

互评作业1: 数据探索性分析与数据预处理

1.实验环境

语言: python3
依赖的包: pandas, matplotlib, scipy, numpy

2.数据说明

dataset1: wine-reviews, 来源<https://www.kaggle.com/zynicide/wine-reviews>
包含两个csv文件:
 winemag-data-130k-v2.csv: 包含14个属性 (3个数值属性, 11个标称属性), 129970条数据记录;
 winemag-data_first150k.csv: 包含11个属性 (3个数值属性, 8个标称属性), 150930条数据记录
dataset2: Consumer & Visitor Insights For Neighborhoods来源<https://www.kaggle.com/safegraph/visit-patterns-by-census-block-group>
包含一个csv文件:
 cbg_patterns.csv: 包含13个属性 (6个数值属性, 7个标称属性), 220736条数据记录

3.数据分析

3.1 数据摘要

数据集1中winemag-data-130k-v2.csv数据包含:
3个数值属性: Uname、points、price;
11个标称属性: country、description、designation、province、region_1、region_2、taster_name、taster_twitter_handle、title、variety、winery;
winemag-data_first150k.csv数据包含:
3个数值属性: Uname、points、price;
8个标称属性: country、description、designation、province、region_1、region_2、variety、winery;
数据集2中cbg_patterns.csv数据包含:
6个数值属性: census_block_group、date_range_start、date_range_end、raw_visit_count、raw_visitor_count、distance_from_home;
6个标称属性: visitor_home_cbgs、visitor_work_cbgs、related_same_day_brand、related_same_month_brand、top_brands、popularity_by_hour、popularity_by_day。
为了方便数据的读取和处理,使用pandas库对数据集进行读取和处理。判断属性的类型之后,使用pandas库中封装的函数获得标称属性的取值频数和数值属性的最大值、最小值、中位数、均值、四分位数及缺失值的个数,程序及结果如下所示:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats
pd.set_option('display.float_format', lambda x: '%.1f' % x, 'max_colwidth',100)

dataload1 = pd.read_csv('winemag-data-130k-v2.csv')#读取表1中的数据
dic1 = {'attribute':[], 'max':[], 'min':[], 'mean':[], 'median':[], 'quantile':[], 'num_NA':[]}#属性, 最大值、最小值、中位数、均值、四分位数及缺失值的个数
count1 = []#频数
for key in dataload1.keys():
    d = dataload1[key]
    if np.issubdtype(d.dtypes,np.number):#判断属性类型
        l = [key,d.max(),d.min(),d.mean(),d.median(),
(d.quantile(0.25),d.quantile(0.5),d.quantile(0.75)),sum(d.isnull())]#统计数值属性的最大、最小、均值、中位数、四分位数及缺失值的个数
        for k,i in zip(dic1.keys(),l):
            dic1[k].append(i)
    else:
        count1.append(d.value_counts())#统计标称属性取值的频数
```

```

dataload2 = pd.read_csv('winemag-data_first150k.csv')#读取表2中的数据
dic2 = {'attribute':[], 'max':[], 'min':[], 'mean':[], 'median':[], 'quantile':[], 'miss_NA':[]}
count2 = []
for key in dataload2.keys():
    d = dataload2[key]
    if np.issubdtype(d.dtypes,np.number):
        l = [key,d.max(),d.min(),d.mean(),d.median(),
(d.quantile(0.25),d.quantile(0.5),d.quantile(0.75)),sum(d.isnull())]#统计数值属性的最大、最小、均值、中位数、四分位数及缺失值的个数
        for k,i in zip(dic2.keys(),1):
            dic2[k].append(i)
    else:
        count2.append(d.value_counts())#统计标称属性取值的频数

dataload3 = pd.read_csv('cbg_patterns.csv')#读取数据集2中的数据
dataload3 =
dataload3.drop(['date_range_start','date_range_end','visitor_work_cbgs','visitor_home_cbgs'],axis=1)
dic3 = {'attribute':[], 'max':[], 'min':[], 'mean':[], 'median':[], 'quantile':[], 'miss_NA':[]}
count3 = []
for key in dataload3.keys():
    d = dataload3[key]
    if np.issubdtype(d.dtypes,np.number):
        l = [key,d.max(),d.min(),d.mean(),d.median(),
(d.quantile(0.25),d.quantile(0.5),d.quantile(0.75)),sum(d.isnull())]#统计数值属性的最大、最小、均值、中位数、四分位数及缺失值的个数
        for k,i in zip(dic3.keys(),1):
            dic3[k].append(i)
    else:
        count3.append(d.value_counts())#统计标称属性取值的频数

```

数据集1中winemag-data-130k-v2.csv数值属性的最大值、最小值、中位数、均值、四分位数及缺失值的个数如下表所示：

```
pd.DataFrame(dic1)
```

	attribute	max	mean	median	min	num_NA	quantile
0	Unnamed: 0	129970.0	64985.0	64985.0	0.0	0	(32492.5, 64985.0, 97477.5)
1	points	100.0	88.4	88.0	80.0	0	(86.0, 88.0, 91.0)
2	price	3300.0	35.4	25.0	4.0	8996	(17.0, 25.0, 42.0)

数据集1中winemag-data_first150k数值属性的最大值、最小值、中位数、均值、四分位数及缺失值的个数如下表所示：

```
pd.DataFrame(dic2)
```

	attribute	max	mean	median	min	missvalue	quantile
0	Unnamed: 0	150929.0	75464.5	75464.5	0.0	0	(37732.25, 75464.5, 113196.75)
1	points	100.0	87.9	88.0	80.0	0	(86.0, 88.0, 90.0)
2	price	2300.0	33.1	24.0	4.0	13695	(16.0, 24.0, 40.0)

数据集2cbg_patterns.csv中数值属性的最大值、最小值、中位数、均值、四分位数及缺失值的个数如下表所示：

```
pd.DataFrame(dic3)
```

	attribute	max	mean	median	min	miss_NA	quantile
0	census_block_group	780309900000.0	287086382055.2	290190012013.5	10010201001.0	1	(131210116244.2, 290190012013.5, 420034897521.7)
1	raw_visit_count	7179900.0	47930.7	30640.0	60.0	106	(17042.0, 30640.0, 56678.0)
2	raw_visitor_count	6113949.0	11820.3	6541.0	50.0	106	(3430.0, 6541.0, 13099.0)
3	distance_from_home	6297845.0	35112.8	14614.0	706.0	217	(8584.0, 14614.0, 31397.75)

数据集1中winemag-data-130k-v2标称属性的取值频数。例如，对于country属性来说，出现频率最高的是US（54504次），其次是France（22903次）、Italy（19540次）。程序及运行结果如下：

