

A dark blue vertical bar on the left side of the page. A blue arrow points to the right from the bar, containing the date.

23-11-2024

Análise de Dados com Vinhos

Several thin, curved lines in shades of blue and grey, resembling stylized grass or reeds, located in the bottom left corner.

Felipe Pacheco
WINE STUDIES

Dedicatória

Dedico este trabalho a todos que, direta ou indiretamente, contribuíram para o meu crescimento pessoal e profissional. Aos meus familiares, que sempre me apoiaram e incentivaram a buscar meus objetivos com dedicação e paixão.

Agradeço também ao Professor Luís Tatin, que compartilhou seus conhecimentos e me ajudaram a expandir meus horizontes, especialmente no campo da ciência de dados.



Sumário

Introdução	3
Países com maior número de vinhos	4
Média de pontos dos vinhos por país, região e variedade.....	5
Correlação entre Preço e Pontuação	6
Vinhos com maior custo-benefício	7
Vinhos mais caros e bem avaliados.....	8
Vinhos com descrições mais longas	9
Variedade mais comum e variação de preços.....	10
Distribuição de Preços por Pontuação	11
Relação entre Idade do Vinho e Pontuação.....	12
Distribuição dos Preços entre Vinhos Orgânicos e Não Orgânicos	13
5 Países com os Vinhos Mais Bem Avaliados.....	14
Comparação de Pontuações dos Vinhos entre Degustadores	15
Média de Pontuação dos Vinhos com Pontuação Acima de 90 por Intervalo de Preço	16
10 Tipos de Uva com Menor Média de Pontuação	17
10 Variedades de Uva nos Vinhos de Alta Qualidade.....	18
Média de Pontos por Número de Avaliações	19
Províncias com Menor Preço Médio	20
Referências Bibliográficas	21

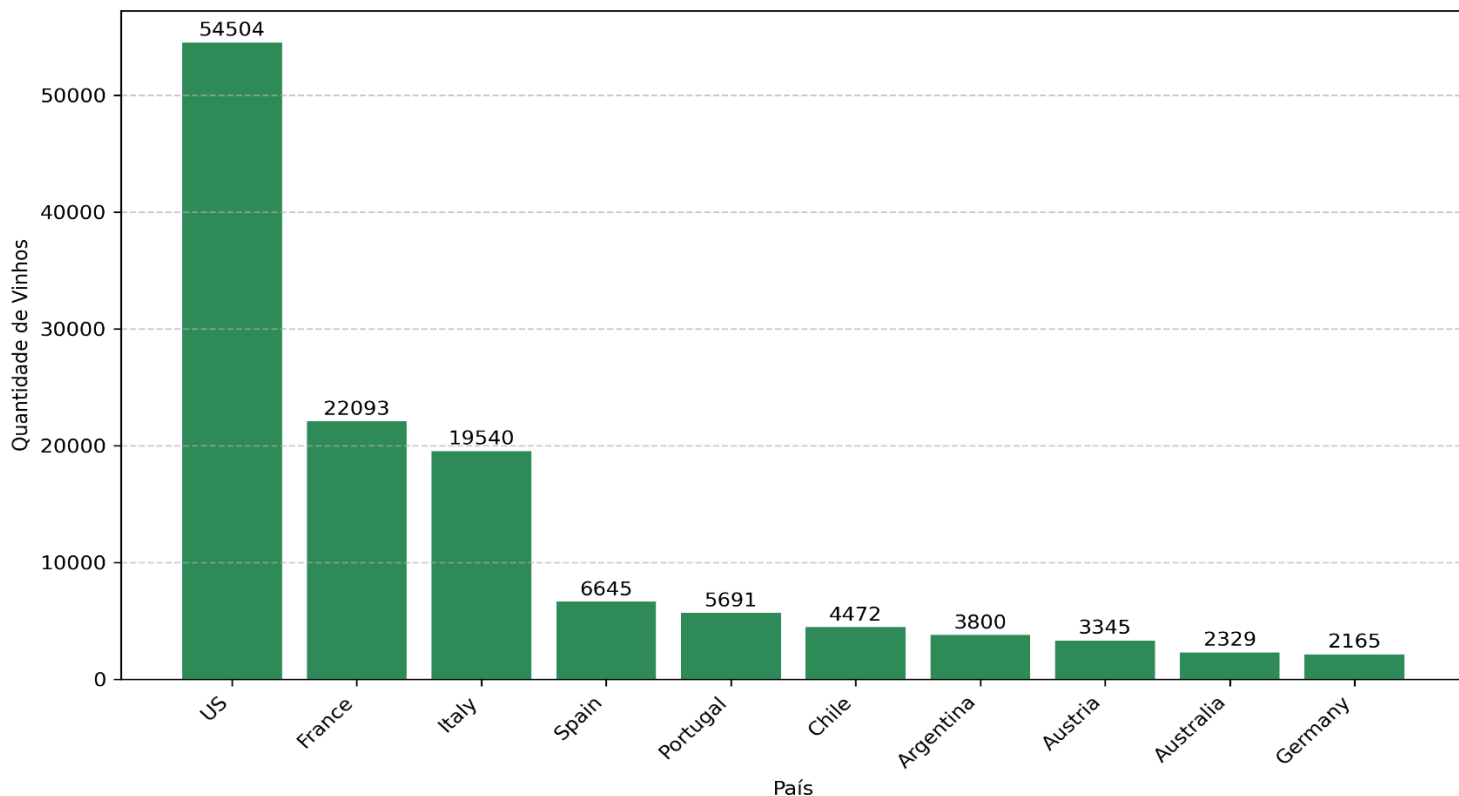
Introdução

A ciência de dados tem se consolidado como uma ferramenta essencial para a análise e interpretação de grandes volumes de informações em diversas áreas, e o mundo dos vinhos não é exceção. Com a crescente quantidade de dados disponíveis sobre vinhos, como suas características sensoriais, processos de produção e preferências dos consumidores, surge uma oportunidade única de aplicar técnicas de análise preditiva e modelagem de dados para descobrir padrões, tendências e melhorar a qualidade do produto.

Este trabalho busca explorar o uso de Python e técnicas de Data Science para analisar dados relacionados a vinhos, com o objetivo de compreender melhor os fatores que influenciam suas características e a forma como são percebidos pelos consumidores. Através de modelos estatísticos e algoritmos, procuramos identificar relações entre variáveis, prever a qualidade de vinhos e oferecer insights valiosos para produtores e entusiastas.

Ao longo deste estudo, abordaremos a preparação dos dados, análise exploratória e técnicas de visualização. O intuito é demonstrar como a ciência de dados pode ser uma aliada poderosa no processo de aprimoramento da produção e apreciação de vinhos, proporcionando resultados que vão além do senso comum.

Países com maior número de vinhos



Este gráfico permite observar a predominância de vinhos originários de alguns países, como os Estados Unidos, França e Itália, enquanto países como Alemanha e Austrália possuem números consideravelmente menores de vinhos registrados. A distribuição desigual pode ser um reflexo da produção de vinhos em larga escala nesses países ou de uma maior presença no mercado de vinhos global.

