



Projeto de Análise de Dados

Avaliações de Vinhos

Projeto de Análise de Dados

Avaliações de Vinhos



Apresentação

- Entendimento do Negócio
- Entendimento dos Dados
- Desenvolvimento do Estudo
- Conclusão

Projeto de Análise de Dados

Avaliações de Vinhos



Apresentação

- **Entendimento do Negócio**
- Entendimento dos Dados
- Desenvolvimento do Estudo
- Conclusão

Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento do Negócio

Contexto

Críticas sobre vinhos com conteúdo de 129971 mil dados de avaliações de vinhos com país (*country*), região (*region*), variedade (*variety*), título (*title*), vinícola (*winery*), preço (*price*), pontuação (*points*), descrição (*description*), designação (*designation*).

Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento do Negócio

Objetivo

Realizar o resumo e interpretação dos dados proporcionando seu conhecimento e entendimento.

Premissas

Origem dos Dados: <https://www.kaggle.com/datasets/zynicide/wine-reviews>

Arquivo: winemag-data-130k-v2.csv(52.91 MB)

Os dados foram extraídos do *WineEnthusiast* durante a semana de 15 de junho de 2017.

Projeto de Análise de Dados

Avaliações de Vinhos



Apresentação

- Entendimento do Negócio
- **Entendimento dos Dados**
- Desenvolvimento do Estudo
- Conclusão

Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Amostra dos Dados

country	description	designation	points	price	province	region_1	region_2	taster_name	taster_twitter_handle	title	variety	winery	Classificacao
Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	NaN	Kerin O'Keefe	@kerinokeefe	Nicosia 2013 Vulkà Bianco (Etna)	White Blend	Nicosia	Acima da Média
Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	NaN	NaN	Roger Voss	@vossroger	Quinta dos Avidagos 2011 Avidagos Red (Douro)	Portuguese Red	Quinta dos Avidagos	Acima da Média
US	Tart and snappy, the flavors of lime flesh and...	NaN	87	14.0	Oregon	Willamette Valley	Willamette Valley	Paul Gregutt	@paulgwine	Rainstorm 2013 Pinot Gris (Willamette Valley)	Pinot Gris	Rainstorm	Acima da Média
US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	87	13.0	Michigan	Lake Michigan Shore	NaN	Alexander Peartree	NaN	St. Julian 2013 Reserve Late Harvest Riesling ...	Riesling	St. Julian	Acima da Média
US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	87	65.0	Oregon	Willamette Valley	Willamette Valley	Paul Gregutt	@paulgwine	Sweet Cheeks 2012 Vintner's Reserve Wild Child...	Pinot Noir	Sweet Cheeks	Acima da Média
...
Germany	Notes of honeysuckle and cantaloupe	Brauneberger Juffer-Sonnenuhr	90	28.0	Mosel	NaN	NaN	Anna Lee C. Iijima	NaN	Dr. H. Thanisch (Erben Müller-Burggraef) 2012	Riesling	Dr. H. Thanisch (Erben Müller-Burggraef)	Excelente

Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Descrição dos Dados

country	País de origem do vinho.
description	Descrição das características do vinho.
designation	A vinha dentro da adega de onde provêm as uvas que dão origem ao vinho.
points	Número de pontos que a <i>WineEnthusiast</i> classificou o vinho de 50 até 100
price	Custo de uma garrafa de vinho em US\$ (no ano de 2017)
province	Província ou estado de origem do vinho
region_1	A área vitivinícola em uma província ou estado (exemplo: Napa)

Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Descrição dos Dados

region_2	Às vezes, há regiões especificadas dentro de uma área vitivinícola (ou seja, Rutherford dentro do Vale de Napa).
taster_name	Nome do testador/avaliador do vinho (enófilo).
taster_twitter_handle	Twitter do testador.
variety	Tipo da uva utilizada na preparação do vinho e responsável por terminar de moldar o sabor de cada bebida, sendo uma das principais responsáveis pela suavidade ou amargor de rótulos específicos.

Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Descrição dos Dados

winery	Vinícola onde o vinho foi produzido.
classificacao	<p>Não constante na origem dos dados. Coluna criada para categorizar a qualidade do vinho conforme a faixa de pontuação abaixo:</p> <p>50-59 -> Inapropriado para Consumo 60-69-> Abaixo da Média 70-79-> Mediano e Inócuo 80-89-> Acima da Média 90-95-> Excelente 96-100-> Extraordinário (Referência: https://www.divvino.com.br/blog/pontuacao-de-vinhos/)</p>

Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados



Lidera na **quantidade**
de avaliações,
representando 41.95%
do total.

Qtde Geral de Avaliações

129.971

Número de países englobados: 43 países

	Frequência	Porcentagem(%)
US	54504	41.955846
France	22093	17.006651
Italy	19540	15.041414
Spain	6645	5.115158
Portugal	5691	4.380793
Chile	4472	3.442436
Argentina	3800	2.925147
Austria	3345	2.574899
Australia	2329	1.792807
Germany	2165	1.666564
. . .		
Brazil	52	0.040028