```
count1
```

```
[US                54504
 France            22093
 Italy             19540
 Spain             6645
 Portugal          5691
 Chile             4472
 Argentina         3800
 Austria           3345
 Australia         2329
 Germany           2165
 New Zealand       1419
 South Africa      1401
 Israel            505
 Greece            466
 Canada            257
 Hungary           146
 Bulgaria          141
 Romania           120
 Uruguay           109
 Turkey            90
 Slovenia           87
 Georgia           86
 England           74
 Croatia           73
 Mexico            70
 Moldova           59
 Brazil            52
 Lebanon           35
 Morocco           28
 Peru              16
 Ukraine           14
 Macedonia         12
 Serbia            12
 Czech Republic    12
 Cyprus            11]
```

India	9
Switzerland	7
Luxembourg	6
Armenia	2
Bosnia and Herzegovina	2
China	1
Egypt	1
Slovakia	1

Name: country, dtype: int64,

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

3

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

3

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 cherry and orange.

3

Cigar box, café 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.

3

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.

3

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 refined tannins.

3

The first release annually from the estate vineyard, this is a solid effort, seemingly infused with scents and flavors of pine needle, pine tar, earth and bark. The fruit is in the background, and leans toward cranberry and brambly blackberry.

2

A nice Chardonnay that shows ripe pineapple, green apple and pear flavors, with added smoky richness from oak, and the finish is dry and crisp. A good price for a wine of this quality.

2

This substantial wine is full bodied, robust in flavor and firmly textured due to ample fine-grained tannins. The fruit flavors are tempting and ripe, while accents of black pepper and dried herbs add complexity.

2

This broad-textured, complex and indulgent Rhône-style blend has cedar, cola and blackberry aromas and concentrated fruit, earth and spice flavors. Milk chocolate richness emerges on the midpalate and plays through the finish.

2

Full-bodied and firmly structured, this offers aromas of dark cherry, mocha, blue flower and a whiff of exotic spice. The concentrated palate doles out black cherry, raspberry jam, raisin, chocolate and baking spice alongside velvety tannins. Drink 2017–2023.

2

Dr. L has developed into the sort of consistent, large-production wine consumers can depend on when faced with an otherwise unknown array of wines. The 2009 is off dry, with crisp acidity, lime and pineapple flavors and a mouthwatering finish. Drink it over the next year or so.

2

Aromas of cured meat, stewed fruit and toasted oak lead the nose. On the ripe palate, notes of powdered sage, vanilla and oak spices accent fleshy black plum. Bracing tannins provide the framework, but drink soon to capture the fruit richness.

2

Though labeled Columbia Valley, this is in fact 100% Merlot from the estate vineyards on Red Mountain. It's chewy and tannic, as is often the case with Red Mountain reds, with black cherry and cassis fruit wrapped by astringent tannins hinting at dark chocolate. In short, a fine steak wine, that needs some fat to cut the tannins.

2

Freixenet's signature Cava is consistent and pure as usual, with aromas of white fruit and white pepper and an overall freshness on the nose. It feels crisp and focused, with a foamy palate that brings apple, lime, peach and spice flavors. The finish is smooth.

2

Exotic hints of dill and clove open on the nose in this elegantly robust wine, which finds a balance in its brightness of red cherry and red currant. Generous tannin takes on a leathery feel on the palate, finding a softness that lingers on the finish.

2

Made from 100% Sangiovese, this opens with a lovely floral fragrance of violet and iris. Ripe sour cherry dominates the palate accented by toasted notes and black pepper. Even though it's not complex, the juicy fruit and smooth tannins make this friendly and delicious.

2

With moderate alcohol, this Pinot is silky and delicate in the mouth, with excellent acidity. Yet it's intense in sunshiney-ripe raspberries and cherries, as well as a hint of mushrooms that will become more pronounced with age. Defines the lighter style of California Pinot, in a gracious, elegant style.

2

Forward and fruity, this is a solid, drink any time Chardonnay. It's built in a blocky chunky mode, but some sleek lime flavors put a polished edge on the underlying acids.

2

This wine opens with funky aromas of tire rubber and nail polish remover that carry over to the palate along with toast and espresso notes. It lacks fruit richness and has a green, bitter finish.

2

This is a bright lemon and grapefruit flavored wine that has lively acidity and the lightest touch of mineral texture. It's crisp, refreshing and will be ready to drink from mid-2016.

2

This wine comes out of the bottle a little cranky, but give it some time in the glass and the initial stink gives way to notes of lime, apple and peach. It's plump in feel, with some dusty minerality and dried spices adding nuance to the sweet fruit.

2

Dry and acidic. If you're looking for a rich, opulent Chardonnay, go elsewhere. This one's citrusy and minerally, in the manner of an elegant Chablis. It offers a useful alternative to the modern, ultraripe, oaky style.

2

There's a slight bit of pleasant funk to the nose of this appellation blend, as well as graphite, sour cherry, soy and shiitake mushroom aromas. The palate is wound around tight acidity, with wild raspberry flavors lifted by incense and pine forest elements.

2

Fig, black plum, dried cherry and chai spice show on the nose of this bottling from River Road south of Salinas. Heavy flavors of dried black cherry and dried cranberry show on the sip, lifted by eucalyptus oil, juniper and black-tea elements.

2

A crisp and fruity blend of the herbaceous Sauvignon Blanc and the creamy Loin de l'Oeil, this has apple and grapefruit flavors. It has a bright feel, with a hint of mineral on the finish.

2

Tropical aromas revolve around lemon-lime. This is acidic without much body. Flavors of lemon, melon, mango and green apple are lean and finish with tangy, pithy citrus flavors.

2

The top-level bottling from the ambitious, sprawling property originally planted by the Mondavis, this vintage-varying blend of 48% Cabernet Sauvignon, 44% Merlot, 4% Petite Sirah and 4% Malbec offers licorice, German chocolate cake, blackberry jam, violet extract and night jasmine on the detailed nose. It's visceral once sipped, with thick jammy blueberry and plum fruits, bloody meats, leather, dark chocolate and coffee grinds.

2

Ambrusco is a softly layered and rich wine that imparts opulent aromas of chocolate, sweet spice, dark fruit, plum, prune and licorice. There's a touch of bitterness in the mouth but otherwise the mouthfeel is thick and dense enough to smooth out any wrinkles.

2

The normally so rich Rosengartl comes in a much more slender body for the 2014 vintage, with zesty relish even. Here and there are glimpses of opulence, in the shadows of pleasantly bitter honeycomb that also contributes to a beautifully defined, linear texture on the midpalate. The finish is whistle-clean. Bravo.

2

..

Dominated by Merlot, this is a fruity wine with finely integrated tannins. It has the red fruits and acidity of the vintage given weight by 13 months of wood aging and the concentration of old vines. Drink from 2019.

1

It doesn't say so on the label, but the grapes come from the estate vineyard in Oak Knoll. It's a moderately cool AVA already, so combined with the chilly vintage, a dry wine with considerable tannins was produced. They're old-fashioned in their astringency, but the ultrarich black currant and dark chocolate flavors, boosted by high alcohol, are thoroughly modern. Made up of Malbec, Cabernet Sauvignon and Petit Verdot. Hold until at least 2018.

1

This Elqui SB smells about as neutral as you'll come across, with a hint of burnt grass to the nose. Oily and dense in feel, this tastes of salt and bitter dried melon.

1

Hints of ripe apricot and ethereal notes of crushed boxwood add interest to the more usual aromatic spectrum of green pear and white pepper. A lovely phenolic grip provides extra dimension and calls for food. Rounded, intriguing but still very refreshing.

1

Lots of forward, ripe blackberry and cherry fruit in this wine, with oak adding richer notes of smoky vanilla and caramel. With smooth tannins and a dry finish, it provides a lot more pleasure than you'd expect at this price.

1

Unusually dry and tangy, this light-bodied wine smells like celery and herbs, and tastes very lean and straightforward.

1

Spicy and smoky with notes of dark red berries, rich red plum guts, tobacco and cocoa on the nose and palate. The finish is long and is loaded with roasted coffee beans, cocoa, and finally a bacony sweetness at the close. Balanced and well-integrated with firm but not overly assertive tannins. Drink now.

1

Mango, presumably, is the dog pictured on the label. This is a heavily-extracted, somewhat bitter Merlot, with fruit flavors of blackberry syrup, alcohol at 15%, and still a fair amount of green bitterness to the tannins. Notes of iodine and anise come into play in the finish.

1

This is a smooth, easy-drinking, chocolaty Cabernet. The ripe red fruit is laced with streaks of black olive, mushroom and loam. Barrel aging in both American and French oak (half new) brings in coffee and chocolate.

1

Soft and simple, with ultraripe blackberry and cherry flavors that veer into currants, dark chocolate and raisins, this Merlot has the benefit of being bone dry, with smooth, fine tannins. Drink now.

1

Ripe yellow pear with the faintest suggestion of something tropical, like passion fruit, leads to a full-fruited, generous palate that is kept in check by structure and zesty acidity.

1

Here's a very pretty white wine from Sicily (made with Grillo grapes) that opens with creamy, light nuances of almond blossom, stone fruit and citrus. Pair this wine with steamed white fish.

1

A pretty pale salmon color is matched by aromas of strawberry and herbs that are light and delicate. The palate brings the suggestion of sugar with an elegant styling.

1

Both cold-tank fermented and barrel-fermented, this white shows steely aromas of lemon and fresh lime, its acidity bristling with mouthwatering succulence and a lightness of body.

1

An easy-pleasing and refreshing wine with ripe aromas of pineapple and guava alongside a citrus-infused grassy note, almost like fresh hops. Lightweight in the mouth with a brisk finish reminiscent of fleshy green apple slices.

1

The wine is initially soft and then begins to show a more steely, mineral character. That gives structure as well as citrus and crisp apple skin flavors. Very tight at this stage, it needs to develop, so give it until 2015.

1

Too sharp in texture, with a cut of acidity that emphasizes the white pepper, and seems to bring out some greenness. Other than that, you'll find a smoothly tannic wine, with blackberry and currant fruit.

1

Earthy aromas include forest floor, black berries, leather and spice. The palate is less expressive and has a green character, showing unripe red cherry alongside astringent tannins that give it an abrupt, bitter close.

1

Blackberry and cassis aromas are bold and come with a coating of oak that brings vanilla and sweet plastic into the frame. A crisp, tightly wound palate shows plum and red-currant flavors in front of a snappy finish with oak spice and chocolate hints.

1

Balsamic aromas of menthol and eucalyptus unfold in the glass and mingle with scents of leather, licorice and dark berry. The concentrated palate delivers dried black cherry, crushed raspberry, black pepper and grilled herb alongside velvety tannins. Drink 2015-2022.

1

Marrenon's luxury cuvée aged in oak is 90% Grenache and 10% Syrah and is full-bodied without being tough or overly muscular. Cherry preserves are wrapped in a coat of coffee- and chocolate-tinged oak. Drink now and over the next couple of years.

1

This wine develops slowly in the mouth, providing a finely structured, rich and elegant character. There's a round white-fruit flavor that's shot through with taut acidity. With its tight finish, it could age a few more months.

1

Tender notes of ripe, juicy pear with more golden inflections of ripe yellow plum and peach on nose and palate close into a central seam of bright, sprightly and immensely toning acidity. Lightness, concentration and lip-smacking freshness are framed by rounded fruit. Structure and grace signal ageing potential. The finish is wonderfully dry and full of verve.

1

Bright aromas of ripe golden apple, Bartlett pear and honeydew melon carry over to the linear palate. Vibrant acidity gives this clean, refreshing finish.

1

A lightweight, dusty-smelling Pinot Noir from near Sancerre. The fruit is very soft, with the dryness giving a final firmness to the aftertaste. Flavors of red cherries have a dusty, perfumed character.

1

A relatively dry kabinett, with slightly nutty scents of vegetable oil followed by flavors of peach pit and underripe melon. But the sum of the parts is actually better than they sound alone, offering fresh, lemony refreshment with mineral underpinnings. Drink now.

1

Some oak aging shows in the smoothness of the wine, which has good concentration as well as young fruitiness. It is tangy with a mineral texture and richness that come from the lees aging in barrel. The wine should mature a little longer, so drink from 2017.

1

Barley sugar and orange peel vie for attention, but both play to the tune set by the fine, distinct lemony acidity, which also provides linearity. Hints of lemon oil and honey complete the picture.

1

Freshly picked roses, violets and woodland berry aromas come together on the nose. The tense palate delivers wild cherry, black raspberry, licorice, ground pepper and tobacco alongside a firm backbone of tightly woven, ultrafine tannins. It's a gorgeous combination of finesse and structure but at the moment it's still tightly tannic and needs time to unwind. Drink 2018-2027.

1

Opening with a pretty honeysuckle nose, this Viognier is completely in balance, with layers of lime and pineapple and restrained weight. Enjoy with Thai papaya salad or ceviche.

1

Name: description, dtype: int64,	
Reserve	2009
Estate	1322
Reserva	1259
Riserva	698
Estate Grown	621
Brut	513
Dry	413
Barrel sample	375
Crianza	343
Estate Bottled	342

Vieilles Vignes	308
Brut Rosé	276
Gran Reserva	261
Barrel Sample	252
Tradition	238
Old Vine	221
Extra Dry	204
Rosé of	172
Rosé	166
Réserve	153
Single Vineyard	131
Vintage	128
Estate Reserve	125
Bien Nacido Vineyard	120
Late Harvest	112
Classic	111
Unoaked	110
Barrel Select	106
Late Bottled Vintage	99
Blanc de Blancs Brut	96

...

La Renardière	1
Self Ranch	1
EBB Extra Brut	1
Afterglow Dessert	1
Kalkmergel Spätlese Trocken	1
Kollmutz Federspiel	1
Eros	1
Scalunera	1
Porer	1
Coste d'Aleyrac Pic Saint-Loup	1
Les Barres	1
Muscadelle	1
Deux Chevaux	1
South Ridge Cuvée Merret Brut	1
Smoke Vineyard	1
Valdisanti	1
Saulheimer Kalkstein Unfiltriert Trocken	1
Butcher's Block	1
The Augur Kick On Ranch Vineyard	1
Vigna Roreto	1
Liebelt Vineyards	1
Rotas de Portugal	1
Selvanova	1
O Grand R	1
Tinto Lena	1
Blanc de Rosé Extra Brut	1
Foz Torto	1
Dal Tralcetto	1
Albero Bobal	1
Lot 287	1

Name: designation, dtype: int64,	
California	36247
washington	8639
Bordeaux	5941
Tuscany	5897
Oregon	5373
Burgundy	3980
Northern Spain	3851
Piedmont	3729
Mendoza Province	3264
Veneto	2716
New York	2688
Alsace	2440
Northeastern Italy	2138
Loire Valley	1856

Sicily & sardinia	1797
Champagne	1613
Southwest France	1503
Southern Italy	1349
South Australia	1349
Provence	1346
Douro	1281
Central Italy	1233
Catalonia	1164
Rhône valley	1081
Beaujolais	1044
Mosel	1039
Alentejano	937
Colchagua Valley	830
Marlborough	798
Virginia	777
...	
Pocerina	1
Casablanca-Curicó valley	1
Muzla	1
Rhode Island	1
Thraki	1
Ticino	1
Dingač	1
Breede River Valley	1
Pitsilia Mountains	1
Slovenska Istra	1
Gladstone	1
Dolenjska	1
Kathikas	1
Central Greece	1
Alenteo	1
Krania Olympus	1
Corinthia	1
Imathia	1
Hawaii	1
Serra do Sudeste	1
Viile Timis	1
China	1
Piekenierskloof	1
Hrvatsko Primorje	1
Vinho da Mesa	1
Lesbos	1
Table wine	1
Vlootenburg	1
Markopoulo	1
Kentucky	1
Name: province, dtype: int64,	
Napa valley	4480
Columbia valley (WA)	4124
Russian River valley	3091
California	2629
Paso Robles	2350
willamette valley	2301
Mendoza	2301
Alsace	2163
Champagne	1613
Barolo	1599
Finger Lakes	1565
Sonoma Coast	1474
Brunello di Montalcino	1470
Rioja	1469
Sonoma County	1246
Toscana	1197
walla walla valley (WA)	1066
chianti Classico	1062

Santa Barbara County	1000
Carneros	999
Yakima Valley	997
Sta. Rita Hills	975
Dry Creek Valley	939
Sicilia	925
Côtes de Provence	859
Santa Ynez Valley	812
Santa Lucia Highlands	806
Lodi	803
Barbaresco	794
Alexander Valley	784

...

Del Veneto	1
Paestum	1
Benaco Bresciano	1
Côtes du Frontonnais	1
Sonoma-Mendocino	1
Jujuy	1
Geelong	1
Tarantino	1
Monterey County-Napa County	1
Bardolino Superiore	1
Geographe	1
Napa-Sonoma-Marin	1
Offida Rosso	1
Vin de Pays des Coteaux de Bessilles	1
Vin Pétillant	1
Aglianico d'Irpinia	1
Corton Perrières	1
Clairette de Bellegarde	1
Colli del Limbara	1
Vallée du Paradis	1
Collioure	1
Cacc'e Mmitte di Lucera	1
Vin Santo di Carmignano	1
Côtes de Montravel	1
Medrano	1
Riverland	1
Prince Edward County	1
Todi	1
Dolcetto di Monferrato	1
Coteaux du Lyonnais	1

Name: region_1, dtype: int64,

Central Coast	11065
Sonoma	9028
Columbia Valley	8103
Napa	6814
Willamette Valley	3423
California Other	2663
Finger Lakes	1777
Sierra Foothills	1462
Napa-Sonoma	1169
Central Valley	1062
Southern Oregon	917
Oregon Other	727
Long Island	680
North Coast	584
Washington Other	534
South Coast	272
New York Other	231

Name: region_2, dtype: int64,

Roger Voss	25514
Michael Schachner	15134
Kerin O'Keefe	10776
Virginie Boone	9537

Paul Gregutt	9532	
Matt Kettmann	6332	
Joe Czerwinski	5147	
Sean P. Sullivan	4966	
Anna Lee C. Iijima	4415	
Jim Gordon	4177	
Anne Krebiehl MW	3685	
Lauren Buzzeo	1835	
Susan Kostrzewa	1085	
Mike DeSimone	514	
Jeff Jenssen	491	
Alexander Peartree	415	
Carrie Dykes	139	
Fiona Adams	27	
Christina Pickard	6	
Name: taster_name, dtype: int64,		
@vossroger	25514	
@wineschach	15134	
@kerinokeefe	10776	
@vboone	9537	
@paulgwine	9532	
@mattkettmann	6332	
@JoeCz	5147	
@wawinereport	4966	
@gordone_cellars	4177	
@AnneInVino	3685	
@laurbuzz	1835	
@suskostrzewa	1085	
@worldwineguys	1005	
@bkfiona	27	
@winewchristina	6	
Name: taster_twitter_handle, dtype: int64,		
Gloria Ferrer NV Sonoma Brut Sparkling (Sonoma County)		11
Korbel NV Brut Sparkling (California)		9
Segura Viudas NV Extra Dry Sparkling (Cava)		8
Ruinart NV Brut Rosé (Champagne)		7
Gloria Ferrer NV Blanc de Noirs Sparkling (Carneros)		7
Segura Viudas NV Aria Estate Extra Dry Sparkling (Cava)		7
Pierre Sparr NV Brut Réserve Sparkling (Crémant d'Alsace)		6
Bailly-Lapierre NV Brut (Crémant de Bourgogne)		6
J Vineyards & Winery NV Brut Rosé Sparkling (Russian River Valley)		6
Jacquart NV Brut Mosaïque (Champagne)		6
Mumm Napa NV Brut Prestige Sparkling (Napa Valley)		6
Korbel NV Sweet Rosé Sparkling (California)		6
Boizel NV Brut Réserve (Champagne)		6
Pommery NV Brut Royal (Champagne)		5
Chandon NV Rosé Sparkling (California)		5
Heidsieck & Co Monopole NV Blue Top Brut (Champagne)		5
Roederer Estate NV Brut Rosé Sparkling (Anderson Valley)		5
Mumm Napa NV Cuvée M Sparkling (Napa Valley)		5
Mailly Grand Cru NV Délice Demi-Sec (Champagne)		5
Thiénot NV Brut (Champagne)		5
Breathless NV Brut Sparkling (North Coast)		5
Korbel NV Blanc de Noirs Sparkling (California)		5
Freixenet NV Cordon Negro Extra Dry Sparkling (Cava)		5
Mailly Grand Cru NV Blanc de Noirs Brut Pinot Noir (Champagne)		5
G. H. Mumm NV Cordon Rouge Brut (Champagne)		5
J. Dumangin Fils NV Le Rosé Premier Cru Brut (Champagne)		5
Henri Abele NV Brut (Champagne)		5
Thiénot NV Brut Rosé (Champagne)		5
Freixenet NV Cordon Negro Brut Sparkling (Cava)		5
Jean Laurent NV Blanc de Noirs Brut Pinot Noir (Champagne)		5
..		
Cereghino Smith 2012 Super Hudson Red (America)		1
Lasseter 2015 Voila Sauvignon Blanc-Semillon (Sonoma Valley)		1
Longship 2015 Wise Man Cabernet Sauvignon (Horse Heaven Hills)		1

Vignobles La Coterie 2010 Pierre du Diable Red (Ventoux)	1
Five Rows 2014 Marmont Vineyards Chardonnay (Lodi)	1
Paumanok 2007 Merlot (North Fork of Long Island)	1
Concha y Toro 2012 Marques de Casa Concha Sauvignon Blanc (Leyda Valley)	1
Chateau St. Jean 2013 Sonoma Estates Pinot Noir (Sonoma Coast)	1
Sangervasio 2011 I Renai Merlot (Toscana)	1
COS 2013 Frappato (Sicilia)	1
Château Clément-Pichon 2014 Haut-Médoc	1
Duca di Salaparuta 2008 Duca Enrico Nero d'Avola (Sicilia)	1
Seven Hills 2014 Seven Hills Vineyard Merlot (Walla Walla Valley (WA))	1
Bouchard Finlayson 2009 Kaaimansgat Chardonnay (Overberg)	1
Caves Velhas 2013 Magna Carta Reserva Red (Alentejano)	1
Herdade Grande 2007 Condado das Vinhas White (Alentejano)	1
Mitolo 2016 Jester Sangiovese Rosé (McLaren Vale)	1
Adami NV Prosecco Superiore di Cartizze	1
Beaulieu Vineyard 2005 Merlot (Napa Valley)	1
Clos des Rocs 2015 En Chantone (Pouilly-Loché)	1
winzer Krems 2012 Chremisa Sandgrube 13 Grüner Veltliner (Niederösterreich)	1
Château Larrivau 2015 Haut-Médoc	1
Falua 2011 Conde de Vimioso Colheita Seleccionada Rosé (Tejo)	1
Paumanok 2014 Festival Merlot-Cabernet (North Fork of Long Island)	1
Barrel Oak 2009 Barrel Fermented Chardonnay (Virginia)	1
J. McClelland 2014 Chardonnay (Oak Knoll District)	1
Round Pond 2014 Proprietary Rosato di Nebbiolo (Rutherford)	1
