# data_mining2

2022 年 4 月 1 日

## 1 引入必要的库

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
from efficient_apriori import apriori
```

---

## 2 分析 Wine Reviews 数据集的 winemag-data_first150k.csv’ 文件

### 2.1 csv 文件路径

[2]:
```
csv_file_path = 'D:/Data/data_mining/1/Wine Reviews/winemag-data_first150k.csv'
```

### 2.2 读取数据，并查看前 10 行数据

[3]:
```
data_frame = pd.read_csv(csv_file_path)
data_frame = data_frame.dropna(how='any').astype(str)
print(data_frame[:10])
```

```
   Unnamed: 0 country                                        description  \
0           0      US  This tremendous 100% varietal wine hails from …
2           2      US  Mac Watson honors the memory of a wine once ma…
3           3      US  This spent 20 months in 30% new French oak, an…
8           8      US  This re-named vineyard was formerly bottled as…
9           9      US  The producer sources from two blocks of the vi…
11         11      US  From 18-year-old vines, this supple well-balan…
12         12      US  A standout even in this terrific lineup of 201…
14         14      US  With its sophisticated mix of mineral, acid an…
```

```
15         15      US  First made in 2006, this succulent luscious Ch…
16         16      US  This blockbuster, powerhouse of a wine suggest…

                     designation points  price    province  \
0                Martha's Vineyard     96  235.0  California
2     Special Selected Late Harvest     96   90.0  California
3                           Reserve     96   65.0      Oregon
8                            Silice     95   65.0      Oregon
9               Gap's Crown Vineyard     95   60.0  California
11  Estate Vineyard Wadensvil Block     95   48.0      Oregon
12                   Weber Vineyard     95   48.0      Oregon
14                   Grace Vineyard     95  185.0      Oregon
15                           Sigrid     95   90.0      Oregon
16                  Rainin Vineyard     95  325.0  California

                    region_1           region_2             variety  \
0                Napa Valley               Napa  Cabernet Sauvignon
2              Knights Valley             Sonoma     Sauvignon Blanc
3           Willamette Valley  Willamette Valley           Pinot Noir
8          Chehalem Mountains  Willamette Valley           Pinot Noir
9                Sonoma Coast             Sonoma           Pinot Noir
11               Ribbon Ridge  Willamette Valley           Pinot Noir
12               Dundee Hills  Willamette Valley           Pinot Noir
14               Dundee Hills  Willamette Valley           Pinot Noir
15          Willamette Valley  Willamette Valley           Chardonnay
16  Diamond Mountain District               Napa  Cabernet Sauvignon

                    winery
0                    Heitz
2                 Macauley
3                    Ponzi
8                 Bergström
9                 Blue Farm
11  Patricia Green Cellars
12  Patricia Green Cellars
14          Domaine Serene
15                Bergström
```

```
16                    Hall
```

### 2.3 选择出标称属性对应的列

[4]:
```
nominal_columns = [1, 3, 6, 7, 8, 9, 10]
data_frame = data_frame[[column for column in data_frame.
 ↪columns[nominal_columns]]]
```

## 3 预处理成适合进行关联规则挖掘的形式

[5]:
```
apriori_data = []
for _, data in data_frame.iterrows():
    apriori_data.append(data)
print(apriori_data[:10])
```

```
[country                       US
designation     Martha's Vineyard
province                California
region_1              Napa Valley
region_2                     Napa
variety         Cabernet Sauvignon
winery                       Heitz
Name: 0, dtype: object, country                                    US
designation    Special Selected Late Harvest
province                           California
region_1                     Knights Valley
region_2                             Sonoma
variety                     Sauvignon Blanc
winery                             Macauley
Name: 2, dtype: object, country                        US
designation               Reserve
province                   Oregon
region_1       Willamette Valley
region_2       Willamette Valley
variety                Pinot Noir
winery                      Ponzi
Name: 3, dtype: object, country                         US
```

```
designation                Silice
province                   Oregon
region_1       Chehalem Mountains
region_2        Willamette Valley
variety                Pinot Noir
winery                  Bergström
Name: 8, dtype: object, country                          US
designation    Gap's Crown Vineyard
province                California
region_1              Sonoma Coast
region_2                    Sonoma
variety                 Pinot Noir
winery                   Blue Farm
Name: 9, dtype: object, country                                  US
designation    Estate Vineyard Wadensvil Block
province                                Oregon
region_1                       Ribbon Ridge
region_2                    Willamette Valley
variety                           Pinot Noir
winery                  Patricia Green Cellars
Name: 11, dtype: object, country                             US
designation            Weber Vineyard
province                      Oregon
region_1                Dundee Hills
region_2           Willamette Valley
variety                   Pinot Noir
winery         Patricia Green Cellars
Name: 12, dtype: object, country                        US
designation       Grace Vineyard
province                  Oregon
region_1            Dundee Hills
region_2       Willamette Valley
variety                Pinot Noir
winery            Domaine Serene
Name: 14, dtype: object, country                        US
designation               Sigrid
province                  Oregon
```

```
region_1       Willamette Valley
region_2       Willamette Valley
variety               Chardonnay
winery                 Bergström
Name: 15, dtype: object, country                                US
designation              Rainin Vineyard
province                      California
region_1       Diamond Mountain District
region_2                           Napa
variety               Cabernet Sauvignon
winery                             Hall
Name: 16, dtype: object]
```

### 3.1 设置显示的最大数目

[6]:
```
max_visual_num = 30
```

## 4 频繁模式

[7]:
```
item_sets, rules = apriori(apriori_data, min_support=0.005, min_confidence=0.3)
print({key: {key2: item_sets[key][key2]
             for key2 in list(item_sets[key].keys())[:max_visual_num //␣
 ↪len(list(item_sets.keys()))]}
      for key in list(item_sets.keys())})
```

```
{1: {('US',): 39241, ('California',): 28557, ('Napa Valley',): 3510, ('Napa',):
5108, ('Cabernet Sauvignon',): 4896, ('Sonoma',): 7786}, 2: {('Alexander
Valley', 'Cabernet Sauvignon'): 302, ('Alexander Valley', 'California'): 783,
('Alexander Valley', 'Sonoma'): 783, ('Alexander Valley', 'US'): 783, ('Amador
County', 'California'): 317, ('Amador County', 'Sierra Foothills'): 317}, 3:
{('Alexander Valley', 'Cabernet Sauvignon', 'California'): 302, ('Alexander
Valley', 'Cabernet Sauvignon', 'Sonoma'): 302, ('Alexander Valley', 'Cabernet
Sauvignon', 'US'): 302, ('Alexander Valley', 'California', 'Sonoma'): 783,
('Alexander Valley', 'California', 'US'): 783, ('Alexander Valley', 'Sonoma',
'US'): 783}, 4: {('Alexander Valley', 'Cabernet Sauvignon', 'California',
'Sonoma'): 302, ('Alexander Valley', 'Cabernet Sauvignon', 'California', 'US'):
302, ('Alexander Valley', 'Cabernet Sauvignon', 'Sonoma', 'US'): 302,
```

```
('Alexander Valley', 'California', 'Sonoma', 'US'): 783, ('Amador County',
'California', 'Sierra Foothills', 'US'): 317, ('Amador County', 'California',
'Sierra Foothills', 'Zinfandel'): 202}, 5: {('Alexander Valley', 'Cabernet
Sauvignon', 'California', 'Sonoma', 'US'): 302, ('Amador County', 'California',
'Sierra Foothills', 'US', 'Zinfandel'): 202, ('Anderson Valley', 'California',
'Mendocino/Lake Counties', 'Pinot Noir', 'US'): 354, ('Bordeaux-style Red
Blend', 'California', 'Napa', 'Napa Valley', 'US'): 385, ('Bordeaux-style Red
Blend', 'Columbia Valley', 'Columbia Valley (WA)', 'US', 'Washington'): 253,
('Cabernet Sauvignon', 'California', 'Central Coast', 'Paso Robles', 'US'):
231}}
```

# 5 关联规则

[8]:
```
rules = rules[:max_visual_num]
```

## 5.1 依次显示关联规则，及其支持度、置信度、Lift 评价和卡方评价

[9]:
```
supports, confidences, lifts, convictions = [], [], [], []
for i, rule in enumerate(rules):
    rules[i] = str(rules[i])
    print(rule)
    # 支持度
    supports.append(rule.support)
    print('support:', supports[-1])
    # 置信度
    confidences.append(rule.confidence)
    print('confidence:', confidences[-1])
    # Lift 评价
    lifts.append(rule.lift)
    print('Lift:', lifts[-1])
    # 卡方评价
    convictions.append(rule.conviction)
    print('Conviction:', convictions[-1])
```

