EFIP 1 – EXAMEN FINAL INTEGRADOR 1

Visualización de Datos Aplicada al Análisis de Vinos



Alumno: Lourdes Ayelen Sanchez Alfaro

Profesor: Pablo Alejandro Virgolini

Fecha: 29/06/2025

## Índice

1. Introducción
2. Metodología de análisis
3. Preparación y limpieza del dataset
4. Desarrollo del análisis visual
   1. Top 10 países por puntuación promedio
   2. Relación entre precio y puntuación
   3. Violin plot por variedad
   4. Distribución de precios
   5. Nube de palabras
   6. Densidad precio/puntuación (hexbin)
   7. Variedades con mejor relación calidad/precio
   8. Países con más vinos premium
   9. Distribución de reseñas por país
   10. Boxplot de puntuaciones por país
   11. Mapa geográfico de puntuación promedio
5. Conclusiones y recomendaciones
6. Referencias bibliográfica

## 1. Introducción

En el marco del Entregable 4 del EFIP 1 de la Licenciatura en Ciencia de Datos, desarrollé un análisis visual aplicado al estudio de reseñas de vinos recopiladas por el portal *Wine Enthusiast*. Este conjunto de datos, conformado por más de 129.000 registros en su versión original, ofrece una valiosa fuente de información sobre las características de los vinos, sus puntuaciones, precios, países de origen, variedades de uva y descripciones detalladas.

Como estudiante de tercer año, abordar este análisis representó la oportunidad de aplicar conceptos teóricos fundamentales de la disciplina, especialmente aquellos vinculados al análisis exploratorio de datos (EDA) y a las técnicas de visualización avanzada. La tarea no solo consistió en generar gráficos, sino en diseñarlos a partir de un criterio analítico sólido, guiado por principios de claridad, relevancia, legibilidad y propósito comunicativo. Cada visualización fue seleccionada con la intención de descubrir y comunicar patrones que no son evidentes a simple vista, y de brindar información útil para la toma de decisiones dentro del sector vitivinícola.

A través de herramientas como pandas, seaborn, matplotlib, plotly y wordcloud, el análisis se llevó a cabo con un enfoque estructurado, que incluyó la limpieza del dataset, la depuración de registros incompletos, la transformación de tipos de datos y la construcción de gráficos significativos y exportables. Si bien el código desarrollado se presenta en un archivo separado, este informe se enfoca en el proceso de razonamiento analítico y teórico que sustentó cada paso del análisis.

Desde una perspectiva aplicada, este trabajo busca evidenciar cómo la ciencia de datos puede facilitar la comprensión de fenómenos complejos en contextos reales. En este caso, el mundo del vino presenta no solo un producto, sino una experiencia sensorial, cultural y económica, cuya representación en datos demanda un abordaje tanto técnico como interpretativo. A lo largo del informe, se vinculan las decisiones tomadas durante el análisis con fundamentos teóricos pertinentes, considerando siempre el contexto del dataset y sus implicancias para distintos actores de la industria (productores, comerciantes, sommeliers y consumidores).

## 2. Fundamentos teóricos

El análisis exploratorio de datos (EDA, por sus siglas en inglés) es una de las etapas más importantes dentro del ciclo de vida de un proyecto de ciencia de datos. Tal como lo plantea John Tukey, uno de sus principales referentes, el EDA permite no solo describir los datos sino también formular hipótesis y descubrir relaciones inesperadas mediante visualizaciones adecuadas. Su objetivo no es simplemente obtener métricas descriptivas, sinocomprender profundamente la estructura, los patrones y las particularidades del conjunto de datos en estudio.

En este trabajo, el EDA se aplicó con el propósito de:

* Identificar patrones generales (por ejemplo, qué países se destacan en calidad de vinos),
* Detectar valores extremos o atípicos (vinos con precios o puntuaciones inusuales),
* Explorar la distribución de variables clave como el precio o la puntuación,
* Descubrir relaciones entre atributos numéricos y categóricos.

### 2.1 Visualización de datos

De acuerdo con autores como Edward Tufte, Alberto Cairo y Stephen Few, la visualización de datos debe cumplir con criterios de claridad, precisión, relevancia y propósito comunicativo. Un gráfico no se elige por estética, sino por su capacidad de facilitar la comprensión de una relación o comportamiento. Por ello, en este análisis cada visualización fue seleccionada según los principios teóricos de diseño gráfico aplicado a datos.

