

ÍNDICE

1. Objetivo	2
2. Consignas	2
2.1. Selección y depuración de datos:	2
2.2. Contextualización y preguntas clave:	3
2.3. Creación de visualizaciones efectivas:	3
2.4. Distribución del Precio por Categorías	8
2.5. Diversidad de Variedades por País	9
3. Conclusión	9

1. Objetivo

El objetivo del presente trabajo es poder explorar y aplicar técnicas avanzadas de visualización de datos con python utilizando el conjunto de datos "[winemag-data-130k-v2.csv](#)". A su vez, este análisis pretende identificar patrones, tendencias y relaciones significativas en las reseñas de vinos aportando herramientas clave para productores y comerciantes en la industria vinícola.

Propósito general:

- Proporcionar información útil sobre las características de los vinos, sus precios, regiones y otras variables relevantes.
- Facilitar la toma de decisiones basadas en los datos para mejorar estrategias de marketing y producción.

2. Consignas

2.1. Selección y depuración de datos:

El primer paso fue realizar la carga de las librerías y la carga del dataset "[winemag-data-130k-v2.csv](#)". Posterior a la carga se realizó la depuración de los datos en línea con la entrega del TP Número 2 replicando el proceso de limpieza e identificación de los valores faltantes. Utilicé la función [isnull\(\).sum\(\)](#) identificando los valores nulos en el conjunto de datos. Tomé la decisión de rellenar estos valores faltantes en lugar de eliminarlos, dado que mantener el mayor número de registros es crucial para garantizar un análisis representativo. Para las columnas categóricas como "country" y "province", los valores faltantes fueron reemplazados por "Sin Información" usando [fillna\(\)](#). En el caso de las variables numéricas, como "price" y "points", reemplacé los valores faltantes por 0.

Por último, se aseguró con la función [pd.to_numeric](#) que las variables "price" y "points" fueran tratadas como variables numéricas. Un paso necesario para la correcta visualización y análisis de estas variables en los gráficos y resúmenes estadísticos.

En estos casos opté por rellenar los valores faltantes porque las columnas afectadas contenían una proporción pequeña de valores faltantes y consideré que su impacto en los análisis posteriores no sería significativo.

2.2. Contextualización y preguntas clave:

Las preguntas clave que guiaron este trabajo práctico se encuentran a continuación:

1. **¿Cuáles son los términos más frecuentes en las descripciones de los vinos?** Este análisis ofrece a comerciantes y productores información valiosa sobre las características más valoradas por los consumidores, facilitando la adaptación de estrategias de marketing.
2. **¿Cuáles son los países con más reseñas?** Identificar los mercados claves en términos de producción y consumo
3. **¿Qué variedades de vinos tienen las mejores puntuaciones promedio?** Destacar las variedades de mayor calidad para priorizar su comercialización.
4. **¿Existen relaciones entre las variables Points (puntos), Price (precio) y otras características?** Se busca poder evaluar las relaciones para entender si el precio influye en la percepción de calidad.
5. **¿Cómo varían las puntuaciones promedio de los vinos entre las principales regiones productoras?** Identificar regiones de alta calidad productora.
6. **¿Cómo se distribuyen los precios de los vinos en diferentes categorías de puntuación?** Responder a la necesidad de comprender la relación de precios y calidad de los vinos.
7. **¿Qué países producen la mayor diversidad de variedades de vino?** Permite identificar mercados con amplia gama de productos o por el contrario aquellos con oportunidades para diversificar.

