



Mundo del vino

Chiara Baudracco Ray

17 de Noviembre del 2024

Índice

Análisis de la industria vinícola	3
Organización	6
Inspección inicial y limpieza de datos	7
Análisis de valores nulos y duplicados	7
Resumen estadístico y gráficos iniciales	10
Gráfico de distribución de puntuaciones de los vinos	12
Gráfico de distribución de precios por país	14
Gráfico de puntaje promedio de vinos por país	16
Gráfico de relación entre precio y puntaje	18
Gráfico de puntaje promedio de vinos por año	20
Gráfico de dispersión: relación entre puntaje y precio	23

Gráfico de cantidad de vinos producidos por región
24

Gráfico de relación entre precio y puntaje de los
vinos 26

Gráfico de precio promedio de vinos por país 29

Gráfico de distribución de precios de vinos con 95+
puntos 31

Análisis de la industria vinícola

La industria del vino es un campo apasionante y complejo que ha fascinado a sommeliers, enólogos y consumidores durante siglos. La capacidad de analizar y clasificar los vinos no solo requiere un profundo conocimiento de las características sensoriales, sino también un entendimiento de las influencias geográficas, varietales y de métodos de producción. En este contexto, el análisis de datos juega un papel esencial, permitiendo a través de técnicas modernas de ciencia de datos explorar y descubrir patrones significativos en grandes conjuntos de información. Este estudio, en particular, se basa en el análisis de un conjunto de datos extenso proveniente de WineEnthusiast, que contiene alrededor de 130,000 reseñas detalladas de vinos, con variables que van

desde la descripción sensorial hasta las regiones de origen y las puntuaciones otorgadas por expertos.

El dataset, denominado [winemag-data-130k-v2.csv](#), incluye varias columnas clave que nos proporcionan una visión integral de cada vino reseñado. Estas columnas cubren desde aspectos geográficos, como el país y la región de origen, hasta características propias del vino, como la variedad de uva, la bodega productora, el precio y la puntuación. La columna de descripción, en particular, ofrece un recurso invaluable para el análisis de lenguaje, ya que en ella se encuentran las notas de cata detalladas, esenciales para los sommeliers al evaluar un vino. Además, los datos sobre el nombre y usuario de Twitter del catador proporcionan un contexto que podría permitir explorar la consistencia y estilo de las reseñas realizadas por distintos críticos.

A través de este estudio, se realiza un análisis exploratorio exhaustivo que abarca varias

fases. Primero, la selección y depuración de datos será esencial para asegurar la calidad y consistencia de la información, eliminando duplicados y datos erróneos. Posteriormente, contextualizar el análisis mediante la formulación de preguntas clave permitirá establecer objetivos claros para comprender patrones y relaciones significativas en los datos.

Una vez depurados y contextualizados los datos, se procederá con la creación de visualizaciones utilizando herramientas de análisis y exploración en Python o R. Estas visualizaciones serán fundamentales para comunicar los descubrimientos de manera clara e intuitiva. A través de gráficos, como histogramas de distribución de puntuaciones o diagramas de dispersión entre precio y calidad, se podrán identificar tendencias y correlaciones que de otro modo podrían pasar desapercibidas.

Finalmente, la fase de interpretación y presentación de resultados constituirá el último paso en el análisis. A través de una síntesis de los hallazgos más significativos, elaboraré una interpretación detallada que no solo explique las relaciones y patrones observados, sino también su utilidad práctica para actores de la industria del vino, como productores y comerciantes.

Organización

El código incluirá:

1. Carga y preparación de los datos.
2. Limpieza de datos (tratamiento de valores nulos, duplicados, etc.).
3. Análisis exploratorio (estadísticas y visualización inicial).
4. Visualización de patrones y relaciones específicas.

Pasos detallados

1. Preparación inicial

Explicación:

1. Se importan bibliotecas como **pandas** y **numpy** para manipulación de datos, y **matplotlib** y **seaborn** para visualización.
2. Cargamos el archivo CSV en un DataFrame de pandas, que es la estructura de datos ideal para analizar tablas de datos.
3. Revisamos las primeras filas con **head()** para asegurarnos de que el archivo se ha cargado correctamente.

2. Inspección inicial y limpieza de datos

Análisis de valores nulos y duplicados

Explicación:

1. **info()** muestra los tipos de datos y detecta valores nulos.

2. Se identifican columnas con valores nulos con `isnull().sum()`, y se eliminan duplicados para evitar datos redundantes.