Se utilizaron diferentes tipos de gráficos según su adecuación al análisis:

* **Gráficos de barras** para comparar valores promedio en categorías (por país o por variedad).
* **Gráficos de dispersión y hexbin** para explorar correlaciones entre variables continuas como precio y puntuación.
* **Violin plots** para observar la distribución interna de puntuaciones dentro de cada variedad de vino.
* **Histogramas** para visualizar la distribución de precios y detectar concentración o dispersión.
* **Nubes de palabras** como herramienta de análisis semántico visual, útil para identificar términos clave en las descripciones.
* **Gráficos de torta** para representar proporciones entre categorías con pocas clases.
* **Mapas coropléticos** para representar promedios geográficos de calidad a nivel país.

Además, la elección de cada visualización respondió al tipo de variable analizada. Por ejemplo, points y price, al ser variables numéricas continuas, fueron representadas mediante gráficos de dispersión y hexbin, que permiten analizar su correlación y concentración. En cambio, variables categóricas como country o variety fueron visualizadas con barras y violines, lo cual se ajusta a las buenas prácticas teóricas para representar distribución o comparación entre categorías.

### 2.2 Análisis textual y representaciones semánticas

Una característica interesante del dataset es la inclusión de descripciones textuales escritas por sommeliers. Este tipo de variable no estructurada permite aplicar técnicas básicas de análisis textual. En este caso, se optó por la generación de una nube de palabras, una técnica simple pero efectiva para visualizar los términos más frecuentes.

Este enfoque se basa en la frecuencia absoluta de palabras, sin necesidad de procesamiento lingüístico complejo. Aun así, ofrece una primera aproximación valiosa para entender qué atributos son más mencionados en la evaluación de los vinos (por ejemplo: *fruit*, *aroma*, *acidity*, *tannin*, *flavor*), lo cual tiene aplicación directa en marketing, posicionamiento sensorial y diseño de producto.

### 2.3 Visualización geográfica

Los mapas coropléticos permiten representar información agregada sobre regiones geográficas. En el caso del dataset de vinos, la variable country permitió construir un mapa de puntuaciones promedio por país. Esta visualización espacial aporta una perspectiva global sobre la calidad percibida de los vinos según su origen, revelando diferencias regionales que pueden ser relevantes tanto desde el punto de vista del consumidor como de los productores.

Desde la teoría, este tipo de representación es muy utilizada en análisis de mercado, logística y segmentación regional. En este caso se generó mediante la librería plotly.express, que permite una presentación clara, interactiva y exportable.

### 2.4 Automatización y reproducibilidad

La creación de análisis replicables y reutilizables es uno de los principios de la ciencia de datos profesional. En este proyecto se incorporó una función personalizada llamada guardar\_y\_mostrar(), que permite automatizar la generación y el guardado de gráficos en formato .png con buena calidad. Esta decisión se enmarca en una lógica de eficiencia, trazabilidad y profesionalismo, ya que facilita el trabajo colaborativo, la generación de informes y la auditoría del análisis.

## 3. Enfoque metodológico

El desarrollo del análisis se basó en una estructura ordenada y lógica que me permitió aplicar lo aprendido durante la carrera, especialmente en temas de limpieza de datos, visualización y criterios analíticos. Aunque el script completo se entrega por separado, en esta sección explico por qué tomé cada decisión y cómo se relaciona con los objetivos del trabajo y con la teoría vista en la materia.

### 3.1 Carga y limpieza de los datos

Uno de los primeros desafíos fue que el archivo casoefip.csv no estaba en un formato clásico. Toda la información aparecía en una sola columna, por lo que fue necesario separarla correctamente usando el delimitador coma. Esto implicó un paso extra de codificación, ya que no se podía aplicar read\_csv() con encabezados como se hace normalmente. Este tipo de problema es bastante común en la práctica profesional, y forma parte de lo que se conoce como preprocesamiento o limpieza de datos, una etapa fundamental en cualquier análisis.

Después de separar correctamente las columnas, revisé los tipos de datos. Las variables numéricas como points (puntos) y price (precio) venían como texto, así que las convertí a valores numéricos usando pd.to\_numeric, controlando errores en el proceso. También decidí eliminar filas con valores faltantes en las columnas clave. Si bien hay otras estrategias como completar con la media o usar interpolaciones, opté por eliminar esos registros para asegurar que los gráficos reflejen información completa y confiable.

