```
In [88]:   import pandas as pd
           df = pd.read_csv("intermediate_data/wineries.csv")
           df.head()
```

Out[88]:

	country	province	region_1	region_2	winery	winery_id
0	Italy	Sicily & Sardinia	Etna	NaN	Nicosia	0
1	Portugal	Douro	NaN	NaN	Quinta dos Avidagos	1
2	US	Oregon	Willamette Valley	Willamette Valley	Rainstorm	2
3	US	Michigan	Lake Michigan Shore	NaN	St. Julian	3
4	US	Oregon	Willamette Valley	Willamette Valley	Sweet Cheeks	4

Fixing "country" and "province" Values

Add in missing "country" and "province" values

First, identify the wineries with missing "country" and "province" information and verify that they're the same

```
In [89]:   idx_missing_location = df[df['country'].isnull()].index
           missing_countries = df["winery"].loc[idx_missing_location]

           idx_missing_location = df[df['province'].isnull()].index
           missing_provinces = df["winery"].loc[idx_missing_location]

           list(missing_countries)==list(missing_provinces)
```

Out[89]:   True

Then, look up the winery names and find the true values.

```
In [90]:   df["winery"].loc[idx_missing_location]
```

Out[90]:
```
838                 Gotsa Family Wines
2636                  Barton & Guestier
3460       Kakhetia Traditional Winemaking
6715                           Tsililis
6837                           Ross-idi
7554                            Orbelus
8931                          St. Donat
9714                     Familia Deicas
10008                     Bartho Eksteen
13507                       Stone Castle
13508                      Teliani Valley
13511                          Undurraga
13912                      Mt. Beautiful
16794                         Neumeister
16956                          Bachelder
18523                            Chilcas
18714                          Santa Ema
19347                             Psagot
20966                    Newton Johnson
21470                     Ktima Voyatzi
25333                            Lismore
25334                            Logodaj
25958                      Somlói Vándor
27408                   Les Frères Dutruy
28854                         El Capricho
29464                   Amiran Vepkhvadze
30376                          Büyülübağ
Name: winery, dtype: object
```

First half: | Winery | Index | Country | Province | | -------- | ------- | ------- | ------- | | Gotsa Family Wines | 838 | Georgia | Kvemo Kartli | | Barton & Guestier | 2636 | France | Gironde | | Kakhetia Traditional Winemaking | 3460 | Georgia | Kakheti | | Tsililis |6715 | Greece | Thessaly | | Ross-idi | 6837 | Bulgaria | Nikolaevo | | Orbelus | 7554 | Bulgaria | Blagoevgrad | | St. Donat | 8931 | Hungary | Veszprem | | Familia Deicas | 9714 | Uruguay | Canelones | | Bartho Eksteen | 10008 | South Africa | Western Cape | | Stone Castle | 13507 | Kosovo | Jakove | | Teliani Valley | 13508 | Georgia | Kakheti | | Undurraga | 13511 | Chile | Santiago | | Mt. Beautiful | 13912 | New Zealand | Canterbury |

```python
In [91]:  df.at[838,"country"] = "Georgia"
          df.at[838,"province"] = "Kvemo Kartli"
          df.at[2636,"country"] = "France"
          df.at[2636,"province"] = "Gironde"
          df.at[3460,"country"] = "Georgia"
          df.at[3460,"province"] = "Kakheti"
          df.at[6715,"country"] = "Greece"
          df.at[6715,"province"] = "Thessaly"
          df.at[6837,"country"] = "Bulgaria"
          df.at[6837,"province"] = "Nikolaevo"
          df.at[7554,"country"] = "Bulgaria"
          df.at[7554,"province"] = "Blagoevgrad"
          df.at[8931,"country"] = "Hungary"
          df.at[8931,"province"] = "Veszprem"
          df.at[9714,"country"] = "Uruguay"
          df.at[9714,"province"] = "Canelones"
          df.at[10008,"country"] = "South Africa"
          df.at[10008,"province"] = "Western Cape"
          df.at[13507     ,"country"] = "Kosovo"
          df.at[13507     ,"province"] = "Jakove"
          df.at[13508     ,"country"] = "Georgia"
          df.at[13508     ,"province"] = "Kakheti"
          df.at[13511     ,"country"] = "Chile"
          df.at[13511     ,"province"] = "Santiago"
          df.at[13912,"country"] = "New Zealand"
          df.at[13912,"province"] = "Canterbury"
```

Second half: | Winery | Index | Country | Province | | | -------- | ------- | ------- | ------- | | | Neumeister | 16794 | Austria | Styria | | Bachelder | 16956 | Canada | Ontario | | Chilcas | 18523 | Chile | Maule | | Santa Ema | 18714 | Chile | Santiago | | Psagot | 19347 | Israel | Judea and Samaria | | Newton Johnson | 20966 | South Africa | Western Cape | | Ktima Voyatzi | 21470 | Greece | Western Macedonia | | Lismore | 25333 | South Africa | Western Cape | | Logodaj | 25334 | Bulgaria | Blagoevgrad | | Somlói Vándor | 25958 | Hungary | Veszprem | | Les Frères Dutruy | 27408 | Switzerland | Vaud | | El Capricho | 28854 | Uruguay | Durazno | | Amiran Vepkhvadze | 29464 | Georgia | Imereti | | Büyülübağ | 30376 | Turkey |Balıkesir |

```python
In [92]:  df.at[16794,"country"] = "Austria"
          df.at[16794,"province"] = "Styria"
          df.at[16956,"country"] = "Canada"
          df.at[16956,"province"] = "Ontario"
          df.at[18523,"country"] = "Chile"
          df.at[18523,"province"] = "Maule"
          df.at[18714,"country"] = "Chile"
          df.at[18714,"province"] = "Santiago"
          df.at[19347,"country"] = "Israel"
          df.at[19347,"province"] = "Judea and Samaria"
          df.at[20966,"country"] = "South Africa"
          df.at[20966,"province"] = "Western Cape"
          df.at[21470,"country"] = "Greece"
          df.at[21470,"province"] = "Western Macedonia"
          df.at[25333,"country"] = "South Africa"
          df.at[25333,"province"] = "Western Cape"
          df.at[25334,"country"] = "Bulgaria"
          df.at[25334,"province"] = "Blagoevgrad"
          df.at[25958,"country"] = "Hungary"
          df.at[25958,"province"] = "Veszprem"
          df.at[27408,"country"] = "Switzerland"
          df.at[27408,"province"] = "Vaud"
          df.at[28854,"country"] = "Uruguay"
          df.at[28854,"province"] = "Durazno"
          df.at[29464,"country"] = "Georgia"
          df.at[29464,"province"] = "Imereti"
          df.at[30376,"country"] = "Turkey"
          df.at[30376,"province"] = "Balıkesir"
```

```python
In [93]:  # Check that there aren't any more null values

          idx_missing_location = df[df['country'].isnull()].index
          missing_countries = df["winery"].loc[idx_missing_location]

          idx_missing_location = df[df['province'].isnull()].index
          missing_provinces = df["winery"].loc[idx_missing_location]

          print(missing_countries)
          print(missing_provinces)

          Series([], Name: winery, dtype: object)
          Series([], Name: winery, dtype: object)
```

