
**Pronóstico de las ventas en Colombia para la
marca de ropa estadounidense UA, basado en el
desempeño de las piezas publicitarias en redes
sociales, entre enero de 2016 y octubre de 2019,
mediante un modelo SARIMAX de series de
tiempo**

Presentado por
**John Andersson
Perez Poveda**



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Fundación Universitaria Los Libertadores

Facultad de Ingeniería y Ciencias Básicas

Especialización en Estadística Aplicada

Bogotá D.C, Colombia

2018

**Pronóstico de las ventas en Colombia para la
marca de ropa estadounidense UA, basado en
el desempeño de las piezas publicitarias en
redes sociales, entre enero de 2016 y octubre
de 2019, mediante un modelo SARIMAX de
series de tiempo**

Presentado por

John Andersson Perez Poveda

en cumplimiento parcial de los requerimientos para optar al título

De

Especialista en Estadística Aplicada

Dirigida por

MSc. Sébastien

Lozano Forero

Profesor

Fundación Universitaria Los Libertadores

Facultad de Ingeniería y Ciencias Básicas

Especialización en Estadística Aplicada

Bogotá D.C, Colombia

2018

Notas de aceptación



LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Firma del presidente del jurado

Firma del jurado

Firma del jurado

Bogotá DC, Diciembre de 2019.



LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

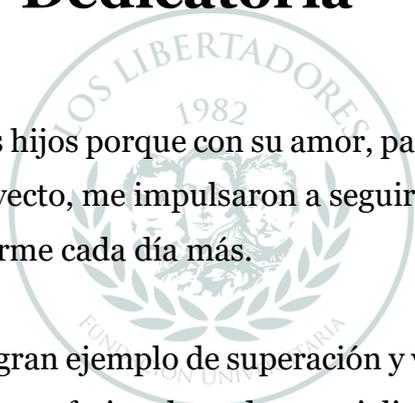
Las directivas de la Fundación Universitaria Los Libertadores, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden únicamente a los autores y a los resultados de su trabajo.

Dedicatoria

Dedicado a mi esposa y a mis hijos porque con su amor, paciencia y apoyo, puedo ver ahora el cumplimiento de este proyecto, me impulsaron a seguir adelante y fueron la motivación para no desfallecer y esforzarme cada día más.

A mi madre y abuela, por su gran ejemplo de superación y valioso aliento en todo momento desde que inicie mis estudios profesionales y de especialización.

A ellos que merecen todos los orgullos, emociones y alegrías que yo les pueda dar; este proyecto es de ellos.



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Agradecimientos

Deseo agradecer en primer lugar a Dios por la sabiduría y entendimiento que me suministró para poder avanzar en el transcurso de este gran proyecto. A mi familia por toda su paciencia y soporte. Al profesor Sébastien Lozano, por su confianza y porque logró de forma práctica, suministrarnos parte de sus conocimientos de forma clara, entendible y aplicable para el desarrollo de este trabajo de grado. A la Fundación Universitaria los Libertadores y al cuerpo docente de la especialización por la excelente formación recibida y a la calidad humana que transmiten.

LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Índice general

Contenido

Introducción.....	1
Planteamiento del Colombia.....	3
Objetivos.....	3
Objetivo General.....	3
Objetivos Específicos.....	3
Justificación.....	4
Soporte Computacional.....	4
Marco Teórico.....	6
Inversión en Publicidad Digital en Colombia.....	6
Relación entre Publicidad y Ventas.....	7
Primeros estudios de la relación y auge.....	7
Curva de Respuesta.....	8
Series de Tiempo.....	9
Metodología Box-Jenkins.....	9
Criterios de Información.....	9
Evaluación de Pronósticos.....	10
Modelo ARIMA y ARIMAX.....	10
Modelo SARIMA y SARIMAX.....	11
Marco Metodológico.....	13
Método.....	13
Enfoque.....	13
Diseño Metodológico.....	14
Metodología Minería de Datos.....	14
Metodología de Análisis Digital Conversacional.....	15
Metodología Estadística para Series de Tiempo.....	15
Acerca de los Datos.....	16
Área de Estudio.....	16
Recolección de Datos.....	16
Población y Muestra.....	17
Análisis y Resultados.....	18
Caracterización de los intereses de los clientes de la marca y su comportamiento de compra..	18
Construcción de la Medición de Desempeño de las Piezas Publicitarias en Redes Sociales.....	21
Modelo SARIMAX.....	23
Conclusiones y Recomendaciones.....	27

Índice de Figuras

Figura 1. Variación de la conversión anual en publicidad digital en Colombia.	6
Figura 2. Curva de respuesta	9
Figura 3. Metodología ASUM-DM.....	15
Figura 4. Metodología Talkers.....	16
Figura 5. Metodología Box-Jenkins.....	17
Figura 6. Nube de palabras de los tópicos de conversación.....	19
Figura 7. Distribución por secciones.....	19
Figura 8. Distribución por tipo de producto.....	20
Figura 9. Distribución por categoría.....	20
Figura 10. Interacción en el tiempo entre el índice de ventas y el indicador de desempeño.....	23
Figura 11. Serie de ventas y autocorrelogramas simple y parcial.....	23
Figura 12. Test Dickey-Fuller.....	24
Figura 13. Serie de ventas y autocorrelogramas simple y parcial diferenciando dos veces.....	24
Figura 14. Prueba de significancia de los parámetros del modelo.....	25
Figura 15. Validación del modelo SARIMAX planteado.....	25
Figura 16. Test Ljung-Box y Jarque-Bera para los residuos	26
Figura 17. Pronostico In-Sample del modelo SARIMAX planteado.....	26

Pronóstico de las ventas en Colombia para la marca de ropa estadounidense UA, basado en el desempeño de las piezas publicitarias en redes sociales, entre enero de 2016 y octubre de 2019, mediante un modelo SARIMAX de series de tiempo

Resumen

Con el auge del internet, las redes sociales se han convertido en una plataforma atractiva para que las compañías de diferentes sectores de la economía, lancen sus campañas publicitarias, ya no solamente con fines comerciales, sino para promover experiencias, fidelizar a los clientes y establecer una identidad con las marcas. Por tal motivo, se hace necesario evaluar y medir con algún grado de certeza, el desempeño de las creatividades expuestas al público, con un indicador que condense las interacciones de los usuarios con las publicaciones creadas para las redes sociales. En el presente trabajo, se aborda el problema de construir dicho indicador y a partir de este, predecir el nivel de ventas para la compañía estadounidense de ropa UA para la sede en Colombia. Para tal fin, adicionalmente se realiza una caracterización de los comportamientos de compra de los clientes actuales y finalmente se predicen las ventas obtenidas como resultado de las piezas publicitarias lanzadas por la compañía en redes sociales, usando modelos auto regresivos integrados de medias móviles estacionales con variables exógenas (SARIMAX).

Capítulo 1

Introducción

Con la llegada de la nueva era digital, el auge del internet y la idea de sociedades conectadas, se ha detonado una revolución en las economías mundiales por los cambios de comportamiento de consumo, cultura, comunicación e intereses en todas las generaciones.

Según los datos suministrados por la agencia creativa española *We Are Social* en su reporte mensual “The Global State Of Digital In October 2019”¹ donde calcula las métricas globales del uso del Internet y de la utilización de redes sociales, revela que para octubre del año 2019, el número total de audiencia activa en Facebook fue de 1.931 millones de cuentas en el mundo y para el caso de Twitter fue de 260,3 millones de cuentas activas, también a nivel mundial. Lo anterior indica que estas plataformas digitales, deben ser objeto de estudio, no solo por las repercusiones en la transformación de las conductas sociales, sino por el impacto económico que generan en cada uno de los países, en su contribución al PIB (Producto Interno Bruto), al empleo, en la necesidad de generar políticas de estado para impulsar el desarrollo de las TIC’s (Tecnologías de la información y la comunicación) y en el comportamiento de la oferta y la demanda, en particular, para las estrategias comerciales y de marketing que deben generar las compañías para sus clientes.

En Colombia y basado en el “Reporte de inversión en publicidad digital”² publicado por *IAB Colombia*, en el año 2018, se invirtió en publicidad digital aproximadamente 849 mil millones de pesos, de los cuales el 27,5% fue dirigido a redes sociales, lo cual corrobora, que la publicidad en estos canales se ha convertido en uno de los recursos más efectivos para conseguir resultados para un negocio, no solo por el alcance masivo, sino porque permite conocer a sus consumidores con mayor profundidad y crear vínculos o experiencias con ellos.

En ese contexto, las organizaciones deben buscar un elemento diferenciador para mantenerse competitivos en un mercado cada vez más cambiante. Tomando esta idea como supuesto, en este trabajo nos enfocaremos en determinar un pronóstico que revele cómo han impactado las estrategias publicitarias en los activos digitales, de la marca estadounidense UA sede en Colombia, en los comportamientos de consumo de sus clientes (ventas). Lo anterior puede ser de interés, para las decisiones futuras que pueda tomar la compañía en sus acciones comunicacionales en el ecosistema digital, además, de ayudar con el entendimiento y cuantificación de las tasas de conversión por publicidad en social media.