### 3.2 Elección de variables para analizar

El dataset incluye muchas columnas, pero no todas son relevantes para este análisis. Me enfoqué en aquellas que me permitieran responder preguntas concretas sobre el vino, como:

* ¿Qué países producen vinos de mejor calidad?
* ¿Existe una relación entre el precio y la puntuación?
* ¿Qué variedades de uva se destacan?
* ¿Qué lenguaje usan los sommeliers en las descripciones?

Por eso seleccioné las columnas country, points, price, variety, description y algunas adicionales como province y winery para posibles segmentaciones. Esta selección estuvo basada en criterios de relevancia, completitud y utilidad visual, algo que es parte del enfoque exploratorio que se enseña en teoría.

### 3.3 Organización del análisis y estructura del script

Decidí dividir el trabajo en etapas, como se recomienda en la práctica profesional:

1. Importación de librerías necesarias para visualización, análisis y texto.
2. Carga y depuración de datos, con conversión de tipos y eliminación de valores nulos.
3. Exploración general, con describe() y value\_counts() para entender el contexto.
4. Visualización de datos, con cada gráfico agrupado por sección y guardado como imagen.
5. Resumen de métricas generales, como promedio de puntuación y precio total.

Esta estructura modular me permitió tener mayor control del análisis y también facilitar su lectura y reutilización.

### 3.4 Automatización del guardado de gráficos

Como uno de los objetivos era exportar las visualizaciones, creé una función llamada guardar\_y\_mostrar() que resume tres pasos en uno: ajustar el diseño del gráfico, guardarlo como imagen .png, y mostrarlo en pantalla. Esto me permitió ahorrar líneas de código y mantener la consistencia en la calidad y formato de los gráficos.

En vez de tener que escribir manualmente savefig() para cada gráfico, simplemente llamaba a la función pasando el nombre del archivo. Esta decisión no solo mejoró el flujo de trabajo, sino que también responde al criterio de eficiencia, reutilización y claridad, que son principios centrales en la ciencia de datos.

### 3.5 Visualización geográfica

Como el dataset incluía la columna country, me pareció importante agregar un mapa para ver las puntuaciones promedio de los vinos según su país de origen. Usé plotly.express porque permite crear mapas de tipo coroplético de forma interactiva y profesional. Además, investigué cómo guardar ese gráfico como imagen estática utilizando fig.write\_image(), lo cual requería instalar el paquete kaleido.

Esta parte fue muy útil para aplicar una visualización espacial, que es algo que en teoría tiene mucho valor para interpretar patrones regionales o hacer análisis de mercado según ubicación.

4. Desarrollo del análisis visual

4.1 Top 10 países por puntuación promedio

¿Por qué este gráfico?

Consideré fundamental comenzar el análisis identificando qué países se destacan por la calidad de sus vinos, medida a través del puntaje promedio asignado por Wine Enthusiast. Esta variable representa una medida estandarizada de evaluación, que funciona como un proxy de calidad percibida. Desde la teoría, los gráficos de barras son la mejor opción cuando se busca comparar valores agregados entre categorías nominales, como los países de origen.

¿Qué muestra?

Este gráfico de barras verticales representa los 10 países con mayor puntuación promedio dentro del conjunto de datos analizado. La altura de cada barra refleja la media de los puntajes (points) asignados a los vinos de cada país.

¿Qué puedo interpretar?

Los resultados muestran que países con tradición vinícola como Francia, Italia y Portugal efectivamente figuran entre los más valorados, pero también emergen con buen rendimiento otros países como Argentina o España. Esto sugiere una competencia internacional equilibrada, y refleja la calidad creciente de productores del Nuevo Mundo.

¿Qué decisiones puedo tomar?

Desde un enfoque estratégico, este análisis puede ayudar a bodegas o distribuidores a realizar benchmarking internacional, comparando su rendimiento con el de otros países. Un importador podría utilizar esta información para orientar su estrategia comercial hacia países con alta reputación promedio, mientras que un productor podría usarla para identificar su posición relativa en el mercado global.

4.2 Relación entre precio y puntuación

¿Por qué este gráfico?

