UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICA

Modelos ARIMA para la agrupación de palabras claves usadas en el Sistema AdWords.

Trabajo Especial de Grado presentado ante la ilustre Universidad Central de Venezuela por la **Br. Yoselin Arvelaiz Fernández** para optar al título de Licenciada en Matemática.

Tutor: Dra. Mairene Colina.

Caracas, Venezuela
Julio 2018

Nosotros, los abajo firmantes, designados por la Universidad Central de Venezuela como integrantes del Jurado Examinador del Trabajo Especial de Grado titulado "**Modelos ARIMA para la agrupación de palabras claves usadas en el Sistema AdWords**", presentado por la **Br. Yoselin Arvelaiz Fernández**, titular de la Cédula de Identidad **22.887.653**, certificamos que este trabajo cumple con los requisitos exigidos por nuestra Magna Casa de Estudios para optar al título de **Licenciada en Matemática**.

Dra. Mairene Colina
Tutor

Dr. José Hernández
Jurado

Solorzano Manuel

Lic. Manuel Solórzano
Jurado

A Dios, a mis padres y a mis hermanas.

Agradecimiento

Primeramente le agradezco a Dios por haberme permitido llegar hasta este punto y haberme dado todo lo necesario para lograr mis objetivos, además de su infinita bondad y misericordia. A mis padres, por haberme apoyado en todo momento, por sus esfuerzos, por sus valores y por sobre todo por el amor. A mis hermanas, por estar siempre presente y por brindarme un apoyo incondicional. A mi tía Viviana, quién desde el principio de la carrera me ha motivado a ser una profesional.

También le agradezco de todo corazón a la Sra. Doris Vásquez, quién me brindo un hogar mientras estudiaba, y no sólo eso, sino por hacerme parte de su familia y haberme inculcado grandes valores cristianos.

No puedo dejar de agradecerle a todos mis amigos y compañeros de la carrera, más aún a José Antonio por motivarme, por apoyarme y ayudarme en momentos difíciles de la carrera.

Agradecerle a todos los profesores, quienes me formaron en la carrera y más aún a mi tutora Mairene Colina por tomarse su tiempo y brindarme todas las herramientas necesarias para abordar éste trabajo.

También a todas aquellas personas que con sus palabras de aliento me ayudaron en el transcurrir de la carrera.

Resumen

El programa de Google AdWords trabaja con lo que se conoce como Posicionamiento SEM (Search Engine Marketing), que quiere decir Marketing en Buscadores.

La estrategia de marketing a través del Posicionamiento SEM, se realiza por medio del uso de **palabras clave** de búsqueda, que el anunciante seleccionará para publicitar su anuncio. En éste trabajo presentaremos un conjunto de datos de una campaña publicitaria de una compañía de ventas de sombreros por internet, que consta de 7155 palabras claves y 12 variables usadas en el sistema AdWords.

A este conjunto de datos se le aplica una reducción de dimensionalidad mediante el Análisis de componentes principales, posteriormente se le aplica la técnica de agrupamiento PAM. Una vez que se consigue la clasificación del conjunto de palabras se destacan las características de estos grupos o clusters y finalmente se ajustan modelos ARIMA y SARIMA a las impresiones de los grupos más ricos en información.

Palabras Claves:

Modelos ARIMA, SARIMA, Series Temporales, Agrupación, PAM, Análisis de Componentes Principales (ACP), Google AdWords.

Índice general

Introducción

Imposible no reconocer que cuando estamos en búsqueda de alguna definición, información, servicio o producto, se recurre a un buscador en Internet. El más popular, dada la cantidad de visitas diarias que recibe, es **Google**, quien por excelencia se ha ganado el puesto y reputación que tiene. Desde hace 14 años que tuvo que modificar sus políticas internas para mejorar sus servicios, permitiendo pagar por anunciar un sitio web en particular.

De esta manera, se lleva a cabo una de las estrategias de marketing más utilizadas en la actualidad, **Google AdWords**, que es el programa que le permitirá publicitar de forma patrocinada, la página web de su negocio por medio de su anunciante potencial para alcanzar el mejor puesto en el ranking de posicionamiento.

El programa de Google AdWords trabaja con lo que se conoce como Posicionamiento SEM (Search Engine Marketing), que quiere decir Marketing en Buscadores. Es decir, Google introduce en su motor de búsqueda, enlaces patrocinados de sitios web, que realizan un pago por una campaña de Adwords, para convertirse en la primera opción de visita de un potencial cliente.

Esta atractiva forma de posicionar una página web, nace de la necesidad de llegar a más potenciales consumidores de manera más rápida y eficaz dada la enorme competencia que existe actualmente en el mercado para los diversos rubros. Por lo tanto, para recurrir a estos enlaces patrocinados, se factura un valor determinado por cada clic realizado por un usuario o cliente en cuestión.

La estrategia de marketing a través del Posicionamiento SEM, se realiza por medio del uso de **palabras clave** de búsqueda, que el anunciante seleccionará para publicitar su anuncio. Es decir, si su empresa se dedica a la venta de muebles para oficina, el experto en SEM, usará palabras clave que se relacionen directamente con sus productos como 'escritorio para oficinas', 'sillas de escritorios', 'muebles de oficina', entre otras. Basándose en los criterios de búsqueda que usan los usuarios o potenciales clientes a través de Google.

Para lograr un Posicionamiento SEM idóneo para su sitio web y al mismo tiempo, para su negocio; la campaña de AdWords debe contemplar utilizar las mejores palabras clave de búsqueda. Por esta razón, en éste trabajo presentaremos un conjunto de datos de una campaña publicitaria de una compañía de ventas de sombreros por internet, que consta de 7155 palabras claves y 12 variables usadas en el sistema AdWords, para agrupar y ajustar modelos que permitan caracterizar las palabras claves.

Para agrupar este conjunto de palabras claves usaremos métodos de agrupamiento cómo: **K-medias** y **PAM** ó K-centroides. Para esto también utilizaremos una importante técnica de reducción de dimensionalidad de un conjunto de datos llamado **Análisis de componentes principales** o ACP, ésta técnica nos permitirá trabajar con variables no correlacionadas que aporten una mayor variabilidad de los datos, permitiendo así reducir la cantidad de datos a estudiar y haciendo más fácil el hecho de agrupar las palabras.

Para ajustar modelos utilizaremos Series de tiempo, específicamente **Modelos ARIMA ó SARIMA** para las **impresiones** de la agrupación de palabras claves con el fin de caracterizar el comportamiento de las impresiones de las palabras claves y poder realizar pronósticos que puedan ser de utilidad para obtener información sí un grupo de palabras tiene un buen rendimiento o no.

En éste trabajo comenzaremos desarrollando en el capítulo 1 algunas Técnicas de Análisis Multivariado de Datos, tales cómo el Análisis de Componentes Principales y el Análisis de Grupos. En el capítulo 2, abordaremos aspectos teóricos de Series de Tiempo y Google AdWors. Finalmente, en el capítulo 3, aplicaremos éstos análisis al conjunto de palabras claves, con el fin de agruparlas en grupos bien segmentados para caracterizar y poder ajustar modelos ARIMA con el propósito de hacer pronósticos sobre éste.

Capítulo 1

Técnicas de Análisis Multivariado de Datos.

En éste capítulo desarrollaremos algunas Técnicas de Análisis Multivarido de Datos, las cuales son: El Análisis de Componentes Principales y el Análisis de Grupo o Agrupamiento, donde haremos mención de definiciones, métodos y diferentes algoritmos de agrupamiento como son: K-medias, PAM y CLARA.

1. Análisis de Componentes Principales.

El análisis de componentes principales (ACP), PCA en sus siglas en Inglés, es una técnica multivariada de datos que se basa principalmente en la reducción de dimensionalidad de un conjunto de datos. Las componentes principales son combinaciones lineales de las variables originales, las cuales no son correlacionadas y ordenadas de modo que los primeros pocos representen la mayor variabilidad de todas las variables originales. En otras palabras, el resultado del análisis de componentes principales podría ser la creación de un conjunto más pequeño de nuevas variables que pueden ser usadas para sustituir las variables originales y en consecuencia proveen una base más simple para el tratamiento de los datos.