Este trabajo se divide en 6 capítulos siendo la presente introducción el primero. En el segundo capítulo se aborda el planteamiento del problema, el cual es desagregado mediante los objetivos y la justificación. En el tercer capítulo se expone de forma breve el marco teórico, abordando temas

¹ (Hootsuite, 2019)

² (IAB Colombia, 2019)

acordes a la publicidad digital en Colombia, la relación entre publicidad y ventas, y la fundamentación de series de tiempo para enfrentar el problema expuesto. La cuarta parte contiene el marco metodológico, aplicado a los datos recopilados de la conversación social de la marca de ropa estadounidense UA con sede en Colombia en canales digitales y los datos de ventas de la compañía, para el periodo comprendido entre enero de 2016 y octubre de 2019. Los resultados de la evaluación de impacto y el modelo de pronóstico son presentados en la quinta sección.

Capítulo 2

Planteamiento del Problema

Recientemente una marca de ropa estadounidense, se encuentra en procesos de expansión, con sedes en varios países, entre ellos Colombia. Con el desafío de incursionar en el mercado, llegar al punto de equilibrio en ventas y ser competitivos frente a otras marcas del mismo sector, ha estado realizando esfuerzos en sus estrategias de mercadeo, particularmente en el ecosistema digital por el gran potencial que ofrece, a través de diferentes canales como las redes sociales, pauta en Google y por correo electrónico. Sin embargo, se presentan algunas situaciones que han limitado sus acciones comunicacionales:

- La empresa no ha creado un indicador que le permita medir el desempeño de cada pieza publicitaria que crea para sus activos digitales en redes sociales.
- La marca no comprende cuanta tracción y fidelización generan cada una de sus creatividades publicadas en Facebook y Twitter.

La empresa necesita establecer si la estrategia comunicativa actual en digital es sostenible y causa un efecto importante en el comportamiento de compra de los consumidores o si por el contrario, el impacto en los clientes, representado en ventas, es mínimo. De ser así, la empresa debería rediseñar su estrategia creativa y de inversión en pauta, si quiere posicionar a sus redes sociales como un canal de servicio, mercadeo, de experiencia y cercanía con los usuarios, que conlleve a generar tracción o futuras compras y fidelidad con la marca.

2.1. Objetivos

Una vez indicados los elementos fundamentales de este trabajo, indicador de desempeño de piezas publicitarias y datos de ventas de la compañía, se plantean los siguientes objetivos fundamentales:

2.1.1. Objetivo General

Consolidar un modelo de pronóstico que permita predecir el nivel de ventas de la marca UA de la sede Colombia a través de los años, utilizando como variable exógena una medición creada para evaluar el desempeño de las piezas publicitarias en redes sociales.

2.1.2. Objetivos Específicos

- Realizar una identificación y categorización de los intereses de los clientes de la marca, basados en sus tópicos de conversación por redes sociales.
- Consolidar la información disponible en los reportes semanales dados por la empresa (área de data y reporting) en un indicador mensual de desempeño de piezas publicitarias.
- Validar un modelo de pronóstico SARIMAX para el comportamiento de ventas de la compañía, utilizando el indicador de desempeño publicitario como variable regresora, aplicando la metodología Box-Jenkins al modelo.

2.2. Justificación

El desarrollo de este trabajo, será de gran apoyo para el área de marketing de la empresa, porque se convertirá en una guía para establecer el direccionamiento de la inversión en pauta en las diferentes campañas publicitarias dentro de los activos digitales que posee.

Crear un modelo de pronóstico, ayudará a comprender la correlación existente entre el desempeño de las acciones comunicacionales y su repercusión en los comportamientos de compra de los consumidores.

Por otra parte, este modelo le indicara a la compañía, la frecuencia de compra e intereses por producto de los clientes, para las distintas épocas del año, basados en los datos transaccionales históricos y el establecimiento de un indicador de desempeño de las piezas publicitarias creadas.

Adicionalmente, resultaría de gran interés poder desarrollar predicciones razonablemente acertadas para la cantidad de ventas generadas por la compañía, en la sede Colombia, para que se puedan tomar acciones en sus planes comunicacionales de mercadeo y servicio. Por ejemplo, si los clientes son más activos en temporadas como diciembre, el presupuesto en elaboración de creatividades para el impulso comercial, debería ser mayor teniendo unos KPI's establecidos.

En resumen, este estudio será una guía para que la marca pueda medir el desempeño de sus campañas en el ecosistema digital desde las métricas en redes sociales y su relación con las ventas, ayudará a generar conocimiento acerca de las microaudiencias de la empresa y forjará ventajas competitivas frente a otras compañías similares de la industria y del comercio textil al alinear los planes de mercadeo al comportamiento de sus usuarios.

Finalmente, en la literatura son limitados los estudios publicados que han incorporado el desarrollo de series de tiempo con la data obtenida en el ecosistema digital, este trabajo

también aportará en el ámbito empresarial y académico como documento guía para el modelamiento de los datos incorporando estas dos áreas del conocimiento.

Capítulo 3

Marco Teórico

Para el desarrollo de este capítulo, es necesario evaluar varios frentes, en primer lugar, la importancia del marketing digital en Colombia, en segundo lugar, las bases teóricas de la relación de las ventas con el lanzamiento de piezas publicitarias; por último la fundamentación del desarrollo de series de tiempo con las que será medido el impacto de la comunicación digital en las conductas de compra de sus usuarios.

3.1. Inversión en Publicidad Digital en Colombia

Esta pequeña sección, revela el incremento anual de la inversión en publicidad digital por diferentes empresas en Colombia, lo que demuestra la importancia de entender para una marca, la efectividad de conversión que están causando sus esfuerzos publicitarios en los diferentes activos digitales.

Según IAB Colombia (Interactive Advertising Bureau) en su reporte de inversión en publicidad digital total 2018³, de una muestra de 150 empresas afiliadas, el crecimiento anual en inversión, desde el 2015 ha sido el siguiente:

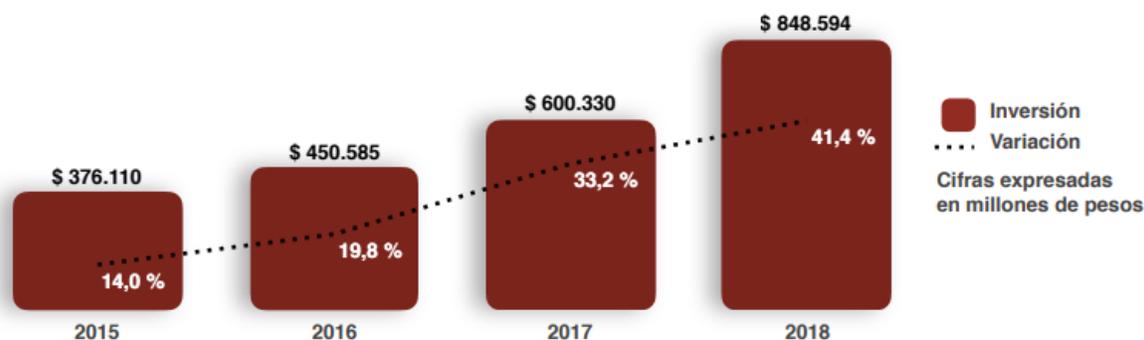


Figura 1. Inversión anual en publicidad digital en Colombia. Información tomada de <http://www.iabcolombia.com/wp-content/uploads/resumen-ejecutivo-inversion-en-publicidad-digital-colombiaaAno-2018.pdf>

³ (IAB Colombia, 2019)

Según Londoño y Mora⁴, muchos medios de comunicación identifican que Internet desempeña un papel fundamental, puesto que por medio de las redes sociales atraen a los consumidores reales o potenciales, dado que allí pueden capturar en el momento preciso mensajes comerciales o de publicidad.

Hay que tener en cuenta que, a la hora de obtener resultados con el marketing digital, se debe resaltar la construcción de marca, las bases de clientes, los resultados de negocio y la generación de valor. El marketing digital desempeña un papel fundamental para las operaciones y genera un valor desde una perspectiva más amplia a la hora de negociar y poner a disposición productos como las prendas de vestir (Villaseca, 2014)

3.2. Relación entre Publicidad y Ventas

La relación de la publicidad con las ventas es uno de los temas más ampliamente estudiados desde hace años. Esto es así, por la gran importancia que tiene para el anunciante, ya que las inversiones en publicidad suelen ser elevadas (como lo reveló la sección anterior) y es difícil poder saber con exactitud qué porcentaje de ventas está asociado directamente a la realización de una campaña.

3.2.1. Primeros Estudios de la Relación y Auge

Son innumerables las investigaciones realizadas sobre la relación entre el desempeño de la publicidad y las ventas, aunque es importante indicar que gran parte de ellos se ubican en la década de los setentas y ochentas, etapa reina de la modelización en el campo del marketing; sumergirse en estos estudios es importante, ya que han servido como base para los análisis realizados en épocas posteriores y a su vez dan sustento a la justificación del problema planteado en este trabajo.

