



Анализ данных по теме «Wine Reviews»

ПОДГОТОВИЛА СМЕШКОВА ЕКАТЕРИНА. 10ИНФУ6

Dataset

Исследование проводилось

над

датасетом, описывающим

сорта различных вин.

Таблица состоит из

150930 строк и

10 столбцов

Ссылка:

https://www.kaggle.com/zynicide/wine-reviews?select=winemag-data_first150k.csv

	country	description	designation	points	price	province	region_1	region_2	variety	winery
0	US	This tremendous 100% varietal wine hails from ...	Martha's Vineyard	96	235.0	California	Napa Valley	Napa	Cabernet Sauvignon	Heitz
1	Spain	Ripe aromas of fig, blackberry and cassis are ...	Carodorum Selección Especial Reserva	96	110.0	Northern Spain	Toro	NaN	Tinta de Toro	Bodega Carmen Rodríguez
2	US	Mac Watson honors the memory of a wine once ma...	Special Selected Late Harvest	96	90.0	California	Knights Valley	Sonoma	Sauvignon Blanc	Macauley
3	US	This spent 20 months in 30% new French oak, an...	Reserve	96	65.0	Oregon	Willamette Valley	Willamette Valley	Pinot Noir	Ponzi
4	France	This is the top wine from La Bégude, named aft...	La Brûlade	95	66.0	Provence	Bandol	NaN	Provence red blend	Domaine de la Bégude
...
150925	Italy	Many people feel Fiano represents southern Ita...	NaN	91	20.0	Southern Italy	Fiano di Avellino	NaN	White Blend	Feudi di San Gregorio
150926	France	Offers an intriguing nose with ginger, lime an...	Cuvée Prestige	91	27.0	Champagne	Champagne	NaN	Champagne Blend	H.Germain
150927	Italy	This classic example comes from a cru vineyard...	Terre di Dora	91	20.0	Southern Italy	Fiano di Avellino	NaN	White Blend	Terredora
150928	France	A perfect salmon shade, with scents of peaches...	Grand Brut Rosé	90	52.0	Champagne	Champagne	NaN	Champagne Blend	Gosset
150929	Italy	More Pinot Grigios should taste like this. A r...	NaN	90	15.0	Northeastern Italy	Alto Adige	NaN	Pinot Grigio	Alois Lageder

Описание набора данных

- COUNTRY (КАТЕГОРИАЛЬНАЯ НОМИНАЛЬНАЯ ПЕРЕМЕННАЯ) – СТРАНА, ИЗ КОТОРОЙ РОДОМ ВИНО
- DESCRIPTION (ОПИСАНИЕ) - НЕСКОЛЬКО ФРАЗ ОТ СОМЕЛЬЕ, ОПИСЫВАЮЩИЕ ВКУС, ЗАПАХ, ВНЕШНИЙ ВИД ВИНА И Т. Д.
- DESIGNATION (КАТЕГОРИАЛЬНАЯ НОМИНАЛЬНАЯ ПЕРЕМЕННАЯ) – НАЗВАНИЕ ВИНОГРАДНИКА
- POINTS (КОЛИЧЕСТВЕННАЯ ДИСКРЕТНАЯ ПЕРЕМЕННАЯ) - КОЛИЧЕСТВО БАЛЛОВ ВИНА ПО ШКАЛЕ 1-100 (WINEENTHUSIAST)
- PRICE (КОЛИЧЕСТВЕННАЯ ДИСКРЕТНАЯ ПЕРЕМЕННАЯ) – ЦЕНА ЗА БУТЫЛКУ ВИНА (\$)
- PROVINCE (КАТЕГОРИАЛЬНАЯ НОМИНАЛЬНАЯ ПЕРЕМЕННАЯ) – ПРОВИНЦИЯ ИЛИ ШТАТ, В КОТОРОМ ПРОИЗВОДИТСЯ ВИНО
- REGION_1 (КАТЕГОРИАЛЬНАЯ НОМИНАЛЬНАЯ ПЕРЕМЕННАЯ) - ВИНОДЕЛЬЧЕСКИЙ РАЙОН
- REGION_2 (КАТЕГОРИАЛЬНАЯ НОМИНАЛЬНАЯ ПЕРЕМЕННАЯ) - БОЛЕЕ КОНКРЕТНЫЕ РЕГИОНЫ, УКАЗАННЫЕ В ПРЕДЕЛАХ ВИНОДЕЛЬЧЕСКОЙ ЗОНЫ
- VARIETY (КАТЕГОРИАЛЬНАЯ НОМИНАЛЬНАЯ ПЕРЕМЕННАЯ) - СОРТ ВИНОГРАДА, ИСПОЛЬЗУЕМОГО ДЛЯ ИЗГОТОВЛЕНИЯ ВИНА
- WINERY (КАТЕГОРИАЛЬНАЯ НОМИНАЛЬНАЯ ПЕРЕМЕННАЯ) – НАЗВАНИЕ ВИНОДЕЛЬНИ