Bibbiano 2010 Vigna del Capannino Gran Selezione (Chianti Classico)	1
Love 2015 Merlot (Vino de la Tierra de Castilla)	1
Barreto 2004 Tempranillo (Lodi)	1
Name: title, dtype: int64,	
Pinot Noir	13272
Chardonnay	11753
Cabernet Sauvignon	9472
Red Blend	8946
Bordeaux-style Red Blend	6915
Riesling	5189
Sauvignon Blanc	4967
Syrah	4142
Rosé	3564
Merlot	3102
Nebbiolo	2804
Zinfandel	2714
Sangiovese	2707
Malbec	2652
Portuguese Red	2466
White Blend	2360
Sparkling Blend	2153
Tempranillo	1810
Rhône-style Red Blend	1471
Pinot Gris	1455
Champagne Blend	1396
Cabernet Franc	1353
Grüner Veltliner	1345
Portuguese White	1159
Bordeaux-style White Blend	1066
Pinot Grigio	1052
Gamay	1025
Gewürztraminer	1012
Viognier	996
Shiraz	836
...	
Syrah-Grenache-Viognier	1
Cabernet Sauvignon-Barbera	1
Sauvignonasse	1
Doña Blanca	1
Moscato di Noto	1
Shiraz-Roussanne	1
Aidani	1

Malbec-Carménère	1
Babosa Negro	1
Syrah-Malbec	1
Premisa	1
Tinta Madeira	1
Siegerrebe	1
Pinot Grigio-Sauvignon Blanc	1
Gamza	1
Kuntra	1
Thrapsathiri	1
País	1
Molinara	1
Malagouzia-Chardonnay	1
Xinisteri	1
Diamond	1
Vermentino Nero	1
Piquepoul Blanc	1
Carcajolu	1
Sauvignon Blanc-Assyrtiko	1
Paralleda	1
Zelen	1
Tsapournakos	1
Vidadillo	1
Name: variety, dtype: int64,	
wines & winemakers	222
Testarossa	218
DFJ Vinhos	215
Williams Selyem	211
Louis Latour	199
Georges Duboeuf	196
Chateau Ste. Michelle	194
Concha y Toro	164
Columbia Crest	159
Kendall-Jackson	130
Siduri	126
Gary Farrell	125
Lynmar	118
Montes	117
Albert Bichot	117
Jean-Luc and Paul Aegerter	113
Trapiche	113
Casa Santos Lima	113
Undurraga	113
Santa Ema	112
Robert Mondavi	112
Chanson Père et Fils	109
Fess Parker	108
V. Sattui	107
Iron Horse	106
Foxen	105
Louis Jadot	104
Chehalem	102
Maryhill	102
Feudi di San Gregorio	102
...	
Kees Kieren	1
Napa Smith	1
Château le Reysse	1
Jacobsdal	1
Vylyan	1
Homage	1
Renato Keber	1
Pago Aylés	1
Kinter Collins	1
Poderi Arancio	1
Crush	1

Owen's Estate	1
Pazo de Arribi	1
Château Cantelaudette	1
Talomas	1
Vinicola Serena	1
La Croix de Renaud	1
Montevannos	1
Longridge Bay View	1
VENN	1
Vin Perdu	1
Broc Cellars	1
valdesil	1
Château Saint Baillon	1
Château de Blagnac	1
Domaine des Elephants	1
Illustrious	1
Barrymore by Carmel Road	1
Duo des Mers	1
Château la Rose de Vitrac	1
Name: winery, dtype: int64]	

数据集1中winemag-data_first150k标称属性的取值频数。例如，对于country属性来说，出现频率最高的是US（62397次），其次是France（23478次）、Italy（21098次）。程序及运行结果如下：

count2

[US	62397
Italy	23478
France	21098
Spain	8268
Chile	5816
Argentina	5631
Portugal	5322
Australia	4957
New Zealand	3320
Austria	3057
Germany	2452
South Africa	2258
Greece	884
Israel	630
Hungary	231
Canada	196
Romania	139
Slovenia	94
Uruguay	92
Croatia	89
Bulgaria	77
Moldova	71
Mexico	63
Turkey	52
Georgia	43
Lebanon	37
Cyprus	31
Brazil	25
Macedonia	16
Serbia	14
Morocco	12
Luxembourg	9
England	9
India	8
Lithuania	8

Czech Republic	6
Ukraine	5
South Korea	4
Bosnia and Herzegovina	4
Switzerland	4
China	3
Egypt	3
Slovakia	3
Albania	2
Japan	2
Montenegro	2
Tunisia	2
US-France	1

Name: country, dtype: int64,

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 mild oak. Highly regular; could use more concentration and density.

6

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.

6

86-88 This could work as a rich wine, because there is good structure and piles of botrytis. It could be delicious, with its lovely dry finish, but that's for the future.

6

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

6

Gibilmore, 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.

5

Sweet cherry and baking vanilla aromas are followed by layers of plum, raspberry and spice in this kosher Bordeaux blend from Israel. The wine offers firm tannins and a lingering finish.

5

Raspberry, red cherries and fresh fruit come together with bright acidity and a soft, full texture. There is a good stalky, tannic core to give structure. A great barbecue wine.

4

Big, fragrant, lightly oily and tropical, with zingy, lively acidity pushing lemon and green melon flavors. The finish is round, a little pithy and solid, and overall this strikes us being the quintessential quaffing white for the summer and early fall.

4

Clean and subtle with intriguing musk scents and a medium-weight palate with orange peel and lemon fruit. While the palate is a bit lazy, there's a hit of acidity on the finish to keep things fresh. Drink now.

4

Ripe, fruity Chardonnay, with toast flavors balancing with smooth apricot and pear juice. The wine is soft, rich and generous, finishing with a light acidity.

4

Seems a bit tight and sinewy, with dried spices and black cherries at the fore. Earth, sous bois and chocolate provide the underpinnings, while crisp cranberry notes emerge on the finish. Imported by Meadowbank Estates.

4

Toasty and bacony on the nose, but there's also a helping of black cherries. Shows more mushroom and forest-floor character with time in the glass, and the fruit moves toward the background. Imported by Chez Suze Imports.

4

Strongly herbal and pungent to the point of sweatiness, with just a bit of citrusy fruit creeping in on the otherwise lightweight finish.

4

Soft, gentle and beguiling in its texture, this Cab has a silk-and-velvet mouthfeel. The flavors are very ripe, all cherries and blackberries and milk chocolate. In fact, the wine seems a little too soft for ageability, so drink up now.

4

Smells spicy and tropical, like something that you would want to drink on a beach vacation. The palate deals pineapple and coconut flavors, and good acidity. (Disclaimer: This review makes this wine sound sweeter and simpler than it actually is, but that's because I can't play steel-drum music in the background, and throw some sand between your toes.) Imported by Epic wines.

4

Quite dry and tasting more of skin than fruit, this has a bitter, alcoholic edge to it. The fruit hints at

banana and light tropical.

4

Cherry-vanilla surrounds a flank of big oak, while classic notes of baked strawberry also come into play. A brawny, full-bodied expression of the variety, given 6% Mourvèdre and 6% Syrah, it has good bones and acidity lurking beneath the power.

4

Recognizably Chenin Blanc, this just manages to give an attractive almond scent allied to sweetness and acidity. Screw cap.

4

Delivers a rich and food friendly blend of Merlot, Cabernet Sauvignon and Syrah. The wine is loaded with cherry and black fruit with a rich, but not over extracted personality.

4

This vividly effervescent sparkling Chardonnay offers notes of bruised apples and pears on the nose and a palate of rustic apple cider and lime pith flavors. A bright, lemony acidity makes this a refreshing, easy-drinking sparkler that would match perfectly with pork in a braised apple sauce.

4

A light, bright and crisp wine, packed with fresh apple juice and really vivid acidity. The fresh prickle on the tongue from this young wine is delicious.

4

Aged five years in cask, this has a dry edge with a light poise to it. It has a good, sweet raisin character, on the dry side.

4

A very good Merlot except for some sharpness. The wine is dry and finely tannic, with black and red cherry, mocha, dried herb and tobacco flavors.

4

This extraordinarily rich Syrah swamps the mouth with jammy blackberries, raspberries and cherries. There's nothing subtle about it. But it's basically dry and properly balanced with acids and tannins, which gives it needed balance. Drink now and through 2014 to experience the youthful opulence.

4

This medium-bodied Merlot wafts fragrantly of sweet, cooked black fruits. On the palate as well, the fruit profile is a bit the prune side. It's a softer expression of Merlot without a lot of acidity or depth, but sturdy tannins keep things in balance.

4

This medium-bodied bargain-priced sparkler is maybe just a touch on the heavy side, but it should satisfy most party crowds. Hints of toast accent pineapple and tropical fruit notes, finishing soft and short. Worth trying in Mimosas or Bellinis.

4

Clean and crisp in acidity, with a rich overlay of smoky, caramelized oak, this Chardonnay has appealing pineapple, peach and lime fruit flavors. There's a honey sweet finish, although the wine itself is totally dry. Good price for all this opulence.

4

This single vineyard vinho Verde is complex and powerful. The texture is creamy, rich, with hints of pineapple as well as grapefruit. There is an edgy texture, balanced with the crispest final acidity.

4

Delicate and pale, this elegant rosé hints at rhubarb and strawberry, dappled with light herbs. Tart and refreshing.

4

This wine shows bright berry fruit and some light shades of smoke and spice that come from careful aging in large oak casks. Pair it with pasta or roast chicken with rosemary potatoes.

4

..

Glimmers of game and pepper infuse the red-fruit-driven core of this wine. Plum, raspberry and cherry, dusted with notes of baking spice, flourish on the palate. Medium weight, with dusty tannins, this is reminiscent of crushed dried berries.

1

Rich and packed with ripe Merlot-based fruit, this is a solid wine, with a berry flavor and tannins. The delicious, juicy flavors speak to the ripe fruit that needs several years to fully mellow and integrate.

1

The bouquet is mildly intense, with soft pear and delicate apple skin as the only distinguishable elements. More apple and bitter citrus appear in the mouth, with almost a hoppy character developing on the finish. Drink now.

1

The ripeness of this wine stands out. The cherry flavors even have a slightly raisined edge. The wine is also fairly tannic. Additional flavor nuances include tangerine zest, wild raspberry, red currants and white pepper. This is a great Hanzell. Give it until 2010, and it should develop interesting bottle complexities for many

years.

1

Grechetto makes for a wonderful alternative to Pinot Grigio for lovers of fresh, easy Italian whites. Citrus and stone fruit aromas are followed by drying mineral tones and a zesty, clean close.

1

Grapefruit peel, ripe apricot and zingy lemon spell freshness, while the palate is focused and precise, exhibiting some of the expected peppery notes. Very tangy with a clean, almost lip-smacking finish.

1

Château du Cros began life in the 12th century, commanding a superb military vantage point over the Garonne river. The surrounding vineyards have produced a fresh and fruity Merlot-Cabernet Sauvignon blend with a mineral bite that gives a strong sense of structure. Its weight and a firm texture act as a foil to the juicy red berries and attractive acidity. The wine is almost drinkable now but will be better from 2016. 1

The warmth of 2009 gives a wine that is open, soft, generous in character. It misses out on much depth of flavor, although it does have some spice and light acidity.

1

An off-dry Chenin, with lush notes of peach, red apple and sweet lime. There are soft hints of sweet grass and spice that unfold on the bouquet and on the finish. Vibrant acidity keeps the flavors fresh and leaves a lingering tingling sensation. An easy choice for drinking now.

1

A big, chewy wine, traditional in style, its dark fruits laden with tannins. It is a food wine, with dense plums, berries and ripe fruit wrapped around a core of dryness. The final taste is dry, leavened with some juiciness.

1

Warm while also fresh, this rounded wine is full of red fruits, lit up by a touch of spice and freshened with acidity. Clean and crisp, it's just bursting with fruit at the end.

1

Here's a wine that rides heavy with smoked meat and cured ham aromas. The nose is full of bacon, rubber, spice and peppery fruit and then it turns to a palate of basic strawberry and raspberry. Short on the finish, with a touch of vanilla.

1

Grass, lime, even asparagus notes show in this decidedly Kiwi-inspired Sauvignon Blanc. Lean and tangy to the max.

1

Atypical for this normally delicate varietal, here's a big, sturdy wine with aromas of black raspberries, licorice, and plums, and the fruity flavors are big and full-bodied, although those Pinot Noir silky tannins are here. There are some disagreeable raisiny, pruny notes in the finish.

1

A good wine, although not up there with the '05. But it could improve with a few years of age. Right now, it's tannic, sharp and fresh in blackberries, cherries and rich oak, with a touch of milk chocolate.

1

Tough in the mouth, and chewy, like a gristly piece of steak you can't quite break down. It's certainly rich in blackberry, black currant and smoky oak flavors, and is properly dry and brisk in acidity. But you have to wonder if that toughness will ever age out.

1

Mainly from Castelão with a little Cabernet Sauvignon, this is a ripe, full-bodied and complex wine. Tannins shoot through the solid black fruits, while acidity laces the texture. Flavors of chocolate, spice and final freshness promising short-term aging. Drink from 2015.

1

A tremendous Petite Sirah, among the finest from Paso Robles. It's a big wine, with massive flavors of blackberries, currants, leather, mocha and smoky oak, with an ultralong, spicy finish. Yet the wine is totally dry, with refined, silky tannins. Yes, the alcohol's high, but it's completely balanced. Best now and through the next decade.

1

There was no Pepper Bridge or Seven Hills Merlot in the mix in 2004, but this as good as any other Leonetti Merlot, which is to say as good as it gets, aided by the inclusion of Cold Creek vineyard fruit. Soft, lush and beautifully aromatic; it's just lightly dusted with chocolate and hints of spicy herb.

1

Spicy and peppery from the start, this cool vintage Pinot is decidedly herbal with a pine needle sharpness. The tart, brambly red fruit flavors show medium concentration and fine balance overall.

1

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 Calaveras County Dalton Vineyard, this is blended with 11% Petite Sirah, also from Dalton.

1

This single-vineyard Cabernet, from a relatively cool site in an historically cool vintage, emphasizes the herbal side of the grape. Rhubarb and tart berry fruit flavors carry light hints of cedar and chocolate as well.

1

Mint, cedar and a bit of tomato leaf accent lovely cassis fruit. Full-bodied, with a slightly dusty mouthfeel, like finely ground cinnamon or cocoa, and picking up hints of those flavors as well. Finishes clean and crisp, with soft tannins.

1

Creamy aromas and layers of almonds, toast and spice make this a fascinating wine that continues to reveal new characters. This is the main white from Vacheron, and the intensity of its fruit comes from the almost biodynamic way in which the family's 98-acre vineyard is cultivated. Imported by North Berkeley Imports.

1

Aromas of syrupy red cherry, vanilla and mint meld with touches of blackberry and pepper. The wine delivers notable concentration, plush texture and a lingering finish.

1

This shows classic Oakville tannins, acidity and rich blackberry and cassis flavors, with a nice, spicy finish. But, it doesn't show the flashy concentration or the ageworthiness of the region's best Cabernets. Give it a decanting and drink up.

1

Spice and cream combine in a rounded, soft, very ripe wine. It has intense flavors of lychee and white peach with a refreshing amount of acidity. The wine is dry, but rich from the ginger and with a peppery final bite.

1

Big, rich and sappy, this tastes like cherry and raspberry jam that has been boiled down and made into wine—a suggestion of its youth. It is very fine, with keen acidity and a noble structure tinged with minerality. Give it 3–5 years to age.

1

This wine doesn't seem to have fully ripened, with green, reedy aromas along with notes of cranberry and cherry Jolly Rancher. The full-bodied palate delivers more but seems disconnected with tart, tangy cranberry flavors.

1

A good Cabernet, if a little gritty, with herb-infused flavors of black currants, plums and spices. Might soften with a few years in the bottle, but not really a longterm ager.

1

Name: description, dtype: int64,	
Reserve	2752
Reserva	1810
Estate	1571
Barrel sample	1326
Riserva	754
Barrel Sample	639
Brut	624
Crianza	503
Estate Grown	449
Estate Bottled	396
Dry	374
Old Vine	331
Gran Reserva	330
Brut Rosé	248
Extra Dry	244
Vieilles vignes	225
Bien Nacido Vineyard	195
Rosé	180
Late Bottled Vintage	171
Réserve	166
Unoaked	161
Late Harvest	161
Vintage	152
Barrel Select	145
Single Vineyard	144
Tradition	141
Grand Reserve	139
Tinto	128
Old Vines	127
Classic	123
...	
Murger des Dents de Chien Premier Cru	1
Millesimato Arcadia	1
Finca Piedra Infinita	1
Pop Rosé Extra Dry	1
Estate Midnight	1

Montevasco	1
Noname Dolceacqua	1
Talianna	1
Clisos	1
Kon-Secco	1
Cuvée Jean-Gabriel	1
No. 9	1
Sha'al Single Vineyard Late Harvest	1
Vina Alberdi Reserva	1
Glenrose Vineyard Roussanne-Grenache Blanc	1
Estate Bottled Rosé of	1
Petit Jammes	1
Serègo Alighieri	1
Selected Vines	1
Privat Reserva Brut Nature	1
Finca Los Nobles Cabernet Bouchet	1
Brut Nature Millesimato	1
Quinta do Gradil	1
Lestruelle	1
Carvalhas	1
D'Antan	1
Talensus Grand Escolha	1
Kill Hill	1
Alpine Vineyard	1
Metterheimer	1
Name: designation, dtype: int64,	
California	44508
Washington	9750
Tuscany	7281
Bordeaux	6111
Northern Spain	4892
Mendoza Province	4742
Oregon	4589
Burgundy	4308
Piedmont	4093
Veneto	3962
South Australia	3004
Sicily & Sardinia	2545
New York	2428
Northeastern Italy	1982
Loire Valley	1786
Alsace	1680
Marlborough	1655
Southwest France	1601
Central Italy	1530
Southern Italy	1439
Champagne	1370
Catalonia	1352
Rhône Valley	1318
Colchagua Valley	1201
Languedoc-Roussillon	1082
Douro	1075
Provence	1021
Port	903
Maipo Valley	895
Other	889
...	
Central Greece	1
Stirling	1
Limnos	1
Ioannina	1
Elim	1
Malgas	1
Beni M'Tir	1
Langenlois	1
Colchagua Costa	1

Rose valley	1
Serra do Sudeste	1
Neuchâtel	1
Central Otago-Marlborough	1
Vale Trentino	1
Maipo valley-Colchagua valley	1
Vino da Tavola della Svizzera Italiana	1
Krania Olympus	1
Slovenska Istra	1
San Clemente	1
Dolenjska	1
Ilia	1
Colares	1
Cape South Coast	1
Lemnos	1
Martinborough Terrace	1
Zitsa	1
Pafos	1
Stereia Ellada	1
Casablanca-Curicó valley	1
Ticino	1
Name: province, dtype: int64,	
Napa Valley	6209
Columbia Valley (WA)	4975
Mendoza	3586
Russian River Valley	3571
California	3462
Paso Robles	3053
Willamette Valley	2096
Rioja	1893
Toscana	1885
Sonoma County	1853
Brunello di Montalcino	1746
Sicilia	1701
Alsace	1574
Sonoma Coast	1473
Carneros	1458
Barolo	1398
Dry Creek Valley	1398
Finger Lakes	1372
Champagne	1369
Santa Barbara County	1319
Walla Walla Valley (WA)	1225
Yakima Valley	1162
Alexander Valley	1139
Chianti Classico	1029
Sta. Rita Hills	983
Sonoma Valley	971
Santa Lucia Highlands	962
Central Coast	950
Ribera del Duero	899
Santa Ynez Valley	898
...	
Central Valley	1
Mazoyeres-Chambertin	1
Sonoma-Santa Barbara-Mendocino	1
Vin de Pays des Pyrénées Orientales	1
Listrac	1
Catamarca	1
Sonoma County-Monterey County	1
Rasteau	1
El Pomar District	1
Bellarine Peninsula	1
Santa Barbara-Monterey	1
Ischia	1
Moselle VQS	1

Geelong	1
Ventura County	1
Blanc Fumé de Pouilly	1
Galluccio	1
Civitella d'Agliano	1
Apremont	1
Aleatico dell'Elba	1
Nevada	1
Erbaluce di Caluso	1
Monterey County-Napa County-Sonoma County	1
Sonoma County-Lake County	1
Recioto della Valpolicella Valpantena	1
Rosso del Salento	1
Arlanza	1
Vesuvio	1
Muscadet Sèvre et Maine Clisson	1
Outer Coastal Plain	1
Name: region_1, dtype: int64,	
Central Coast	13057
Sonoma	11258
Columbia Valley	9157
Napa	8801
California Other	3516
Willamette Valley	3181
Mendocino/Lake Counties	2389
Sierra Foothills	1660
Napa-Sonoma	1645
Finger Lakes	1510
Central Valley	1115
Long Island	771
Southern Oregon	662
Oregon Other	661
North Coast	632
Washington Other	593
South Coast	198
New York Other	147
Name: region_2, dtype: int64,	
Chardonnay	14482
Pinot Noir	14291
Cabernet Sauvignon	12800
Red Blend	10062
Bordeaux-style Red Blend	7347
Sauvignon Blanc	6320
Syrah	5825
Riesling	5524
Merlot	5070
Zinfandel	3799
Sangiovese	3345
Malbec	3208
White Blend	2824
Rosé	2817
Tempranillo	2556
Nebbiolo	2241
Portuguese Red	2216
Sparkling Blend	2004
Shiraz	1970
Corvina, Rondinella, Molinara	1682
Rhône-style Red Blend	1505
Pinot Gris	1365
Barbera	1365
Cabernet Franc	1363
Sangiovese Grosso	1346
Pinot Grigio	1305
Viognier	1263
Bordeaux-style White Blend	1261
Champagne Blend	1238

Port	1058
...	
Blauburgunder	1
Carineña	1
Viura-Verdejo	1
Irsai Oliver	1
Karacakiz	1
Malvoisie	1
Listán Negro	1
Terret Blanc	1
Petit Meslier	1
Chinuri	1
Moscate! de Alejandría	1
Groppello	1
Grenache Gris	1
Trousseau Gris	1
Albarín	1
Chardonnay-Pinot Gris	1
Cabernet Moravia	1
Freisa	1
Tinta Francisca	1
Pinotage-Merlot	1
Syrah-Malbec	1
Muskat	1
Durello	1
Catalanesca	1
Aidani	1
Sangiovese Cabernet	1
Moscato di Noto	1
Sauvignonasse	1
Chardone!l	1
Vitovska	1
Name: variety, dtype: int64,	
Williams Selyem	374
Testarossa	274
DFJ vinhos	258
Chateau Ste. Michelle	225
Columbia Crest	217
Concha y Toro	216
Kendall-Jackson	216
Trapiche	205
Bouchard Père & Fils	203
Kenwood	191
De Loach	189
Joseph Drouhin	189
Georges Duboeuf	188
Cameron Hughes	172
Wines & Winemakers	169
Albert Bichot	167
Robert Mondavi	166
Louis Latour	154
Dry Creek Vineyard	153
D'Arenberg	153
Morgan	153
Concannon	151
Martin Ray	149
Errazuriz	148
L'Ecole No. 41	144
Gary Farrell	144
Olivier Leflaive	143
Iron Horse	142
Montes	142
Waterbrook	142
...	
Vignobles Dom Brial	1
Sadie Family	1

Château Haut Maurin	1
Oro de Castilla	1
Quinta do Côa	1
Vivencias	1
Château Sainte-Croix	1
Château Beauregard Lagupeau	1
Celler Piñol	1
Shane wines	1
Domaine de la Coume du Roy	1
1850	1
Lowden Hills	1
2Plank	1
Matošević	1
Red Bucket	1
Barefoot Refresh	1
Château Jean Faux	1
Château Tour d'Auron	1
Truth	1
Drassanes	1
Garzón	1
The Tribe	1
Quinta da Pedra	1
Garden Creek	1
Tusk 'n Red	1
Pierre Dupond	1
Château Rocher-Bonregard	1
Domaine Normand	1
Château Lacombe Cadiot	1
Name: winery, dtype: int64]	

数据集2标称属性的取值频数。程序及运行结果如下：