```
{Alexander Valley} -> {Cabernet Sauvignon} (conf: 0.386, supp: 0.008, lift:
3.091, conv: 1.425)
support: 0.007696032211207665
```

confidence: 0.38569604086845466
Lift: 3.091319105334769
Conviction: 1.4247548350651855
{Alexander Valley} -> {California} (conf: 1.000, supp: 0.020, lift: 1.374, conv:
272266252.134)
support: 0.01995361993832981
confidence: 1.0
Lift: 1.3741289351122317
Conviction: 272266252.13424736
{Alexander Valley} -> {Sonoma} (conf: 1.000, supp: 0.020, lift: 5.040, conv:
801585076.833)
support: 0.01995361993832981
confidence: 1.0
Lift: 5.039943488312355
Conviction: 801585076.8329043
{Alexander Valley} -> {US} (conf: 1.000, supp: 0.020, lift: 1.000, conv: 0.000)
support: 0.01995361993832981
confidence: 1.0
Lift: 1.0
Conviction: 0.0
{Amador County} -> {California} (conf: 1.000, supp: 0.008, lift: 1.374, conv:
272266252.134)
support: 0.008078285466731224
confidence: 1.0
Lift: 1.3741289351122317
Conviction: 272266252.13424736
{Amador County} -> {Sierra Foothills} (conf: 1.000, supp: 0.008, lift: 34.727,
conv: 971203588.084)
support: 0.008078285466731224
confidence: 1.0
Lift: 34.726548672566366
Conviction: 971203588.0838917
{Amador County} -> {US} (conf: 1.000, supp: 0.008, lift: 1.000, conv: 0.000)
support: 0.008078285466731224
confidence: 1.0
Lift: 1.0
Conviction: 0.0

{Amador County} -> {Zinfandel} (conf: 0.637, supp: 0.005, lift: 8.858, conv: 2.558)
support: 0.005147677174383935
confidence: 0.637223974763407
Lift: 8.857706692770403
Conviction: 2.5582173853605465
{Anderson Valley} -> {California} (conf: 1.000, supp: 0.013, lift: 1.374, conv: 272266252.134)
support: 0.012767258734486888
confidence: 1.0
Lift: 1.3741289351122314
Conviction: 272266252.13424736
{Mendocino/Lake Counties} -> {Anderson Valley} (conf: 0.332, supp: 0.013, lift: 26.022, conv: 1.478)
support: 0.012767258734486888
confidence: 0.3322281167108753
Lift: 26.021883289124666
Conviction: 1.4783981843088079
{Anderson Valley} -> {Mendocino/Lake Counties} (conf: 1.000, supp: 0.013, lift: 26.022, conv: 961570806.045)
support: 0.012767258734486888
confidence: 1.0
Lift: 26.021883289124666
Conviction: 961570806.0446981
{Anderson Valley} -> {Pinot Noir} (conf: 0.707, supp: 0.009, lift: 3.719, conv: 2.761)
support: 0.009021176830356005
confidence: 0.7065868263473054
Lift: 3.7187732903292128
Conviction: 2.760593996548355
{Anderson Valley} -> {US} (conf: 1.000, supp: 0.013, lift: 1.000, conv: 0.000)
support: 0.012767258734486888
confidence: 1.0
Lift: 1.0
Conviction: 0.0
{Bordeaux-style Red Blend} -> {California} (conf: 0.625, supp: 0.026, lift: 0.859, conv: 0.727)

support: 0.026248056879284422
confidence: 0.6253794778384942
Lift: 0.8593520359232536
Conviction: 0.72677879427569
{Bordeaux-style Red Blend} -> {Columbia Valley} (conf: 0.306, supp: 0.013, lift: 2.076, conv: 1.229)
support: 0.0128437093855916
confidence: 0.30601092896174864
Lift: 2.075743278027308
Conviction: 1.2285174154338194
{Bordeaux-style Red Blend} -> {US} (conf: 1.000, supp: 0.042, lift: 1.000, conv: 0.000)
support: 0.04197140745648684
confidence: 1.0
Lift: 1.0
Conviction: 0.0
{Bordeaux-style Red Blend} -> {Washington} (conf: 0.318, supp: 0.013, lift: 2.039, conv: 1.238)
support: 0.013353380392956346
confidence: 0.31815421979356406
Lift: 2.0393155404964465
Conviction: 1.2378014617210407
{Cabernet Franc} -> {California} (conf: 0.523, supp: 0.006, lift: 0.719, conv: 0.571)
support: 0.006090568538008715
confidence: 0.5229759299781181
Lift: 0.7186363577501605
Conviction: 0.5707599860757511
{Cabernet Franc} -> {US} (conf: 1.000, supp: 0.012, lift: 1.000, conv: 0.000)
support: 0.011645982518284448
confidence: 1.0
Lift: 1.0
Conviction: 0.0
{Cabernet Sauvignon} -> {California} (conf: 0.827, supp: 0.103, lift: 1.136, conv: 1.570)
support: 0.10313192834025636
confidence: 0.8265931372549019

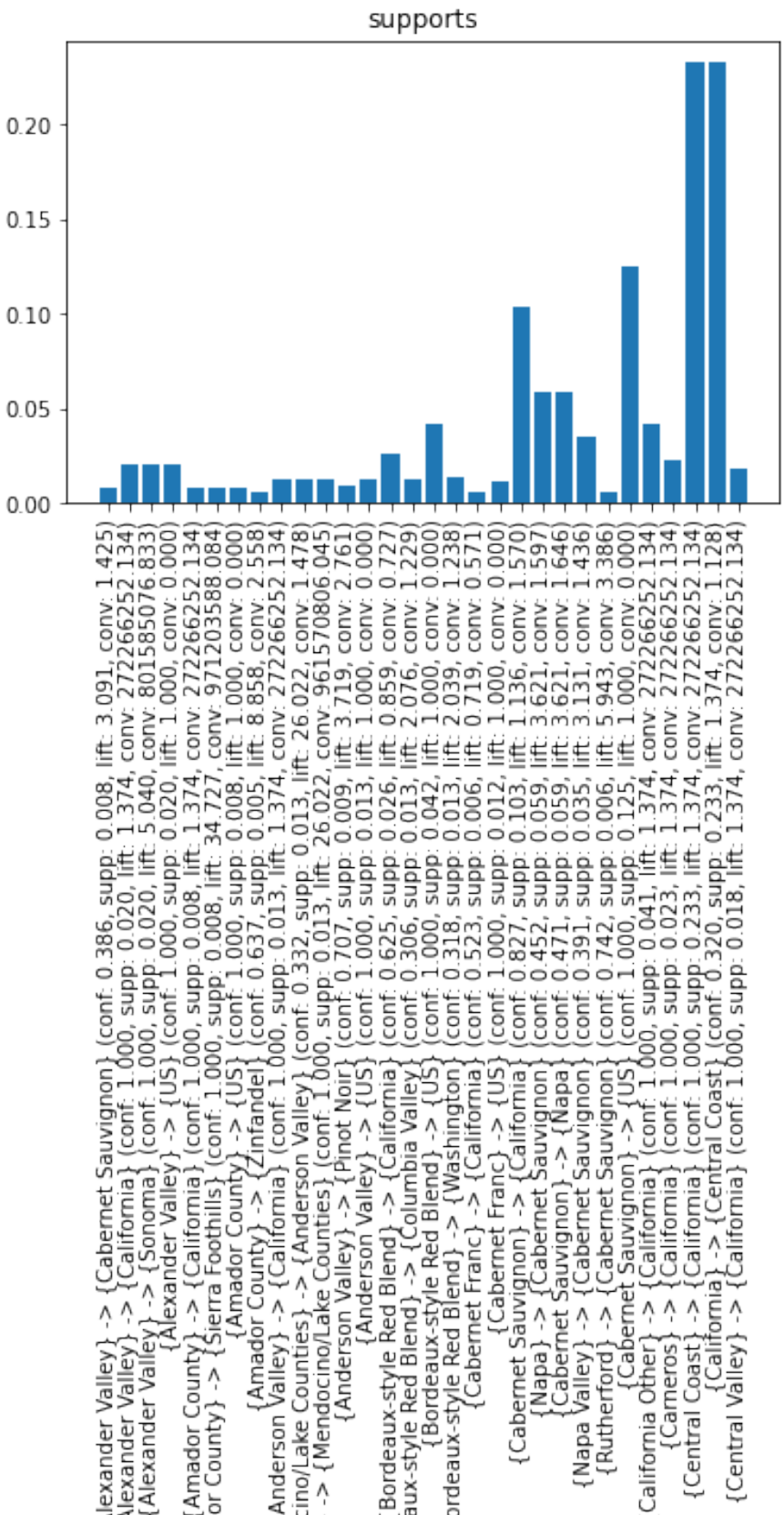
Lift: 1.1358455474671572
Conviction: 1.5701007806384704
{Napa} -> {Cabernet Sauvignon} (conf: 0.452, supp: 0.059, lift: 3.621, conv:
1.597)
support: 0.058816034249891694
confidence: 0.45184025058731403
Lift: 3.6214590018988537
Conviction: 1.5966742117387953
{Cabernet Sauvignon} -> {Napa} (conf: 0.471, supp: 0.059, lift: 3.621, conv:
1.646)
support: 0.058816034249891694
confidence: 0.4714052287581699
Lift: 3.6214590018988537
Conviction: 1.645551697437333
{Napa Valley} -> {Cabernet Sauvignon} (conf: 0.391, supp: 0.035, lift: 3.131,
conv: 1.436)
support: 0.03493794755485334
confidence: 0.3905982905982906
Lift: 3.130610196357745
Conviction: 1.43621608285308
{Rutherford} -> {Cabernet Sauvignon} (conf: 0.742, supp: 0.006, lift: 5.943,
conv: 3.386)
support: 0.00614153563874519
confidence: 0.7415384615384616
Lift: 5.943364127702362
Conviction: 3.3863163518274764
{Cabernet Sauvignon} -> {US} (conf: 1.000, supp: 0.125, lift: 1.000, conv:
0.000)
support: 0.12476746260288983
confidence: 1.0
Lift: 1.0
Conviction: 0.0
{California Other} -> {California} (conf: 1.000, supp: 0.041, lift: 1.374, conv:
272266252.134)
support: 0.04133431869728091
confidence: 1.0
Lift: 1.3741289351122317