ОПИСАНИЕ СТОЛБЦОВ

1	data.dtypes	
	country	object
	description	object
	designation	object
	points	int64
	price	float64
	province	object
	region_1	object
	region_2	object
	variety	object
	winery	object

Типы данных в столбцах

1	data.isnull().sum()	
	country	5
	description	0
	designation	45735
	points	0
	price	13695
	province	5
	region_1	25060
	region_2	89977
	variety	0
	winery	0

Количество пустых значений в столбцах

Сортировка

Для того, чтобы нам было удобнее работать с данными, отсортируем вина от лучших к худшим (согласно рейтингу WineEnthusiast), а внутри этих значений также произведем сортировку по алфавиту для страны, в которой производится вино

	country	description	designation	points	price	province	region_1	region_2	variety	winery
19355	Australia	Not a Cellar Selection in the traditional sens...	Rare	100	300.0	Victoria	Rutherglen	NaN	Muscat	Chambers Rosewood Vineyards
84035	Australia	Not a Cellar Selection in the traditional sens...	Rare	100	300.0	Victoria	Rutherglen	NaN	Muscat	Chambers Rosewood Vineyards
119195	Australia	Not a Cellar Selection in the traditional sens...	Rare	100	300.0	Victoria	Rutherglen	NaN	Muscat	Chambers Rosewood Vineyards
2145	France	Full of ripe fruit, opulent and concentrated, ...	NaN	100	848.0	Bordeaux	Pessac-Léognan	NaN	Bordeaux-style White Blend	Château Haut-Brion
26296	France	A wine that has created its own universe. It h...	Clos du Mesnil	100	1400.0	Champagne	Champagne	NaN	Chardonnay	Krug
...
60260	Uruguay	Peanut and leather aromas, with a candied swee...	NaN	80	7.0	Canelones	NaN	NaN	Cabernet Sauvignon	Ariano
60257	Uruguay	Dark in color, with black plum along with damp...	NaN	80	9.0	Canelones	NaN	NaN	Cabernet Sauvignon	Toscanini
119481	Uruguay	Not very lively or fruity. Vague aromas and fl...	Identity Tannat	80	15.0	Canelones	NaN	NaN	Rosé	Gimenez Mendez
94521	Uruguay	Not very lively or fruity. Vague aromas and fl...	Identity Tannat	80	15.0	Canelones	NaN	NaN	Rosé	Gimenez Mendez
42381	Uruguay	Not very lively or fruity. Vague aromas and fl...	Identity Tannat	80	15.0	Canelones	NaN	NaN	Rosé	Gimenez Mendez

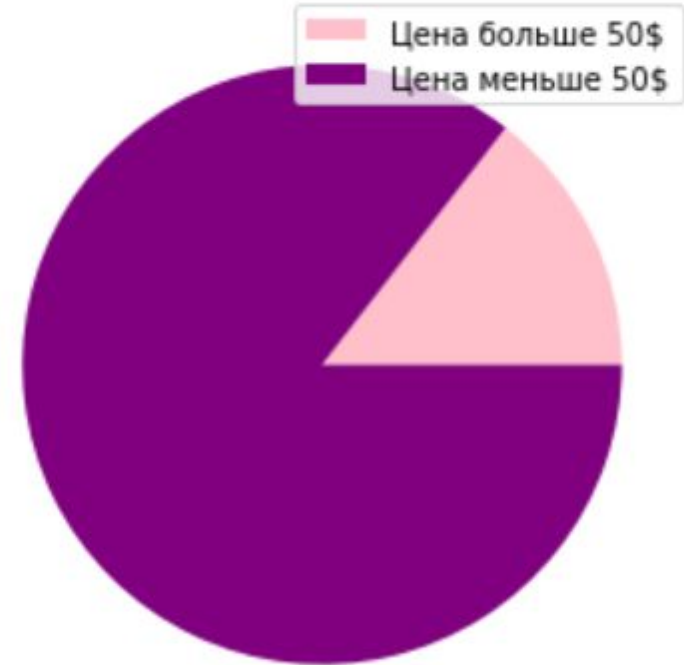
Рассмотрим цены за бутылку вина и проверим, преобладает ли количество сортов вин, цена которых больше 50\$