Para manejar los valores nulos en cada columna, seguí algunas reglas comunes en la imputación y limpieza de datos en un dataset de vinos. Una guía con el código en Python para cada columna, junto con una breve justificación.

1. **country y province**: Como estos dos valores suelen ser importantes para el análisis de los vinos, eliminamos las filas con valores nulos en estas columnas, ya que no es un número alto y la falta de este dato podría afectar al análisis.

2. **designation, region_1, region_2**: Dado que son datos descriptivos y tienen un alto número de valores nulos, es posible llenarlos con un valor indicativo como "Desconocido" o NaN, en lugar de eliminarlos, para evitar perder demasiada información.

3. **price**: Este es un dato importante para el análisis. Para los valores nulos en esta columna, podemos utilizar la mediana para completarlos, ya que el precio puede variar mucho y la mediana es menos sensible a valores extremos.

4. **taster_name** y **taster_twitter_handle**: Son datos descriptivos que no son esenciales para todos los análisis. Podemos imputarlos con "Desconocido" o eliminar estas columnas si no son relevantes para el proyecto.

5. **variety**: Hay solo un valor nulo, por lo que es seguro eliminar esta fila.

Este código limpia el dataset al reducir los valores nulos y deja columnas importantes para el análisis sin mucha pérdida de datos.

Análisis exploratorio de los datos

Resumen estadístico y gráficos iniciales

`describe()` genera estadísticas clave, incluyendo promedio, desviación estándar, etc.

Creamos gráficos de histograma para `points` y `price` para visualizar la distribución de puntuaciones y precios de los vinos.

Análisis de gráficos

Estas visualizaciones aportan una visión más profunda y detallada del dataset de vinos, permitiendo descubrir patrones específicos por país, variedad y puntaje. Cada gráfico ayuda a responder preguntas importantes sobre la calidad, accesibilidad y variabilidad de los vinos en función de factores como el país de origen, la variedad, el precio y el año.

Gráfico de Distribución de Precios de los Vinos

- En este gráfico se observa una distribución asimétrica con una gran concentración de vinos en el rango de precios bajos (0-100 dólares), y una

Mundo del Vino

minoría que alcanza precios elevados, con algunos vinos que superan los 1500 dólares.

- **Conclusión:** La mayoría de los vinos tienen precios accesibles, lo que sugiere un mercado muy competitivo en el rango bajo. Para los comerciantes, esta información es útil para ajustar el enfoque de marketing, destacando los atributos que justifican los precios altos en caso de vinos de gama alta. Para los productores, podría ser estratégico enfocarse en vinos con buena relación calidad-precio para captar a la mayoría de los consumidores.

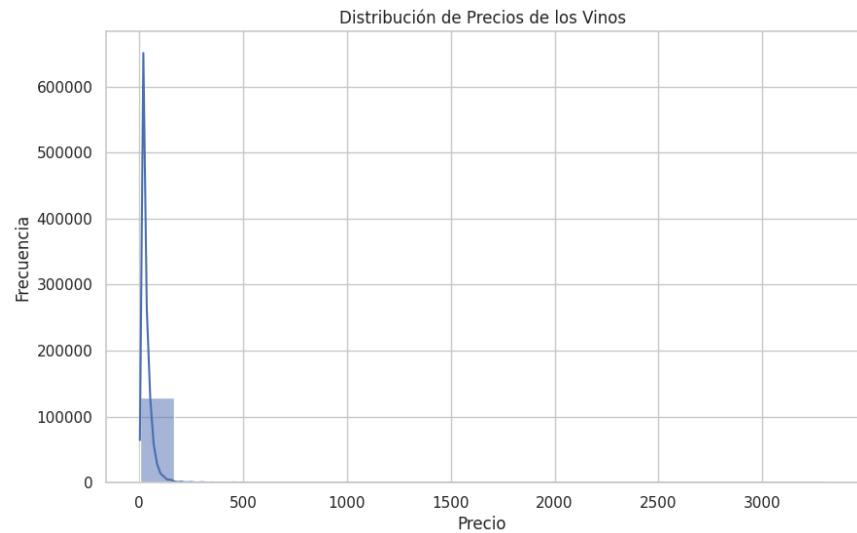


Gráfico de Distribución de Puntuaciones de los Vinos

Este gráfico muestra la frecuencia de las puntuaciones asignadas a los vinos en una escala de 80 a 100 puntos. Observamos una distribución normal, con un pico principal alrededor de 87-88 puntos, que es la puntuación más común. Esto sugiere que la mayoría de los vinos reciben una calificación alta, pero dentro de un rango limitado, lo cual puede deberse a la calidad estándar que se espera en esta industria.