```
Conviction: 272266252.13424736
{Carneros} -> {California} (conf: 1.000, supp: 0.023, lift: 1.374, conv:
272266252.134)
support: 0.022833261129940625
confidence: 1.0
Lift: 1.3741289351122319
Conviction: 272266252.13424736
{Central Coast} -> {California} (conf: 1.000, supp: 0.233, lift: 1.374, conv:
272266252.134)
support: 0.2331744858693713
confidence: 1.0
Lift: 1.3741289351122317
Conviction: 272266252.13424736
{California} -> {Central Coast} (conf: 0.320, supp: 0.233, lift: 1.374, conv:
1.128)
support: 0.2331744858693713
confidence: 0.3204118079630213
Lift: 1.3741289351122317
Conviction: 1.1283679174939745
{Central Valley} -> {California} (conf: 1.000, supp: 0.018, lift: 1.374, conv:
272266252.134)
support: 0.01842460691623557
confidence: 1.0
Lift: 1.3741289351122317
Conviction: 272266252.13424736
```

# 6 可视化

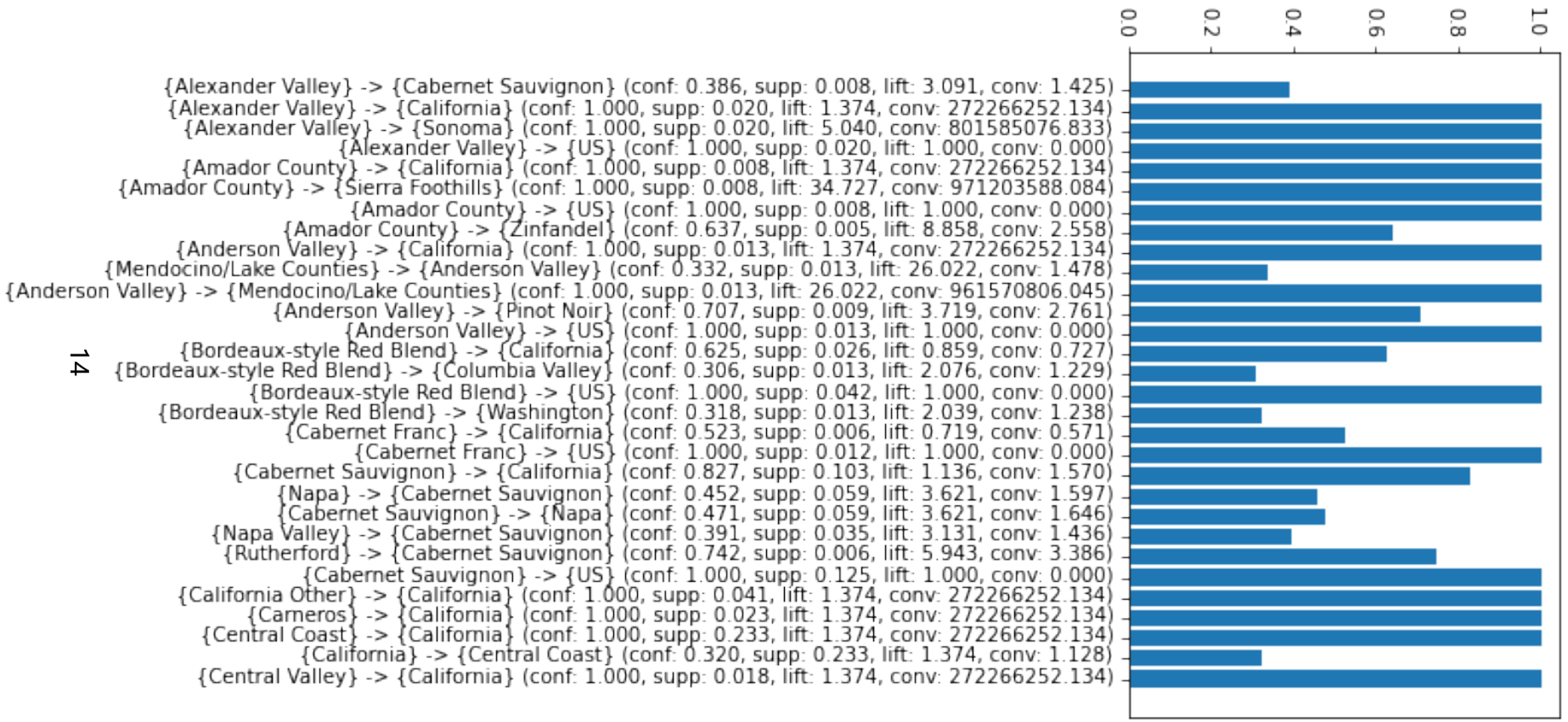
## 6.1 支持度

[10]:
```
plt.title('supports')
plt.xticks(range(len(rules)), rules, rotation=90)
plt.bar(range(len(supports)), supports)
plt.show()
```


supports
0.20
0.15
0.10
0.05
0.00
{Alexander Valley} -> {Cabernet Sauvignon} (conf: 0.386, supp: 0.008, lift: 3.091, conv: 1.425)
{Alexander Valley} -> {California} (conf: 1.000, supp: 0.020, lift: 1.374, conv: 272266252.134)
{Alexander Valley} -> {Sonoma} (conf: 1.000, supp: 0.020, lift: 5.040, conv: 801585076.833)
{Alexander Valley} -> {US} (conf: 1.000, supp: 0.020, lift: 1.000, conv: 0.000)
{Amador County} -> {California} (conf: 1.000, supp: 0.008, lift: 1.374, conv: 272266252.134)
{Amador County} -> {Sierra Foothills} (conf: 1.000, supp: 0.008, lift: 34.727, conv: 971203588.084)
{Amador County} -> {US} (conf: 1.000, supp: 0.008, lift: 1.000, conv: 0.000)
{Amador County} -> {Zinfandel} (conf: 0.637, supp: 0.005, lift: 8.858, conv: 2.558)
{Anderson Valley} -> {California} (conf: 1.000, supp: 0.013, lift: 1.374, conv: 272266252.134)
{Mendocino/Lake Counties} -> {Anderson Valley} (conf: 0.332, supp: 0.013, lift: 26.022, conv: 1.478)
{Anderson Valley} -> {Mendocino/Lake Counties} (conf: 1.000, supp: 0.013, lift: 26.022, conv: 961570806.045)
{Anderson Valley} -> {Pinot Noir} (conf: 0.707, supp: 0.009, lift: 3.719, conv: 2.761)
{Anderson Valley} -> {US} (conf: 1.000, supp: 0.013, lift: 1.000, conv: 0.000)
{Bordeaux-style Red Blend} -> {California} (conf: 0.625, supp: 0.026, lift: 0.859, conv: 0.727)
{Bordeaux-style Red Blend} -> {Columbia Valley} (conf: 0.306, supp: 0.013, lift: 2.076, conv: 1.229)
{Bordeaux-style Red Blend} -> {US} (conf: 1.000, supp: 0.042, lift: 1.000, conv: 0.000)
{Bordeaux-style Red Blend} -> {Washington} (conf: 0.318, supp: 0.013, lift: 2.039, conv: 1.238)
{Cabernet Franc} -> {California} (conf: 0.523, supp: 0.006, lift: 0.719, conv: 0.571)
{Cabernet Franc} -> {US} (conf: 1.000, supp: 0.012, lift: 1.000, conv: 0.000)
{Cabernet Sauvignon} -> {California} (conf: 0.827, supp: 0.103, lift: 1.136, conv: 1.570)
{Napa} -> {Cabernet Sauvignon} (conf: 0.452, supp: 0.059, lift: 3.621, conv: 1.597)
{Cabernet Sauvignon} -> {Napa} (conf: 0.471, supp: 0.059, lift: 3.621, conv: 1.646)
{Napa Valley} -> {Cabernet Sauvignon} (conf: 0.391, supp: 0.035, lift: 3.131, conv: 1.436)
{Rutherford} -> {Cabernet Sauvignon} (conf: 0.742, supp: 0.006, lift: 5.943, conv: 3.386)
{Cabernet Sauvignon} -> {US} (conf: 1.000, supp: 0.125, lift: 1.000, conv: 0.000)
{California Other} -> {California} (conf: 1.000, supp: 0.041, lift: 1.374, conv: 272266252.134)
{Carneros} -> {California} (conf: 1.000, supp: 0.023, lift: 1.374, conv: 272266252.134)
{Central Coast} -> {California} (conf: 1.000, supp: 0.233, lift: 1.374, conv: 272266252.134)
{California} -> {Central Coast} (conf: 0.320, supp: 0.233, lift: 1.374, conv: 1.128)
{Central Valley} -> {California} (conf: 1.000, supp: 0.018, lift: 1.374, conv: 272266252.134)

### 6.2 置信度

[11]:
```
plt.title('confidences')
plt.xticks(range(len(rules)), rules, rotation=90)
plt.bar(range(len(confidences)), confidences)
plt.show()
```


{Alexander Valley} -> {Cabernet Sauvignon} (conf: 0.386, supp: 0.008, lift: 3.091, conv: 1.425)
{Alexander Valley} -> {California} (conf: 1.000, supp: 0.020, lift: 1.374, conv: 272266252.134)
{Alexander Valley} -> {Sonoma} (conf: 1.000, supp: 0.020, lift: 5.040, conv: 801585076.833)
{Alexander Valley} -> {US} (conf: 1.000, supp: 0.020, lift: 1.000, conv: 0.000)
{Amador County} -> {California} (conf: 1.000, supp: 0.008, lift: 1.374, conv: 272266252.134)
{Amador County} -> {Sierra Foothills} (conf: 1.000, supp: 0.008, lift: 34.727, conv: 971203588.084)
{Amador County} -> {US} (conf: 1.000, supp: 0.008, lift: 1.000, conv: 0.000)
{Amador County} -> {Zinfandel} (conf: 0.637, supp: 0.005, lift: 8.858, conv: 2.558)
{Anderson Valley} -> {California} (conf: 1.000, supp: 0.013, lift: 1.374, conv: 272266252.134)
{Mendocino/Lake Counties} -> {Anderson Valley} (conf: 0.332, supp: 0.013, lift: 26.022, conv: 1.478)
{Anderson Valley} -> {Mendocino/Lake Counties} (conf: 1.000, supp: 0.013, lift: 26.022, conv: 961570806.045)
{Anderson Valley} -> {Pinot Noir} (conf: 0.707, supp: 0.009, lift: 3.719, conv: 2.761)
{Anderson Valley} -> {US} (conf: 1.000, supp: 0.013, lift: 1.000, conv: 0.000)
{Bordeaux-style Red Blend} -> {California} (conf: 0.625, supp: 0.026, lift: 0.859, conv: 0.727)
{Bordeaux-style Red Blend} -> {Columbia Valley} (conf: 0.306, supp: 0.013, lift: 2.076, conv: 1.229)
{Bordeaux-style Red Blend} -> {US} (conf: 1.000, supp: 0.042, lift: 1.000, conv: 0.000)
{Bordeaux-style Red Blend} -> {Washington} (conf: 0.318, supp: 0.013, lift: 2.039, conv: 1.238)
{Cabernet Franc} -> {California} (conf: 0.523, supp: 0.006, lift: 0.719, conv: 0.571)
{Cabernet Franc} -> {US} (conf: 1.000, supp: 0.012, lift: 1.000, conv: 0.000)
{Cabernet Sauvignon} -> {California} (conf: 0.827, supp: 0.103, lift: 1.136, conv: 1.570)
{Napa} -> {Cabernet Sauvignon} (conf: 0.452, supp: 0.059, lift: 3.621, conv: 1.597)
{Cabernet Sauvignon} -> {Napa} (conf: 0.471, supp: 0.059, lift: 3.621, conv: 1.646)
{Napa Valley} -> {Cabernet Sauvignon} (conf: 0.391, supp: 0.035, lift: 3.131, conv: 1.436)
{Rutherford} -> {Cabernet Sauvignon} (conf: 0.742, supp: 0.006, lift: 5.943, conv: 3.386)
{Cabernet Sauvignon} -> {US} (conf: 1.000, supp: 0.125, lift: 1.000, conv: 0.000)
{California Other} -> {California} (conf: 1.000, supp: 0.041, lift: 1.374, conv: 272266252.134)
{Carneros} -> {California} (conf: 1.000, supp: 0.023, lift: 1.374, conv: 272266252.134)