Check for "country" and "province" duplicates

```
In [94]:   sorted(list(df['country'].unique()))
```

```
Out[94]:   ['Argentina',
            'Armenia',
            'Australia',
            'Austria',
            'Bosnia and Herzegovina',
            'Brazil',
            'Bulgaria',
            'Canada',
            'Chile',
            'China',
            'Croatia',
            'Cyprus',
            'Czech Republic',
            'Egypt',
            'England',
            'France',
            'Georgia',
            'Germany',
            'Greece',
            'Hungary',
            'India',
            'Israel',
            'Italy',
            'Kosovo',
            'Lebanon',
            'Luxembourg',
            'Macedonia',
            'Mexico',
            'Moldova',
            'Morocco',
            'New Zealand',
            'Peru',
            'Portugal',
            'Romania',
            'Serbia',
            'Slovakia',
            'Slovenia',
            'South Africa',
            'Spain',
            'Switzerland',
            'Turkey',
            'US',
            'Ukraine',
            'Uruguay']
```

```
In [111…   locations = {}
           for i in range(len(df)):
               location = (df['country'].loc[i],df['province'].loc[i])
               if location not in locations:
                   locations[location] = [i]
               else:
                   locations[location].append(i)

           locations_list=sorted(locations)

           print(len(locations_list))

           curr_country=None
           for k in range(len(locations_list)):
               if locations_list[k][0]!=curr_country:
                   print()
                   curr_country = locations_list[k][0]
               print(locations_list[k])
```

```
('Argentina', 'Mendoza Province')
('Argentina', 'Unknown')

('Armenia', 'Unknown')

('Australia', 'New South Wales')
('Australia', 'South Australia')
('Australia', 'Tasmania')
('Australia', 'Unknown')
('Australia', 'Victoria')
('Australia', 'Western Australia')

('Austria', 'Burgenland')
('Austria', 'Carnuntum')
('Austria', 'Eisenberg')
('Austria', 'Kamptal')
('Austria', 'Kremstal')
('Austria', 'Leithaberg')
('Austria', 'Mittelburgenland')
('Austria', 'Neusiedlersee')
('Austria', 'Niederösterreich')
('Austria', 'Steiermark')
('Austria', 'Styria')
('Austria', 'Südburgenland')
('Austria', 'Südoststeiermark')
('Austria', 'Südsteiermark')
('Austria', 'Thermenregion')
('Austria', 'Traisental')
('Austria', 'Unknown')
('Austria', 'Vienna')
('Austria', 'Wachau')
('Austria', 'Wagram')
('Austria', 'Weinland Österreich')
('Austria', 'Weinviertel')
('Austria', 'Wiener Gemischter Satz')
('Austria', 'Österreichischer Perlwein')
('Austria', 'Österreichischer Sekt')

('Bosnia and Herzegovina', 'Mostar')

('Brazil', 'Campanha')
('Brazil', 'Pinto Bandeira')
('Brazil', 'Santa Catarina')
('Brazil', 'Serra Gaúcha')
('Brazil', 'Serra do Sudeste')
('Brazil', 'Unknown')
('Brazil', 'Vale Trentino')
('Brazil', 'Vale dos Vinhedos')

('Bulgaria', 'Black Sea Coastal')
('Bulgaria', 'Blagoevgrad')
('Bulgaria', 'Danube River Plains')
('Bulgaria', 'Nikolaevo')
('Bulgaria', 'Thracian Valley')
('Bulgaria', 'Unknown')

('Canada', 'British Columbia')
('Canada', 'Ontario')
('Canada', 'Unknown')

('Chile', 'Aconcagua Costa')
('Chile', 'Aconcagua Valley')
('Chile', 'Apalta')
('Chile', 'Buin')
('Chile', 'Bío Bío Valley')
('Chile', 'Cachapoal Valley')
('Chile', 'Casablanca Valley')
('Chile', 'Cauquenes Valley')
('Chile', 'Central Valley')
('Chile', 'Choapa Valley')
('Chile', 'Coelemu')
('Chile', 'Colchagua Costa')
('Chile', 'Colchagua Valley')
('Chile', 'Curicó Valley')
('Chile', 'Elqui Valley')
('Chile', 'Itata Valley')
('Chile', 'Leyda Valley')
('Chile', 'Limarí Valley')
```

```
('Chile', 'Lolol Valley')
('Chile', 'Loncomilla Valley')
('Chile', 'Lontué Valley')
('Chile', 'Maipo Valley')
('Chile', 'Malleco')
('Chile', 'Marchigue')
('Chile', 'Maule Valley')
('Chile', 'Molina')
('Chile', 'Peumo')
('Chile', 'Pirque')
('Chile', 'Puente Alto')
('Chile', 'Rapel Valley')
('Chile', 'Requinoa')
('Chile', 'Rio Claro')
('Chile', 'Sagrada Familia')
('Chile', 'San Antonio')
('Chile', 'San Clemente')
('Chile', 'Santa Cruz')
('Chile', 'Santiago')
('Chile', 'Unknown')

('China', 'Unknown')

('Croatia', 'Dalmatian Coast')
('Croatia', 'Dingač')
('Croatia', 'Hrvatsko Primorje')
('Croatia', 'Hvar')
('Croatia', 'Istria')
('Croatia', 'Korčula')
('Croatia', 'Krk')
('Croatia', 'Kutjevo')
('Croatia', 'Middle and South Dalmatia')
('Croatia', 'North Dalmatia')
('Croatia', 'Peljesac')
('Croatia', 'Podunavlje')
('Croatia', 'Unknown')

('Cyprus', 'Commandaria')
('Cyprus', 'Kathikas')
('Cyprus', 'Lemesos')
('Cyprus', 'Pafos')
('Cyprus', 'Pitsilia Mountains')
('Cyprus', 'Unknown')

('Czech Republic', 'Moravia')

('Egypt', 'Unknown')

('England', 'Unknown')

('France', 'Alsace')
('France', 'Beaujolais')
('France', 'Bordeaux')
('France', 'Burgundy')
('France', 'Champagne')
('France', 'Gironde')
('France', 'Languedoc-Roussillon')
('France', 'Loire Valley')
('France', 'Provence')
('France', 'Rhône Valley')
('France', 'Southwest France')
('France', 'Unknown')

('Georgia', 'Imereti')
('Georgia', 'Kakheti')
('Georgia', 'Kvemo Kartli')
('Georgia', 'Unknown')

('Germany', 'Ahr')
('Germany', 'Baden')
('Germany', 'Franken')
('Germany', 'Landwein Rhein')
('Germany', 'Mittelrhein')
('Germany', 'Mosel')
('Germany', 'Nahe')
('Germany', 'Pfalz')
('Germany', 'Rheingau')
('Germany', 'Rheinhessen')
('Germany', 'Unknown')
('Germany', 'Württemberg')
```

```
('Greece', 'Achaia')
('Greece', 'Agioritikos')
('Greece', 'Amindeo')
('Greece', 'Arcadia')
('Greece', 'Atalanti Valley')
('Greece', 'Attica')
('Greece', 'Beotia')
('Greece', 'Central Greece')
('Greece', 'Cephalonia')
('Greece', 'Chalkidiki')
('Greece', 'Corinth')
('Greece', 'Corinthia')
('Greece', 'Crete')
('Greece', 'Cyclades')
('Greece', 'Drama')
('Greece', 'Epanomi')
('Greece', 'Florina')
('Greece', 'Goumenissa')
('Greece', 'Halkidiki')
('Greece', 'Imathia')
('Greece', 'Ismarikos')
('Greece', 'Krania Olympus')
('Greece', 'Lakonia')
('Greece', 'Lesbos')
('Greece', 'Letrinon')
('Greece', 'Limnos')
('Greece', 'Macedonia')
('Greece', 'Mantinia')
('Greece', 'Markopoulo')
('Greece', 'Mavrodaphne de Cephalonie')
('Greece', 'Mavrodaphne of Patras')
('Greece', 'Messinia')
('Greece', 'Monemvasia')
('Greece', 'Mount Athos')
('Greece', 'Muscat of Kefallonian')
('Greece', 'Muscat of Patras')
('Greece', 'Naoussa')
('Greece', 'Nemea')
('Greece', 'Pangeon')
('Greece', 'Patras')
('Greece', 'Peloponnese')
('Greece', 'Rapsani')
('Greece', 'Retsina')
('Greece', 'Samos')
('Greece', 'Santorini')
('Greece', 'Sithonia')
('Greece', 'Sterea Ellada')
('Greece', 'Thessaly')
('Greece', 'Thraki')
('Greece', 'Unknown')
('Greece', 'Vin de Pays de Velvendo')
('Greece', 'Western Macedonia')

('Hungary', 'Eger')
('Hungary', 'Mátra')
('Hungary', 'Sopron')
('Hungary', 'Szekszárd')
('Hungary', 'Tokaj')
('Hungary', 'Unknown')
('Hungary', 'Veszprem')
('Hungary', 'Villány')

('India', 'Nashik')

('Israel', 'Dan')
('Israel', 'Ella Valley')
('Israel', 'Galil')
('Israel', 'Galilee')
('Israel', 'Golan Heights')
('Israel', 'Haut-Judeé')
('Israel', 'Jerusalem Hills')
('Israel', 'Judea and Samaria')
('Israel', 'Judean Hills')
('Israel', 'Negev')
('Israel', 'Samson')
('Israel', 'Shomron')
('Israel', 'Unknown')

('Italy', 'Central Italy')
```

```
('Italy', 'Lombardy')
('Italy', 'Northeastern Italy')
('Italy', 'Northwestern Italy')
('Italy', 'Piedmont')
('Italy', 'Sicily & Sardinia')
('Italy', 'Southern Italy')
('Italy', 'Tuscany')
('Italy', 'Unknown')
('Italy', 'Veneto')

('Kosovo', 'Jakove')

('Lebanon', 'Bekaa Valley')
('Lebanon', 'Unknown')

('Luxembourg', 'Moselle Luxembourgeoise')

('Macedonia', 'Tikves')

('Mexico', 'San Antonio de las Minas Valley')
('Mexico', 'San Vicente')
('Mexico', 'Valle de Guadalupe')

('Moldova', 'Cahul')
('Moldova', 'Codru Region')
('Moldova', 'Unknown')

('Morocco', 'Guerrouane')
('Morocco', 'Unknown')
('Morocco', 'Zenata')

('New Zealand', 'Awatere Valley')
('New Zealand', 'Canterbury')
('New Zealand', 'Central Otago')
('New Zealand', 'East Coast')
('New Zealand', 'Gisborne')
('New Zealand', 'Gladstone')
('New Zealand', "Hawke's Bay")
('New Zealand', 'Kumeu')
('New Zealand', 'Marlborough')
('New Zealand', 'Martinborough')
('New Zealand', 'Nelson')
('New Zealand', 'South Island')
('New Zealand', 'Unknown')
('New Zealand', 'Waiheke Island')
('New Zealand', 'Waipara')
('New Zealand', 'Wairarapa')
('New Zealand', 'Wairau Valley')
('New Zealand', 'Waitaki Valley')

('Peru', 'Ica')

('Portugal', 'Alenquer')
('Portugal', 'Alentejo')
('Portugal', 'Algarve')
('Portugal', 'Bairrada')
('Portugal', 'Beira Atlantico')
('Portugal', 'Beira Interior')
('Portugal', 'Beiras')
('Portugal', 'Bucelas')
('Portugal', 'Colares')
('Portugal', 'Douro')
('Portugal', 'Duriense')
('Portugal', 'Dão')
('Portugal', 'Estremadura')
('Portugal', 'Lisboa')
('Portugal', 'Madeira')
('Portugal', 'Minho')
('Portugal', 'Moscatel de Setúbal')
('Portugal', 'Moscatel do Douro')
('Portugal', 'Obidos')
('Portugal', 'Palmela')
('Portugal', 'Península de Setúbal')
('Portugal', 'Port')
('Portugal', 'Ribatejo')
('Portugal', 'Setubal')
('Portugal', 'Tejo')
('Portugal', 'Terras do Dão')
('Portugal', 'Terras do Sado')
('Portugal', 'Trás-os-Montes')
```

```
('Portugal', 'Távora-Varosa')
('Portugal', 'Unknown')
('Portugal', 'Vinho Espumante')
('Portugal', 'Vinho Licoroso')
('Portugal', 'Vinho Verde')
('Portugal', 'Vinho da Mesa')

('Romania', 'Dealu Mare')
('Romania', 'Dealurile Hușilor')
('Romania', 'Dealurile Munteniei')
('Romania', 'Jidvei')
('Romania', 'Murfatlar')
('Romania', 'Panciu')
('Romania', 'Recas')
('Romania', 'Sebes')
('Romania', 'Tarnave')
('Romania', 'Unknown')
('Romania', 'Viile Timis')
('Romania', 'Vânju Mare')

('Serbia', 'Fruška Gora')
('Serbia', 'Pocerina')
('Serbia', 'Župa')

('Slovakia', 'Muzla')

('Slovenia', 'Brda')
('Slovenia', 'Dolenjska')
('Slovenia', 'Goriska Brda')
('Slovenia', 'Kras')
('Slovenia', 'Primorska')
('Slovenia', 'Slovenska Istra')
('Slovenia', 'Unknown')
('Slovenia', 'Vipavska Dolina')
('Slovenia', 'Štajerska')

('South Africa', 'Bot River')
('South Africa', 'Breede River Valley')
('South Africa', 'Cape Agulhas')
('South Africa', 'Cape Peninsula')
('South Africa', 'Cape South Coast')
('South Africa', 'Cederberg')
('South Africa', 'Ceres Plateau')
('South Africa', 'Coastal Region')
('South Africa', 'Constantia')
('South Africa', 'Darling')
('South Africa', 'Devon Valley')
('South Africa', 'Durbanville')
('South Africa', 'Eilandia')
('South Africa', 'Elgin')
('South Africa', 'Elim')
('South Africa', 'Franschhoek')
('South Africa', 'Groenekloof')
('South Africa', 'Helderberg')
('South Africa', 'Hemel en Aarde')
('South Africa', 'Jonkershoek Valley')
('South Africa', 'Lutzville Valley')
('South Africa', 'Malgas')
('South Africa', 'Northern Cape')
('South Africa', 'Olifants River')
('South Africa', 'Overberg')
('South Africa', 'Paardeberg')
('South Africa', 'Paarl')
('South Africa', 'Philadelphia')
('South Africa', 'Piekenierskloof')
('South Africa', 'Polkadraai Hills')
('South Africa', 'Robertson')
('South Africa', 'Simonsberg-Paarl')
('South Africa', 'Simonsberg-Stellenbosch')
('South Africa', 'Stellenbosch')
('South Africa', 'Swartland')
('South Africa', 'Tulbagh')
('South Africa', 'Unknown')
('South Africa', 'Vlootenburg')
('South Africa', 'Walker Bay')
('South Africa', 'Wellington')
('South Africa', 'Western Cape')

('Spain', 'Andalucia')
('Spain', 'Catalonia')
```

```
('Spain', 'Central Spain')
('Spain', 'Galicia')
('Spain', 'Levante')
('Spain', 'Northern Spain')
('Spain', 'Spanish Islands')
('Spain', 'Unknown')

('Switzerland', 'Neuchâtel')
('Switzerland', 'Ticino')
('Switzerland', 'Unknown')
('Switzerland', 'Valais')
('Switzerland', 'Vaud')

('Turkey', 'Aegean')
('Turkey', 'Ankara')
('Turkey', 'Balıkesir')
('Turkey', 'Cappadocia')
('Turkey', 'Elazığ')
('Turkey', 'Elazığ-Diyarbakir')
('Turkey', 'Thrace')
('Turkey', 'Unknown')
('Turkey', 'Urla')

('US', 'Arizona')
('US', 'California')
('US', 'Colorado')
('US', 'Connecticut')
('US', 'Hawaii')
('US', 'Idaho')
('US', 'Illinois')
('US', 'Iowa')
('US', 'Kentucky')
('US', 'Massachusetts')
('US', 'Michigan')
('US', 'Missouri')
('US', 'Nevada')
('US', 'New Jersey')
('US', 'New Mexico')
('US', 'New York')
('US', 'North Carolina')
('US', 'Ohio')
('US', 'Oregon')
('US', 'Pennsylvania')
('US', 'Rhode Island')
('US', 'Texas')
('US', 'Unknown')
('US', 'Vermont')
('US', 'Virginia')
('US', 'Washington')

('Ukraine', 'Unknown')

('Uruguay', 'Atlantida')
('Uruguay', 'Canelones')
('Uruguay', 'Durazno')
('Uruguay', 'Juanico')
('Uruguay', 'Montevideo')
('Uruguay', 'Progreso')
('Uruguay', 'San Jose')
('Uruguay', 'Unknown')
```