Conclusión: La concentración de puntuaciones en el rango alto indica que los vinos en este conjunto de datos generalmente cumplen con una calidad aceptable o superior. Para los productores, esto significa que una mejora en la calidad podría incrementar la calificación, situándose en la parte superior de la curva (hacia los 95-100 puntos), lo cual puede ser atractivo para los consumidores que buscan vinos de élite.

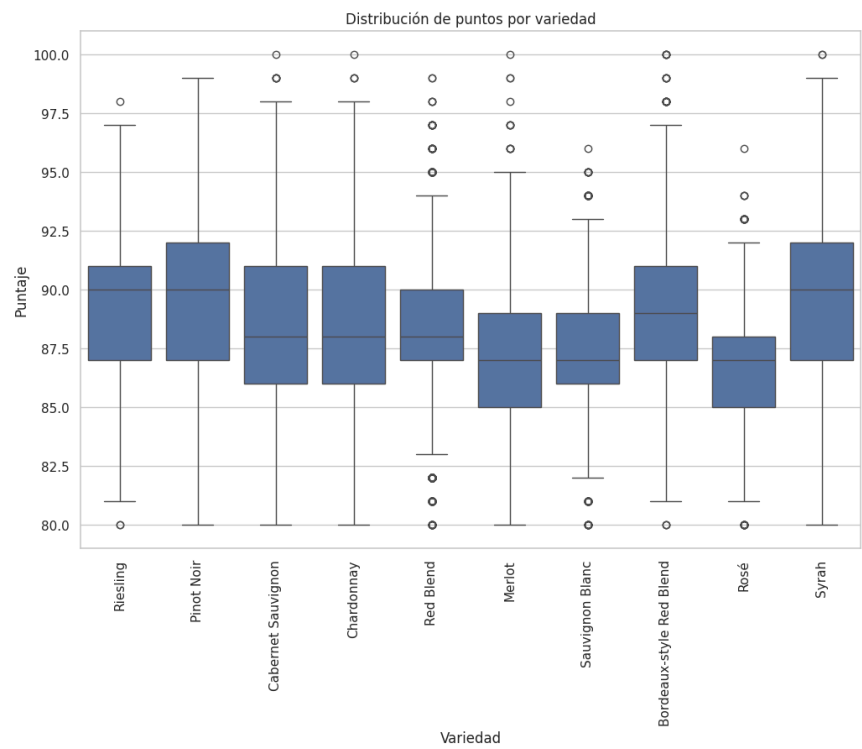


Gráfico de Distribución de Precios por País

Este gráfico ilustra cómo varían los precios de los vinos según el país. Vemos que países como Italia, Portugal, y Estados Unidos tienen vinos que

alcanzan precios altos, mientras que otros como Argentina, España y Chile tienen precios más bajos.

Conclusión: La variabilidad en los precios entre países refleja diferencias en las percepciones de calidad, costos de producción y posicionamiento de mercado. Esto puede indicar que ciertos países tienen vinos premium muy valorados (como Italia), mientras que otros se especializan en vinos asequibles. Para los comerciantes, esta información puede orientar las estrategias de importación y exportación. Para los productores, conocer esta variabilidad puede ayudarles a adaptar su producción según la demanda internacional y los precios de referencia en su categoría.