```
count3
```

```
[[]
35542
["Dunkin' Donuts"]
7010
["starbucks"]
4178
["walmart"]
3054
["mcdonalds"]
2911
["wawa"]
2831
["7-Eleven US"]
2403
["Publix Super Markets"]
2272
["QuikTrip"]
1812
["Kroger"]
1256
["Chick-fil-A"]
1188
["Chevron"]
1058
["Sheetz"]
903
["Safeway"]
```



```
854
["Costco Wholesale Corp."]
849
["Meijer"]
843
["ARCO"]
836
["H-E-B"]
786
["Kwik Trip"]
782
["Circle K Stores"]
732
["King Soopers"]
711
["Speedway"]
621
["Circle K Stores","QuikTrip"]
621
["Shell Oil"]
612
["Dunkin' Donuts","Stop & Shop"]
598
["Giant Eagle"]
575
["Wawa","Dunkin' Donuts"]
536
["Family Dollar Stores"]
496
["Target"]
477
["mcdonalds","Speedway"]
473

...
["Shell Oil","Marathon Petroleum","Hardee's","Dollar General"]
1
["mcdonalds","Society of St. Vincent de Paul","Speedway","Kroger","Family Dollar Stores"]
1
["walmart","Whataburger","Exxon Mobil","Shell Oil","Sonic","Brookshire's Grocery Company"]
1
["Price Chopper","Stewart's Shops","Kinney Drugs"]
1
["Hy-Vee","Health Mart","Casey's General Stores","Culver's","Fareway Stores"]
1
["mcdonalds","Kroger","Dollar General","Shell Oil"]
1
["Food Lion","Sheetz","Shell Oil"]
1
["Costco Wholesale Corp.,"walmart","Holiday Station","starbucks","mcdonalds"]
1
["Hampton","ConocoPhillips","mcdonalds","Exxon Mobil","Quality Inn","Maverik","Family Dollar Stores"]
1
["Ralphs","99 Cents Only Stores","Costco Wholesale Corp."]
1
["Exxon Mobil","mcdonalds","CITGO Petroleum"]
1
["mcdonalds","7-Eleven US","Ralphs"]
1
["Walgreens","Dollar General","Shell Oil","Family Dollar Stores"]
1
["walmart","Safeway","Mobil","starbucks","mcdonalds"]
1
["mcdonalds","Sam's Club","Speedway"]
1
["Stewart's Shops","Dunkin' Donuts","Applebee's","Sunoco","Dollar General"]
```

```

1
["Ascend Resort Collection","Buffalo Wild Wings","mcdonalds","Grand Hyatt","Merrell","United States Postal
Service (USPS)","VFW (Veterans of Foreign Wars)","Ramada","Culver's"] 1
["ConocoPhillips","Safeway","King Soopers","Taco Bell","The Home Depot","Walgreens"]
1
["Circle K Stores","mcdonalds","Giant Eagle","Dollar General","Goodwill Industries"]
1
["Safeway","Chevron","starbucks","Sephora"]
1
["Pay Less Super Markets","mcdonalds","Spirit Halloween","Speedway"]
1
["Fairway Market","Edible Arrangements","The Home Depot","King Kullen Grocery Co.","Lord and
Taylor","Carvel","Dunkin' Donuts","Planet Fitness","Stop & Shop","Red Mango"] 1
["Bealls Florida","Sunglass Hut","Hilton International","Circle K Stores"]
1
["Dairy Queen","Sunoco","Dollar General"]
1
["walmart","Whataburger","Buc-ee's","Exxon Mobil"]
1
["Kwik Shop","Raymond James Financial","SUBWAY","Dollar General","Good Neighbor Pharmacy","Sonic","Pizza
Hut","Godfather's Pizza","Do It Best","Health Mart"] 1
["walmart","America's Thrift Store","Chevron","Burger King US","Boys & Girls Clubs of America","Circle K
Stores","Dollar General","Jack's Family Restaurants"] 1
["walmart","Dollar General","Sonic","Texaco","Chevron","mcdonalds","Exxon Mobil"]
1
["Auntie Anne's","SUBWAY","starbucks"]
1
["walmart","Tractor Supply Co.","mcdonalds","Speedway","Dollar General","Cracker Barrel"]
1
Name: related_same_day_brand, dtype: int64,
[]
2983
["SmartStyle Family Hair Salons"]
335
["Department of Veterans Affairs"]
156
["H&R Block"]
134
["Sprint"]
112
["MasterCuts"]
92
["Supercuts"]
62
["Dunkin' Donuts"]
55
["Safeway"]
45
["Yogen Fr  z"]
34
["Yogen Fr  z","Department of Veterans Affairs"]
33
["H-E-B","Whataburger","walmart","Stripes Convenience Stores","Valero Energy","Exxon
Mobil","mcdonalds","Circle K Stores","Sonic","SUBWAY"] 32
["CVS","SmartStyle Family Hair Salons"]
32
["Cold Stone Creamery"]
32
["walmart","Sonic","Phillips 66","mcdonalds","Braum's Ice Cream and Dairy Stores","Chick-fil-
A","OnCue","Walgreens","Target","starbucks"] 26
["Fresenius Medical Care North America"]
25
["Publix Super Markets"]
25
["Dunkin' Donuts","ShopRite"]
24

```