Una de los primeros estudios en este sentido, es el elaborado por Benesch⁵, que investigó cómo las innovaciones en la publicidad (ilustraciones, medios utilizados, posición del anuncio, utilización del color, además de otros elementos), podían repercutir en las ventas.

Años después, el uso de modelos empezó a ser común para estudiar estas relaciones. Así, Bass y Clarke utilizaron diferentes modelos econométricos para analizar la tardanza que el impacto de la comunicación tenía en las ventas de un producto sin azúcar para el control de aumento de peso (Bass & Clarke, 1972).

En los ochentas, se ejecutaron análisis comparativos, con el objetivo de entender las distintas contribuciones presupuestarias de las áreas de ventas y de publicidad, a la organización de la comunicación en 500 compañías industriales y de productos de gran consumo (Dubinsky, Barry & Kerin, 1981). Además, también se indagó la publicidad comparativa, que hasta entonces había sido estudiada desde el punto de vista de las actitudes, creencias, intenciones,

⁴ (Londoño & Mora, 2018)

⁵ (Benesch, 1952)

entre otros, que generaba en el consumidor, y que entonces se analizó para profundizar en los efectos que dicha publicidad comparativa tenía en las ventas (Demirdjian, 1983).

Hacia la mitad de la década de los ochentas, la gran diversidad de métodos existentes para estudiar la dependencia de la comunicación con las ventas, hizo que hubiesen análisis detallados para ahondar en la idoneidad de cada uno de esos modelos, como ocurrió con un caso realizado por Assmus, Farley y Lehmann⁶, que tomaron 128 modelos econométricos que verificaban el impacto de la publicidad en las ventas, tanto a corto como a largo plazo.

Adicionalmente se plasmó, un artículo basado en los efectos que la publicidad tiene en el ciclo de vida del producto; este estudio específicamente se efectuó con marcas individuales, marcas que competían dentro de una misma categoría, así como marcas de diferentes categorías, y en todas ellas se percibió un efecto similar de la publicidad en el ciclo de vida del producto, siendo éste decreciente y gradual (Holak & Tang, 1990).

Se podrían seguir contando más trabajos perpetuados en esta temática, sin embargo el objetivo era nombrar algunos de la época dorada del boom del marketing, reitero, para dar sustento al problema que varios personajes a través de los años han querido solucionar por diferentes medios y apoyados en distintas áreas del conocimiento y herramientas, para resolver la relación existente entre la publicidad y las ventas. Este trabajo busca entender intrínsecamente dicha relación para una empresa puntual, pero en un ambiente diferente, que cambia varias de las reglas de juego existentes y es en el ecosistema digital.

3.2.2. Curva de Respuesta

La curva de respuesta surgió para la década de los setentas, intentando crear una relación entre las ventas a lo largo del tiempo y el momento en que es lanzada una campaña publicitaria. Para ello, es necesario detallar las condiciones que apoyaron a la construcción de esta curva:

- Las ventas crecen si aumenta la publicidad (frecuencia) y decrecen si disminuye la publicidad, pero a distintas velocidades.
- La curva de respuesta a la publicidad debe ser cóncava o con forma “S”.
- Las ventas son positivas con niveles de publicidad nula.
- La publicidad de la competencia afecta las ventas negativamente (frecuentemente).
- La efectividad de la inversión publicitaria cambia en el tiempo (cambios en copia, medios, etc.).
- Si la publicidad se mantiene constante, el impacto es menos positivo.

Observemos de manera gráfica lo anteriormente planteado:

⁶ (Assmus, Farley, & Lehmann, 1984)

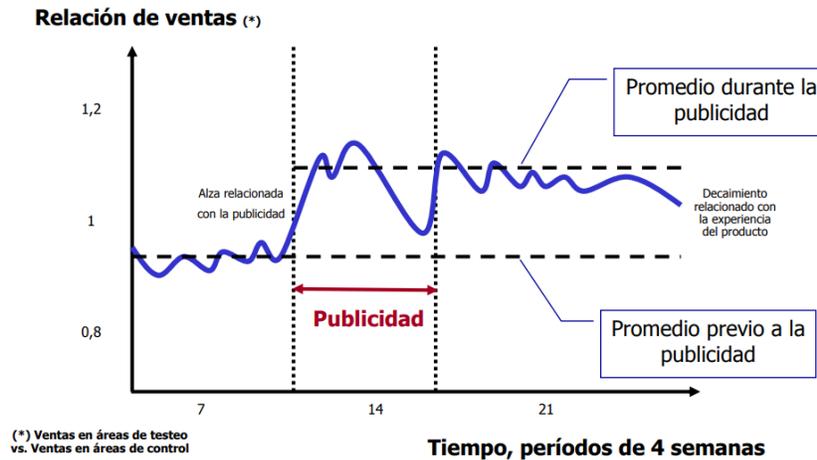


Figura 2. Curva de respuesta. Fuente: (Fritis, Polit, Ramirez, & Young, 2018).

3.3. Series de Tiempo

En este apartado se presentan de forma global los modelos estadísticos tradicionales de series de tiempo.

3.3.1. Metodología Box-Jenkins

El modelamiento que se introducirá en esta sección, consiste en ajustar modelos a la estructura general SARIMA $(p,d,q) \times (P,D,Q)_s$ y alguna de sus variaciones. La estrategia para la construcción del modelo se basa en un ciclo iterativo y utiliza los propios datos para la elección de la estructura de este modelo (Lozano, 2018).

Las fases de esta metodología son cuatro: Identificación, evaluación, diagnóstico y pronóstico. La descripción y ciclos de estas fases serán ampliadas en la sección 4.3.3., de este trabajo.

3.3.2. Criterios de Información

Una de las aproximaciones para la selección de un modelo es la comparación simultánea de todos los modelos a través del criterio de información de Akaike (AIC) o al criterio de información bayesiana (BIC). En los dos casos se escoge el modelo que presenta el valor más pequeño entre todos los modelos evaluados, siendo este el que presenta el mejor ajuste a los datos.

El criterio de información Bayesiana (BIC) propuesto por Schwarz en 1978⁷, es una medida alternativa para el AIC para obtener información sobre el mejor modelo que se puede utilizar. El BIC se define como

⁷ (Schwarz, 1978)

$$\text{BIC} = -2\log(L) + k \log(T),$$

Donde k es el número de parámetros que varían ($k = p+q$ en los modelos ARMA) y T es la cantidad de datos evaluados en el periodo de estudio.

Como el BIC a diferencia del criterio AIC es consistente⁸, usaremos en este trabajo esta medida.

3.3.3. Evaluación de Pronósticos

Según Lozano (2018)⁹, para evaluar los pronósticos generados por los modelos propuestos se debe proceder de la siguiente manera. Al considerar una serie de tiempo con T observaciones, serán retiradas de ésta serie, las últimas n observaciones. Se genera el modelo con las $T - n$ observaciones y al generar los pronósticos, se evaluará la diferencia de los valores pronosticados y los valores verdaderos contenidos en la serie (los n puntos que fueron retirados). Considere

$$e_j = y_j - \hat{y}_j,$$

Donde \hat{y}_j denota el pronóstico y_j .

La capacidad predictiva del modelo estimado puede ser evaluada utilizando la medida de error medio (entre otras), donde $k = T - n$ o la autocorrelación de los errores con rezago 1 (ACF1).

Error Medio (ME):

$$ME = \frac{1}{n-1} \sum_{j=k+1}^T e_j$$

3.3.4. Modelo ARIMA y ARIMAX

Guevara y Guevara (2019)¹⁰, mencionan que para los procesos autorregresivos integrados de medias móviles ARIMA (p, d, q), en las series temporales no estacionarias se incumplen los supuestos de media y varianza constante en el tiempo. En concordancia con lo anterior, para poder realizar un análisis de la variable de interés se requiere diferenciarla d veces. Matemáticamente se encuentran definidos como:

$$\phi(B)(1-B)^d Y_t = \theta(B)\epsilon_t$$

Siendo d el orden de diferenciación.

Por otra parte, para los modelos autorregresivos integrados de medias móviles con variables

⁸ (Lozano, 2018)

⁹ (Lozano, 2018)

¹⁰ (Guevara & Guevara, 2019)

exógenas ARIMAX (p, d, q), las variables exógenas permiten realizar un mejor estudio en las series de tiempo. Se encuentran descritos por:

$$\phi(B)(1 - B)^d Y_t = X_t + \theta(B)\varepsilon_t$$

Donde X_t hace referencia a la variable exógena.

3.3.5. Modelo SARIMA y SARIMAX

Basado en el libro de Tsay (2005)¹¹, el modelo SARIMA es una extensión del modelo ARIMA, que se utiliza frecuentemente cuando sospechamos que un modelo puede tener un efecto estacional.

