



Facultade de Humanidades e Documentación



UNIVERSIDADE DA CORUÑA

TRABALLO DE FIN DE GRAO

GRAO EN XESTIÓN INDUSTRIAL DA MODA

# Analizando a percepción dunha empresa empregando business analytics

**Estudante:** Valeria Añón Pensado

**Titor:** Ángel Manuel González Rueda

Javier Tarrío Saavedra

Ferrol, Xuño de 2023

## Resumo

Analizar a percepción dunha marca é esencial para comprender as preferencias e necesidades dos consumidores, avaliar a imaxe da marca, identificar fortalezas e debilidades, medir a satisfacción do cliente e tomar decisións estratéxicas baseadas en datos. Estes coñecementos son fundamentais para construír unha marca sólida, diferenciada e exitosa no mercado.

O presente traballo de fin de grao ten como obxectivo principal analizar a percepción dunha empresa de comercio electrónico empregando *business analytics*, máis concretamente o software R, para analizar unha base de datos pública con información sobre os seus usuarios. A empresa analizada céntrase no modelo de negocio C2C (consumer-to-consumer), onde os consumidores interactúan directamente entre si para comprar e vender produtos.

Neste estudo, recompílanse datos relevantes sobre os consumidores, incluíndo variables demográficas, comportamentos de compra e venda e preferencias. Estes datos son procesados e analizados empregando o software R, coñecido pola súa ampla gama de capacidades analíticas e ferramentas estatísticas.

A través da análise dos datos dos consumidores, búscase identificar patróns, tendencias e percepcións clave relacionadas coa marca. Lévese a cabo unha análise descritiva explorando variables como o uso de algunha app da marca, o título civil e o xénero dos consumidores, o seu idioma de preferencia e o seu país de procedencia, e o emprego de unha foto de perfil. Todas estas variables son relacionadas co número de compras e vendas, para así coñecer os diferentes comportamentos de compra e venda existentes entre os usuarios. Ademais, empréganse outras análises, como a correlación entre variables e diferentes tipos de diagramas que, de maneira máis visual e ordenada, nos permiten comprender con claridade os resultados.

A análise realizada empregando o software R permite obter resultados significativos sobre a percepción da marca entre os seus usuarios. Estes resultados poden ser empregados pola empresa para tomar decisións estratéxicas, como mellorar a experiencia do usuario, optimizar a plataforma de comercio electrónico, desenvolver campañas de marketing máis efectivas ou fortalecer a relación cos consumidores.

**Palabras clave:** percepción de marca, business analytics, C2C, análise de datos, comportamento dos usuarios, usuarios activos.

## Resumen

Analizar la percepción de una marca es esencial para comprender las preferencias y necesidades de los consumidores, evaluar la imagen de la marca, identificar fortalezas y debilidades, medir la satisfacción del cliente y tomar decisiones estratégicas basadas en datos. Estos conocimientos son fundamentales para construir una marca sólida, diferenciada y exitosa en el mercado.

El presente trabajo de fin de grado tiene como objetivo principal analizar la percepción de una empresa de comercio electrónico empleando *business analytics*, más concretamente el software R, para analizar una base de datos pública con información sobre sus usuarios. La empresa analizada se centra en el modelo de negocio C2C (consumer-to-consumer), donde los consumidores interactúan directamente entre sí para comprar y vender productos.

En este estudio, se recopilan datos relevantes sobre los consumidores, incluyendo variables demográficas, comportamientos de compra y venta y preferencias. Estos datos son procesados y analizados empleando el software R, conocido por su amplia gama de capacidades analíticas y herramientas estadísticas.

A través del análisis de los datos de los consumidores, se busca identificar patrones, tendencias y percepciones clave relacionadas con la marca. Se lleva a cabo un análisis descriptivo explorando variables como el uso de alguna app de la marca, el título civil y el género de los consumidores, su idioma de preferencia y su país de procedencia, y el empleo de una foto de perfil. Todas estas variables son relacionadas con el número de compras y ventas, para así conocer los diferentes comportamientos de compra y venta existentes entre los usuarios. Además, se emplean otros análisis, como la correlación entre variables y diferentes tipos de diagramas que, de manera más visual y ordenada, nos permiten comprender con claridad los resultados.

El análisis realizado empleando el software R permite obtener resultados significativos sobre la percepción de la marca entre sus usuarios. Estos resultados pueden ser empleados por la empresa para tomar decisiones estratégicas, como mejorar la experiencia del usuario, optimizar la plataforma de comercio electrónico, desarrollar campañas de marketing más efectivas o fortalecer la relación con los consumidores.

**Palabras clave:** percepción de marca, business analytics, C2C, análisis de datos, comportamiento de los usuarios, usuarios activos.

## **Abstract**

Analysing brand perception is essential to understand consumer preferences and needs, assess brand image, identify strengths and weaknesses, measure customer satisfaction and make data-driven strategic decisions. This knowledge is fundamental to build a strong, differentiated and successful brand in the marketplace.

The main objective of this thesis is to analyse the perception of an e-commerce company using business analytics, more specifically R software, to analyse a public database with information about its users. The company analysed focuses on the C2C (consumer-to-consumer) business model, where consumers interact directly with each other to buy and sell products.

In this study, relevant consumer data is collected, including demographic variables, buying and selling behaviours and preferences. This data is processed and analysed using R software, known for its wide range of analytical capabilities and statistical tools.

Through the analysis of consumer data, we seek to identify key patterns, trends and perceptions related to the brand. A descriptive analysis is conducted by exploring variables such as the use of a brand app, consumers' marital status and gender, their language preference and country of origin, and the use of a profile picture. All these variables are associated with the number of purchases and sales, in order to understand the different buying and selling behaviours among users. In addition, other analyses are used, such as the correlation between variables and different types of diagrams that, in a more visual and orderly way, allow us to clearly understand the results.

The analysis carried out using R software allows us to obtain significant results on the perception of the brand among its users. These results can be used by the company to make strategic decisions, such as improving the user experience, optimising the e-commerce platform, developing more effective marketing campaigns or strengthening the relationship with consumers.

**Keywords:** brand perception, business analytics, C2C, data analysis, user behaviour, active users.

# Táboa de contidos

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>INTRODUCCIÓN</b>  | <b>1</b>  |
| <b>2</b> | <b>FUNDAMENTOS TEÓRICOS</b>  | <b>3</b>  |
| 2.1      | O MODELO C2C   | 3         |
| 2.2      | CONCEPTO E IMPORTANCIA DA PERCEPCIÓN DE MARCA  | 3         |
| 2.3      | MÉTODOS PREVIOS DE MEDICIÓN DA PERCEPCIÓN DE MARCA   | 4         |
| 2.4      | BUSINESS ANALYTICS   | 6         |
| <b>3</b> | <b>METODOLOXÍA</b>   | <b>7</b>  |
| 3.1      | OBTENCIÓN DE DATOS   | 7         |
| 3.2      | PROCESAMIENTO DE DATOS   | 9         |
| <b>4</b> | <b>RESULTADOS E DISCUSIÓN</b>  | <b>12</b> |
| 4.1      | LECTURA E RESUMO XERAL DE DATOS  | 12        |
| 4.2      | DATOS DESCRIPTIVOS DOS USUARIOS  | 12        |
| 4.3      | CORRELACIÓN ENTRE VARIABLES  | 18        |
| 4.4      | COMPRADORES E VENDEDORES SEGUNDO O SEU TÍTULO CIVIL: DIAGRAMA DE CAIXA   | 21        |
| 4.5      | PORCENTAXES DE USUARIOS ACTIVOS E USUARIOS REXISTRADOS QUE NON REALIZAN NINGUNHA TRANSACCIÓN AGRUPADOS SEGUNDO O SEU XÉNERO E TÍTULO CIVIL | 24        |
| <b>5</b> | <b>CONCLUSIÓNS</b>   | <b>30</b> |
|          | <b>REFERENCIAS BIBLIOGRÁFICAS</b>  | <b>32</b> |
|          | <b>ANEXO</b>   | <b>34</b> |

---

# 1 Introducción

A percepción que os consumidores teñen dunha marca é un factor clave para o éxito da mesma, xa que inflúe na decisión dos mesmos clientes sobre o consumo ou uso dela. Para analizar esta percepción é fundamental utilizar ferramentas e técnicas que permitan recompilar e estudar datos relevantes. Esta investigación destacará a importancia da información recompilada dunha base de datos como medio para comprender e valorar a percepción que os usuarios teñen dunha marca.

O obxectivo xeral desta investigación é analizar como a información obtida dunha base de datos pode proporcionar unha visión profunda e significativa da percepción dos consumidores cara unha marca en particular. A través dun enfoque baseado en datos, explorárase como as empresas poden aproveitar esta ampla e útil fonte de información para mellorar a súa imaxe de marca e desenvolver estratexias que permitan mellorar en diferentes áreas. Pois a imaxe dunha marca considérase un factor importante, tal e como demostran as enormes sumas de diñeiro que gastan as empresas no desenvolvemento e na medición da súa imaxe corporativa/de marca. Sen embargo, pouco se sabe sobre a relación entre a percepción da marca e o comportamento do comprador (Romaniuk & Sharp, 2003).

Para analizar os datos obtidos empregárase a *business analytics*, unha disciplina que combina a análise de datos e as técnicas analíticas avanzadas para obter información útil e tomar decisións estratéxicas. A través da aplicación de ferramentas e métodos de *business analytics*, buscarase comprender en profundidade como os consumidores perciben unha marca e como esta percepción afecta ao seu comportamento e á súa relación coa empresa. En particular, empregárase o software R para analizar unha base de datos detallada sobre os usuarios da marca para obter información valiosa e significativa. Trátase da linguaxe estatística que os expertos en datos de todo o mundo empregan para todo, dende a elaboración de mapas de amplas tenencias sociais e de marketing en liña ata o desenvolvemento de modelos financeiros e climáticos que axudan a impulsar as nosas economías e comunidades (*What is R?*, s. f.). Mediante a análise destes datos, preténdese afondar na comprensión de como os usuarios perciben a marca e como esta percepción afecta ás súas actitudes e comportamentos.

A base de datos empregada para realizar o estudo contén información dos usuarios dunha marca de tipo C2C (consumer-to-consumer, en galego consumidor-a-consumidor), entre os que se inclúen datos demográficos, como o país de procedencia ou o xénero de cada un; pero tamén datos relacionados coa súa actividade coa marca, como o número de produtos comprados e vendidos ou o número de produtos engadido á súa lista de desexos. O modelo de negocio C2C refírese a transaccións comerciais directas entre consumidores individuais, onde os usuarios actúan como vendedores e compradores a través de plataformas en liña especializadas. Este enfoque permitirá explorar a percepción de marca nun contexto específico, onde a interacción directa entre consumidores desenvolve un papel clave no proceso de compra e venda. Polo tanto, a percepción dunha marca deste tipo adquire un papel moi importante, xa que inflúe na confianza e na disposición dos usuarios para participar en transaccións con outros consumidores, a diferenza doutros casos, nos que só se require depositar a confianza nunha marca xa recoñecida.

O emprego de *business analytics* para analizar a percepción dunha marca ofrece varias vantaxes e oportunidades. En primeiro lugar, permite recompilar e analizar grandes volumes de datos de maneira eficiente, o que facilita a identificación de patróns, tendencias e relacións entre variables relevantes para a percepción de marca. Da mesma forma, a análise de datos permite obter información cuantitativa e cualitativa sobre como os consumidores perciben unha marca, o que proporciona unha visión máis completa e detallada das súas preferencias, actitudes e comportamentos.

---

En resumo, este traballo de fin de grao enfócase na análise da percepción dunha marca C2C empregando *business analytics* para o seu desenvolvemento. A través da análise de datos e emprego do software R, buscarase comprender como os consumidores perciben a marca no contexto dun modelo de negocio C2C, e como esta percepción pode influír no seu comportamento de cara a marca.

Ao longo deste traballo, presentaranse os fundamentos teóricos, a metodoloxía empregada, os resultados obtidos e as conclusións relevantes. Espérase que este estudo contribúa ao coñecemento existente no campo da análise da percepción de marca a través de *business analytics*, e proporcione recomendacións prácticas para as organizacións que buscan mellorar a súa imaxe de marca e a súa relación cos consumidores.

---

## 2 Fundamentos teóricos

Nos fundamentos teóricos deste traballo, estudaranse conceptos clave relacionados coas marcas e a súa percepción. Entre os que se inclúen o funcionamento do modelo C2C (consumidor a consumidor) e o concepto e a importancia da percepción de marca. Ademais, exploraranse métodos previos de medición da percepción de marca e realizarase unha descrición sobre o concepto de business analytics.

### 2.1 O modelo C2C

C2C, entre consumidores, o comercio electrónico consumer to consumer (consumidor a consumidor), é o tipo de comercio que se leva a cabo entre dous ou máis consumidores mediante a rede, onde uns actúan como vendedores e outros como compradores. O obxectivo desta modalidade é facilitar a comercialización de produtos entre particulares. Este modelo está emerxendo con forza en Internet e permite a reutilización de produtos en desuso e unha comunicación máis fluida. Neste tipo de comercio electrónico destacan as poxas online, algunhas das empresas máis coñecidas de C2C son por exemplo eBay ou Mercadolibre, estas realizan unha actividade de intermediación entre os consumidores e cobran unha pequena tarifa polo servizo (Sancho Esper, 2017).

Este modelo presenta diversas vantaxes tanto para os consumidores como para os vendedores. Permite o acceso a unha ampla gama de produtos que están dispoñibles no mercado e, en moitos casos, a oportunidade de obter produtos únicos e exclusivos que os consumidores non poden atopar noutros canais de venta. Ademais, se o número de vendedores é moi elevado, pódense atopar prezos moi competitivos. Por outro lado, existe unha interacción directa entre consumidores polo que se establece unha relación máis personalizada. Sen embargo, isto tamén pode ser unha desvantaxe, pois pode supoñer tamén unha dificultade para establecer confianza entre comprador e vendedor, xa que os consumidores poden sentirse inseguros comprándolle a vendedores descoñecidos ou sen ningunha reputación establecida. Outra desvantaxe relacionada tamén pode ser a falta de responsabilidade despois da venta. Os consumidores poden enfrontar dificultades para solucionar calquera problema co produto comprado, xa que esa responsabilidade recae nos vendedores individuais e non en unha empresa recoñecida.

### 2.2 Concepto e importancia da percepción de marca

A percepción de marca refírese á imaxe que os consumidores teñen dunha marca específica, a cal se constrúe a través das experiencias que os consumidores desenvolven en relación coa marca. A importancia da percepción de marca atópase na súa capacidade para influír nas actitudes e comportamentos dos consumidores cara a marca. Non só se trata da mensaxe que a túa empresa queira comunicar, senón de como a reciben e interpretan os usuarios (Crandi, 2022).

A percepción de marca é un elemento clave para o éxito empresarial, como somos percibidos polos usuarios actuais e potenciais determinará o que van pensar de nós, a súa disposición e en última instancia se van percibir o servizo con maior calidade ou valor, se van ser máis fieis ou se o van recomendar de forma positiva (Alguacil Jiménez, 2017).

Unha percepción positiva pode diferenciar a unha empresa dos seus competidores. Nun mercado saturado e moi competitivo, unha marca ben percibida destaca como a opción preferida por parte dos consumidores. Isto pode producir unha vantaxe competitiva, xa que os consumidores están máis dispostos a elixir unha marca que perciben como fiable.



---

A percepción de marca tamén ten un papel importante na lealdade por parte dos consumidores. Pois cando estes teñen unha percepción positiva sobre unha marca, están máis dispostos a establecer unha relación de longa duración con ela. Pódese dicir que a confianza nunha marca convértese nunha maior lealdade dos clientes, o que leva a repetidas interaccións con ela, como veñen sendo máis compras e vendas. Ademais, na era dixital na que nos atopamos a día de hoxe, necesítase moito máis que realizar unha venta para gañar a fidelidade dos clientes. Por isto mesmo, é necesario crear unha estratexia para analizar como a túa audiencia percibe a túa marca (Crandi, 2022).