2.3. Creación de visualizaciones efectivas:

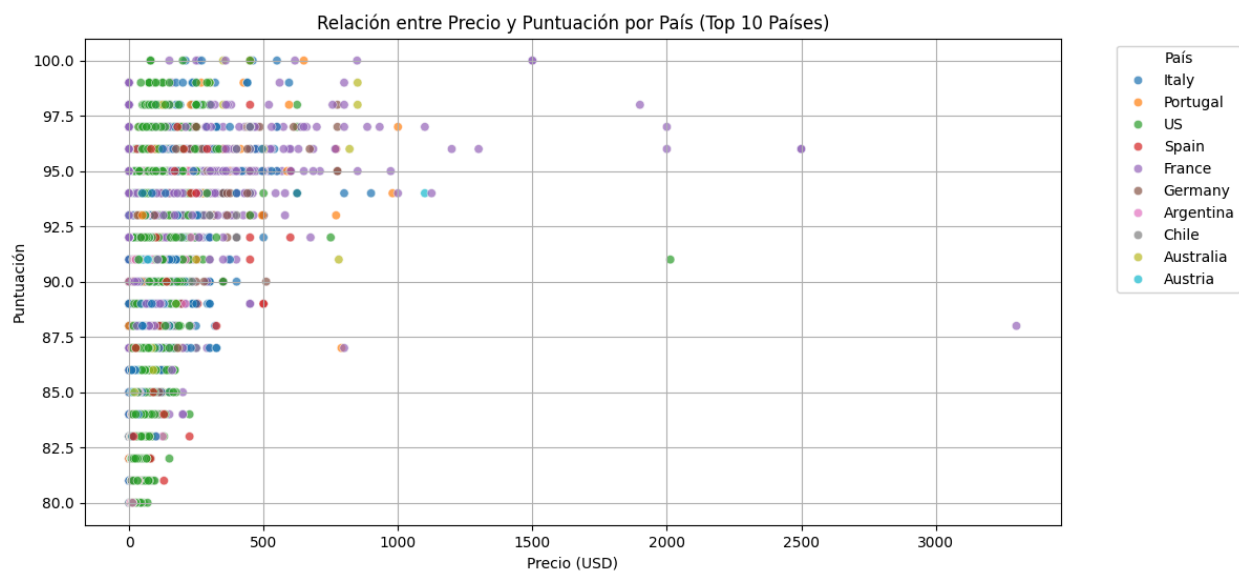
2.3.1. Términos más comunes en las descripciones

Por medio del uso de `from wordcloud import WordCloud, STOPWORDS` generé una nube de palabras utilizando la variable description (Descripción). Esta visualización permitió identificarlos términos más frecuentes en las reseñas. Se incluyeron palabras

mercado. Excluir términos genéricos (como "wine") garantiza que la nube esté enfocada en características clave.

2.3.2. Visualización: Relación entre precio y puntuación por país

Este gráfico muestra cómo varían las puntuaciones y los precios entre los países principales (Top10) identificando países con puntajes de vinos altos a precios accesibles, ofreciendo información para estrategias para la clasificación y selección de productos.