Por definición, el proceso autorregresivo estacional integrado de media móvil - SARIMA (p,d,q)(P,D,Q)_s - es un proceso multiplicativo de dos procesos ARMA de la serie de tiempo diferenciada, definido como:

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - \sum_{j=1}^P \Phi_j L^{j \times s})(1 - L)^d(1 - L^s)^D x_t = (1 + \sum_{i=1}^q \theta_i L^i)(1 + \sum_{j=1}^Q \Theta_j L^{j \times s}) a_t$$

$$y_t = (1 - L)^d(1 - L^s)^D$$

Donde:

- X_t es la salida original no estacional en el tiempo t.
- Y_t es la salida diferenciada (estacional) en el tiempo t.
- d es el orden de integración no estacional de las series temporales.
- p es el orden del componente AR no estacional.
- P es el orden del componente AR estacional.
- q es el orden del componente MA no estacional
- Q es el orden del componente MA estacional.
- s es la duración de la estacionalidad
- D es el orden de integración de las series de tiempo estacionales.
- a_t es la innovación, choque o término de error en el tiempo t.

Asumiendo Y_t con media estacional a largo plazo de μ , entonces teniendo la expectativa de ambos lados, podemos expresarlo ϕ_0 de la siguiente manera:

$$\phi_0 = (1 - \phi_1 - \phi_2 - \dots - \phi_p)(1 - \Phi_1 - \Phi_2 - \dots - \Phi_P)$$

Entonces, el proceso SARIMA (p,d,q)(P,D,Q)_s pueden ahora ser expresado como:

$$z_t = y_t - \mu$$

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - \sum_{j=1}^P \Phi_j L^{j \times s}) z_t = (1 + \sum_{i=1}^q \theta_i L^i)(1 + \sum_{j=1}^Q \Theta_j L^{j \times s}) a_t$$

Por otra parte, el modelo estacional autorregresivo integrado de media móvil exógena

¹¹ (Tsay, 2005)

SARIMAX es un modelo de regresión lineal que utiliza un proceso de tipo SARIMA, es decir este modelo es útil en los casos donde sospechamos que los residuos pueden exhibir una tendencia estacional o patrón (Beckwith, 1972).

$$w_t = y_t - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \dots - \beta_b x_{b,t}$$

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - \sum_{j=1}^P \Phi_j L^{j \times s})(1 - L)^d (1 - L^s)^D w_t - \eta = (1 + \sum_{i=1}^q \theta_i L^i)(1 + \sum_{j=1}^Q \Theta_j L^{j \times s}) a_t$$

$$a_t \sim \text{i.i.d} \sim \Phi(0, \sigma^2)$$

Donde:

- L es el operador de rezago (mejor conocido como back-shift).
- Yt es la salida observada en el tiempo t.
- Xk,t es la variable de entrada exógena k-ésimo en el tiempo t.
- βk es el valor del coeficiente de la variable de entrada k-ésima exógena (explicativa).
- b es el número de variables de entradas exógenas.
- wt son los residuos de la regresión de auto-correlación.
- p es el orden del componente AR no estacional.
- P es el orden del componente RA estacional.
- q es el orden del componente MA no estacional.
- Q es el orden del componente estacional MA.
- s es la duración de la estacionalidad.
- D es el orden de integración de estacionalidad de la series de tiempo.
- η es una constante en el modelo SARIMA
- at es la innovación, shock o término de error en el tiempo t.

Reordenar los términos de la ecuación anterior y asume los resultados diferenciados (tanto estacionales y no estacionales) en una serie de tiempo estacionaria (zt) se obtiene lo siguiente:

$$z_t = (1 - L)^d (1 - L^s)^D w_t$$

$$\mu = E[z_t] = \frac{\eta}{(1 - \phi_1 - \phi_2 - \dots - \phi_p)(1 - \Phi_1 - \Phi_2 - \dots - \Phi_P)}$$

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - \sum_{j=1}^P \Phi_j L^{j \times s})(1 - L)^d (1 - L^s)^D (w_t - \mu) = (1 + \sum_{i=1}^q \theta_i L^i)(1 + \sum_{j=1}^Q \Theta_j L^{j \times s}) a_t$$

Capítulo 4

Marco Metodológico

4.1. Método

Investigación Descriptiva

La investigación descriptiva, se refiere a la parte previa al análisis de datos con el fin de buscar factores, características y otros rasgos importantes del tema analizar. También comprende el registro, análisis e interpretación de la naturaleza actual de los datos y trabaja sobre hechos que están ocurriendo (Hernandez Sampieri, 2006).

El presente proyecto cabe dentro de este tipo de investigación , por una lado, porque describe y evalúa las características del comportamiento de compra de los usuarios que han recibido publicidad por medios digitales de la marca UA en Colombia, analizando los datos reunidos de cada una de las variables a investigar, y por otra parte, porque puede establecer una posible conducta de las ventas obtenidas a futuro, por dicho comportamiento de consumo a partir de la incidencia de las acciones comunicacionales de la compañía, mediante herramientas y técnicas estadísticas para series de tiempo.

4.2. Enfoque

Esta investigación está sustentada dentro de los paradigmas cuantitativo y cualitativo, es decir, un enfoque de carácter mixto.

Es cuantitativo, porque su objetivo es poner a prueba teorías o hipótesis, identificar relaciones causales, busca diferencias entre los grupos o patrones. Se aíslan variables, se controlan factores potenciales e influencias extrañas y se excluyen hipótesis opuestas. Adicionalmente como su nombre lo dice, confía en la medición numérica y en el uso de la estadística, en este caso series de tiempo, para establecer con exactitud patrones de conducta de los usuarios de la marca UA.

Es cualitativo, porque el objeto de estudio es un fenómeno social donde se trata de entender las conductas e intereses de una población y porque argumenta la pregunta de investigación con la recolección de los datos y el análisis e interpretación que se le pueda dar a estos.

Este enfoque ofrece varias ventajas, como lo menciona Todd, Werlich y Mckeown (2004)¹²:

- Incrementa la confianza en que los resultados, son una representación fiel, genuina y fidedigna de lo que ocurre con el fenómeno estudiado, ayuda a clarificar y teorizar el planteamiento del problema.
- Los modelos mixtos logran que se exploren y se exploten mejor los datos, los cuales serán analizados y presentados como resultado de la investigación.

4.3. Diseño Metodológico

Para este trabajo fueron adoptadas tres metodologías, la primera concerniente al proceso de minería de datos, la segunda relacionada al análisis conversacional de los usuarios en marketing digital y la tercera, la cual abarca la mayor parte de este proyecto y que contempla las técnicas estadísticas de series de tiempo a usar.

4.3.1. Metodología Minería de Datos

Se utiliza la metodología ASUM-DM creada por IBM, que es reconocida para realizar proyectos de minería de datos, enfatiza varias de las nuevas prácticas en la ciencia de datos como el uso de volúmenes de información muy grandes, la incorporación de análisis de texto en el modelado predictivo y la automatización de algunos procesos.¹³

Se compone de diez fases como se muestra a continuación en la siguiente figura, con sus respectivas descripciones:

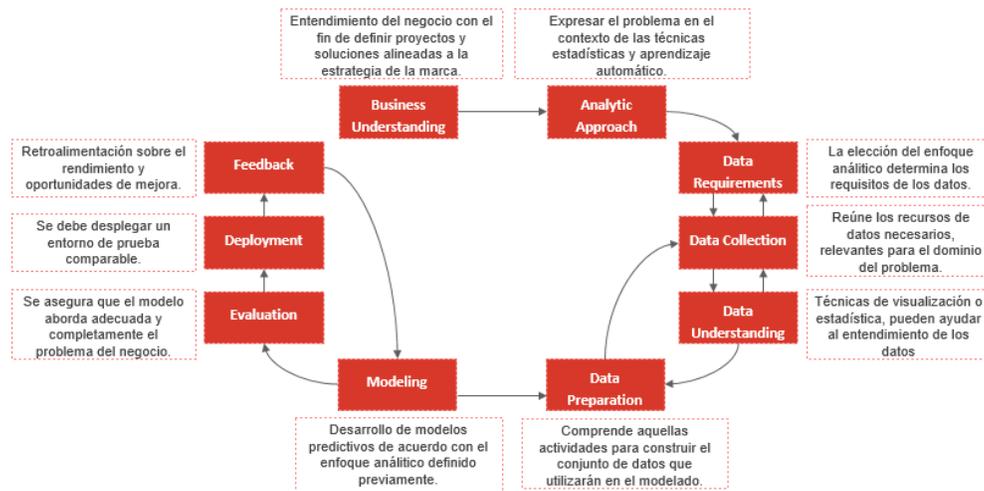


Figura 3. Elaboración propia. Metodología ASUM-DM. Información tomada de <https://www.ibmbigdatahub.com/blog/why-we-need-methodology-data-science>

¹² (Lagarda, 2008) cita a “Todd, Werlich y Mckeown (2004)”.

¹³ (IBM, 2015)

4.3.2. Metodología de Análisis Digital Conversacional

Para poder evaluar los intereses de los usuarios a partir de sus tópicos de conversación en los activos digitales, se adoptara la metodología de TALKERS creada por la agencia publicitaria BRM. Esta metodología permite realizar una búsqueda directa de las menciones a la marca por redes sociales, clasificarla en segmentos y sub-segmentos, hacer una balanza de opinión de ellas y adicionalmente concluir las motivaciones que llevaron a un usuario a publicar estas menciones.