Definición 1.1. El *Análisis de Componentes Principales* (ACP) describe la variación en un conjunto de variables correlacionadas, $x = (x_1, x_2, ..., x_q)$, en términos de una conjunto nuevo de variables no correlacionadas, $y = (y_1, y_2, ..., y_q)$, la cuál cada una es combinación lineal de la variable x.

Cada una de las nuevas variables tienen un grado de importancia de acuerdo al orden, es decir, la primera nueva variable y_1 representa la mayor variabilidad de los datos originales, y_2 se elige para representar la mayor cantidad posible de la variación restante, sujeta a no estar correlacionada con y_1, y así sucesivamente. Las nuevas variables definidas por este proceso, $(y_1, y_2, ..., y_q)$ son las componentes principales.

El análisis de componentes principales busca la proyección en la cual los datos queden mejor representados en término de mínimos cuadrados. Esta convierte un conjunto de observaciones de variables posiblemente correlacionadas en un conjunto de valores sin correlación lineal.

Algebráicamente, la primera componente principal de las observaciones, y_1, es la combinación lineal de $x_1, x_2, ..., x_q$

$$y_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1q}x_q,$$

Donde, $a_{11}, a_{12}, ..., a_{1q}$ son los pesos, que matemáticamente se determinan para maximizar la variación de la composición lineal, tal que la suma al cuadrado de los pesos sea igual a uno, es decir, $\sum_{i=1}^{q} a_{1i}^2 = 1$.

Para la segunda componente principal, implica encontrar un segundo vector de peso $a_{21}, a_{22}, ..., a_{2q}$ tal que la varianza de

$$y_2 = a_{21}x_1 + a_{22}x_2 + ... + a_{2q}x_q,$$

es maximizada bajo la restricción de estar no correlacionada con la primera componente principal y_1 y $\sum_{i=1}^{q} a_{2i}^2 = 1$.

Este proceso continúa hasta tantas componentes cómo variables desee calcular. Sin embargo, las dos primeras componentes principales usualmente tienen la mayor variabilidad en las variables y en consecuencia son de gran interés ya que pueden proveer la mayor información de la estructura de los datos.

El ACP se emplea sobre todo en *análisis exploratorio de datos* para construir modelos predíctivos. El ACP comporta el cálculo de la descomposición en autovalores de la matriz covarianza, normalmente tras centrar los datos en la media de cada atributo.

Es importante mencionar que existen dos formas básicas de aplicar el ACP:

1. *Método basado en la matriz de correlación,* cuando los datos no son dimensionalmente homogéneos o el orden de magnitud de las variables aleatorias medidas no es el mismo.
2. *Método basado en la matriz de covarianza*, que se usa cuando los datos son dimensionalmente homogéneo y presentan valores medios similares.

2. Análisis de Grupos.

El análisis de grupos es uno de los métodos impotantes de la minería de datos para descubrir conocimiento de un conjunto de datos multidimensional. El objetivo de agrupar es identificar patrones o grupos de objetos similares dentro de un conjunto de dato de interés.

2.1. Medidas de distancia de agrupamiento.

La clasificación de observaciones dentro de grupos requiere algunos métodos para calcular la distancia o disimilitud entre cada par de observaciones. El resultado de éste cálculo es conocido como una disimilitud o matriz de distancia.

La escogencia de una medida de distancia es un paso crítico para la agrupación de las observaciones. Ésto define cómo se calcula la similitud de dos elementos (x, y) y cómo influirá en la forma de los grupos.

Hay muchos métodos para calcular ésta información de distancia. A continuación se describirán las medidas de distancias más comunes.

Los métodos clásicos para medidas de distancia son *Distancia Euclídea* y la *Distancia Manhattan*. Se definen como sigue:

1. *Distancia Euclídea*

$$d_{euc}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}.$$

2. *Distancia Manhattan*

$$d_{man}(x, y) = \sum_{i=1}^{n} |(x_i - y_i)|.$$

Donde, x y y son dos vectores de longitud n.

Otras medidas de disimilitud son:

1. *Distancia de correlación Pearson.*
2. *Distancia de correlación Eisen.*
3. *Distancia de correlación Spearman.*
4. *Distancia de correlación Kendall.*

Distancias basadas en correlaciones, las cuales son ampliamente utilizadas para el análisis de datos de expresión génica.

2.2. Estandarización de los datos.

El valor de la medida de distancia está íntimamente relacionado con la escala en la que se realizan las mediciones. Por lo tanto, las variables a menudo se escalan antes de medir las diferencias de interrelación. Esto es recomendable particularmente cuando las variables tienen medidas en escalas diferentes (Por ejemplo: kilometros, kilogramos, centimetros,...); de otra manera las medidas de disimilitud obtenidas pueden ser afectadas.

La meta es hacer las variables comparables. Generalmente las variables están escaladas para tener,

i. Desviación estándar cero.

ii. Media cero.

Al escalar variables, los datos se pueden transformar de la siguiente manera:

$$\frac{x_i - \text{center}(x)}{\text{scale}(x)}.$$

Donde center(x) puede ser la media o mediana de los valores x y scale(x) puede ser la desviación estándar (SD), el rango intercuartil o la desviación media absoluta.

2.3. Particionamiento de grupo.

Son métodos de agrupamientos usados para clasificar observaciones, dentro de un conjunto de datos, en grupos múltiples basados en sus similitudes.

Los grupos de particionamiento comúnmente utilizados, son:

- **Agrupamiento K-medias.**
- **Agrupamiento K-centroides o PAM** (Particionamiento alrededor de centroides).
- **Algoritmo CLARA** (Agrupamientos de grandes aplicaciones).

2.4. Agrupamiento K-medias.

El agrupamiento K-medias (MacQueen, 1967) es el algoritmo de aprendizaje no supervisado más comúnmente usado para particionar un conjunto de datos dado en un conjunto de k grupos llamados *k clusters*, donde k representa el número de grupos pre-especificados por el analista. Esto clasifica objetos en múltiples grupos (es decir, clusters), tales que los objetos dentro del mismo cluster o grupo son tan similares posibles, mientras que los objetos de diferentes clusters son tan disimilares posibles. En el agrupamiento K-medias, cada cluster es representado por centros llamados *centroides* los cuales corresponden a la media de los puntos asignados a cada cluster.

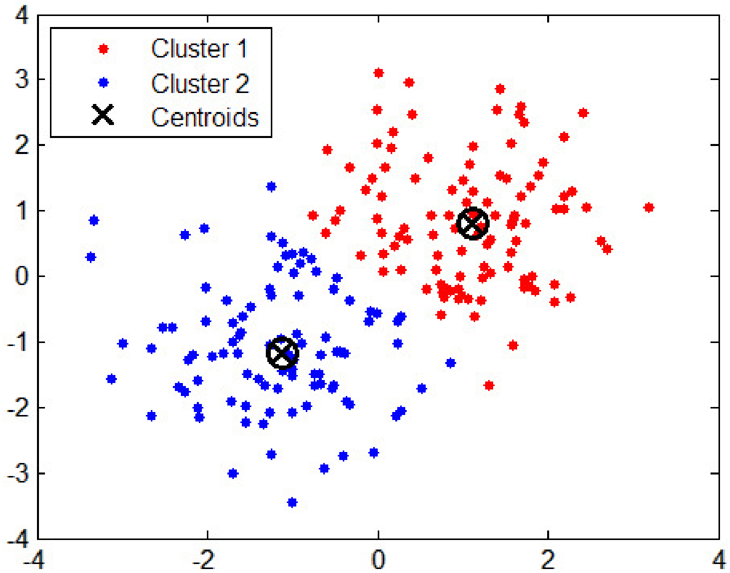

FIGURA 1.1. Agrupamiento K-medias para dos grupos o dos clusters.

La idea básica detrás del agrupamiento K-medias consiste en definir clusters para minimizar la variación total dentro del cluster.

Hay varias algoritmos de K-medias disponibles. El algoritmo estándar es el algoritmo de Hartigan-Wong (1979), la cual define la variación dentro del cluster como la suma de las distancias al cuadrado (distancia Euclídea), distancia entre puntos del clusters y el centro correspondiente:

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2,$$

- x_i un punto de los datos perteneciente al cluster C_k.
- μ_k la media de los puntos asignados al cluster C_k.