There are two main issues here:

- potential duplicate provinces
- non-descriptive province information, like the country name or "other"

The following are the "(country,province)" pairs that potentially have issues:

Potential duplicate provinces\ ~('Austria', 'Südoststeiermark')~ -- upon further investigation, not a duplicate\ ~('Austria', 'Südsteiermark')~ -- upon further investigation, not a duplicate\ ('Austria', 'Wagram')\ ('Austria', 'Wagram-Donauland') -- Wagram\ ('Chile', 'Casablanca & Leyda Valleys') -- Casablanca Valley\ ('Chile', 'Casablanca Valley')\ ('Chile', 'Casablanca-Curicó Valley') -- Casablanca Valley\ ('Chile', 'Curicó Valley')\ ('Chile', 'Curicó and Leyda Valleys') -- Curicó Valley\ ('Chile', 'Curicó and Maipo Valleys') -- Curicó Valley\ ('Chile', 'Leyda Valley')\ ('Chile', 'Leyda Valley-Maipo Valley') -- Leyda Valley\ ('Chile', 'Maipo Valley')\ ('Chile', 'Maipo Valley-Colchagua Valley') -- Maipo Valley\ ('Chile', 'Colchagua Valley')\ ('Chile', 'Maule') -- Maule Valley\ ('Chile', 'Maule Valley')\ ('Germany', 'Mosel')\ ('Germany', 'Mosel-Saar-Ruwer') -- Mosel\ ('Greece', 'Amindeo')\ ('Greece', 'Amyndeon') -- Amindeo\ ~('Greece', 'Corinth')~ -- upon further investigation, not a duplicate\ ('Greece', 'Corinthia')\ ('Greece', 'Korinthia') -- Corinthia\ ('Greece', 'Pageon') -- typo of

Pangeon\ ('Greece', 'Pangeon')\ ('Greece', 'Thessalikos') -- Thessaly\ ('Greece', 'Thessaly')\ ('Hungary', 'Tokaj')\ ('Hungary', 'Tokaji') -- Tokaj\ ~('Israel', 'Galil')~ -- upon further investigation, not a duplicate\ ('Israel', 'Galilee')\ ('Israel', 'Upper Galilee') -- Galilee\ ~('Israel', 'Judea and Samaria')~ -- upon further investigation, not a duplicate\ ~('Israel', 'Judean Hills')~ -- upon further investigation, not a duplicate\ ('Israel', 'Negev')\ ('Israel', 'Negev Hills') -- Negev\ ('New Zealand', 'Waipara')\ ('New Zealand', 'Waipara Valley') -- Waipara\ ~('New Zealand', 'Wairarapa')~ -- upon further investigation, not a duplicate\ ~('New Zealand', 'Wairau Valley')~ -- upon further investigation, not a duplicate\ ('Portugal', 'Alentejano') -- Alentejo\ ('Portugal', 'Alentejo')\ ('Portugal', 'Alenteo') -- Alentejo\ ('Portugal', 'Ribatejano') -- Ribatejo\ ('Portugal', 'Ribatejo')\ ('Portugal', 'Vinho Espumante')\ ('Portugal', 'Vinho Espumante de Qualidade') -- Vinho Espumante\ ('Romania', 'Viile Timis')\ ('Romania', 'Viile Timisului') -- Viile Timis\ ('South Africa', 'Breede River Valley')\ ('South Africa', 'Breedekloof') -- Breede River Valley\ ~('South Africa', 'Simonsberg-Paarl')~ -- upon further investigation, not a duplicate\ ~('South Africa', 'Simonsberg-Stellenbosch')~ -- upon further investigation, not a duplicate\ ~('Turkey', 'Elazığ')~ -- upon further investigation, not a duplicate\ ~('Turkey', 'Elazığ-Diyarbakir')~ -- upon further investigation, not a duplicate\ ('Turkey', 'Thrace')\ ('Turkey', 'Urla-Thrace') -- Urla\ ('US', 'Oregon')\ ('US', 'Washington')\ ('US', 'Washington-Oregon') -- see below section for resolution

Non-descriptive or incorrect provinces\ ('Argentina', 'Other')\ ('Armenia', 'Armenia')\ ('Australia', 'Australia Other')\ ('Austria', 'Austria')\ ('Brazil', 'Brazil')\ ('Bulgaria', 'Bulgaria')\ ('Canada', 'Canada Other')\ ('Chile', 'Chile')\ ('China', 'China')\ ('Croatia', 'Croatia')\ ('Cyprus', 'Cyprus')\ ('Egypt', 'Egypt')\ ('England', 'England')\ ('France', 'France Other')\ ('Georgia', 'Georgia')\ ('Germany', 'Germany')\ ('Greece', 'Greece')\ ('Hungary', 'Hungary')\ ('Israel', 'Israel')\ ('Italy', 'Italy Other')\ ('Lebanon', 'Lebanon')\ ('Moldova', 'Moldova')\ ('Morocco', 'Morocco')\ ('New Zealand', 'New Zealand')\ ('Portugal', 'Portugal')\ ('Romania', 'Romania')\ ('Slovenia', 'Slovenia')\ ('South Africa', 'South Africa')\ ('Spain', 'Spain Other')\ ('Switzerland', 'Switzerland')\ ('Turkey', 'Turkey')\ ('US', 'America')\ ('Ukraine', 'Ukraine')\ ('Uruguay', 'Uruguay')\ ('Portugal', 'Portuguese Table Wine')\ ('Portugal', 'Table wine')

```
In [96]:   # Check where "Washington-Oregon" occurs
           df.loc[locations[('US', 'Washington-Oregon')]]
```