{Central Coast} -> {California} (conf: 1.000, supp: 0.233, lift: 1.374, conv: 272266252.134)
{California} -> {Central Coast} (conf: 0.320, supp: 0.233, lift: 1.374, conv: 1.128)
{Central Valley} -> {California} (conf: 1.000, supp: 0.018, lift: 1.374, conv: 272266252.134)
0.0
0.2
0.4
0.6
0.8
1.0
confidences

### 6.3 Lifts 评价

[12]:
```
plt.title('lifts')
plt.xticks(range(len(rules)), rules, rotation=90)
plt.bar(range(len(lifts)), lifts)
plt.show()
```

lifts
35
30
25
20
15
10
5
0
{Alexander Valley} -> {Cabernet Sauvignon} (conf: 0.386, supp: 0.008, lift: 3.091, conv: 1.425)
{Alexander Valley} -> {California} (conf: 1.000, supp: 0.020, lift: 1.374, conv: 272266252.134)
{Alexander Valley} -> {Sonoma} (conf: 1.000, supp: 0.020, lift: 5.040, conv: 80158507 6.833)
{Alexander Valley} -> {US} (conf: 1.000, supp: 0.020, lift: 1.000, conv: 0.000)
{Amador County} -> {California} (conf: 1.000, supp: 0.008, lift: 1.374, conv: 272266252.134)
{Amador County} -> {Sierra Foothills} (conf: 1.000, supp: 0.008, lift: 34.727, conv: 97120358 8.084)
{Amador County} -> {US} (conf: 1.000, supp: 0.008, lift: 1.000, conv: 0.000)
{Anderson Valley} -> {Zinfandel} (conf: 0.637, supp: 0.005, lift: 8.858, conv: 2.558)
{Anderson Valley} -> {California} (conf: 1.000, supp: 0.013, lift: 1.374, conv: 272266252.134)
{Mendocino/Lake Counties} -> {Anderson Valley} (conf: 0.332, supp: 0.013, lift: 26.022, conv: 1.478)
{Anderson Valley} -> {Mendocino/Lake Counties} (conf: 1.000, supp: 0.013, lift: 26.022, conv: 96157080 6.045)
{Anderson Valley} -> {Pinot Noir} (conf: 0.707, supp: 0.009, lift: 3.719, conv: 2.761)
{Anderson Valley} -> {US} (conf: 1.000, supp: 0.013, lift: 1.000, conv: 0.000)
{Bordeaux-style Red Blend} -> {California} (conf: 0.625, supp: 0.026, lift: 0.859, conv: 0.727)
{Bordeaux-style Red Blend} -> {Columbia Valley} (conf: 0.306, supp: 0.013, lift: 2.076, conv: 1.229)
{Bordeaux-style Red Blend} -> {US} (conf: 1.000, supp: 0.042, lift: 1.000, conv: 0.000)
{Bordeaux-style Red Blend} -> {Washington} (conf: 0.318, supp: 0.013, lift: 2.039, conv: 1.238)
{Cabernet Franc} -> {California} (conf: 0.523, supp: 0.006, lift: 0.719, conv: 0.571)
{Cabernet Franc} -> {US} (conf: 1.000, supp: 0.012, lift: 1.000, conv: 0.000)
{Cabernet Sauvignon} -> {California} (conf: 0.827, supp: 0.103, lift: 1.136, conv: 1.570)
{Napa} -> {Cabernet Sauvignon} (conf: 0.452, supp: 0.059, lift: 3.621, conv: 1.597)
{Cabernet Sauvignon} -> {Napa} (conf: 0.471, supp: 0.059, lift: 3.621, conv: 1.646)
{Napa Valley} -> {Cabernet Sauvignon} (conf: 0.391, supp: 0.035, lift: 3.131, conv: 1.436)
{Rutherford} -> {Cabernet Sauvignon} (conf: 0.742, supp: 0.006, lift: 5.943, conv: 3.386)
{Cabernet Sauvignon} -> {US} (conf: 1.000, supp: 0.125, lift: 1.000, conv: 0.000)
{California Other} -> {California} (conf: 1.000, supp: 0.041, lift: 1.374, conv: 272266252.134)
{Carneros} -> {California} (conf: 1.000, supp: 0.023, lift: 1.374, conv: 272266252.134)
{Central Coast} -> {California} (conf: 1.000, supp: 0.233, lift: 1.374, conv: 272266252.134)
{California} -> {Central Coast} (conf: 0.320, supp: 0.233, lift: 1.374, conv: 1.128)
{Central Valley} -> {California} (conf: 1.000, supp: 0.018, lift: 1.374, conv: 272266252.134)

### 6.4 卡方评价

[13]:
```
plt.title('convictions')
plt.xticks(range(len(rules)), rules, rotation=90)
plt.bar(range(len(convictions)), convictions)
plt.show()
```

convictions
1e9
1.0
0.8
0.6
0.4
0.2
0.0
{Alexander Valley} -> {Cabernet Sauvignon} (conf: 0.386, supp: 0.008, lift: 3.091, conv: 1.425)
{Alexander Valley} -> {California} (conf: 1.000, supp: 0.020, lift: 1.374, conv: 272266252.134)
{Alexander Valley} -> {Sonoma} (conf: 1.000, supp: 0.020, lift: 5.040, conv: 80158076.833)
{Alexander Valley} -> {US} (conf: 1.000, supp: 0.020, lift: 1.000, conv: 0.000)
{Amador County} -> {California} (conf: 1.000, supp: 0.008, lift: 1.374, conv: 272266252.134)
{Amador County} -> {Sierra Foothills} (conf: 1.000, supp: 0.008, lift: 34.727, conv: 971203588.084)
{Amador County} -> {US} (conf: 1.000, supp: 0.008, lift: 1.000, conv: 0.000)
{Amador County} -> {Zinfandel} (conf: 0.637, supp: 0.005, lift: 8.858, conv: 2.558)
{Anderson Valley} -> {California} (conf: 1.000, supp: 0.013, lift: 1.374, conv: 272266252.134)
{Mendocino/Lake Counties} -> {Anderson Valley} (conf: 0.332, supp: 0.013, lift: 26.022, conv: 1.478)
{Anderson Valley} -> {Mendocino/Lake Counties} (conf: 1.000, supp: 0.013, lift: 26.022, conv: 961570806.045)
{Anderson Valley} -> {Pinot Noir} (conf: 0.707, supp: 0.009, lift: 3.719, conv: 2.761)
{Anderson Valley} -> {US} (conf: 1.000, supp: 0.013, lift: 1.000, conv: 0.000)
{Bordeaux-style Red Blend} -> {California} (conf: 0.625, supp: 0.026, lift: 0.859, conv: 0.727)
{Bordeaux-style Red Blend} -> {Columbia Valley} (conf: 0.306, supp: 0.013, lift: 2.076, conv: 1.229)
{Bordeaux-style Red Blend} -> {US} (conf: 1.000, supp: 0.042, lift: 1.000, conv: 0.000)
{Bordeaux-style Red Blend} -> {Washington} (conf: 0.318, supp: 0.013, lift: 2.039, conv: 1.238)
{Cabernet Franc} -> {California} (conf: 0.523, supp: 0.006, lift: 0.719, conv: 0.571)
{Cabernet Franc} -> {US} (conf: 1.000, supp: 0.012, lift: 1.000, conv: 0.000)
{Cabernet Sauvignon} -> {California} (conf: 0.827, supp: 0.103, lift: 1.136, conv: 1.570)
{Napa} -> {Cabernet Sauvignon} (conf: 0.452, supp: 0.059, lift: 3.621, conv: 1.597)
{Cabernet Sauvignon} -> {Napa} (conf: 0.471, supp: 0.059, lift: 3.621, conv: 1.646)
{Napa Valley} -> {Cabernet Sauvignon} (conf: 0.391, supp: 0.035, lift: 3.131, conv: 1.436)
{Rutherford} -> {Cabernet Sauvignon} (conf: 0.742, supp: 0.006, lift: 5.943, conv: 3.386)
{Cabernet Sauvignon} -> {US} (conf: 1.000, supp: 0.125, lift: 1.000, conv: 0.000)
{California Other} -> {California} (conf: 1.000, supp: 0.041, lift: 1.374, conv: 272266252.134)
{Carneros} -> {California} (conf: 1.000, supp: 0.023, lift: 1.374, conv: 272266252.134)
{Central Coast} -> {California} (conf: 1.000, supp: 0.233, lift: 1.374, conv: 272266252.134)
{California} -> {Central Coast} (conf: 0.320, supp: 0.233, lift: 1.374, conv: 1.128)
{Central Valley} -> {California} (conf: 1.000, supp: 0.018, lift: 1.374, conv: 272266252.134)

# 7 分析 Wine Reviews 数据集的 winemag-data-130k-v2.csv’ 文件

## 7.1 csv 文件路径

[14]:
```
csv_file_path = 'D:/Data/data_mining/1/Wine Reviews/winemag-data-130k-v2.csv'
```

## 7.2 读取数据，并查看前 10 行数据

[15]:
```
data_frame = pd.read_csv(csv_file_path)
data_frame = data_frame.dropna(how='any').astype(str)
print(data_frame[:10])
```