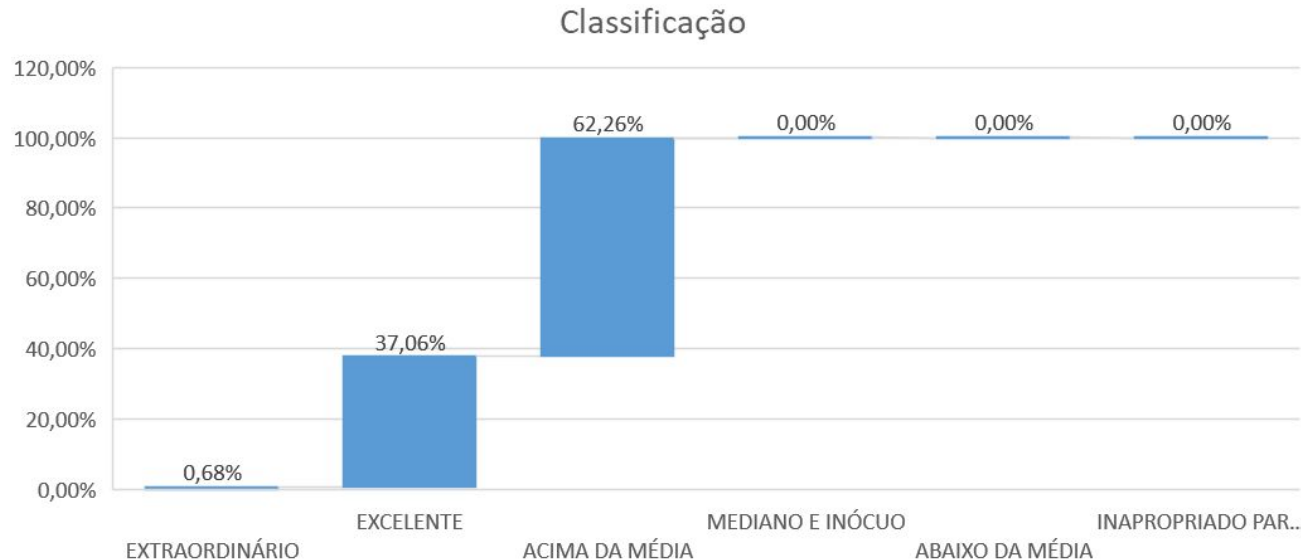
Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Classificação



CLASSIFICAÇÃO	QUANTIDADE DE AVALIAÇÕES
EXTRAORDINÁRIO	881
EXCELENTE	48164
ACIMA DA MÉDIA	80926
MEDIANO E INÓCUO	0
ABAIXO DA MÉDIA	0
INAPROPRIADO PARA CONSUMO	0

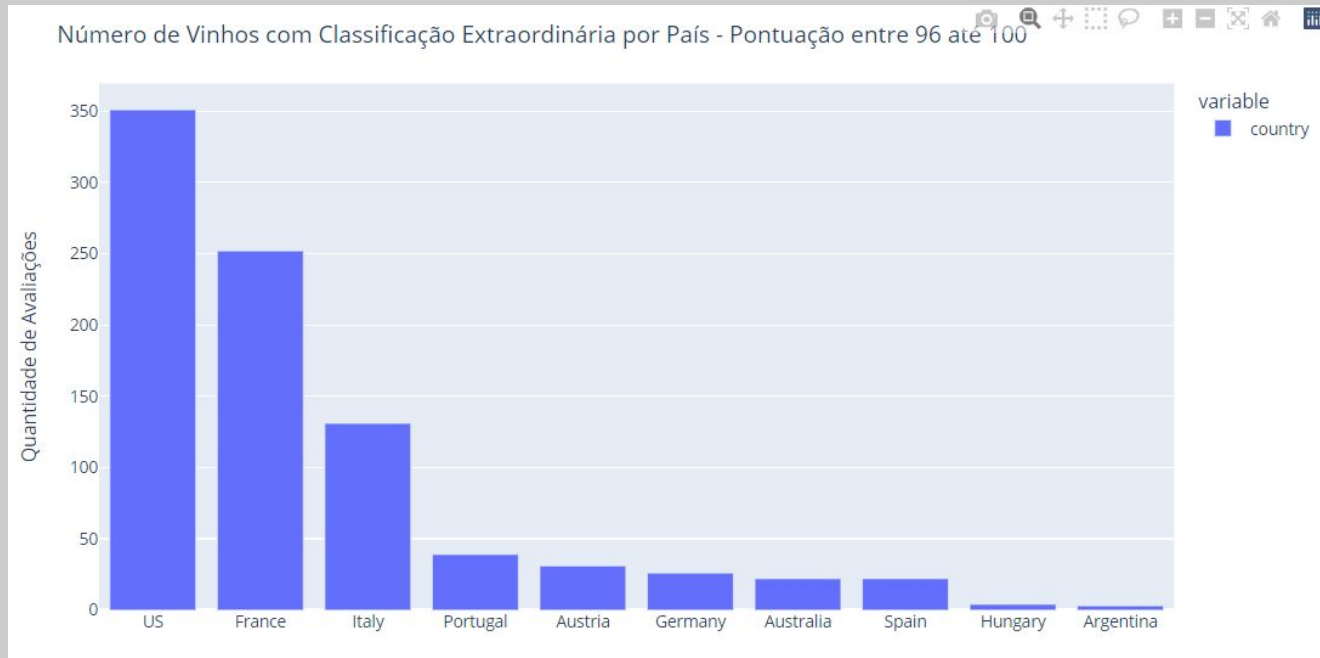
Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Vinhos Extraordinários