A percepción de marca tamén pode afectar nos prezos que os consumidores están dispostos a pagar polos produtos ou servizos dunha empresa. Unha marca ben percibida leva a unha maior valoración por parte dos consumidores, o que pode xustificar prezos máis altos. Isto contribúe ás marxes de beneficio da empresa e á súa rentabilidade. No caso dunha empresa C2C (consumidor a consumidor), isto podería reflexarse no pago de algunha taxa por produto comprado ou vendido. Se os clientes teñen unha percepción positiva da marca, aínda que sexa máis custosa que outros competidores, estarán dispostos a pagar un prezo maior.

Por último, a percepción de marca pode ser un factor determinante na imaxe e na reputación dunha empresa. Unha marca ben percibida pode xerar unha tanto unha imaxe como unha reputación positiva no mercado. Isto pode atraer por exemplo a inversores ou talento de calidade, fortalecendo así a posición competitiva da marca.

En conclusión, a percepción de marca é un factor moi importante na relación entre marca e consumidor. A súa importancia atópase na súa capacidade para diferenciar unha marca, xerar confianza e lealdade, xustificar prezos máis altos e construír unha imaxe e reputación positiva. Neste estudo, a análise da percepción de marca a partir de datos é fundamental para comprender mellor as preferencias e comportamentos dos usuarios, e así poder desenvolver estratexias efectivas para aumentar o éxito da mesma.

### 2.3 Métodos previos de medición da percepción de marca

É importante revisar os métodos previos de medición da percepción de marca empregados noutros estudos. Estes métodos proporcionan unha base para comprender como se investigou este tema anteriormente, o cal nos axuda a entender como enfocar este estudo.

O uso de **enquisas e cuestionarios** é unha ferramenta comunmente empregada para medir a percepción dunha marca. Estes métodos permiten recompilar datos de maneira estruturada, o que facilita a análise e a interpretación dos resultados. Son especialmente útiles para obter información directa dos consumidores sobre como perciben unha marca en particular. A través de preguntas adaptadas a cada circunstancia, pódense explorar diferentes aspectos da percepción de marca, como a imaxe de marca ou a lealdade cara a ela.

Unha das principais vantaxes deste método é a capacidade para obter datos dunha mostra representativa da poboación. Permite seleccionar unha mostra adecuada para aplicar as enquisas ou os cuestionarios, obtendo unha visión xeral de como se percibe a marca entre os consumidores. Ademais, as enquisas e cuestionarios poden adaptarse para tratar aspectos específicos da percepción da marca, incluíndo só preguntas relacionadas cos temas que se desexan tratar. Tamén é posible empregar preguntas hipotéticas para avaliar como os consumidores percibirían a marca se se desen diferentes situacións.

---

As **entrevistas individuais** ou os coñecidos como **focus groups** (grupos de discusión), proporcionan unha perspectiva máis profunda e detallada de como os consumidores perciben unha marca en particular.

Nas entrevistas lévase a cabo unha conversación directa e en profundidade con un participante, o que permite explorar actitudes e experiencias persoais dos consumidores entrevistados en relación coa marca.

Por outro lado, un **focus group** ou grupo de discusión é un método de investigación cualitativa que reúne participantes dunha entrevista, na cal se expoñen opinións sobre produtos ou servizos. Normalmente reúne de 5 a 10 persoas e sempre conta cun moderador que lidera o intercambio de ideas (Lisboa, 2019). Esta ferramenta permite explorar aspectos compartidos entre varios consumidores dunha mesma marca, o que proporciona unha visión xeral da súa percepción por parte dos consumidores.

Moitas veces as respostas abertas das entrevistas ou dos **focus groups** poden complementar os resultados obtidos a través de enquisas ou cuestionarios.

O **Net Promoter Score (NPS)** foi introducido por Reichheld na *Harvard Business Review* en 2003 nun artigo titulado “The One Number You Need to Grow”, no cal o definía da seguinte maneira:

Preguntando a unha mostra estatisticamente válida de clientes “Que probabilidade hai de que recomende a nosa empresa a un amigo ou compañeiro?” permíteche calcular o teu Net Promoter Score: a proporción [sic] de promotores fronte a detractores. Baseándose nas súas respostas nunha escala de 0 a 10, agrupa aos teus clientes en “promotores” (de 9 a 10, moi propensos a recomendar), “pasivamente satisfeitos” (de 7 a 8) e “detractores” (de 0 a 6, moi pouco propensos a recomendar). A continuación, resta a porcentaxe de detractores da porcentaxe de promotores (Fisher & Kordupleski, 2018).

Como mencionan Fisher e Kordupleski no seu estudo, o NPS conta con algúns problemas, entre os cales: non proporciona datos acerca de que facer para mellorar; só se centra en manter clientes, e non en conseguir outros novos; e céntrase no interno, sen ter en conta a influencia de factores externos.

O **indicador de satisfacción do cliente (CSAT)** é unha técnica empregada globalmente para medir a satisfacción do cliente. Caracterízase por ser discreta e sinxela. Impleméntase por medio de enquisas nas que se cualifica o produto ou servizo cun número que pode ir do un ao tres, do un ao cinco ou do un ao dez, sendo o un a cualificación máis baixa e o número máis alto, a máis alta.[...] A súa desvantaxe está en que, dada a rapidez coa que se responden, provoca incertidume en torno ao estado de ánimo co que foi respondida, polo que só se recomenda para cuestións de curto prazo (Bustamante Lazcano, 2023).

O **Click Through Rate (CTR)** ou proporción de clicks indica o número de clicks sobre un enlace con respecto a o seu número de impresións. Calcúlase dividindo o número de clicks obtidos nun enlace entre o número de impresións (número de vistas por usuario) e multiplícase por 100 (Sordo, 2021).

Esta técnica presenta o inconveniente de que só analiza o número de clicks con respecto ás visitas, polo que non proporciona información acerca do comportamento dos usuarios coa marca despois de ese momento.

---

Moitas marcas realizan un **seguimento de redes sociais** continua, que consiste nas análises de información obtida a partir das medición de cantidades de interaccións, como *likes*, *views*, reaccións, seguidores, comentarios... Este método permite obter información en grandes cantidades, debido ao elevado número de usuarios nas diferentes redes.

A exposición á opinión da rede global é unha labor que implica responsabilidade sobre o que se da a coñecer e como se da a coñecer. O potencial é infinito, pola cantidade de persoas que circulan en internet (Bustamante Lazcano, 2023).

Outro método similar é o **seguimento de comentarios noutros canais especializados**. Ao igual que a análise de comentarios en redes sociais, moitas marcas teñen en conta comentarios recibidos por outros medios, como poden ser as reseñas de Google ou, en algúns casos, na propia web da marca.

Neste ámbito ten importancia o concepto coñecido como Electronic Word-of-mouth (eWOM), calquera declaración positiva ou negativa realizada por clientes potenciais, reais ou antigos sobre un produto ou empresa, que se pon a disposición dunha multitude de persoas e institución a través de Internet (Reza Jalilvand et al., 2010).

Ao analizar a percepción dunha marca, en moitas ocasións tamén se ten en conta o coñecido como **valor de marca baseado no consumidor**, debido a que moitos investigadores afirman que as marcas son os activos máis valiosos dunha empresa. O valor de marca é o conxunto de asociacións e comportamentos por parte dos consumidores da marca, membros da canle, e da empresa matriz que lle permite á marca obter un maior volume ou maiores marxes dos que obtería sen a marca e que outorga á marca unha vantaxe forte, sostible e diferenciada fronte aos competidores (Farreen Y.L. & Lee, 2011). Trátase dun concepto asociado á lealdade dos clientes, xa que un valor positivo indica un maior número de clientes fieis.

## 2.4 Business analytics

A Business Analytics ou analítica de negocio é o arte e a ciencia de descubrir coñecementos -mediante o uso de sofisticados métodos matemáticos, estatísticos, de aprendizaxe automático e de ciencia de redes, xunto cunha variedade de datos e coñecementos de expertos- para apoiar unha toma de decisións mellor e máis rápida. Polo tanto, a business analytics pode considerarse un instrumento para a toma de decisións e a resolución de problemas (Delen & Ram, 2018). Aplicado á análise de datos dunha marca en concreto, esta ferramenta permite recompilar, analizar e visualizar grandes volumes de datos relacionados coa marca e o seu entorno, co obxectivo de obter resultados significativos e tomar decisións estratéxicas fundamentadas. Ademais, permite identificar patróns e tendencias entre os seus consumidores, proporcionando información sobre como perciben a propia marca. Isto permite ás empresas adaptar as súas estratexias para satisfacer así as necesidades dos seus clientes de maneira efectiva.

---

## 3 Metodoloxía

Este estudo ten como obxectivo principal expoñer un método específico para obter conclusións significativas sobre a percepción dunha marca mediante datos dos seus usuarios extraídos dunha base de datos. O método introducido podería ser aplicado por pequenas empresas que, a pesar de non ter coñecementos e recursos suficientes para levar a cabo unha análise do estilo, desexan sacar proveito da información recompilada da propia actividade dos seus usuarios.

Nos seguintes apartados descríbese a metodoloxía aplicada en cada unha das fases do procedemento proposto neste estudo. No Anexo inclúese unha descrición detallada dos pasos seguidos e das opcións do software R seleccionadas.

### 3.1 Obtención de datos

Para levar a cabo esta investigación, recompiláronse os datos de kaggle, unha plataforma web que reúne a comunidade Data Science máis grande do mundo, con máis de 536 mil membros activos en 194 países, recibe máis de 150 mil publicacións por mes. [...] Con máis de 50 mil conxuntos de datos públicos e 400 mil notas públicas dispoñibles para todo o mundo (*Kaggle: todo lo que hay que saber sobre esta plataforma, 2021*).

Os datos empregados pertencen a unha exitosa tenda de moda online C2C con máis de 10 millóns de usuarios rexistrados. Lanzada por primeira vez en Europa arredor de 2009 e máis tarde expandida por todo o mundo. Estes datos só inclúen información de 98.913 usuarios rexistrados, sen ter en conta a aqueles usuarios que non están rexistrados, polo que poden ver os artigos da páxina, pero non realizar compras ou vendas na plataforma.

Os datos recollidos pertencen ao intervalo de tempo entre os días 2 e 30 de Novembro de 2019, aproximadamente 10 anos despois do lanzamento da páxina web, aínda que a maioría deles pertencen a un período do transcurso dunha semana.

A base de datos consta de 3 arquivos diferentes en formato CSV:

- [6M-0K-99K.users.dataset.public.csv](#) – este ficheiro contén un conxunto de datos de usuarios e información sobre o seu comportamento social e de compras.
- [comparison\\_of\\_sellers\\_by\\_gender\\_and\\_country.csv](#) – comparación de vendedores por país e xénero. Podería empregarse para determinar que podería aprender un país ou un xénero de outro.
- [countries\\_with\\_top\\_sellers\\_fashion\\_c2c.csv](#) – datos estatísticos sobre os países con mellores vendedores, considerándoos como aqueles con 20 ou máis produtos vendidos. Representan aproximadamente un 10% do total de vendedores.

Para esta investigación, só se empregou o arquivo [6M-0K-99K.users.dataset.public.csv](#). Na Táboa 1 descríbense as diferentes variables do conxunto de datos empregado para o estudo:

Táboa 1. Variables do conxunto de datos [6M-0K-99K.users.dataset.public.csv](#)

|  |
|--|
| <a href="#">6M-0K-99K.users.dataset.public.csv</a> |
|--|

| <b>VARIABLE</b>      | <b>DEFINICIÓN</b>  |
|----------------------|--|
| #identifierHash      | O código identificador de cada usuario.  |
| #type                | Neste caso todos son usuarios.   |
| #country             | País do usuario.   |
| #language            | Idioma de preferencia do usuario.  |
| #socialNbFollowers   | Número de usuarios subscritos á actividade deste usuario.  |
| #socialNbFollows     | Número de contas de usuario ás que este usuario segue.   |
| #socialProductsLiked | Número de produtos que lle gustaron a este usuario.  |
| #productsListed      | Número de produtos non vendidos que este usuario subiu.  |
| #productsSold        | Número de produtos vendidos por este usuario.  |
| #productsPassRate    | Porcentaxe de produtos que cumpren a descrición do produto (os produtos vendidos son revisados polo equipo da tenda antes de ser enviados ao comprador). |
| #productsWished      | Número de produtos que este usuario engadiu á súa lista de desexos.  |
| #productsBought      | Número de produtos que este usuario comprou.   |
| #gender              | Xénero do usuario.   |
| #civilityGenderId    | 'civilityTitle' como número enteiro.   |
| #civilityTitle       | Título de civismo.   |
| #hasAnyApp           | O usuario utilizou algunha vez algunha aplicación oficial da tenda.  |
| #hasAndroidApp       | O usuario utilizou algunha vez a aplicación oficial de Android.  |
| #hasIosApp           | O usuario utilizou algunha vez a aplicación oficial de IOS.  |
| #hasProfilePicture   | O usuario ten unha imaxe de perfil personalizada.  |
| #daysSinceLastLogin  | Número de días desde o último inicio de sesión.  |
| #seniority           | Número de días desde que o usuario se rexistrou.   |
| #seniorityAsMonths   | 'seniority' en meses.  |
| #seniorityAsYears    | 'seniority' en anos.   |
| #countryCode         | País do usuario (ISO_3166-1).  |

---

## 3.2 Procesamento de datos

Unha vez recompilados os datos, comeza o seu procesamento, agrupándoos e clasificándoos para comprender de maneira máis profunda a percepción da marca. Así, en primeiro lugar é necesario seleccionar as variables que se utilizarán para cada unha das diferentes funcións: análise descritiva, correlación entre variables e análises dos comportamentos de compra e venta en diferentes segmentos de usuarios.

O software R será o empregado para levar a cabo tanto a análise descritiva como a realización das diferentes matrices de correlación entre variables: tendo en conta todos os usuarios activos, os usuarios que só compran, e os que só venden. Trátase dun conxunto de programas integrados para o manexo de datos, simulacións, cálculos e realización de gráficos (Zambrano, 2019). En concreto, tanto para a análise descritiva como para as correlacións entre variables, a ferramenta R Studio foi a empregada para todas as funcións desenvolvidas nestes apartados, un entorno de desenvolvemento integrado para a linguaxe de programación R.

Por outro lado, para as análises dos comportamentos de compra e venta dos usuarios empregárase R Commander, unha Interface Gráfica de Usuario (GUI en inglés), creada por John Fox, que permite acceder a moitas capacidades do entorno estatístico R sen que o usuario teña que coñecer a linguaxe de comandos propio deste entorno (Santana, s. f.). Esta ferramenta permitirá levar a cabo diferentes tipos de diagramas nos que se poderán comparar os diferentes comportamentos de compra e venta dos usuarios en función do segmento ao que pertencen.

### 3.2.1 Análise descritiva

Neste apartado levarase a cabo unha exploración de variables numéricas relevantes, co obxectivo de obter unha comprensión máis profunda dos datos. Comezarase creando unha táboa de frecuencias para cada variable e, a partir desta, un gráfico que permita visualizar os resultados obtidos de maneira máis clara e sinxela. Mediante as variables elixidas resolveranse as seguintes cuestións: cantos usuarios empregaron algunha aplicación oficial da tenda, o número de usuarios que se identifican baixo cada título civil, cal é o idioma de preferencia dos usuarios da web, cal é o país de procedencia de cada un dos usuarios, cantos usuarios teñen imaxe de perfil, e o número de usuarios identificados baixo cada un dos xéneros.

Unha vez analizada cada variable de xeito individual, procederase a estudar a relación entre cada unha delas co número de produtos comprados e vendidos polos usuarios. Isto permitirá observar se certos parámetros repercuten nas decisións de compra e venta. Por exemplo, poderíase investigar se os usuarios que teñen imaxe de perfil tenden a vender ou comprar máis produtos en comparación con aqueles que non a teñen.

Unha vez visualizados os resultados, procederase a discutir hipóteses formuladas sobre os resultados obtidos. Analizaranse as diferentes situacións e exporanse interpretacións baseadas nos resultados obtidos a partir da análise das variables. Esta discusión permitirá afondar na comprensión dos factores que inflúen nos comportamentos de compra e venta dos usuarios, así como na identificación de posibles áreas de mellora.