Mundo del Vino

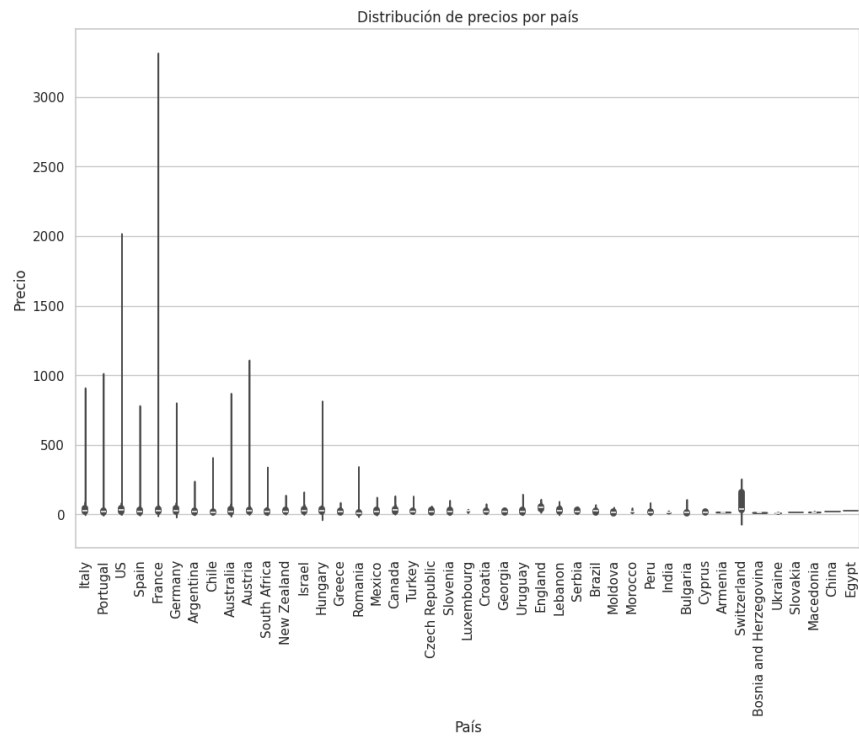


Gráfico de puntaje promedio de vinos por país

En este gráfico de barras, podemos ver los puntajes promedio de vinos distribuidos por país. Los países están ordenados de mayor a menor

puntaje promedio, lo que facilita la comparación entre ellos. A continuación se destacan algunos puntos:

- Países líderes: Países como Inglaterra, India y Austria tienen los puntajes promedio más altos, lo que sugiere una buena calidad de vino según las reseñas.
- Diversidad de calidad: La variación en los puntajes entre los países indica diferencias en la percepción de calidad del vino en cada región.
- Posibles áreas de mejora: Los países con puntajes más bajos, como Egipto y Perú, podrían enfocarse en mejorar ciertos aspectos de la producción o distribución para incrementar la calidad percibida de sus vinos.

Este análisis puede ser útil para los productores y comerciantes en términos de identificar competidores con altos puntajes y regiones en donde podría ser valioso invertir en la mejora de la calidad.

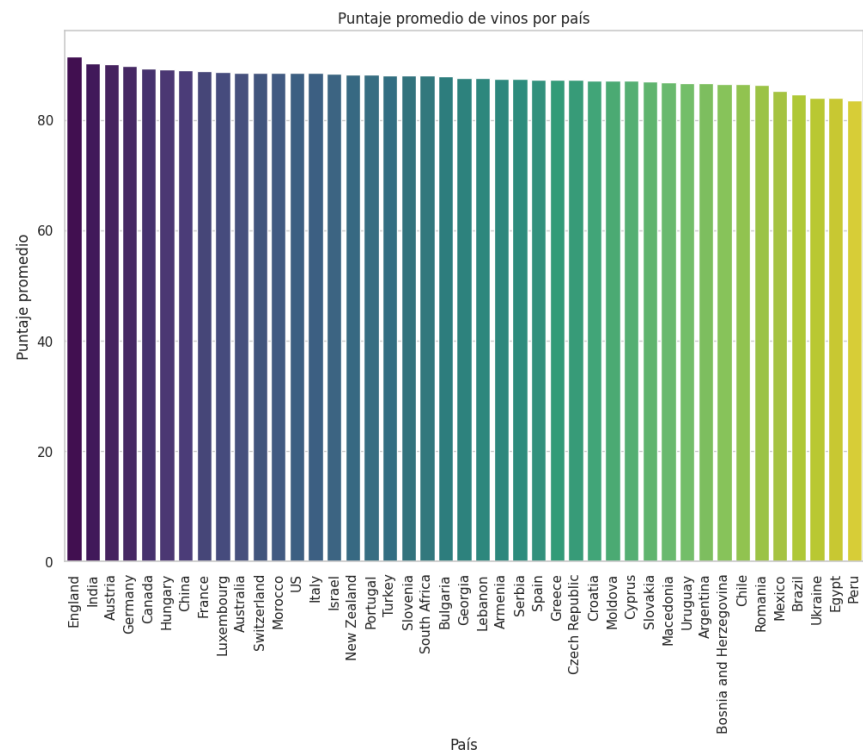
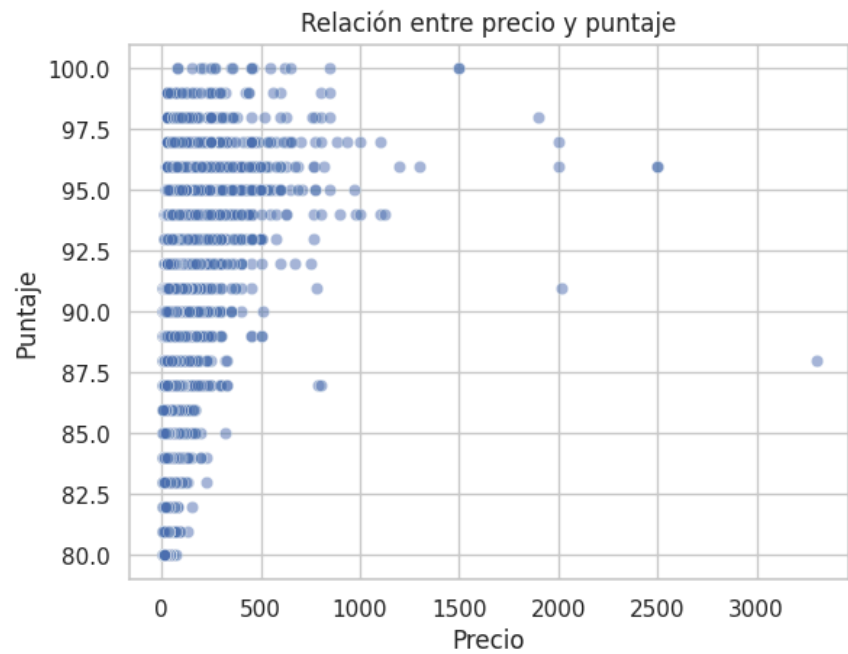