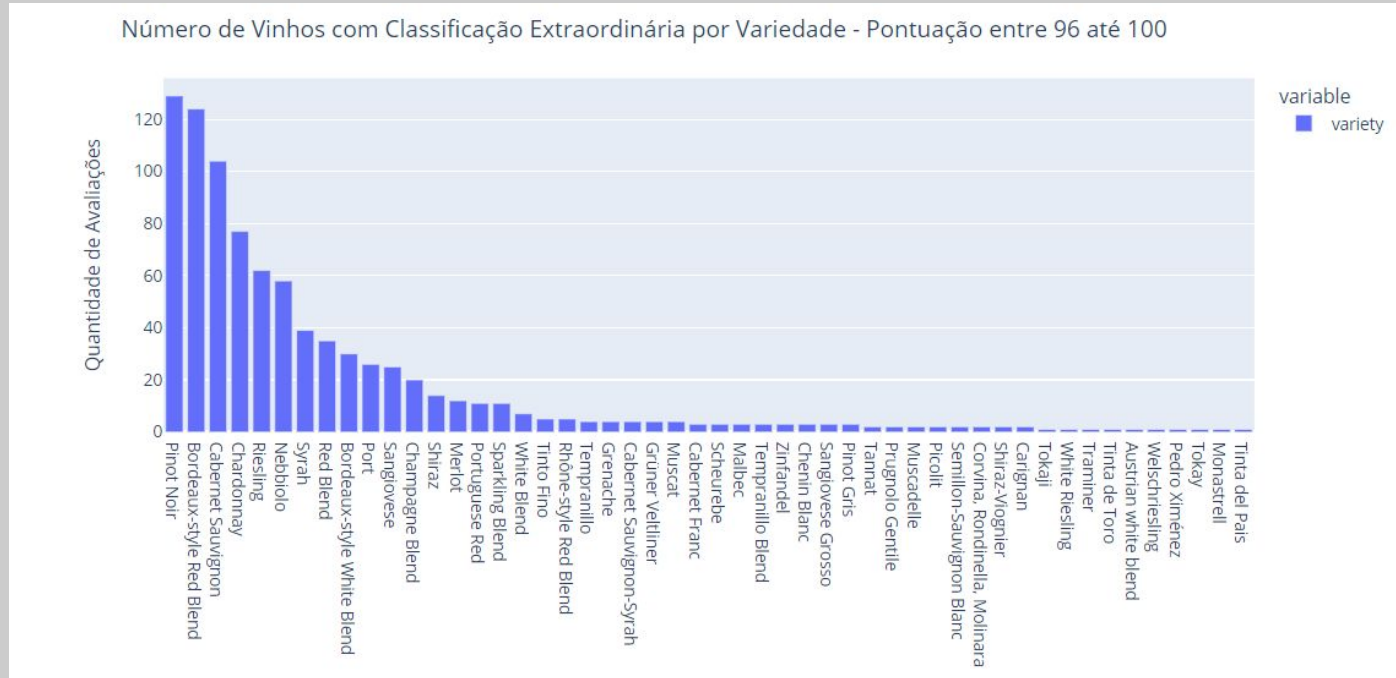
Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Vinhos Extraordinários



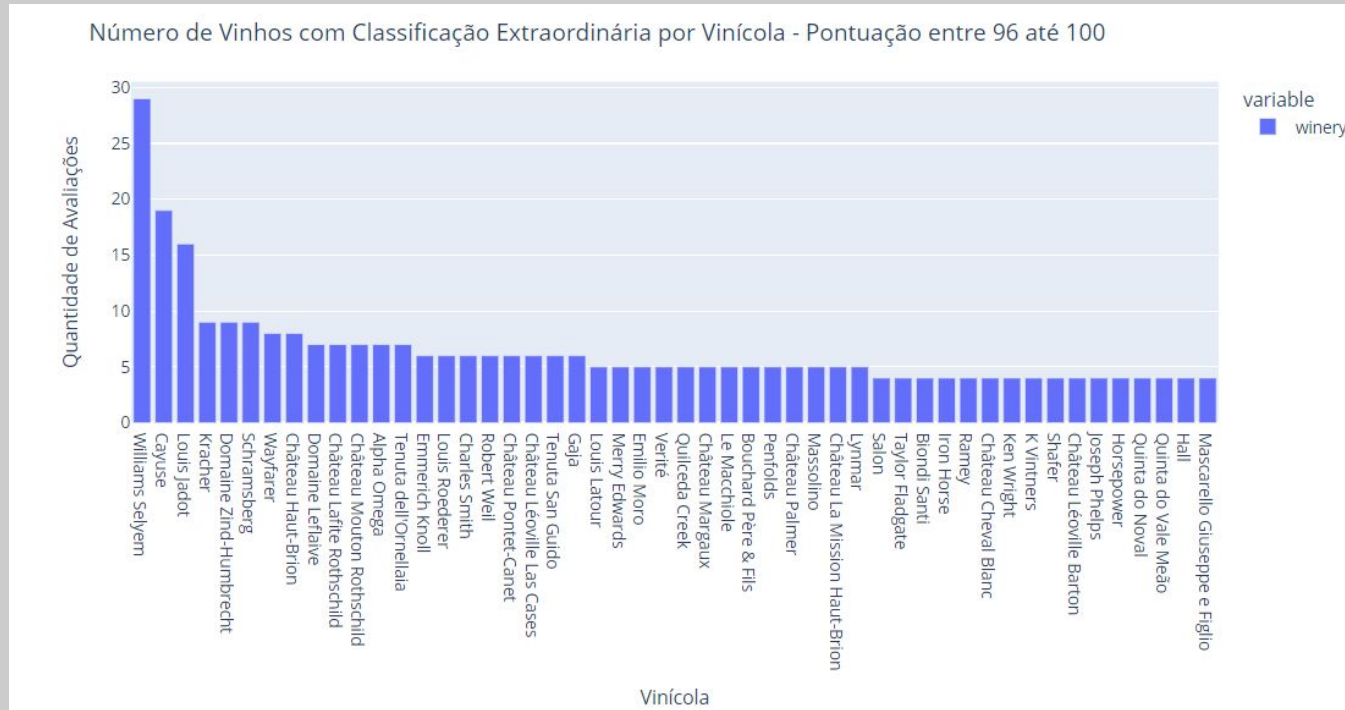
Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Vinhos Extraordinários