На диаграмме видно, что у большинства рассмотренных сортов вин цена за бутылку менее 50\$

Создадим дополнительный столбец, в котором «1» будет соответствовать вину, цена которого больше 50\$, а «0» – меньше

Посчитаем точные значения

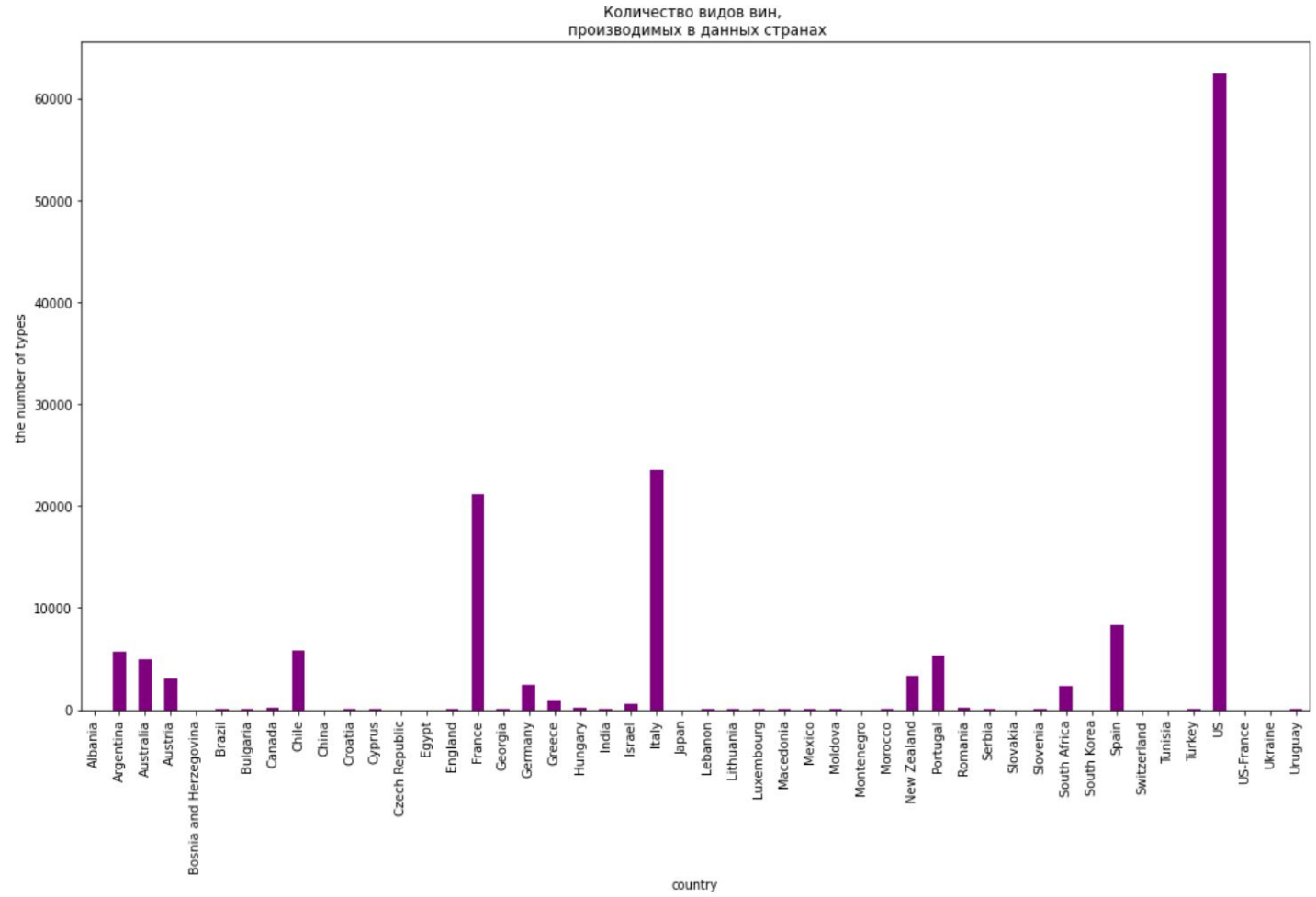
Prices



```
1 data['more_than_50'].value_counts()
0    128098
1     22832
```