---

### 3.2.2 Correlación entre variables

A análise de correlación entre variables supón un papel moi importante para comprender as relacións existentes entre diferentes factores que poden influír na percepción dunha marca. Ademais, esta técnica pódenos axudar a predicir o valor dunha variable a partir do valor doutra, definir patróns de comportamentos similares ou correlativos en grupos de valores e ademais, comparar medidas cuxo contido sexa do tipo de dato continuo (*Clases online de Data analyst: Técnicas y herramientas de informes avanzado | LinkedIn Learning, antes Lynda.com, 2021*). A través desta análise podemos determinar se dúas variables presentan unha correlación positiva (cando o aumento dunha das variables supón o aumento da outra), negativa (cando o aumento dunha das variables supón a diminución da outra), ou se non teñen ningún tipo de relación.

Neste estudo, analizarase a correlación entre variables tendo en conta a todos os usuarios rexistrados na plataforma. Unha vez realizada esta análise, estudarase a correlación entre variables primeiro tendo en conta só ós usuarios que realizaron polo menos unha compra, e en segundo lugar tendo en conta só ós usuarios que realizaron unha ou máis vendas. Polo que, para ás dúas últimas análises será necesario levar a cabo unha limpeza de datos, eliminando primeiro aqueles usuarios que non realizaron ningunha compra (con valor 0 na variable #productsBought), e seguidamente aqueles que non realizaron ningunha venda (con valor 0 na variable #productsSold).

### 3.2.3 Compradores e vendedores segundo o seu título civil: análise mediante un diagrama de caixa.

O diagrama de caixa converteuse na técnica estándar para presentar o resumo de 5 números, que consta dos valores mínimo e máximo do intervalo, os cuartiles superior e inferior e a mediana. Esta colección de valores é unha forma rápida de resumir a distribución dun conxunto de datos. Ademais, esta representación reducida que ofrece o resumo de 5 números proporciona unha forma máis directa de comparar conxuntos de datos, xa que só é necesario analizar estes valores característicos (Potter, s. f.).

Este tipo de diagrama consiste nun rectángulo chamado caixa, que representa o rango no que se atopan o 50% dos datos, coñecido como rango intercuartílico (IQR), a mediana e liñas estendidas dende a caixa chamadas bigotes. Os valores atípicos represéntanse como puntos individuais máis alá dos bigotes, este indica observacións inusuais dentro do conxunto de datos.

Neste estudo, analizaranse o número de compras realizadas polos consumidores identificados baixo cada un dos títulos civís, e o número de vendas realizadas polos vendedores agrupados segundo os distintos títulos civís, tendo en conta só os usuarios que realizaron unha ou máis compras e os que realizaron unha ou máis vendas respectivamente.

Ao igual que nas matrices de correlación, para levar a cabo estas análises, será necesario realizar unha limpeza de datos previa. En primeiro lugar, eliminaranse os datos de aqueles usuarios que non realizasen ningunha compra, xa que non aportan información relevante para a análise do número de compras. En segundo lugar, eliminaranse os datos de todos aqueles usuarios que non realizase ningunha venda, xa que estes dificultarían a interpretación dos patróns e tendencias relacionados con dita variable.

---

### 3.2.4 Porcentaxes de usuarios activos e usuarios que só actúan como visitantes agrupados segundo o seu xénero e título civil.

Neste último apartado analizaranse as porcentaxes de usuarios activos na plataforma, primeiro diferenciados por xénero e a continuación diferenciados segundo o seu estado civil. Levaranse a cabo análises dos usuarios que compraron un ou máis produtos, dos que venderon algún produtos, e de aqueles que realizaron polo menos unha vez ambas as transaccións, unha ou máis compras e vendas.

Un mercado non é un todo homoxéneo. Está composto por centos, miles e incluso millóns de individuos, empresas ou organizacións que son diferentes os uns dos outros en función da súa localización, nivel socioeconómico, cultura, preferencias de compra, estilo, personalidade, capacidade de compra, etc. (Thompson, 2005).

Por eses motivos, nace a imperiosa necesidade de dividir o mercado en grupos cuxos integrantes teñan certas características que os asemellen e permitan á empresa deseñar e implementar unha mestura de mercadotecnia para todo o grupo, pero a un custo moito menor e con resultados máis satisfactorios que se o fixeran para todo o mercado (Thompson, 2005).

Estas análises proporcionarán unha visión máis detallada dos resultados obtidos en función de como se segmentan os usuarios. Ao dividir o conxunto de datos en grupos máis pequenos, poderase comprender mellor a importancia e o impacto de cada un dos grupos nos que se segmentaron.

A comprender en maior profundidade os diferentes segmentos nos que se dividen os usuarios, xurde a oportunidade de deseñar ofertas e solucións máis precisas e relevantes para satisfacer as necesidades e preferencias específicas de cada segmento. Adaptando as estratexias a cada grupo, maximízase a satisfacción do cliente e auméntase a eficacia das accións e decisións tomadas.

A segmentación de datos desenvolve un papel clave na análise de datos e na comprensión dos patróns e comportamentos dos usuarios. Ao dividir o conxunto de datos en grupos máis pequenos e homoxéneos, pódense identificar características distintivas, preferencias e necesidades particulares de cada segmento. Isto permite unha maior personalización das estratexias e accións, o que á súa vez conduce a unha maior efectividade e éxito na satisfacción dos clientes e na consecución dos obxectivos da empresa.



---

## 4 Resultados e discusión

### 4.1 Lectura e resumo xeral de datos.

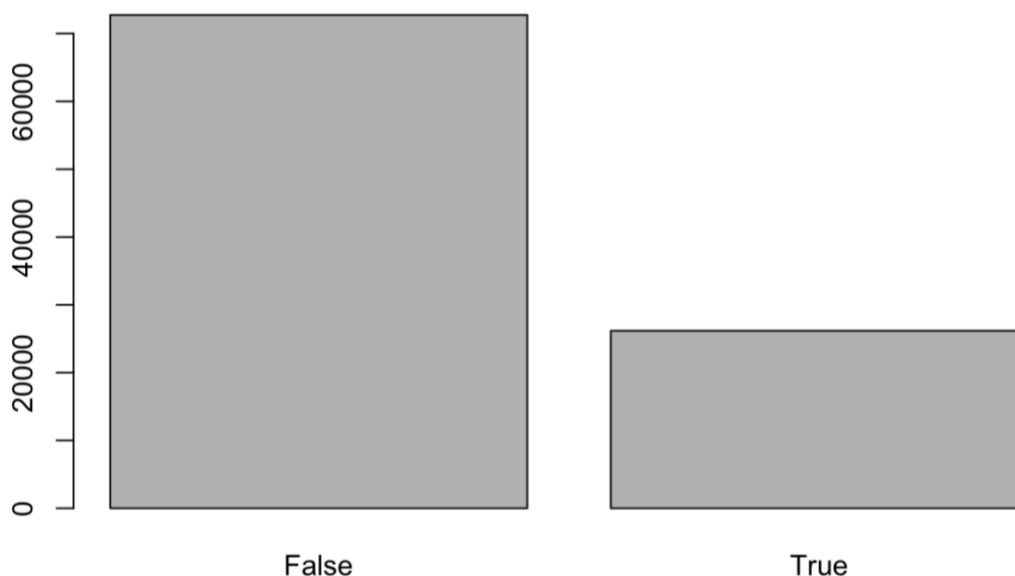
O ficheiro empregado na análise, [users.dataset.public](#), contén un conxunto de datos de usuarios e información sobre o seu comportamento social e de compras. Unha vez importado o ficheiro CSV a R Studio, realizárase un resumo dos datos para obter unha visión xeral das diferentes variables.

### 4.2 Datos descritivos dos usuarios.

A continuación, presentarase unha análise descritiva dos usuarios da web empregando R Studio, co obxectivo de obter unha comprensión detallada das características e comportamentos dos usuarios en dita plataforma. Para lograr este obxectivo, empregáronse diversas funcións dispoñibles en R Studio que permitirán realizar análises estatísticas coas súas respectivas visualizacións. Tamén se levará a cabo unha relación entre variables, para identificar patróns e tendencias de compra e venta entre usuarios.

#### 4.2.1 Uso da app.

A variable *hasAnyApp* permítenos coñecer se o usuario utilizou algunha vez algunha aplicación oficial da tenda. Tras realizar una táboa de frecuencia podemos observar representado no diagrama de barras (*Figura 1*) como *True* aos usuarios que empregan algunha app, e representados como *False* os que non fan uso de app. Como se pode observar, a maioría dos usuarios da web non fan uso de ningunha aplicación da tenda.



*Figura 1. Emprego de algunha app oficial.*

Podemos relacionar o uso de aplicación co número de produtos comprados (*productsBought*) e vendidos (*productsSold*). Agrupando os datos de compra e venta dos usuarios entre os que utilizan e os que non utilizan algunha aplicación concluímos en que os usuarios que usan a aplicación realizan máis compras e vendas que os que non a teñen instalada.

Polo tanto, tendo en conta a todos os usuarios, tanto os que compran e venden como os que non, vemos que a maioría deles non fan uso de ningunha app oficial da marca (*Táboa 2*). Sen embargo, ao

comparar os produtos comprados e vendidos entre os usuarios que utilizan algunha app e os que non, vemos que tanto a media de produtos vendidos como a de produtos comprados é superior en aqueles usuarios que si fan uso de algunha app oficial. Isto pode deberse a que aquelas persoas que só actúan como visitantes e non realizan ningunha compra ou venda prefiren o uso da web xa que non precisa ser descargada e é accesible desde calquera dispositivo. No caso dos usuarios que si son activos na plataforma, é dicir, que compraron ou venderon polo menos un produto, probablemente prefiran o uso da app pola maior capacidade e rendemento.

Táboa 2. Relación entre uso da app e produtos vendidos e comprados.

| Ten App | Produtos vendidos | Media de produtos vendidos | Produtos comprados | Media de produtos comprados |
|---------|-------------------|----------------------------|--------------------|-----------------------------|
| Non     | 3.233             | 0,0444                     | 7.957              | 0,109                       |
| Si      | 8.794             | 0,336                      | 9.049              | 0,346                       |

#### 4.2.2 Título civil.

A variable *civilityTitle* clasifica o estado civil dos usuarios entre *miss* (muller solteira), *mr* (home casado) e *mrs* (muller casada). No diagrama de barras (*Figura 2*) podemos observar que unha gran maioría dos usuarios rexistrados son mulleres casadas.

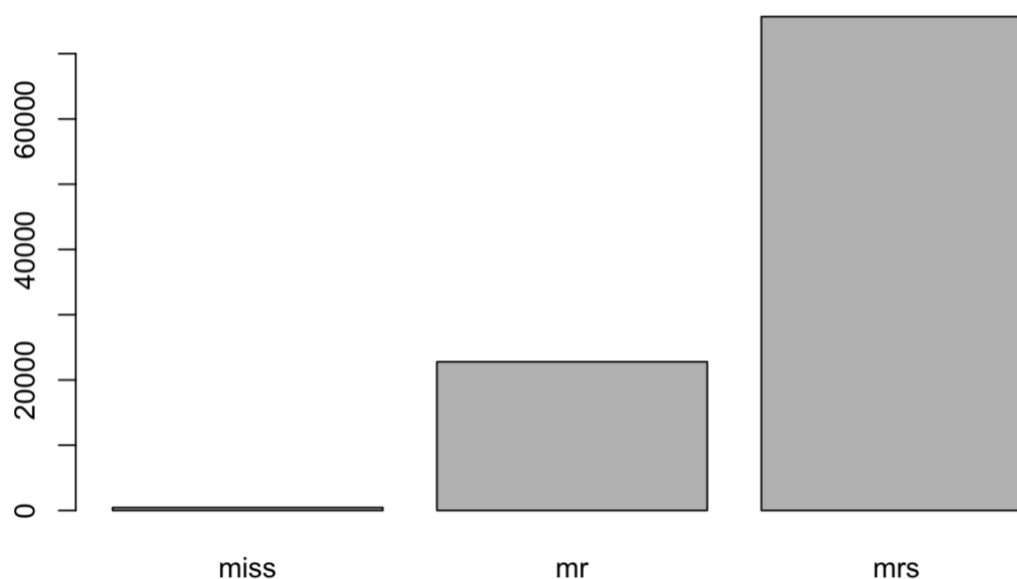


Figura 2. Clasificación dos usuarios en función do estado civil.

Relacionando o estado civil dos usuarios co número de produtos comprados e vendidos obtemos a conclusión de que as mulleres solteiras, a pesar de ser os usuarios menos frecuentes, son as que presentan unha media de compra e venda maior que a do resto dos usuarios, e polo tanto un maior poder adquisitivo. Polo contrario, os usuarios máis activos, as mulleres casadas, compran e venden en menor cantidade que o resto de usuarios (*Táboa 3*).

Isto pode deberse a factores como á dispoñibilidade de tempo, xa que as mulleres solteiras, ao non ter responsabilidades como a de coidar a unha familia, poden ter máis tempo libre para dedicar a

compras e vendas. Da mesma maneira, poden dispoñer de unha maior flexibilidade financeira ao non ter que compartir os seus ingresos nin tomar decisións financeiras en conxunto. Ademais, as mulleres solteiras na maioría dos casos representan un grupo de menor idade que as casadas, polo que en xeral teñen un mellor manexo de plataformas en liña, o que facilita a realización destas transaccións.

Táboa 3. Relación entre o título civil e os produtos vendidos e comprados.

| Título civil | Produtos vendidos | Media de produtos vendidos | Produtos comprados | Media de produtos comprados |
|--------------|-------------------|----------------------------|--------------------|-----------------------------|
| Miss         | 574               | 1,31                       | 287                | 0,657                       |
| Mr           | 2.673             | 0,117                      | 4.262              | 0,187                       |
| Mrs          | 8.780             | 0,116                      | 12.457             | 0,165                       |

#### 4.2.3 Idioma usado na web.

A variable *language* indica o idioma de preferencia do usuario, entre os que atopamos o alemán (de), inglés (en), español (es), francés (fr) e italiano (it). Os resultados da táboa de frecuencias realizada demostran que o idioma máis empregado polos usuarios é o inglés, con case o dobre de usuarios que o segundo idioma máis empregado, o francés. A *Figura 3* representa os resultados da táboa de frecuencias.

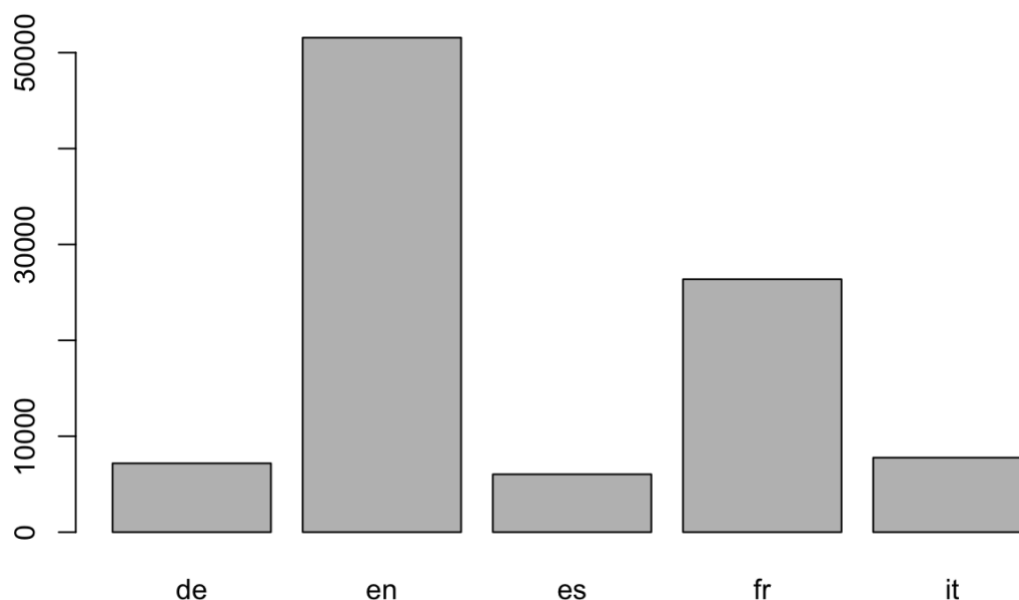


Figura 3. Idiomas de preferencia do usuarios.