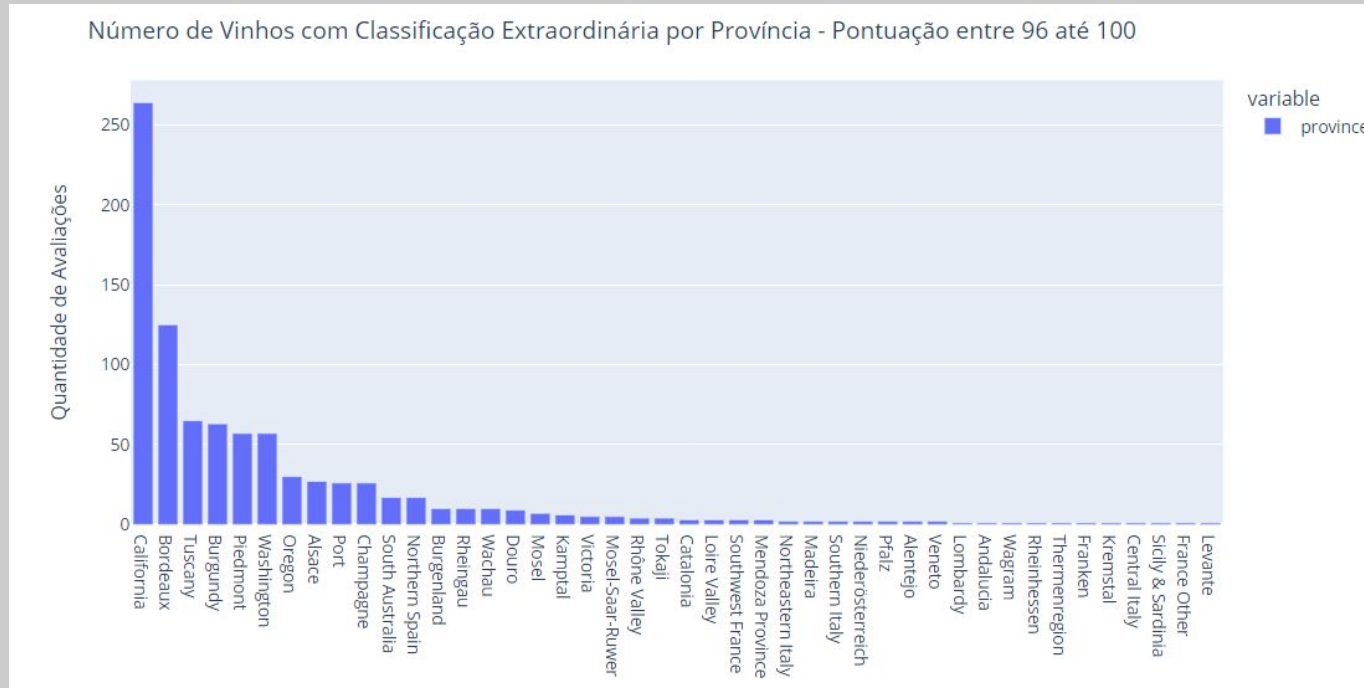
Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Vinhos Extraordinários



Projeto de Análise de Dados

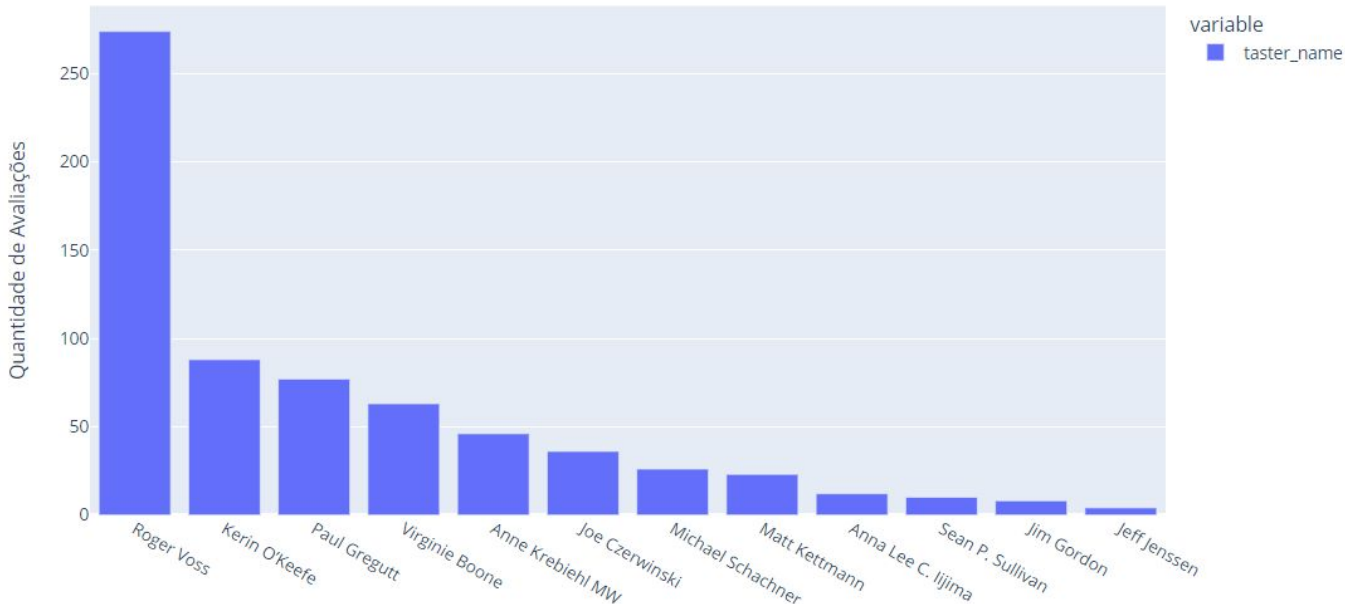
Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Vinhos Extraordinários

Número de Vinhos com Classificação Extraordinária por Testador - Pontuação entre 96 até 100