Média de pontos dos vinhos por país, região e variedade

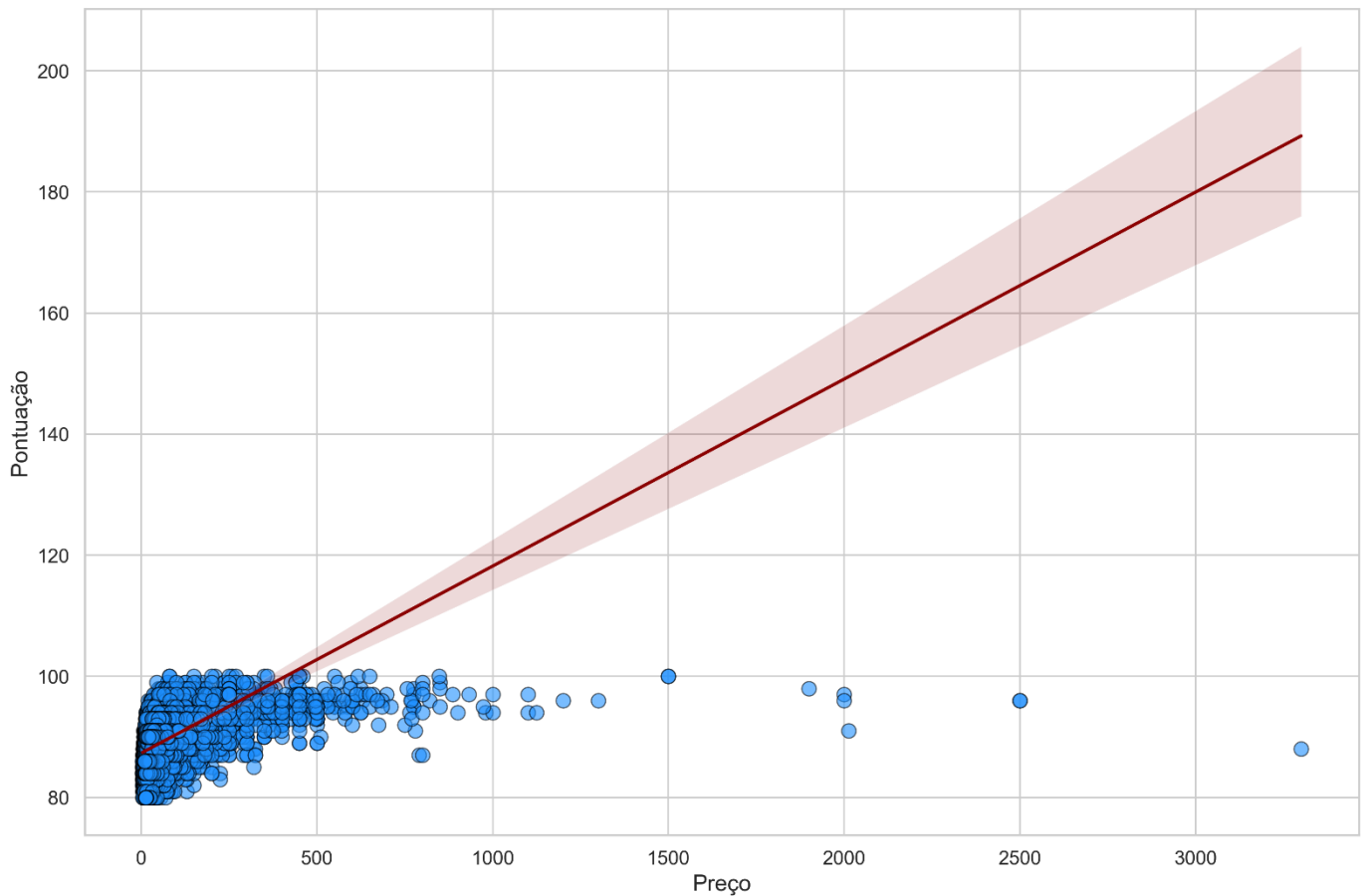
Info Vinho	Media Pontos
Italy - Vin Santo di Montepulciano - Prugnolo Gentile	99.5
Italy - Bolgheri Sassicaia - Red Blend	96.6
France - Bienvenues Bâtard-Montrachet - Chardonnay	96.5
France - Corton-Pougets - Pinot Noir	96.5
US - Napa-Monterey-Mendocino - Sparkling Blend	96.0
France - Clos de Lambrays - Pinot Noir	96.0
Australia - McLaren Vale - Cabernet-Shiraz	96.0
US - Walla Walla Valley (OR) - Cabernet Sauvignon-Merlot	96.0
US - Alexander Valley - Sémillon	96.0
France - Montrachet - Chardonnay	96.0

Os vinhos com as pontuações mais altas são destacados, refletindo a excelência dos mesmos em termos de qualidade, e ajudam a identificar quais regiões, países e variedades estão se destacando no mercado. A tabela serve como um guia para os consumidores e profissionais da área, oferecendo uma curadoria dos vinhos mais bem avaliados.

O país mais presente na tabela é a Itália, com duas entradas no topo da lista.

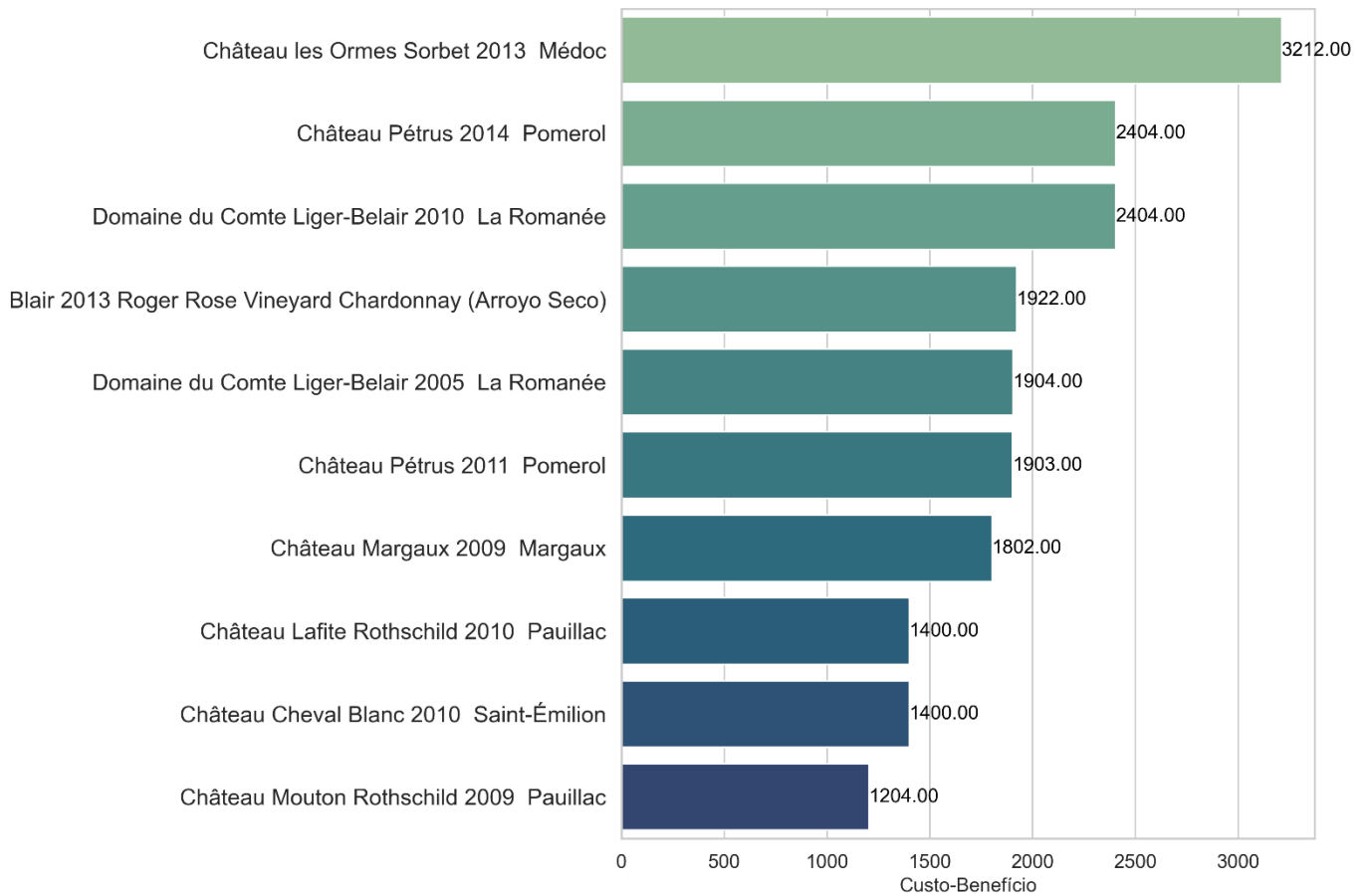
A variedade de uva também desempenha um papel importante, com tipos como Chardonnay e Cabernet Sauvignon aparecendo entre as variedades mais pontuadas.