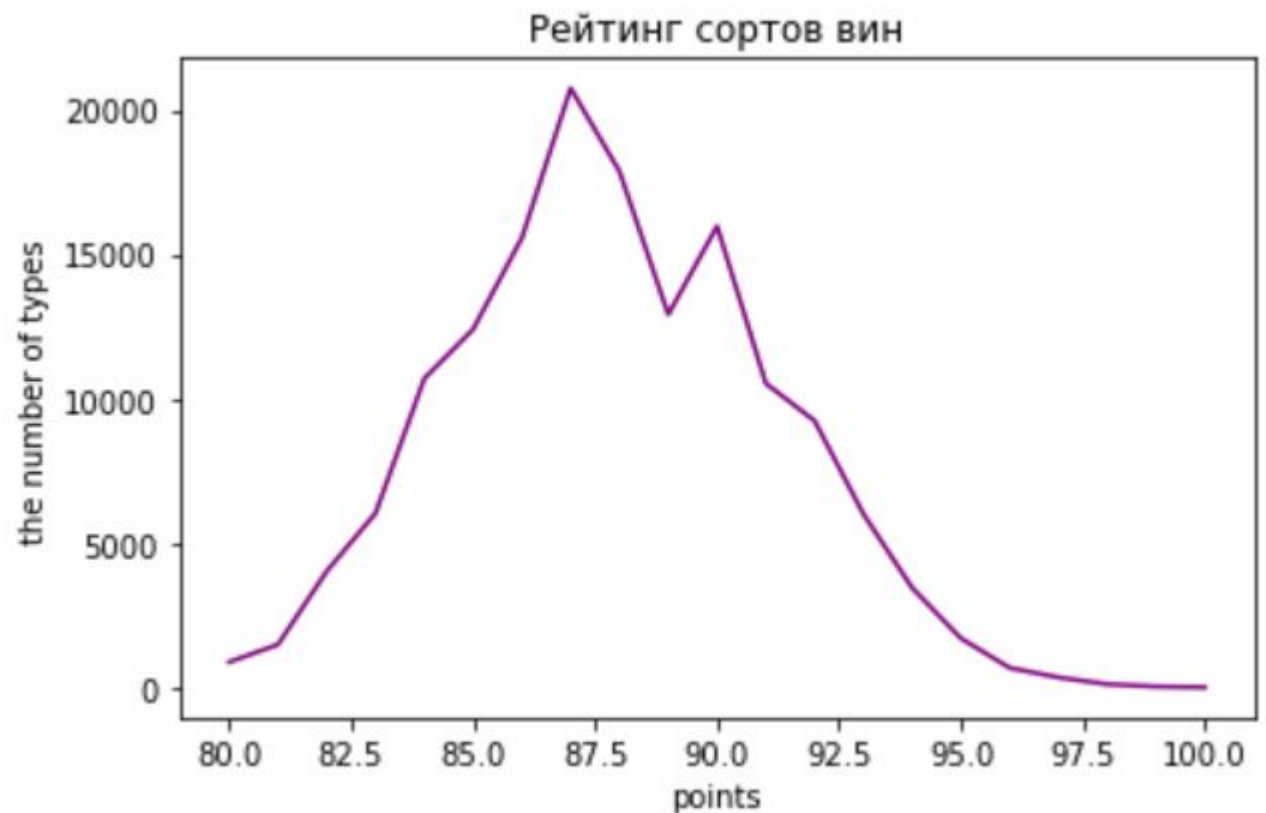

Построим гистограмму для того, чтобы проанализировать, в каких странах производится наибольшее количество вин

В США производится наибольшее количество вин. Следом идет Италия, и Франция практически наравне с ней