Out[96]:

	country	province	region_1	region_2	winery	winery_id
8278	US	Washington-Oregon	NaN	NaN	J. Scott Cellars	8278
26253	US	Washington-Oregon	NaN	NaN	Corvidae	26253
26873	US	Washington-Oregon	NaN	NaN	Ken Wright	26873
30370	US	Washington-Oregon	NaN	NaN	ENSO	30370

After further investigation of each of the above wineries:

J. Scott Cellars should be Oregon \ Corvidae should be Washington\ Ken Wright should be Oregon\ ENSO should be Oregon

```
In [97]:   # Cleaning of these four winery instances:
           df.at[8278,"province"] = "Oregon"
           df.at[26253,"province"] = "Washington"
           df.at[26873,"province"] = "Oregon"
           df.at[30370,"province"] = "Oregon"
```

Cleaning changes to be made for duplicate provinces

('Austria', 'Wagram-Donauland') -- Wagram\ ('Chile', 'Casablanca & Leyda Valleys') -- Casablanca Valley\ ('Chile', 'Casablanca-Curicó Valley') -- Casablanca Valley\ ('Chile', 'Curicó and Leyda Valleys') -- Curicó Valley\ ('Chile', 'Curicó and Maipo Valleys') -- Curicó Valley\ ('Chile', 'Leyda Valley-Maipo Valley') -- Leyda Valley\ ('Chile', 'Maipo Valley-Colchagua Valley') -- Maipo Valley\ ('Chile', 'Maule') -- Maule Valley\ ('Germany', 'Mosel-Saar-Ruwer') -- Mosel\ ('Greece', 'Amyndeon') -- Amindeo\ ('Greece', 'Korinthia') -- Corinthia\ ('Greece', 'Pageon') -- Pangeon\ ('Greece', 'Thessalikos') -- Thessaly\ ('Hungary', 'Tokaji') -- Tokaj\ ('Israel', 'Upper Galilee') -- Galilee\ ('Israel', 'Negev Hills') -- Negev\ ('New Zealand', 'Waipara Valley') -- Waipara\ ('Portugal', 'Alentejano') -- Alentejo\ ('Portugal', 'Alenteo') -- Alentejo\ ('Portugal', 'Ribatejano') -- Ribatejo\ ('Portugal', 'Vinho Espumante de Qualidade') -- Vinho Espumante\ ('Romania', 'Viile Timisului') -- Viile Timis\ ('South Africa', 'Breedekloof') -- Breede River Valley\ ('Turkey', 'Urla-Thrace') -- Urla

```
In [98]:   values_to_replace = [
               ("Wagram-Donauland","Wagram"),
               ("Casablanca & Leyda Valleys","Casablanca Valley"),
               ("Casablanca-Curicó Valley","Casablanca Valley"),
               ("Curicó and Leyda Valleys","Curicó Valley"),
               ("Curicó and Maipo Valleys","Curicó Valley"),
               ("Leyda Valley-Maipo Valley","Leyda Valley"),
               ("Maipo Valley-Colchagua Valley","Maipo Valley"),
               ("Maule","Maule Valley"),
               ("Mosel-Saar-Ruwer","Mosel"),
               ("Amyndeon","Amindeo"),
               ("Korinthia","Corinthia"),
               ("Pageon","Pangeon"),
```

```
        ("Thessalikos","Thessaly"),
        ("Tokaji","Tokaj"),
        ("Upper Galilee","Galilee"),
        ("Negev Hills","Negev"),
        ("Waipara Valley","Waipara"),
        ("Alentejano","Alentejo"),
        ("Alenteo","Alentejo"),
        ("Ribatejano","Ribatejo"),
        ("Vinho Espumante de Qualidade","Vinho Espumante"),
        ("Viile Timisului","Viile Timis"),
        ("Breedekloof","Breede River Valley"),
        ("Urla-Thrace","Urla")
    ]
    for old,new in values_to_replace:
        for idx in df.index[df['province']==old]:
            df.at[idx,"province"]=new
```

Cleaning changes for non-descriptive or incorrect provinces:

Change all to "Other"

```
In [99]:  vague_provinces = ['Other', 'Armenia', 'Australia Other', 'Austria', 'Brazil', 'Bulgaria', 'Canada Other', 'Chile', 'Chi
                             'Croatia', 'Cyprus', 'Egypt', 'England', 'France Other', 'Georgia', 'Germany', 'Greece', 'Hungary', '
                             'Italy Other', 'Lebanon', 'Moldova', 'Morocco', 'New Zealand', 'Portugal', 'Romania', 'Slovenia',
                             'South Africa', 'Spain Other', 'Switzerland', 'Turkey', 'America', 'Ukraine', 'Uruguay',
                             'Portuguese Table Wine', 'Table wine']
          for prov in vague_provinces:
              for idx in df.index[df['province']==prov]:
                  df.at[idx,"province"]="Unknown"
```