Correlação entre Preço e Pontuação



A correlação de 0.416 indica uma relação moderada positiva entre o preço e a pontuação dos vinhos, mas há uma dispersão significativa nos dados, o que implica que outros fatores além do preço influenciam a pontuação. Ou seja, à medida que o preço aumenta, a pontuação também tende a aumentar, embora não seja uma relação perfeita.

Vinhos com maior custo-benefício



Este gráfico apresenta os vinhos com maior custo-benefício, considerando a diferença entre o preço e os pontos atribuídos a cada vinho. A métrica de custo-benefício foi calculada subtraindo os pontos de cada vinho do seu preço, com o intuito de identificar os vinhos que oferecem a melhor relação entre qualidade (representada pelos pontos) e custo (representado pelo preço).

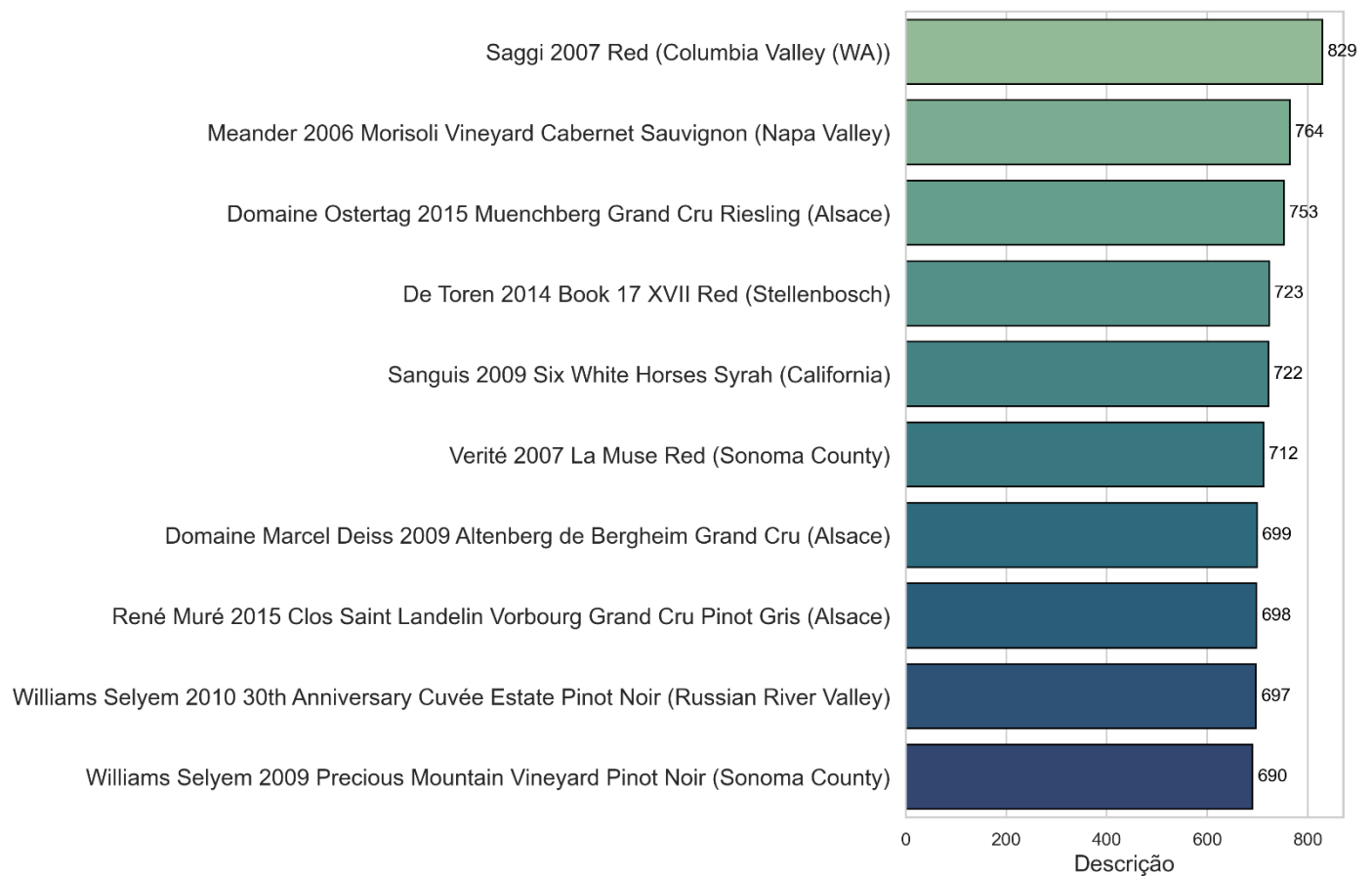
Vinhos mais caros e bem avaliados

Vinho	Preço	Pontos
Château les Ormes Sorbet 2013 Médoc	3300.0	88
Château Pétrus 2014 Pomerol	2500.0	96
Domaine du Comte Liger-Belair 2010 La Romanée	2500.0	96
Blair 2013 Roger Rose Vineyard Chardonnay	2013.0	91
Domaine du Comte Liger-Belair 2005 La Romanée	2000.0	96
Château Pétrus 2011 Pomerol	2000.0	97
Château Margaux 2009 Margaux	1900.0	98
Château Lafite Rothschild 2010 Pauillac	1500.0	100
Château Cheval Blanc 2010 Saint-Émilion	1500.0	100
Château Mouton Rothschild 2009 Pauillac	1300.0	96

A tabela apresenta uma análise dos 10 vinhos mais caros e bem avaliados.

