

Memoria del Proyecto: Relación entre Precio y Calidad del Vino

Introducción

El análisis exploratorio de datos (EDA) se centró en la relación entre el precio, la puntuación y la calidad-precio del vino, considerando también variedades de uva y países como factores diferenciadores. Este trabajo busca identificar patrones y tendencias que ayuden a comprender los factores que influyen en la percepción del valor de los vinos, así como determinar las características comunes de los vinos mejor valorados.

Datos Utilizados

- **Origen:** Dataset con más de 250,000 referencias de vino.
- **Variables Clave:**
 - **Precio:** Precio por botella.
 - **Puntuación:** Puntuaciones dadas por expertos (0-100).
 - **Calidad-Precio:** Relación entre la puntuación y el precio.
 - **Variedades:** Tipos de uvas utilizadas.
 - **Países:** Región de producción del vino.

Metodología

Limpieza de Datos

1. Fusión de dos bases de datos usando un "left join".
2. Eliminación de columnas irrelevantes, como **taster_name** y **taster_twitter_handle**.
3. Relleno de valores faltantes:
 - **region_2:** Rellenado con el valor de **region_1**.
 - **designation:** Etiquetado como "Desconocido".
4. Eliminación de duplicados y filas con valores nulos críticos.

Análisis

Análisis Univariante

- Exploración de distribuciones de precio y puntuación.

- Identificación de países con mayor número de variedades.

Análisis Bivariante

- Relación entre precio y puntuación promedio por país.
- Análisis de la relación calidad-precio por país.

Análisis Multivariante

- Identificación de patrones en puntuación y precio según variedades y países.
- Evaluación de las 5 variedades más caras por país.
- Análisis de los 50 vinos mejor y peor valorados por país.

Visualizaciones Clave

1. **Gráfico de Dispersión:**
 - Relación entre precio promedio y puntuación promedio por país.
 - Observación: Francia e Italia destacan por sus altos precios y puntuaciones.
2. **Gráfico de Barras:**
 - Relación calidad-precio promedio por país.
 - Argentina y España lideran en relación calidad-precio.
3. **Top 5 Variedades Más Caras:**
 - Identificación por país (Argentina, Australia, Canadá, España, Francia, Italia y EE.UU.).
4. **Gráficos de Barras Horizontales:**
 - Top 50 vinos más valorados y peor valorados por país.

Hipótesis

1. Relación entre precio, puntuación y calidad-precio

Hipótesis: Los vinos con un precio más alto tienen una mayor puntuación, pero el ratio de calidad-precio puede ser más favorable en vinos de precio medio o bajo.

2. Variedades y regiones como factores diferenciadores

Hipótesis: Las variedades de uva autóctonas de regiones vinícolas prestigiosas tendrán un precio más alto debido a su rareza y reputación, pero los vinos de otras variedades locales pueden ofrecer mejor calidad a precios más bajos.

3. Qué características comunes tienen los vinos más top

Hipótesis: Los vinos más exclusivos comparten características como un proceso de vinificación meticuloso, una puntuación elevada, y el uso de variedades de uva limitadas y de alta calidad provenientes de regiones vinícolas famosas.

6. Relación entre el precio y la ubicación geográfica

Hipótesis: Los precios de los vinos varían considerablemente según la ubicación geográfica, siendo los vinos provenientes de regiones vinícolas famosas como Burdeos y Borgoña más caros que aquellos de zonas menos reconocidas, incluso si las puntuaciones son similares.

Conclusiones

1. Relación Precio-Puntuación:

- Francia e Italia presentan los precios promedio más altos (40 €) y puntuaciones superiores (88.5 puntos).
- Argentina y España ofrecen alternativas con excelente calidad-precio (25 € y 87 puntos).

2. Variedades y Regiones:

- Variedades históricas, como Pinot Noir (Francia) y Nebbiolo (Italia), destacan por su reputación y precios elevados.
- Malbec (Argentina) y Tempranillo (España) ofrecen gran calidad a precios accesibles.

3. Vinos Destacados:

- Los vinos más exclusivos provienen de Francia e Italia, respaldados por su prestigio histórico.
- Argentina y España sobresalen por democratizar la calidad del vino, haciendo accesibles productos destacados.

Anexos

1. Código Fuente:

- Disponible en el repositorio de GitHub: **main**

2. Gráficos:

- Ver carpeta **src/img**

3. Gráficos Adicionales:

- Ver carpeta **src/notebook/Notebook_otros_graficos**

4. Datos:

- Muestra representativa disponible en **src/data(EDA_wine_data)**