Se compone de cinco fases que son representadas y explicadas en la siguiente figura:



Figura 4. Elaboración propia. Metodología TALKERS.

4.3.3. Metodología Estadística para Series de Tiempo

Para el caso del análisis de series de tiempo, la metodología a usar es la de Box Jenkins publicada por George Box y Gwilyn Jenkins en la década de los 70's en el libro *Time Series Analysis: Forecasting and Control*.¹⁴

El objetivo de la metodología Box Jenkins es identificar y estimar un modelo estadístico que puede ser interpretado como generador de la información de la muestra. En este sentido, si el modelo estimado es usado para la predicción debe suponerse que las características de la serie son constantes en el tiempo, especialmente para los periodos futuros. Por lo tanto, la predicción se efectúa sobre una base válida que el modelo es estacionario o estable.

La metodología Box- Jenkins, también conocida como metodología ARIMA, consiste en un proceso iterativo,* compuesto por cuatro fases:

Identificación: Consiste en identificar las características de la serie de tiempo y encontrar, a través de un autocorrelograma simple y parcial los valores adecuados para p,d,q.

¹⁴ (Box, Jenkins, & Reinsel, 1976)

Estimación: Tras hallar estos tres valores adecuados para el modelo, se procede a estimar los parámetros para cada uno de ellos.

Exámenes de diagnóstico: Habiendo seleccionado el modelo ARIMA, se evalúa si el modelo se ajusta a los datos, de forma razonable.

Pronóstico: Se pronostica para un lapso de tiempo determinado.

Las fases son representadas en la siguiente figura:

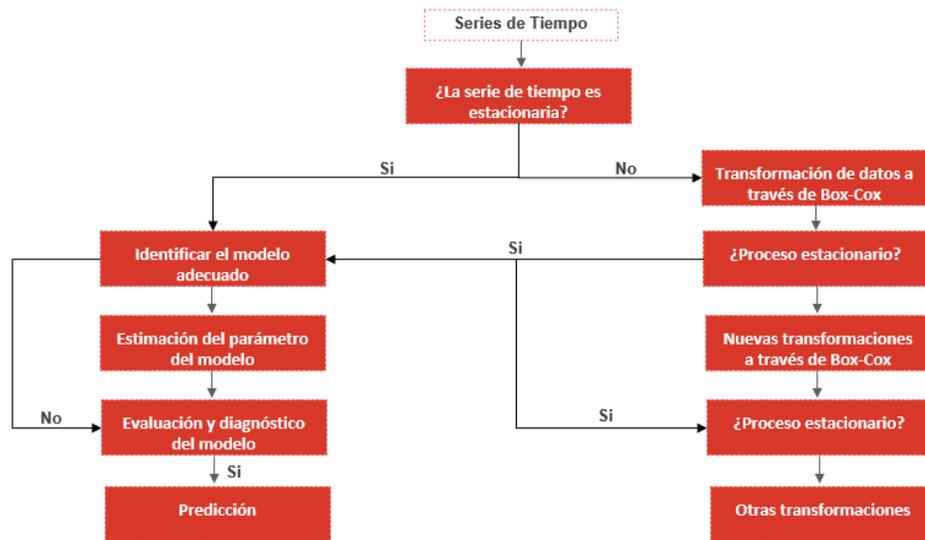


Figura 5. Elaboración propia. Metodología Box-Jenkins. Fuente: (Box, Jenkins, & Reinsel, 1976)

4.4. Acerca de los Datos

4.4.1. Área de Estudio

El área de estudio comprende la publicidad y el marketing digital, dentro de los cuales se consideran las diferentes métricas de una publicación en redes sociales, los datos transaccionales de los clientes y las tasas de conversión de la marca UA sede Colombia. El análisis de las variables en conjunto permite identificar la pertinencia de las acciones comunicacionales de la marca en la generación de tracción o ventas potenciales.

4.4.2. Recolección de Datos

Los datos históricos de ventas fueron suministrados por la compañía a solicitud y con carta de confidencialidad para datos específicos; los datos de desempeño de las piezas publicitarias fueron descargados directamente de las redes sociales Twitter y Facebook, con

el acceso a cada una de las cuentas, que también fueron facilitados por la marca y finalmente la información conversacional y de los intereses de los usuarios fue descargada, apoyándose en la plataforma de social listening Social Studio que permite traquear la información pública de los usuarios en redes.

4.4.3. Población y muestra

El universo del presente trabajo corresponde, a los datos de ventas de la empresa desde su llegada a Colombia a finales de octubre del año 2015. Como los activos digitales para Facebook y Twitter de la marca comenzaron a funcionar en enero del año 2016, mes en el cual inició, a su vez, el proceso publicitario por estos canales, la data comprende desde este mes hasta octubre de 2019, agrupándola de manera mensual. De esta misma manera, la información conversacional e indicadores fueron agrupados con periodicidad mensual para estos cuatro años.

Capítulo 5

Análisis y Resultados

En el presente capítulo serán presentados los resultados de la caracterización de los intereses de los clientes de la marca y su comportamiento de compra, la construcción del indicador de desempeño de las piezas publicitarias para las redes sociales de Facebook y Twitter y los modelos en series de tiempo que fueron obtenidos a partir de los datos disponibles.

Siguiendo la estrategia usada por Lozano (2018)¹⁵ para la evaluación de los modelos que serán presentados, se utilizará el esquema por fuera de la muestra, lo que significa que los datos serán divididos en dos grupos, el primero será llamado de in-sample y contendrá la información en el periodo de enero de 2016 hasta mayo de 2019. Los datos de los 5 meses restantes (junio 2019 hasta octubre de 2019) serán llamados de out-sample. La idea es calcular los pronósticos con los datos de in-sample con el fin de corroborar si los modelos se ajustan bien, deberán ser razonablemente parecidos a los verdaderos valores presentes en el out-sample.

5.1. Caracterización de los Intereses de los Clientes de la Marca y su Comportamiento de Compra

En la siguiente sección se procede a realizar la caracterización mencionada. Los datos que se van a compartir son solo una proporción de los datos reales de la compañía, dado que la información de las ventas y el stock **real** son solo de interés de sus trabajadores para los objetivos estratégicos, tácticos, operativos y comerciales de la marca. Sin embargo, el mostrar una proporción (la misma para todos los datos) será de utilidad, pues el fin es identificar las tendencias de compra de los usuarios y sus intereses.

Basado en la metodología de talkers nombrada en el capítulo 4 de este trabajo, donde a través de la data conversacional por las redes sociales se identifican los intereses del cliente y sus tópicos de conversación, se elaboró la siguiente nube de palabras, limpiando todos los conectores, artículos y preposiciones que no agregan valor a dicho análisis.

Luego se procederá a describir la data transaccional y sus diferentes comportamientos en el periodo de tiempo del gripe de información In-sample.

¹⁵ (Lozano, 2018)



Figura 6: Nube de palabras de los tópicos de conversación de los usuarios. Elaboración propia. Contenidos tomados de las conversaciones de Facebook y Twitter de la marca.

Se observa que los usuarios al hablar de la marca, se encuentran interesados por sus nuevas colecciones, preguntando por ropa deportiva, formal y para ir para la playa. Dentro de los intereses de los hombres se encuentran prendas de vestir como zapatos, chaquetas, blazers y camisas. Las mujeres preguntan por accesorios y ropa para la temporada de vacaciones en diciembre. Adicionalmente, la comunidad indaga a la marca por facilidades de pago en las distintas ciudades y tiempos de entrega de sus pedidos.

En el periodo evaluado, la marca realizó 645 publicaciones entre Facebook y Twitter, veamos la distribución por secciones del tipo de ropa que la comunidad consultó en las redes sociales:

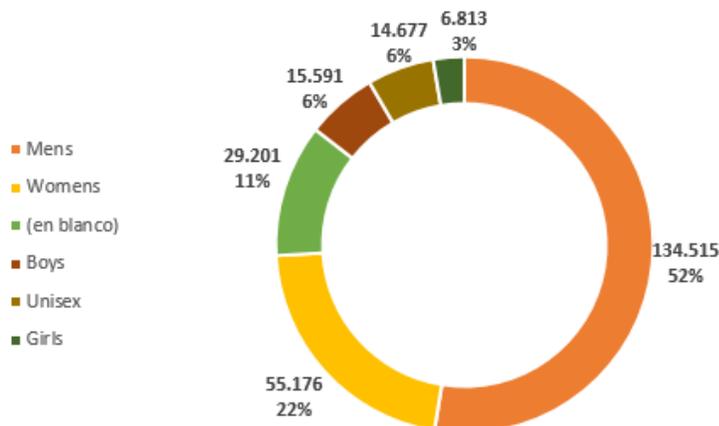


Figura 7: Distribución por secciones. Elaboración propia. Contenidos tomados de las conversaciones de Facebook y Twitter de la marca.

