

Datos, pero que sean moda

Anónimo

Abstract Este proyecto de investigación explora los entresijos de la industria de la moda uruguaya a través de un análisis exhaustivo de la MoWeek (Semana de la Moda Uruguaya), empleando técnicas de web scraping para recopilar un amplio conjunto de datos. El estudio se centra en las categorías de ropa, los precios, las tallas, los colores y la información sobre sostenibilidad, así como en una colección de imágenes relacionales al conjunto de datos, con el fin de obtener información valiosa sobre la industria. Los objetivos principales consisten en analizar las tendencias de precios en las distintas categorías de ropa, emplear agrupaciones de *k-means* para determinar los colores predominantes y evaluar la proporción de producción nacional y las prácticas sostenibles entre las marcas. Al examinar cuantitativamente estos factores, la investigación revela patrones quizá, desconocidos y proporciona una comprensión holística de la industria de la moda uruguaya. Los resultados contribuyen tanto a la comprensión académica como a las aplicaciones prácticas, permitiendo a los profesionales de la industria, a los responsables políticos y a los entusiastas de la moda tomar decisiones informadas en relación con las estrategias de marketing, las preferencias de los consumidores y las iniciativas de sostenibilidad. Esta propuesta pretende presentar los resultados de la investigación en el *LatinR*, difundiendo las metodologías, los puntos clave y las implicaciones a un público más amplio. En última instancia, este análisis pretende fomentar el crecimiento, inspirar nuevas investigaciones y promover prácticas sostenibles en la industria de la moda uruguaya.

Palabras clave: Scraping - K-means clustering - Moda

Introducción

La industria de la moda es una poderosa expresión de la identidad cultural y las preferencias de los consumidores. Este proyecto de investigación se adentra en el panorama de la moda uruguaya, comenzando como piloto, en la MoWeek (Semana de la Moda Uruguaya). Aprovechando técnicas de *web scraping*, reuní un exhaustivo conjunto de datos de vestimenta que va desde categorías de ropa, precios, tallas, colores e información sobre sostenibilidad y producción nacional, así como la colección de imágenes correspondientes.

El objetivo principal de este estudio era obtener información valiosa sobre la industria de la moda uruguaya, analizando aspectos clave como las tendencias de precios en las distintas categorías de ropa. Mediante el empleo de algoritmos avanzados, como la agrupación de *k-means*, exploré los patrones inherentes a las imágenes recopiladas para determinar los colores predominantes asociados a categorías de moda específicas, es decir, las tendencias de la moda.

Además de investigar las tendencias estéticas, mi objetivo era a su vez, arrojar luz sobre la proporción de producción nacional y las prácticas de sostenibilidad predominantes entre las marcas participantes. Mediante una cuidadosa extracción y análisis de datos, examiné hasta qué punto la sostenibilidad y la producción local desempeñan un papel en el panorama de la moda uruguaya, proporcionando una evaluación del compromiso de la industria con las prácticas responsables y éticas.

Las conclusiones de este proyecto de investigación contribuyen a la comprensión académica de la industria de la moda uruguaya y aportan valiosos conocimientos a los profesionales del sector, los responsables políticos y los aficionados a la moda. Al combinar el análisis cuantitativo con las evaluaciones cualitativas, descubrimos tendencias y patrones, lo que permite a las partes interesadas tomar decisiones informadas con respecto a las estrategias de marketing, las preferencias de los consumidores y las iniciativas de sostenibilidad.

Esta propuesta pretende presentar la culminación de mis esfuerzos de investigación en *LatinR*, ofreciendo una plataforma para compartir mis hallazgos, metodologías y las implicaciones que tienen tanto para la comunidad académica como para la industria. Creo que este análisis exhaustivo de la industria de la moda uruguaya podría poner sobre la mesa una comprensión más profunda de su dinámica, el comportamiento del consumidor y la sostenibilidad.

Elaboración del proyecto: desafíos y oportunidades

El proyecto se forjó ante el interés de analizar la moda local de una manera más rigurosa y sistemática. ¿El catalizador perfecto? La Moweeek, y por ende, el encuentro de diseño más importante de Uruguay. Previo al evento “MoWeek” se realiza todos los años la venta online en la plataforma “MarketPlace”, un anticipo web al evento Moweeek y con los mismos productos que se venderán en el evento. Esto me permitió no solo extraer datos sobre qué tipo de prenda era sino su marca, precio, talles y colores disponibles como si eran producción nacional y/o sustentable¹, sino también fotos de las mismas.

¹Un producto es sustentable es considerado un producto que fue producido con materiales sostenibles. Es información recogida por Moweeek.

Esto implicó un desafío en varios frentes, en primer lugar, el diseño del *scraping*, pensando en un loop que no solo extraiga datos en texto plano sino imágenes en carpetas según la categoría a la que pertenezca la prenda y que esté vinculada por nombre al conjunto de datos. Por otra parte, de limpieza de datos, si bien la página de MoWeek era bastante “homogénea”, datos como colores y talles venían en un solo bloque por su carácter ‘html’ y tuvieron que ser limpiados muy cuidadosamente. Por ejemplo, los talles y colores hay que separarlos, cuando todo vienen en distintos formatos y compatibilizarlos.

A su vez, analizar los colores tendencia para cada categoría, es decir, tapados y chaquetas, blusas y tops, pantalones, polleras, etc. presentó no solo un desafío a nivel de escoger la metodología sino de poder presentar la información de una manera lo suficientemente interesante para el público en general. La metodología escogida finalmente fue *k-means clustering* para determinar colores predominantes en las imágenes, es decir, colores “tendencia” en cada categoría. Recogí los clusters de color de cada imagen, después de aplicar el *clustering* de *k-means* a cada imagen, obtuve los clusters que representan los colores principales. Cada cluster está formado por el centroide (color representativo) y los puntos de datos asociados. Luego, combiné los clusters de color, es decir, combine los clusters de color de todas las imágenes en una única colección (por categoría). A esta altura tenía una lista de centroides que representan los colores principales de todas las imágenes. Determiné la frecuencia de cada color, es decir, “conté” el número de puntos de datos asociados a cada centroide de color. Esto me dio una indicación de la frecuencia con la que aparece cada color en las imágenes analizadas. Clasifiqué los colores en función de su frecuencia en orden descendente. Esta clasificación me ayudó a identificar los colores más dominantes en el conjunto de datos de imágenes. Luego, seleccioné los colores principales: elegí el número deseado de colores principales que deseaba incluir en el resumen del espectro de colores. Por ejemplo, como quería resumir los cinco colores principales, seleccioné los cinco primeros colores de la lista ordenada. Por último, visualicé el espectro de colores como una paleta de colores.

Por último, es importante destacar que este proyecto está siendo ampliado actualmente a marcas más masivas y no solo pertenecientes a la “élite” de la moda (MoWeek). El trabajo de armonización de los datos ha planteado un desafío importante en cuanto a qué datos y cómo los ofrece cada página de vestimenta. Desde compatibilizar las categorías de ropa, hasta determinar qué tejido se considera sustentable según la MoWeek para aplicarlo a marcas de consumo más masivo. Este proyecto es el que deseo presentar en el LatinR.