Ao relacionar o idioma empregado polos usuarios co número de produtos comprados e vendidos por cada un, vemos que os usuarios do idioma máis empregado, o inglés, non son nin os que teñen maior media de produtos comprados nin vendidos, sendo os usuarios con maior media de produtos comprados os que empregan o alemán, e con maior media de produtos vendidos os que prefiren como idioma o italiano (*Táboa 4*).

Táboa 4. Relación entre o idioma de preferencia e os produtos vendidos e comprados.

| Idioma   | Produtos vendidos | Media de produtos vendidos | Produtos comprados | Media de produtos comprados |
|----------|-------------------|----------------------------|--------------------|-----------------------------|
| Alemán   | 539               | 0,0751                     | 1.705              | 0,238                       |
| Inglés   | 4.877             | 0,0946                     | 8.918              | 0,173                       |
| Español  | 901               | 0,149                      | 797                | 0,132                       |
| Francés  | 3.004             | 0,114                      | 4.386              | 0,166                       |
| Italiano | 2.706             | 0,348                      | 1.200              | 0,155                       |

#### 4.2.4 País de procedencia

Se agrupamos os usuarios segundo o seu lugar de procedencia obtemos como resultado un total de 200 países diferentes (Figura 4).

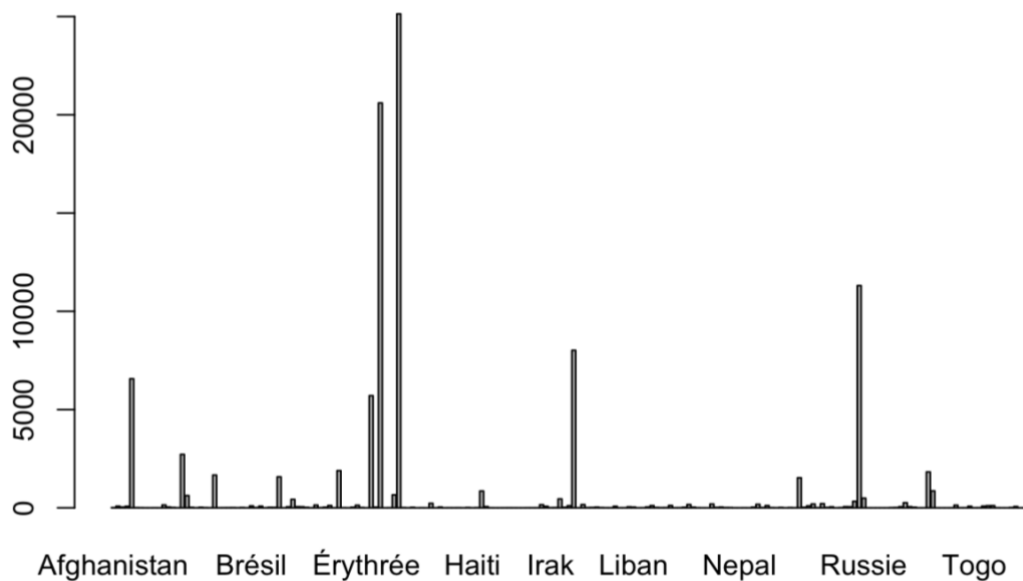


Figura 4. Clasificación dos usuarios segundo o seu país de procedencia.

Debido a que o elevado número de diferentes países impide unha comprensión clara dos resultados, seleccionaremos aqueles países de procedencia que contan con máis de 5.000 usuarios (Táboa 5).

Relacionando o país de procedencia dos usuarios co número de produtos vendidos e comprados, podemos observar que os usuarios máis predominantes son os procedentes de Francia, país que se corresponde co lugar de creación da empresa. Sen embargo este non é o país nin coa media máis alta de produtos vendidos nin comprados, pois no primeiro caso é Italia e no segundo Alemaña.

Táboa 5. Relación entre o país de procedencia dos usuarios e os produtos vendidos e comprados.

| País                  | Número de usuarios | Produtos vendidos | Media de produtos vendidos | Produtos comprados | Media de produtos comprados |
|-----------------------|--------------------|-------------------|----------------------------|--------------------|-----------------------------|
| <b>Alemaña</b>        | 6.567              | 536               | 0,081620222                | 1.635              | 0,24897213                  |
| <b>España</b>         | 5.706              | 990               | 0,173501577                | 1.028              | 0,18016123                  |
| <b>Estados Unidos</b> | 20.602             | 967               | 0,046937191                | 2.370              | 0,11503738                  |
| <b>Francia</b>        | 25.135             | 3.014             | 0,119912473                | 3.573              | 0,14215238                  |
| <b>Italia</b>         | 8.015              | 2.820             | 0,351840299                | 1.221              | 0,15233936                  |
| <b>Reino Unido</b>    | 11.310             | 1.806             | 0,159681698                | 2.174              | 0,19221927                  |

Se comparamos os datos do país de procedencia cos do idioma de preferencia do usuario analizados anteriormente (Táboa 4) vemos que os resultados de ambas análises coinciden, sendo os usuarios que empregan o italiano e proceden de Italia aqueles coa maior media de produtos vendidos, e os usuarios que empregan o alemán e proceden de Alemaña os que presentan a maior media de produtos comprados, o que indica que en xeral os usuarios tenden a seleccionar como idioma de uso da plataforma o oficial do seu país.

#### 4.2.5 Imaxe de perfil

Realizando unha táboa de frecuencia coa variable *hasProfilePicture*, podemos coñecer que cantidade de usuarios a proporcionan. No seguinte diagrama de barras (Figura 5) aparecen representados como *True* os usuarios que teñen imaxe de perfil, e como *False* os que non a teñen. Como podemos observar, unha gran maioría de usuarios proporciona algunha fotografía no perfil.

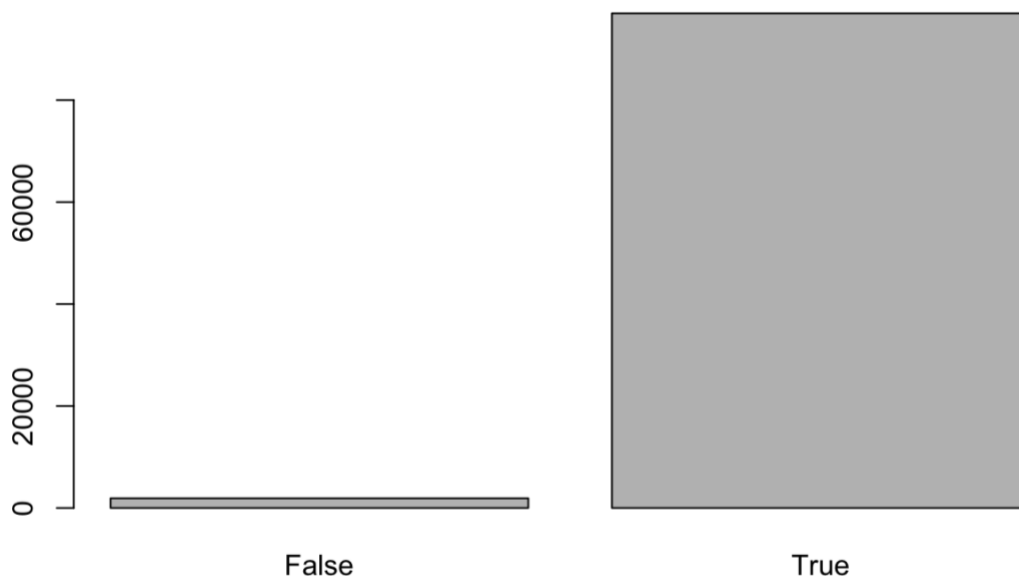


Figura 5. Usuarios con imaxe de perfil.

Se relacionamos o feito de ter ou non ter imaxe de perfil co número de produtos comprados ou vendidos por cada usuario chegamos á conclusión de que os usuarios con foto de perfil son os que máis compran ou venden (Táboa 6). Isto pode deberse a que ao proporcionar unha identidade visual ao usuario, aumenta a sensación de confianza e credibilidade entre os compradores e vendedores.

Táboa 6. Relación entre o emprego de imaxe de perfil e os produtos vendidos e comprados.

| Ten imaxe de perfil | Produtos vendidos | Media de produtos vendidos | Produtos comprados | Media de produtos comprados |
|---------------------|-------------------|----------------------------|--------------------|-----------------------------|
| Non                 | 5.409             | 2,85                       | 3.482              | 1,84                        |
| Sí                  | 6.618             | 0,0682                     | 13.524             | 0,139                       |

#### 4.2.6 Xénero

Mediante a variable *gender*, podemos coñecer que cantidade de usuarios se identifican baixo o xénero feminino ou baixo o xénero masculino. No diagrama de barras (Figura 6) aparecen representados coa letra F os usuarios de xénero feminino, e como M os de xénero masculino. Pódese observar unha gran diferenza entre os dous xéneros, sendo moito maior a proporción de usuarios de xénero feminino.

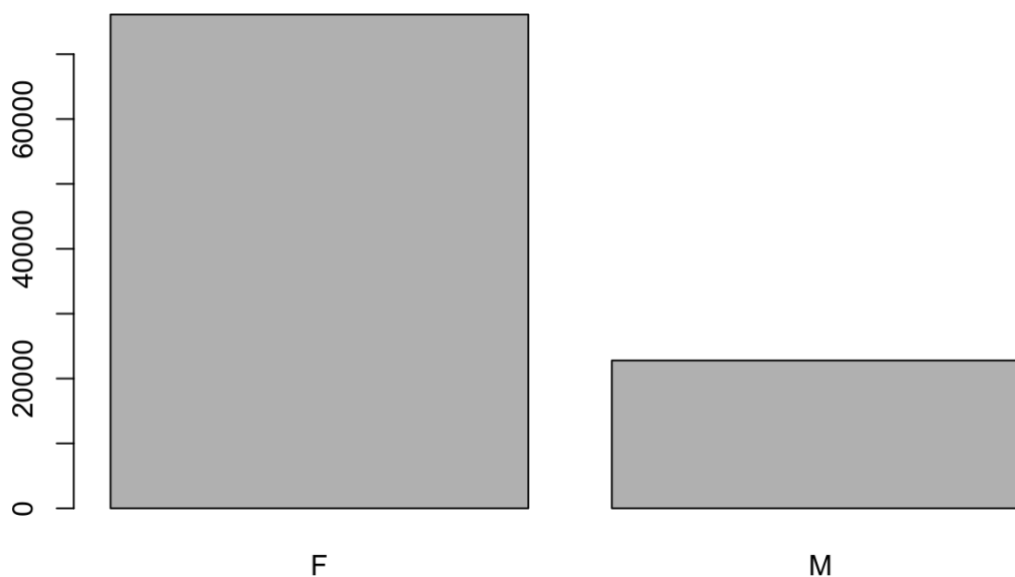


Figura 6. Distribución dos usuarios segundo o seu xénero

Relacionando o xénero dos usuarios co número de produtos comprados e vendidos obtemos a conclusión de que os usuarios de xénero feminino son os que presentan unha media de venta superior. Sen embargo, o xénero masculino, a pesar de ser os usuarios menos frecuentes, son os que presentan unha media de compra maior que a do xénero feminino (Táboa 7).

Poderíamos relacionar o dato da media de venta coa análise dos usuarios segundo o seu título civil, no que se obtivo o resultado que indicaba que os usuarios que máis vendas realizaban eran as mulleres solteiras. En canto á media de produtos comprados, neste caso a superioridade do xénero masculino pode deberse a que as mulleres solteiras eran as que máis compraban, pero as que menos o facían eran as mulleres casadas, polo que ao unir os dous grupos no xénero feminino pode provocar que esta media diminúa.

Táboa 7. Relación entre o xénero dos usuarios e os produtos vendidos e comprados.

| Xénero    | Produtos vendidos | Media de produtos vendidos | Produtos comprados | Media de produtos comprados |
|-----------|-------------------|----------------------------|--------------------|-----------------------------|
| Feminino  | 9.354             | 0,123                      | 12.744             | 0,167                       |
| Masculino | 2.673             | 0,117                      | 4.262              | 0,187                       |

### 4.3 Correlación entre variables.

Para a análise da correlación entre variables, elimináronse aquelas variables que non eran o suficientemente relevantes. Por un lado, o identificador de cada usuario (*identifierHash*) e todas aquelas variables que non son numéricas (*type*, *country*, *language*, *gender*, *civilityGenderId*, *civilityTitle*, *hasAnyApp*, *hasAndroidApp*, *hasIosApp*, *hasProfilePicture*, *countryCode*), e por outro as variables *seniorityAsMonths* e *seniorityAsYears*, xa que representan o mesmo que a variable *seniority* pero en diferentes unidades de tempo.

### 4.3.1 Todos os usuarios.

A continuación podemos observar o resultado da correlación (Figura 7) unha vez realizada a limpeza de variables:

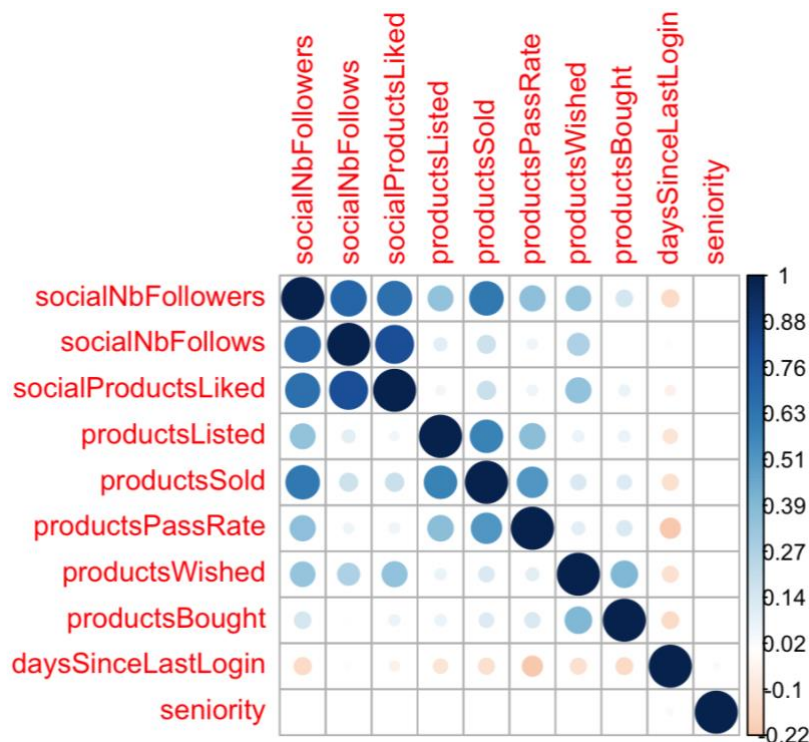


Figura 7. Usuarios da marca: correlación entre variables.

Para cada usuario, o número de usuarios seguidos presenta unha correlación positiva co número de seguidores, e o mesmo sucede co número de produtos que lle gusta a cada usuario.

Desde o punto de vista dos vendedores, os produtos vendidos están positivamente correlacionados coa taxa de aprobación do produto e co número de produtos subidos á páxina, así como co número de seguidores. Isto quere dicir que aqueles usuarios que publican máis produtos son os máis propensos a vender máis e, ao mesmo tempo, a ter máis seguidores. De maneira máis débil, as vendas presentan unha correlación positiva co número de usuarios seguidos, os produtos que lles gustan e os que desexan, e os produtos comprados, o que indica unha lixeira relación tamén entre estas variables.

Por outro lado, as compras correlaciónanse positivamente cos produtos desexados, indicando que aqueles usuarios que engaden máis produtos á lista de desexos son os que máis compran. Sen embargo, a correlación entre os produtos comprados e os produtos gustados é moito máis débil, indicándonos unha clara diferenza entre os produtos desexados e gustados por cada usuario, pois estes non engaden os mesmos produtos a ambas listas. Con estes datos podemos coñecer mellor que produtos son máis propensos a ser comprados por cada consumidor, sendo os da lista de produtos desexados os máis probables a ser comprados e os da lista de gustados aqueles que, a pesar de chamar a atención do cliente, supoñan unha menor probabilidade de rematar nunha decisión de compra.