Gráfico de relación entre precio y puntaje

En el gráfico de dispersión, observamos la relación entre el precio y el puntaje de los vinos. Algunos puntos clave son:

- Tendencia general: La mayoría de los vinos se encuentran en el rango de precios bajos a medianos (0 - 500), con puntajes elevados, lo que sugiere que muchos vinos económicos logran buena calidad.
- Valores atípicos: Existen algunos vinos con precios superiores a 1000 y 2000, con puntajes altos, pero no parece haber una relación directa entre precio y puntaje.
- Oportunidades de mercado: Los productores podrían evaluar si aumentar el precio de algunos vinos con buen puntaje (y precio bajo) podría ser una estrategia viable.

Este gráfico es útil para comprender cómo el precio afecta la percepción de calidad y si existen oportunidades para ajustar precios según los puntajes.



*Gráfico de puntaje promedio de vinos por
año*

Este gráfico de línea muestra la evolución de los puntajes promedio de los vinos a lo largo del tiempo. Algunos puntos importantes:

- Tendencias de calidad: Observamos fluctuaciones en la calidad del vino, con picos de puntaje en varios puntos a lo largo de los años. Esto puede reflejar cambios en técnicas de producción o preferencias de los consumidores.
- Épocas destacadas: Los puntajes altos en décadas pasadas, como alrededor de 1950, podrían indicar épocas de alta calidad en producción vinícola.
- Decadencia en años recientes: La caída en los puntajes en los años más recientes podría señalar la necesidad de revisar prácticas actuales o adaptarse mejor a las demandas modernas.

Mundo del Vino

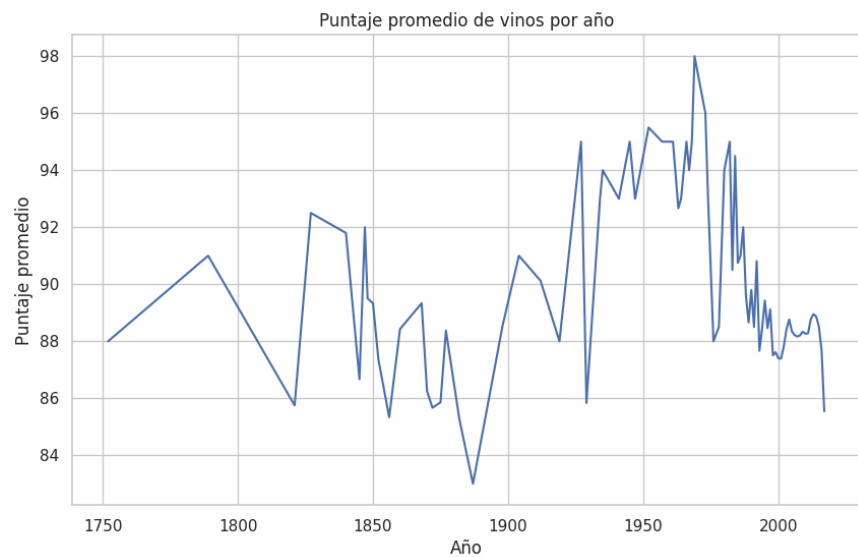


Gráfico de dispersión: relación entre puntaje y precio

- Descripción: Este gráfico muestra cómo se relacionan los puntajes de calidad con los precios de los vinos.
- Interpretación:
 - Correlación positiva: Una relación ascendente entre precio y puntaje indicaría que los

consumidores valoran la calidad y están dispuestos a pagar más por vinos bien evaluados.