```

["starbucks","Costco Wholesale Corp.","Chevron","Safeway","Target","Shell Oil","mcdonalds","In-N-Out
Burger","SUBWAY","7-Eleven US"]
24
["SUBWAY"]
24
["starbucks","Safeway","Chevron","Costco Wholesale Corp.","Target","In-N-Out Burger","Shell
Oil","mcdonalds","SUBWAY","76"]
23
["wawa"]
22
["Circle K Stores","QuikTrip","walmart","mcdonalds","Fry's Food & Drug Stores","Jack in the
Box","Walgreens","SUBWAY","starbucks","Chevron"]
22
["CVS"]
21
["Sheetz"]
21
["starbucks","Costco Wholesale Corp.","Chevron","Safeway","Target","Shell Oil","In-N-Out
Burger","mcdonalds","SUBWAY","7-Eleven US"]
20
["starbucks","mcdonalds","ARCO","Chevron","In-N-Out Burger","Target","Jack in the Box","SUBWAY","walmart","7-
Eleven US"]
19
["MasterCuts","Little Caesars"]
19
["starbucks","Costco Wholesale Corp.","Target","Chevron","Safeway","mcdonalds","7-Eleven US","In-N-Out
Burger","SUBWAY","Shell Oil"]
19
["Casey's General Stores"]
19
...
["starbucks","mcdonalds","Target","walmart","Taco Bell","ARCO","Save Mart Supermarkets","Costco Wholesale
Corp.","Chevron","76"]
1
["Dunkin' Donuts","starbucks","mcdonalds","SUBWAY","Popeyes Louisiana Kitchen","Rite Aid","Modell's Sporting
Goods","Nathan's Famous","Duane Reade","Baskin Robbins"]
1
["walmart","mcdonalds","Kwik Shop","Hy-Vee","Taco Bell","Casey's General Stores","Menard's","Scooter,Âs
Coffee and Yogurt","SUBWAY","starbucks"]
1
["H-E-B","walmart","Shell Oil","Whataburger","Exxon Mobil","mcdonalds","Chick-fil-
A","starbucks","Sonic","Walgreens"]
1
["walmart","mcdonalds","Dollar General","BP","Chick-fil-A","Taco Bell","Family Dollar
Stores","Wendy's","Circle K Stores","Sonic"]
1
["Dunkin' Donuts","ShopRite","Tesla Motors","mcdonalds","7-Eleven US","Walgreens","CVS","Urban Edge
Properties","Wendy's","The Home Depot"]
1
["Sheetz","walmart","mcdonalds","Sunoco","Dollar General","Dairy Queen","Wendy's","Giant Eagle","Quality
Inn","Rite Aid"]
1
["Hannaford Supermarkets","Dunkin' Donuts","Irving Oil","CITGO Petroleum","Circle K Stores","Rite Aid"]
1
["walmart","Sonic","Exxon Mobil","Dollar General","Whataburger","mcdonalds","Braum's Ice Cream and Dairy
Stores","Dairy Queen","Valero Energy","Root 66 Propane"]
1
["walmart","Publix Super Markets","mcdonalds","wawa","Chick-fil-A","Shell Oil","Walgreens","7-Eleven
US","Circle K Stores","CVS"]
1
["Hy-Vee","Casey's General Stores","walmart","BP","mcdonalds","Menard's","Walgreens","Scooter,Âs
Coffee and Yogurt","Taco Bell","Burger King US"]
1
["ARCO","starbucks","In-N-Out Burger","Jack in the Box","Chevron","Target","Costco Wholesale Corp.","Carl's
Jr.","Del Taco","7-Eleven US"]
1
["mcdonalds","Meijer","Marathon Petroleum","walmart","Dollar General","Speedway","Phillips 66","Taco
Bell","Goodwill Industries","Family Dollar Stores"]
1
["mcdonalds","Speedway","Giant Eagle","Dollar General","Sunoco","walmart","Marc's","Dollar Tree","BP","Shell
oil"]
1
["starbucks","Safeway","Costco Wholesale Corp.","Chevron","76","mcdonalds","7-Eleven US","Fred Meyer
Jewelers","Jack in the Box","Shell Oil"]
1
["walmart","Loaf 'N Jug","Exxon Mobil","ConocoPhillips","starbucks","Kaiser
Permanente","SUBWAY","Safeway","mcdonalds","Sonic"]
1
["starbucks","mcdonalds","Chevron","Target","ARCO","In-N-Out Burger","Costco Wholesale Corp.","Ralphs","7-
Eleven US","SUBWAY"]
1
["walmart","mcdonalds","starbucks","Chick-fil-A","Target","Kroger","Walgreens","Sonic","Shell Oil","Food
City"]
1
["walmart","Dollar General","Jack's Family Restaurants","mcdonalds","Chevron","Shell Oil","America's Thrift
Store","Chick-fil-A","Taco Bell","Piggly Wiggly"]
1
["mcdonalds","walmart","Walgreens","SUBWAY","BP","Dunkin' Donuts","Dollar Tree","Burger King US","Family
Dollar Stores","CITGO Petroleum"]
1

```

```

["walmart", "Casey's General Stores", "mcdonalds", "Dollar General", "Sinclair Oil", "Sonic", "SUBWAY", "Phillips
66", "United States Postal Service (USPS)", "Dairy Queen"] 1
["Dunkin' Donuts", "starbucks", "mcdonalds", "SUBWAY", "Rite Aid", "Baskin Robbins", "BP", "Burger King US", "7-Eleven
US", "Modell's Sporting Goods"] 1
["Chick-fil-A", "walmart", "mcdonalds", "QuikTrip", "RaceTrac", "Publix Super Markets", "Zaxby's", "Waffle
House", "Dollar General", "Wendy's"] 1
["walmart", "Phillips 66", "Sonic", "Braum's Ice Cream and Dairy Stores", "Dollar General", "Harps Food
Store", "mcdonalds", "SUBWAY", "Casey's General Stores", "Lowe's"] 1
["walmart", "Shell Oil", "mcdonalds", "Chevron", "Exxon Mobil", "Walgreens", "Whataburger", "Jack in the Box", "Chick-
fil-A", "Fiesta Mart"] 1
["starbucks", "Target", "In-N-Out Burger", "Costco Wholesale
Corp.", "mcdonalds", "Chevron", "Vons", "Ralphs", "SUBWAY", "Taco Bell"]
1
["mcdonalds", "Meijer", "SUBWAY", "Speedway", "Shell Oil", "BP", "walmart", "Pilot Travel Centers", "Burger King
US", "Walgreens"] 1
["Publix Super Markets", "Wawa", "7-Eleven US", "CITGO Petroleum", "Family Dollar Stores", "Shell
Oil", "walmart", "CVS", "Bealls Florida", "mcdonalds"] 1
["Dunkin' Donuts", "ShopRite", "CVS", "starbucks", "T.J. Maxx", "Stop & Shop", "mcdonalds", "Wendy's", "Target", "The
Home Depot"] 1
["Dunkin' Donuts", "Stop & Shop", "CVS", "Cumberland
Farms", "mcdonalds", "walmart", "SUBWAY", "Mobil", "Walgreens", "Dollar Tree"]
1
Name: related_same_month_brand, dtype: int64,
[]
74540
["United States Postal Service (USPS)"]
5352
["Aflac (American Family Life Assurance)"]
1517
["Dollar General"]
1308
["National Association for the Education of Young Children (NAEYC)"]
1071
["Cricket Wireless"]
1036
["The American Legion"]
1011
["Shell Oil"]
794
["Boys & Girls Clubs of America"]
712
["7-Eleven US"]
665
["CITGO Petroleum"]
646
["VFW (Veterans of Foreign Wars)"]
635
["Family Dollar Stores"]
592
["Marathon Petroleum"]
563
["BP"]
480
["Sunoco"]
477
["Chevron"]
473
["United States Postal Service (USPS)", "The American Legion"]
434
["Dollar General", "United States Postal Service (USPS)"]
398
["SUBWAY"]
387
["Auto wares"]
384
["State Farm"]

```

```

371
["Exxon Mobil"]
365
["Valero Energy"]
363
["Mobil"]
349
["Walgreens"]
326
["CrossFit"]
301
["CVS"]
290
["YMCA"]
255
["Dunkin' Donuts"]
254

...
["Planet Fitness","Vallarta Supermarkets","Dollar Rent A Car","Panda Express","DeL
Taco","mcdonalds","Fatburger","SUBWAY","Denny's","Avis Rent A Car"]
1
["Chevron","Hardee's"]
1
["RaceWay","Dairy Queen","America's Thrift Store","YRC Freight"]
1
["Target","Pure Barre","Lenny's Sub Shop","Williams-Sonoma","Pottery Barn","Pottery Barn Kids","United States
Postal Service (USPS)","Allstate Insurance","Merle Norman Cosmetics"] 1
["Burger King US","CVS","Family Dollar Stores","CITGO Petroleum","Helena Chemical","Fraternal Order of
Eagles","Allstate Insurance","Boost Mobile"] 1
["United States Postal Service (USPS)","Rita's Italian Ice","Best Western"]
1
["Ross Stores","Walgreens","O'Charley's","PetSmart","Tractor Supply Co.,"GNC (General Nutrition
Centers)","Wendy's","Hobby Lobby Stores","Waffle House","Best Western"] 1
["Vision Source","84 Lumber","Miracle-Ear","United States Postal Service (USPS)"]
1
["mcdonalds","Dairy Queen","Sonic","Circle K Stores","Arby's","Phillips 66","Pizza Hut","OnCue","Sirloin
Stockade","KFC US"] 1
["CrossFit","Maytag"]
1
["Pilot Travel Centers","SUBWAY","ARCO","Denny's","Rodeway Inn"]
1
["Department of Veterans Affairs","Sunoco","United States Postal Service (USPS)"]
1
["76","True Value Hardware","Dollar General","Cricket Wireless"]
1
["Chick-fil-A","Einstein Brothers","Fresh√ns","Encompass Health"]
1
["Circle K Stores","South Carolina DMV","American Job Centers","LifeSafer","East Coast Metal
Distributors","CrossFit"] 1
["Dunkin' Donuts","T-Mobile"]
1
["CVS","Rita's Italian Ice","Dunkin' Donuts","Panera Bread","Mobil","Drayer Physical Therapy Institute","Shell
Oil","Edward Jones","H&R Block","Raymond James Financial"] 1
["Circle K Stores","Auto Wares"]
1
["mcdonalds","Good Neighbor Pharmacy","IGA (Independent Grocers Alliance)","Marathon Petroleum","Dollar
General","Family Dollar Stores","Indiana DMV","CVS","Dairy Queen","Napa Auto Parts"] 1
["Society of St. Vincent de Paul","Hardee's","Marathon Petroleum","Shred-it","GAF Materials"]
1
["Hardee's","Sherwin-Williams","PPG Pittsburgh Paints","Cricket Wireless","Boost Mobile"]
1
["Exxon Mobil","Circle K Stores","Edward Jones"]
1
["Shell Oil","Parker Hannifin"]
1

```

```

["Love's Travel Stops and Country Stores","mcdonalds","SUBWAY","Shell Oil","Farmers Insurance Group"]
1
["Club Fitness","Goodwill Industries","BP","HoneyBaked Ham","American Family Insurance (AmFam)"]
1
["Walgreens","Burger King US","AutoZone","Sunoco","Advance Auto Parts","Dollar General","Dairy
Queen","Domino's Pizza","Circle K Stores","Family Dollar Stores"]
1
["National Association for the Education of Young Children (NAEYC)","Enterprise Rent-A-Car"]
1
["Waste Management","Penske"]
1
["starbucks","Speedway","Anytime Fitness","Little Caesars","SUBWAY","Rite Aid","Hungry Howie's Pizza &
Subs","Olga's Kitchen","Martinizing Dry Cleaners"]
1
["mcdonalds","Sports Clubs Network","starbucks","AT&T"]
1
Name: top_brands, dtype: int64,
[]
106
[1576,1360,1242,1282,1357,1824,2434,5385,6447,5447,5631,6411,7176,7129,7404,9139,8619,8669,8065,5720,4151,3330
,2471,1982]
1
[1600,1620,1624,1581,1602,1577,1728,1860,1914,1757,1814,1815,1985,1975,2063,2278,2298,2385,2296,1957,2000,1803
,1768,1679]
1
[1036,1047,1066,1037,1026,1073,1133,1326,1141,1080,1105,1120,1127,1052,1104,1563,1346,1450,1306,1175,1224,1125
,1125,1070]
1
[3060,2957,2945,2932,3085,3746,4480,7752,6432,5920,6006,6309,6523,6362,6657,7858,6931,7311,6648,5538,5158,4264
,3631,3186]
1
[870,852,841,872,887,885,829,2465,2256,2109,2209,2379,2267,2210,2534,2713,1893,2150,1953,1756,1471,1382,1072,9
17]
1
[2819,2775,2800,2794,2794,2861,3414,4601,3004,2796,2850,2995,2911,2952,3975,3695,3996,4405,3976,3583,3549,3397
,3157,2894]
1
[585,587,567,565,567,574,710,634,509,460,439,451,462,474,570,667,764,791,764,712,765,674,667,621]
1
[3726,3278,3142,3117,3216,3762,5496,7178,7128,6771,7638,8962,10215,10484,11292,11815,12729,13781,13897,11435,9
077,7244,5464,4388]
1
[2513,2514,2484,2413,2398,2403,2380,2226,1853,1515,1532,1455,1458,1542,1584,1580,1838,1946,2090,2378,2431,2505
,2555,2492]
1
[618,622,602,627,613,616,682,1105,985,758,774,730,744,699,844,834,868,920,981,902,778,693,658,642]
1
[1138,1037,1024,999,1016,1155,1273,4079,2972,2415,2464,2738,2759,2589,2452,4047,2725,3082,2588,2067,1779,1538,
1391,1174]
1
[3914,3070,2658,2697,2987,4993,10714,21798,22400,13297,10452,9916,9702,10238,11937,12985,15461,20730,19204,127
29,9171,7219,6391,5413]
1
[13951,13056,12453,12226,12780,14146,17472,16888,14885,12963,13235,13636,14898,15610,16550,18300,18069,19730,1
9510,19019,18218,17329,16504,15086]
1
[584,574,568,551,568,656,911,2405,1873,1331,1243,1276,1354,1402,1640,2035,1904,1864,1490,1030,879,898,729,624]
1
[4771,4584,4503,4548,4718,5202,7169,11284,10509,9655,10169,10774,11241,11102,11572,12827,13375,13771,13055,108
69,9805,7969,6330,5789]
1
[2368,2248,2191,2137,2155,2344,3220,5113,4379,3837,3648,3719,4055,4002,4120,4462,4585,5020,4610,3410,3012,2889
,2692,2506]
1
[1039,670,476,388,634,1724,3963,7063,6976,6443,6990,7756,8530,8547,8634,10137,10597,11358,10211,8539,6138,4093
,2318,1338]
1
[2649,2207,2091,2064,2332,3377,5890,8556,7538,7090,6957,7649,8174,7908,9164,10259,10428,10210,8921,7432,6361,5
532,4179,3272]
1
[899,894,867,914,929,1116,1412,1744,1576,1620,1734,1882,1975,1970,2044,2333,2265,2145,1929,1647,1487,1255,1131
,983]
1
[2217,2074,2039,2033,2113,2369,3308,4085,3642,4106,4765,5638,6115,5743,5779,6584,6753,6849,6023,4980,4147,3238
,2914,2510]
1
[1903,1887,1874,1876,1883,1902,2106,2557,2630,2195,2179,2150,2236,2190,2251,2450,2720,2588,2507,2243,2171,2143
,2035,1960]
1
[1281,1244,1228,1219,1273,1338,1575,1904,1673,1340,1260,1217,1404,1469,1467,1906,1915,2069,2040,1795,1722,1577
,1456,1402]
1
[2825,2592,2519,2506,2632,2752,3048,4157,4895,3725,3693,3717,3869,4003,4567,5652,5527,5720,4981,3926,3607,3421
,3261,3151]
1
[5311,4955,4818,4852,5112,5966,8213,12473,11978,9813,10006,9969,10662,10881,11562,13067,14724,16262,13401,1106
1,9956,8168,6921,6018]
1
[3864,3651,3496,3583,3697,3988,4863,7042,6395,6642,7040,7080,7335,7508,7433,7965,7593,7969,6809,5868,5595,5302

```

,4673,4263]	1
[1147,1126,1089,1087,1156,1153,1204,1556,1108,963,922,983,989,1000,1070,1098,1256,1230,1254,1198,1191,1277,1177,1160]	1
[1477,1370,1320,1347,1451,1752,2240,5120,2882,2595,2599,2956,3136,2904,3149,4626,4179,4399,3201,2837,2383,2064,1860,1672]	1
[2789,2379,2268,2275,2653,3221,5646,7730,6943,5560,5933,6373,6754,6999,7614,9720,9510,10070,8845,6724,5507,4884,4095,3461]	1
[1300,1259,1282,1233,1297,1532,1797,2375,2091,1942,2075,2098,2253,2324,2271,2507,2817,3018,2804,2414,2100,1753,1658,1490]	1
...	
[1588,1538,1529,1580,1530,1637,1560,2071,1852,1644,1617,1560,1702,1699,1851,1937,2135,2110,1966,1867,1799,1780,1668,1605]	1
[1380,1228,1100,1104,1153,1250,1576,3944,4854,5500,5939,6696,7358,7005,7025,6948,6802,6300,5376,4512,3734,2766,2145,1665]	1
[1073,986,959,961,1051,1359,2021,3606,2722,2350,2351,2460,2590,2484,3241,3287,3222,3534,2963,2745,2448,2211,1531,1270]	1
[2129,2021,2028,2080,2157,2265,2924,3847,3459,3320,3187,3181,3356,3057,3136,3538,3466,3606,3613,2993,2880,2686,2436,2272]	1
[2480,2060,1857,1802,2341,3851,5706,8717,8497,7001,6669,7261,7815,8063,9840,11750,11927,11463,10013,6944,5757,5197,4359,3189]	1
[1910,1794,1789,1825,1860,2201,2980,3668,3101,2612,2537,2691,2897,2843,3376,3841,4346,4655,3788,3113,2891,2662,2378,2071]	1
[1127,1136,1106,1147,1081,1182,1462,1790,2046,1888,1906,1906,1840,1847,1886,1965,1588,1282,1164,1134,1159,1233,1214,1179]	1
[1087,1074,1083,1081,1109,1111,1222,1339,1169,1089,1031,988,1003,986,1064,1157,1236,1420,1257,1163,1251,1201,1180,1146]	1
[2117,2014,1927,1939,1958,1998,2341,2438,2080,1971,1971,2091,2221,2320,2582,2716,2697,2809,2847,2791,2709,2541,2294,2158]	1
[3309,2921,2774,2755,2900,3518,6149,9738,13455,11602,13194,15380,17367,17171,16525,17244,19041,17300,15335,12250,9532,7151,5433,4133]	1
[1215,1180,1057,1046,1067,1296,1416,2077,2244,2413,2674,3081,3327,3121,2981,3270,3291,3265,2922,2489,2220,2001,1596,1417]	1
[1100,1086,1047,1032,1048,1123,1188,1707,1306,1177,1262,1306,1431,1375,1550,1957,1948,2032,1784,1597,1421,1393,1343,1232]	1
[3438,3047,2790,2803,2948,4133,6058,8168,8090,8271,8410,8921,9213,9537,10238,11462,11509,11070,9833,8181,7068,6396,5182,4054]	1
[1101,1063,1053,1065,1131,1227,1694,1688,1680,1506,1667,1679,1678,1657,1849,2036,2075,1986,1676,1435,1335,1275,1223,1099]	1
[1723,1529,1466,1344,1370,1665,2401,3665,3882,3724,3799,3850,4249,4279,4268,4548,4491,4450,3942,3484,2868,2439,2173,1907]	1
[6327,6297,6265,6196,6638,6924,7873,9980,8590,7263,7259,7342,7556,7526,8035,9388,9891,10027,8953,8354,7990,7306,6962,6843]	1
[895,859,870,844,848,938,1072,1757,1460,1177,1192,1225,1305,1207,1354,1861,1628,1977,1625,1381,1206,1133,1042,953]	1
[1155,979,897,733,685,787,1208,3144,3752,3632,3983,4465,5422,4736,4876,4871,4436,5052,4700,3976,3019,2487,1743,1326]	1
[2995,2937,2887,2910,2992,3198,3630,4458,3735,3264,3235,3396,3414,3459,3620,4162,4648,5073,4695,3970,3670,3547,3311,3093]	1
[2247,2159,2164,2014,2019,2013,2338,2851,2160,1923,1684,1703,1740,1820,1936,2103,2431,2667,2654,2604,2361,2330,2278,2302]	1
[6173,5652,5598,5492,6308,8599,14251,19311,14847,14361,14497,15694,16539,16914,18901,21807,23759,23931,19408,15596,13488,11084,9012,7324]	1
[3044,2898,2695,2754,2842,3276,4195,5279,5318,4308,4367,4451,4788,4965,5075,6313,6120,6414,6597,5964,5146,4844,4248,3475]	1
[1108,1079,1064,1056,1120,1242,1513,3570,2606,2495,2489,2720,2819,2783,2993,3471,2704,2901,2647,2285,1993,1771,1404,1244]	1
[1708,1560,1531,1504,1721,2062,3334,5329,3702,3123,3478,4053,4339,3793,4882,5171,5740,5673,4640,3912,3304,2692,2362,1947]	1
[4030,3499,3464,3199,3498,5465,12207,17111,14023,11763,11542,11531,12139,11961,12900,15204,17270,18092,15781,12389,9469,7860,6616,5253]	1
[2424,2347,2301,2293,2276,2234,2572,2859,2800,2252,2026,2030,2103,2145,2228,2700,3069,3021,3037,2909,2898,2893,2698,2585]	1
[1776,1638,1610,1588,1823,2826,3990,5347,5006,4232,4018,4065,4522,4396,5430,6389,6983,7208,6028,4010,3463,3133,2594,2053]	1
[1515,1419,1339,1292,1289,1309,1689,2977,3586,2858,3156,3485,3672,3786,3813,4087,3952,4171,3495,2887,2465,2173	