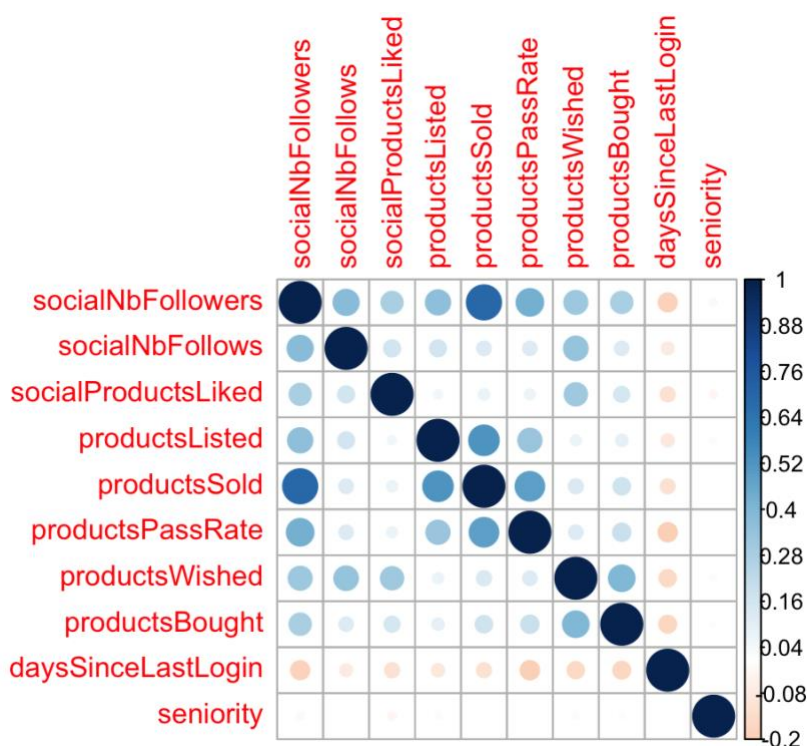
No caso da actividade dos usuarios, a variable *daysSinceLastLogin* está correlacionada negativamente co resto das variables, o que nos confirma que os usuarios que máis tempo levan sen utilizar a páxina son os menos activos nela, sendo os que realizan menos compras e vendas.



Tras analizar esta matriz de correlación chegamos á conclusión de que o número de seguidores é un factor moi importante á hora de vender máis produtos. Un maior número de seguidores indica que ese usuario é popular entre o resto, que os seus produtos destacan entre os demais. Ademais, o resto dos usuarios poden sentirse máis seguros á hora de realizar unha compra con un perfil con moitos seguidores, xa que indica que moitos outros confiaron nos seus produtos. Tamén se chega á conclusión de que para ter unha idea de que tipo de produtos compraría cada consumidor é importante ter en conta a lista de desexos de cada un, sendo os incluídos nela os produtos con máis probabilidade de ser comprados. Esta información tamén pode ser de axuda á hora de recomendar outros produtos que poderían resultar interesantes para cada usuario, xa que proporciona unha idea dos seus gustos ou necesidades.

### 4.3.2 Usuarios que realizan polo menos unha compra.

Son moitas as persoas que, a pesar de ser usuarios rexistrados da marca, non realizan ningunha compra. Para levar a cabo unha análise máis precisa desde o punto de vista dos compradores, na *Figura 8* podemos ver a correlación entre variables tendo en conta só aqueles usuarios que realizaron unha ou máis compras:



*Figura 8. Usuarios que realizaron polo menos unha compra: correlación entre variables.*

Analizando esta matriz de correlación dende a perspectiva dos compradores, sorprende ver que o número de produtos comprados e o número de usuarios seguidos non presentan unha correlación positiva moi forte, xa que tendo en conta que o número de seguidores inflúe na venda de máis produtos, o normal sería ver que produtos comprados e contas seguidas se correlacionasen máis. A resposta pode atoparse na fidelidade a certos usuarios, pois un maior número de compras non ten por que supoñer un maior número de usuarios seguidos, xa que un usuario pode mercar en repetidas ocasións produtos dun mesmo vendedor, converténdose nun cliente fiel e confiando na calidade dos seus produtos.

### 4.3.3 Usuarios que realizan polo menos unha venta.

Ao igual que na análise de correlacións dos usuarios que realizaron algunha compra, na *Figura 9* podemos ver unha matriz de correlación elaborada a partir dos datos de aqueles usuarios que realizaron polo menos unha venta:

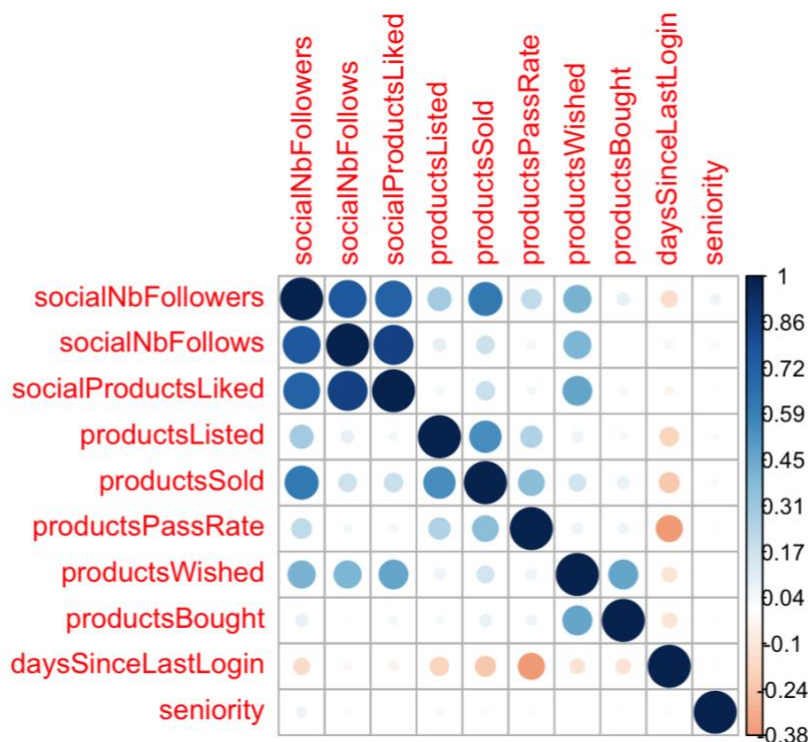


Figura 9. Usuarios que realizaron polo menos unha venta: correlación entre variables.

Se comparamos esta matriz coa *Figura 7*, na que se tiñan en conta todos os usuarios, non se aprecian cambios significativos. De igual maneira, analizando as correlacións dende unha perspectiva de vendas e con unha mostra de datos máis concreta, podemos observar que os seguidores dun usuario son un factor moi importante nas súas vendas, e que un maior número de produtos subidos tamén supón un maior número de produtos vendidos. Algo curioso é que a taxa de aprobación dos produtos non presente unha correlación aínda máis forte cos produtos vendidos, o que quere dicir que algúns produtos, a pesar de ser mercados por algún cliente, non puideron ser entregados a este porque non cumprían os requirimentos necesarios.

## 4.4 Compradores e vendedores segundo o seu título civil: diagrama de caixa.

### 4.4.1 Usuarios que realizaron polo menos unha compra.

Para obter unha visión xeral da distribución dos valores de compra dos usuarios representados baixo os distintos títulos civís, realizamos un diagrama de caixa (*Figura 10*):

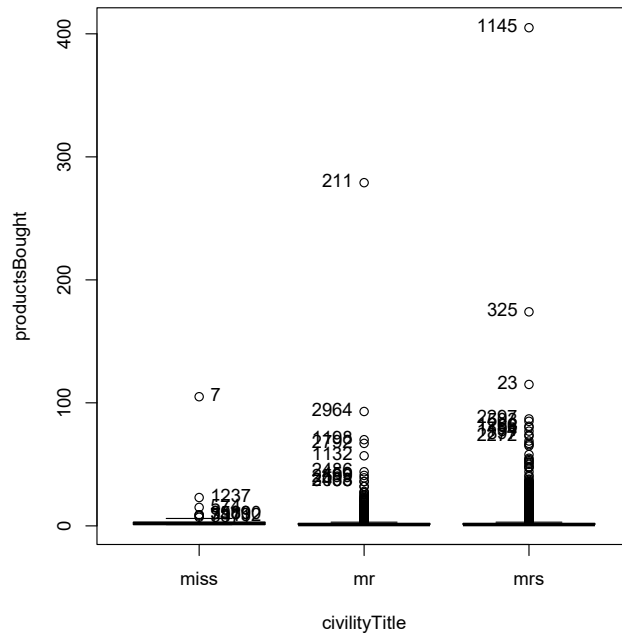


Figura 10. Usuarios que realizaron polo menos unha compra: número de produtos comprados segundo o título civil do usuario.

No diagrama obsérvanse varios valores atípicos, sendo o valor máis disperso o correspondente a unha muller casada con un número de produtos mercados de aproximadamente 400 produtos. Debido ao gran número e dispersión dos valores atípicos, resulta difícil interpretar os resultados do diagrama, polo que na *Figura 11* vemos os resultados correspondentes a un intervalo menor (usuarios que realizaron entre 1 e 20 compras):

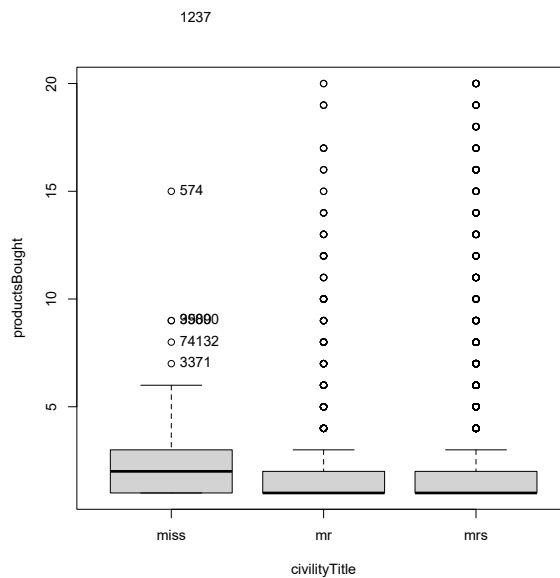


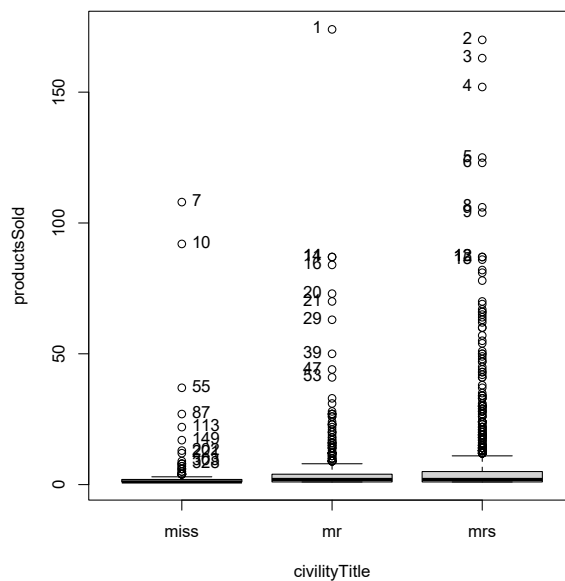
Figura 11. Usuarios que realizaron polo menos unha compra: número de produtos comprados segundo o título civil do usuario, tomando só os valores inferiores a 20.

A *Figura 11* móstranos que case o 100% tanto dos homes casados (mr) como das mulleres casadas (mrs) realiza un número de compras inferior a 4, sitúándose a mediana nun valor próximo a unha compra por usuario. No caso das mulleres solteiras (miss) os valores elévanse un pouco máis con case a totalidade dos datos nun intervalo inferior a 6 e a mediana entre 2 e 3 produtos comprados.

Os homes e mulleres casadas presentan uns valores moi similares, nos que a maioría dos usuarios realiza un número reducido de compras, pero nos que atopamos un número de valores atípicos moi disperso. Por outro lado, a maioría das mulleres solteiras realiza máis compras que os outros dous grupos. Sen embargo, neste caso os valores atípicos que podemos atopar non chegan a ser tan dispersos como os anteriores, situándose o máis elevado nun valor próximo a 100, como podemos ver na *Figura 10*.

#### 4.4.2 Usuarios que realizaron polo menos unha venta.

Neste caso, realizamos un gráfico de caixa analizando o número de vendas dos usuarios distribuídos segundo o seu título civil e tendo só en conta a aqueles que realizaron unha ou máis vendas (*Figura 12*):



*Figura 12. . Usuarios que realizaron polo menos unha venta: número de produtos comprados segundo o título civil do usuario.*

Ao contrario que na análise do número de compras, neste caso o valor atípico máis elevado pertence ao grupo dos homes casados, con un valor superior a 150 vendas. Ao igual que no caso da *Figura 10*, resulta complicado interpretar os resultados do diagrama, polo que na *Figura 13* podemos ver o número de vendas (nun intervalo de entre 1 e 20) que realizaron os usuarios identificados baixo cada título civil:

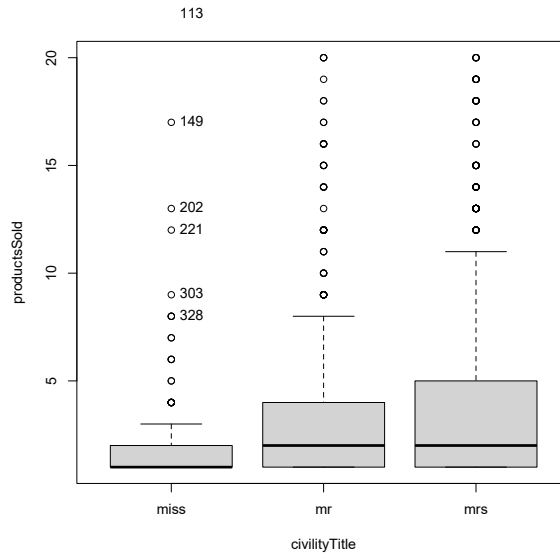


Figura 13. Usuarios que realizaron polo menos unha venta: número de produtos comprados segundo o título civil do usuario, tomando só os valores inferiores a 20.

Os resultados son moi diferentes aos obtidos no gráfico referido ás compras dos usuarios, pois neste caso, como vemos na *Figura 13*, son as mulleres solteiras as que presentan valores inferiores ao resto dos grupos. Agora, as mulleres e homes casados presentan unha mediana similar, de aproximadamente 3 produtos vendidos. No caso dos homes casados, case o 100% dos datos atópanse en valores inferiores a 9, e no das mulleres casadas, este número aumenta ata chegar aos 11-12 produtos. Outro dato a destacar é que aínda que no grupo dos homes casados atopamos o valor atípico máis elevado, o resto dos valores correspóndense todos a números inferiores a 100. No caso das mulleres casadas, estes valores están máis repartidos, estando o máis elevado moi preto do maior valor dos homes casados.

Por outro lado, as mulleres solteiras presentan unha mediana de aproximadamente 2 produtos vendidos, e case a totalidade dos seus valores representan un número inferior a 4.

#### 4.5 Porcentaxes de usuarios activos e usuarios rexistrados que non realizan ningunha transacción agrupados segundo o seu xénero e título civil.

Como xa se viu anteriormente, non todos os usuarios rexistrados na plataforma chegan a realizar transaccións nela, polo que resulta moi interesante coñecer que grupos son máis e menos activos para así poder desenvolver estratexias centradas en incitar a aqueles que nunca o fixeron a realizar compras e vendas. Para iso, neste apartado daranse a coñecer, mediante gráficos de barras, as porcentaxes de usuarios que realizan ou non algunha compra ou venta, clasificándoos tanto por xénero como polo título civil, para así obter datos máis concretos.