Cada observación x_i se asigna a un cluster dado de tal manera que la suma de la distancia al cuadrado de las observaciones a su centro sea mínima.

Se define la variación total dentro de los clusters como:

$$tot.withinss = \sum_{k=1}^{k} W(C_k) = \sum_{k=1}^{k} \sum_{x_i \in C_k} (x_i - \mu_k)^2.$$

Las variaciones dentro del cluster miden la compacidad de la agrupación y queremos que se lo más pequeña posible.

2.5. Agrupamiento K-centroides.

El algoritmo K-centroides es un enfoque de agrupamiento relacionado con el agrupamiento K-medias para particionar un conjunto de datos en k grupos o clusters. En el algoritmo K-centroides, cada cluster es representado por uno de los puntos de los datos en el cluster, éstos puntos son llamados centroides del cluster o *medoids*.

El término centroide o *medoid* se refiere a un objeto dentro de un cluster cuyo promedio de disimilitud entre él y los demás miembros es mínimo. Éste corresponde al punto más céntrico del cluster. Éstos objetos (uno por cluster) pueden considerarse un ejemplo representativo de los miembros de ese grupo que pueden ser útiles en algunas situaciones. Recuerde que, en el agrupamiento K-medias, el centro de un cluster se calcula como la media de todos los puntos en el cluster.

K-centroides es una alternativa robusta del agrupamiento K-medias. Esto significa que, el algoritmo K-centroides es menos sensible a los ruidos y valores atípicos, comparado con el agrupamiento K-medias, porque usa centroides en lugar de medias.

El método más común en el agrupamiento K-centroides es el algoritmo **PAM** (Particiones alrededor de centroides, Kaufmann & Rousseeuw, 1990).

El uso de las medias implica que el agrupamiento K-medias es altamente sensible a los valores atípicos. Esto puede afectar levemente las asignaciones de las observaciones al cluster o al grupo. Un algoritmo más robusto es proporcionado por el algoritmo **PAM**.

Capítulo 2

Series de Tiempo y Google AdWords.

En éste capítulo desarrollaremos aspectos teóricos de Series de Tiempo y Google AdWords, donde haremos mención de definiciones, critérios, características, modelos de regresión y modelos ARIMA para las series de tiempo y también mostraremos algunas terminologías del Sistema AdWords como: palabras claves, clics, impresiones,..., entre otras.

1. Series de Tiempo.

El análisis de los datos experimentales que se han observado en diferentes puntos en el tiempo conduce a problemas nuevos y únicos en la modelización e inferencia estadística. La correlación introducida por el muestreo de puntos adyacentes en el tiempo puede restringir severamente la aplicabilidad de los muchos métodos estadísticos convencionales que tradicionalmente dependen de la suposición de que estas observaciones adyacentes son independientes e idénticamente distribuidas. El enfoque sistemático por el cual se trata de responder a las preguntas matemáticas y estadísticas planteadas por estas correlaciones se conoce comúnmente como análisis de series de tiempo.

El impácto del análisis de series de tiempo en las aplicaciones científicas puede documentarse parcialmente produciendo una lista abreviada de los diversos campos en los que pueden surgir problemas importantes de series de tiempo. Por ejemplo, muchas series de tiempo ocurren en el campo de la economía, donde están continuamente expuestos a cotizaciones bursátiles diarias o cifras mensuales de desempleo. Los científicos sociales siguen series poblacionales, como tasa de natalidad o matrícula escolar.

1.1. Características de una Serie de Tiempo.

Definición 2.1. Una *serie de tiempo* es un conjunto de observaciones x_t, cada una registrada a un tiempo específico t.

Definición 2.2. Un *modelo de series de tiempo* para los datos observados $\{x_t\}$ es una especificación de una distribución conjunta (o posiblemente solo de medias y covarianzas) de una sucesión de variables aleatorias $\{X_t\}$ de las cuales $\{x_t\}$ es una realización.

El análisis clásico de las series de tiempo se basa en la suposición de que los valores que toma la variable de observación es la consecuencia de tres componentes, cuya actuación conjunta da como resultados los valores medidos, estos componentes son:

- *Componente tendencia.* Se puede definir como un cambio a largo plazo que se produce en la relación al nivel medio, o el cambio a largo plazo de la media. La tendencia se identifica con un movimiento suave de la serie a largo plazo.
- *Componente estacional.* Muchas series de tiempo presentan cierta periodicidad o dicho de otro modo, variación de cierto período (semestral, mensual,etc.). Por ejemplo las Ventas al Detalle en Puerto Rico aumentan por los meses de noviembre y diciembre por las festividades navideñas. Estos efectos son fáciles de entender y se pueden medir explícitamente o incluso se pueden eliminar de la serie de datos, a este proceso se le llama desestacionalización de la serie.
- *Componente aleatoria.* Esta componente no responde a ningún patrón de comportamiento, sino que es el resultado de factores fortuitos o aleatorios que inciden de forma aislada en una serie de tiempo.

De las tres componentes anteriores las dos primeras son componentes determinísticas, mientras que la última es aleatoria. Así, se puede denotar la serie de tiempo como,

$$X_t = T_t + E_t + \epsilon_t$$

donde T_t es la tendencia, E_t es la componente estacional y ϵ_t es la componente aleatoria.

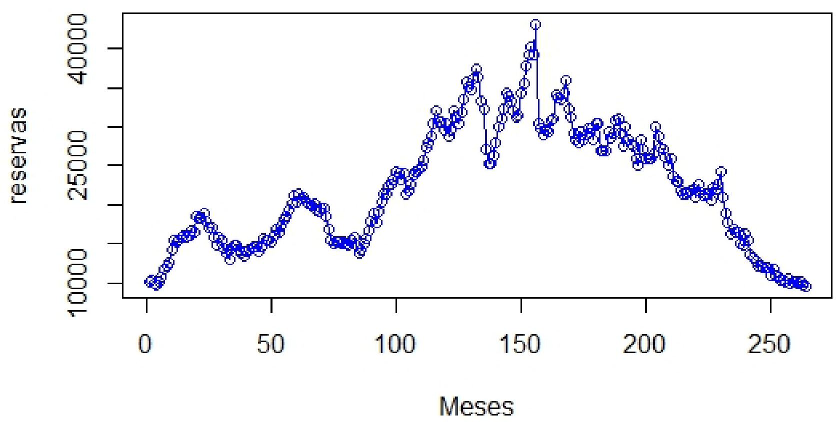

FIGURA 2.1. Serie de tiempo para las Reservas Internacionales de Enero de 1996 hasta Diciembre de 2017.

Modelos Estadísticos para Series de Tiempo.

El objetivo primario en el análisis de series de tiempo es desarrollar modelos matemáticos que provean una descripción apropiada para los datos muestrales.

DEFINICIÓN 2.3. Un *proceso estocástico* es una familia de variables aleatorias indexadas $x(w,t)$ ó $x_t(w)$ donde t pertenece a un conjunto de índices T y w pertenece a un espacio muestral Ω. Si $t = t^*$ fijo, $x(w,t^*)$ es una variable aleatoria. Si $w = w^*$ fijo, $x(w^*,t)$ es una función de t, y se llama una realización del proceso. Una **serie de tiempo** es la realización de un proceso estocástico.

Un ejemplo muy importante para la realización de un proceso estocástico es el **Ruido Blanco.**

Ruido Blanco.

Una manera sencilla de generar series de tiempo puede ser considerando una sucesión de variables aleatorias no-correlacionadas, w_t con media 0 y varianza σ_w^2. Las series de tiempo generadas de esta manera son usadas como modelos para ruido en aplicaciones para Ingenería, donde ellas son llamadas **ruido blanco**, denotaremos este proceso como $w_t \sim wn(0, \sigma_w^2)$.

También se requerirá que el ruido sea una colección de variables aleatorias independientes e idénticamente distribuidas con media 0 y varianza σ_w^2. Distinguiremos este caso diciendo que es ruido blanco independiente o escribiendo $w_t \sim iid(0, \sigma_w^2)$. Un muy usado ruido blanco es el ruido blanco gaussiano, donde w_t son variables aleatorias normales con media 0 y varianza σ_w^2 e identificadas como $w_t \sim iidN(0, \sigma_w^2)$.

