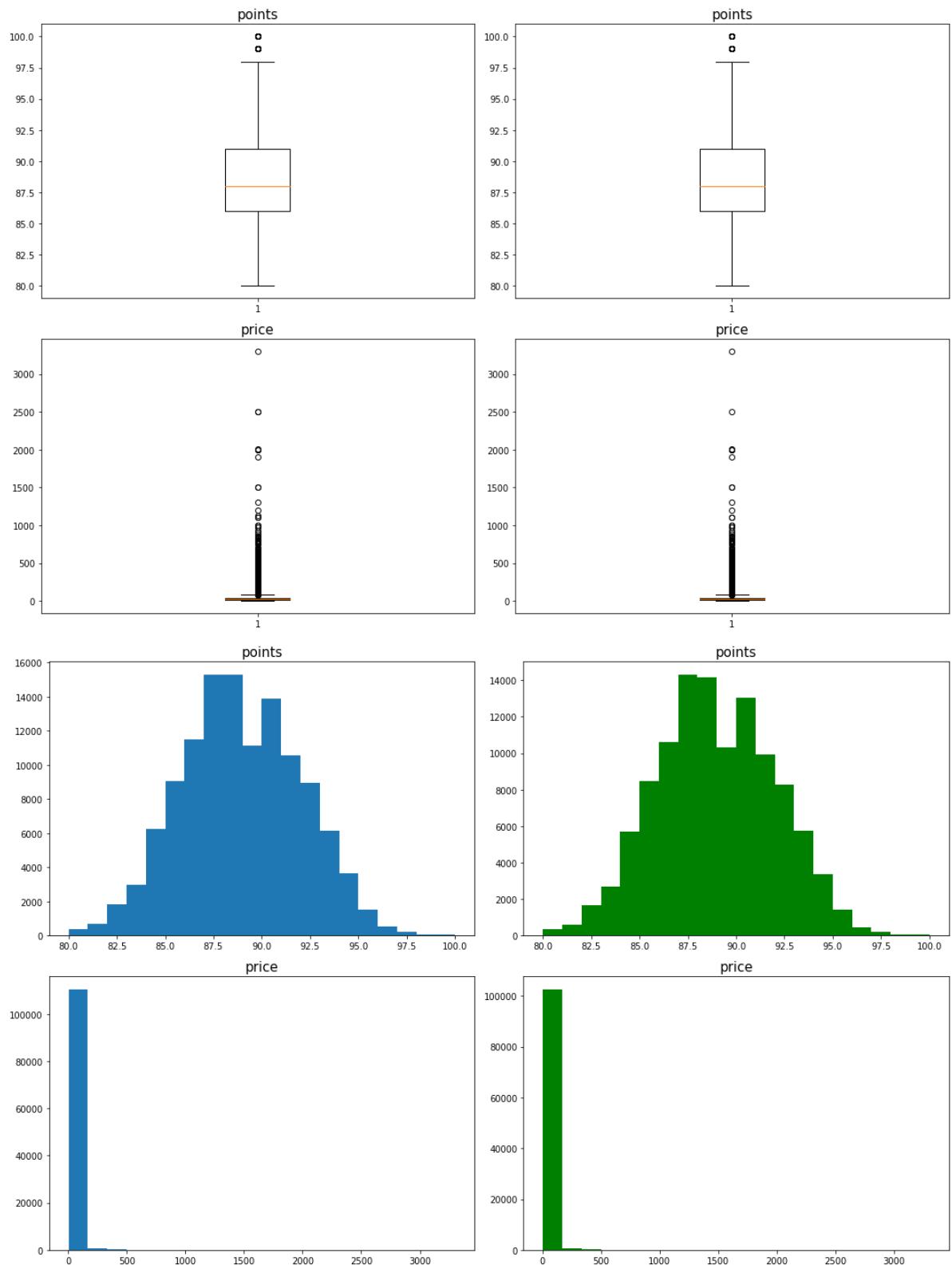


new_variety



new_winery





[分析]预测后的数据分布与原始数据差距不大，由此可知本方法预测合理