```
,1933,1725] 1
[1130,988,929,848,858,965,1374,2415,2412,2469,2749,3771,4448,4120,3781,4007,3940,4406,4522,4123,3062,2392,1832,1365] 1
[1222,1153,1129,1125,1110,1152,1283,1918,1688,1515,1515,1519,1597,1530,1672,1968,1830,1875,1700,1553,1446,1398,1368,1321] 1
Name: popularity_by_hour, dtype: int64,
{}
106
{"Monday":6438,"Tuesday":6354,"Wednesday":6500,"Thursday":6048,"Friday":7319,"Saturday":7769,"Sunday":7271}
1
{"Monday":21175,"Tuesday":21622,"Wednesday":21511,"Thursday":18240,"Friday":20405,"Saturday":20310,"Sunday":17299} 1
{"Monday":599,"Tuesday":681,"Wednesday":1116,"Thursday":942,"Friday":905,"Saturday":516,"Sunday":527}
1
{"Monday":3902,"Tuesday":3624,"Wednesday":3574,"Thursday":3263,"Friday":5802,"Saturday":5250,"Sunday":5376}
1
{"Monday":2283,"Tuesday":2420,"Wednesday":2295,"Thursday":1983,"Friday":2185,"Saturday":1776,"Sunday":1676}
1
{"Monday":2912,"Tuesday":2738,"Wednesday":2636,"Thursday":2307,"Friday":3339,"Saturday":4816,"Sunday":3066}
1
{"Monday":3139,"Tuesday":3019,"Wednesday":3003,"Thursday":2522,"Friday":2679,"Saturday":2835,"Sunday":2452}
1
{"Monday":7753,"Tuesday":8360,"Wednesday":9184,"Thursday":6545,"Friday":7379,"Saturday":7557,"Sunday":7374}
1
{"Monday":5682,"Tuesday":5738,"Wednesday":5509,"Thursday":4441,"Friday":4750,"Saturday":3499,"Sunday":3593}
1
{"Monday":8677,"Tuesday":8898,"Wednesday":9209,"Thursday":7372,"Friday":8358,"Saturday":7056,"Sunday":6045}
1
{"Monday":3512,"Tuesday":3267,"Wednesday":3107,"Thursday":2699,"Friday":3295,"Saturday":3700,"Sunday":3228}
1
{"Monday":29059,"Tuesday":30160,"Wednesday":27965,"Thursday":24167,"Friday":24779,"Saturday":23592,"Sunday":22265} 1
{"Monday":828,"Tuesday":817,"Wednesday":771,"Thursday":631,"Friday":770,"Saturday":640,"Sunday":684}
1
{"Monday":3721,"Tuesday":4107,"Wednesday":3839,"Thursday":3037,"Friday":3159,"Saturday":2325,"Sunday":2490}
1
{"Monday":8408,"Tuesday":8326,"Wednesday":7828,"Thursday":6761,"Friday":7698,"Saturday":8375,"Sunday":8233}
1
{"Monday":4982,"Tuesday":5010,"Wednesday":4984,"Thursday":4260,"Friday":4417,"Saturday":4052,"Sunday":3982}
1
{"Monday":19562,"Tuesday":20190,"Wednesday":20165,"Thursday":16194,"Friday":17918,"Saturday":17638,"Sunday":14357} 1
{"Monday":2258,"Tuesday":2105,"Wednesday":2280,"Thursday":2232,"Friday":2718,"Saturday":2380,"Sunday":2612}
1
{"Monday":3125,"Tuesday":3229,"Wednesday":3254,"Thursday":2444,"Friday":2593,"Saturday":2378,"Sunday":2396}
1
{"Monday":2750,"Tuesday":2909,"Wednesday":2770,"Thursday":2358,"Friday":2566,"Saturday":2071,"Sunday":1781}
1
{"Monday":11637,"Tuesday":11939,"Wednesday":11210,"Thursday":9765,"Friday":10260,"Saturday":8359,"Sunday":7326}
} 1
{"Monday":1582,"Tuesday":1664,"Wednesday":1530,"Thursday":1398,"Friday":1421,"Saturday":1246,"Sunday":1143}
1
{"Monday":33478,"Tuesday":33344,"Wednesday":32836,"Thursday":27772,"Friday":31436,"Saturday":28127,"Sunday":29800} 1
{"Monday":4906,"Tuesday":5094,"Wednesday":4982,"Thursday":4201,"Friday":4679,"Saturday":4152,"Sunday":3429}
1
{"Monday":2894,"Tuesday":2848,"Wednesday":2727,"Thursday":2297,"Friday":2480,"Saturday":2492,"Sunday":2197}
1
{"Monday":6451,"Tuesday":6577,"Wednesday":6175,"Thursday":5155,"Friday":5586,"Saturday":5581,"Sunday":4412}
1
{"Monday":4624,"Tuesday":4359,"Wednesday":4030,"Thursday":3788,"Friday":4875,"Saturday":3931,"Sunday":4073}
1
{"Monday":3405,"Tuesday":3439,"Wednesday":3335,"Thursday":2659,"Friday":2711,"Saturday":2775,"Sunday":2599}
1
{"Monday":10617,"Tuesday":10588,"Wednesday":10113,"Thursday":8570,"Friday":9928,"Saturday":8204,"Sunday":8389}
1
```



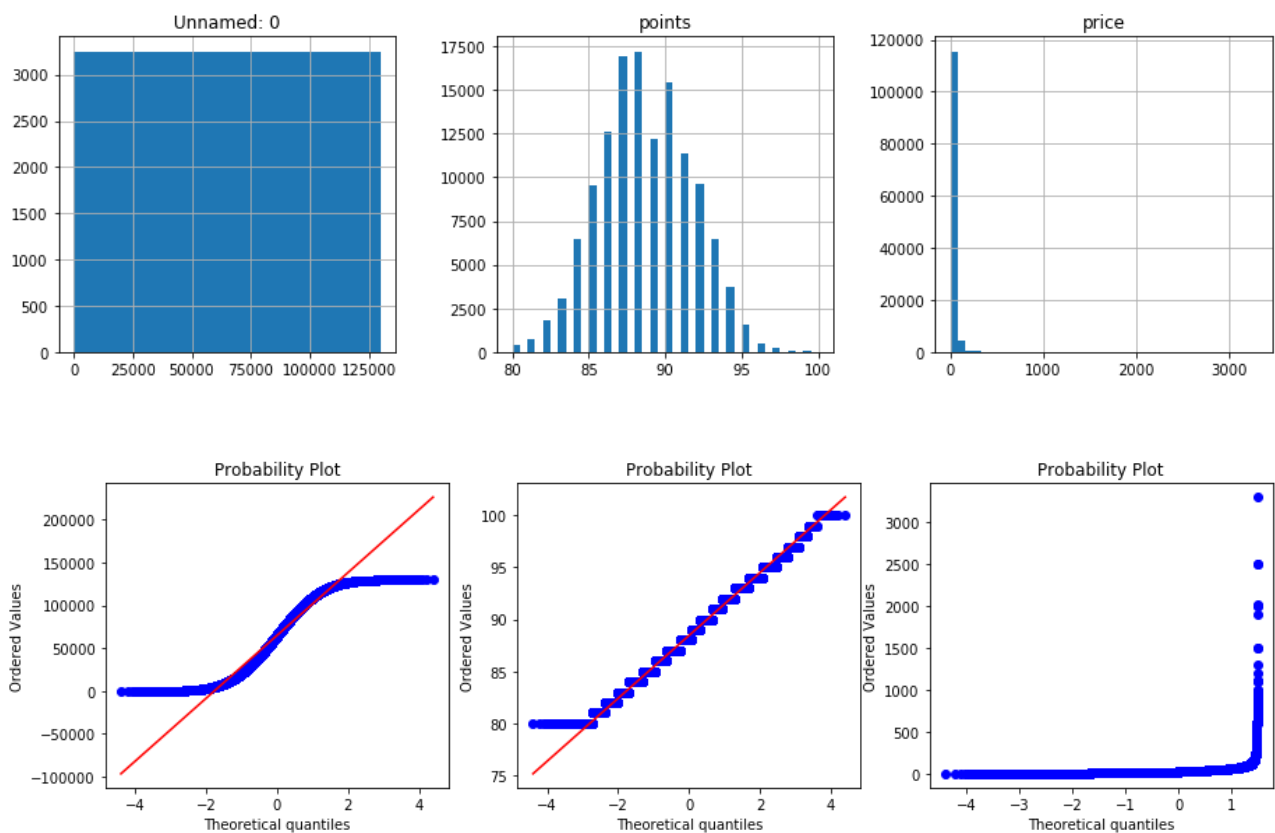
```
...
{"Monday":6703,"Tuesday":6890,"Wednesday":6458,"Thursday":5583,"Friday":6205,"Saturday":6447,"Sunday":5328}
1
{"Monday":1630,"Tuesday":1606,"Wednesday":1516,"Thursday":1419,"Friday":1481,"Saturday":1389,"Sunday":1378}
1
{"Monday":1841,"Tuesday":1781,"Wednesday":1777,"Thursday":1459,"Friday":1658,"Saturday":1555,"Sunday":1620}
1
{"Monday":2323,"Tuesday":2326,"Wednesday":2257,"Thursday":1919,"Friday":2024,"Saturday":1778,"Sunday":1639}
1
{"Monday":3309,"Tuesday":3273,"Wednesday":3131,"Thursday":2605,"Friday":2941,"Saturday":2582,"Sunday":2668}
1
{"Monday":6806,"Tuesday":6920,"Wednesday":7017,"Thursday":5832,"Friday":6001,"Saturday":5775,"Sunday":3886}
1
{"Monday":2220,"Tuesday":2320,"Wednesday":2375,"Thursday":1825,"Friday":2174,"Saturday":1839,"Sunday":1535}
1
{"Monday":3855,"Tuesday":4114,"Wednesday":4001,"Thursday":3142,"Friday":3406,"Saturday":2408,"Sunday":2172}
1
{"Monday":2858,"Tuesday":3018,"Wednesday":3044,"Thursday":2672,"Friday":3411,"Saturday":2971,"Sunday":3203}
1
{"Monday":9256,"Tuesday":9058,"Wednesday":9614,"Thursday":7439,"Friday":7429,"Saturday":6056,"Sunday":5109}
1
{"Monday":10901,"Tuesday":11810,"Wednesday":11876,"Thursday":9410,"Friday":9575,"Saturday":9961,"Sunday":8117}
1
{"Monday":25359,"Tuesday":24922,"Wednesday":24467,"Thursday":20749,"Friday":23199,"Saturday":20968,"Sunday":18
616}
1
{"Monday":4031,"Tuesday":4150,"Wednesday":4065,"Thursday":3214,"Friday":3780,"Saturday":3976,"Sunday":3766}
1
{"Monday":16942,"Tuesday":17263,"Wednesday":17418,"Thursday":14677,"Friday":16843,"Saturday":14802,"Sunday":12
179}
1
{"Monday":3528,"Tuesday":3317,"Wednesday":3291,"Thursday":2819,"Friday":3171,"Saturday":2777,"Sunday":2737}
1
{"Monday":4945,"Tuesday":5065,"Wednesday":5148,"Thursday":4132,"Friday":4689,"Saturday":4177,"Sunday":3483}
1
{"Monday":5828,"Tuesday":6142,"Wednesday":5814,"Thursday":5176,"Friday":5553,"Saturday":5282,"Sunday":4985}
1
{"Monday":5061,"Tuesday":4787,"Wednesday":4851,"Thursday":3979,"Friday":4053,"Saturday":3758,"Sunday":3892}
1
{"Monday":11003,"Tuesday":10826,"Wednesday":11327,"Thursday":9276,"Friday":10247,"Saturday":9578,"Sunday":8325}
}
1
{"Monday":727,"Tuesday":692,"Wednesday":732,"Thursday":561,"Friday":671,"Saturday":882,"Sunday":983}
1
{"Monday":5529,"Tuesday":5869,"Wednesday":6051,"Thursday":4842,"Friday":5224,"Saturday":4532,"Sunday":4555}
1
{"Monday":3733,"Tuesday":3659,"Wednesday":3670,"Thursday":3054,"Friday":4025,"Saturday":4938,"Sunday":4143}
1
{"Monday":8469,"Tuesday":7936,"Wednesday":8532,"Thursday":6234,"Friday":7474,"Saturday":7301,"Sunday":6697}
1
{"Monday":730,"Tuesday":678,"Wednesday":647,"Thursday":549,"Friday":640,"Saturday":802,"Sunday":656}
1
{"Monday":5191,"Tuesday":5412,"Wednesday":5188,"Thursday":4164,"Friday":4849,"Saturday":3445,"Sunday":3199}
1
{"Monday":11199,"Tuesday":11003,"Wednesday":10740,"Thursday":9261,"Friday":11040,"Saturday":8815,"Sunday":7840}
}
1
{"Monday":2902,"Tuesday":2958,"Wednesday":2811,"Thursday":2249,"Friday":2538,"Saturday":2409,"Sunday":2269}
1
{"Monday":8245,"Tuesday":8568,"Wednesday":8507,"Thursday":7193,"Friday":7604,"Saturday":6538,"Sunday":5539}
1
{"Monday":2819,"Tuesday":2883,"Wednesday":2753,"Thursday":2309,"Friday":2653,"Saturday":2403,"Sunday":2062}
1
{"Monday":4450,"Tuesday":4160,"Wednesday":4493,"Thursday":3451,"Friday":4060,"Saturday":3743,"Sunday":2699}
1
Name: popularity_by_day, dtype: int64]
```

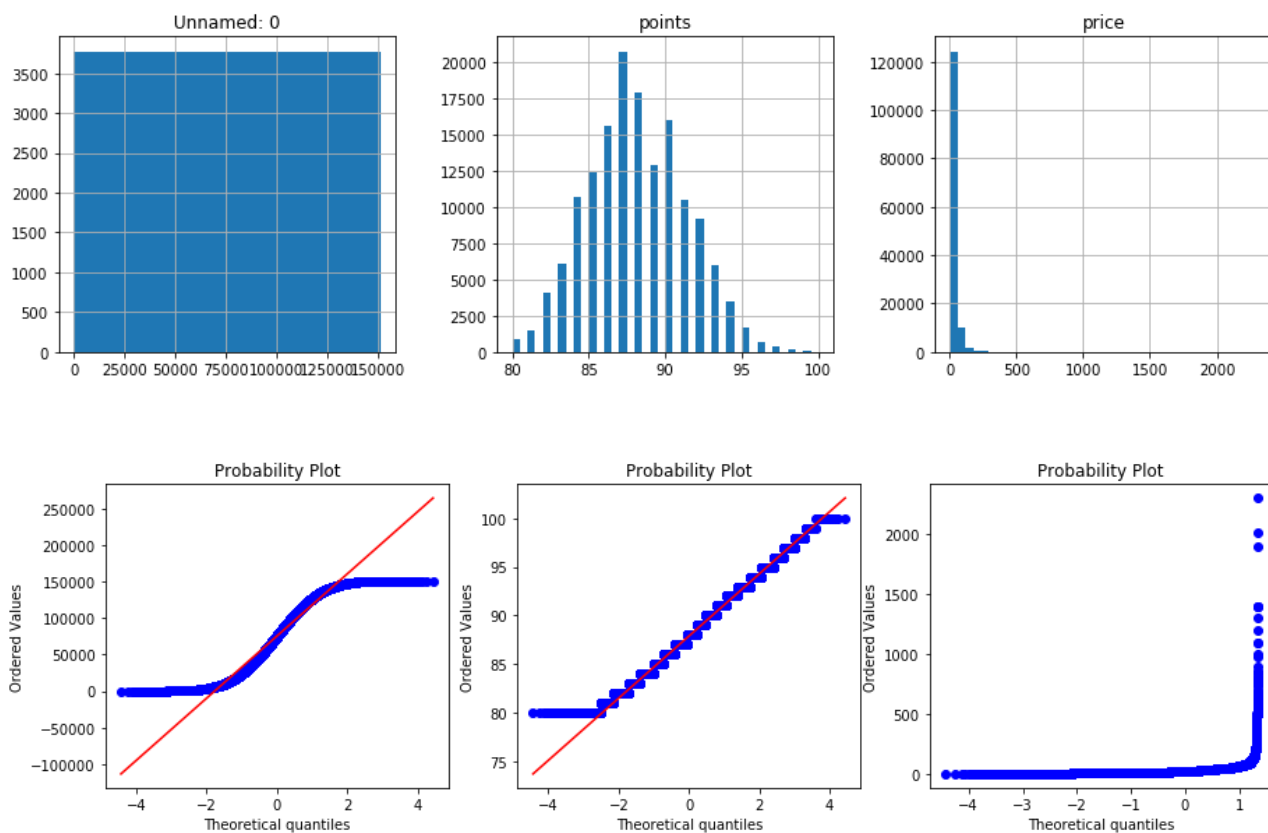
3.2 数据可视化

对于两个数据集的全部数值属性，使用matplotlib库进行可视化，绘制直方图、qq图及盒图，检查数据分布及离群点，程序及结果如下所示：