Series de Tiempo Estacionarias.

- **Estacionarias.** Una serie es estacionaria cuando es estable a lo largo del tiempo, es decir, cuando la media y la varianza son constantes en el tiempo. Esto se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo.
- **No estacionaria.** Son series en la cuales la tendencia y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o a decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.

1.2. Regresión de Series de Tiempo.

El modelo lineal y sus aplicaciones son al menos tan dominantes en el contexto de series de tiempo como en la estadística clásica. Los modelos de regresión son importantes para modelos en el dominio de tiempo y frecuencia que discutiremos en la siguiente sección. La idea principal depende de poder expresar una serie respuesta X_t como una combinación lineal de entradas $z_{t_1}, z_{t_2}, ..., z_{t_q}$. La estimación de los coeficientes $\beta_1, \beta_2, ..., \beta_q$ de la combinación por mínimos cuadrados proporciona un método para modelar X_t en términos de las entradas.

Una regresión clásica se puede expresar como:

$$X_t = \beta_1 z_{t_1} + \beta_2 z_{t_2} + ... + \beta_q z_{t_q} + w_t,\ t = 1, 2, ..., n.$$

Influenciada por una colección de series independientes $z_{t_1}, z_{t_2}, ..., z_{t_q}$ fijas y conocidas, donde $\beta_1, \beta_2, ..., \beta_q$ son los coeficientes de regresión fijos y desconocidos, $\{w_t\}$ es un error aleatorio o un proceso de ruido consistente de variables normales independientes e idénticamente distribuidas con media cero y varianza σ_w^2.

Suponga que consideramos un modelo de regresión con k coeficientes y denotemos el estimador de máxima verosimilitud para la varianza como

$$\hat{\sigma_k^2} = \frac{RSS_k}{n} \tag{2.1}$$

donde RSS_k denota la suma residual de cuadrados bajo el modelo con k coeficientes de regresión.

Definición 2.4. **Criterio de Información de Akaike (AIC).**
El *Criterio de Información de Akaike* se define como:

$$AIC = \ln \hat{\sigma_k^2} + \frac{n+2k}{n}$$

donde $\hat{\sigma_k^2}$ está dado por (2.1) y k es el número de parámetros en el modelo.

El **criterio de información de Akaike** (AIC) es una medida de la calidad relativa de un modelo estadístico, para un conjunto de datos. Como tal, el AIC proporciona un medio para la selección del modelo. AIC se basa en la entropía de información: se ofrece una estimación relativa de la información perdida cuando se utiliza un modelo determinado para representar el proceso que genera los datos. AIC no proporciona una prueba de un modelo en el sentido de probar una hipótesis nula, es decir AIC puede decir nada acerca de la calidad del modelo en un sentido absoluto. Si todos los modelos candidatos encajan mal, AIC no dará ningún aviso de ello.

El valor de k que minimiza AIC especifica el mejor modelo. La idea es que la minimización $\hat{\sigma_k^2}$ sea razonablemente objetiva.

En el criterio AIC, el sesgo es aproximado por el número de parámetros los cuales son constantes y no tienen variabilidad. Para el modelo de regresión, la correción del sesgo del logaritmo de la verosimilitud se define como

Definición 2.5. **AIC con sesgo corregido (AICc).**

$$AICc = \ln \hat{\sigma_k^2} + \frac{n+k}{n-k-2}$$

donde $\hat{\sigma_k^2}$ está dado por (2.1), k es el número de parámetros en el modelo y n es el tamaño de la muestra.

De manera similar que en el AIC, se selecciona el modelo con el menor valor de AICc.

Definición 2.6. **Criterio de Información de Schwarz (SIC).**

$$SIC = \ln \hat{\sigma_k^2} + \frac{k \ln n}{n}.$$

SIC es también llamado Criterio de Información Bayesiano (BIC) es un criterio para la selección de modelos entre un conjunto finito de modelos. Se basa, en parte, en la función de probabilidad y está estrechamente relacionado con el Criterio de Información Akaike (AIC).

1.3. Modelos ARIMA.

El modelo de regresión clásico fue desarrollado para el caso estático, es decir, solo permitimos que la variable dependiente sea influenciada por los valores actuales de las variables independientes. En el caso de series de tiempo, es deseable permitir que la variable dependiente sea influenciada por los valores pasados de las variables independientes y posiblemente por su propio valor pasado. Si el presente se puede modelar de manera plausible en términos de solo los valores pasados de las entradas independientes, tenemos la posibilidad atractiva de que la previsión sea posible.

Los modelos autoregresivos se basan en la idea de que el valor actual de la serie, x_t, se puede explicar como una función de p valores pasados, $x_{t-1}, x_{t-2}, ..., x_{t-p}$, donde p determina el número de pasos en el pasado necesarios para pronosticar el valor actual.

Definición 2.7. Un **modelo autoregresivo de orden** p, abreviado $\mathbf{AR}(p)$, es de la forma,

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + w_t \tag{2.2}$$

donde x_t es estacionario y $\phi_1, \phi_2, ..., \phi_p$ son constantes ($\phi_p \neq 0$). A menos que se declare lo contrario, se asume que w_t es un ruido blanco gaussiano de media cero y varianza σ_w^2. Si la media μ de x_t no es cero, reemplazamos x_t por $x_t - \mu$ en (2.2), es decir,

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + ... + \phi_p(x_{t-p} - \mu) + w_t$$

o escribimos

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + w_t,$$

donde $\alpha = \mu(1 - \phi_1 - \phi_2 - ... - \phi_p)$.

Una forma útil de escribir el modelo $\mathbf{AR}(p)$ es usando el **operador backshift**,

$$(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)x_t = w_t,$$

de forma simplificada

$$\phi(B)x_t = w_t$$

donde $\phi(B)$ es el **operador autoregresivo**.

Como una alternativa a la representación autoregresiva en la cual x_t del lado izquierdo de la ecuación se asume como una combinación lineal, en los modelos de promedio móvil de orden q abreviados $\mathbf{MA}(q)$ asumimos el ruido blanco del lado derecho de la ecuación que los define como una combinación lineal de los datos observados.

Definición 2.8. El **modelo de promedio móvil de orden** q o modelo $\mathbf{MA}(q)$ se define como,

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + ... + \theta_p w_{t-p} \tag{2.3}$$

donde hay q pasos en el promedio móvil y $\theta_1, \theta_2, ..., \theta_p$ ($\theta_p \neq 0$) son parámetros. El ruido w_t se asume como un ruido blanco gaussiano.

Usando el **operador backshift** podemos escribir el modelo $\mathbf{MA}(q)$ como sigue,

$$(1 + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q)w_t = x_t,$$

de forma simplificada

$$\theta(B)w_t = x_t,$$

donde $\theta(B)$ es el **operador de promedio móvil.**

Ahora procederemos con un desarrollo más general de modelos autoregresivos, de promedio móvil y mezcla de ambos modelos para series de tiempo estacionarias.

Definición 2.9. Una serie de tiempo $\{x_t; t = 0, \pm 1, \pm 2, ...\}$ es un **proceso autoregresivo de promedio móvil**, denotado **ARMA**(p, q), si es estacionario y

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + ... + \theta_p w_{t-p} \tag{2.4}$$

con $\phi_p \neq 0$, $\theta_p \neq 0$ y $\sigma_w^2 > 0$. Los parámetros p y q son llamados ordenes autoregresivos y de promedio móvil respectivamente. Si la media μ de x_t no es cero, hacemos $\alpha = \mu(1 - \phi_1 - \phi_2 - ... - \phi_p)$ y escribimos el modelo como

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + ... + \theta_p w_{t-p}.$$

A menos que se declare lo contrario, $\{w_t; t = 0, \pm 1, \pm 2, ...\}$ es una sucesión de ruido blanco gaussiano.

El modelo **ARMA**(p, q) también se puede escribir de la siguiente forma,

$$\phi(B) x_t = \theta(B) w_t.$$

Función de Autocorrelación (ACF) y Función de Autocorrelación Parcial (PACF).