```
   Unnamed: 0 country                                        description  \
4           4      US  Much like the regular bottling from 2012, this…
10         10      US  Soft, supple plum envelopes an oaky structure …
23         23      US  This wine from the Geneseo district offers aro…
25         25      US  Oak and earth intermingle around robust aromas…
35         35      US  As with many of the Erath 2010 vineyard design…
60         60      US  Syrupy and dense, this wine is jammy in plum a…
62         62      US  The aromas are brooding, with notes of barrel …
64         64      US  There are intriguing touches to the nose of th…
67         67      US  A blend of Merlot and Cabernet Franc, this win…
71         71      US  Big oak defines this robustly dense and extrac…

                          designation points  price    province  \
4   Vintner's Reserve Wild Child Block     87   65.0      Oregon
10                      Mountain Cuvée     87   19.0  California
23                Signature Selection      87   22.0  California
25                King Ridge Vineyard      87   69.0  California
35                             Hyland      86   50.0      Oregon
60                             Estate      86  100.0  California
62               Alder Ridge Vineyard      86   25.0  Washington
64                        Golden Horn      86   26.0  California
```

```
67                          Inspired     86   46.0  Washington
71                          Old Vine     86   40.0  California

                region_1           region_2       taster_name  \
4      Willamette Valley  Willamette Valley      Paul Gregutt
10           Napa Valley               Napa    Virginie Boone
23           Paso Robles      Central Coast     Matt Kettmann
25         Sonoma Coast              Sonoma    Virginie Boone
35           McMinnville  Willamette Valley      Paul Gregutt
60           Napa Valley               Napa    Virginie Boone
62  Columbia Valley (WA)    Columbia Valley  Sean P. Sullivan
64     Santa Ynez Valley      Central Coast     Matt Kettmann
67  Columbia Valley (WA)    Columbia Valley  Sean P. Sullivan
71      Alexander Valley              Sonoma    Virginie Boone

   taster_twitter_handle                                          title  \
4            @paulgwine   Sweet Cheeks 2012 Vintner's Reserve Wild Child…
10               @vboone  Kirkland Signature 2011 Mountain Cuvée Caberne…
23         @mattkettmann  Bianchi 2011 Signature Selection Merlot (Paso …
25               @vboone  Castello di Amorosa 2011 King Ridge Vineyard P…
35             @paulgwine          Erath 2010 Hyland Pinot Noir (McMinnville)
60               @vboone  Okapi 2013 Estate Cabernet Sauvignon (Napa Val…
62         @wawinereport  Ram 2014 Alder Ridge Vineyard Cabernet Franc (…
64         @mattkettmann  Sevtap 2015 Golden Horn Sauvignon Blanc (Santa…
67         @wawinereport  Basel Cellars 2013 Inspired Red (Columbia Vall…
71               @vboone  Eco Terreno 2013 Old Vine Cabernet Sauvignon (…

                     variety               winery
4                  Pinot Noir         Sweet Cheeks
10         Cabernet Sauvignon   Kirkland Signature
23                     Merlot              Bianchi
25                 Pinot Noir  Castello di Amorosa
35                 Pinot Noir                Erath
60         Cabernet Sauvignon                Okapi
62             Cabernet Franc                  Ram
64           Sauvignon Blanc                Sevtap
67  Bordeaux-style Red Blend        Basel Cellars
```

```
71          Cabernet Sauvignon          Eco Terreno
```

### 7.3 选择出标称属性对应的列

[16]:
```
nominal_columns = [1, 3, 6, 7, 8, 9, 10, 12, 13]
data_frame = data_frame[[column for column in data_frame.
 ↪columns[nominal_columns]]]
```

## 8 预处理成适合进行关联规则挖掘的形式

[17]:
```
apriori_data = []
for _, data in data_frame.iterrows():
    apriori_data.append(data)
print(apriori_data[:10])
```

```
[country                                                US
designation              Vintner's Reserve Wild Child Block
province                                             Oregon
region_1                                  Willamette Valley
region_2                                  Willamette Valley
taster_name                                    Paul Gregutt
taster_twitter_handle                           @paulgwine
variety                                        Pinot Noir
winery                                         Sweet Cheeks
Name: 4, dtype: object, country                                    US
designation                   Mountain Cuvée
province                          California
region_1                         Napa Valley
region_2                                Napa
taster_name                   Virginie Boone
taster_twitter_handle                @vboone
variety                   Cabernet Sauvignon
winery                    Kirkland Signature
Name: 10, dtype: object, country                                     US
designation               Signature Selection
province                           California
region_1                          Paso Robles
```