Una de las preguntas más comunes en el mundo del vino es si mayor precio implica mayor calidad. Elegí un gráfico de dispersión porque es la herramienta teórica más adecuada para visualizar relaciones entre dos variables numéricas continuas, como lo son price y points. Además, apliqué una escala logarítmica para el eje de precios, dado que hay mucha variabilidad y algunos valores extremos.

¿Qué muestra?

Cada punto representa un vino. Su posición horizontal indica su precio (log-scale) y la vertical, su puntuación. El gráfico permite observar si existe una correlación entre ambas variables.

¿Qué puedo interpretar?

Se observan vinos con precios bajos que reciben puntajes altos, lo cual indica que la calidad no siempre está ligada al costo. A su vez, algunos vinos caros obtienen puntajes relativamente bajos, lo cual puede ser indicio de posicionamiento de marca más que de valor real percibido.

¿Qué decisiones puedo tomar?

Este gráfico es útil para comerciantes o distribuidores que buscan vinos de alta calidad a bajo precio. También puede servir para diseñar estrategias de precios más eficientes, y para consumidores que quieren obtener el mejor valor por su dinero.

4.3 Violin plot de puntuaciones por variedad (Top 10)

¿Por qué este gráfico?

Los gráficos de violín permiten visualizar la distribución de los datos dentro de cada categoría, mostrando tanto la densidad como los rangos y valores extremos. Los elegí para analizar cómo varían las puntuaciones dentro de las 10 variedades de uva más frecuentes, en vez de limitarme a mostrar el promedio.

¿Qué muestra?

Cada “violín” representa una variedad de vino, y muestra cómo se distribuyen los puntajes dentro de esa categoría. Cuanto más ancho, mayor densidad. También permite ver la mediana y la presencia de outliers.

¿Qué puedo interpretar?

Variedades como Pinot Noir o Riesling presentan una mayor dispersión, lo que sugiere que pueden haber tanto vinos muy buenos como regulares bajo esa etiqueta. Otras variedades tienen una distribución más concentrada, lo que indica consistencia en calidad.

¿Qué decisiones puedo tomar?

Para un productor, puede ser útil elegir variedades que tengan menor dispersión si busca mantener estándares constantes. Por otro lado, bodegas que quieran diferenciarse podrían enfocarse en uvas con más variabilidad, ya que ofrecen más margen para innovación o posicionamiento de alta gama.

4.4 Distribución de precios

¿Por qué este gráfico?

Quise analizar cómo se comporta el precio en el mercado del vino. Desde la teoría, los histogramas son ideales para observar la distribución de una variable continua. Usé una escala logarítmica porque los precios en este dataset varían desde vinos muy accesibles hasta opciones de lujo.

¿Qué muestra?

El gráfico muestra la frecuencia con la que aparecen distintos rangos de precio en el dataset. La gran mayoría de los vinos se concentra en la franja de 10 a 40 dólares, mientras que solo unos pocos superan los 100.

¿Qué puedo interpretar?

El mercado de vinos parece estar muy centrado en la gama media. Esto es valioso para entender cómo se comporta la oferta y qué expectativas de precios tienen los consumidores y los productores.

¿Qué decisiones puedo tomar?

Sirve para definir políticas de pricing, especialmente para bodegas que están decidiendo en qué franja posicionar su producto. También puede ayudar a diseñar gamas diferenciadas: económica, estándar y premium.

4.5 Nube de palabras de descripciones

¿Por qué este gráfico?

Quise explorar la dimensión textual del dataset, que muchas veces se deja de lado. Las descripciones que escriben los sommeliers contienen mucha información sensorial y comercial. La nube de palabras es una forma sencilla pero poderosa de identificar los términos más frecuentes en el lenguaje de las reseñas.

¿Qué muestra?

Las palabras que más aparecen en las descripciones de vinos. Algunas de las más destacadas son: fruit, acidity, tannin, aroma, flavor, entre otras.

¿Qué puedo interpretar?

Estas palabras reflejan los atributos más valorados o más comunes en la experiencia del vino. También revelan cómo se describe un vino en términos de sabor, cuerpo y aroma.

¿Qué decisiones puedo tomar?

Una bodega puede usar esta información para definir el lenguaje de sus etiquetas o campañas publicitarias. También puede adaptar su perfil sensorial a lo que los catadores tienden a destacar en sus reseñas.