La función de autocorrelación parcial, al igual que la función de autocorrelación, transmiten información vital con respecto a la estructura de dependencia de un proceso estacionario.

Para definir la función de autocorrelación (ACF), definiremos la función de autocovarianza para una serie estacionaria,

Definición 2.10. La **función de autocovarianza de una serie de tiempo estacionaria** se escribirá como:

$$\gamma(h) = E\left[(x_{t+h} - \mu)(x_t - \mu)\right] \tag{2.5}$$

donde h representa el tiempo de traslación o salto.

Definición 2.11. La **función de autocorrelación (ACF)** de una serie de tiempo estacionaria será descrita, usando (2.5), como

$$\rho(h) = \frac{\gamma(t+h, t)}{\sqrt{\gamma(t+h, t+h)\gamma(t, t)}} = \frac{\gamma(h)}{\gamma(0)}.$$

La ACF mide la predictibilidad lineal de una serie de tiempo en tiempo t. Se tiene que $-1 \leq \rho(h) \leq 1$ para todo h.

Definición 2.12. La **función de autocorrelación parcial (PACF)** de un proceso estacionario x_t, denotada ϕ_{hh}, para $h = 1, 2, ...$ es

$$\phi_{11} = corr(x_1, x_0) = \rho(1)$$

y

$$\phi_{hh} = corr(x_h - x_h^{h-1}, x_0 - x_0^{h-1}), h \geq 2.$$

Tanto $(x_h - x_h^{h-1})$ como $(x_0 - x_0^{h-1})$ son no correlacionados con $\{x_1, x_2, ..., x_{h-1}\}$. Por estacionaridad, la PACF ϕ_{hh} es la correlación entre x_t y x_{t-h} con la dependencia lineal $\{x_{t-1}, ..., x_{t-(h-1)}\}$ removida en cada uno.

Modelos Integrados para Datos no estacionarios.

En muchas situaciones, se puede pensar que las series de tiempo están compuestas por dos componentes, un componente de tendencia no estacionario y un componente estacionario de media cero. Por ejemplo, consideremos el siguiente modelo

$$x_t = \mu_t + y_t,$$

donde $\mu_t = \beta_0 + \beta_1 t$ y y_t es estacionario. Diferenciar tal proceso conducirá a un proceso estacionario:

$$\Delta x_t = x_t - x_{t-1} = \beta_1 + y_t - y_{t-1} = \beta_1 + \Delta y_t.$$

El integrado **ARMA**, o modelo **ARIMA**, es una ampliación de una clase de modelos **ARMA** para incluir la diferenciación.

Definición 2.13. Un proceso x_t se dice que es **ARIMA**(p, d, q) si

$$\Delta^d x_t = (1 - B)^d x_t$$

es **ARMA**(p, q). En general, escribiremos el modelo como

$$\phi(B)(1 - B)^d x_t = \theta(B) w_t. \tag{2.6}$$

Si $E(\Delta^d x_t) = \mu$, escribimos el modelo como

$$\phi(B)(1-B)^d x_t = \delta + \theta(B) w_t,$$

donde $\delta = \mu(1 - \phi_1 - \phi_2 - ... - \phi_p)$.

Modelos Multiplicativos ARIMA Estacionales.

En esta sección presentaremos modificaciones hechas al modelo **ARIMA** para dar cuenta del comportamiento estacional y no estacionario. A menudo, la dependencia en el pasado tiende a ocurrir fuertemente en múltiplos de algunos subyacentes retrasos estacionales. Por ejemplo, en datos económicos mensuales, hay una fuerte componente anual que ocurre en rezagos que son múltiplos de $s = 12$, esto debido a las fuertes conexiones de las actividades del año. Por ésta razón, es apropiado introducir polinomios autoregresivos y de promedio móvil que se identifican con los retrasos estacionales. Resultando un modelo estacional autoregresivo de promedio móvil, llamado **ARMA**$(P, Q)_s$, que toma la siguiente forma

$$\Phi_P(B^s) x_t = \Theta_Q(B^s) w_t.$$

donde $\Phi_P(B^s)$ y $\Theta_Q(B^s)$ son los operadores autoregresivo estacional y promedio móvil estacional de orden P y Q, respectivamente, con período estacional s.

En general, podemos combinar los operadores estacionales y no estacionales dentro de un modelo multiplicativo estacional autoregresivo de promedio móvil, denotado por **ARMA**$(p, q) \times (P, Q)_s$, y se escribe

$$\Phi_P(B^s)\phi(B) x_t = \Theta_Q(B^s)\theta(B) w_t.$$

Definición 2.14. Una **diferencia estacional** de orden D se define como

$$\Delta_s^D x_t = (1 - B^s)^D x_t,$$

donde $D = 1, 2, ...$ toma valores enteros positivos.

Introduciendo éstas ideas dentro de un modelo general, tenemos la siguiente definición.

Definición 2.15. El **modelo multiplicativo estacional integrado autoregresivo de promedio móvil**, o modelo **SARIMA**, es dado por,

$$\Phi_P(B^s)\phi(B)\Delta_s^D \Delta^d x_t = \alpha + \Theta_Q(B^s)\theta(B) w_t,$$

donde w_t es un proceso usual de ruido blanco Gaussiano. El modelo general es denotado por $\mathbf{ARIMA}(p, d, q) \times (P, D, Q)_s$. Los componentes autoregresivos y de promedio móvil son representados por $\phi(B)$ y $\theta(B)$ de orden p y q, respectivamente, y las componentes estacionales autoregresivas y de promedio móvil por $\Phi_P(B^s)$ y $\Theta_Q(B^s)$ de orden P y Q y las componentes de diferencias estacionales por $\Delta^d = (1 - B)^d$ y $\Delta_s^D = (1 - B^s)^D$.

2. Google AdWords.

Google AdWords es un servicio y programa de la empresa de **Google** que se utiliza para ofrecer publicidad patrocinada a potenciales anunciantes. A través de **AdWords**, se pueden crear anuncios en línea para llegar a los usuarios en el momento exacto en que se interesen por los productos y servicios que se ofrecen. **Google AdWords** es un producto que puede utilizarse para promocionar empresas, vender productos o servicios, dar a conocer y aumentar el tráfico de los sitios web.

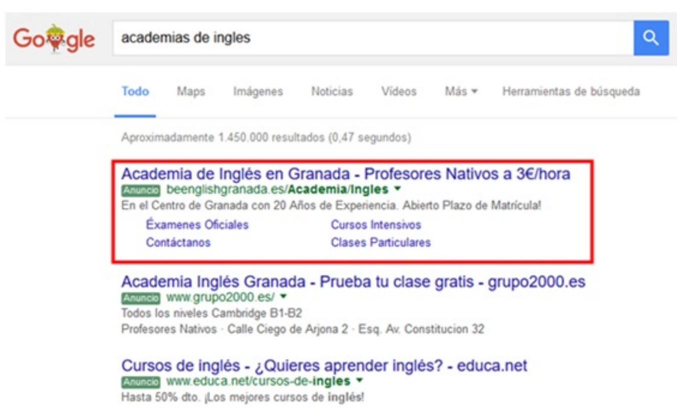

FIGURA 2.2. Anuncios para la búsqueda de Academias en Inglés.