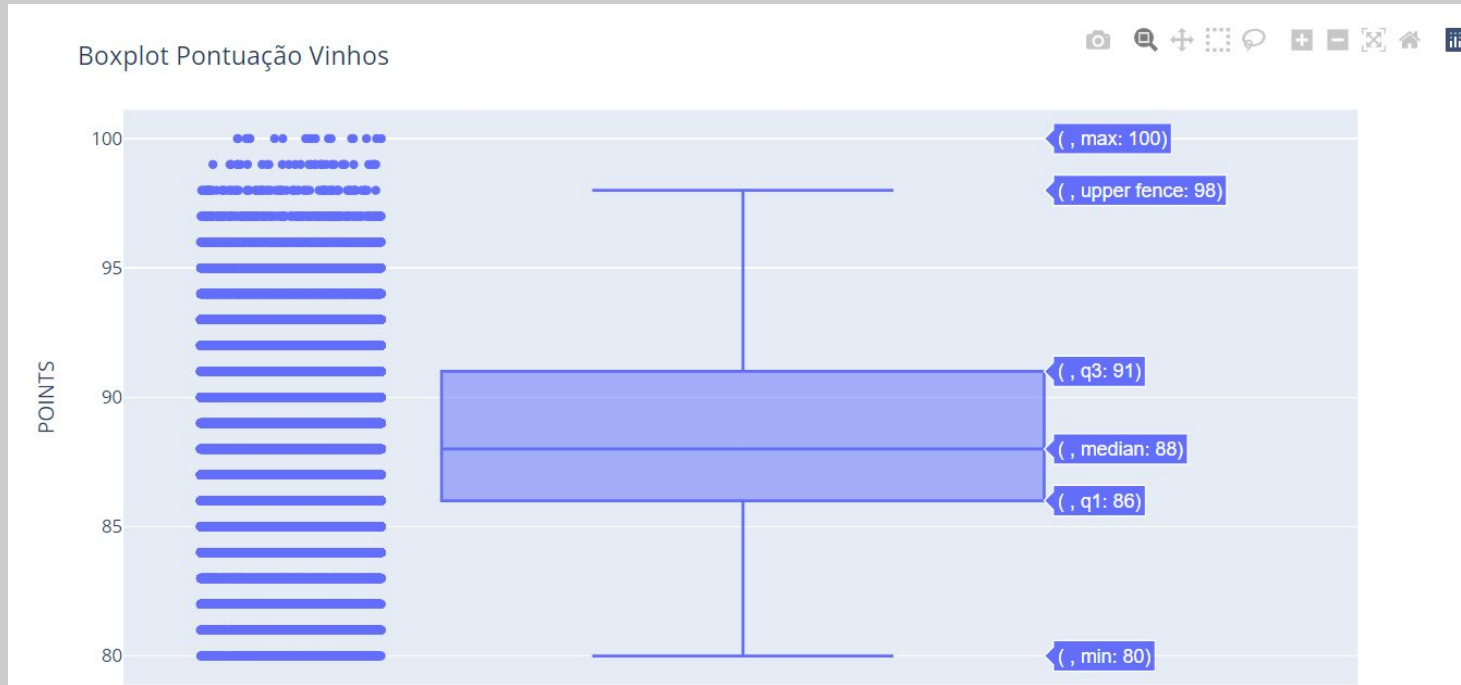
Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Distribuição da Pontuação



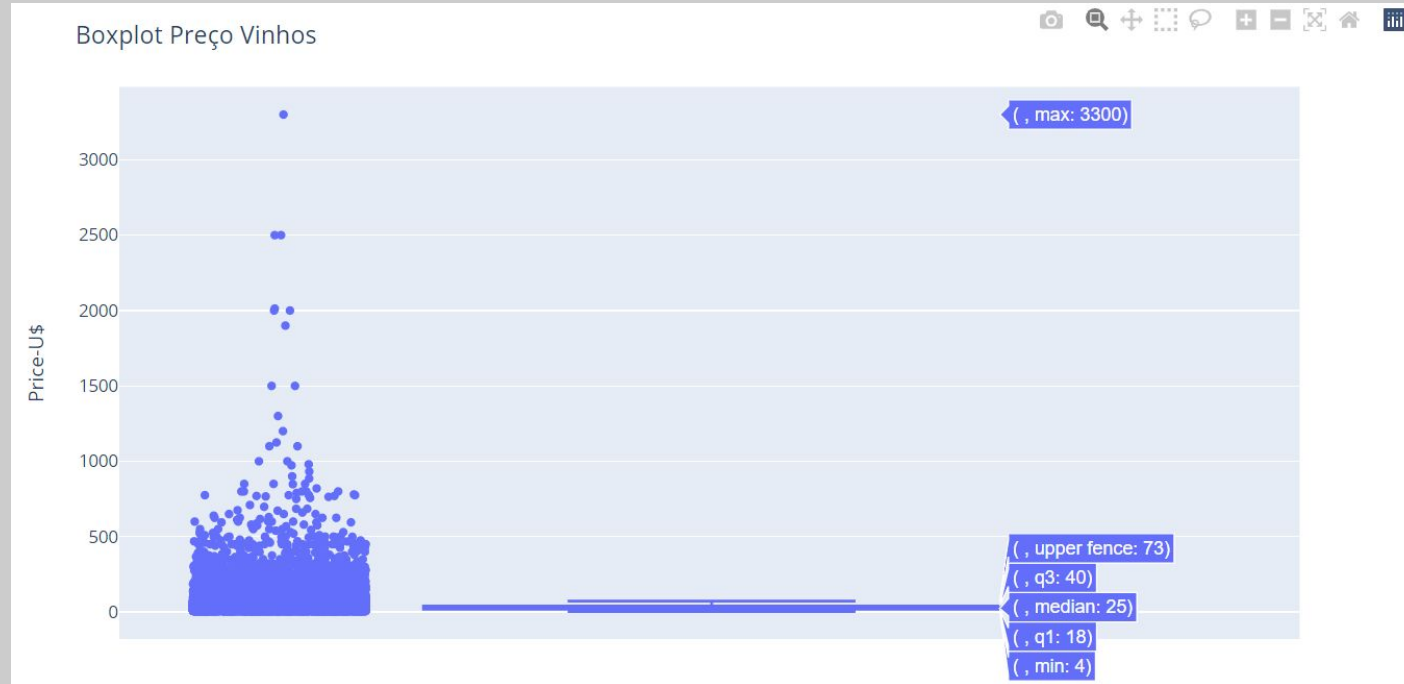
Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Distribuição do Valor (Price)