El 58% de la comunidad preguntó por ropa relacionada al género masculino (hombres y

niños), mientras que el 33% indagó por prendas para mujer y niñas, lo cual evidencia un potencial para seguir mejorando en esta línea de ropa femenina.

Al calcular la proporción de consultas por tipo de producto que indagan los clientes, se tuvo que:

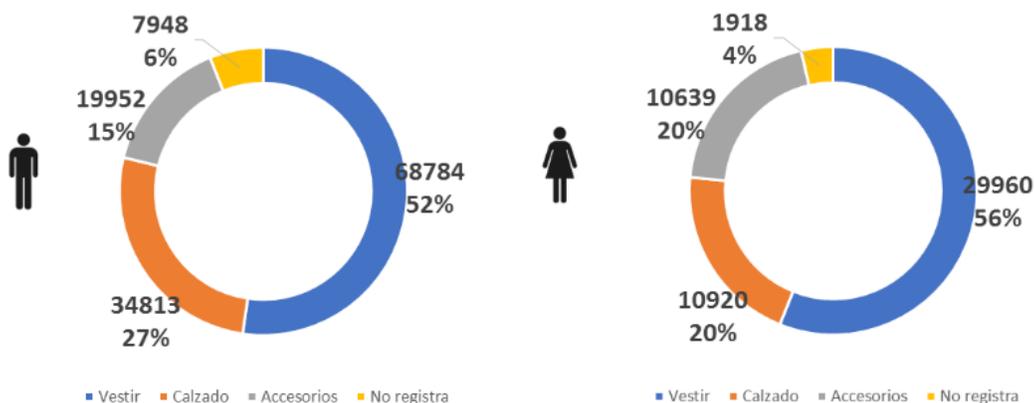


Figura 8: Distribución por tipo de producto. Elaboración propia. Contenidos tomados de las conversaciones de Facebook y Twitter de la marca.

Los usuarios tanto hombres como mujeres preguntan casi en la misma proporción por prendas de vestir, siendo el porcentaje de las mujeres un poco mayor. Por otra parte se evidencia que los hombres están más interesados en el calzado que le ofrece la marca mientras que las mujeres en los accesorios que puedan comprar.

Si indagamos por la categoría de las prendas de vestir:

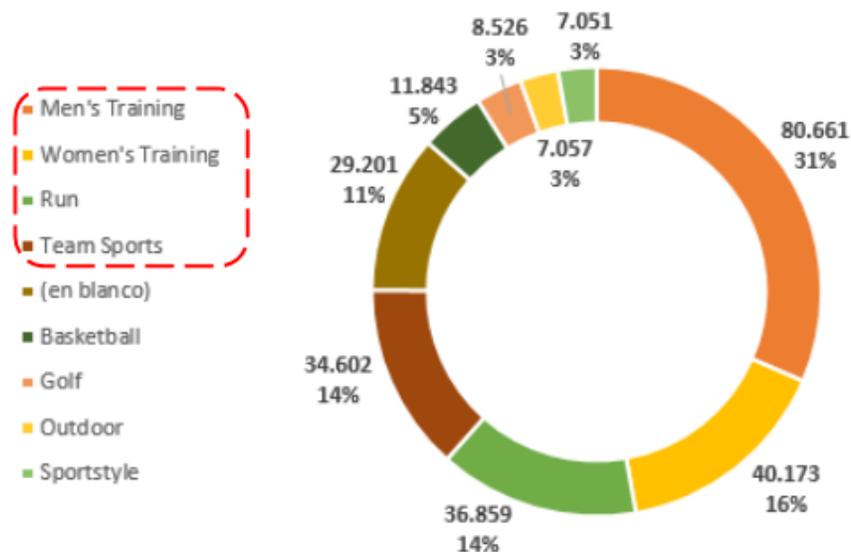


Figura 9: Distribución por categoría. Elaboración propia. Contenidos tomados de las conversaciones de Facebook y Twitter de la marca.

Se entiende que existe una atracción de la comunidad hacia las prendas de tipo deportivo que ofrece la compañía, lo cual puede ser un insumo importante para sus publicaciones futuras.

Con la información anterior capturada de las conversaciones de las redes sociales, se identificaron algunos de los intereses de los consumidores y es una guía para el equipo creativo de la marca para potenciar las interacciones con sus diferentes publicaciones.

5.2. Construcción de la Medición de Desempeño de las Piezas Publicitarias en Redes Sociales

Uno de los grandes objetivos de este trabajo es proponer un indicador que condense las métricas obtenidas en redes sociales como Facebook y Twitter y que refleje el desempeño obtenido de una pieza publicitaria. Este indicador es una construcción propia, diseñada para reflejar las percepciones e interacciones que un cliente aporta a la marca. Para poder proceder es necesario hacer claridad sobre algunos términos:

- **Publicación Pautada:** Publicaciones en las cuales se paga por una segmentación de audiencia para que la publicidad sea mostrada a estos usuarios y no la pasen por alto.
- **Publicación Orgánica:** Es aquel que se publica en redes sociales y no tiene ningún coste.
- **Alcance:** El número de personas únicas que vieron la publicación.
- **Impresiones:** El número de veces que la publicación fue vista sin importar el número de repeticiones que la pueda visualizar el mismo usuario.
- **Shares:** Número de veces que se comparte una publicación.
- **Views:** Número de visualizaciones de un contenido de formato video o gif. Se tendrán en cuenta solamente las visualizaciones que superen el 95% de la reproducción
- **Retweet:** Número de Tweets copiados en el perfil propio.

Para el caso de la marca UA al tener solo activos digitales en Twitter y Facebook en Colombia, la medición se realiza con las siguientes ecuaciones:

Twitter:

En Twitter, es de conocimiento la no existencia de alcance en sus métricas; por eso el indicador de base son las impresiones; pero teniendo en cuenta que su número de impresiones se ve afectada si el contenido es orgánico o pagado, se da una ponderación según sea el caso:

- Si el contenido es “pagado”, la formula a usar sería:

$$\text{Impacto} = 30\% \left[\frac{\text{Impresiones}}{\text{Máx Impresiones}} \right] + 15\% \left[\frac{\text{Interacciones}}{\text{Máx Interacciones}} \right] + 25\% \left[\frac{\text{Interacciones}}{\text{Impresiones}} \right] + 15\% \left[\frac{\text{Retweets}}{\text{Interacciones}} \right] + 15\% \left[\frac{\text{Interacciones}}{\text{último día mes} - \text{día publicación}} \right]$$

- Si el contenido es “orgánico”, la formula a usar sería:

$$\text{Impacto} = 40\% \left[\frac{\text{Impresiones}}{\text{Máx Impresiones}} \right] + 15\% \left[\frac{\text{Interacciones}}{\text{Máx Interacciones}} \right] + 20\% \left[\frac{\text{Interacciones}}{\text{Impresiones}} \right] + 10\% \left[\frac{\text{Retweets}}{\text{Interacciones}} \right] + 15\% \left[\frac{\text{Interacciones}}{\text{último día mes} - \text{día publicación}} \right]$$

Las **interacciones** a medir son las siguientes:



Facebook:

El medir un top de publicaciones por engagement lleva a ubicar contenido de gran alcance como el peor contenido, siempre que no responda con un mayor número de interacciones. Por esta razón, se generó un indicador de impacto diferenciando su ponderación, según si el contenido es pago u orgánico:

- Si el contenido es “pautado”, la formula a usar sería:

$$\text{Impacto} = 40\% \left[\frac{\text{Alcance}}{\text{Máx Mes}} \right] + 10\% \left[\frac{\text{Interacciones}}{\text{Max Mes}} \right] + 20\% \left[\frac{\text{Usuarios que interactúan}}{\text{Máx Mes}} \right] + 15\% \left[\frac{\text{interacciones}}{\text{Impresiones}} \right] + 15\% \left[\frac{\text{Interacciones}}{\text{último día mes} - \text{día publicación}} \right]$$

- Si el contenido es “orgánico”, la formula a usar sería:

$$\text{Impacto} = 50\% \left[\frac{\text{Alcance}}{\text{Máx Mes}} \right] + 10\% \left[\frac{\text{Interacciones}}{\text{Max Mes}} \right] + 15\% \left[\frac{\text{Usuarios que interactúan}}{\text{Máx Mes}} \right] + 10\% \left[\frac{\text{interacciones}}{\text{Impresiones}} \right] + 15\% \left[\frac{\text{Interacciones}}{\text{último día mes} - \text{día publicación}} \right]$$

Las **interacciones** a medir son las siguientes:



Para las **reacciones**, se debe tener en cuenta lo siguiente:



De esta manera se calcula el desempeño de una pieza publicitaria según las interacciones de los usuarios con la creatividad por red social.

Finalmente, este indicador se construye combinando los resultados de las publicaciones **pautadas** tanto de Facebook como de Twitter de la siguiente manera:

$$\text{Indicador desempeño} = (0,5 \times \text{Impacto en } \textcircled{f}) + (0,5 \times \text{Impacto en } \textcircled{t})$$

Este indicador fue calculado con los datos reales de las redes sociales de la marca UA sede Colombia para desarrollar la siguiente sección de este capítulo, teniendo en cuenta solamente lo pautado en ambas redes sociales.