- Vinos sobrevalorados/subvalorados:

Algunos puntos fuera de la tendencia podrían representar vinos con precios demasiado altos o bajos en comparación con su calidad percibida, lo que ofrece oportunidades para ajustes de mercado.

- Decisiones de producción: Los

productores podrían utilizar esta información para fijar precios competitivos y posicionar sus vinos en segmentos específicos.

Mundo del Vino

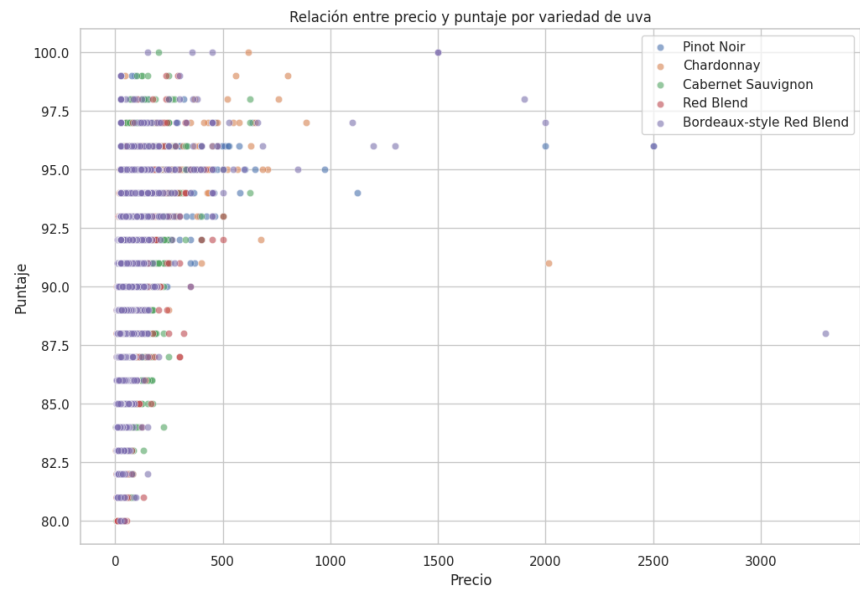


Gráfico Cantidad de vinos producidos por región

Este gráfico (barras, pastel u otro) analiza la cantidad de vinos producidos según diferentes regiones.

Interpretación:

1. Concentración de la producción:

Algunas regiones destacan como líderes en producción de vinos. Estas áreas probablemente cuentan con una combinación de factores favorables, como clima, suelo y tradición vinícola.

2. Diferenciación regional: Las

regiones con menor producción podrían centrarse en vinos boutique o de nicho, ofreciendo una oportunidad para diferenciarse en calidad y exclusividad.

3. Oportunidades de expansión: Si hay

regiones emergentes con una producción creciente, esto puede indicar nuevas áreas con potencial competitivo.

Relevancia para productores y comerciantes:

- Los productores de regiones menos

reconocidas pueden desarrollar estrategias para posicionar su oferta de manera diferenciada.

- Los comerciantes pueden utilizar esta información para diversificar su portafolio y ofrecer vinos de regiones no tradicionales.

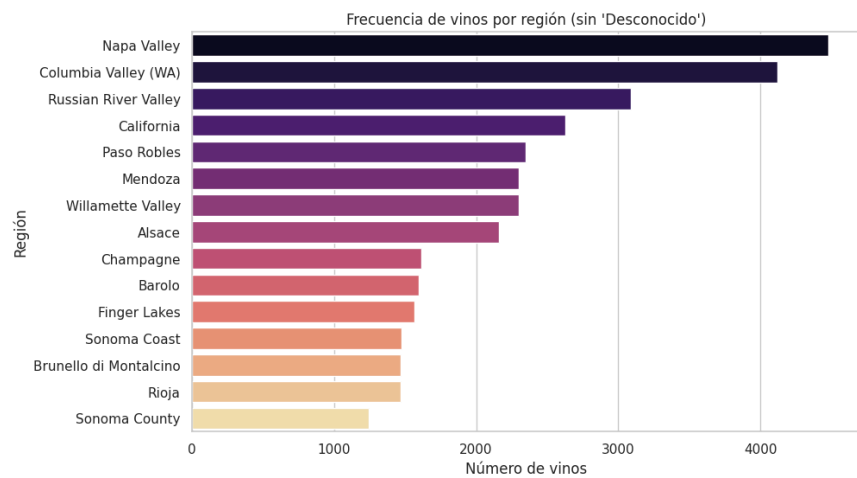


Gráfico 3: Relación entre precio y puntaje de los vinos

Este gráfico (posiblemente un diagrama de dispersión) evalúa cómo se relacionan los puntajes con los precios de los vinos.