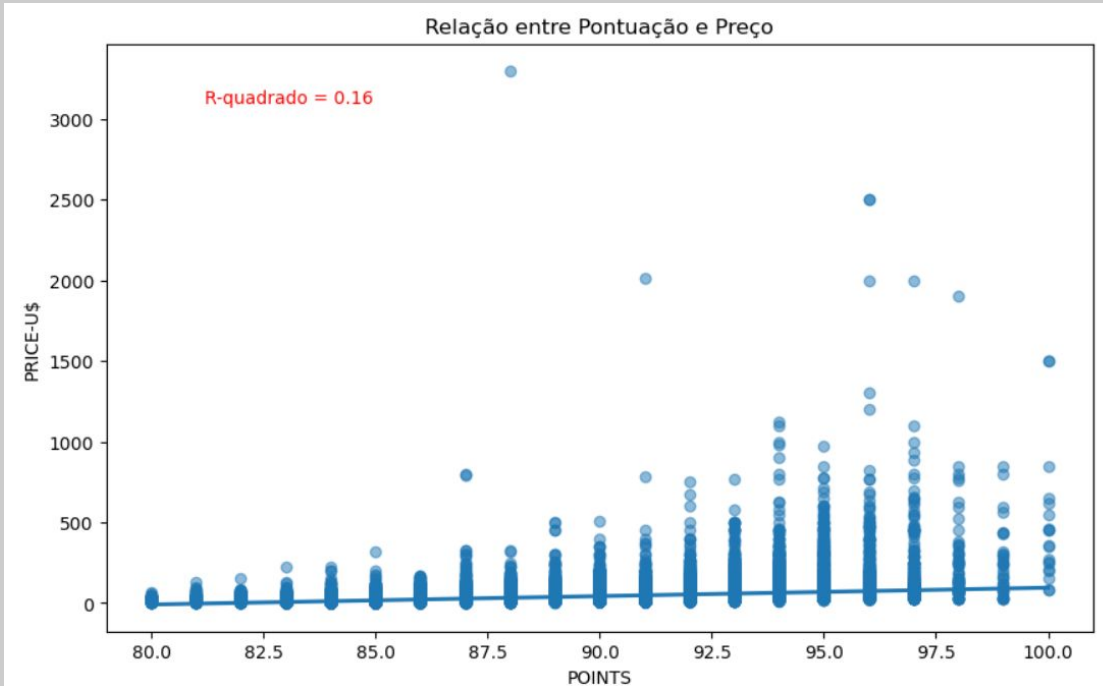
Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Correlação Preço (Price) e Pontuação (Points)



FRACA

É importante considerar a existência de outros fatores (outras variáveis) que possam influenciar na precificação do vinho.

Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados



Qtde Geral de Avaliações	Qtde Avaliações Brasil		Classificação
129.971	52	0,04%	Acima da Média (100%) <i>©Carla G. B. Teixeira</i>



Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados

Medidas Resumo - Brasil			x	Medidas Resumo - Global		
	points	price			points	price
count	52.000000	52.000000		count	129971.000000	129971.000000
mean	84.673077	23.884615		mean	88.447138	34.646083
std	2.340782	10.504255		std	3.039730	39.664385
min	80.000000	10.000000		min	80.000000	4.000000
25%	83.000000	15.000000		25%	86.000000	18.000000
50%	85.000000	23.000000		50%	88.000000	25.000000
75%	86.000000	27.250000		75%	91.000000	40.000000
max	89.000000	60.000000		max	100.000000	3300.000000
median	85.0000	23.00000		median	88.0000	25.00000