Double check list of provinces again

```
In [100…  locations_new = {}
          for i in range(len(df)):
              location = (df['country'].loc[i],df['province'].loc[i])
              if location not in locations_new:
                  locations_new[location] = [i]
              else:
                  locations_new[location].append(i)

          locations_list_new=sorted(locations_new)

          curr_country=None
          for k in range(len(locations_list_new)):
              if locations_list_new[k][0]!=curr_country:
                  print()
                  curr_country = locations_list_new[k][0]
              print(locations_list_new[k])
```

```
('Argentina', 'Mendoza Province')
('Argentina', 'Unknown')

('Armenia', 'Unknown')

('Australia', 'New South Wales')
('Australia', 'South Australia')
('Australia', 'Tasmania')
('Australia', 'Unknown')
('Australia', 'Victoria')
('Australia', 'Western Australia')

('Austria', 'Burgenland')
('Austria', 'Carnuntum')
('Austria', 'Eisenberg')
('Austria', 'Kamptal')
('Austria', 'Kremstal')
('Austria', 'Leithaberg')
('Austria', 'Mittelburgenland')
('Austria', 'Neusiedlersee')
('Austria', 'Niederösterreich')
('Austria', 'Steiermark')
('Austria', 'Styria')
('Austria', 'Südburgenland')
('Austria', 'Südoststeiermark')
('Austria', 'Südsteiermark')
('Austria', 'Thermenregion')
('Austria', 'Traisental')
('Austria', 'Unknown')
('Austria', 'Vienna')
('Austria', 'Wachau')
('Austria', 'Wagram')
('Austria', 'Weinland Österreich')
('Austria', 'Weinviertel')
('Austria', 'Wiener Gemischter Satz')
('Austria', 'Österreichischer Perlwein')
('Austria', 'Österreichischer Sekt')

('Bosnia and Herzegovina', 'Mostar')

('Brazil', 'Campanha')
('Brazil', 'Pinto Bandeira')
('Brazil', 'Santa Catarina')
('Brazil', 'Serra Gaúcha')
('Brazil', 'Serra do Sudeste')
('Brazil', 'Unknown')
('Brazil', 'Vale Trentino')
('Brazil', 'Vale dos Vinhedos')

('Bulgaria', 'Black Sea Coastal')
('Bulgaria', 'Blagoevgrad')
('Bulgaria', 'Danube River Plains')
('Bulgaria', 'Nikolaevo')
('Bulgaria', 'Thracian Valley')
('Bulgaria', 'Unknown')

('Canada', 'British Columbia')
('Canada', 'Ontario')
('Canada', 'Unknown')

('Chile', 'Aconcagua Costa')
('Chile', 'Aconcagua Valley')
('Chile', 'Apalta')
('Chile', 'Buin')
('Chile', 'Bío Bío Valley')
('Chile', 'Cachapoal Valley')
('Chile', 'Casablanca Valley')
('Chile', 'Cauquenes Valley')
('Chile', 'Central Valley')
('Chile', 'Choapa Valley')
('Chile', 'Coelemu')
('Chile', 'Colchagua Costa')
('Chile', 'Colchagua Valley')
('Chile', 'Curicó Valley')
('Chile', 'Elqui Valley')
('Chile', 'Itata Valley')
('Chile', 'Leyda Valley')
('Chile', 'Limarí Valley')
('Chile', 'Lolol Valley')
('Chile', 'Loncomilla Valley')
```

```
('Chile', 'Lontué Valley')
('Chile', 'Maipo Valley')
('Chile', 'Malleco')
('Chile', 'Marchigue')
('Chile', 'Maule Valley')
('Chile', 'Molina')
('Chile', 'Peumo')
('Chile', 'Pirque')
('Chile', 'Puente Alto')
('Chile', 'Rapel Valley')
('Chile', 'Requinoa')
('Chile', 'Rio Claro')
('Chile', 'Sagrada Familia')
('Chile', 'San Antonio')
('Chile', 'San Clemente')
('Chile', 'Santa Cruz')
('Chile', 'Santiago')
('Chile', 'Unknown')

('China', 'Unknown')

('Croatia', 'Dalmatian Coast')
('Croatia', 'Dingač')
('Croatia', 'Hrvatsko Primorje')
('Croatia', 'Hvar')
('Croatia', 'Istria')
('Croatia', 'Korčula')
('Croatia', 'Krk')
('Croatia', 'Kutjevo')
('Croatia', 'Middle and South Dalmatia')
('Croatia', 'North Dalmatia')
('Croatia', 'Peljesac')
('Croatia', 'Podunavlje')
('Croatia', 'Unknown')

('Cyprus', 'Commandaria')
('Cyprus', 'Kathikas')
('Cyprus', 'Lemesos')
('Cyprus', 'Pafos')
('Cyprus', 'Pitsilia Mountains')
('Cyprus', 'Unknown')

('Czech Republic', 'Moravia')

('Egypt', 'Unknown')

('England', 'Unknown')

('France', 'Alsace')
('France', 'Beaujolais')
('France', 'Bordeaux')
('France', 'Burgundy')
('France', 'Champagne')
('France', 'Gironde')
('France', 'Languedoc-Roussillon')
('France', 'Loire Valley')
('France', 'Provence')
('France', 'Rhône Valley')
('France', 'Southwest France')
('France', 'Unknown')

('Georgia', 'Imereti')
('Georgia', 'Kakheti')
('Georgia', 'Kvemo Kartli')
('Georgia', 'Unknown')

('Germany', 'Ahr')
('Germany', 'Baden')
('Germany', 'Franken')
('Germany', 'Landwein Rhein')
('Germany', 'Mittelrhein')
('Germany', 'Mosel')
('Germany', 'Nahe')
('Germany', 'Pfalz')
('Germany', 'Rheingau')
('Germany', 'Rheinhessen')
('Germany', 'Unknown')
('Germany', 'Württemberg')

('Greece', 'Achaia')
```

```
('Greece', 'Agioritikos')
('Greece', 'Amindeo')
('Greece', 'Arcadia')
('Greece', 'Atalanti Valley')
('Greece', 'Attica')
('Greece', 'Beotia')
('Greece', 'Central Greece')
('Greece', 'Cephalonia')
('Greece', 'Chalkidiki')
('Greece', 'Corinth')
('Greece', 'Corinthia')
('Greece', 'Crete')
('Greece', 'Cyclades')
('Greece', 'Drama')
('Greece', 'Epanomi')
('Greece', 'Florina')
('Greece', 'Goumenissa')
('Greece', 'Halkidiki')
('Greece', 'Imathia')
('Greece', 'Ismarikos')
('Greece', 'Krania Olympus')
('Greece', 'Lakonia')
('Greece', 'Lesbos')
('Greece', 'Letrinon')
('Greece', 'Limnos')
('Greece', 'Macedonia')
('Greece', 'Mantinia')
('Greece', 'Markopoulo')
('Greece', 'Mavrodaphne de Cephalonie')
('Greece', 'Mavrodaphne of Patras')
('Greece', 'Messinia')
('Greece', 'Monemvasia')
('Greece', 'Mount Athos')
('Greece', 'Muscat of Kefallonian')
('Greece', 'Muscat of Patras')
('Greece', 'Naoussa')
('Greece', 'Nemea')
('Greece', 'Pangeon')
('Greece', 'Patras')
('Greece', 'Peloponnese')
('Greece', 'Rapsani')
('Greece', 'Retsina')
('Greece', 'Samos')
('Greece', 'Santorini')
('Greece', 'Sithonia')
('Greece', 'Sterea Ellada')
('Greece', 'Thessaly')
('Greece', 'Thraki')
('Greece', 'Unknown')
('Greece', 'Vin de Pays de Velvendo')
('Greece', 'Western Macedonia')

('Hungary', 'Eger')
('Hungary', 'Mátra')
('Hungary', 'Sopron')
('Hungary', 'Szekszárd')
('Hungary', 'Tokaj')
('Hungary', 'Unknown')
('Hungary', 'Veszprem')
('Hungary', 'Villány')

('India', 'Nashik')

('Israel', 'Dan')
('Israel', 'Ella Valley')
('Israel', 'Galil')
('Israel', 'Galilee')
('Israel', 'Golan Heights')
('Israel', 'Haut-Judeé')
('Israel', 'Jerusalem Hills')
('Israel', 'Judea and Samaria')
('Israel', 'Judean Hills')
('Israel', 'Negev')
('Israel', 'Samson')
('Israel', 'Shomron')
('Israel', 'Unknown')

('Italy', 'Central Italy')
('Italy', 'Lombardy')
('Italy', 'Northeastern Italy')
```

```
('Italy', 'Northwestern Italy')
('Italy', 'Piedmont')
('Italy', 'Sicily & Sardinia')
('Italy', 'Southern Italy')
('Italy', 'Tuscany')
('Italy', 'Unknown')
('Italy', 'Veneto')

('Kosovo', 'Jakove')

('Lebanon', 'Bekaa Valley')
('Lebanon', 'Unknown')

('Luxembourg', 'Moselle Luxembourgeoise')

('Macedonia', 'Tikves')

('Mexico', 'San Antonio de las Minas Valley')
('Mexico', 'San Vicente')
('Mexico', 'Valle de Guadalupe')

('Moldova', 'Cahul')
('Moldova', 'Codru Region')
('Moldova', 'Unknown')

('Morocco', 'Guerrouane')
('Morocco', 'Unknown')
('Morocco', 'Zenata')

('New Zealand', 'Awatere Valley')
('New Zealand', 'Canterbury')
('New Zealand', 'Central Otago')
('New Zealand', 'East Coast')
('New Zealand', 'Gisborne')
('New Zealand', 'Gladstone')
('New Zealand', "Hawke's Bay")
('New Zealand', 'Kumeu')
('New Zealand', 'Marlborough')
('New Zealand', 'Martinborough')
('New Zealand', 'Nelson')
('New Zealand', 'South Island')
('New Zealand', 'Unknown')
('New Zealand', 'Waiheke Island')
('New Zealand', 'Waipara')
('New Zealand', 'Wairarapa')
('New Zealand', 'Wairau Valley')
('New Zealand', 'Waitaki Valley')

('Peru', 'Ica')

('Portugal', 'Alenquer')
('Portugal', 'Alentejo')
('Portugal', 'Algarve')
('Portugal', 'Bairrada')
('Portugal', 'Beira Atlantico')
('Portugal', 'Beira Interior')
('Portugal', 'Beiras')
('Portugal', 'Bucelas')
('Portugal', 'Colares')
('Portugal', 'Douro')
('Portugal', 'Duriense')
('Portugal', 'Dão')
('Portugal', 'Estremadura')
('Portugal', 'Lisboa')
('Portugal', 'Madeira')
('Portugal', 'Minho')
('Portugal', 'Moscatel de Setúbal')
('Portugal', 'Moscatel do Douro')
('Portugal', 'Obidos')
('Portugal', 'Palmela')
('Portugal', 'Península de Setúbal')
('Portugal', 'Port')
('Portugal', 'Ribatejo')
('Portugal', 'Setubal')
('Portugal', 'Tejo')
('Portugal', 'Terras do Dão')
('Portugal', 'Terras do Sado')
('Portugal', 'Trás-os-Montes')
('Portugal', 'Távora-Varosa')
('Portugal', 'Unknown')
```

```
('Portugal', 'Vinho Espumante')
('Portugal', 'Vinho Licoroso')
('Portugal', 'Vinho Verde')
('Portugal', 'Vinho da Mesa')

('Romania', 'Dealu Mare')
('Romania', 'Dealurile Huşilor')
('Romania', 'Dealurile Munteniei')
('Romania', 'Jidvei')
('Romania', 'Murfatlar')
('Romania', 'Panciu')
('Romania', 'Recas')
('Romania', 'Sebes')
('Romania', 'Tarnave')
('Romania', 'Unknown')
('Romania', 'Viile Timis')
('Romania', 'Vânju Mare')

('Serbia', 'Fruška Gora')
('Serbia', 'Pocerina')
('Serbia', 'Župa')

('Slovakia', 'Muzla')

('Slovenia', 'Brda')
('Slovenia', 'Dolenjska')
('Slovenia', 'Goriska Brda')
('Slovenia', 'Kras')
('Slovenia', 'Primorska')
('Slovenia', 'Slovenska Istra')
('Slovenia', 'Unknown')
('Slovenia', 'Vipavska Dolina')
('Slovenia', 'Štajerska')

('South Africa', 'Bot River')
('South Africa', 'Breede River Valley')
('South Africa', 'Cape Agulhas')
('South Africa', 'Cape Peninsula')
('South Africa', 'Cape South Coast')
('South Africa', 'Cederberg')
('South Africa', 'Ceres Plateau')
('South Africa', 'Coastal Region')
('South Africa', 'Constantia')
('South Africa', 'Darling')
('South Africa', 'Devon Valley')
('South Africa', 'Durbanville')
('South Africa', 'Eilandia')
('South Africa', 'Elgin')
('South Africa', 'Elim')
('South Africa', 'Franschhoek')
('South Africa', 'Groenekloof')
('South Africa', 'Helderberg')
('South Africa', 'Hemel en Aarde')
('South Africa', 'Jonkershoek Valley')
('South Africa', 'Lutzville Valley')
('South Africa', 'Malgas')
('South Africa', 'Northern Cape')
('South Africa', 'Olifants River')
('South Africa', 'Overberg')
('South Africa', 'Paardeberg')
('South Africa', 'Paarl')
('South Africa', 'Philadelphia')
('South Africa', 'Piekenierskloof')
('South Africa', 'Polkadraai Hills')
('South Africa', 'Robertson')
('South Africa', 'Simonsberg-Paarl')
('South Africa', 'Simonsberg-Stellenbosch')
('South Africa', 'Stellenbosch')
('South Africa', 'Swartland')
('South Africa', 'Tulbagh')
('South Africa', 'Unknown')
('South Africa', 'Vlootenburg')
('South Africa', 'Walker Bay')
('South Africa', 'Wellington')
('South Africa', 'Western Cape')

('Spain', 'Andalucia')
('Spain', 'Catalonia')
('Spain', 'Central Spain')
('Spain', 'Galicia')
```

```
('Spain', 'Levante')
('Spain', 'Northern Spain')
('Spain', 'Spanish Islands')
('Spain', 'Unknown')

('Switzerland', 'Neuchâtel')
('Switzerland', 'Ticino')
('Switzerland', 'Unknown')
('Switzerland', 'Valais')
('Switzerland', 'Vaud')

('Turkey', 'Aegean')
('Turkey', 'Ankara')
('Turkey', 'Balıkesir')
('Turkey', 'Cappadocia')
('Turkey', 'Elazığ')
('Turkey', 'Elazığ-Diyarbakir')
('Turkey', 'Thrace')
('Turkey', 'Unknown')
('Turkey', 'Urla')

('US', 'Arizona')
('US', 'California')
('US', 'Colorado')
('US', 'Connecticut')
('US', 'Hawaii')
('US', 'Idaho')
('US', 'Illinois')
('US', 'Iowa')
('US', 'Kentucky')
('US', 'Massachusetts')
('US', 'Michigan')
('US', 'Missouri')
('US', 'Nevada')
('US', 'New Jersey')
('US', 'New Mexico')
('US', 'New York')
('US', 'North Carolina')
('US', 'Ohio')
('US', 'Oregon')
('US', 'Pennsylvania')
('US', 'Rhode Island')
('US', 'Texas')
('US', 'Unknown')
('US', 'Vermont')
('US', 'Virginia')
('US', 'Washington')

('Ukraine', 'Unknown')

('Uruguay', 'Atlantida')
('Uruguay', 'Canelones')
('Uruguay', 'Durazno')
('Uruguay', 'Juanico')
('Uruguay', 'Montevideo')
('Uruguay', 'Progreso')
('Uruguay', 'San Jose')
('Uruguay', 'Unknown')
```