Interpretación:

1. Relación calidad-precio: En muchos casos, vinos con puntajes altos tienen precios elevados, lo cual refleja una percepción de mayor valor por parte de los consumidores.

2. Vinos de alta calidad a precios accesibles: Si hay puntos que destacan por tener altos puntajes pero precios relativamente bajos, podrían representar grandes oportunidades de mercado.

3. Segmentación de mercado: Este gráfico puede ayudar a identificar segmentos de precio según calidad, útil para posicionar productos en diferentes nichos.

Relevancia para productores y comerciantes:

- Los productores pueden usar esta relación para fijar precios más competitivos o justificar valores más altos.
- Los comerciantes pueden destacar los vinos que ofrecen mejor calidad en relación a su

precio, atrayendo consumidores conscientes del valor.

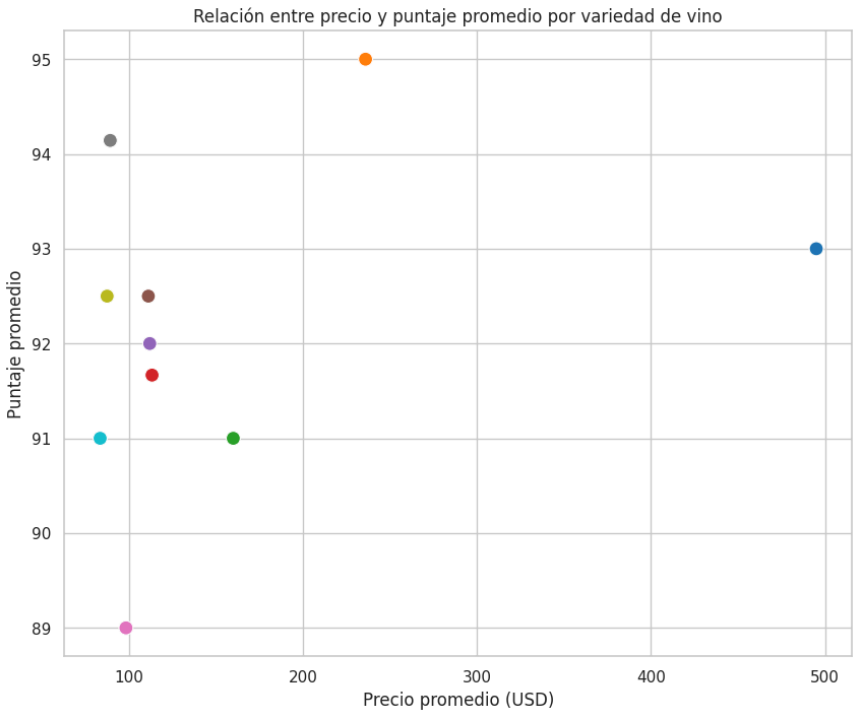


Gráfico de Top 10 variedades de vino por puntaje promedio

Interpretación:

- Este gráfico muestra las diez variedades de vino mejor calificadas, según sus puntajes promedio.
- Tinta del País encabeza la lista, seguida de variedades como Gelber Traminer y Terrantez, indicando su destacada calidad percibida.
- Las puntuaciones promedio altas sugieren que estas variedades son ideales para promocionar entre consumidores que buscan vinos premium.
- Para los productores, estas variedades representan una oportunidad para enfocarse en uvas de alto prestigio, y para los comerciantes, una guía para priorizar su inventario.

Conclusión clave:

Promover las variedades mejor calificadas como sinónimo de calidad puede aumentar la

percepción del valor del vino y generar mayores márgenes de ganancia.