2.1. Terminología Básica de AdWords.

Algunas terminologías son,

- **Palabras Clave:** Son palabras o frases que describen su producto o servicio y que elige para ayudar a determinar cuándo y dónde puede aparecer su anuncio. Cuando alguién realiza una búsqueda en Google, su anuncio podría aparecer en función de la similitud de sus palabras clave con los términos de búsqueda de la persona, así como función de los tipos de concordancia de las palabras clave. Contar con una lista eficaz de palabras clave puede ayudarlo a mejorar el rendimiento de sus anuncios. Las palabras clave de mala calidad pueden dar como resultado costos más altos y anuncios con posiciones bajas.
- **Clic:** Cuando alguién hace clic en su anuncio. Los clics le permiten saber si su anuncio es atráctivo para las personas que lo ven. Los anuncios relevantes y muy orientados tienen más probabilidades de recibir clics.
- **Impresiones:** Se refiere a la frecuencia con la que se muestra su anuncio. Una impresión se cuenta cada vez que su anuncio aparece en una página de resultados de la búsqueda.
- **Porcentajes de Clics (CTR):** Porcentaje que indica la frecuencia con la que los usuarios que ven tu anuncio acaban haciendo clic en él. El porcentaje de clics (CTR) se puede usar para medir el rendimiento de las palabras claves y los anuncios. El CTR se calcula dividiendo el número de clics que recibe tu anuncio entre el número de veces que se muestra:

$$\frac{Clics}{Impresiones} = CTR.$$

- **Nivel de Calidad:** Es una estimación de la calidad de los anuncios, las palabras clave y la páginas de destino. Los anuncios de mayor calidad pueden significar precios más bajos y mejores posiciones del anuncio.
- **Costo por clic (CPC):** Significa que paga por cada clic que se hace en sus anuncios.
- **Costo promedio por clic:** El importe promedio que se le cobró por un clic en su anuncio.
- **Oferta de CPC máximo:** Es una oferta que establece el importe más alto que está dispuesto a pagar un clic en su anuncio. Si un usuario hace clic en su anuncio, ese

clic no le costará más que la oferta costo máximo por clic. Por lo general, una oferta más alta permite que su anuncio aparezca en una posición superior en la página.

- **Campaña:** Es un grupo de anuncios que comparten un presupuesto, una orientación geográfica y otros parámetros de configuración. A menudo, las campañas se usan para organizar las categorias de productos o servicios que ofrece.
- **Conversión:** Una acción que se cuenta cuando una persona interactúa con su anuncio y, después, lleva a cabo una acción que usted definió como valiosa para su empresa, como una compra en línea o una llamada a su empresa desde un teléfono celular.
- **Posición promedio:** Una estadística que describe la clasificación habitual de su anuncio en comparación con otros anuncios. Esta clasificación determina el orden de aparición de los anuncios en la página.

Capítulo 3

Modelos ARIMA para la agrupación de palabras claves.

1. Conjunto de datos.

El conjunto de datos fué extraido de una campaña publicitaria de una compañía de ventas de sombreros por internet, que consta de 7155 palabras claves y 12 variables usadas en el Sistema Adwords, filtradas por hora en un período de dos meses. Septiembre de 2017 hasta Octubre de 2017 específicamente.

Antes de agrupar las palabras claves del conjunto de datos, aplicaremos el análisis de componentes principales (ACP) que permite escoger las variables menos correlacionadas para reducir la dimensionalidad de los datos y así poder facilitar la agrupación.

2. ACP.

Cómo se describió anteriormente el análisis de componentes principales se considera una técnica para transformar un conjunto de variables observadas en un nuevo conjunto de variables que no están correlacionadas entre sí. La variación en las q variables originales solo se explica por completo por todas las componentes principales. La utilidad de éstas variables transformadas, sin embargo, se deriva de su propiedad de tener en cuenta la varianza en proporciones decrecientes. El análisis de componentes principales proporciona un gran resumen de los datos (reduciendo la dimensionalidad de los datos) que pueden ser útil en análisis posteriores. Ahora, la pregunta que nos debemos hacer es ¿cuántas componentes principales se necesitan para proporcionar un resumen adecuado de un conjunto de datos determinado? Hay muchas técnicas para escoger el número de componentes adecuado, una de ellas es el **diagrama de pantalla** ó *scree diagram* en Inglés.

El **diagrama de pantalla** sugiere la gráfica de las varianzas de cada una de las componentes principales. Cada una de éstas varianzas corresponden a un codo, es decir, un cambio en la pendiente desde la superficie. Se recomienda escoger el número de componentes principales si éstas tienen varianzas con pendientes muy inclinadas en el diagrama de pantalla, ya que que ésto representa una mayor variabilidad.

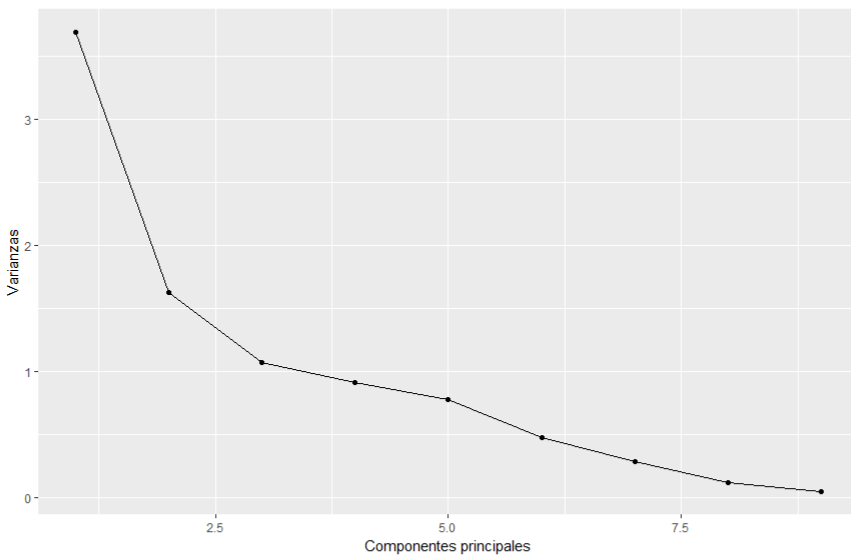

FIGURA 3.1. Varianzas de las componentes principales.

En la gráfica anterior se observa las varianzas de las componentes principales, donde las primeras dos componentes tienen las pendientes más inclinadas, por lo que la proporción acumulativa de la varianza de las dos primeras componentes es 0.5907 que es el 59 % de información de los datos. Las variables que aportan más información en la dos primeras componentes principales, son:

Variables	Pesos
costo	0.4941
clic	0.4844
oferta cpc-maximo	0.5070
impresiones	0.4513
calidad del anuncio	-0.5013

La proyección de los datos en las dos primeras componentes principales se muestra en la siguiente gráfica.

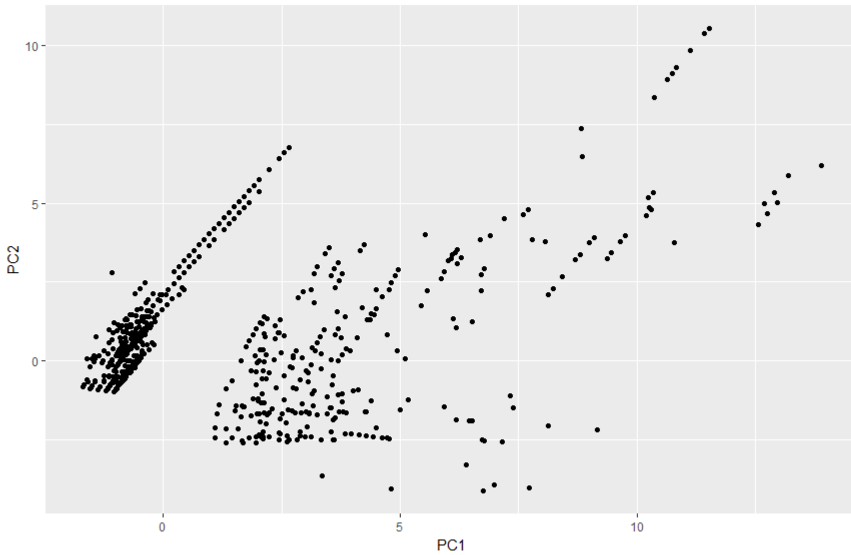

FIGURA 3.2. Proyección de los datos a través de las dos primeras componentes principales.

3. Agrupación con K-centroides.

Haciendo uso de las variables que aportan más información en las dos primeras componentes y usando el agrupamiento K-centroides o PAM tenemos la siguiente agrupación,

Agrupación con las variables costo, clic, oferta cpc-máximo e impresiones.