```
region_2                       Central Coast
taster_name                     Matt Kettmann
taster_twitter_handle          @mattkettmann
variety                               Merlot
winery                               Bianchi
Name: 23, dtype: object, country                                  US
designation              King Ridge Vineyard
province                           California
region_1                         Sonoma Coast
region_2                                Sonoma
taster_name                    Virginie Boone
taster_twitter_handle                 @vboone
variety                            Pinot Noir
winery                    Castello di Amorosa
Name: 25, dtype: object, country                                 US
designation                             Hyland
province                                Oregon
region_1                         McMinnville
region_2                  Willamette Valley
taster_name                     Paul Gregutt
taster_twitter_handle          @paulgwine
variety                          Pinot Noir
winery                                Erath
Name: 35, dtype: object, country                                  US
designation                             Estate
province                           California
region_1                          Napa Valley
region_2                                 Napa
taster_name                    Virginie Boone
taster_twitter_handle                 @vboone
variety                   Cabernet Sauvignon
winery                                  Okapi
Name: 60, dtype: object, country                                   US
designation              Alder Ridge Vineyard
province                           Washington
region_1                  Columbia Valley (WA)
region_2                       Columbia Valley
```

```
taster_name                   Sean P. Sullivan
taster_twitter_handle            @wawinereport
variety                         Cabernet Franc
winery                                     Ram
Name: 62, dtype: object, country                                  US
designation                      Golden Horn
province                          California
region_1                  Santa Ynez Valley
region_2                      Central Coast
taster_name                   Matt Kettmann
taster_twitter_handle         @mattkettmann
variety                     Sauvignon Blanc
winery                                Sevtap
Name: 64, dtype: object, country                                       US
designation                                Inspired
province                                 Washington
region_1                      Columbia Valley (WA)
region_2                           Columbia Valley
taster_name                       Sean P. Sullivan
taster_twitter_handle                @wawinereport
variety                   Bordeaux-style Red Blend
winery                                Basel Cellars
Name: 67, dtype: object, country                                    US
designation                            Old Vine
province                              California
region_1                     Alexander Valley
region_2                                Sonoma
taster_name                    Virginie Boone
taster_twitter_handle                  @vboone
variety                    Cabernet Sauvignon
winery                              Eco Terreno
Name: 71, dtype: object]
```

### 8.1 设置显示的最大数目

[18]:
```
max_visual_num = 30
```

## 9 频繁模式

[19]:
```
item_sets, rules = apriori(apriori_data, min_support=0.005, min_confidence=0.3)
print({key: {key2: item_sets[key][key2]
             for key2 in list(item_sets[key].keys())[:max_visual_num //␣
 ↪len(list(item_sets.keys()))]}
      for key in list(item_sets.keys())})
```

```
{1: {('US',): 22387, ('Oregon',): 3489, ('Willamette Valley',): 2603, ('Paul
Gregutt',): 5989}, 2: {('@gordone_cellars', 'Amador County'): 142,
('@gordone_cellars', 'Cabernet Sauvignon'): 141, ('@gordone_cellars',
'California'): 1915, ('@gordone_cellars', 'California Other'): 467}, 3:
{('@gordone_cellars', 'Amador County', 'California'): 142, ('@gordone_cellars',
'Amador County', 'Jim Gordon'): 142, ('@gordone_cellars', 'Amador County',
'Sierra Foothills'): 142, ('@gordone_cellars', 'Amador County', 'US'): 142}, 4:
{('@gordone_cellars', 'Amador County', 'California', 'Jim Gordon'): 142,
('@gordone_cellars', 'Amador County', 'California', 'Sierra Foothills'): 142,
('@gordone_cellars', 'Amador County', 'California', 'US'): 142,
('@gordone_cellars', 'Amador County', 'Jim Gordon', 'Sierra Foothills'): 142},
5: {('@gordone_cellars', 'Amador County', 'California', 'Jim Gordon', 'Sierra
Foothills'): 142, ('@gordone_cellars', 'Amador County', 'California', 'Jim
Gordon', 'US'): 142, ('@gordone_cellars', 'Amador County', 'California', 'Sierra
Foothills', 'US'): 142, ('@gordone_cellars', 'Amador County', 'Jim Gordon',
'Sierra Foothills', 'US'): 142}, 6: {('@gordone_cellars', 'Amador County',
'California', 'Jim Gordon', 'Sierra Foothills', 'US'): 142, ('@gordone_cellars',
'California', 'California Other', 'Jim Gordon', 'Red Blend', 'US'): 133,
('@gordone_cellars', 'California', 'Central Coast', 'Jim Gordon', 'Livermore
Valley', 'US'): 171, ('@gordone_cellars', 'California', 'Central Valley', 'Jim
Gordon', 'Lodi', 'US'): 249}, 7: {('@mattkettmann', 'Cabernet Sauvignon',
'California', 'Central Coast', 'Matt Kettmann', 'Paso Robles', 'US'): 113,
('@mattkettmann', 'California', 'Central Coast', 'Chardonnay', 'Matt Kettmann',
'Sta. Rita Hills', 'US'): 131, ('@mattkettmann', 'California', 'Central Coast',
'Matt Kettmann', 'Paso Robles', 'Red Blend', 'US'): 143, ('@mattkettmann',
```

'California', 'Central Coast', 'Matt Kettmann', 'Pinot Noir', 'Santa Cruz Mountains', 'US'): 114}}

# 10 关联规则

[20]:
```
rules = rules[:max_visual_num]
```

## 10.1 依次显示关联规则，及其支持度、置信度、Lift 评价和卡方评价

[21]:
```
supports, confidences, lifts, convictions = [], [], [], []
for i, rule in enumerate(rules):
    rules[i] = str(rules[i])
    print(rule)
    # 支持度
    supports.append(rule.support)
    print('support:', supports[-1])
    # 置信度
    confidences.append(rule.confidence)
    print('confidence:', confidences[-1])
    # Lift 评价
    lifts.append(rule.lift)
    print('Lift:', lifts[-1])
    # 卡方评价
    convictions.append(rule.conviction)
    print('Conviction:', convictions[-1])
```

```
{Amador County} -> {@gordone_cellars} (conf: 0.667, supp: 0.006, lift: 7.794, conv: 2.743)
support: 0.006342966900433287
confidence: 0.6666666666666666
Lift: 7.793559617058311
Conviction: 2.7433778449882515
{@gordone_cellars} -> {California} (conf: 1.000, supp: 0.086, lift: 1.735, conv: 423772725.242)
support: 0.08554071559387144
confidence: 1.0
Lift: 1.7354263565891472
```

Conviction: 423772725.2423281
{California Other} -> {@gordone_cellars} (conf: 0.784, supp: 0.021, lift: 9.160, conv: 4.225)
support: 0.0208603207218475
confidence: 0.7835570469798657
Lift: 9.160047838505616
Conviction: 4.224943651069659
{Central Valley} -> {@gordone_cellars} (conf: 0.572, supp: 0.015, lift: 6.691, conv: 2.139)
support: 0.015187385536248715
confidence: 0.5723905723905723
Lift: 6.6914400752520855
Conviction: 2.1385386364840486
{El Dorado} -> {@gordone_cellars} (conf: 0.709, supp: 0.007, lift: 8.285, conv: 3.140)
support: 0.006521642024389154
confidence: 0.7087378640776699
Lift: 8.285386194833837
Conviction: 3.139643532348265
{Jim Gordon} -> {@gordone_cellars} (conf: 1.000, supp: 0.086, lift: 11.690, conv: 914459284.406)
support: 0.08554071559387144
confidence: 1.0
Lift: 11.690339425587467
Conviction: 914459284.4061285
{@gordone_cellars} -> {Jim Gordon} (conf: 1.000, supp: 0.086, lift: 11.690, conv: 914459284.406)
support: 0.08554071559387144
confidence: 1.0
Lift: 11.690339425587467
Conviction: 914459284.4061285
{Livermore Valley} -> {@gordone_cellars} (conf: 0.552, supp: 0.008, lift: 6.449, conv: 2.039)
support: 0.007638361549113325
confidence: 0.5516129032258065
Lift: 6.448542070243409
Conviction: 2.039441564990453

{Lodi} -> {@gordone_cellars} (conf: 0.542, supp: 0.011, lift: 6.342, conv:
1.999)
support: 0.011122526466252737
confidence: 0.5424836601307189
Lift: 6.341818119763136
Conviction: 1.9987467172618485
{North Coast} -> {@gordone_cellars} (conf: 0.717, supp: 0.009, lift: 8.387,
conv: 3.236)
support: 0.008844418635815428
confidence: 0.717391304347826
Lift: 8.38654784879101
Conviction: 3.235778994910467
{Sierra Foothills} -> {@gordone_cellars} (conf: 0.645, supp: 0.026, lift: 7.539,
conv: 2.575)
support: 0.02604189931656765
confidence: 0.6449115044247787
Lift: 7.539234386191918
Conviction: 2.575299659735418
{@gordone_cellars} -> {Sierra Foothills} (conf: 0.304, supp: 0.026, lift: 7.539,
conv: 1.380)
support: 0.02604189931656765
confidence: 0.3044386422976501
Lift: 7.539234386191918
Conviction: 1.379633025871729
{@gordone_cellars} -> {US} (conf: 1.000, supp: 0.086, lift: 1.000, conv: 0.000)
support: 0.08554071559387144
confidence: 1.0
Lift: 1.0
Conviction: 0.0
{California} -> {@mattkettmann} (conf: 0.347, supp: 0.200, lift: 1.735, conv:
1.225)
support: 0.20007147004958234
confidence: 0.3472093023255814
Lift: 1.7354263565891472
Conviction: 1.2253981736792245
{@mattkettmann} -> {California} (conf: 1.000, supp: 0.200, lift: 1.735, conv:
423772725.242)

support: 0.20007147004958234
confidence: 1.0
Lift: 1.7354263565891472
Conviction: 423772725.2423281
{Central Coast} -> {@mattkettmann} (conf: 0.910, supp: 0.191, lift: 4.546, conv: 8.846)
support: 0.19051235091794344
confidence: 0.909575602473875
Lift: 4.546253407587105
Conviction: 8.846379328907632
{@mattkettmann} -> {Central Coast} (conf: 0.952, supp: 0.191, lift: 4.546, conv: 16.546)
support: 0.19051235091794344
confidence: 0.952221478008484
Lift: 4.546253407587105
Conviction: 16.546097209474468
{Edna Valley} -> {@mattkettmann} (conf: 0.981, supp: 0.007, lift: 4.905, conv: 42.929)
support: 0.007057667396256756
confidence: 0.9813664596273292
Lift: 4.905079466773168
Conviction: 42.929495470122745
{Matt Kettmann} -> {@mattkettmann} (conf: 1.000, supp: 0.200, lift: 4.998, conv: 799928529.950)
support: 0.20007147004958234
confidence: 1.0
Lift: 4.998213887028355
Conviction: 799928529.9504176
{@mattkettmann} -> {Matt Kettmann} (conf: 1.000, supp: 0.200, lift: 4.998, conv: 799928529.950)
support: 0.20007147004958234
confidence: 1.0
Lift: 4.998213887028355
Conviction: 799928529.9504176
{Paso Robles} -> {@mattkettmann} (conf: 0.991, supp: 0.041, lift: 4.955, conv: 93.492)
support: 0.041407959976772234

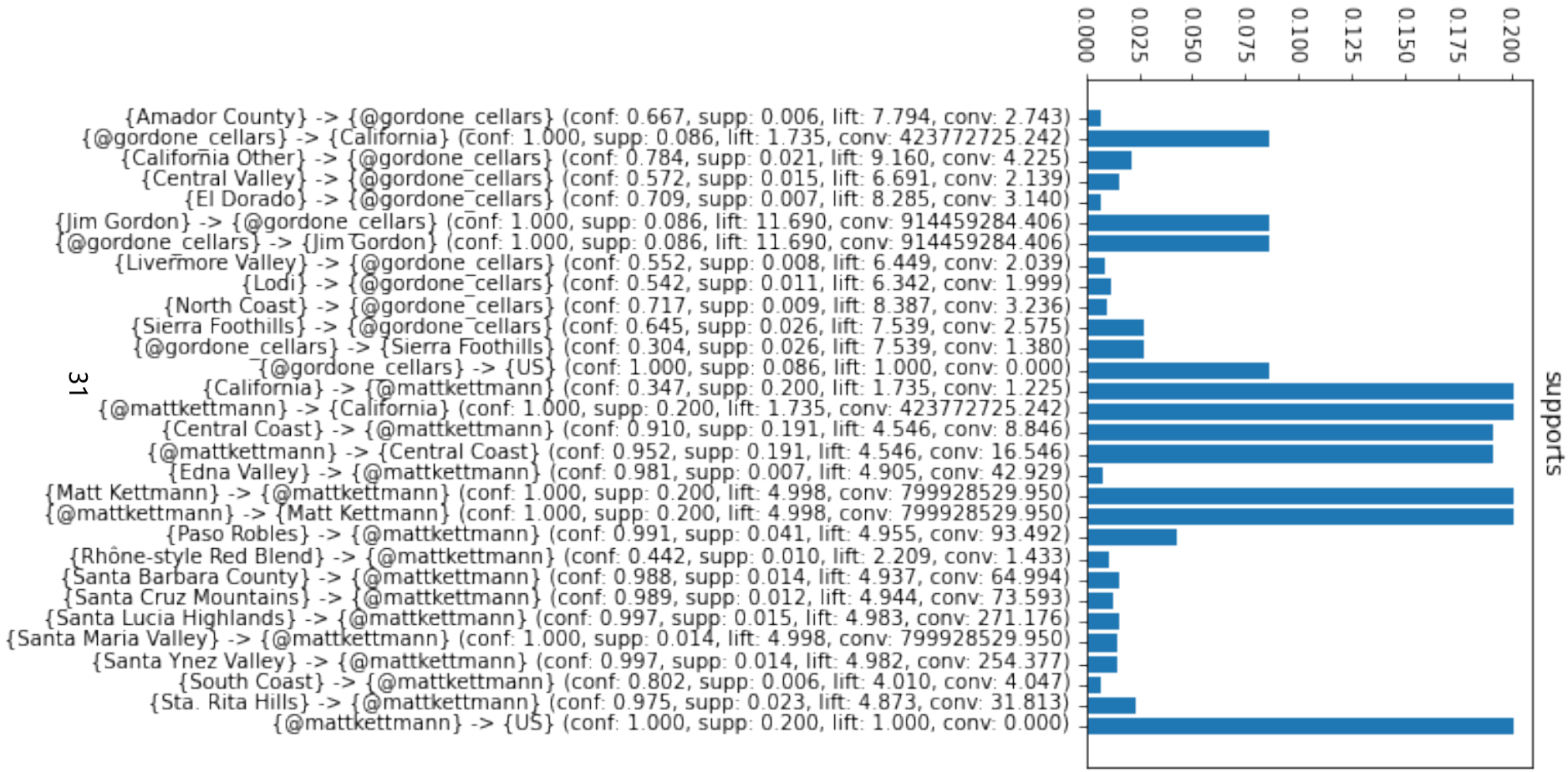
confidence: 0.9914438502673797
Lift: 4.955448420615278
Conviction: 93.49163601112022
{Rhône-style Red Blend} -> {@mattkettmann} (conf: 0.442, supp: 0.010, lift:
2.209, conv: 1.433)
support: 0.010184482065484434
confidence: 0.4418604651162791
Lift: 2.208513112872994
Conviction: 1.4332052802600055
{Santa Barbara County} -> {@mattkettmann} (conf: 0.988, supp: 0.014, lift:
4.937, conv: 64.994)
support: 0.014338678697458347
confidence: 0.9876923076923076
Lift: 4.936697408418775
Conviction: 64.99418777769344
{Santa Cruz Mountains} -> {@mattkettmann} (conf: 0.989, supp: 0.012, lift:
4.944, conv: 73.593)
support: 0.01219457720998794
confidence: 0.9891304347826086
Lift: 4.94388547521283
Conviction: 73.59341798484364
{Santa Lucia Highlands} -> {@mattkettmann} (conf: 0.997, supp: 0.015, lift:
4.983, conv: 271.176)
support: 0.015098047974270782
confidence: 0.9970501474926253
Lift: 4.983469893261309
Conviction: 271.17567972463235
{Santa Maria Valley} -> {@mattkettmann} (conf: 1.000, supp: 0.014, lift: 4.998,
conv: 799928529.950)
support: 0.014160003573502479
confidence: 1.0
Lift: 4.998213887028355
Conviction: 799928529.9504176
{Santa Ynez Valley} -> {@mattkettmann} (conf: 0.997, supp: 0.014, lift: 4.982,
conv: 254.377)
support: 0.014160003573502479
confidence: 0.9968553459119497