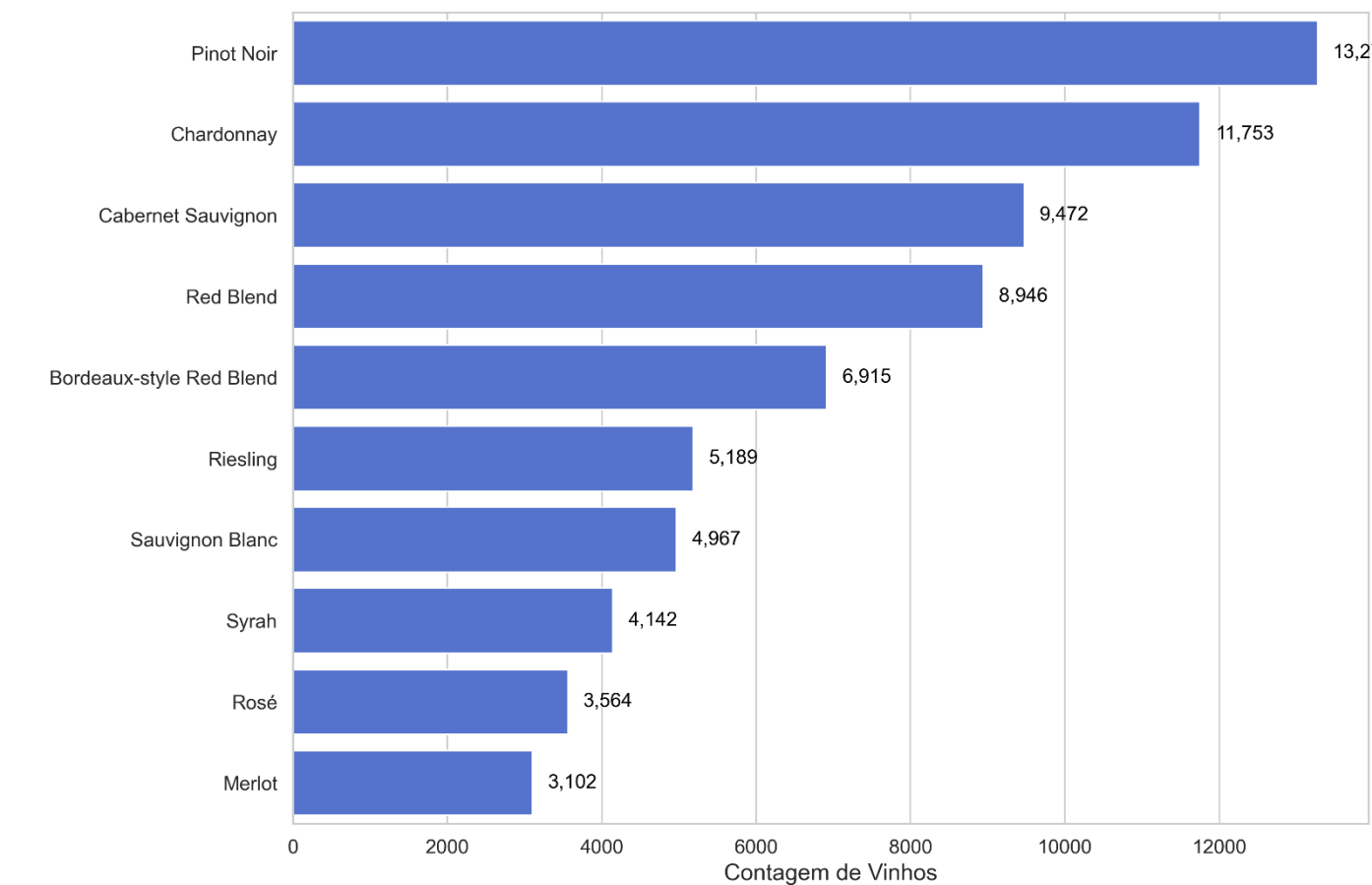
Os vinhos listados representam o ápice de qualidade, sendo provenientes de regiões renomadas como Bordeaux (França) e Romanée (França), bem como exemplares destacados de Califórnia (EUA). A tabela evidencia a relação entre o preço elevado e as altas avaliações, mostrando a predominância de vinhos franceses em quesitos de excelência e prestígio.

Vinhos com descrições mais longas



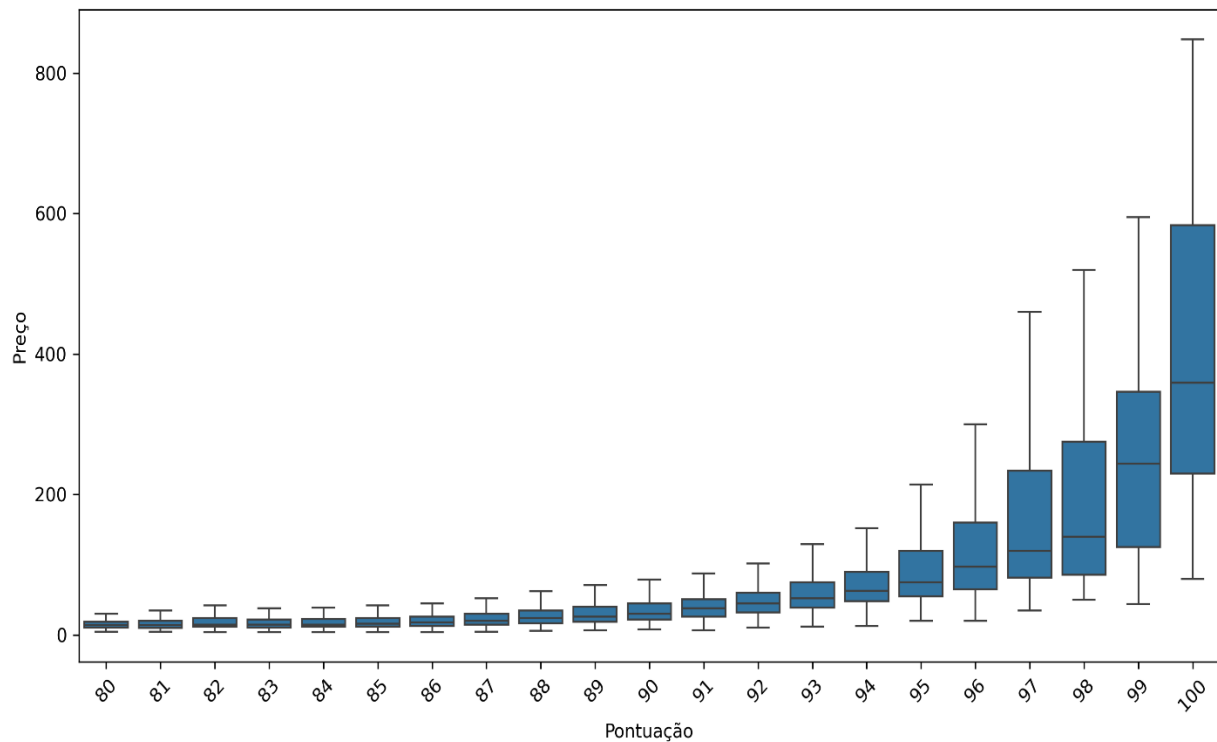
Esse gráfico destaca os vinhos cujas descrições são mais detalhadas, sendo o vinho "Saggi 2007 Red (Columbia Valley (WA))" o que possui a descrição mais longa (829 caracteres). Os valores variam de 690 a 829 caracteres, mostrando a diversidade na extensão das descrições fornecidas.