4.6 Hexbin: Densidad precio vs. puntuación

¿Por qué este gráfico?

Después del gráfico de dispersión, quería mejorar la legibilidad en zonas con mucha superposición. El gráfico hexbin permite representar la densidad de puntos en una grilla hexagonal, mostrando dónde se concentran más vinos según precio y puntuación.

¿Qué muestra?

Se visualiza una alta concentración de vinos entre 85 y 90 puntos, con precios entre 10 y 30 USD. El resto del espacio queda más disperso.

¿Qué puedo interpretar?

Confirma que la mayoría del mercado se concentra en ese segmento: calidad aceptable a precio moderado. También permite detectar zonas vacías o “nichos” poco explotados.

¿Qué decisiones puedo tomar?

Una bodega puede usar esta información para posicionar estratégicamente nuevos productos. También sirve para identificar oportunidades comerciales no saturadas.

### 4.7 Mejores variedades calidad/precio

¿Por qué este gráfico?  
 Uno de los objetivos del análisis era identificar vinos que ofrezcan mayor valor relativo. Para eso, construí un indicador llamado precio\_por\_punto, dividiendo el precio promedio por la puntuación promedio de cada variedad. Cuanto más bajo es este valor, mejor la relación calidad/precio. El gráfico de barras me permitió comparar fácilmente entre variedades.

¿Qué muestra?  
 Muestra las 10 variedades de uva con mejor rendimiento en términos de relación calidad/precio, considerando solo aquellas con más de 100 registros para asegurar validez estadística.

¿Qué puedo interpretar?  
 Variedades como Bobal o Tempranillo ofrecen buenos puntajes a bajo costo. Esto puede ser útil tanto para consumidores que buscan optimizar su gasto como para empresas que quieren posicionar productos en ese segmento.

¿Qué decisiones puedo tomar?  
 Es ideal para una bodega que quiere introducir una nueva variedad con buen potencial comercial. También ayuda a diseñar promociones de vinos “inteligentes” desde lo económico, destacando su eficiencia en calidad respecto al precio.

4.8 Países con más vinos premium (≥95 puntos)

¿Por qué este gráfico?  
 Si bien el promedio es útil, también es importante saber qué países logran vinos excepcionalmente bien puntuados. Usé un gráfico de barras para visualizar la cantidad de vinos con 95 puntos o más, lo que representa el rango premium según Wine Enthusiast.

¿Qué muestra?  
 Los 10 países con mayor número de vinos en esta categoría de excelencia.

¿Qué puedo interpretar?  
 Francia y EE. UU. lideran en vinos premium, pero aparecen también países como Italia y Portugal, lo que refuerza su posicionamiento como productores de alta gama. Argentina también figura, aunque en menor cantidad, lo que igualmente resalta su potencial competitivo.

¿Qué decisiones puedo tomar?  
 Un importador premium puede enfocarse en estos países para ofrecer vinos de alta gama. Una bodega argentina, por ejemplo, puede usar esta información como argumento comercial (“nuestros vinos compiten a nivel mundial en calidad”).

### 4.9 Gráfico de torta (pie chart) de reseñas por país (Top 5)

¿Por qué este gráfico?  
 Quise visualizar cómo se distribuyen las reseñas del dataset según país. Los gráficos de torta son útiles cuando queremos representar partes de un todo, en este caso, la proporción de reseñas pertenecientes a los cinco países más relevantes.

¿Qué muestra?  
 Muestra el porcentaje de reseñas que corresponde a cada uno de los cinco países con más registros.

¿Qué puedo interpretar?  
 Estados Unidos representa una porción significativa, lo cual puede estar influido por la fuente de datos. Sin embargo, también muestra el peso de Francia, Italia y otros países en el conjunto de análisis.

¿Qué decisiones puedo tomar?  
 Sirve para tener contexto sobre el sesgo potencial del análisis (si hay sobre-representación de un país). También puede ser útil para campañas internacionales o para ajustar decisiones de segmentación de mercado.

### 4.10 Boxplot de puntuaciones por país (Top 5)

¿Por qué este gráfico?  
 El boxplot es ideal para comparar distribuciones, medianas y presencia de outliers entre categorías. Lo usé para observar cómo varían las puntuaciones dentro de los 5 países con más reseñas.