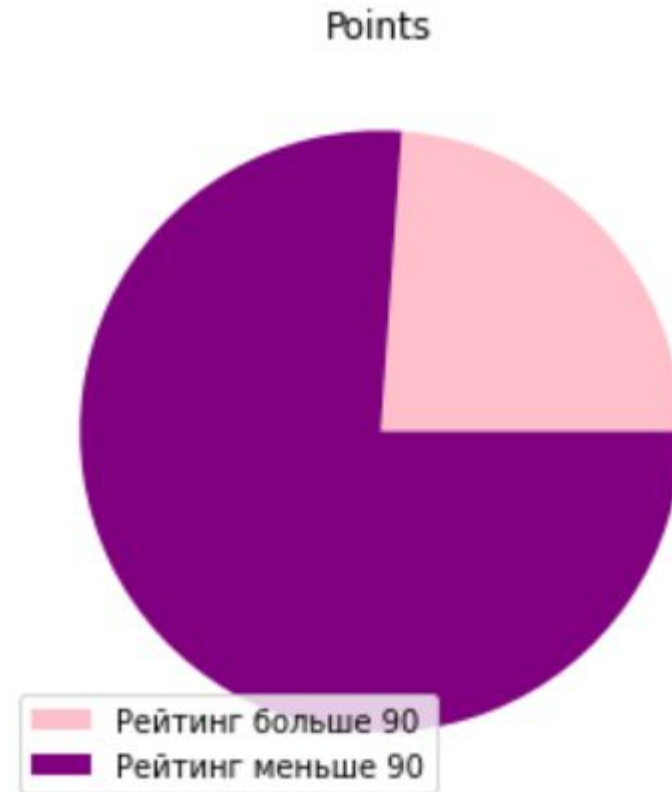
Построим график,
чтобы определить,
какой рейтинг сортов
вин преобладает в
нашем датасете

У большинства вин
рейтинг WineEnthusiast
около 87.5



Построим круговую диаграмму для того, чтобы определить, какой вид вин преобладает в нашем датасете: с рейтингом более или менее 90 баллов

У большинства вин рейтинг WineEnthusiast меньше 90 баллов



Создадим сводную таблицу, в которой будет отражено, сколько сортов вин производится в каждой из стран

Наименьшее количество вин создается в Албании, Китае, Египте, Японии, Черногории и Тунисе

country	
Albania	2
Argentina	5631
Australia	4957
Austria	3057
Bosnia and Herzegovina	4
Brazil	25
Bulgaria	77
Canada	196
Chile	5816
China	3
Croatia	89
Cyprus	31
Czech Republic	6
Egypt	3
England	9
France	21098
Georgia	43
Germany	2452
Greece	884
Hungary	231

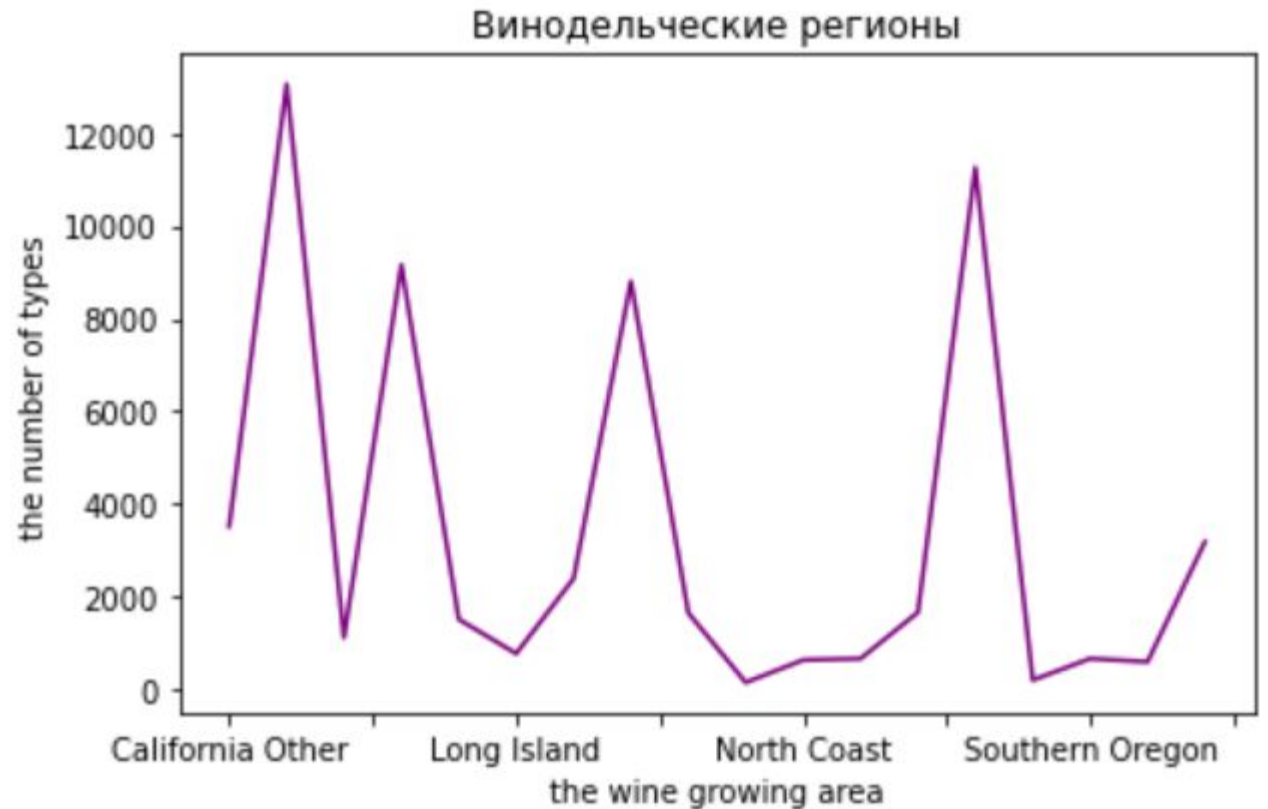
В данной сводной таблице отразим количество видов вин, создаваемых в винодельне того или иного винодельческого района