Variedade mais comum e variação de preços



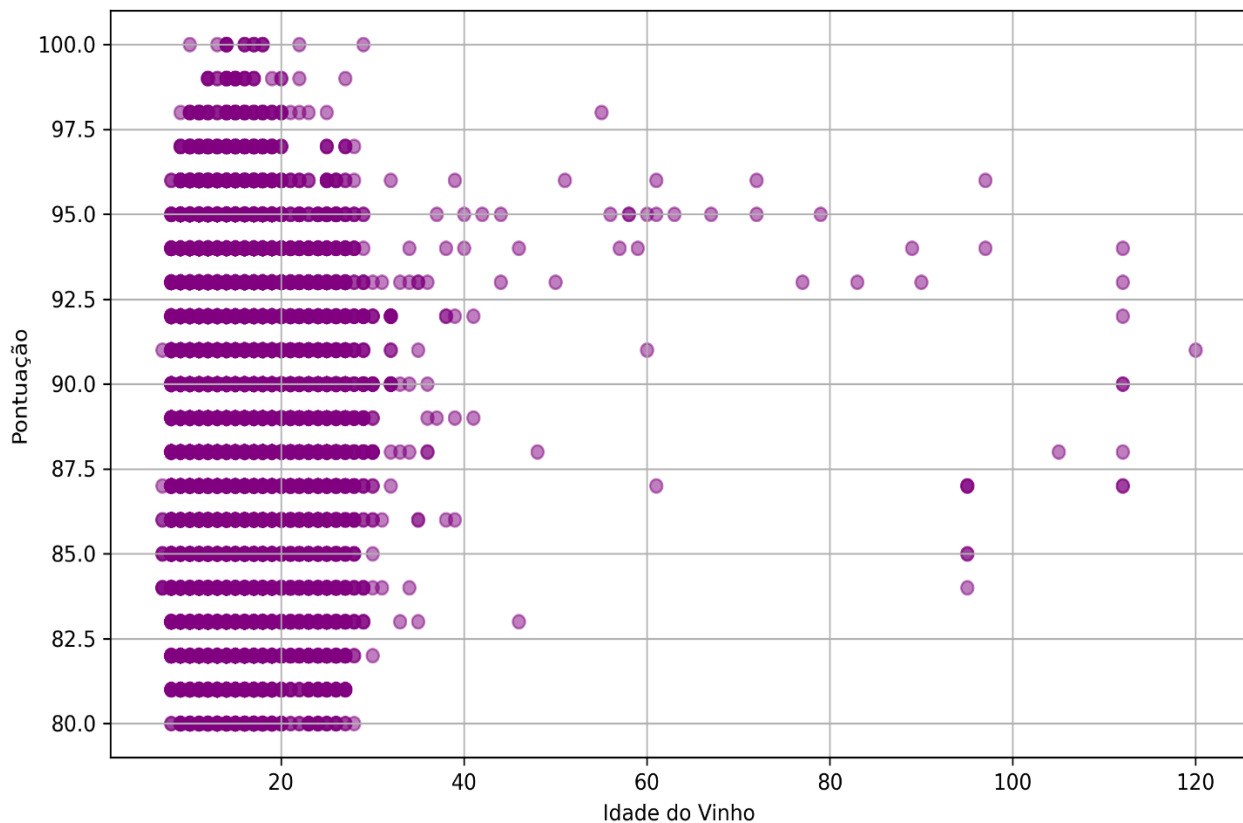
Este gráfico proporciona uma visão clara da popularidade relativa de cada variedade de vinho, ajudando na análise e comparação de quantidades entre elas.

Distribuição de Preços por Pontuação



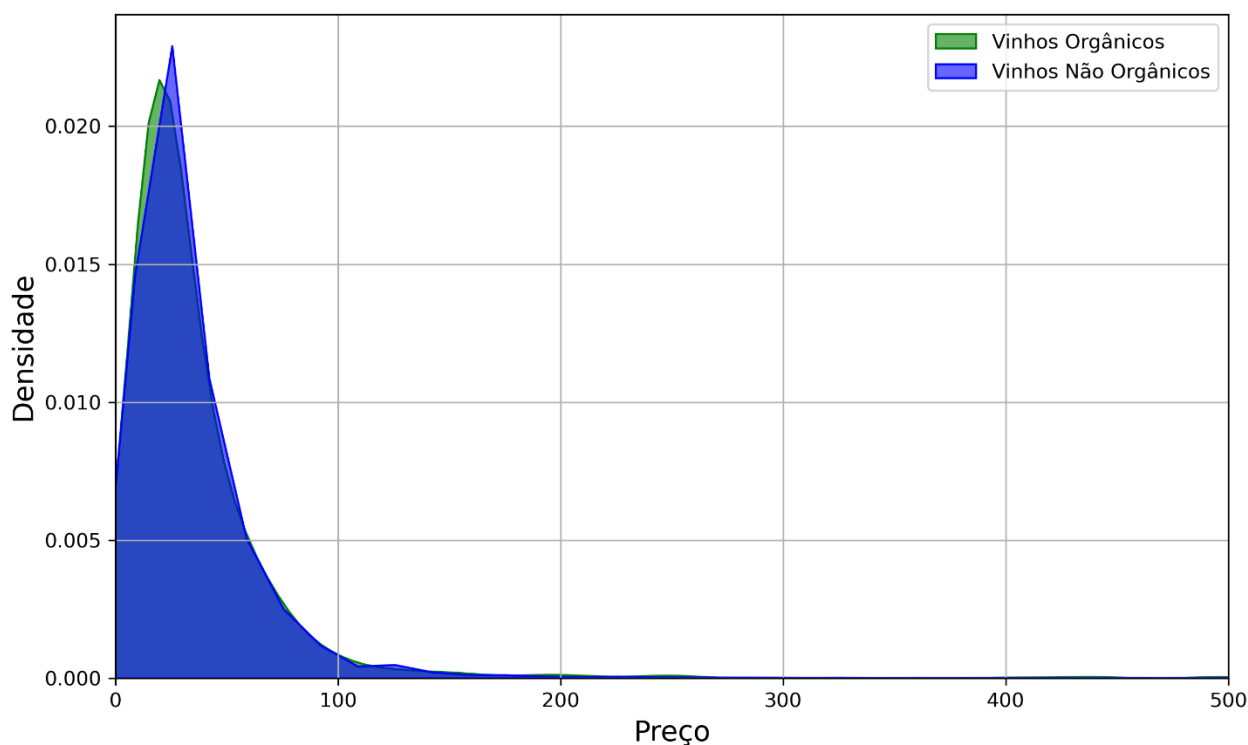
Com este gráfico, percebemos que vinhos com pontuações mais altas geralmente têm preços mais elevados. Isso faz sentido, pois vinhos melhores tendem a custar mais. No entanto, há uma sobreposição em algumas pontuações intermediárias, indicando que vinhos com preços variados podem receber pontuações semelhantes.

