

## ESTUDIO DEL CONTENIDO E IMPACTO DE LAS VALORACIONES ONLINE DE VINOS EN AMAZON.ES

**Emilio Pindado<sup>a</sup>, Silverio Alarcón<sup>a\*</sup>**

<sup>a</sup> *Departamento de Economía Agraria, Estadística y Gestión de Empresas, ETSIAAB, Universidad Politécnica de Madrid, (Madrid, emilio.pindado@upm.es, silverio.alarcon@upm.es).*

### Resumen

Las valoraciones publicadas en plataformas como Amazon influyen en las decisiones de los consumidores y en las ventas. Sin embargo, la investigación sobre su influencia en la compra de alimentos es escasa. Por ello, este estudio analiza 2.197 revisiones en Amazon.es de sus 100 vinos más vendidos. Se realiza un análisis exploratorio del contenido publicado por los compradores mediante minería de datos y análisis de tópicos. Mediante modelos de regresión se analiza el impacto de las valoraciones de los usuarios, sentimientos expresados, precio y número de comentarios en la distribución de ventas. Los resultados revelan cómo los consumidores ponen de relieve la relación precio-calidad y características organolépticas como atributos a destacar. Finalmente, entre los factores que influyen las ventas destaca el número de comentarios y la distribución de las valoraciones, estando aquellos con mayor porcentaje de 5 cinco estrellas en las posiciones más altas de la clasificación de los 100 vinos más vendidos, mientras que los sentimientos expresados no son significativos.

**Palabras clave:** web scraping, amazon, vino, comportamiento del consumidor.

### 1. Introducción

El comercio de alimentos en plataformas online se ha incrementado en la última década, y más aún a partir de 2020 con la crisis de COVID-19 y los confinamientos (Chang y Meyerhoefer, 2021). Ahora bien, a pesar de este incremento, todavía se desconocen los principales factores que determinan las decisiones de compra online de alimentos (Heng et al., 2018). Para cada referencia las plataformas ofrecen comentarios suministrados por los consumidores, valoraciones de los mismos en una escala de una a cinco estrellas, y el ranking de ventas. La literatura ha señalado el efecto que tienen estas fuentes de información sobre las decisiones de compra de los consumidores (Kordrostami y Rahmani, 2020). No obstante, la influencia de la información generada por los consumidores puede variar en función de la tipología del producto y el contenido de las mismas (Etumnu et al., 2020). En este sentido, los estudios analizando el efecto que tiene el contenido generado por los consumidores sobre las ventas de vinos en plataformas online son escasos ya que la mayoría se han centrado en analizar el conocimiento previo de los consumidores o en la segmentación de los mismos (Ellis y Caruana, 2018; Moon y Kamakura, 2017). Por lo tanto, el objetivo de este trabajo es, por una parte, explorar los tópicos más frecuentes en los comentarios de los vinos más vendidos en España a través de Amazon, y por otra, estudiar las relaciones entre la información que suministra la plataforma y las ventas. Analizar las relaciones entre estas variables puede mejorar el conocimiento sobre el uso de la información y el comportamiento de los consumidores online, permitiendo desarrollar estrategias óptimas de producción y marketing.

### 2. Antecedentes

Una breve revisión de la literatura empírica sobre este tema indica que las distribuciones de las valoraciones son asimétricas hacia la derecha, o positivamente desequilibradas, es decir, son considerablemente superiores las proporciones de valoraciones con 4 y 5 estrellas. Esto se debe a una auto-selección de los compradores (Schoenmüller et al., 2019). Es decir, los consumidores más propensos a estar satisfechos con un producto son sus primeros compradores y sus primeros valoradores. Por tanto, se produce un sesgo en las valoraciones de estos en relación con la población en general. También es frecuente encontrar una polarización de las distribuciones de las valoraciones online. Pues son los consumidores con experiencias extremas, o muy buenas o muy malas, los que están más motivados para realizar las valoraciones. Mientras que los medianamente satisfechos reportarán información en menos ocasiones y estarán infrarrepresentados (Etumnu et al., 2020). Respecto a los comentarios, se tienen los mismos problemas de sesgos y polarización. En estas condiciones y dada la heterogeneidad de los consumidores no está clara la relación entre la información suministrada por las plataformas y las ventas. Los potenciales consumidores pueden desconfiar de las valoraciones y comentarios o pueden pensar que están realizados por personas con gustos muy diferentes a los suyos (Chevalier y Mayzlin, 2006), de forma que más estrellas o buenas revisiones no se corresponden con mayores ventas. En cualquier caso, el tipo de producto y el mercado, o las características de los consumidores, pueden ser decisivas en este contexto. Y de ahí surge la dificultad de generalizar y la

necesidad de realizar estudios para productos concretos si se quiere conocer el comportamiento del consumidor online.