Наибольшее количество вин производится в винодельнях D'Arenberg и Longview

region_1	winery	
Abruzzo	Canaletto	2
	Cataldi Madonna	2
	Farnese	5
	Feudo Antico	5
	Adelaida District	Adelaida
Adelaida District	Adelaida Cellars	3
	Bella Luna	1
	Brecon Estate	1
	Chateau Lettau	1
	Daou	2
	Halter Ranch	2
	Lone Madrone	3
	Peachy Canyon	4
	Pulchella	1
	Summerwood	1
	kukkula	5
	Adelaide	D'Arenberg
Groom		3
Hewitson		3
Longview		6
Nepenthe		7
	Nova Vita	5

В данном графике отразим, сколько сортов вин производится в том или ином винодельческом регионе (стоит отметить, что здесь рассматривались данные из region_2, т. е. более конкретные места)

В Калифорнии и Южном Орегоне производится наибольшее количество вин



Создадим сводную таблицу, в которой будет отражено количество видов вин определенной цены, производимых из того или иного сорта винограда

У большинства сортов существует только один вид вина, соответствующий той или иной цене

price	variety	
4.0	Cabernet Sauvignon	1
	Chardonnay	1
	Malbec	1
	Malbec-Syrah	1
	Merlot	1
	Pinot Grigio	1
	Portuguese Red	1
	Tempranillo	1
	Zinfandel	1
5.0	Cabernet Sauvignon	1
	Chardonnay	1
	Chenin Blanc	1
	Lambrusco	1
	Malbec	1
	Malbec-Bonarda	1
	Merlot	1
	Merlot-Malbec	1
	Moscato	1
	Pinot Grigio	1
	Pinot Noir	1

ВЫВОДЫ

- У БОЛЬШИНСТВА ВИН ЦЕНА ЗА БУТЫЛКУ МЕНЕЕ 50\$
- В США ПРОИЗВОДИТСЯ НАИБОЛЬШЕЕ КОЛИЧЕСТВО ВИДОВ ВИН. СЛЕДОМ ИДЕТ ИТАЛИЯ, И ФРАНЦИЯ ПРАКТИЧЕСКИ НАРАВНЕ С НЕЙ
- У БОЛЬШИНСТВА ВИН РЕЙТИНГ WINEENTHUSIAST МЕНЬШЕ 90 БАЛЛОВ
- НАИМЕНЬШЕЕ КОЛИЧЕСТВО ВИН СОЗДАЕТСЯ В АЛБАНИИ, КИТАЕ, ЕГИПТЕ, ЯПОНИИ, ЧЕРНОГОРИИ И В ТУНИСЕ
- НАИБОЛЬШЕЕ КОЛИЧЕСТВО ВИН ПРОИЗВОДИТСЯ В ВИНОДЕЛЬНЯХ D'ARENBERG И LONGVIEW
- В КАЛИФОРНИИ И ЮЖНОМ ОРЕГОНЕ ПРОИЗВОДИТСЯ НАИБОЛЬШЕЕ КОЛИЧЕСТВО ВИН
- У БОЛЬШИНСТВА СОРТОВ СУЩЕСТВУЕТ ТОЛЬКО ОДИН СОРТ ВИНА, СООТВЕТСТВУЮЩИЙ ТОЙ ИЛИ ИНОЙ ЦЕНЕ
- У БОЛЬШИНСТВА ВИН РЕЙТИНГ WINEENTHUSIAST ОКОЛО 87.5



Спасибо за внимание!