Relação entre Idade do Vinho e Pontuação



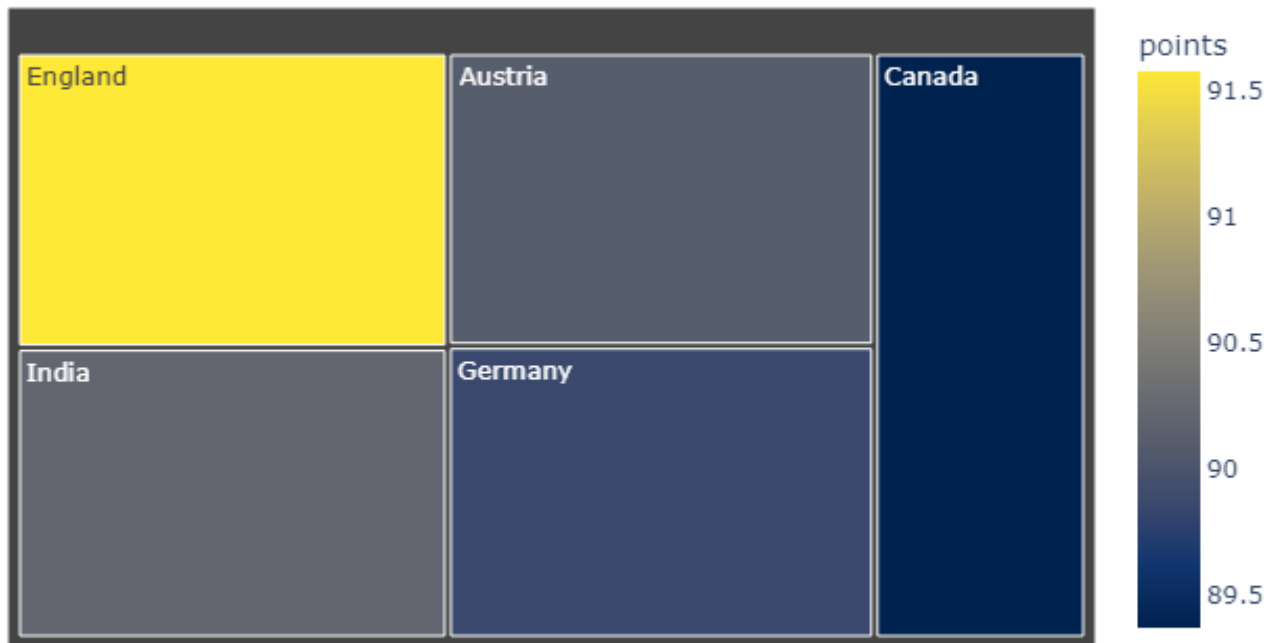
Neste gráfico, estamos analisando a relação entre a idade do vinho (quanto tempo se passou desde sua safra) e a pontuação que ele recebeu. A idade do vinho é calculada subtraindo o ano da safra do ano atual. A pontuação é a avaliação do vinho com base na sua qualidade, como determinada pelos críticos.

Distribuição dos Preços entre Vinhos Orgânicos e Não Orgânicos



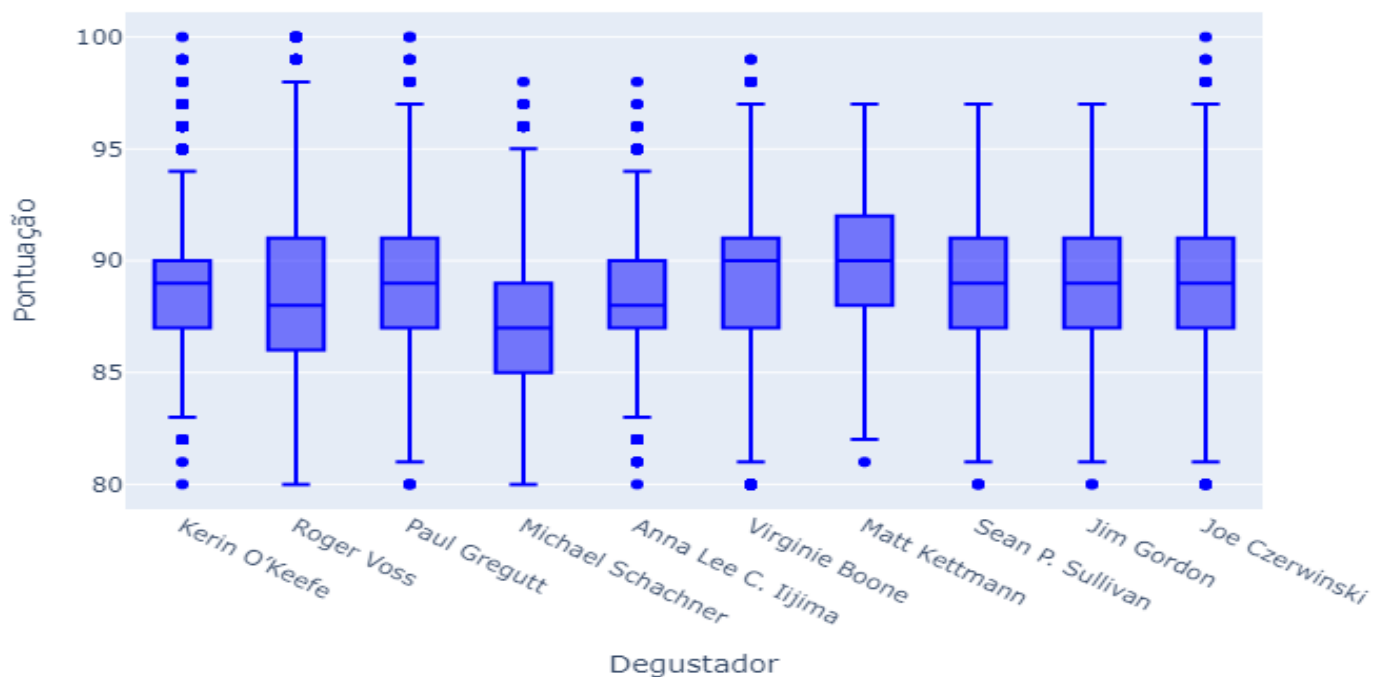
O gráfico que estamos visualizando é um gráfico de densidade (KDE), onde comparamos as distribuições de preços entre vinhos orgânicos e não orgânicos. A principal característica desse gráfico é que ele nos permite entender de forma suave e clara como os preços dos vinhos estão distribuídos para esses dois grupos.

5 Países com os Vinhos Mais Bem Avaliados



Esse tipo de gráfico é bastante útil para mostrar comparações rápidas entre categorias de dados, como, neste caso, os países produtores de vinho. Ele destaca tanto a diferença em termos de média de pontuação, como também visualiza a proporção de cada país dentro do grupo.

Comparação de Pontuações dos Vinhos entre Degustadores



O gráfico fornece uma comparação direta entre os degustadores com relação às pontuações dos vinhos que eles avaliaram. Podemos observar se algum degustador tende a dar pontuações mais altas ou mais baixas, ou se a distribuição das pontuações é mais ou menos consistente entre os degustadores.

Outliers podem ser identificados facilmente, permitindo perceber se algum degustador tem uma tendência a avaliar vinhos de maneira muito diferente dos outros

Esse gráfico é útil para entender a variabilidade nas avaliações dos vinhos e pode ser utilizado para perceber se há alguma inconsistência nas pontuações atribuídas por diferentes degustadores.