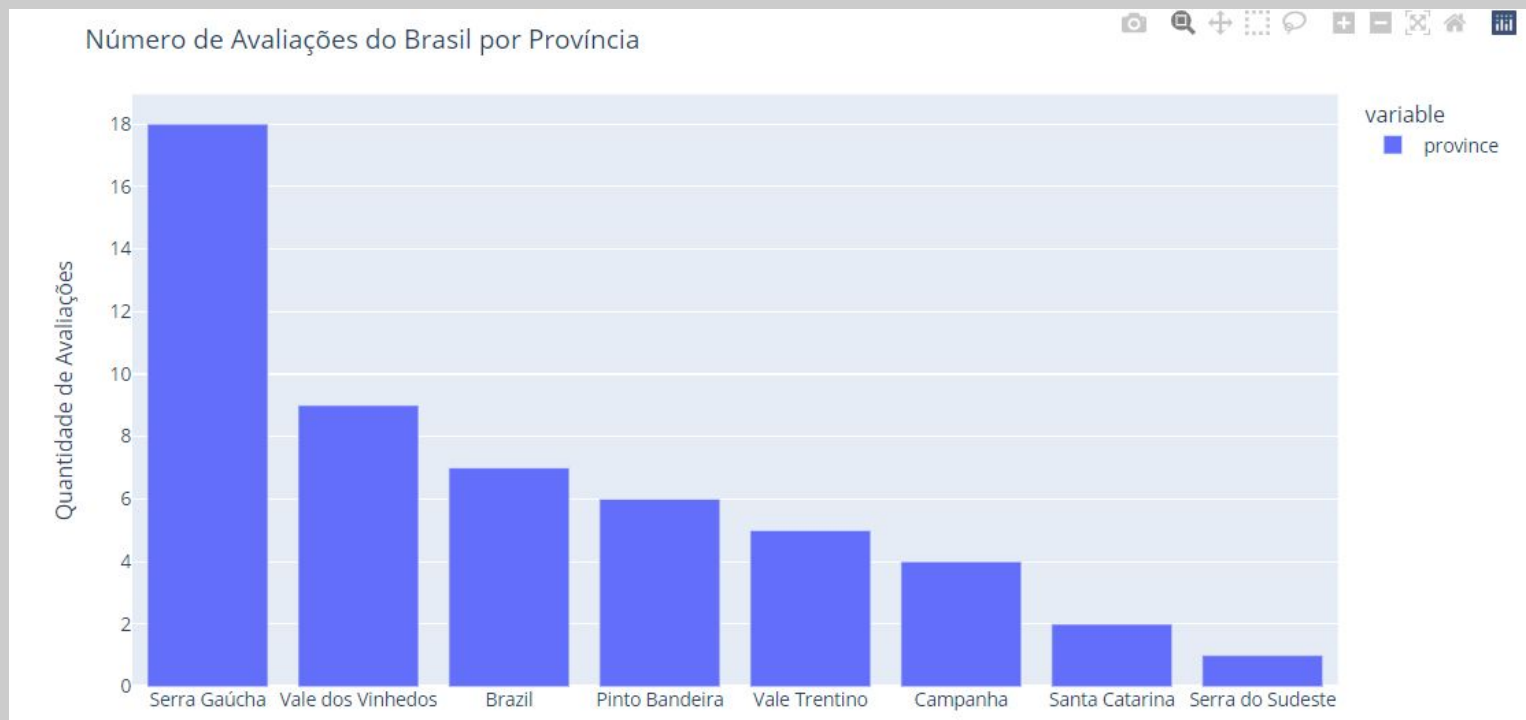
Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Brasil - Província





Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Brasil - Vinícola





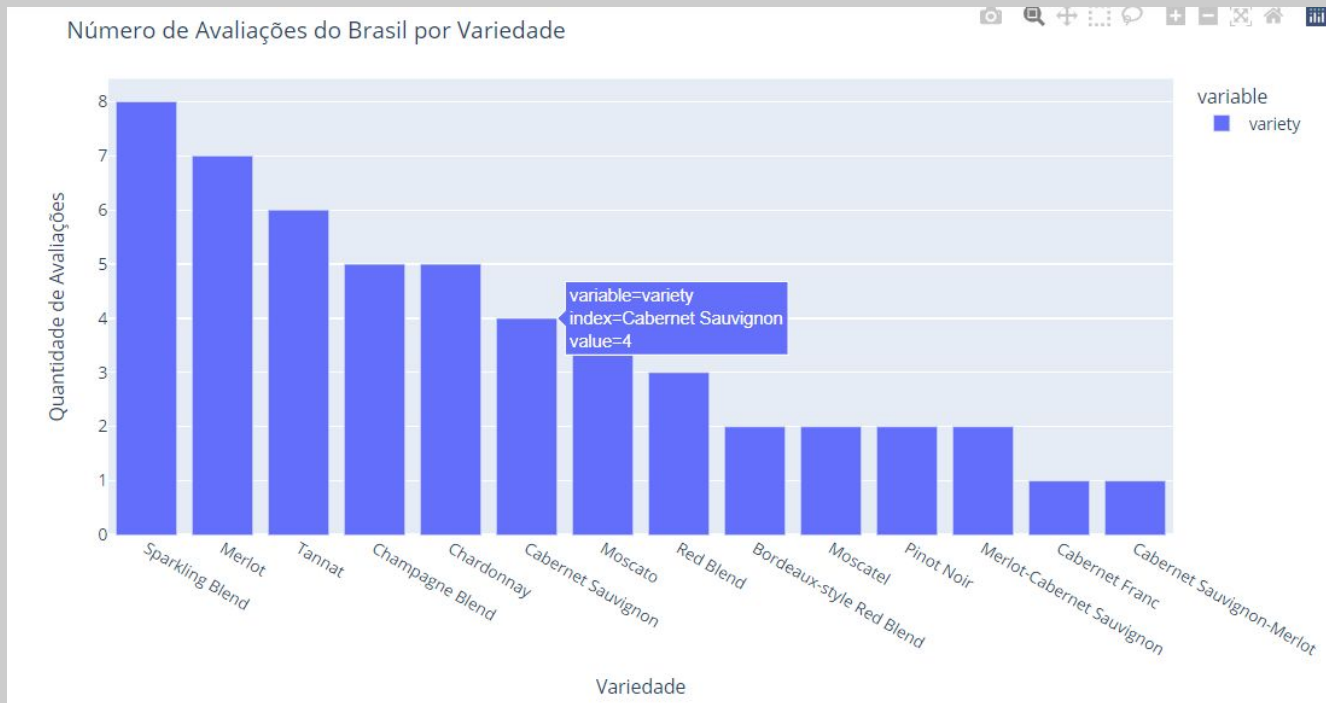
Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Brasil - Variedade





Projeto de Análise de Dados

Avaliações de Vinhos



Entendimento dos Dados

Análise Exploratória dos Dados - Brasil - Testador



Projeto de Análise de Dados

Avaliações de Vinhos



Apresentação

- Entendimento do Negócio
- Entendimento dos Dados
- **Desenvolvimento do Estudo**
- Conclusão