### 3. Metodología

Los datos se han obtenido de la web de Amazon.es aplicando técnicas *web scraping* mediante el paquete “rvest” usando el software R. La información descargada corresponde con los 100 vinos más vendidos en España a través de Amazon a principios de 2021. Después de eliminar algunas referencias por estar fuera de la categoría analizada se obtuvieron 93 vinos. Para ellos, se obtuvieron la posición en el ranking de ventas, precio, número de comentarios y valoración. Además, se descargaron los comentarios de los productos seleccionados, obteniendo 2.197 revisiones. Para alcanzar el objetivo del trabajo, se realizó un análisis de contenido de los comentarios y posteriormente se modelizó mediante MCO el efecto de la información generada por los compradores sobre el ranking de ventas. La variable dependiente son las ventas de vino, que se aproxima mediante el ranking de ventas. Dado que a mayores ventas corresponde un valor menor en el ranking, se ha tomado (K-ranking), siendo K una constante cualquiera (Etumnu et al., 2020). Las variables explicativas empleadas han sido el número de comentarios, el precio y la calidad expresada por otros consumidores (Chevalier y Mayzlin, 2006). Esta última se puede capturar mediante la valoración media, la proporción de buenas valoraciones (5 y 4 estrellas), la proporción de malas valoraciones (1 y 2 estrellas) o un indicador del sentimiento de los comentarios, caracterizando los mismos en una escala de sentimientos utilizando el diccionario NRC mediante el paquete “syuzhet” en R. El Cuadro 1 presenta los estadísticos básicos de estas variables.

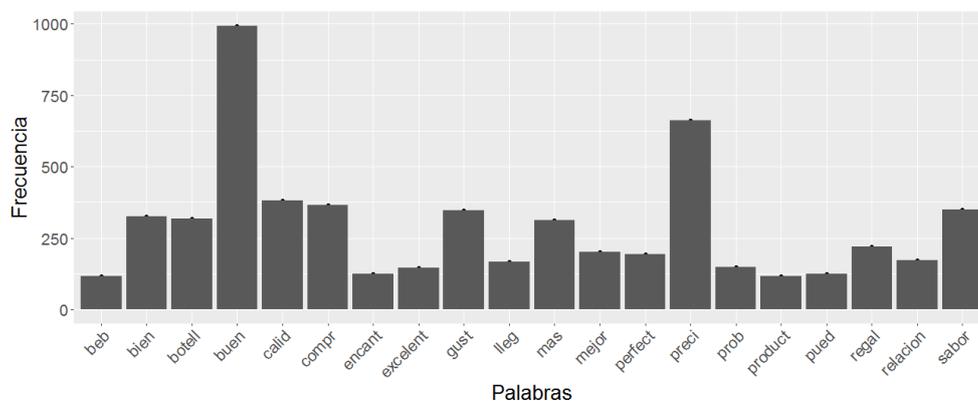
Cuadro 1. Estadísticos de las variables

	Media	Mediana	Mínimo	Máximo	Desv. std.
Ranking de ventas	50,87	50,00	1,00	100,00	29,83
Nº comentarios	142,09	43,00	1,00	3378,00	435,50
Precio	20,07	16,42	1,79	59,99	14,09
Valoración media (Estrellas)	4,34	4,40	3,00	5,00	0,33
Proporción 5 y 4 estrellas	0,84	0,84	0,41	1,00	0,11
Proporción 1 y 2 estrellas	0,08	0,07	0,00	0,40	0,08
Valencia sentimientos	1,19	1,00	-1,20	7,50	1,14

### 4. Resultados

El análisis del contenido de las revisiones refleja cómo los principales atributos señalados por los consumidores de vino en Amazon.es están relacionados con la relación precio calidad de los productos adquiridos, sabor y presentación (Gráfico 1). Estos resultados se ven reforzados por el análisis de tópicos realizado mediante el modelo LDA (Latent Dirichlet allocation). El número óptimo de temáticas identificado es de 3 siguiendo la metodología de Cao et al. (2009). El Cuadro 2 presenta los lexemas que conforman los tópicos identificados.

Gráfico 1. Términos frecuentes (lexemas) comentarios top 100 vinos Amazon.es



Cuadro 2. Términos frecuentes (lexemas) tópicos identificados

Tópico 1 – Precio Calidad	<i>buen, preci, calid, bien, perfect, relacion, excelent, com, gust, calidadpreci</i>
Tópico 2 – Características organolépticas	<i>sabor, gust, gran, tom, mejor, agrad, suav, afrut, tint, mes</i>

---

Tópico 3 – Experiencia de compra *compr, botell, mas, regal, pued, caj, lleg, mal, prob, ped*