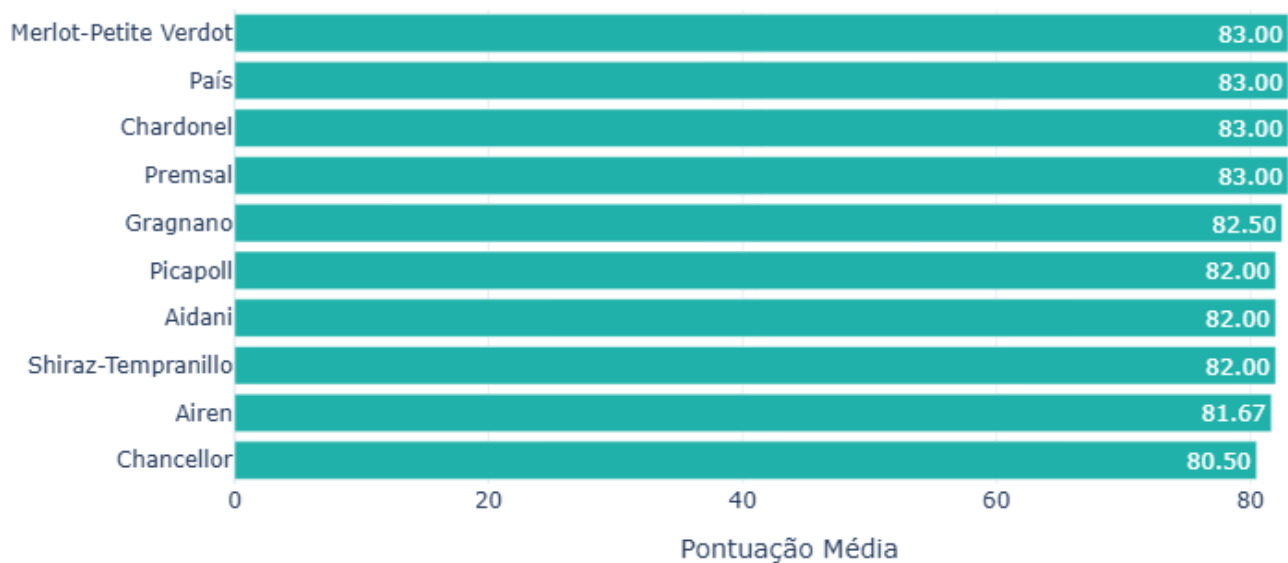
Média de Pontuação dos Vinhos com Pontuação Acima de 90 por Intervalo de Preço



Este gráfico permite analisar como o preço impacta a pontuação média dos vinhos. Através dele, é possível identificar rapidamente se os vinhos mais caros tendem a ter pontuações mais altas ou se há alguma faixa de preço com uma média de pontuação superior.

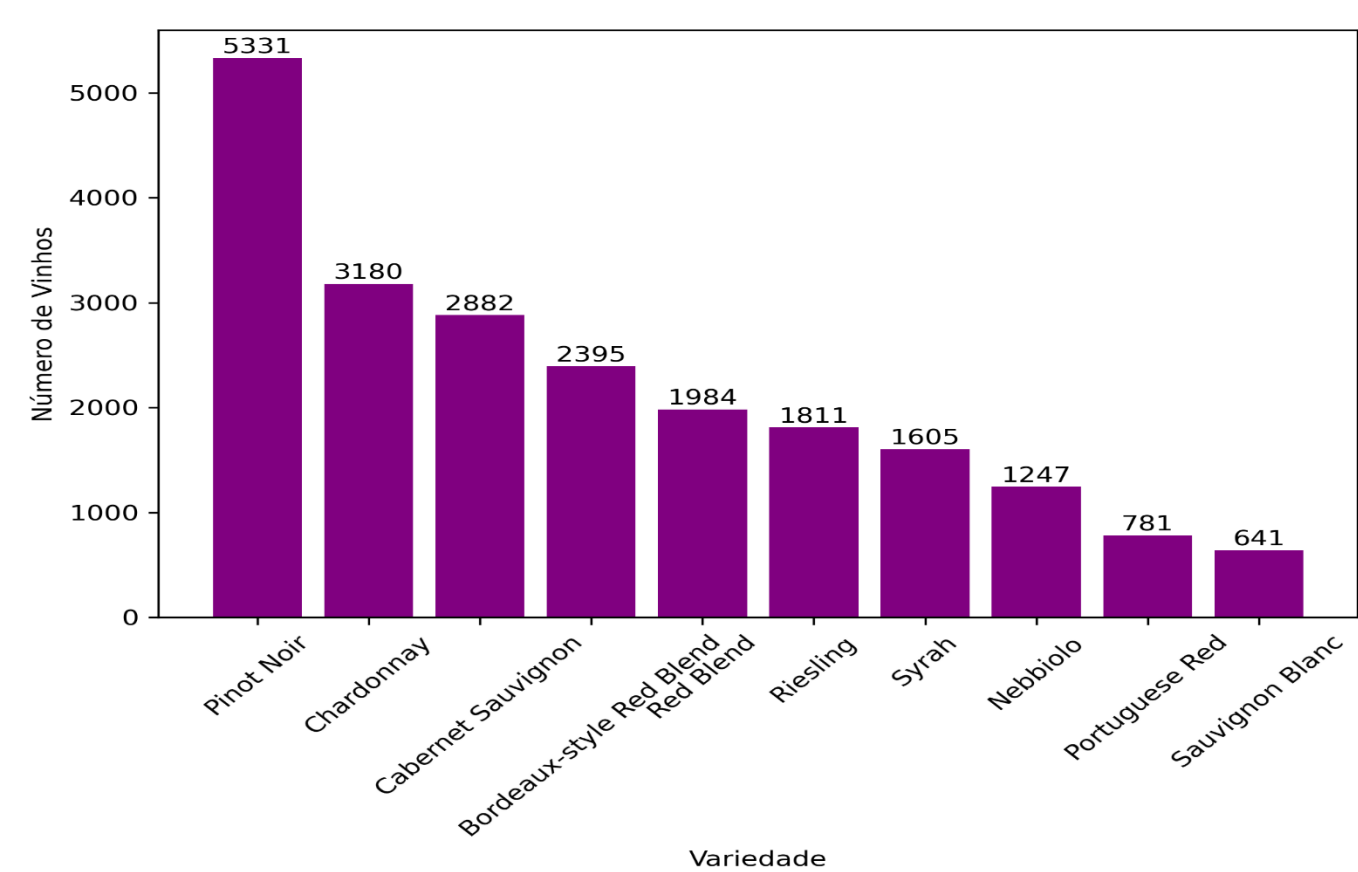
Por exemplo, se você notar que as faixas de preço mais altas possuem uma média de pontuação maior, isso pode indicar que vinhos mais caros tendem a ser avaliados mais positivamente pelos degustadores. Por outro lado, se não houver uma diferença significativa nas pontuações entre as faixas de preço, isso pode sugerir que o preço não é o único fator determinante para as boas avaliações dos vinhos.