Projeto de Análise de Dados

Avaliações de Vinhos

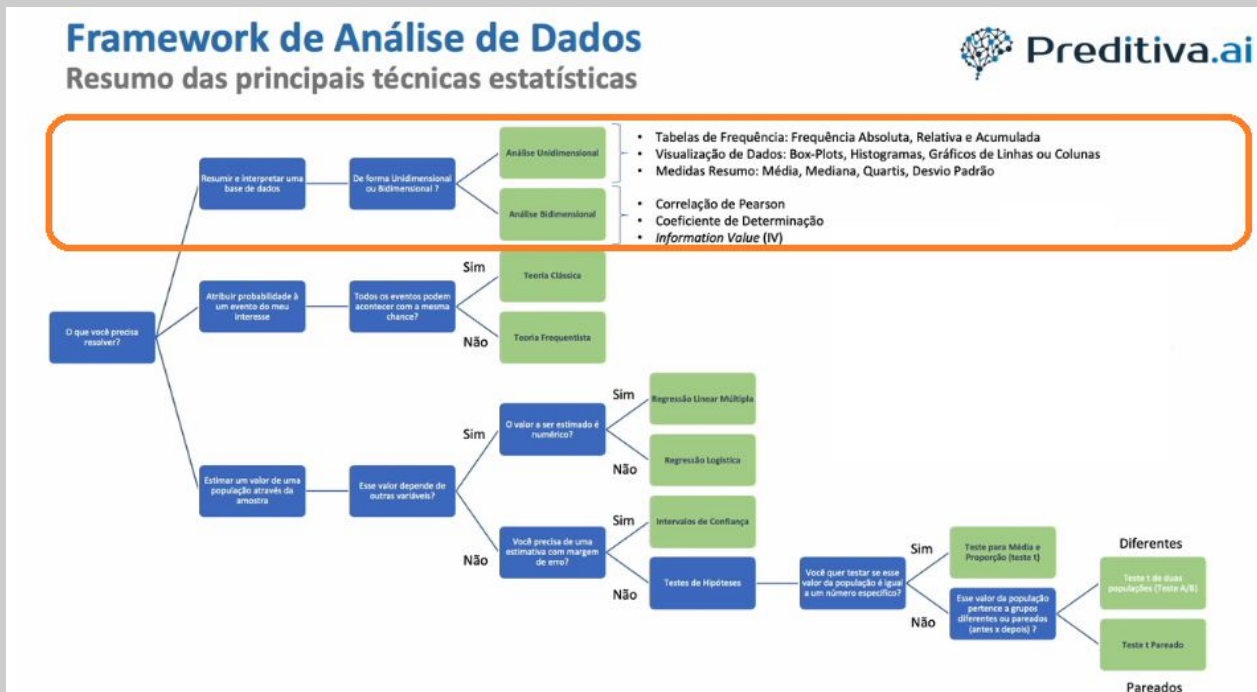


Desenvolvimento do Estudo

Técnicas Estatísticas Utilizadas

Estatística descritiva nas análises unidimensionais e bidimensionais.

Linguagem
Python



Projeto de Análise de Dados

Avaliações de Vinhos



Apresentação

- Entendimento do Negócio
- Entendimento dos Dados
- Desenvolvimento do Estudo
- **Conclusão**

Projeto de Análise de Dados

Avaliações de Vinhos



Conclusão

A amostra analisada traz uma concentração de avaliações nos US (41% das avaliações). A seguir temos França (17%) e Itália (15%). O Brasil representa 0.04% das avaliações, com apenas 52 observações e todas com classificação Acima da Média (pontuação entre 80 a 89). A Casa Perini na Serra Gaúcha obteve maior quantidade de avaliações.

Com as observações altamente concentradas em um único país (US), a chance de algum tipo de informação seja maior de aparecer neste país. Citamos como exemplo, a frequência da classificação do vinho como extraordinário, onde obtemos um total de 881 avaliações (0,6%). A concentração de vinhos avaliados como extraordinários está no US, seguido da França e Itália.

Projeto de Análise de Dados

Avaliações de Vinhos



Conclusão

As variedades de vinhos mais avaliadas estão no Pinot Noir (10% - 13272 avaliações), seguido do Chardonnay (9% - 11753 avaliações) e Cabernet Sauvignon (7% - 9472 avaliações) .

O preço dos vinhos possui valores extremos de US\$ 4.00 (menor) e US\$ (3300) (maior valor). A mediana ficou em US\$ 25.00 considerando toda a amostra.

A análise de correlação entre a pontuação e preço resultou ser muito fraca nos métodos apurados. Desta forma não podemos afirmar ou inferir que quanto maior a pontuação de um vinho maior será o seu preço ou vice versa, podendo existir outras variáveis que precisam ser analisadas para inferência de preço.

Projeto de Análise de Dados

Avaliações de Vinhos



Conclusão

A *WineEnthusiast* apresentou o [TOP 100 BEST BY de 2023](#), onde uma equipe avaliou, as cegas, mais de 30.000 vinhos. “O vinho nº 1 do TOP 100 é um Syrah de 93 pontos e US\$ 15 amplamente disponível da vinícola J. Lohr da Califórnia, sobre o qual o revisor Matt Kettmann diz: “Este é um Syrah muito satisfatório por um preço bastante impressionante””.

WineEnthusiast



Projeto de Análise de Dados

Avaliações de Vinhos



Conclusão

Traços comparativos podem ser melhor aplicados por país ou por conjunto de países que detêm características e frequências similares.

De acordo com o propósito do trabalho, conclui-se que obtivemos as respostas necessárias para o resumo, interpretação e conhecimento das informações.

Portfólio deste Estudo em:

<https://github.com/Carla-G-B-Teixeira/Portfolio/tree/Portfolio/Estudos%20de%20Caso/Vinhos>

<https://drive.google.com/drive/folders/1xLQ5fSciadCSrOxWwUVEzrGdZtWeOiiw>