Attempt to clean "region_1" and "region_2" values

Check for duplicates

```python
In [110…  sorted(list(df['region_1'].dropna().unique()))
```

```
Out[110]:   ['Abruzzo',
             'Adelaida District',
             'Adelaide',
             'Adelaide Hills',
             'Adelaide Plains',
             "Aglianico d'Irpinia",
             'Aglianico del Beneventano',
             'Aglianico del Taburno',
             'Aglianico del Vulture',
             'Agrelo',
             'Albana di Romagna',
             'Alcamo',
             "Aleatico dell'Elba",
             'Alella',
             'Alexander Valley',
             'Alghero',
             'Alicante',
             'Almansa',
             'Aloxe-Corton',
             'Alpilles',
             'Alpine Valleys',
             'Alsace',
             'Alta Langa',
             'Alta Mesa',
             'Alta Valle della Greve',
             'Alto Adige',
             'Alto Adige Terlano',
             'Alto Adige Valle Isarco',
             'Alto Valle del Río Negro',
             'Altos de Mendoza',
             'Amador County',
             'Amador-Napa',
             'Amarone della Valpolicella',
             'Amarone della Valpolicella Classico',
             'Amarone della Valpolicella Valpantena',
             'Ancient Lakes',
             'Anderson Valley',
             'Anjou',
             'Anjou Villages',
             'Anjou Villages Brissac',
             'Antelope Valley of the California High Desert',
             'Applegate Valley',
             'Apremont',
             'Arbois',
             'Ardèche',
             'Argentina',
             'Arizona',
             'Arribes del Duero',
             'Arroyo Grande Valley',
             'Arroyo Seco',
             'Asolo Prosecco Superiore',
             'Asprinio di Aversa',
             'Assisi',
             'Asti',
             'Atlantique',
             'Atlas Peak',
             'Aude Hauterive',
             'Augusta',
             'Australia',
             'Australia-New Zealand',
             'Auxey-Duresses',
             'Ballard Canyon',
             'Bandol',
             'Banyuls',
             'Barbaresco',
             "Barbera d'Alba",
             "Barbera d'Alba Superiore",
             "Barbera d'Asti",
             "Barbera d'Asti Superiore",
             "Barbera d'Asti Superiore Nizza",
             'Barbera del Monferrato',
             'Barbera del Monferrato Superiore',
             'Barbera di Piemonte',
             'Barco Reale di Carmignano',
             'Bardolino',
             'Bardolino Chiaretto',
             'Bardolino Classico',
             'Bardolino Classico Superiore',
             'Bardolino Superiore',
             'Barolo',
```

'Barossa',
'Barossa Valley',
'Barossa-Adelaide Hills',
'Barossa-Langhorne Creek',
'Barsac',
'Basilicata',
'Beamsville Bench',
'Beaujolais',
'Beaujolais Blanc',
'Beaujolais Rosé',
'Beaujolais-Leynes',
'Beaujolais-Villages',
'Beaune',
'Beechworth',
'Ben Lomond Mountain',
'Benaco Bresciano',
'Bendigo',
'Beneventano',
'Bennett Valley',
'Bergerac',
'Bergerac Rosé',
'Bergerac Sec',
'Bianco di Custoza',
'Bienvenues Bâtard-Montrachet',
'Bierzo',
'Biferno Rosso',
'Big Valley District-Lake County',
'Bizkaiko Txakolina',
'Blanquette de Limoux',
'Blaye Côtes de Bordeaux',
'Boca',
'Bolgheri',
'Bolgheri Sassicaia',
'Bolgheri Superiore',
'Bonnes-Mares',
'Bonnezeaux',
'Bordeaux',
'Bordeaux Blanc',
'Bordeaux Clairet',
'Bordeaux Côtes de Francs ',
'Bordeaux Rosé',
'Bordeaux Supérieur',
'Bourgogne',
'Bourgogne Aligoté',
'Bourgogne Côtes d’Auxerre',
'Bourgogne Epineuil',
'Bourgogne Gamay',
'Bourgogne Hautes Côtes de Beaune',
'Bourgogne Hautes Côtes de Nuits',
'Bourgogne Vézelay',
'Bourgueil',
"Brachetto d'Acqui",
'Bramaterra',
'Breganze',
'Brindisi',
'British Columbia',
'Brouilly',
'Brulhois',
'Brunello di Montalcino',
'Buenos Aires',
'Bullas',
'Burgundy',
'Buzet',
'Bâtard-Montrachet',
'Béarn',
'Cabardes',
"Cabernet d'Anjou",
'Cabernet de Saumur',
'Cacc’e Mmitte di Lucera',
'Cadillac',
'Cadillac Côtes de Bordeaux',
'Cafayate',
'Cahors',
'Calabria',
'Calatayud',
'Calaveras County',
'Calchaquí Valley',
'California',
'California Other',
'California-Oregon',

'Calistoga',
'Campania',
'Campi Flegrei',
'Campo de Borja',
'Canada',
'Cannonau di Sardegna',
'Canon-Fronsac',
'Capay Valley',
'Carema',
'Carignano del Sulcis',
'Cariñena',
'Carmel Valley',
'Carmignano',
'Carneros',
'Carneros-Napa Valley',
'Casorzo',
'Castel del Monte',
'Castilla La Mancha',
'Castillon Côtes de Bordeaux',
'Catalanesca del Monte Somma',
'Catalonia',
'Catalunya',
'Catamarca',
'Cava',
'Cayuga Lake',
'Central Coast',
'Central Ranges',
'Central Valley',
'Central Victoria',
"Cerasuolo d'Abruzzo",
'Cerasuolo di Vittoria',
'Cerasuolo di Vittoria Classico',
'Cesanese del Piglio',
'Chablis',
'Chalk Hill',
'Chalone',
'Chambertin',
'Chambertin Clos de Bèze',
'Chambolle-Musigny',
'Champagne',
'Chapelle-Chambertin',
'Charentais',
'Charmes-Chambertin',
'Chassagne-Montrachet',
'Chehalem Mountains',
'Chelan County',
'Chevalier-Montrachet',
'Cheverny',
'Chianti',
'Chianti Classico',
'Chianti Colli Aretini',
'Chianti Colli Fiorentini',
'Chianti Colli Pisani',
'Chianti Colli Senesi',
'Chianti Montalbano',
'Chianti Montespertoli',
'Chianti Rufina',
'Chianti Superiore',
'Chignin',
'Chignin-Bergeron',
'Chiles Valley',
'Chinon',
'Chiroubles',
'Chorey-lès-Beaune',
'Château-Chalon',
'Châteaumeillant',
'Châteauneuf-du-Pape',
'Chénas',
'Cienega Valley',
'Cigales',
'Cilento',
'Circeo',
'Cirò',
'Cirò Classico',
"Civitella d'Agliano",
'Clairette de Bellegarde',
'Clairette de Die',
'Clairette de Die Méthode Dioise Ancestrale',
'Clare Valley',
'Clare Valley-Coonawarra',

'Clare Valley-Eden Valley',
'Clarendon',
'Clarksburg',
'Clear Lake',
'Clements Hills',
'Clos Saint-Denis',
'Clos de Lambrays',
'Clos de Tart',
'Clos de Vougeot',
'Clos de la Roche',
'Cochise County',
"Coda di Volpe d'Irpinia",
'Cole Ranch',
'Colli Aprutini',
'Colli Bolognesi',
'Colli Etruria Centrale',
'Colli Euganei',
'Colli Martani',
'Colli Orientali del Friuli',
'Colli Perugini',
'Colli Piacentini',
'Colli Tortonesi',
'Colli Trevigiani',
'Colli del Limbara',
'Colli della Toscana Centrale',
'Colli di Conegliano',
'Colli di Faenza',
'Colli di Luni',
'Colli di Salerno',
'Colline Lucchesi',
'Colline Novaresi',
'Colline Pescaresi',
'Colline Teatine',
'Colline Teramane',
'Collines Rhôdaniennes',
'Collio',
'Collioure',
'Colorado',
'Columbia Gorge (OR)',
'Columbia Gorge (WA)',
'Columbia Valley',
'Columbia Valley (OR)',
'Columbia Valley (WA)',
'Columbia Valley-Walla Walla Valley',
'Conca de Barberà',
'Condrieu',
'Conegliano Valdobbiadene Prosecco Superiore',
'Conero',
'Contea di Sclafani',
'Contessa Entellina',
'Contra Costa County',
'Controguerra',
'Coombsville',
'Coonawarra',
'Copertino',
'Corbières',
'Corbières-Boutenac',
'Cornas',
'Corse',
'Corse Porto Vecchio',
'Cortese di Gavi',
'Corton',
'Corton Grèves',
'Corton Les Renardes',
'Corton Perrières',
'Corton Vergennes',
'Corton-Bressandes',
'Corton-Charlemagne',
'Corton-Pougets',
'Corton-Rognet',
'Cortona',
'Costa Toscana',
"Costa d'Amalfi",
'Coste della Sesia',
'Costers del Segre',
'Costières de Nîmes',
'Coteaux Bourguignons',
'Coteaux Varois',
'Coteaux Varois en Provence',
"Coteaux d'Aix-en-Provence",