10 Tipos de Uva com Menor Média de Pontuação



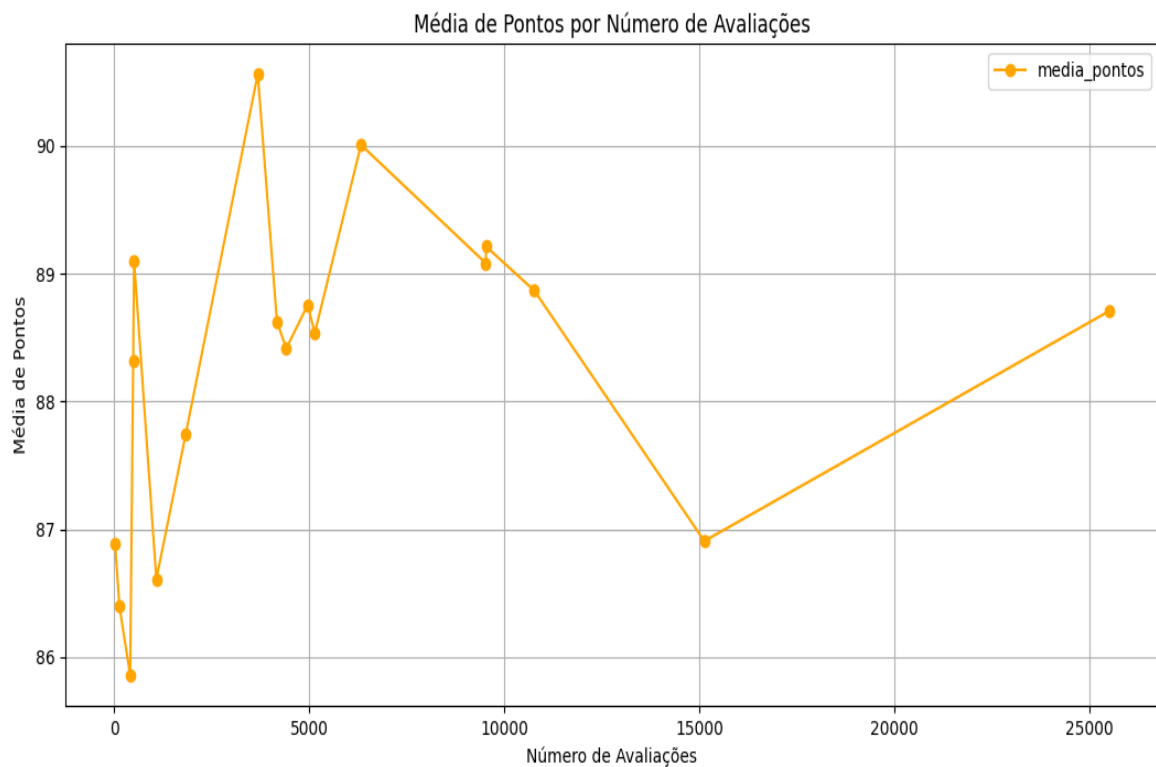
O gráfico de barras horizontais apresenta as 10 variedades de uva com as menores médias de pontuação de acordo com as avaliações registradas na base de dados. Cada barra representa um tipo de uva, e seu comprimento corresponde à pontuação média obtida pelos vinhos feitos a partir dessa variedade.

10 Variedades de Uva nos Vinhos de Alta Qualidade



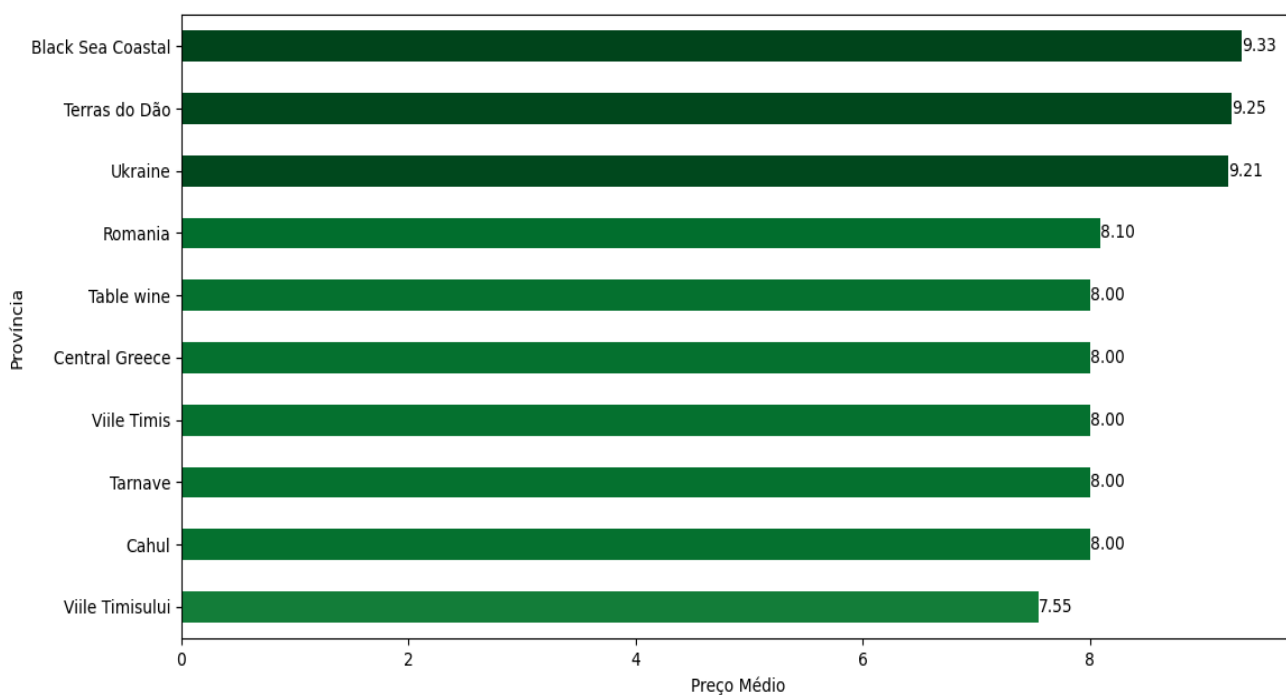
O gráfico de barras apresentado ilustra as 10 maiores variedades de uva encontradas nos vinhos de alta qualidade (com pontuação superior a 90). As barras verticais representam a quantidade de vinhos de alta qualidade que possuem cada variedade de uva, enquanto o eixo horizontal lista as diferentes variedades. A cor das barras foi definida como roxa, proporcionando uma visualização clara e atraente.

Média de Pontos por Número de Avaliações



Este gráfico de linhas mostra como a **média de pontos** dada pelos degustadores varia com o **número de avaliações feitas**. A linha conecta cada degustador, com base no número de vinhos que ele avaliou e sua média de pontos. A ideia é observar se há uma correlação entre a experiência do degustador (representada pelo número de avaliações) e sua média de pontos.

Províncias com Menor Preço Médio



O gráfico exibido mostra as **10 Províncias e Regiões com Menor Preço Médio de Vinho**. Ele é um **gráfico de barras horizontal** com a variável **preço médio** representando o valor médio de cada vinho nas diferentes províncias e regiões, e o **nome da província**, representando os diferentes locais onde os vinhos foram produzidos ou avaliados.

Referências Bibliográficas

pandas: Pandas Documentation. Available at: <https://pandas.pydata.org/pandas-docs/stable/>

matplotlib: Matplotlib Documentation. Available at: <https://matplotlib.org/stable/contents.html>

seaborn: Seaborn Documentation. Available at: <https://seaborn.pydata.org/>

Storytelling com Dados um Guia Sobre Visualização de Dados Para Profissionais de Negócios