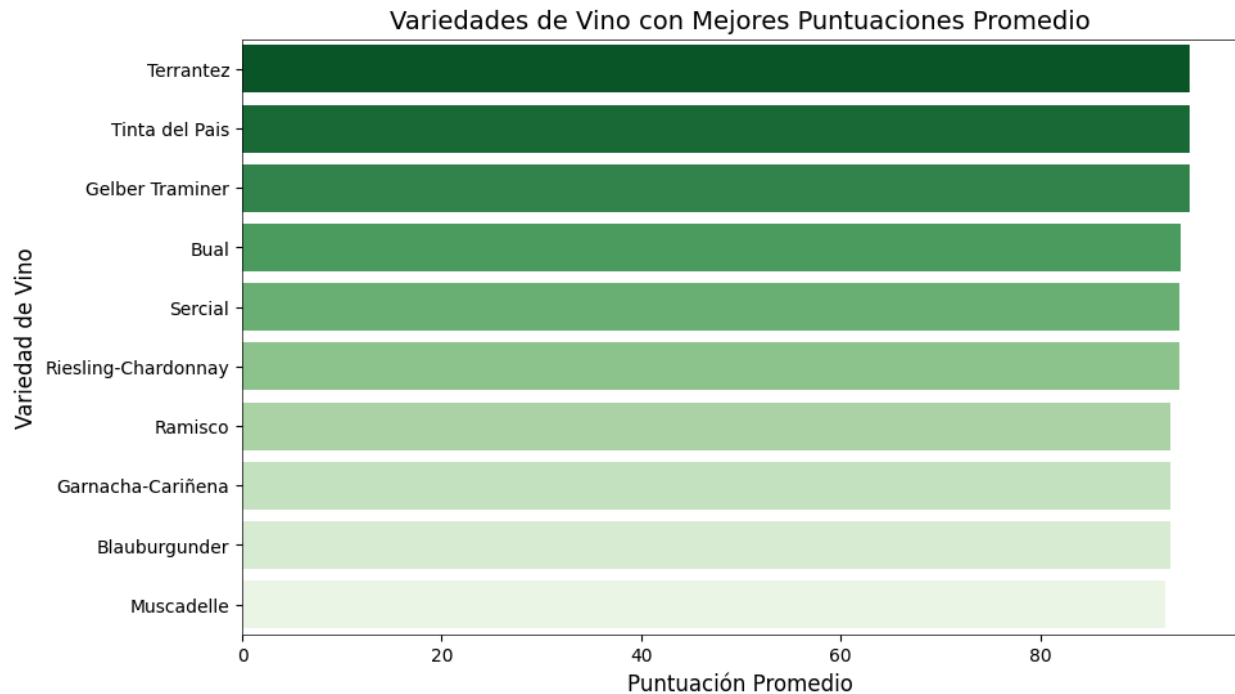


Países como el caso de US (Estados Unidos) ofreciendo vinos muy bien puntuados a precios accesibles o bien identificando los vinos más puntuados provenientes de US, Italia y Francia. Entre ellos Italia con más de 40.000 reseñas evidencia una amplia oferta y diversidad, ayudando a pensar en priorizar vinos italianos en el inventario o stock sería una buena estrategia para los comerciantes. Id

Identificar estos datos puede generar oportunidades a desarrollar en los mercados para los comerciantes como el ejemplo nombrado en el párrafo anterior y en el caso de los productores para poder determinar a dónde enfocar en las distintas regiones y saber cuál o cuáles serán su competencia.

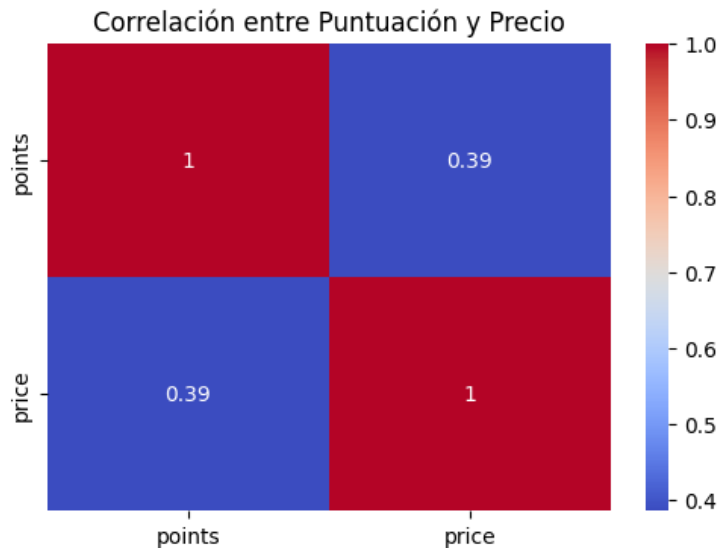
2.3.3. Variedad de vino con mejores puntuaciones promedio

Se generó un gráfico de barras que muestra las variedades mejor puntuadas, indicando preferencias del mercado por calidad tales como “Terrantez”, “Tinta del país” y “Gelber Traminer” como las principales variedades de vinos permite a productores enfocar su oferta en vinos de mayor calidad. Esto permite a bodegas optimizar su oferta y a comerciantes priorizar variedades con alta aceptación.