5.2. Modelo SARIMAX

Al considerar el comportamiento de las ventas de la compañía UA sede Colombia, para el periodo comprendido entre enero de 2016 y mayo de 2019, se observa que históricamente estas presentan volatilidad y un comportamiento cíclico a lo largo del tiempo. En la figura 10 se presentan la serie del índice de ventas e indicador de desempeño de las piezas publicitarias (se calculó el promedio de resultados de las publicaciones por mes). Visualmente, se puede observar, que en general, el indicador propuesto de desempeño de las piezas publicitarias, parece dar buena cuenta de la variable de interés, ya que es similar en algunos picos y en la tendencia que presenta.

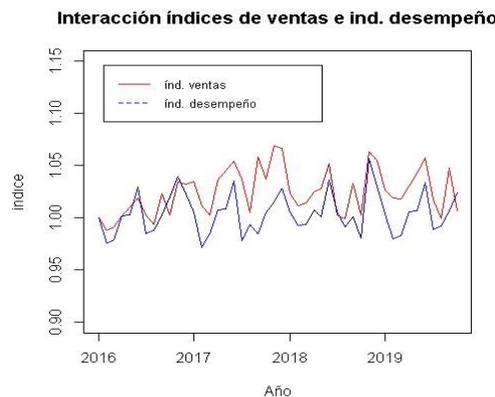


Figura 10: Interacción en el tiempo entre los índices considerados. Elaboración propia.

Para la etapa de identificación, de acuerdo al esquema Box-Jenkins es necesario considerar los autocorrelogramas simples y parciales de los datos como se muestra en la figura 11.

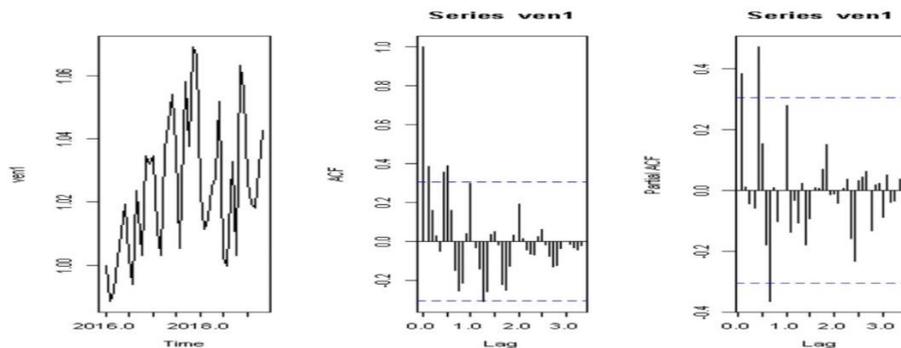


Figura 11: Serie de ventas y autocorrelogramas simple y parcial. Elaboración propia.

Se puede observar en la serie de ventas (grafico izquierda), que la serie tiene un comportamiento aparentemente estacional año tras año, donde los datos más altos se encuentran para los meses de final de año. Sin embargo, la serie no se ve constante en su media por ende se puede considerar que la serie no es estacionaria. Al evaluar los correlogramas, el ACF y PACF corroboran los picos secuenciales de una serie estacional. La no convergencia a cero y algunos rezagos que se salen del intervalo de confianza, también nos hacen suponer que la serie no es estacionaria.

La figura 12 muestra los resultados de la prueba de Dickey-Fuller, aportando un p-valor del 0,0652, es decir que el 6,5% del modelo tiene raíz unitaria, por tanto, no se puede rechazar la hipótesis nula para un nivel de significancia del 5% y se corrobora que el proceso no es estacionario.

```

Augmented Dickey-Fuller Test

data: ven1
Dickey-Fuller = -3.4375, Lag order = 3, p-value = 0.06526
alternative hypothesis: stationary

```

Figura 12: Test Dickey-Fuller. Elaboración propia.

La figura 13, presenta los datos con dos diferencias tanto para su componente estacional como para su componente no estacional. Esto se realizó con el objetivo de tener una media constante en la serie y convertirla en estacionaria.

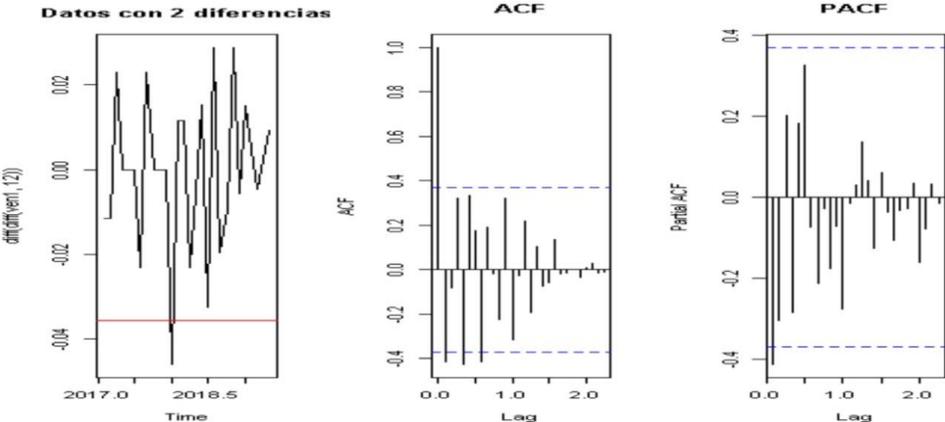


Figura 13: Serie de ventas y autocorrelogramas simple y parcial después de diferenciar dos veces. Elaboración propia.

Después de un proceso de búsqueda del modelo adecuado, en donde fueron tenidos en cuenta los resultados de los autocorrelogramas, la prueba de raíz unitaria, el comportamiento de los residuos, el criterio de información BIC y la menor cantidad de parámetros, se llegó a que el mejor modelo para describir y pronosticar el nivel de ventas de la compañía UA sede Colombia es un modelo SARIMAX (11,1,0) x (1,1,0)12 que incorpora los valores del

indicador de desempeño de las piezas publicitarias en redes sociales, un valor contemporáneo y uno rezagado. Se encuentra que todos los parámetros son estadísticamente significativos, como se muestra en la figura 14 y que el valor del BIC fue de 146,99.

ar11	sar1	A1	A2
2.150816	-3.049685	-2.544410	-4.797345

Figura 14: Prueba de significancia del modelo ajustado con la incorporación de la variable exógena. Elaboración propia.

En la figura 15, se presentan los insumos para el diagnóstico del modelo. Se muestran los datos originales versus los datos ajustados, revelando un buen ajuste. Según el ACF y PACF de los residuos, no se evidencia algún problema de autocorrelación serial. En el qqplot de los residuos, aparentemente los datos no se distribuyen de manera normal, lo cual corroboraremos con el test de Jarque-Bera, porque se salen del umbral en las colas.

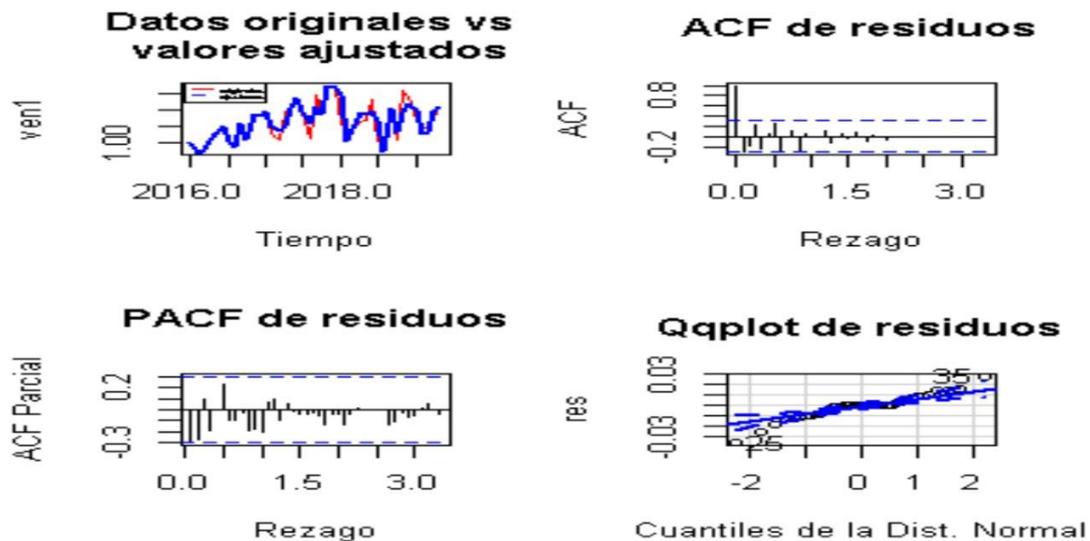


Figura 15: Validación del modelo SARIMAX planteado en esta sección. Elaboración propia.