La mejor agrupación se obtuvo al considerar tres grupos. La segmentación de las palabras claves en tres grupos con respecto a las dos primeras componentes principales de los datos dimensionalmente reducidos a éstas cuatro variables es la siguiente:

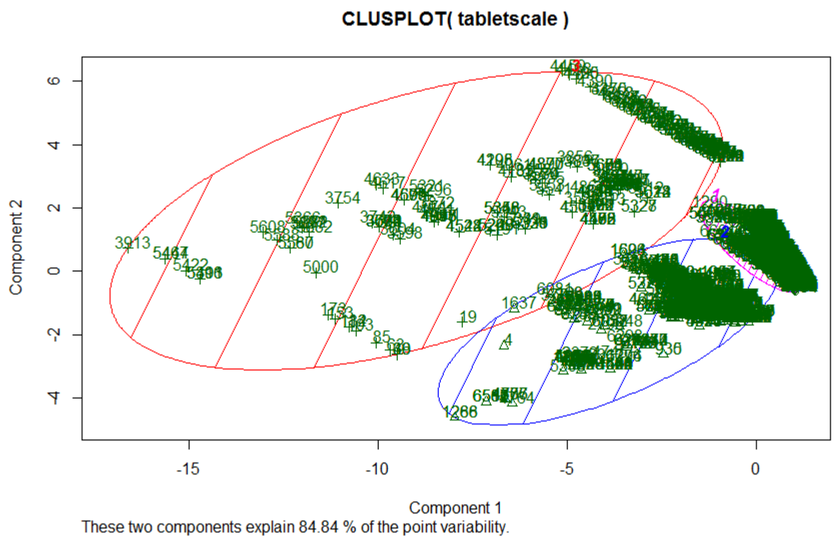

Figura 3.3. Agrupación de grupos de tres de las palabras claves usando **PAM**.

La proyección de los datos en grupos de tres mediante las componentes principales se puede apreciar en la siguiente figura,

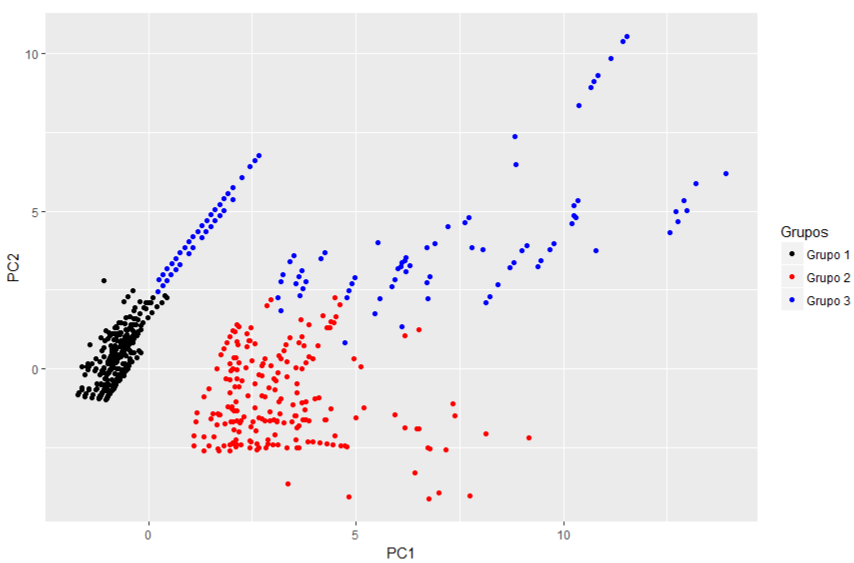

Figura 3.4. Proyección de los datos en grupos de tres.

La distribución de las palabras claves en cada grupo es la siguiente:

Grupo 1	Grupo 2	Grupo 3
5736	1130	289

Tanto en la Figura 3.3 como en la Figura 3.4 se puede apreciar como las palabras se proyectan en tres grupos bien diferenciados. Y ésta apreciación la confirmaremos en la siguiente sección.

Validación.

La validación del cluster consiste en medir la bondad de los resultados de la agrupación. Uno de los métodos para validación de una agrupación es el **Método de la silueta**.

El análisis de la silueta mide qué tan bien se agrupa una observación y estima la distancia promedio entre los grupos.

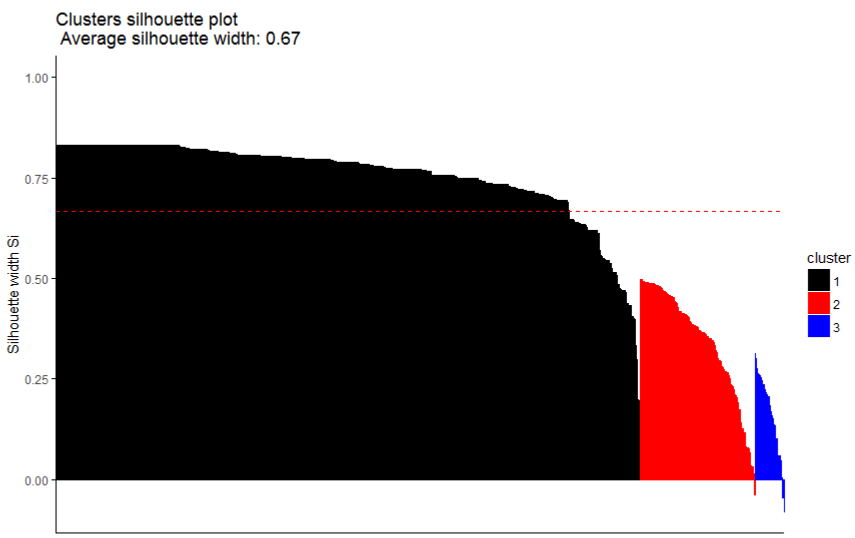

FIGURA 3.5. Gráfica de la silueta para el agrupamiento de tres grupos.

De acuerdo a la gráfica de la Silueta en la Figura 3.5, el agrupamiento en tres grupos tiene un buena estructura con un promedio de 0.67 de calidad de agrupamiento, esto quiere decir, que un 67 % de las observaciones en promedio encajan en los grupos al que han sido asignadas. Los promedios de la silueta por cada grupo son los siguientes:

Grupo	Tamaño	Promedio
1	5736	0.76
2	1130	0.34
3	289	0.16

Con ésta información ahora procederemos a ver las características de cada grupo para con ello explorar cuáles son las variables que los hacen distinguibles entre sí.

3.1. Grupo 1.

Este grupo consta de 5736 palabras claves y la palabra con mayor frecuencia se observó 328 veces. El siguiente resumen muestra algunas características del grupo de palabras,

	oferta cpc-max	impresiones	clics	ctr	costo	prom-cpc	pos-prom	calidad	conversion
Mín	0.18	1	0	0	0	0	1	0	0
Mediana	1	1	0	0	0	0	1	8	0
Media	0.99	1.6	0	0	0	0	1	8.6	0
Máx	2.8	12	0	0	0	0	5	10	0

Una característica importante para este grupo es que no hubo clic y por esto tampoco ctr, costo, promedio-cpc o conversiones. Podemos también destacar que la media de las impresiones es de 1.6 con un máximo de 12, la media de la posición promedio es 1 y el promedio de la calidad del anuncio es de 8.6, por lo que éste grupo de palabras no tiene clics pero si tiene buenas posiciones y buena calidad de los anuncios.

Para este grupo las variables más correlacionadas son la oferta cpc-máximo y la posición promedio con -0.31 de relación.

3.2. Grupo 2.

Este grupo consta de 1130 palabras claves y la palabra con mayor frecuencia se observó 159 veces.

	oferta cpc-max	impresiones	clics	ctr	costo	prom-cpc	pos-prom	calidad	conversion
Mín	0.19	1	1	6.7	0.01	0.01	1	6	0
Mediana	1	1	1	100	0.75	0.64	1	10	0
Media	1.07	2.6	1.13	75.8	0.82	0.7	1.13	9	0
Máx	2.2	17	3	100	3.9	1.8	3.5	10	0

A diferencia del grupo anterior éste grupo tuvo al menos un clic. Además, se caracteriza por no tener conversión. Una característica importante de recalcar es que la media de las impresiones aumentaron de 1.6 a 2.6 con respecto al grupo anterior y tiene un máximo de 17 impresiones. También vale la pena mencionar que el grupo tiene en promedio 75.8 % de ctr, es decir, la cantidad de clics que obtuvo cada palabra representa el 75.8 % en promedio de la cantidad de impresiones que obtuvieron éstas palabras. Al igual que el grupo anterior éste grupo tiene buenas posiciones y buena calidad de los anuncios.