```
Lift: 4.982496233295561
Conviction: 254.37719163228593
{South Coast} -> {@mattkettmann} (conf: 0.802, supp: 0.006, lift: 4.010, conv:
4.047)
support: 0.00616429177647742
confidence: 0.8023255813953488
Lift: 4.010194862848331
Conviction: 4.046697248689409
{Sta. Rita Hills} -> {@mattkettmann} (conf: 0.975, supp: 0.023, lift: 4.873,
conv: 31.813)
support: 0.022513065618439274
confidence: 0.9748549323017408
Lift: 4.872533460468648
Conviction: 31.812541041329393
{@mattkettmann} -> {US} (conf: 1.000, supp: 0.200, lift: 1.000, conv: 0.000)
support: 0.20007147004958234
confidence: 1.0
Lift: 1.0
Conviction: 0.0
```

# 11 可视化

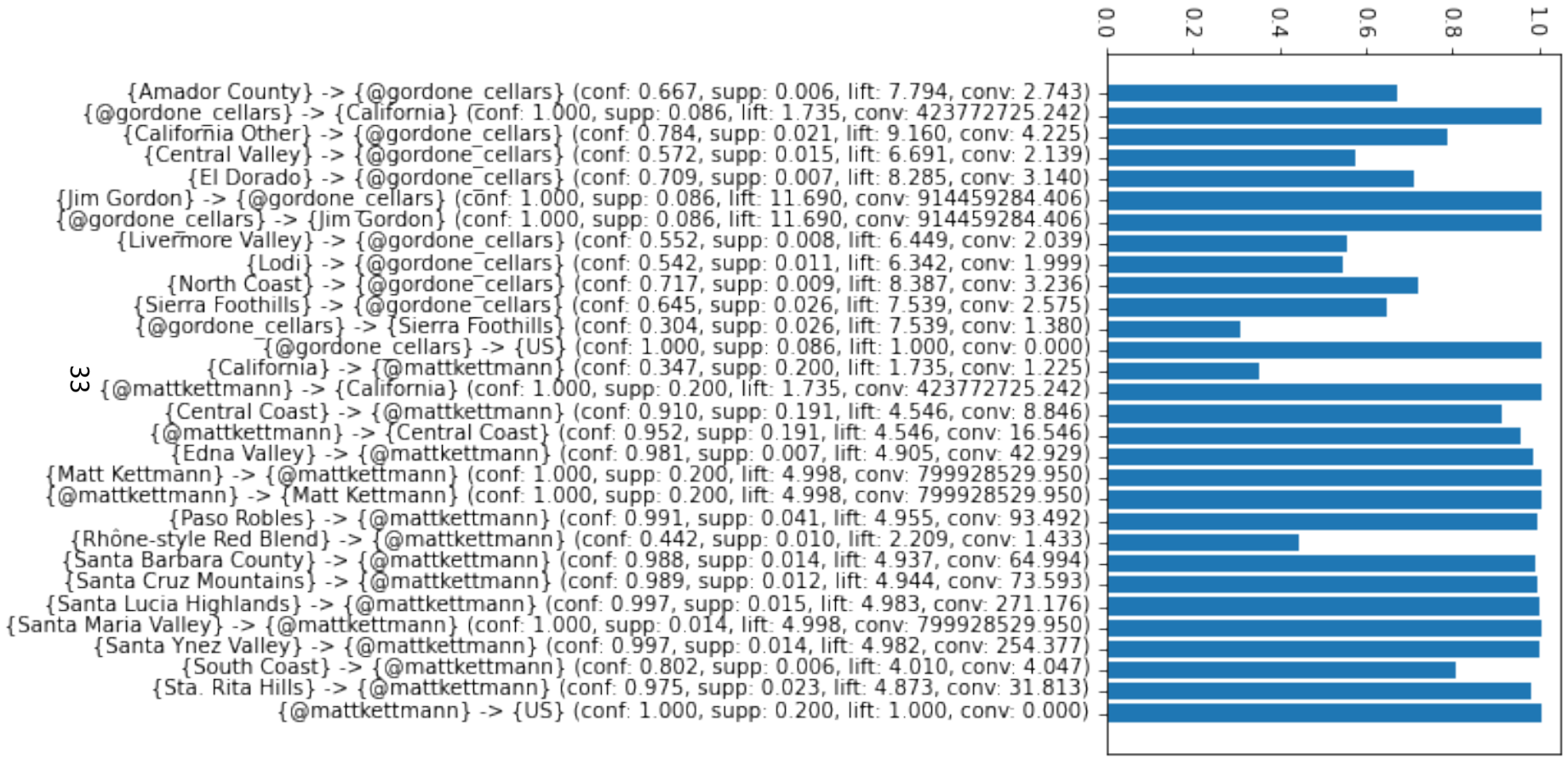
## 11.1 支持度

[22]:
```
plt.title('supports')
plt.xticks(range(len(rules)), rules, rotation=90)
plt.bar(range(len(supports)), supports)
plt.show()
```


{Amador County} -> {@gordone_cellars} (conf: 0.667, supp: 0.006, lift: 7.794, conv: 2.743)
{@gordone_cellars} -> {California} (conf: 1.000, supp: 0.086, lift: 1.735, conv: 423772725.242)
{California Other} -> {@gordone_cellars} (conf: 0.784, supp: 0.021, lift: 9.160, conv: 4.225)
{Central Valley} -> {@gordone_cellars} (conf: 0.572, supp: 0.015, lift: 6.691, conv: 2.139)
{El Dorado} -> {@gordone_cellars} (conf: 0.709, supp: 0.007, lift: 8.285, conv: 3.140)
{Jim Gordon} -> {@gordone_cellars} (conf: 1.000, supp: 0.086, lift: 11.690, conv: 914459284.406)
{@gordone_cellars} -> {Jim Gordon} (conf: 1.000, supp: 0.086, lift: 11.690, conv: 914459284.406)
{Livermore Valley} -> {@gordone_cellars} (conf: 0.552, supp: 0.008, lift: 6.449, conv: 2.039)
{Lodi} -> {@gordone_cellars} (conf: 0.542, supp: 0.011, lift: 6.342, conv: 1.999)
{North Coast} -> {@gordone_cellars} (conf: 0.717, supp: 0.009, lift: 8.387, conv: 3.236)
{Sierra Foothills} -> {@gordone_cellars} (conf: 0.645, supp: 0.026, lift: 7.539, conv: 2.575)
{@gordone_cellars} -> {Sierra Foothills} (conf: 0.304, supp: 0.026, lift: 7.539, conv: 1.380)
{@gordone_cellars} -> {US} (conf: 1.000, supp: 0.086, lift: 1.000, conv: 0.000)
{California} -> {@mattkettmann} (conf: 0.347, supp: 0.200, lift: 1.735, conv: 1.225)
{@mattkettmann} -> {California} (conf: 1.000, supp: 0.200, lift: 1.735, conv: 423772725.242)
{Central Coast} -> {@mattkettmann} (conf: 0.910, supp: 0.191, lift: 4.546, conv: 8.846)
{@mattkettmann} -> {Central Coast} (conf: 0.952, supp: 0.191, lift: 4.546, conv: 16.546)
{Edna Valley} -> {@mattkettmann} (conf: 0.981, supp: 0.007, lift: 4.905, conv: 42.929)
{Matt Kettmann} -> {@mattkettmann} (conf: 1.000, supp: 0.200, lift: 4.998, conv: 799928529.950)
{@mattkettmann} -> {Matt Kettmann} (conf: 1.000, supp: 0.200, lift: 4.998, conv: 799928529.950)
{Paso Robles} -> {@mattkettmann} (conf: 0.991, supp: 0.041, lift: 4.955, conv: 93.492)
{Rhône-style Red Blend} -> {@mattkettmann} (conf: 0.442, supp: 0.010, lift: 2.209, conv: 1.433)
{Santa Barbara County} -> {@mattkettmann} (conf: 0.988, supp: 0.014, lift: 4.937, conv: 64.994)
{Santa Cruz Mountains} -> {@mattkettmann} (conf: 0.989, supp: 0.012, lift: 4.944, conv: 73.593)
{Santa Lucia Highlands} -> {@mattkettmann} (conf: 0.997, supp: 0.015, lift: 4.983, conv: 271.176)
{Santa Maria Valley} -> {@mattkettmann} (conf: 1.000, supp: 0.014, lift: 4.998, conv: 799928529.950)
{Santa Ynez Valley} -> {@mattkettmann} (conf: 0.997, supp: 0.014, lift: 4.982, conv: 254.377)
{South Coast} -> {@mattkettmann} (conf: 0.802, supp: 0.006, lift: 4.010, conv: 4.047)
{Sta. Rita Hills} -> {@mattkettmann} (conf: 0.975, supp: 0.023, lift: 4.873, conv: 31.813)
{@mattkettmann} -> {US} (conf: 1.000, supp: 0.200, lift: 1.000, conv: 0.000)
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
supports

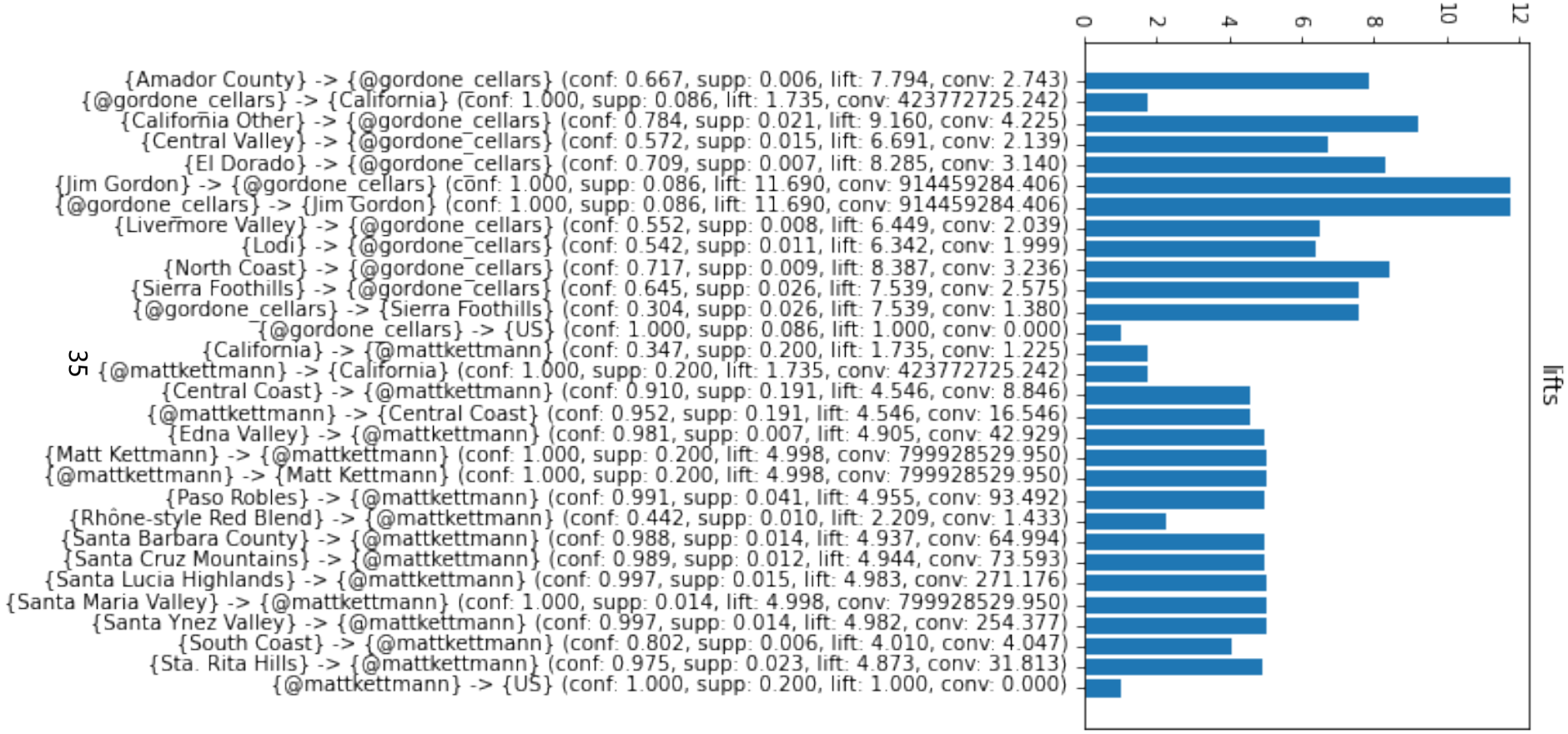
### 11.2 置信度

[23]:
```
plt.title('confidences')
plt.xticks(range(len(rules)), rules, rotation=90)
plt.bar(range(len(confidences)), confidences)
plt.show()
```


{Amador County} -> {@gordone_cellars} (conf: 0.667, supp: 0.006, lift: 7.794, conv: 2.743)
{@gordone_cellars} -> {California} (conf: 1.000, supp: 0.086, lift: 1.735, conv: 423772725.242)
{California Other} -> {@gordone_cellars} (conf: 0.784, supp: 0.021, lift: 9.160, conv: 4.225)
{Central Valley} -> {@gordone_cellars} (conf: 0.572, supp: 0.015, lift: 6.691, conv: 2.139)
{El Dorado} -> {@gordone_cellars} (conf: 0.709, supp: 0.007, lift: 8.285, conv: 3.140)
{Jim Gordon} -> {@gordone_cellars} (conf: 1.000, supp: 0.086, lift: 11.690, conv: 914459284.406)
{@gordone_cellars} -> {Jim Gordon} (conf: 1.000, supp: 0.086, lift: 11.690, conv: 914459284.406)
{Livermore Valley} -> {@gordone_cellars} (conf: 0.552, supp: 0.008, lift: 6.449, conv: 2.039)
{Lodi} -> {@gordone_cellars} (conf: 0.542, supp: 0.011, lift: 6.342, conv: 1.999)
{North Coast} -> {@gordone_cellars} (conf: 0.717, supp: 0.009, lift: 8.387, conv: 3.236)
{Sierra Foothills} -> {@gordone_cellars} (conf: 0.645, supp: 0.026, lift: 7.539, conv: 2.575)
{@gordone_cellars} -> {Sierra Foothills} (conf: 0.304, supp: 0.026, lift: 7.539, conv: 1.380)
{@gordone_cellars} -> {US} (conf: 1.000, supp: 0.086, lift: 1.000, conv: 0.000)
{California} -> {@mattkettmann} (conf: 0.347, supp: 0.200, lift: 1.735, conv: 1.225)
{@mattkettmann} -> {California} (conf: 1.000, supp: 0.200, lift: 1.735, conv: 423772725.242)
{Central Coast} -> {@mattkettmann} (conf: 0.910, supp: 0.191, lift: 4.546, conv: 8.846)
{@mattkettmann} -> {Central Coast} (conf: 0.952, supp: 0.191, lift: 4.546, conv: 16.546)
{Edna Valley} -> {@mattkettmann} (conf: 0.981, supp: 0.007, lift: 4.905, conv: 42.929)
{Matt Kettmann} -> {@mattkettmann} (conf: 1.000, supp: 0.200, lift: 4.998, conv: 799928529.950)
{@mattkettmann} -> {Matt Kettmann} (conf: 1.000, supp: 0.200, lift: 4.998, conv: 799928529.950)
{Paso Robles} -> {@mattkettmann} (conf: 0.991, supp: 0.041, lift: 4.955, conv: 93.492)
{Rhône-style Red Blend} -> {@mattkettmann} (conf: 0.442, supp: 0.010, lift: 2.209, conv: 1.433)
{Santa Barbara County} -> {@mattkettmann} (conf: 0.988, supp: 0.014, lift: 4.937, conv: 64.994)
{Santa Cruz Mountains} -> {@mattkettmann} (conf: 0.989, supp: 0.012, lift: 4.944, conv: 73.593)
{Santa Lucia Highlands} -> {@mattkettmann} (conf: 0.997, supp: 0.015, lift: 4.983, conv: 271.176)
{Santa Maria Valley} -> {@mattkettmann} (conf: 1.000, supp: 0.014, lift: 4.998, conv: 799928529.950)