Con el ánimo de confirmar la pertinencia del modelo elegido, se procede a realizar la validación de los supuestos de autocorrelación serial y normalidad de los residuos.

En el primer caso, se realiza una prueba de Ljung-Box. Los resultados revelados en la figura 16 permiten concluir que los errores son independientes para un nivel de significancia del 5% y, por ende, no existe autocorrelación serial de los mismos aceptando la hipótesis nula.

En segundo lugar, la prueba Jarque-Bera, para un nivel de significancia del 5%, sugiere que como el p-valor es de 0,008756, los residuos no se distribuyen normalmente y se configuran como un ruido blanco.

Box-Ljung test

```
data: res
X-squared = 3.8068, df = 1, p-value = 0.05104
Jarque Bera Test
```

```
data: res
X-squared = 9.4761, df = 2, p-value = 0.008756
```

Figura 16: Test Ljung-Box y Jarque-Bera para los residuos. Elaboración propia.

Finalmente, se evaluarán los pronósticos del modelo para ver qué tan cercanos son, a los datos contenidos en el gripo Out-Sample. En la figura 17, se presentan en rojo los datos originales y en azul los pronósticos generados con el modelo.

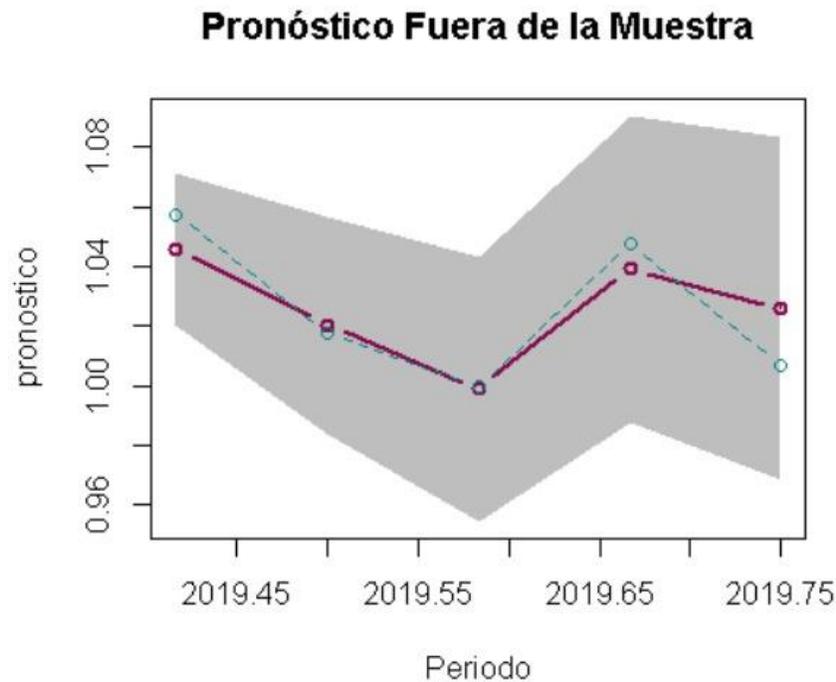


Figura 17: Pronostico in-sample del modelo SARIMAX (11, 1,0) x (1, 1,0)12. Elaboración propia.

Se evidencia que los valores estimados se encuentran razonablemente cerca a los originales, de esta manera el modelo propuesto puede ser utilizado por la compañía para pronosticar la variación en los niveles de ventas a partir de la medición del desempeño de sus piezas publicitarias en el ecosistema digital con el indicador propuesto en la segunda sección del presente capítulo.

Capítulo 6

Conclusiones y Recomendaciones

La tracción y tasa de conversión en ventas que puede traer una campaña publicitaria lanzada en los activos digitales de una marca es un fenómeno poco estudiado en Colombia y requiere la formalización, consignación y publicación de los hallazgos y métodos utilizados, para tener más herramientas en nuestro país que nos acerquen al entendimiento y construcción del ecosistema digital.

Así como muchas compañías la marca estadounidense UA con sede en Colombia, necesita no solo entender un poco de los tópicos de conversación que se generan alrededor de su marca desde las redes sociales, sino que necesitaba evaluar el impacto de sus acciones comunicacionales en los comportamientos de compra de los consumidores.

Este trabajo, llevo a entender que la marca está siendo reconocida y aceptada por la población masculina, en particular por su línea deportiva y calzado; adicionalmente la comunidad habla positivamente de los productos de la compañía y están interesados en nuevas colecciones y opciones de pago, esto genera un parte de tranquilidad pero a la vez un gran reto para todas las áreas internas con las secciones que no están siendo tan visibles y acogidas por los usuarios.

Por otra parte fue creado un indicador que consolida la información de las métricas de redes sociales como Facebook y Twitter. Dicho indicador, presenta un comportamiento razonable para entender la variable de interés, que en este trabajo, es el nivel de ventas de la compañía, sin embargo, sería bueno para un trabajo posterior agregar variables al modelo como la inversión y el volumen de publicaciones por mes, para hacer más robusto el estudio.

Se encontró un modelo SARIMAX para la variación de las ventas de la compañía, logrando que los pronósticos de dicho modelo fueran evaluados y corroborados a través del esquema in-sample y out-sample.

Se concluye que el indicador construido, se muestra competitivo para el entendimiento o generación rápida de pronósticos para las ventas en función del desempeño de piezas publicitarias, que puede ser replicado en otras compañías con la intención de incursionar en el mundo digital para establecer sus estrategias de marketing.

Bibliografía

- Assmus, G., Farley, J., & Lehmann, D. (1984). How Advertising Affects Sales: Meta-Analysis of Econometric Results. *Journal of Marketing Research*, 21(1), 65-70.
- Bass, F., & Clarke, D. (1972). Testing Distributed Lag Models of Advertising Effect. *Journal of Marketing Research*, 9(3), 298.
- Beckwith, N. (1972). Multivariate Analysis of Sales Responses of Competing Brands to Advertising. *Multivariate Analysis of Sales Responses of Competing Brands to Advertising. Journal of Marketing Research*, 9(2), 168.
- Benesch, C. (1952). Sales Tests as a Measure of Advertising Effectiveness. . *Journal of Marketing*, 17(2), 178.
- Box, G., Jenkins, G., & Reinsel, G. (1976). *Time Series Analysis: Forecasting and Control*. New Jersey: John Wiley & Sons.
- Demirdjian, Z. (1983). Sales effectiveness of comparative advertising: an experimental field investigation. *Journal of Consumer Research*, 10 (3), 362-364.
- Dubinsky, A., Barry, T., & Kerin, R. (1981). The salesadvertising interface in promotion planning. *Journal of Advertising*, 10 (3), 35-41.
- Fritis, N., Polit, E., Ramirez, M., & Young, W. (2018). *Modelos para el apoyo de*. Chile: Universidad de Chile.
- Guevara, I., & Guevara, N. (Septiembre de 2019). *Pronóstico de la relación entre tasa de subempleo subjetivo por competencias y la tasa desempleo en Colombia entre enero del 2010 hasta diciembre de 2018 mediante un modelo ARIMAX de series de tiempo*. Bogotá, Colombia: Trabajo de grado: Especialización en Estadística Aplicada. Fundación Universitaria los Libertadores.
- Hernandez Sampieri, R., & Fernandez Collado, C. (2006). *Metodología de la investigación*. México: (4ª ed.). HcGraw-Hill).

- Holak, S., & Tang, Y. (1990). Advertising's Effect on the Product Evolutionary Cycle. *Journal of Marketing*, 54(3), 16.
- Hootsuite. (23 de 10 de 2019). *wearesocial*. Obtenido de <https://wearesocial.com/blog/2019/10/the-global-state-of-digital-in-october-2019>.
- IAB Colombia. (29 de 03 de 2019). *iabcolombia*. Obtenido de <http://www.iabcolombia.com/guias-y-estandares/informe-anual/>
- IBM. (24 de 08 de 2015). *ibmbigdatahub*. Obtenido de <https://www.ibmbigdatahub.com/blog/why-we-need-methodology-data-science>
- Ihaka, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5.
- Lagarda, R. (2008). Seminario Teórico de comunicación educativa I. *Enfoque Mixto*. México.
- Londoño, S., & Mora, Y. (2018). Modelos estadísticos sobre la eficacia del marketing digital. *Revista EAN*, 167-186.
- Lozano, S. (2018). *Propuesta de un índice SIPSA para el pronóstico de la inflación de Alimentos. Evidencia empírica*. Bogotá, Colombia: Trabajo de grado: Especialización en Estadística Aplicada. Fundación Universitaria los Libertadores.
- Lu"tkepohl, H. (2005). *New Introduction To Multiple Time Series Analysis*. Springer.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6.
- Tsay, R. (2005). *Analysis of Financial Time Series*. John Wiley & SONS. ISBN 0-471-690740.
- Tsay, R. (2013). *Multivariate Time Series Analysis: With R and Financial Applications*. Wiley, 2 edition.
- Villaseca, D. (2014). *Innovación y marketing de servicios en*. Madrid: ESIC.