"Coteaux d'Ancenis",
'Coteaux de Verdon',
'Coteaux du Giennois',
'Coteaux du Languedoc',
'Coteaux du Languedoc Pic Saint Loup',
'Coteaux du Languedoc-Pézenas',
'Coteaux du Layon',
'Coteaux du Layon Beaulieu',
'Coteaux du Layon Chaume',
'Coteaux du Layon Saint Aubin',
'Coteaux du Lyonnais',
'Coteaux du Vendômois',
'Cour-Cheverny',
'Criots-Bâtard-Montrachet',
'Crozes-Hermitage',
"Crémant d'Alsace",
'Crémant de Bordeaux',
'Crémant de Bourgogne',
'Crémant de Jura',
'Crémant de Limoux',
'Crémant de Loire',
'Cucamonga Valley',
'Currency Creek',
'Currency Creek-McLaren Vale',
'Curtefranca',
'Custoza',
'Custoza Superiore',
'Cérons',
'Cévennes',
'Côte Chalonnaise',
'Côte Roannaise',
'Côte Rôtie',
'Côte de Beaune',
'Côte de Beaune-Villages',
'Côte de Brouilly',
'Côte de Nuits-Villages',
'Côtes Catalanes',
'Côtes de Bergerac',
'Côtes de Bordeaux',
'Côtes de Bourg',
'Côtes de Castillon',
'Côtes de Duras',
'Côtes de Forez',
'Côtes de Gascogne',
'Côtes de Montravel',
'Côtes de Nuits Villages',
'Côtes de Provence',
'Côtes de Provence La Londe',
'Côtes de Provence Sainte-Victoire',
'Côtes de Saint-Mont',
'Côtes de Thau',
'Côtes de Thongue',
'Côtes du Frontonnais',
'Côtes du Jura',
'Côtes du Lot',
'Côtes du Luberon',
'Côtes du Marmandais',
'Côtes du Rhône',
'Côtes du Rhône Villages',
'Côtes du Roussillon',
'Côtes du Roussillon Les Aspres',
'Côtes du Roussillon Villages',
'Côtes du Roussillon Villages Caramany',
'Côtes du Roussillon Villages Tautavel',
'Côtes du Tarn',
'Côtes-du-Ventoux',
'Del Veneto',
'Delia Nivolelli',
'Delle Venezie',
'Diamond Mountain District',
"Diano d'Alba",
'Dogliani',
'Dogliani Superiore',
"Dolcetto d'Acqui",
"Dolcetto d'Alba",
"Dolcetto d'Alba Superiore",
"Dolcetto d'Asti",
"Dolcetto di Diano d'Alba",
'Dolcetto di Dogliani',
'Dolcetto di Dogliani Superiore',

'Dolcetto di Monferrato',
'Dominio de Valdepusa',
'Donnici',
'Dry Creek Valley',
'Dundee Hills',
'Dunnigan Hills',
'Echézeaux',
'Eden Valley',
'Edna Valley',
'El Dorado',
'El Dorado County',
'El Pomar District',
'Elkton Oregon',
'Eloro',
'Emilia',
'Emilia-Romagna',
'Emporadà-Costa Brava',
'Empordà',
'Entre-Deux-Mers',
'Eola-Amity Hills',
'Epomeo',
'Erbaluce di Caluso',
'Erice',
'Est! Est!! Est!!! di Montefiascone',
'Etna',
'Extremadura',
'Fair Play',
'Falanghina del Beneventano',
'Falanghina del Molise',
'Falanghina del Sannio',
'Falerio',
'Falerio dei Colli Ascolani',
'Falerno del Massico',
'Famatina Valley',
'Fara',
'Faro',
'Faugères',
'Fiano di Avellino',
'Fiddletown',
'Finca Élez',
'Finger Lakes',
'Fitou',
'Fixin',
'Fleurie',
'Fleurieu Peninsula',
'Forlì',
'Fort Ross-Seaview',
'France',
'Franciacorta',
'Francs Côtes de Bordeaux',
'Frankland River',
'Frascati',
'Frascati Superiore',
'Friuli',
'Friuli Aquileia',
'Friuli Colli Orientali',
'Friuli Grave',
'Friuli Isonzo',
'Friuli Venezia Giulia',
'Fronsac',
'Fronton',
'Gaillac',
'Gaillac Doux',
'Galatina',
'Galluccio',
'Gambellara Classico',
'Gard',
'Garda',
'Garda Classico',
'Gattinara',
'Gavi',
'Gavi di Gavi',
'Geelong',
'Geographe',
'Gers',
'Getariako Txakolina',
'Gevrey-Chambertin',
'Ghemme',
'Gigondas',
'Gioia del Colle',

'Gippsland',
'Givry',
'Goulburn Valley',
'Grampians',
'Grand River Valley',
'Grand Valley',
'Grands-Echezeaux',
'Grave del Friuli',
'Graves',
'Great Southern',
'Greco di Tufo',
'Green Valley',
'Guenoc Valley',
'Gundagai',
'Gutturnio Classico Superiore',
'Gutturnio Colli Piacentini',
'Happy Canyon of Santa Barbara',
'Haut-Médoc',
'Hautes Cotes de Beaune',
'Hautes Côtes de Nuits',
'Hawaii',
'Heathcote',
'Henty',
'Hermitage',
'High Valley',
'Hilltops',
'Horse Heaven Hills',
'Howell Mountain',
'Hudson River Region',
'Humboldt County',
'Hunter Valley',
'Idaho',
'Ile de Beauté',
'Illinois',
'Iowa',
'Irancy',
'Irpinia',
'Irrouléguy',
'Ischia',
'Isle St. George',
'Isola dei Nuraghi',
'Isonzo del Friuli',
'Italy',
'Jerez',
'Jujuy',
'Juliénas',
'Jumilla',
'Jurançon',
'Jurançon Sec',
'Kangaroo Island',
'Kelsey Bench-Lake County',
'Kentucky',
'King Valley',
'Knights Valley',
'La Clape',
'La Consulta',
'La Grande Rue',
'La Mancha',
'La Rioja',
'La Romanée',
'Lacryma Christi del Vesuvio',
'Ladoix',
'Lago di Corbara',
'Lake Chelan',
'Lake County',
'Lake County-Mendocino County',
'Lake County-Sonoma County',
'Lake Erie',
'Lake Erie North Shore',
'Lake Michigan Shore',
'Lake-Sonoma-Napa',
'Lalande de Pomerol',
'Lambrusco Grasparossa di Castelvetro',
'Lambrusco Reggiano',
'Lambrusco Salamino di Santa Croce',
"Lambrusco dell'Emilia",
'Lambrusco di Modena',
'Lambrusco di Sorbara',
'Lamezia',
'Landes',

'Langhe',
'Langhorne Creek',
'Langhorne Creek-Limestone Coast',
'Languedoc',
'Lanzarote',
'Latricières-Chambertin',
'Laudun',
'Lazio',
'Leelanau Peninsula',
'Lenswood',
'Les Baux de Provence',
'Lessini Durello',
'Lessona',
'Lime Kiln Valley',
'Limestone Coast',
'Limoux',
'Lirac',
'Lison-Pramaggiore',
'Listrac-Médoc',
'Livermore Valley',
'Loazzolo',
'Lodi',
'Lombardy',
'Long Island',
'Los Carneros',
'Lot',
'Luberon',
'Lugana',
'Lugana Superiore',
'Luján de Cuyo',
'Lussac Saint-Émilion',
'Macon-Bussières',
'Madera',
'Madiran',
'Madrid',
'Maipú',
'Malibu Coast',
'Malibu-Newton Canyon',
'Malvasia delle Lipari',
'Mamertino',
'Manchuela',
'Manzanilla-Sanlúcar de Barrameda',
'Maranges',
'Marca Trevigiana',
'Marche',
'Marcillac',
'Maremma',
'Maremma Toscana',
'Margaret River',
'Margaux',
'Marin County',
'Marsala',
'Marsannay',
'Martina',
'Massachusetts',
'Maury',
'Maury Sec',
'Mazis-Chambertin',
'Mazoyeres-Chambertin',
'McDowell Valley',
'McLaren Vale',
'McLaren Vale-Adelaide Hills',
'McLaren Vale-Langhorne Creek',
'McLaren Vale-Padthaway',
'McMinnville',
'Mediterranée',
'Medrano',
'Mendocino',
'Mendocino County',
'Mendocino Ridge',
'Mendocino-Amador',
'Mendocino-Lake',
'Mendocino-Lake County-Napa',
'Mendoza',
'Menetou-Salon',
'Menfi',
'Mentrida',
'Mercurey',
'Meursault',
'Meursault-Blagny',

'Michigan',
'Middleburg',
'Minervois',
'Minervois La Liviniere',
'Missouri',
'Mitterberg',
'Mokelumne River',
'Molise',
'Monbazillac',
'Monferrato',
'Monferrato Bianco',
'Monica di Sardegna',
'Monreale',
'Montagne-Saint-Émilion',
'Montagny',
'Montecarlo',
'Montecucco',
'Montefalco',
'Montefalco Rosso',
'Montefalco Sagrantino',
'Montello e Colli Asolani',
'Montenetto di Brescia',
"Montepulciano d'Abruzzo",
"Montepulciano d'Abruzzo Cerasuolo",
"Montepulciano d'Abruzzo Colline Teramane",
'Monteregio di Massa Marittima',
'Monterey',
'Monterey County',
'Monterey County-Napa County',
'Monterey County-Napa County-Sonoma County',
'Monterey-San Luis Obispo-Santa Barbara',
'Monterey-Santa Barbara',
'Monterey-Santa Barbara-Sonoma',
'Monterey-Santa Cruz',
'Monterrei',
'Montescudaio',
'Monthélie',
'Monti Lessini',
'Monticello',
'Montilla-Moriles',
'Montlouis-sur-Loire',
'Montrachet',
'Montravel',
'Montsant',
'Moon Mountain District Sonoma County',
'Morellino di Scansano',
'Morey-Saint-Denis',
'Morgon',
'Mornington Peninsula',
'Moscadello di Montalcino',
"Moscato d'Asti",
'Moscato di Noto',
'Moscato di Pantelleria',
'Moscato di Trani',
'Moulin-à-Vent',
'Moulis',
'Moulis-en-Médoc',
'Mount Barker',
'Mount Barker & Margaret River',
'Mount Harlan',
'Mount Veeder',
'Mt. Harlan',
'Mudgee',
'Murgia',
'Murray-Darling',
'Muscadet',
'Muscadet Coteaux de la Loire',
'Muscadet Côtes de Grandlieu',
'Muscadet Sèvre et Maine',
'Muscadet Sèvre et Maine Clisson',
'Muscat de Beaumes de Venise',
'Muscat de Rivesaltes',
'Muscat de Saint-Jean de Minervois',
'Musigny',
'Málaga',
'Mâcon',
'Mâcon Chaintré',
'Mâcon La Roche Vineuse',
'Mâcon Solutré',
'Mâcon-Azé',