2.3.4. Correlación entre variables Precio y Puntaje

Generé un mapa de calor de correlación utilizando la función [sns.heatmap](#) para poder evaluar la relación entre las puntuaciones “Points” y el precio “Price” mostrando los coeficientes de correlación entre ambas variables, proporcionando información sobre si los vinos caros tienden a ser mejor valorados.



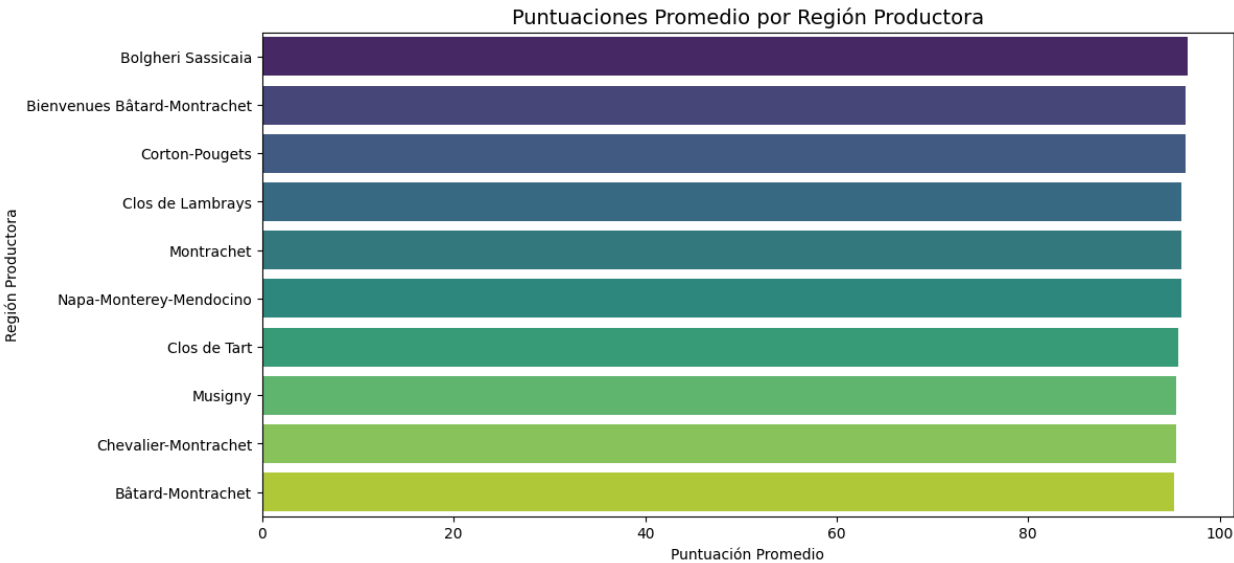
La correlación oscila entre -1 y +1. Los valores más cercanos a cero (0) significa que no hay tendencia lineal entre las dos variables. Mientras más cercana a 1 sea la correlación, más positivamente correlacionadas estarán; es decir, a medida que aumenta una, aumenta la otra, y cuanto más cercana a 1, más fuerte es esta relación. Una correlación cercana a -1 es similar, pero en lugar de aumentar ambas, una variable disminuirá a medida que aumenta la otra. Las diagonales son todas 1/verde oscuro porque esos cuadrados están correlacionando cada variable consigo misma (por lo que es una correlación perfecta).

En términos sencillos, este gráfico nos dice lo estrechamente relacionados que están el precio y las puntuaciones de los vinos. El cuadrado azul oscuro, significa que los vinos mejor valorados tienden a ser más caros, y viceversa.

Una correlación positiva indica que los vinos mejor puntuados tienden a ser más caros. Este análisis permite entender cómo el precio afecta la percepción de calidad en el mercado.