```
dataload1.hist(layout=(1,3),bins=40,figsize=(15,4))
l = ['Unnamed: 0','points','price']
plt.figure(figsize=(15, 4))
for i in range(len(l)):
    plt.subplot(130+i+1)
    stats.probplot(dataload1[l[i]],plot=plt)
plt.show()
dataload2.hist(layout=(1,3),bins=40,figsize=(15,4))
plt.figure(figsize=(15, 4))
for i in range(len(l)):
    plt.subplot(130+i+1)
    stats.probplot(dataload2[l[i]],plot=plt)
plt.show()
```

```
D:\Anaconda3\lib\site-packages\scipy\stats\_distn_infrastructure.py:875: RuntimeWarning: invalid value
encountered in greater
    return (self.a < x) & (x < self.b)
D:\Anaconda3\lib\site-packages\scipy\stats\_distn_infrastructure.py:875: RuntimeWarning: invalid value
encountered in less
    return (self.a < x) & (x < self.b)
D:\Anaconda3\lib\site-packages\scipy\stats\_distn_infrastructure.py:1814: RuntimeWarning: invalid value
encountered in less_equal
    cond2 = cond0 & (x <= self.a)
```

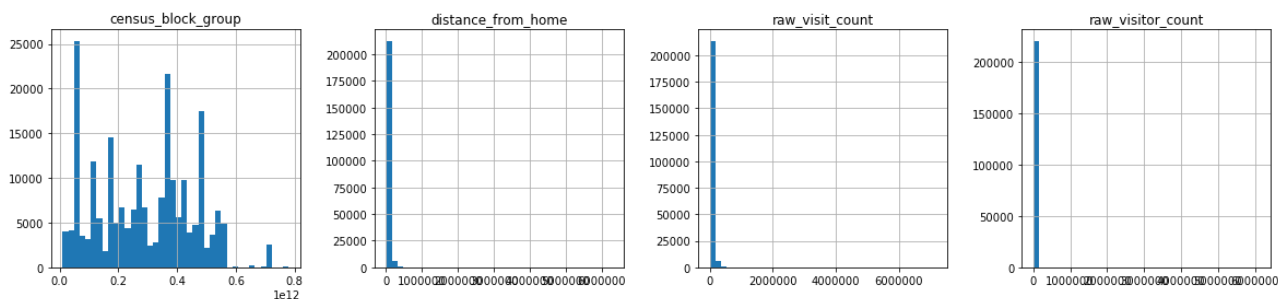


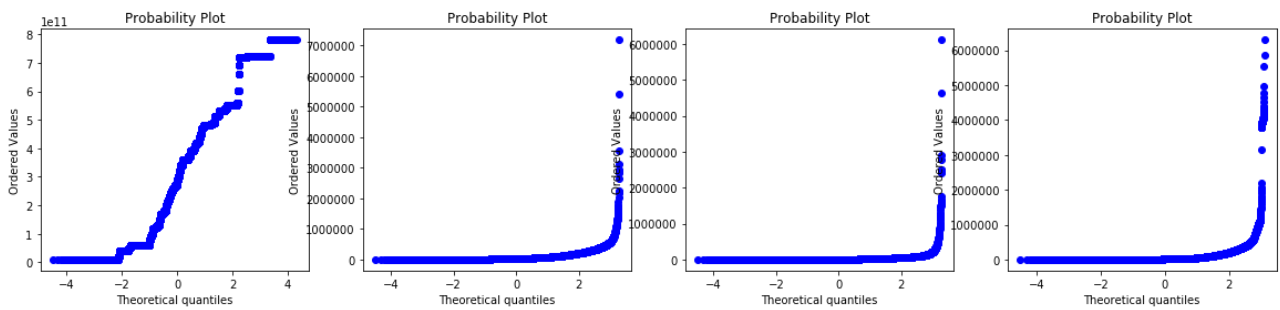


根据数据集1中数值属性绘制直方图及qq图，从图中可以看出：属性points基本满足正态分布，另外两个属性Unnamed:0和price不满足正态分布。

```
dataload3.hist(layout=(1,4),bins=40,figsize=(20,4))
l = ['census_block_group','raw_visit_count','raw_visitor_count','distance_from_home']
plt.figure(figsize=(20, 4))
for i in range(len(l)):
    plt.subplot(140+i+1)
    stats.probplot(dataload3[l[i]],plot=plt)
plt.show()
```

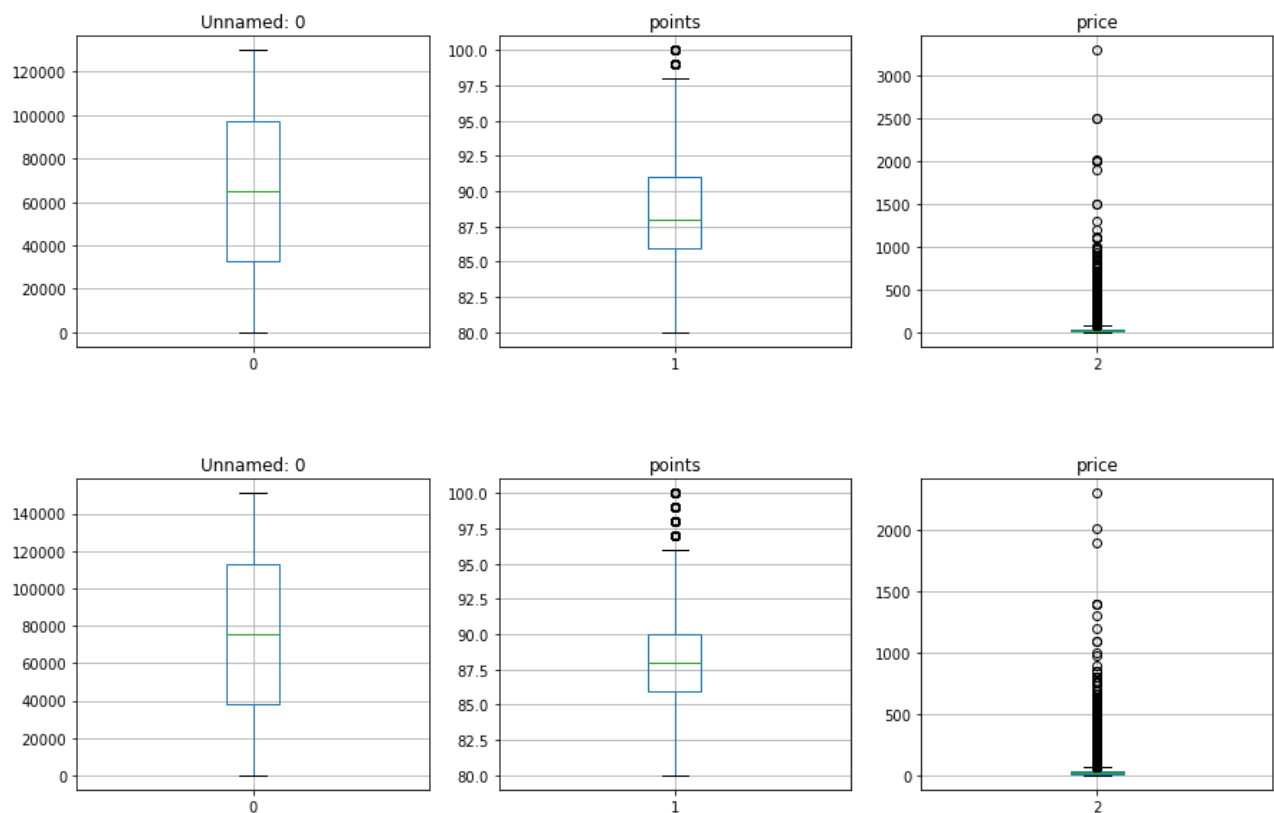
```
D:\Anaconda3\lib\site-packages\scipy\stats\_distn_infrastructure.py:875: RuntimeWarning: invalid value
encountered in greater
    return (self.a < x) & (x < self.b)
D:\Anaconda3\lib\site-packages\scipy\stats\_distn_infrastructure.py:875: RuntimeWarning: invalid value
encountered in less
    return (self.a < x) & (x < self.b)
D:\Anaconda3\lib\site-packages\scipy\stats\_distn_infrastructure.py:1814: RuntimeWarning: invalid value
encountered in less_equal
    cond2 = cond0 & (x <= self.a)
```





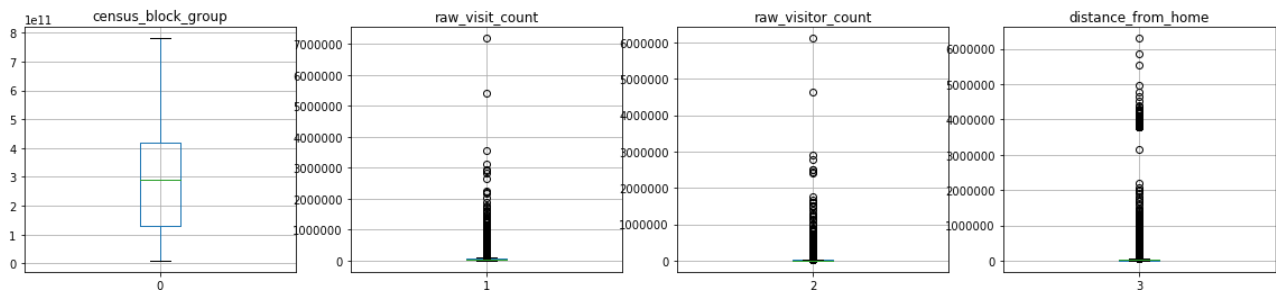
根据数据集2中数值属性绘制直方图及qq图，从图中可以看出：4个属性都不满足正态分布。