'Mâcon-Chardonnay',
'Mâcon-Charnay',
'Mâcon-Cruzille',
'Mâcon-Davayé',
'Mâcon-Fuissé',
'Mâcon-Loché',
'Mâcon-Lugny',
'Mâcon-Mancey',
'Mâcon-Milly Lamartine',
'Mâcon-Pierreclos',
'Mâcon-Prissé',
'Mâcon-Péronne',
'Mâcon-Uchizy',
'Mâcon-Vergisson',
'Mâcon-Verze',
'Mâcon-Villages',
'Mâcon-Vinzelles',
'Médoc',
'Naches Heights',
'Nagambie Lakes',
'Napa',
'Napa County',
'Napa County-Lake County',
'Napa County-Sonoma County',
'Napa County-Sonoma County-Lake County',
'Napa County-Sonoma County-San Joaquin County',
'Napa Valley',
'Napa Valley-Paso Robles',
'Napa-Amador',
'Napa-Carneros',
'Napa-Mendocino-Sonoma',
'Napa-Mendocino-Sonoma-Marin',
'Napa-Monterey-Mendocino',
'Napa-Sonoma',
'Napa-Sonoma-Marin',
'Nardò',
'Nasco di Cagliari',
'Navarra',
"Nebbiolo d'Alba",
'Neuquén',
'Nevada',
'Nevada County',
'New Jersey',
'New Mexico',
'New South Wales',
'New York',
'Niagara Escarpment',
'Niagara Peninsula',
'Niagara-On-The-Lake',
'North Carolina',
'North Coast',
'North Fork of Long Island',
'Northern Sonoma',
'Noto',
'Nuits-St.-Georges',
'Nuragus di Cagliari',
'Oak Knoll District',
'Oakville',
'Offida Passerina',
'Offida Pecorino',
'Offida Rosso',
'Ohio',
'Okanagan Valley',
'Old Mission Peninsula',
'Oltrepò Pavese',
'Orange',
'Orcia',
'Oregon',
'Oregon Other',
'Orvieto',
'Orvieto Classico',
'Orvieto Classico Superiore',
'Outer Coastal Plain',
'Pacherenc du Vic Bilh',
'Padthaway',
'Paestum',
'Pago de Arínzano',
'Paicines',
'Palette',
'Paso Robles',

'Paso Robles Highlands District',
'Paso Robles Willow Creek District',
'Passito di Noto',
'Passito di Pantelleria',
'Patagonia',
'Patrimonio',
'Pauillac',
'Pavia',
"Pays d'Oc",
'Pays de Loire',
'Pedernal Valley',
'Pemberton',
'Penedès',
'Penisola Sorrentina',
'Pennsylvania',
'Perdriel',
'Pernand-Vergelesses',
'Pessac-Léognan',
'Petit Chablis',
'Piave',
'Piccadilly Valley',
'Picpoul de Pinet',
'Piedmont',
'Piemonte',
'Pine Mountain-Cloverdale Peak',
'Pla de Bages',
'Placer County',
'Polk County',
'Pomerol',
'Pomino',
'Pommard',
'Pompeiano',
'Potter Valley',
'Pouilly-Fuissé',
'Pouilly-Fumé',
'Pouilly-Loché',
'Pouilly-Vinzelles',
'Pouilly-sur-Loire',
'Premieres Côtes de Blaye',
'Premieres Côtes de Bordeaux',
'Primitivo del Salento',
'Primitivo del Tarantino',
'Primitivo di Manduria',
'Prince Edward County',
'Priorat',
'Prosecco',
'Prosecco Superiore di Cartizze',
'Prosecco Treviso',
'Prosecco del Veneto',
'Prosecco di Conegliano',
'Prosecco di Conegliano e Valdobbiadene',
'Prosecco di Valdobbiadene',
'Prosecco di Valdobbiadene Superiore',
'Provincia di Pavia',
'Puget Sound',
'Puglia',
'Puisseguin Saint-Émilion',
'Puligny-Montrachet',
'Pyrenees',
'Pécharmant',
'Périgord',
'Quarts de Chaume',
'Quincy',
'Ramandolo',
'Ramona Valley',
'Rasteau',
'Rattlesnake Hills',
'Recioto della Valpolicella',
'Recioto della Valpolicella Classico',
'Recioto di Soave',
'Recioto di Soave Classico',
'Red Hills Lake County',
'Red Mountain',
'Redwood Valley',
'Reuilly',
'Ribbon Ridge',
'Ribeira Sacra',
'Ribeiro',
'Ribera del Duero',
'Ribera del Guadiana',

```
'Ribera del Júcar',
'Rioja',
'River Junction',
'Riverina',
'Riverland',
'Rivesaltes',
'Rivesaltes Rosé',
'Riviera Ligure di Ponente',
'Roccamonfina',
'Rockpile',
'Roero',
'Roero Arneis',
'Rogue Valley',
'Romagna',
'Romanée-St.-Vivant',
'Rosazzo',
'Rosso Conero',
'Rosso Piceno',
'Rosso Piceno Superiore',
'Rosso del Veronese',
'Rosso di Cerignola',
'Rosso di Montalcino',
'Rosso di Montepulciano',
'Rosso di Toscana',
"Rosé d'Anjou",
'Rosé de Loire',
'Roussette de Savoie',
'Rubicone',
'Ruché di Castagnole Monferrato',
'Rueda',
'Rully',
'Russian River Valley',
'Rutherford',
'Rutherglen',
'Régnié',
'Rías Baixas',
'Río Negro Valley',
'Sable de Camargue',
'Saddle Rock-Malibu',
'Sagrantino di Montefalco',
'Saint Guilhem le Désert',
'Saint-Amour',
'Saint-Aubin',
'Saint-Bris',
'Saint-Chinian',
'Saint-Chinian-Roquebrun',
'Saint-Croix-du-Mont',
'Saint-Estèphe',
'Saint-Georges-Saint-Émilion',
'Saint-Joseph',
'Saint-Julien',
'Saint-Mont',
'Saint-Nicolas-de-Bourgueil',
'Saint-Péray',
'Saint-Romain',
'Saint-Véran',
'Saint-Émilion',
'Sainte-Foy Bordeaux',
'Salento',
'Salice Salentino',
'Salina',
'Salta',
'San Antonio Valley',
'San Benito County',
'San Bernabe',
'San Carlos',
'San Diego County',
'San Francisco Bay',
'San Francisco Bay-Livermore Valley',
'San Gimignano',
'San Juan',
'San Lucas',
'San Luis Obispo',
'San Luis Obispo County',
'San Marino',
'San Martino della Battaglia',
'San Rafael',
'Sancerre',
'Sangiovese di Romagna',
'Sangiovese di Romagna Superiore',
```

```
     'Sannio',
     "Sant' Agata dei Goti",
     "Sant'Antimo",
     'Santa Barbara',
     'Santa Barbara County',
     'Santa Barbara-Monterey',
     'Santa Clara County',
     'Santa Clara Valley',
     'Santa Cruz County',
     'Santa Cruz Mountains',
     'Santa Lucia Highlands',
     'Santa Margarita Ranch',
     'Santa Maria Valley',
     'Santa Ynez Valley',
     'Santenay',
     'Sardinia',
     'Sardon de Duero',
     'Saumur',
     'Saumur-Champigny',
     'Saussignac',
     'Sauternes',
     'Savennières',
     'Savennières-Roche-Aux-Moines',
     'Savigny-lès-Beaune',
     'Savoie',
     'Sebino',
     'Seneca Lake',
     'Sforzato di Valtellina',
     'Shenandoah Valley',
     'Shenandoah Valley (CA)',
     'Short Hills Bench',
     'Sicilia',
     'Sierra Foothills',
     'Sierra Pelona Valley',
     'Sierra de Salamanca',
     'Similkameen Valley',
     'Siracusa',
     'Snipes Mountain',
     'Soave',
     'Soave Classico',
     ...]
```

In [118…
```python
print(len(list(df['region_1'].dropna().unique())),"unique values")
print(df["region_1"].isnull().sum(),"empty values")
```

```
1229 unique values
3719 empty values
```

In [115…
```python
sorted(list(df['region_2'].dropna().unique()))
```

Out[115]:
```
['California Other',
 'Central Coast',
 'Central Valley',
 'Columbia Valley',
 'Finger Lakes',
 'Long Island',
 'Napa',
 'Napa-Sonoma',
 'New York Other',
 'North Coast',
 'Oregon Other',
 'Sierra Foothills',
 'Sonoma',
 'South Coast',
 'Southern Oregon',
 'Washington Other',
 'Willamette Valley']
```

Conclusion: There are too many regions to manually check each possible problematic pair/trio using Google and Google Maps. I'd just recommend not using these columns since theyre so incomplete anyway.

Write result to csv file

In [119…
```python
df.to_csv("intermediate_data/wineries_cleaned.csv")
```