¿Qué muestra?  
 La mediana, los cuartiles y los valores extremos de la puntuación (points) en cada país. A diferencia de los promedios, acá se puede observar si un país tiene mayor dispersión, sesgo o extremos atípicos.

¿Qué puedo interpretar?  
 Por ejemplo, países como Italia y Portugal muestran distribuciones consistentes, mientras que otros tienen mayor dispersión. Esto puede estar relacionado con políticas de producción más o menos estandarizadas.

¿Qué decisiones puedo tomar?  
 Ayuda a entender qué tan uniforme es la calidad dentro de un país. Puede influir en estrategias de importación o diseño de portfolios (por ejemplo, elegir países con menor variabilidad si se busca asegurar calidad constante).

### Mapa geográfico de puntuación promedio por país

¿Por qué este gráfico?  
 Incorporé un mapa tipo choropleth porque permite analizar los datos desde una perspectiva espacial, algo cada vez más importante en visual analytics. Este tipo de representación es ideal para comparar países en el plano global.

¿Qué muestra?  
 La puntuación promedio por país, codificada por color: cuanto más intenso, mayor la puntuación. El mapa permite identificar patrones regionales, como el liderazgo de Europa occidental.

¿Qué puedo interpretar?  
 Francia, Italia y Portugal se destacan por su alta calidad promedio, pero también se observa el aporte significativo de países del Nuevo Mundo como Argentina, Chile o Sudáfrica.

¿Qué decisiones puedo tomar?  
 Muy útil para análisis de mercado geográfico, posicionamiento de marca internacional y detección de oportunidades emergentes por país.

## 5. Conclusiones generales y recomendaciones estratégicas

El análisis realizado sobre el dataset de reseñas de vinos mediante técnicas avanzadas de visualización permitió extraer conclusiones significativas, no solo desde el punto de vista descriptivo, sino también con aplicaciones directas para la toma de decisiones en la industria vitivinícola.

En primer lugar, se identificaron patrones geográficos claros: países como Francia, Italia y Portugal mantienen un alto estándar de calidad, mientras que otros como Argentina y Estados Unidos emergen como competidores sólidos, incluso en el segmento premium. Esto permite a productores locales o regionales evaluar su posicionamiento competitivo y definir estrategias de internacionalización o benchmarking.

Por otro lado, el cruce entre precio y puntuación reveló que no existe una relación lineal entre ambas variables. Esta observación, respaldada por visualizaciones como el scatter plot y el hexbin, habilita oportunidades de mercado para vinos de alta calidad a precios moderados, que pueden ser aprovechadas por distribuidores, vinotecas o campañas de marketing dirigidas al consumidor final.

El análisis por variedad, tanto en términos de puntuación como de relación calidad/precio, mostró que ciertos tipos de uva ofrecen ventajas competitivas claras. Esto es particularmente útil para bodegas en etapa de diseño de portfolio, ya que pueden elegir variedades que combinen consistencia y buena percepción sin necesidad de invertir en posicionamiento costoso.

A nivel cualitativo, la nube de palabras permitió extraer atributos sensoriales clave que se repiten en las reseñas, lo que aporta valor al diseño de etiquetas, fichas técnicas o estrategias de comunicación más alineadas con el lenguaje del consumidor y el sommelier.

Finalmente, las visualizaciones utilizadas, todas seleccionadas con fundamento teórico, cumplieron su rol de transformar datos complejos en información comprensible, accionable y estéticamente clara. Este trabajo demuestra cómo la Ciencia de Datos no solo organiza información, sino que la traduce en ventajas competitivas concretas cuando se la aplica correctamente al análisis de productos reales como el vino.

1. Bibliografia

* Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press.

Respaldó la elección de gráficos claros y comparativos como barras, tortas y boxplots.

* McKinney, W. (2018). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (2ª ed.). O'Reilly Media.

Fuente principal para el tratamiento y análisis de datos con Pandas y NumPy.

* Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2ª ed.). Graphics Press.

Fundamentó principios de claridad visual y eficacia comunicativa en los gráficos usados.

* Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.

Aunque no se usó R, sus fundamentos influyen en librerías como Seaborn y su enfoque estético.

* OpenAI. (2025). *Asistencia de IA en ciencia de datos aplicada a la industria*. OpenAI.

Herramienta de soporte para interpretación y fundamentación didáctica del análisis.