```
l = ['Unnamed: 0', 'points', 'price']
plt.figure(figsize=(15, 4))
for i in range(len(l)):
    plt.subplot(130+i+1)
    data1.boxplot(column=i)
    plt.title(l[i])
plt.show()
plt.figure(figsize=(15, 4))
for i in range(len(l)):
    plt.subplot(130+i+1)
    data2.boxplot(column=i)
    plt.title(l[i])
plt.show()
```



根据数据集1绘制的盒图可以看出，属性Unnamed:0不存在离散点，属性points存在少数离散点，属性price存在较多离散点。

```
l = ['census_block_group', 'raw_visit_count', 'raw_visitor_count', 'distance_from_home']
plt.figure(figsize=(20, 4))
for i in range(len(l)):
    plt.subplot(140+i+1)
    data103.boxplot(column=i)
    plt.title(l[i])
plt.show()
```



根据数据集2绘制的盒图可以看出，属性census_block_group不存在离散点，其余属性存在较多离散点。

4. 数据缺失处理

数据收集可能因为人为原因或者其他因素导致数据有缺失现象，这时候需要对数据进行缺失处理，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

借助 python 下 impute 包进行数据缺失值填补，impute 不仅高效实现了利用一些基本数值（均值、众数、中位数）进行数据填补的方法，还可以基于属性相关关系进行数据填补，以及基于 KNN 进行数据对象相似度聚类计算进行数据填补的函数，可以方便快速地辅助数据缺失值填补。得到处理后的数据与原数据进行可视化对比，程序及结果如下所示：

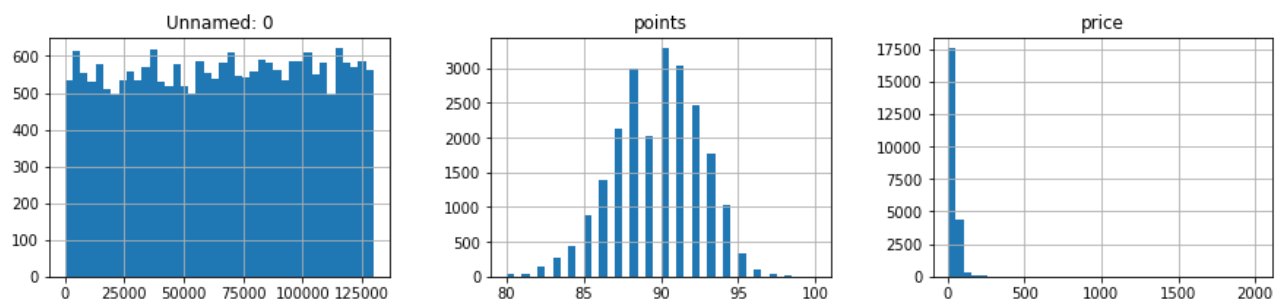
```
import impute as impy
data_drop = data103.dropna()#将缺失值剔除
data_drop.hist(layout=(1,3),bins=40,figsize=(15,3))

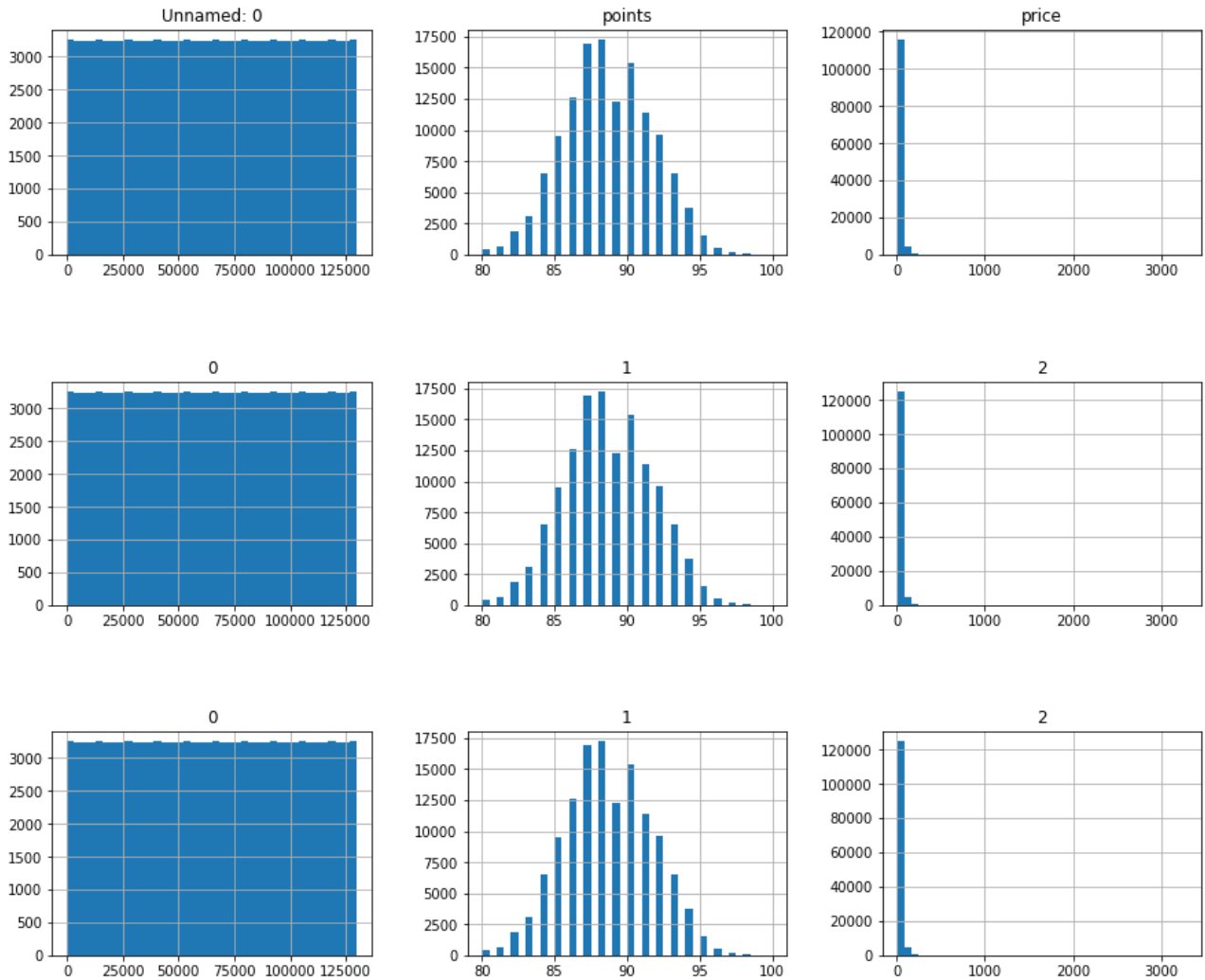
data_mode = data103.fillna(data103.mode())#用最高频率纸填补缺失值
data_mode.hist(layout=(1,3),bins=40,figsize=(15,3))

data = data103[['Unnamed: 0','points','price']]
nd = np.array(data)
filled_mice = impy.mice(nd)#通过属性的相关关系来填补缺失值
data_mice = pd.DataFrame(filled_mice)
data_mice.hist(layout=(1,3),bins=40,figsize=(15,3))

filled_knn = impy.fast_knn(nd,k=3)#通过数据对象之间的相似性来填补缺失值
data_knn = pd.DataFrame(filled_knn)
data_knn.hist(layout=(1,3),bins=40,figsize=(15,3))

plt.show()
```





对数据集1进行缺失值处理得到以上直方图，从图中可以看出：当缺失值数占可用数据比例较大时，删除缺失值会对原始数据分布造成较大影响，其他方法对数据影响程度较小，可以保持原始的数据分布情况。

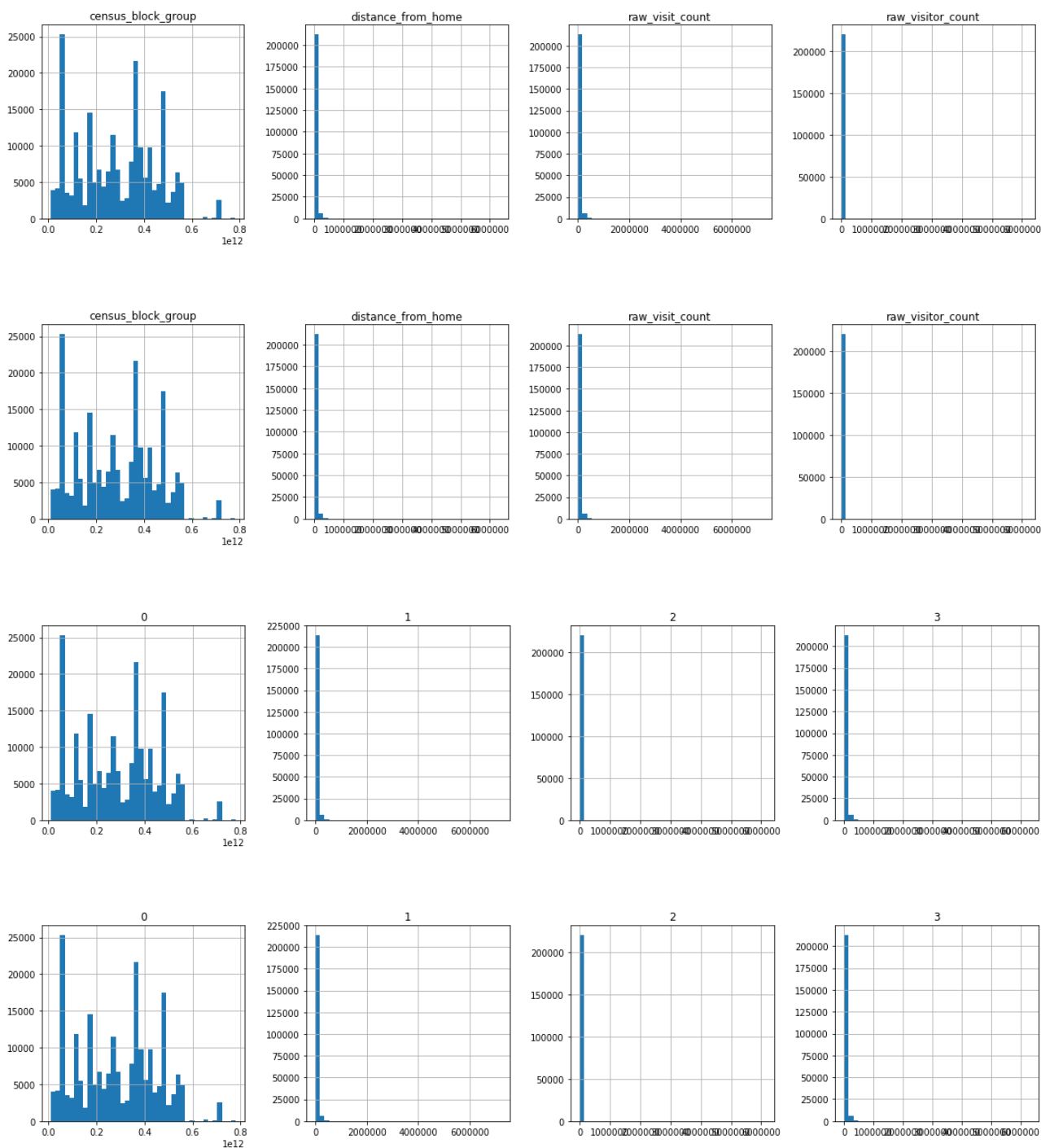
```
import impyute as impy
data_drop = data1.dropna()#将缺失值剔除
data_drop.hist(layout=(1,4),bins=40,figsize=(20,4))

data_mode = data1.fillna(data1.mode())#用最高频率纸填补缺失值
data_mode.hist(layout=(1,4),bins=40,figsize=(20,4))

data = data1[['census_block_group','raw_visit_count','raw_visitor_count','distance_from_home']]
nd = np.array(data)
filled_mice = impy.mice(nd)#通过属性的相关关系来填补缺失值
data_mice = pd.DataFrame(filled_mice)
data_mice.hist(layout=(1,4),bins=40,figsize=(20,4))

filled_knn = impy.fast_knn(nd,k=3)#通过数据对象之间的相似性来填补缺失值
data_knn = pd.DataFrame(filled_knn)
data_knn.hist(layout=(1,4),bins=40,figsize=(20,4))

plt.show()
```



对数据集2进行缺失值处理得到以上直方图，从图中可以看出：删除缺失值会对原始数据分布造成较大影响，其他方法对数据影响程度较小，可以保持原始的数据分布情况。

总结4种处理缺失值方法的适用情况：

* 将缺失部分剔除：适用于缺失值数据占可用数据比例较少的情况，否则删除缺失值会对原始数据分布造成较大影响

* 用最高频率值来填补缺失值：填补缺失值最简便和快捷的方法就是使用一些代表中心趋势的值。而代表中心趋势的值反映了变量分布的最常见值，因此中心区是指是最好的选择。有多个代表数据中心趋势的指标：平均值、中位数、众数等。对于正态分布来说，所有的观测值都较好的聚集在平均值周围，平均值就是最好的选择。然而，对于偏态分布，或者说有离群值的变量来说，选择平均值就不好。偏态分布的大部分数值都聚集在变量分布的一侧，因此不能选择平均值。另一方面，离群值（极值）的存在会扭曲平均值（比如：1、2、3、30的平均值是 9）。对于偏态分布或者有离群值的分布而言，中位数是更好的代表数据中心趋势的指标。所以最后还得看数据的分布来选择一些代表中心趋势的值。

* 通过属性的相关关系来填补缺失值：如果某两变量之间具有明显的一定相关关系，那么我们可以根据另一个未确实的值来预测所缺失的那一个值。

* 通过数据对象之间的相似性来填补缺失值：根据KNN算法找到任何案例最近的k个邻居，在K在最近邻案例中通过设定函数值（一般会选取均值、中位数、众数等）来填充缺失值