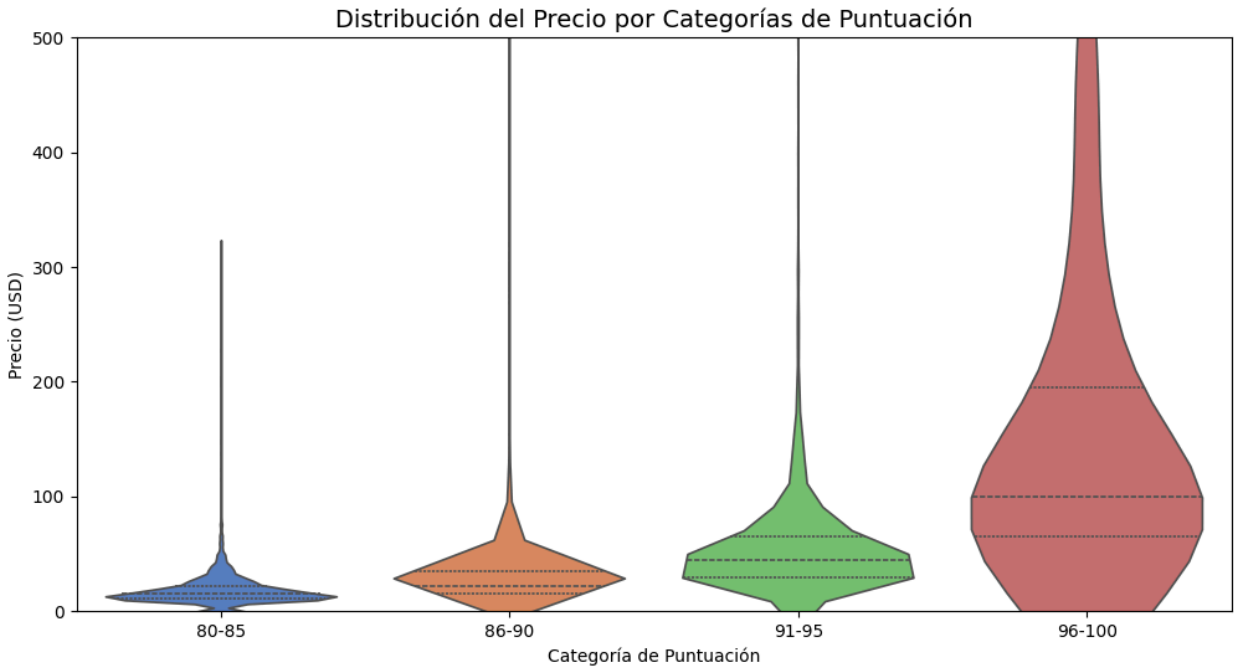
2.3.5. Puntuaciones promedio por región productora

Identifica regiones productoras con puntuaciones promedio más altas. Este gráfico ayuda a destacar regiones que producen vinos de alta calidad, facilitando decisiones estratégicas sobre importación y distribución.



2.4. Distribución del Precio por Categorías

Este gráfico realizado representa la densidad de precios de vino por categoría de puntuación lo que permite a los comerciantes identificar rangos de precios típicos para cada nivel de categoría o de calidad. En el caso de los productores a posicionar sus productos dentro de un rango de precio competitivo y alineado con la puntuación:

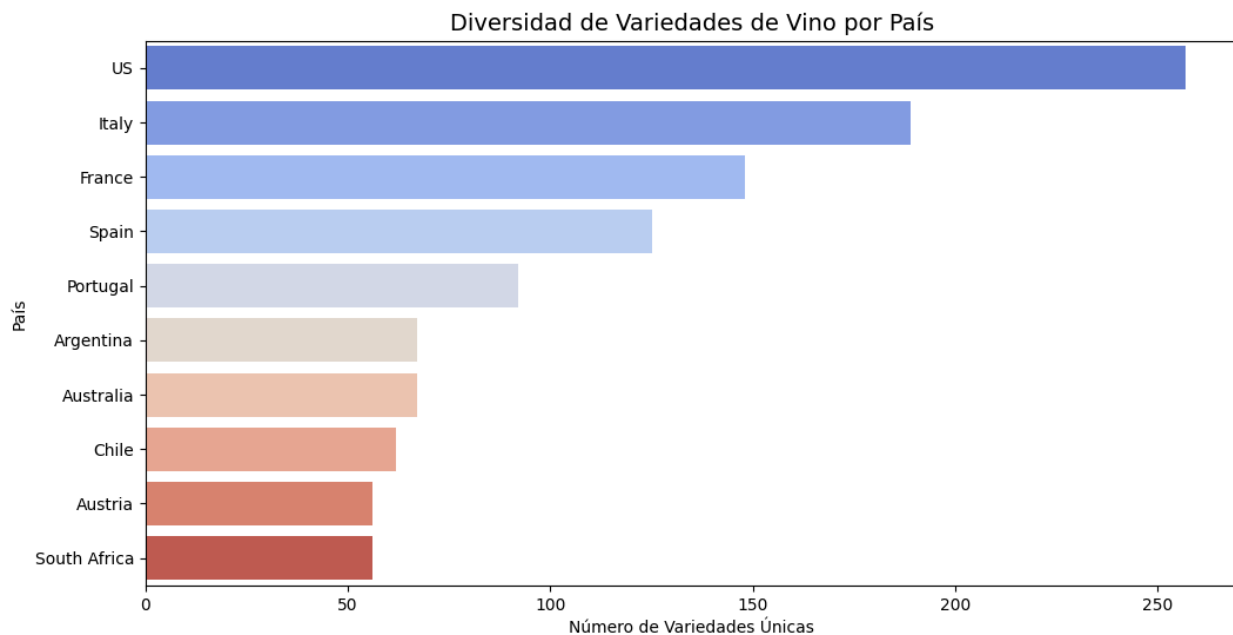


Este gráfico permite identificar valores según categorías como por ejemplo la diferencia entre los vinos en la categoría de 96-100 puntos donde tienen precios

promedio superiores a \$100 USD, mientras que los de 91-95 puntos muestran un rango de \$30-\$90 USD.

2.5. Diversidad de Variedades por País

Se generó el gráfico que muestra al Top 10 de países con mayor densidad de vinos en relación a los distintos tipos de uvas que producen. Comerciantes pueden priorizar países con mayor densidad para ofrecer una gama amplia de opciones, mientras que productores pueden identificar oportunidades para expandir su oferta competir con otros productores:



Los comerciantes podrían priorizar los vinos de US, Italia y Francia entre los más destacados con flexibilidad y la riqueza que los caracteriza.

3. Conclusión

A través de este análisis, hemos logrado realizar un análisis profundo y exhaustivo del dataset proporcionando una comprensión integral de las reseñas y sus implicancias en la industria del vino. Por medio de técnicas avanzadas de visualización de los datos en python se identificaron patrones, relaciones y tendencias con información clave para productores y comerciantes.

Resultados claves en resumen de lo analizado:

1. **Marketing:** Resaltar características comunes en campañas puede atraer a consumidores que valoran estas cualidades.
2. **Selección de productos:** Priorizar variedades y regiones con altas puntuaciones promedio garantiza una oferta competitiva.
3. **Estrategias de precios:** Identificar segmentos donde los vinos de alta calidad tienen precios accesibles puede abrir nuevas oportunidades de mercado.

Este análisis no solo tiene como objetivo identificar la información clave, sino que a su vez el ofrecer aplicaciones prácticas para mejorar la toma de decisiones, optimizar su producción, comercialización y estrategias de marketing, alineándose con las expectativas del mercado y maximizando su impacto en un entorno competitivo.