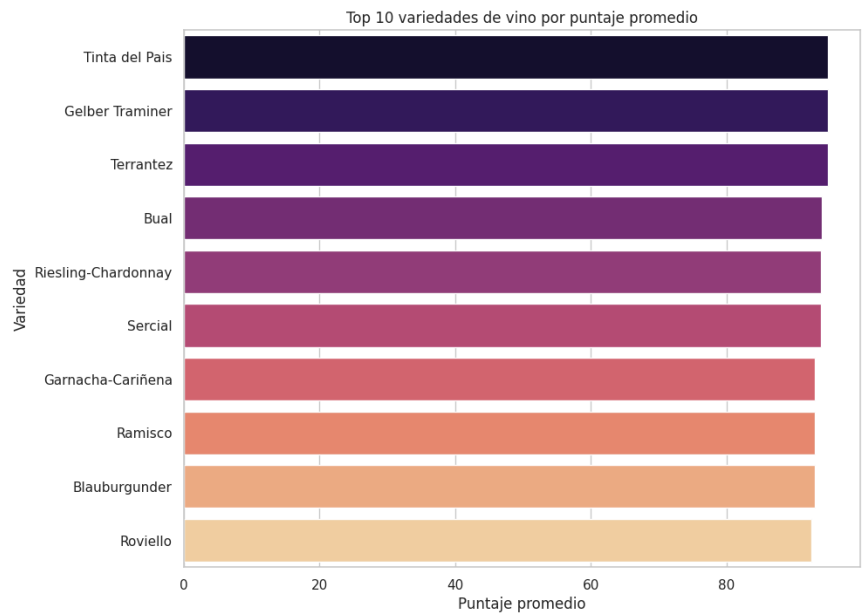


Gráfico de precio promedio de vinos por país

Interpretación:

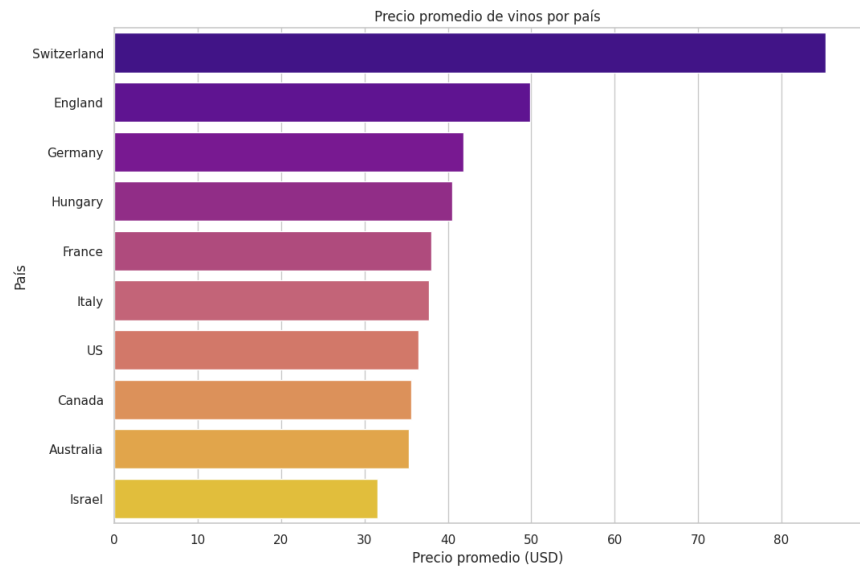
- Suiza lidera en el precio promedio de sus vinos, superando significativamente a otros países como Inglaterra, Alemania y Hungría.
- Los países europeos dominan los precios altos, reflejando su reputación en el mercado global del vino.
- La diferencia de precios podría estar relacionada con factores como costos de producción, reputación histórica y posicionamiento de mercado.
- Los vinos de Italia, Francia y EE. UU., aunque más económicos que los suizos, representan una opción de calidad a precios más accesibles.

Conclusión clave:

Los comerciantes pueden usar esta información para ajustar su estrategia de precios según el origen del vino, mientras que los productores en países menos caros podrían destacar

Mundo del Vino

la relación calidad-precio como una ventaja competitiva.



*Gráfico de distribución de precios de vinos
con 95+ puntos*

Interpretación:

- La mayoría de los vinos altamente calificados (95 puntos o más) se encuentran en el

rango de precios de \$0 a \$100 USD, indicando accesibilidad relativa para vinos de alta calidad.

- Existe una disminución gradual en la frecuencia a medida que los precios aumentan, con pocos vinos en el rango de \$300 a \$500.
- Esto sugiere que los consumidores están dispuestos a pagar más por vinos excepcionales, pero hay una barrera de entrada en precios muy altos.

Conclusión clave:

Para productores, ofrecer vinos de alta calidad en el rango de \$50 a \$100 podría maximizar las ventas. Los comerciantes pueden destacar estos vinos como opciones premium accesibles para