---

El Cuadro 3 presenta las estimaciones de cuatro modelos, que difieren en la variable que recoge la calidad del producto expresada por los consumidores online. Se ha excluido la variable precio debido a la escasa correlación que se ha detectado. Y se ha incluido la interacción del logaritmo del número de comentarios por la variable de calidad, mejorando la bondad del ajuste. Los tests F muestran que las variables explicativas usadas son en conjunto significativas. Si bien la variabilidad del ranking de ventas explicada es aproximadamente una quinta parte de la total, excepto en el modelo (4) que es menor. Este resultado señala la heterogeneidad del consumidor online y la dificultad de encontrar variables que modelicen su comportamiento. Se han realizado también tests Breusch-Pagan, no encontrando excesiva heterocedasticidad, siendo las estimaciones robustas a esta cuestión. Los dos primeros modelos muestran que tanto el número de comentarios como la calidad, medida como número medio de estrellas (1) o proporción de 5 y 4 estrellas (2), son positivas y significativas. Por tanto, los vinos más vendidos son los que tienen más comentarios y mejores valoraciones. El modelo (3) pone de manifiesto que las ventas tienen una relación negativa con los vinos que tienen peores valoraciones. Y en estas situaciones el número de comentarios no presenta relación significativa con las ventas. Por otra parte, el modelo (4) muestra que la valencia de los sentimientos no es una buena variable explicativa del ranking de ventas.

Cuadro 3. Regresiones  $\log(K\text{-ranking de ventas})$  MCO

	(1)	(2)	(3)	(4)
Constante	-4,850 (3,013)	-1,761 (2,155)	3,072*** (0,354)	2,835*** (0,445)
$\log(N^\circ \text{ comentarios})$	2,485*** (0,818)	1,602*** (0,601)	0,109 (0,102)	0,241* (0,126)
Valoración media (Estrellas)	1,720** (0,674)			
Proporción 5 y 4 estrellas		4,908** (2,323)		
Proporción 1 y 2 estrellas			-7,395** (3,677)	
Valencia sentimientos				-0,047 (0,117)
Interacción	-0,508*** (0,184)	-1,511** (0,659)	2,482** (1,124)	-0,014 (0,046)
Num.Obs.	93	93	93	92
R2	0,221	0,192	0,204	0,113
F	8,428***	7,066***	7,622***	3,750**

Desviaciones típicas entre paréntesis. \*  $p < 0,1$ , \*\*  $p < 0,05$ , \*\*\*  $p < 0,01$

## 5. Conclusiones

Este trabajo muestra como los principales atributos destacados por los consumidores de vinos en Amazon.es son los relacionados con la relación precio-calidad de los productos adquiridos y sus características organolépticas. Estos resultados están en línea con estudios anteriores analizando los factores determinantes en la elección de vinos en el mercado español (Albisu y Zeballos, 2014). Destacan también, consecuencia de ser producto de experiencia, los comentarios respecto a la compra o motivo de la misma (Chua y Banerjee, 2016). El análisis de los factores determinantes de las ventas refleja como los consumidores compran más aquellos vinos con mayores revisiones y buenas valoraciones. Por lo tanto, los resultados sugieren que el consumidor online de vinos presenta un comportamiento similar al de otros productos (Chevalier y Mayzlin, 2006; Etumnu et al., 2020). Sin embargo, se han encontrado evidencias de la nula sensibilidad a los sentimientos expresados por otros consumidores, algo que sí ocurre en productos como los libros (Hu et al., 2014).

## Bibliografía

Albisu, L. M. y Zeballos, M. G. (2014). Consumo de vino en España. Tendencias y comportamiento del consumidor. Cajamar.

- Cao, J., Xia, T., Li, J., Zhang, Y. y Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7-9), 1775-1781.
- Chang, H. H. y Meyerhoefer, C. D. (2021). COVID-19 and the Demand for Online Food Shopping Services: Empirical Evidence from Taiwan. *American Journal of Agricultural Economics*, 103(2), 448-465.
- Chevalier, J. A. y Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345-354.
- Chua, A. Y. y Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior*, 54, 547-554.
- Ellis, D. y Caruana, A. (2018). Consumer wine knowledge: components and segments. *International Journal of Wine Business Research*, 30(3), 277-29.
- Etumnu, C. E., Foster, K., Widmar, N. O., Lusk, J. L. y Ortega, D. L. (2020). Does the distribution of ratings affect online grocery sales? Evidence from amazon. *Agribusiness*, 36(4), 501-521.
- Heng, Y., Gao, Z., Jiang, Y. y Chen, X. (2018). Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach. *Journal of Retailing and Consumer Services*, 42, 161-168.
- Kordrostami, E. y Rahmani, V. (2020). Investigating conflicting online review information: evidence from Amazon.com. *Journal of Retailing and Consumer Services*, 55, 102125.
- Moon, S. y Kamakura, W. A. (2017). A picture is worth a thousand words: Translating product reviews into a product positioning map. *International Journal of Research in Marketing*, 34(1), 265-285.
- Schoenmüller, V., Netzer, O. y Stahl, F. (2019). The extreme distribution of online reviews: Prevalence, drivers and implications. *Columbia Business School Research Paper*, 18-10.