#### 4.5.1 Usuarios que realizan polo menos unha compra.

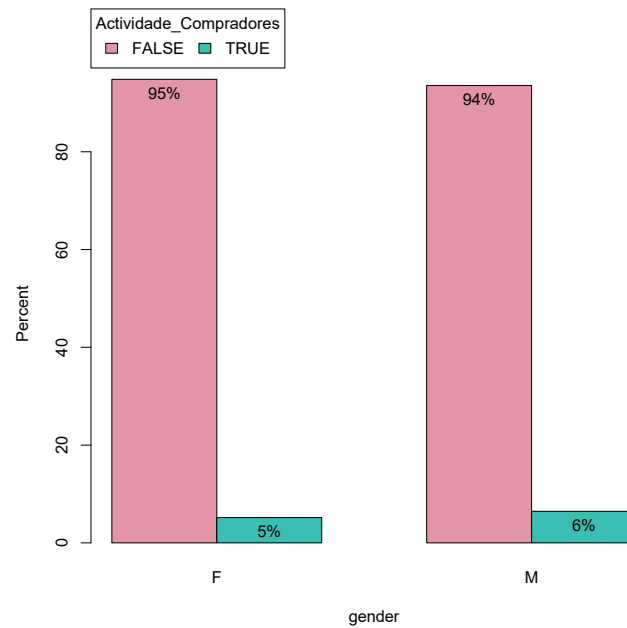


Figura 14. Actividade de compra dos usuarios Segundo o seu xénero.

Na Figura 14 podemos ver representada a actividade de compra en cada un dos xéneros dos usuarios. No caso do xénero feminino vemos que só un 5% dos usuarios chegou a realizar algunha compra. No caso do xénero masculino a porcentaxe de usuarios que chegaron a realizar polo menos unha compra é do 6%.

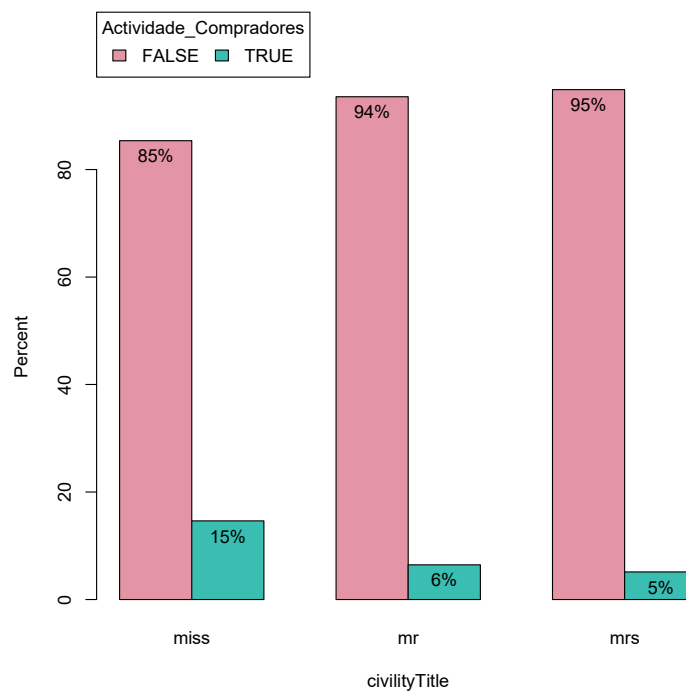
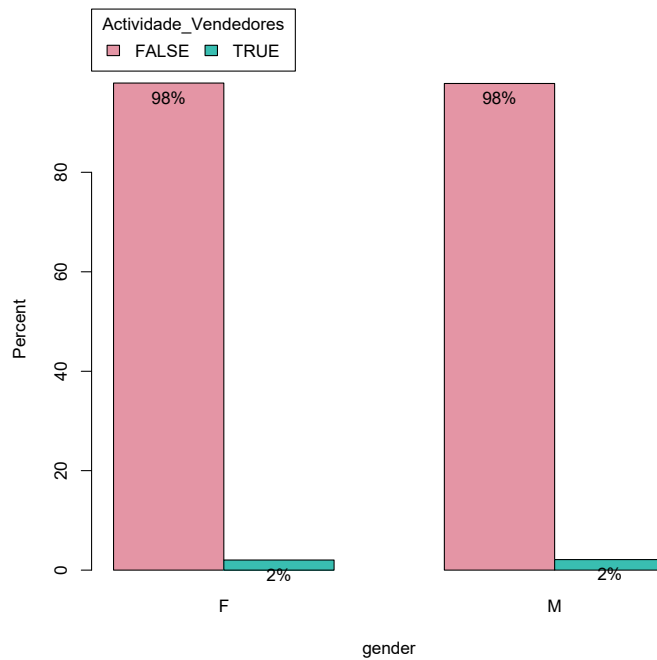


Figura 15. Actividade de compra dos usuarios segundo o seu título civil.

---

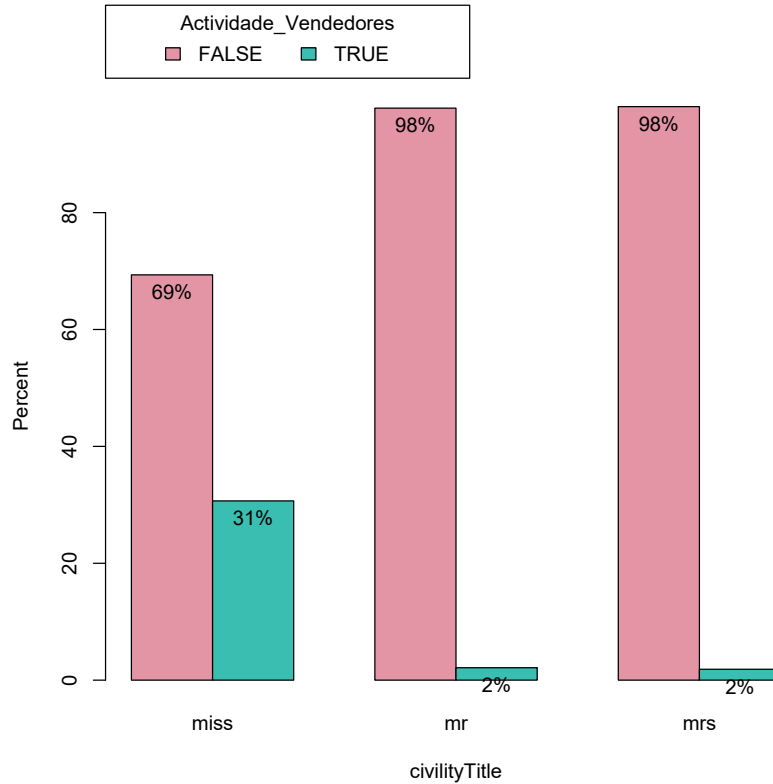
Se clasificamos os usuarios segundo o seu título civil, a *Figura 15* móstranos que a porcentaxe de usuarios que realizan algunha compra sitúase no 6% para os homes, coincidindo cos datos do xénero masculino, pois só se fai uso dun título civil para todo o xénero. En canto ás mulleres, vemos que o 15% das mulleres solteiras chega a realizar algunha compra, mentres que no caso das mulleres casadas, este número non supera o 5%.

#### 4.5.2 Usuarios que realizan algunha venta.



*Figura 16. Actividade de venta dos usuarios segundo o seu xénero.*

A *Figura 16* contén as porcentaxes dos usuarios que realizaron ou non algunha venta na plataforma, clasificándoos segundo o seu xénero. Neste caso vemos que os datos coinciden nos dous xéneros, pois tanto os usuarios do xénero feminino como do masculino que chegan a realizar algunha venta supoñen un 2% dos usuarios de cada xénero.



*Figura 17. Actividade de venta dos usuarios segundo o seu título civil.*

Se analizamos as porcentaxes da actividade de vendas clasificando os usuarios segundo o seu título civil (*Figura 17*) vemos que, como xa se explicou previamente os valores do xénero masculino e dos homes coinciden. No caso das mulleres, só un 2% das casadas chega a realizar algunha compra. Por outro lado, un 31% das mulleres solteiras realiza vendas na plataforma.



### 4.5.3 Usuarios que realizan polo menos unha compra e unha venta.

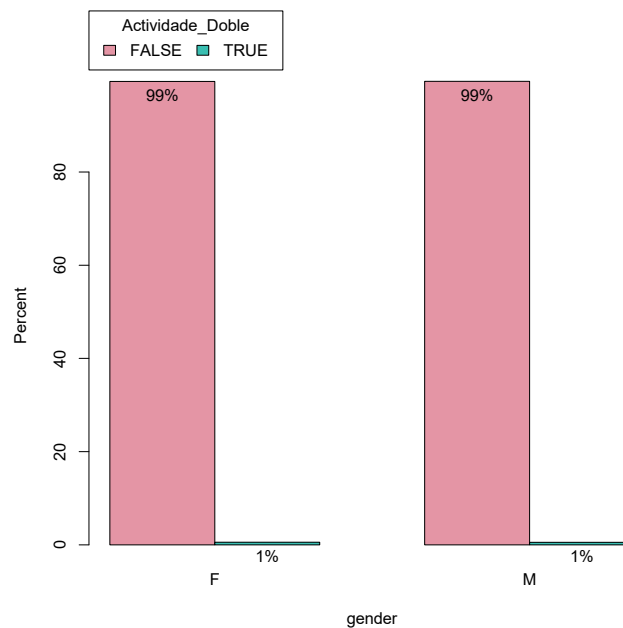


Figura 18. Actividade de compra e venta dos usuarios segundo o seu xénero.

A Figura 18 correspóndese coas porcentaxes de usuarios, clasificados segundo o seu xénero, que realizan ou non polo menos unha compra e unha venta. Como vemos só o 1% de cada un dos xéneros levou a cabo como mínimo unha compra e unha venta, pois son moitos os usuarios que empregan a plataforma só para realizar compras ou só para vender, xunto con aqueles que actúan unicamente como visitantes e non realizan ningún tipo de transacción.

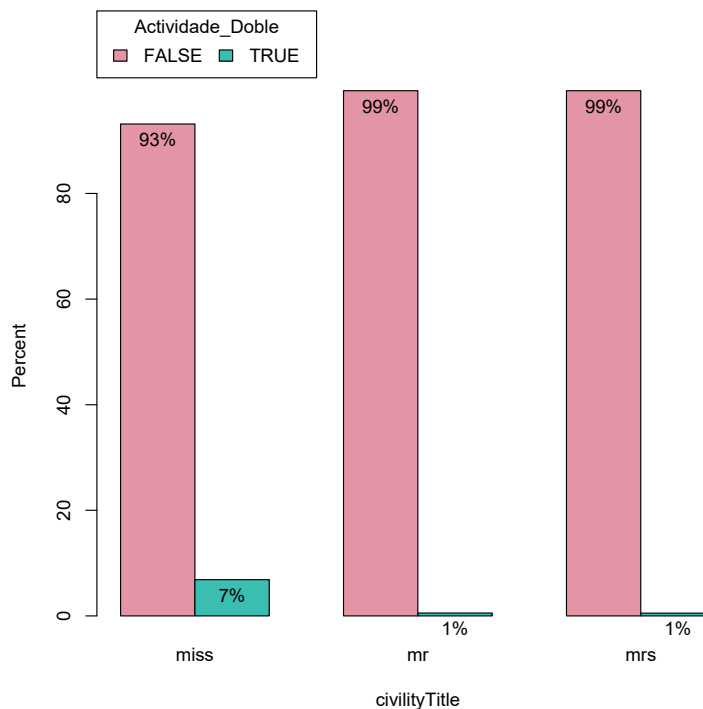


Figura 19. Actividade de compra e venta dos usuarios segundo o seu título civil.

---

Agrupando os datos de aqueles usuarios que realizaron polo menos tanto unha compra como unha venda segundo o seu estado civil na *Figura 19* podemos ver que, fronte ao 1% tanto dos homes como das mulleres casadas, un 7% das mulleres solteiras levaron a cabo ambos os dous tipos de transaccións, sendo os usuarios máis activos na plataforma.

Tras a análise das porcentaxes de compra e venda nos gráficos anteriores, tanto por separado como de maneira conxunta, chégase á conclusión de que levando a cabo este tipo de análise agrupando os usuarios por xénero non nos proporciona indicacións moi precisas, xa que os datos son moi similares para ambos os dous xéneros. Sen embargo, ao analizar os datos clasificándoos polo título civil dos usuarios obtemos conclusións máis claras, nas que observamos que as mulleres solteiras son os usuarios con unha actividade de compra e venda máis destacable, xa que o número de usuarios do grupo que realiza calquera tipo de transacción é o máis elevado nos tres casos.

Analizando a diferenza entre os resultados dos gráficos nos que se dividen os usuarios por xénero e nos que se dividen por título civil, obsérvase que no caso do título civil, ao dividir os usuarios en máis grupos, os resultados amosan de xeito máis preciso o comportamento de cada un dos grupos. Isto quere dicir que mediante a segmentación de datos podemos coñecer as actitudes e preferencias de grupos de consumidores con características en común, o que permite dirixirse a cada un deles dunha maneira personalizada e así satisfacer mellor as necesidades de cada un. En este caso, podemos ver que as mulleres solteiras presentan unha actividade diferente das mulleres casadas. Sen embargo, os homes aparecen agrupados nun único conxunto, polo que obtemos unha visión xeral do seu comportamento, a pesar de que pode que entre eles existan numerosas diferenzas.

---

## 5 Conclusións

Durante o transcurso deste traballo de fin de grao explorouse de maneira exhaustiva o emprego de ferramentas de análises de datos e o software R para analizar os datos dos usuarios dunha empresa C2C fundada en Francia e expandida por todo o mundo, obtidos a través dunha plataforma web na que se reúnen numerosas bases de datos públicas. A través de este estudo, logrouse obter unha visión detallada e significativa da percepción da empresa e o seu papel no mercado.

O emprego de *business analytics* demostrou ser fundamental na comprensión e avaliación dos datos empresariais. Mediante a análise de datos co emprego do software R, unha ferramenta poderosa e versátil, para levar a cabo diversas técnicas de análises de datos, como análises descritivas ou correlacións entre variables, puidéronse identificar patróns, tendencias e relacións clave que inflúen na percepción da empresa por parte dos seus clientes e usuarios.

Os resultados obtidos proporcionan unha valiosa información sobre a percepción de esta empresa C2C. Puidéronse identificar diferentes comportamentos nos clientes, así como tanto os puntos fortes da empresa ou as áreas que requiren melloras. Isto permitiría á empresa tomar decisións informadas e estratéxicas para fortalecer a súa posición no mercado, mellorar a satisfacción do cliente e optimizar as súas operacións.

Tras levar a cabo a análise dos diferentes datos obtivéronse unha serie de conclusións, entre as máis relevantes atopamos as seguintes:

1. Os usuarios que chegan a realizar algunha transacción empregan máis a app da tenda que a web, o que nos indica que á hora de realizar compras e vendas, ben sendo por facilidade de uso, seguridade ou comodidade, os usuarios prefiren esta opción fronte á web.
2. Os usuarios máis numerosos son os que teñen o inglés como idioma de preferencia. Sen embargo, os que empregan o alemán e o italiano son os máis activos, os primeiros en compras e os segundos en vendas. Os datos das transaccións segundo o idioma coinciden cos das transaccións segundo o país de procedencia, o que nos indica que os usuarios, normalmente, prefiren o uso do idioma oficial do seu país.
3. Os usuarios con imaxe de perfil realizan máis compras e vendas que os que non a teñen, polo que se entende que un usuario con imaxe de perfil é percibido de maneira máis positiva polos compradores.
4. O feito de ter máis seguidores ten un efecto positivo nas vendas.
5. Á hora de suxerir outras opcións aos compradores, é relevante ter en conta a lista de desexos de cada un e ofrecer suxerencias en base a esta.
6. As mulleres casadas forman o grupo máis numeroso entre os usuarios. Sen embargo, son as mulleres solteiras as máis activas na plataforma, polo que convén deseñar estratexias diferentes para cada un dos grupos.

Ademais, destacouse a importancia da segmentación de datos na análise empresarial. Mediante a división dos datos en segmentos máis pequenos e homoxéneos, puidéronse identificar características e patróns específicos dentro de cada grupo, o que permite unha personalización e adaptación máis efectiva das estratexias empresariais. Este reflexión puido obterse ao ver como as porcentaxes das actividades de compra e venda variaban en función da segmentación elixida, sendo os resultados máis precisos canto máis se segmentaban os datos.