{Santa Ynez Valley} -> {@mattkettmann} (conf: 0.997, supp: 0.014, lift: 4.982, conv: 254.377)
{South Coast} -> {@mattkettmann} (conf: 0.802, supp: 0.006, lift: 4.010, conv: 4.047)
{Sta. Rita Hills} -> {@mattkettmann} (conf: 0.975, supp: 0.023, lift: 4.873, conv: 31.813)
{@mattkettmann} -> {US} (conf: 1.000, supp: 0.200, lift: 1.000, conv: 0.000)
0.0
0.2
0.4
0.6
0.8
1.0
confidences

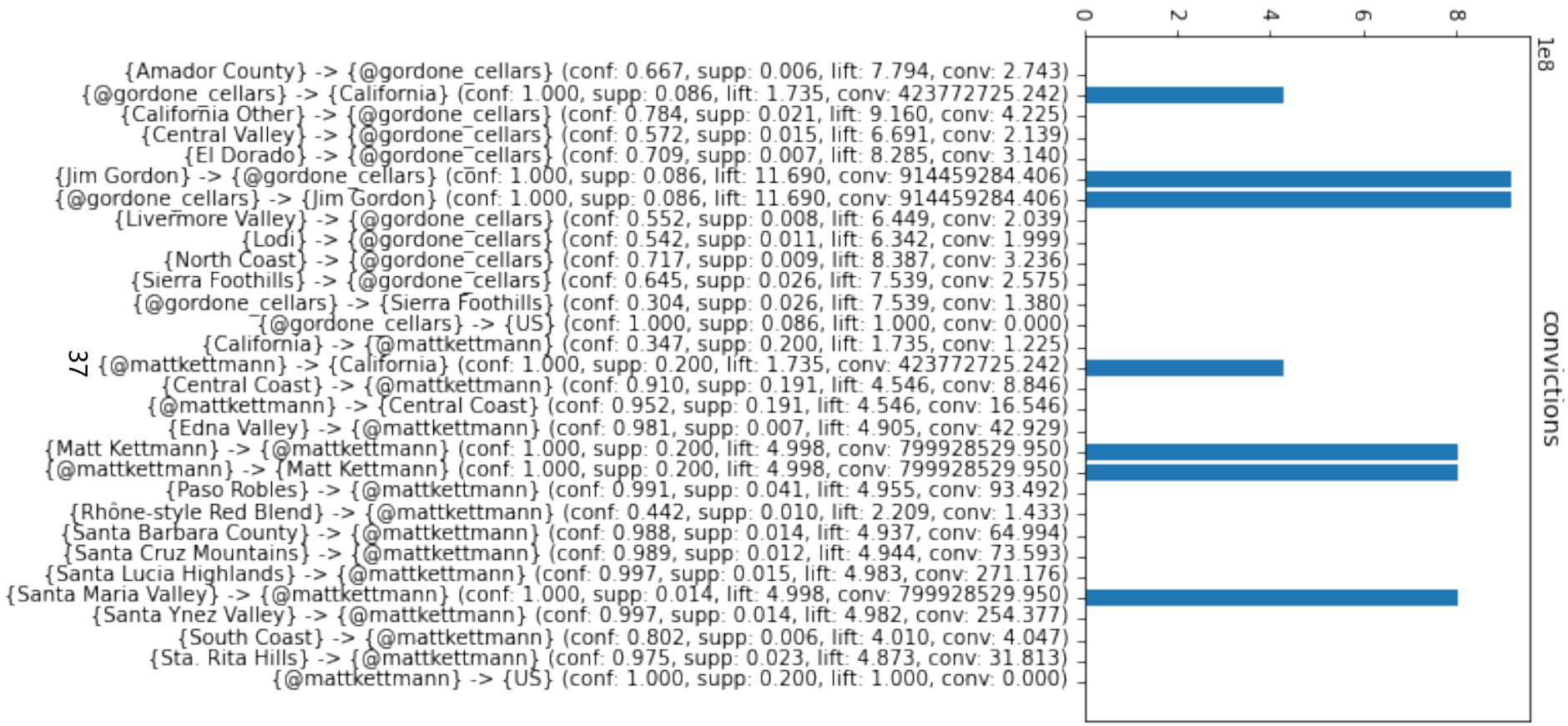
### 11.3 Lifts 评价

[24]:
```
plt.title('lifts')
plt.xticks(range(len(rules)), rules, rotation=90)
plt.bar(range(len(lifts)), lifts)
plt.show()
```


{Amador County} -> {@gordone_cellars} (conf: 0.667, supp: 0.006, lift: 7.794, conv: 2.743)
{@gordone_cellars} -> {California} (conf: 1.000, supp: 0.086, lift: 1.735, conv: 423772725.242)
{California Other} -> {@gordone_cellars} (conf: 0.784, supp: 0.021, lift: 9.160, conv: 4.225)
{Central Valley} -> {@gordone_cellars} (conf: 0.572, supp: 0.015, lift: 6.691, conv: 2.139)
{El Dorado} -> {@gordone_cellars} (conf: 0.709, supp: 0.007, lift: 8.285, conv: 3.140)
{Jim Gordon} -> {@gordone_cellars} (conf: 1.000, supp: 0.086, lift: 11.690, conv: 914459284.406)
{@gordone_cellars} -> {Jim Gordon} (conf: 1.000, supp: 0.086, lift: 11.690, conv: 914459284.406)
{Livermore Valley} -> {@gordone_cellars} (conf: 0.552, supp: 0.008, lift: 6.449, conv: 2.039)
{Lodi} -> {@gordone_cellars} (conf: 0.542, supp: 0.011, lift: 6.342, conv: 1.999)
{North Coast} -> {@gordone_cellars} (conf: 0.717, supp: 0.009, lift: 8.387, conv: 3.236)
{Sierra Foothills} -> {@gordone_cellars} (conf: 0.645, supp: 0.026, lift: 7.539, conv: 2.575)
{@gordone_cellars} -> {Sierra Foothills} (conf: 0.304, supp: 0.026, lift: 7.539, conv: 1.380)
{@gordone_cellars} -> {US} (conf: 1.000, supp: 0.086, lift: 1.000, conv: 0.000)
{California} -> {@mattkettmann} (conf: 0.347, supp: 0.200, lift: 1.735, conv: 1.225)
{@mattkettmann} -> {California} (conf: 1.000, supp: 0.200, lift: 1.735, conv: 423772725.242)
{Central Coast} -> {@mattkettmann} (conf: 0.910, supp: 0.191, lift: 4.546, conv: 8.846)
{@mattkettmann} -> {Central Coast} (conf: 0.952, supp: 0.191, lift: 4.546, conv: 16.546)
{Edna Valley} -> {@mattkettmann} (conf: 0.981, supp: 0.007, lift: 4.905, conv: 42.929)
{Matt Kettmann} -> {@mattkettmann} (conf: 1.000, supp: 0.200, lift: 4.998, conv: 799928529.950)
{@mattkettmann} -> {Matt Kettmann} (conf: 1.000, supp: 0.200, lift: 4.998, conv: 799928529.950)
{Paso Robles} -> {@mattkettmann} (conf: 0.991, supp: 0.041, lift: 4.955, conv: 93.492)
{Rhône-style Red Blend} -> {@mattkettmann} (conf: 0.442, supp: 0.010, lift: 2.209, conv: 1.433)
{Santa Barbara County} -> {@mattkettmann} (conf: 0.988, supp: 0.014, lift: 4.937, conv: 64.994)
{Santa Cruz Mountains} -> {@mattkettmann} (conf: 0.989, supp: 0.012, lift: 4.944, conv: 73.593)
{Santa Lucia Highlands} -> {@mattkettmann} (conf: 0.997, supp: 0.015, lift: 4.983, conv: 271.176)
{Santa Maria Valley} -> {@mattkettmann} (conf: 1.000, supp: 0.014, lift: 4.998, conv: 799928529.950)
{Santa Ynez Valley} -> {@mattkettmann} (conf: 0.997, supp: 0.014, lift: 4.982, conv: 254.377)
{South Coast} -> {@mattkettmann} (conf: 0.802, supp: 0.006, lift: 4.010, conv: 4.047)
{Sta. Rita Hills} -> {@mattkettmann} (conf: 0.975, supp: 0.023, lift: 4.873, conv: 31.813)
{@mattkettmann} -> {US} (conf: 1.000, supp: 0.200, lift: 1.000, conv: 0.000)
0
2
4
6
8
10
12
lifts

### 11.4 卡方评价

[25]:
```
plt.title('convictions')
plt.xticks(range(len(rules)), rules, rotation=90)
plt.bar(range(len(convictions)), convictions)
plt.show()
```


{Amador County} -> {@gordone_cellars} (conf: 0.667, supp: 0.006, lift: 7.794, conv: 2.743)
{@gordone_cellars} -> {California} (conf: 1.000, supp: 0.086, lift: 1.735, conv: 423772725.242)
{California Other} -> {@gordone_cellars} (conf: 0.784, supp: 0.021, lift: 9.160, conv: 4.225)
{Central Valley} -> {@gordone_cellars} (conf: 0.572, supp: 0.015, lift: 6.691, conv: 2.139)
{El Dorado} -> {@gordone_cellars} (conf: 0.709, supp: 0.007, lift: 8.285, conv: 3.140)
{Jim Gordon} -> {@gordone_cellars} (conf: 1.000, supp: 0.086, lift: 11.690, conv: 914459284.406)
{@gordone_cellars} -> {Jim Gordon} (conf: 1.000, supp: 0.086, lift: 11.690, conv: 914459284.406)
{Livermore Valley} -> {@gordone_cellars} (conf: 0.552, supp: 0.008, lift: 6.449, conv: 2.039)
{Lodi} -> {@gordone_cellars} (conf: 0.542, supp: 0.011, lift: 6.342, conv: 1.999)
{North Coast} -> {@gordone_cellars} (conf: 0.717, supp: 0.009, lift: 8.387, conv: 3.236)
{Sierra Foothills} -> {@gordone_cellars} (conf: 0.645, supp: 0.026, lift: 7.539, conv: 2.575)
{@gordone_cellars} -> {Sierra Foothills} (conf: 0.304, supp: 0.026, lift: 7.539, conv: 1.380)
{@gordone_cellars} -> {US} (conf: 1.000, supp: 0.086, lift: 1.000, conv: 0.000)
{California} -> {@mattkettmann} (conf: 0.347, supp: 0.200, lift: 1.735, conv: 1.225)
{@mattkettmann} -> {California} (conf: 1.000, supp: 0.200, lift: 1.735, conv: 423772725.242)
{Central Coast} -> {@mattkettmann} (conf: 0.910, supp: 0.191, lift: 4.546, conv: 8.846)
{@mattkettmann} -> {Central Coast} (conf: 0.952, supp: 0.191, lift: 4.546, conv: 16.546)
{Edna Valley} -> {@mattkettmann} (conf: 0.981, supp: 0.007, lift: 4.905, conv: 42.929)
{Matt Kettmann} -> {@mattkettmann} (conf: 1.000, supp: 0.200, lift: 4.998, conv: 799928529.950)
{@mattkettmann} -> {Matt Kettmann} (conf: 1.000, supp: 0.200, lift: 4.998, conv: 799928529.950)
{Paso Robles} -> {@mattkettmann} (conf: 0.991, supp: 0.041, lift: 4.955, conv: 93.492)
{Rhône-style Red Blend} -> {@mattkettmann} (conf: 0.442, supp: 0.010, lift: 2.209, conv: 1.433)
{Santa Barbara County} -> {@mattkettmann} (conf: 0.988, supp: 0.014, lift: 4.937, conv: 64.994)
{Santa Cruz Mountains} -> {@mattkettmann} (conf: 0.989, supp: 0.012, lift: 4.944, conv: 73.593)
{Santa Lucia Highlands} -> {@mattkettmann} (conf: 0.997, supp: 0.015, lift: 4.983, conv: 271.176)
{Santa Maria Valley} -> {@mattkettmann} (conf: 1.000, supp: 0.014, lift: 4.998, conv: 799928529.950)
{Santa Ynez Valley} -> {@mattkettmann} (conf: 0.997, supp: 0.014, lift: 4.982, conv: 254.377)
{South Coast} -> {@mattkettmann} (conf: 0.802, supp: 0.006, lift: 4.010, conv: 4.047)
{Sta. Rita Hills} -> {@mattkettmann} (conf: 0.975, supp: 0.023, lift: 4.873, conv: 31.813)
{@mattkettmann} -> {US} (conf: 1.000, supp: 0.200, lift: 1.000, conv: 0.000)
0
2
4
6
8
1e8
convictions