Este estudo demostrou o valor e a relevancia da análise de datos e a *business analytics* no entorno empresarial. Os datos recompilados e analizados proporcionaron unha base sólida para a toma de

---

decisións estratéxicas e a mellora continua da empresa. Ademais, o emprego do software R demostrou ser unha ferramenta poderosa e eficiente para a análise de datos empresariais.

Como recomendación para futuras investigacións, Estas técnicas poden proporcionar una visión aínda máis profunda e precisa da percepción da empresa e axudar a identificar oportunidades emerxentes e desafíos potenciais.

Como ampliacións ao traballo realizado, sería interesante coñecer que valor deben obter os produtos na taxa de aprobación para poder ser entregados aos clientes, xa que coñecemos que de media os produtos teñen unha puntuación dun 82,23 sobre 100. Sen embargo, non se coñece se só os que conseguen un valor de 100 a superan ou se existe outro valor mínimo para que o produto poida seguir o seu camiño. Esta dúbida nace tras observar que a correlación entre os produtos vendidos e a taxa de aprobación non era moi forte.

Tamén sería interesante levar a cabo análises similares empregando como base os dous conxuntos de datos pertencentes á mesma empresa que non se chegaron a analizar. Un deses conxuntos de datos consiste nunha comparación dos usuarios vendedores por país e xénero. Esa información podería empregarse para determinar que podería aprender un país ou xénero de outro. O conxunto de datos restante contén datos estatísticos sobre os países con mellores vendedores, sendo estes os que realizaron 20 ou máis vendas de produtos. A análise destes datos permitiría coñecer que leva a eses usuarios a realizar numerosas vendas, e así desenvolver estratexias para poder incrementar o número de vendas do resto dos usuarios. Ademais, proporcionaría datos moi interesantes sobre unha pequena proporción dos usuarios, analizando o comportamento de ese que aparecían representados como valores atípicos nos diagramas de caixa.

En resumen, este traballo demostrou que a análise da percepción dunha empresa a través do uso de *business analytics* e o software R é esencial para comprender mellor a súa posición no mercado, identificar áreas de mellora e tomar decisións estratéxicas informadas. Estas ferramentas e enfoques poden ser aplicados en diversos sectores e organizacións, brindando una vantaxe competitiva e promovendo un crecemento sostible no entorno empresarial actualmente dinámico e desafiante.

---

## Referencias bibliográficas

- Alguacil Jiménez, M. (2017). *Análisis de la percepción de marca y su relación con el rendimiento percibido en servicios deportivos: comparación entre público y privado*. <https://roderic.uv.es/bitstream/handle/10550/59685/TESIS%20DOCTORAL%20Mario%20Alguacil%20Jiménez.%20Roderic%20UV.pdf?sequence=1&isAllowed=y>
- Bustamante Lazcano, J. A. (2023). *Métodos de recolección de información para el análisis de la satisfacción del cliente*. <https://revistas.uvp.mx/index.php/ixmati/article/view/7/6>
- Clases online de Data analyst: Técnicas y herramientas de informes avanzado | LinkedIn Learning, antes Lynda.com*. (2021). <https://www.linkedin.com/learning/data-analyst-tecnicas-y-herramientas-de-informes-avanzado>
- Crandi, A. (2022). *Qué es la percepción de marca y cómo medirlo*. <https://academia.crandi.com/marketing-digital/percepcion-de-marca/>
- Delen, D., & Ram, S. (2018). Research challenges and opportunities in business analytics. *Journal of Business Analytics*, 1, 2-12. <https://doi.org/10.1080/2573234x.2018.1507324>
- Farreen Y.L., C., & Lee, G. C. (2011). *Customer-based brand equity: a literature review*. [https://www.researchgate.net/profile/Fayrene-Chieng/publication/312661241\\_Customer-based\\_brand\\_equity\\_a\\_literature\\_review/links/5aeba9b7a6fdcc8508b6dddde/Customer-based-brand-equity-a-literature-review.pdf](https://www.researchgate.net/profile/Fayrene-Chieng/publication/312661241_Customer-based_brand_equity_a_literature_review/links/5aeba9b7a6fdcc8508b6dddde/Customer-based-brand-equity-a-literature-review.pdf)
- Fisher, N. I., & Kordupleski, R. E. (2018). Good and bad market research: A critical review of Net Promoter Score. *Applied Stochastic Models in Business and Industry*, 35, 138-151. <https://doi.org/10.1002/asmb.2417>
- Kaggle: todo lo que hay que saber sobre esta plataforma*. (2021). <https://datascientest.com/es/kaggle-todo-lo-que-hay-que-saber-sobre-esta-plataforma#:~:text=Kaggle%20es%20una%20plataforma%20web>
- Lisboa, R. (2019). *¿Qué es un focus group o grupo focal y para qué sirve?* <https://rockcontent.com/es/blog/focus-group/>
- Potter, K. (s. f.). *Methods for Presenting Statistical Information: The Box Plot*. [https://people.montefiore.uliege.be/kvansteen/MATH0008-2/ac20122013/Class2Oct/Supplementary%20info\\_AppendixA\\_Methods%20for%20presenting%20statistical%20information.pdf](https://people.montefiore.uliege.be/kvansteen/MATH0008-2/ac20122013/Class2Oct/Supplementary%20info_AppendixA_Methods%20for%20presenting%20statistical%20information.pdf)
- Reza Jalilvand, M., Shekarchizadeh Esfahani, S., & Samiei, N. (2010). *Electronic word-of-mouth: challenges and opportunities*. [https://www.sciencedirect.com/science/article/pii/S1877050910003832?ref=pdf\\_download&fr=RR-2&rr=7dac50f978a61ba7](https://www.sciencedirect.com/science/article/pii/S1877050910003832?ref=pdf_download&fr=RR-2&rr=7dac50f978a61ba7)
- Romaniuk, J., & Sharp, B. (2003). Measuring brand perceptions: Testing quantity and quality. *Journal of Targeting, Measurement and Analysis for Marketing*, 11, 218-229. <https://doi.org/10.1057/palgrave.jt.5740079>
- Sancho Esper, F. M. (2017). *Percepción del e-commerce en el sector textil español*. [https://rua.ua.es/dspace/bitstream/10045/67267/1/Percepcion\\_del\\_ecommerce\\_en\\_el\\_sector\\_textil\\_espan\\_Reig\\_Torregrosa\\_Rafael.pdf](https://rua.ua.es/dspace/bitstream/10045/67267/1/Percepcion_del_ecommerce_en_el_sector_textil_espan_Reig_Torregrosa_Rafael.pdf)
- Santana, A. (s. f.). *Introducción al uso de R-commander*. <https://www.uv.es/conesa/CursoR/material/Manual-R-commander.pdf>
- Sordo, A. I. (2021). *Cómo medir y mejorar la percepción de los clientes*. <https://blog.hubspot.es/service/mejorar-percepcion-clientes>

---

Thompson, I. (2005). *La Segmentación del Mercado*.  
[https://www.accioneduca.org/admin/archivos/clases/material/segmentacion-y-mercado\\_1563828020.pdf](https://www.accioneduca.org/admin/archivos/clases/material/segmentacion-y-mercado_1563828020.pdf)

*What is R?* (s. f.). <https://mran.microsoft.com/documents/what-is-r>

Zambrano, R. (2019). *Qué es R y por qué utilizarlo*. <https://openwebinars.net/blog/que-es-r-y-por-que-utilizarlo/>

---

## Anexo

### Procesamento de datos

Antes de comezar, é necesarios descargar e instalar o software R de <https://cran.r-project.org/mirrors.html>.

A continuación, realizaranse as análises pertencentes aos apartados 'lectura e resumo xeral de datos', 'datos descriptivos dos usuarios' e 'correlación entre variables' empregando R Studio, o cal pode ser descargado de <https://www.rstudio.com/products/rstudio/>.

Lectura e resumo xeral de datos:

- Unha vez aberto R Studio, cargamos os datos:

```
datos <- read.csv("~/Desktop/TFG/datos/6M-0K-99K.users.dataset.public.csv")
```

- Facemos un resumo dos datos:

```
summary(datos)

##  identifierHash          type          country          language
##  Min.   :-9.223e+18   Length:98913   Length:98913   Length:98913
##  1st Qu.: -4.623e+18   Class :character   Class :character   Class :character
##  Median :-1.338e+15   Mode  :character   Mode  :character   Mode  :character
##  Mean   :-6.692e+15
##  3rd Qu.: 4.616e+18
##  Max.    : 9.223e+18
##  socialNbFollowers socialNbFollows   socialProductsLiked productsListed
##  Min.   : 3.000   Min.   : 0.000   Min.   : 0.00   Min.   : 0.0000
##  1st Qu.: 3.000   1st Qu.: 8.000   1st Qu.: 0.00   1st Qu.: 0.0000
##  Median : 3.000   Median : 8.000   Median : 0.00   Median : 0.0000
##  Mean   : 3.432   Mean   : 8.426   Mean   : 4.42   Mean   : 0.0933
##  3rd Qu.: 3.000   3rd Qu.: 8.000   3rd Qu.: 0.00   3rd Qu.: 0.0000
##  Max.   :744.000   Max.   :13764.000   Max.   :51671.00   Max.   :244.0000
##  productsSold      productsPassRate  productsWished    productsBought
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.000   Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.000   1st Qu.: 0.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.000   Median : 0.0000
##  Mean   : 0.1216   Mean   : 0.8123   Mean   : 1.563   Mean   : 0.1719
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.000   3rd Qu.: 0.0000
##  Max.   :174.0000   Max.   :100.0000   Max.   :2635.000   Max.   :405.0000
##  gender            civilityGenderId  civilityTitle     hasAnyApp
##  Length:98913     Min.   :1.000   Length:98913     Length:98913
##  Class :character  1st Qu.:2.000   Class :character  Class :character
##  Mode  :character  Median :2.000   Mode  :character  Mode  :character
##                  Mean   :1.774
##                  3rd Qu.:2.000
##                  Max.   :3.000
##  hasAndroidApp    hasIosApp         hasProfilePicture  daysSinceLastLogin
##  Length:98913     Length:98913     Length:98913     Min.   : 11.0
##  Class :character  Class :character  Class :character  1st Qu.:572.0
##  Mode  :character  Mode  :character  Mode  :character  Median :694.0
##                  Mean   :581.3
##                  3rd Qu.:702.0
##                  Max.   :709.0
##  seniority        seniorityAsMonths seniorityAsYears   countryCode
##  Min.   :2852     Min.   : 95.07   Min.   :7.92   Length:98913
##  1st Qu.:2857     1st Qu.: 95.23   1st Qu.:7.94   Class :character
##  Median :3196     Median :106.53   Median :8.88   Mode  :character
##  Mean   :3064     Mean   :102.13   Mean   :8.51
```

```
## 3rd Qu.:3201 3rd Qu.:106.70 3rd Qu.:8.89
## Max. :3205 Max. :106.83 Max. :8.90
```

Datos descriptivos dos usuarios:

- **Uso da App:**

Distribución de usuarios que usan algunha APP:

```
tabfreq <- table(datos$hasAnyApp)
tabfreq

##
## False True
## 72739 26174

barplot(tabfreq)
```

Obtemos o diagrama de barras *Figura 1*.

Analizamos a relación entre o uso da APP e produtos comprados e vendidos:

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
## filter, lag

## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union

datos_app_compraventa <- datos[,c("hasAnyApp", "productsSold", "productsBought")]
df <- datos_app_compraventa %>% group_by(hasAnyApp) %>%
  summarise(sum_sold = sum(productsSold),
            mean_sold = mean(productsSold),
            sum_bought = sum(productsBought),
            mean_bought = mean(productsBought)
            )

print(df)

## # A tibble: 2 × 5
##   hasAnyApp sum_sold mean_sold sum_bought mean_bought
##   <chr>      <int>      <dbl>      <int>      <dbl>
## 1 False      3233      0.0444      7957      0.109
## 2 True       8794      0.336       9049      0.346
```

Estes datos aparecen representados na *Táboa 2*.

- **Título civil:**

Distribución dos usuarios en función do seu título civil:

```
tabfreq <- table(datos$civilityTitle)
tabfreq

##
## miss mr mrs
## 437 22792 75684
```



```
barplot(tabfreq)
```

Obtemos o diagrama de barras *Figura 2*.

Analizamos a relación entre o título civil e produtos comprados e vendidos:

```
library(dplyr)
datos_civil_compraventa <- datos[,c("civilityTitle", "productsSold", "productsBought"
)]
df <- datos_civil_compraventa %>% group_by(civilityTitle) %>%
  summarise(sum_sold = sum(productsSold),
            mean_sold = mean(productsSold),
            sum_bought = sum(productsBought),
            mean_bought = mean(productsBought)
            )
print(df)

## # A tibble: 3 × 5
##   civilityTitle sum_sold mean_sold sum_bought mean_bought
##   <chr>          <int>     <dbl>     <int>     <dbl>
## 1 miss           574       1.31         287       0.657
## 2 mr            2673       0.117        4262       0.187
## 3 mrs           8780       0.116       12457       0.165
```

Estes datos aparecen representados na *Táboa 3*.

- **Idioma usado na web:**

Táboa de frecuencias dos idiomas de preferencia dos usuarios da web:

```
tabfreq <- table(datos$language)
tabfreq

##
##   de   en   es   fr   it
## 7178 51564 6033 26372 7766

barplot(tabfreq)
```

Obtemos o diagrama de barras *Figura 3*.

Analizamos a relación entre o idioma usado na web e produtos comprados e vendidos:

```
tab <- table(datos$language, datos$civilityTitle)
library(dplyr)
datos_idio_compraventa <- datos[,c("language", "productsSold", "productsBought")]
df <- datos_idio_compraventa %>% group_by(language) %>%
  summarise(sum_sold = sum(productsSold),
            mean_sold = mean(productsSold),
            sum_bought = sum(productsBought),
            mean_bought = mean(productsBought)
            )
print(df)

## # A tibble: 5 × 5
##   language sum_sold mean_sold sum_bought mean_bought
##   <chr>     <int>     <dbl>     <int>     <dbl>
## 1 de         539       0.0751     1705     0.238
## 2 en        4877       0.0946     8918     0.173
## 3 es         901       0.149       797     0.132
## 4 fr        3004       0.114     4386     0.166
## 5 it        2706       0.348     1200     0.155
```

Estes datos aparecen representados na *Táboa 4*.

- **Páís de procedencia:**

Táboa de frecuencias do país de procedencia dos usuarios da web:

```

tabfreq <- table(datos$country)
tabfreq

##
##           Afghanistan           Afrique du Sud
##                10                83
##           Albanie                Algérie
##                37                76
##           Allemagne                Andorre
##           6567                9
##           Angola                Anguilla
##                6                2
##           Antarctique                Antigua et Barbuda
##                1                3
##           Antilles néerlandaises                Arabie Saoudite
##                1                143
##           Argentine                Arménie
##                39                16
##           Aruba                Australie
##                3                2719
##           Autriche                Azerbaïdjan
##           624                22
##           Bahamas                Bahreïn
##                3                25
##           Bangladesh                Barbade
##                9                3
##           Belgique                Belize
##           1666                4
##           Benin                Bermude
##                6                7
##           Biélorussie                Bolivie
##                10                4
##           Bosnie et Herzegovine                Botswana
##                14                2
##           Brésil                Brunei Darussalam
##                95                12
##           Bulgarie                Burkina Faso
##                82                3
##           Cambodge                Cameroun
##                23                20
##           Canada                Chad
##           1577                1
##           Chili                Chine
##                49                428
##           Chypre                Colombie
##                55                50
##           Congo                Corée du Nord
##                21                1
##           Corée du Sud                Costa Rica
##                138                8
##           Cote D'Ivoire                Croatie
##                30                116
##           Cuba                Danemark
##                2                1892
##           Djibouti                Dominique
##                6                2
##           Egypte                Émirats arabes unis
##                28                131
##           Équateur                Érythrée
##                13                1
##           Espagne                Estonie

```

|    |                                |  |
|----|--------------------------------|--|
| ## | 5706                           | 33                                     |
| ## | Etats-Unis                     | Éthiopie                               |
| ## | 20602                          | 1                                      |
| ## | Fidji                          | Finlande                               |
| ## | 4                              | 661                                    |
| ## | France                         | Gabon                                  |
| ## | 25135                          | 11                                     |
| ## | Gambie                         | Géorgie                                |
| ## | 2                              | 23                                     |
| ## | Géorgie du Sud / Iles Sandwich | Ghana                                  |
| ## | 1                              | 4                                      |
| ## | Gibraltar                      | Grèce                                  |
| ## | 1                              | 229                                    |
| ## | Grenade                        | Guadeloupe                             |
| ## | 2                              | 38                                     |
| ## | Guam                           | Guatemala                              |
| ## | 2                              | 4                                      |
| ## | Guernsey                       | Guinée                                 |
| ## | 3                              | 5                                      |
| ## | Guyana                         | Guyane                                 |
| ## | 1                              | 14                                     |
| ## | Haiti                          | Honduras                               |
| ## | 2                              | 8                                      |
| ## | Hong Kong                      | Hongrie                                |
| ## | 854                            | 60                                     |
| ## | Île Bouvet                     | Île de Man                             |
| ## | 3                              | 1                                      |
| ## | Île Norfolk                    | Îles Canaries                          |
| ## | 1                              | 2                                      |
| ## | Iles Cayman                    | Îles Féroé                             |
| ## | 3                              | 2                                      |
| ## | Îles Malouines                 | Iles mineures éloignées des États-Unis |
| ## | 1                              | 1                                      |
| ## | Îles Turques-et-Caïques        | Îles Vierges (U.S.)                    |
| ## | 3                              | 3                                      |
| ## | Îles Vierges britanniques      | Inde                                   |
| ## | 3                              | 157                                    |
| ## | Indonésie                      | Irak                                   |
| ## | 75                             | 8                                      |
| ## | Iran                           | Irlande                                |
| ## | 12                             | 452                                    |
| ## | Islande                        | Israël                                 |
| ## | 25                             | 105                                    |
| ## | Italie                         | Jamaïque                               |
| ## | 8015                           | 2                                      |
| ## | Japon                          | Jersey                                 |
| ## | 165                            | 10                                     |
| ## | Jordanie                       | Kazakhstan                             |
| ## | 17                             | 34                                     |
| ## | Kenya                          | Kirghizistan                           |
| ## | 11                             | 4                                      |
| ## | Kiribati                       | Kowait                                 |
| ## | 3                              | 77                                     |
| ## | Laos                           | Le Salvador                            |
| ## | 3                              | 3                                      |
| ## | Lettonie                       | Liban                                  |
| ## | 48                             | 36                                     |
| ## | Libye                          | Liechtenstein                          |
| ## | 1                              | 3                                      |
| ## | Lituanie                       | Luxembourg                             |
| ## | 40                             | 111                                    |
| ## | Macau                          | Macédoine                              |
| ## | 16                             | 8                                      |
| ## | Madagascar                     | Malaisie                               |
| ## | 10                             | 119                                    |
| ## | Maldives                       | Mali                                   |
| ## | 3                              | 5                                      |
| ## | Malte                          | Maroc                                  |

|    |                                 |                           |
|----|---------------------------------|---------------------------|
| ## | 26                              | 167                       |
| ## | Martinique                      | Maurice                   |
| ## | 31                              | 5                         |
| ## | Mauritanie                      | Mayotte                   |
| ## | 2                               | 1                         |
| ## | Mexique                         | Moldavie                  |
| ## | 191                             | 20                        |
| ## | Monaco                          | Mongolie                  |
| ## | 36                              | 12                        |
| ## | Myanmar                         | Namibie                   |
| ## | 11                              | 2                         |
| ## | Nepal                           | Nicaragua                 |
| ## | 1                               | 3                         |
| ## | Niger                           | Nigeria                   |
| ## | 1                               | 32                        |
| ## | Norvège                         | Nouvelle Calédonie        |
| ## | 178                             | 13                        |
| ## | Nouvelle Zélande                | Oman                      |
| ## | 113                             | 9                         |
| ## | Ouganda                         | Pakistan                  |
| ## | 4                               | 17                        |
| ## | Palau                           | Panama                    |
| ## | 1                               | 12                        |
| ## | Paraguay                        | Pays-Bas                  |
| ## | 4                               | 1529                      |
| ## | Pérou                           | Philippines               |
| ## | 26                              | 98                        |
| ## | Pologne                         | Polynésie Française       |
| ## | 193                             | 8                         |
| ## | Portugal                        | Puerto Rico               |
| ## | 212                             | 19                        |
| ## | Qatar                           | République centrafricaine |
| ## | 47                              | 1                         |
| ## | République dominicaine          | République tchèque        |
| ## | 10                              | 56                        |
| ## | Réunion                         | Roumanie                  |
| ## | 51                              | 329                       |
| ## | Royaume-Uni                     | Russie                    |
| ## | 11310                           | 490                       |
| ## | Rwanda                          | Saint Kitts et Nevis      |
| ## | 1                               | 1                         |
| ## | Saint Vincent et les Grenadines | Sainte Lucie              |
| ## | 1                               | 1                         |
| ## | Samoa                           | Samoa américaines         |
| ## | 1                               | 9                         |
| ## | Sénégal                         | Serbie                    |
| ## | 15                              | 54                        |
| ## | Singapour                       | Slovaquie                 |
| ## | 260                             | 54                        |
| ## | Slovénie                        | Sri Lanka                 |
| ## | 28                              | 1                         |
| ## | St Barthélémy                   | Suède                     |
| ## | 1                               | 1826                      |
| ## | Suisse                          | Suriname                  |
| ## | 859                             | 1                         |
| ## | Svalbard et Jan Mayen Islands   | Swaziland                 |
| ## | 1                               | 1                         |
| ## | Syrie                           | Taiwan                    |
| ## | 2                               | 133                       |
| ## | Tadjikistan                     | Tanzanie                  |
| ## | 4                               | 5                         |
| ## | Thaïlande                       | Togo                      |
| ## | 78                              | 3                         |
| ## | Trinidad et Tobago              | Tunisie                   |
| ## | 5                               | 85                        |
| ## | Turquie                         | Ukraine                   |
| ## | 105                             | 110                       |
| ## | Uruguay                         | Uzbekistan                |

```
##           4           2
##           Vanuatu       Venezuela
##           1           14
##           Viet Nam      Zaire
##           64           1
##           Zambia       Zimbabwe
##           1           5

barplot(tabfreq)
```

Obtemos o diagrama de barras *Figura 4*.

Analizamos a relación entre o país de procedencia dos usuarios e produtos comprados e vendidos:

```
library(dplyr)
datos_pais_compraventa <- datos[,c("country", "productsSold", "productsBought")]
df <- datos_pais_compraventa %>% group_by(country) %>%
  summarise(sum_sold = sum(productsSold),
            mean_sold = mean(productsSold),
            sum_bought = sum(productsBought),
            mean_bought = mean(productsBought)
            )
print(df)

## # A tibble: 200 × 5
##   country          sum_sold mean_sold sum_bought mean_bought
##   <chr>              <int>     <dbl>     <int>     <dbl>
## 1 Afghanistan         0         0         0         0
## 2 Afrique du Sud      0         0         0         0
## 3 Albanie             0         0         0         0
## 4 Algérie             0         0         0         0
## 5 Allemagne          536     0.0816     1635     0.249
## 6 Andorre             0         0         0         0
## 7 Angola             0         0         0         0
## 8 Anguilla           0         0         0         0
## 9 Antarctique        0         0         0         0
## 10 Antigua et Barbuda 0         0         0         0
## # i 190 more rows
```

Estes datos aparecen representados na *Táboa 5*.

- **Foto de perfil:**

Distribución de usuarios que teñen foto de perfil:

```
tabfreq <- table(datos$hasProfilePicture)
tabfreq

##
## False  True
## 1895 97018

barplot(tabfreq)
```

Obtemos o diagrama de barras *Figura 5*.

Analizamos a relación entre o emprego da imaxe de perfil e produtos comprados e vendidos:

```
library(dplyr)
datos_app_compraventa <- datos[,c("hasProfilePicture", "productsSold", "productsBought")]
df <- datos_app_compraventa %>% group_by(hasProfilePicture) %>%
  summarise(sum_sold = sum(productsSold),
            mean_sold = mean(productsSold),
```

```

        sum_bought = sum(productsBought),
        mean_bought = mean(productsBought)
    )
print(df)

## # A tibble: 2 × 5
##   hasProfilePicture sum_sold mean_sold sum_bought mean_bought
##   <chr>             <int>    <dbl>    <int>    <dbl>
## 1 False             5409    2.85     3482    1.84
## 2 True              6618    0.0682  13524   0.139

```

Estes datos aparecen representados na *Táboa 6*.

- **Xénero:**

Táboa de frecuencias do xénero dos usuarios da web:

```

tabfreq <- table(datos$gender)
tabfreq

##
##      F      M
## 76121 22792

barplot(tabfreq)

```

Obtemos o diagrama de barras *Figura 6*.

Analizamos a relación entre o xénero e produtos comprados e vendidos:

```

library(dplyr)
datos_app_compraventa <- datos[,c("gender", "productsSold", "productsBought")]
df <- datos_app_compraventa %>% group_by(gender) %>%
  summarise(sum_sold = sum(productsSold),
            mean_sold = mean(productsSold),
            sum_bought = sum(productsBought),
            mean_bought = mean(productsBought)
  )
print(df)

## # A tibble: 2 × 5
##   gender sum_sold mean_sold sum_bought mean_bought
##   <chr>    <int>    <dbl>    <int>    <dbl>
## 1 F        9354    0.123    12744    0.167
## 2 M        2673    0.117     4262    0.187

```

Estes datos aparecen representados na *Táboa 7*.

Correlación entre variables:

- **Todos os usuarios:**

Analizamos a correlación entre variables tendo en conta os datos de todos os usuarios, independentemente de si realizaron algunha transacción ou non:

```

datos_contvars <- datos[,c(5:12, 20:21)]
cormat <- cor(datos_contvars)

library(corrplot)

## corrplot 0.92 loaded

```

```
corrplot(cormat,  
         is.corr = TRUE,  
         col.lim=c(min(cormat),max(cormat)))
```

Matriz de correlación visible na *Figura 7*.

- **Usuarios que realizan polo menos unha compra:**

Realizamos unha limpeza de datos, eliminando aqueles usuarios que non chegaron a realizar ningunha compra:

```
datos_compradores = datos[datos$productsBought>0,]
```

Analizamos a correlación entre variables tendo en conta só os datos dos usuarios que realizaron como mínimo unha compra:

```
datos_compradores <- datos_compradores[,c(5:12,20:21)]  
cormat <- cor(datos_compradores)  
  
library(corrplot)  
corrplot(cormat,  
         is.corr = TRUE,  
         col.lim=c(min(cormat),max(cormat)))
```

Matriz de correlación visible na *Figura 8*.

- **Usuarios que realizan polo menos unha venta:**

Realizamos unha limpeza de datos, eliminando aqueles usuarios que non chegaron a realizar ningunha venta:

```
datos_vendedores = datos[datos$productsSold>0,]
```

Analizamos a correlación entre variables tendo en conta só os datos dos usuarios que realizaron como mínimo unha venta:

```
datos_vendedores <- datos_vendedores[,c(5:12,20:21)]  
cormat <- cor(datos_vendedores)  
  
library(corrplot)  
corrplot(cormat,  
         is.corr = TRUE,  
         col.lim=c(min(cormat),max(cormat)))
```

Matriz de correlación visible na *Figura 9*.

A partir de esta fase, as análises levanse a cabo empregando R Commander. Para iniciar o programa insertamos o seguinte na consola de R:

```
library(Rcmdr)
```

Unha vez aberto cargamos os mesmos datos empregados en R Studio.

```
C:/Users/Valeria/Desktop/TFG/datos/6M-0K-99K.users.dataset.public.csv
```

---

Compradores e vendedores segundo o seu título civil: diagrama de caixa.

- **Usuarios que realizaron polo menos unha compra.**

Eliminamos os datos de aqueles usuarios que non chegaron a realizar ningunha compra:

```
datos_compra = datos[datos$productsBought>0,]
```

Realizamos un diagrama de caixa segmentando mediante o título civil os usuarios que realizaron unha ou máis compras:

```
Boxplot(productsBought~civilityTitle, data=datos_compra, id=list(method="y"))
```

Obtemos a *Figura 10*.

Volvemos a realizar o diagrama, pero esta vez seleccionando un intervalo de entre 1 e 20 compras:

```
Boxplot(productsBought~civilityTitle, data=datos_compra, id=list(method="y"), ylim=c(1,20))
```

Obtemos a *Figura 11*.

- **Usuarios que realizaron polo menos unha venta.**

Eliminamos os datos de aqueles usuarios que non chegaron a realizar ningunha venta:

```
datos_venta = datos[datos$productsSold>0,]
```

Realizamos un diagrama de caixa segmentando mediante o título civil os usuarios que realizaron unha ou máis ventas:

```
Boxplot(productsSold~civilityTitle, data=datos_venta, id=list(method="y"))
```

Obtemos a *Figura 12*.

Volvemos a realizar o diagrama, pero esta vez seleccionando un intervalo de entre 1 e 20 compras:

```
Boxplot(productsSold~civilityTitle, data=datos_venta, id=list(method="y"), ylim=c(1,20))
```

Obtemos a *Figura 13*.

Porcentaxes de usuarios activos e usuarios que só actúan como visitantes agrupados segundo o seu xénero e título civil.

- **Usuarios que realizan polo menos unha compra.**

```
datos$Actividade_Compradores = datos$productsBought>0
```



---

Realizamos un diagrama de barras segmentando mediante o xénero os usuarios que realizaron unha ou máis compras:

```
with(datos, Barplot(gender, by=ActividadeCompradores, style="parallel", legend.pos="above", xlab="gender", ylab="Percent", scale="percent", label.bars=TRUE))
```

Obtemos a *Figura 14*.

Realizamos un diagrama de barras segmentando mediante o título civil os usuarios que realizaron unha ou máis compras:

```
with(datos, Barplot(civilityTitle, by=ActividadeCompradores, style="parallel", legend.pos="above", xlab="civilityTitle", ylab="Percent", scale="percent", label.bars=TRUE))
```

Obtemos a *Figura 15*.

- **Usuarios que realizan polo menos unha venta.**

```
datos$ActividadeVendedores = datos$productsSold>0
```

Realizamos un diagrama de barras segmentando mediante o xénero os usuarios que realizaron unha ou máis ventas:

```
with(datos, Barplot(gender, by=ActividadeVendedores, style="parallel", legend.pos="above", xlab="gender", ylab="Percent", scale="percent", label.bars=TRUE))
```

Obtemos a *Figura 16*.

Realizamos un diagrama de barras segmentando mediante o título civil os usuarios que realizaron unha ou máis compras:

```
with(datos, Barplot(civilityTitle, by=ActividadeVendedores, style="parallel", legend.pos="above", xlab="civilityTitle", ylab="Percent", scale="percent", label.bars=TRUE))
```

Obtemos a *Figura 17*.

- **Usuarios que realizan polo menos unha compra e unha venta.**

```
datos$ActividadeDoble = (datos$productsSold>0 & datos$productsBought>0)
```

Realizamos un diagrama de barras segmentando mediante o xénero os usuarios que realizaron unha ou máis compras e ventas:

```
with(datos, Barplot(gender, by=ActividadeDoble, style="parallel", legend.pos="above", xlab="gender", ylab="Percent", scale="percent", label.bars=TRUE))
```

Obtemos a *Figura 18*.

---

Realizamos un diagrama de barras segmentando mediante o título civil os usuarios que realizaron unha ou máis compras e ventas:

```
with(datos, Barplot(civilityTitle, by=Actividade_Doble, style="parallel", legend.p  
s="above", xlab="civilityTitle", ylab="Percent", scale="percent", label.bars=TRUE))
```

Obtemos a *Figura 19*.