Las variables más correlacionadas son el costo con el promedio-cpc con una relación de 0.8, seguida de impresiones y ctr con -0.7 de relación.

3.3. Grupo 3.

Este grupo consta de 289 registros los cuales corresponden a una sola palabra.

	oferta cpc-max	impresiones	clics	ctr	costo	prom-cpc	pos-prom	calidad	conversion
Mín	1.3	1	0	0	0	0	1	7	0
Mediana	5	7	1	8.3	0.55	0.55	1	7	0
Media	4.9	9.4	0.94	10.50	1	0.59	1	7.3	0.07
Máx	5	33	4	100	4.8	2.38	1.1	8	1

Con respecto al grupo 2 éste grupo tuvo menos clics, con una media de 0.94, sin embargo, obtuvo más impresiones, con una media de 9.4, por lo que le da motivo de tener un 10.5 % de ctr.

A diferencia de los grupos anteriores en este grupo hubo conversiones, en total 22. Ésto puede ser debido a que se tiene una oferta cpc-máximo con promedio de 4.9, el cual es superior al de los dos grupos anteriores y por tener más impresiones.

Las variables más correlacionadas son oferta cpc-máximo y posición promedio con una relación de -1, seguida de clics con costo con una relación de -0.81.

Cómo podemos notar se consiguió segmentar el conjunto de palabras claves en tres grupos claramente diferenciados por los clics, las impresiones y las conversiones. Vale la pena destacar que si bien las variables clic e impresiones se encontraban entre las variables seleccionadas por el método ACP la variable conversión no.

Para caracterizar un poco mas estos grupos de palabras, en la siguiente sección estudiaremos el comportamiento de las impresiones de dos de estos grupos.

4. Modelos ARIMA.

En ésta sección trabajaremos con las impresiones de dos de los grupos más importantes, considerados así dada las características de las palabras claves que los conforman, las cuales fueron expuestas en la sección precedente. Las impresiones indican la frecuencia con la que se muestra un anuncio y pueden ser particularmente importantes en las campañas de desarrollo de la marca, ya que pueden ser un indicador de la cantidad de clientes que muestran un interés real en un anuncio sí éstos dan clic en el, de manera que pueden ayudar a evaluar qué tan atractivo es el anuncio para clientes potenciales. Vale la pena resaltar que según lo observado aquellas palabras que aparecen en los primeros lugares de la búsqueda, consiguen por lo general más impresiones.

Los grupos que vamos a trabajar son los grupos 1 y 3, por tener éstos características como impresiones de palabras sin obtener clic e impresiones de palabras que obtuvieron clic y además tuvieron conversión como es el caso del grupo 3.

4.1. Grupo 1.

Éste grupo consta de 5736 palabras claves con características ya mencionadas en la sección anterior. En éste contexto, se quiere ajustar un modelo ARIMA para las impresiones del grupo, de manera que se pueda hacer análisis a futuro. Para esto, se transformaron las impresiones del grupo en una serie de tiempo con una frecuencia de doce.

En la Figura 3.6 se ilustra la serie de tiempo de las impresiones de las palabras claves del grupo 1,

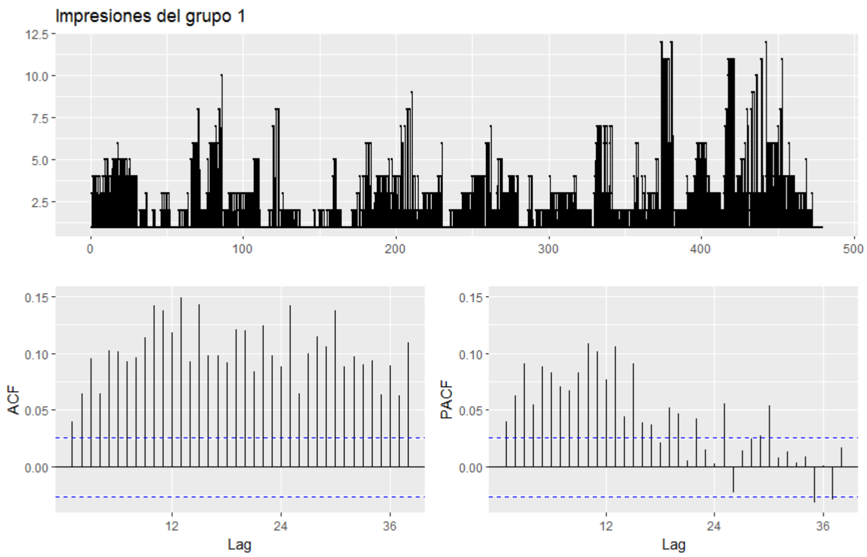

Figura 3.6. Serie de tiempo de las impresiones de las palabras claves del grupo 1 junto con las ACF y PACF.

Podemos observar en la Figura 3.6 que la tendencia y la estacionalidad no están marcadas, sin embargo ésto no significa que no tenga tendencia ni estacionalidad, también se observa en la gráfica de ACF que los valores decaen muy lentamente corroborando la no estacionaridad de la serie. Una de las técnicas para obtener una serie estacionaria es descomponiendo la serie en tres partes, que son: tendencia, estacionalidad y la componente aleatoria. En la Figura 3.7 se muestra la serie removida, es decir, la serie sin la componente de tendencia y estacionalidad.

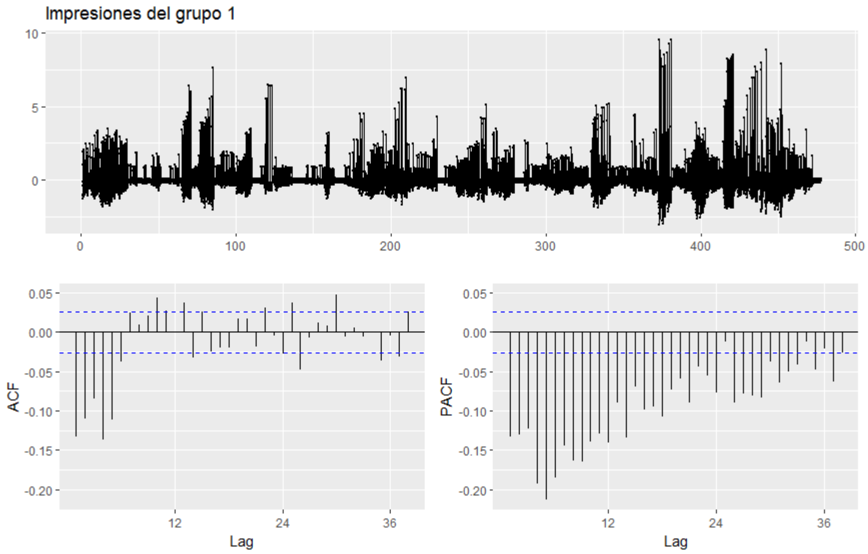

FIGURA 3.7. Serie de tiempo removida de las impresiones de las palabras claves del grupo 1 junto con las ACF y PACF.

Observando la ACF para la serie removida, tenemos que después de cinco pasos (Lags) los valores caen en la banda de confianza y en la PACF los valores decrecen lentamente.

Modelo.

Usando el comando *auto.arima()* en R, tenemos que el modelo que mejor se ajusta a los datos es el modelo $\textbf{ARIMA}(0,0,1)\times(0,0,1)_{12}$.

Los parámetros de éste modelo son:

$$\theta_1 = -0{,}1983, \Theta_1 = 0{,}0148.$$

Así, el modelo puede ser reescrito como,

$$x_t = \Theta_1(B^{12})\theta(B)w_t.$$

Donde, $\Theta_1(B^{12})$ y $\theta(B)$ son los operadores de promedio móvil estacional y de promedio móvil.

Validación.

Para la validación de éste modelo se realizó un análisis de residuos usando la función de autocorrelación (ACF). De acuerdo al análisis de la gráfica de la ACF, sí por lo menos un valor está fuera de la banda de confianza, el modelo posiblemente no es el adecuado.

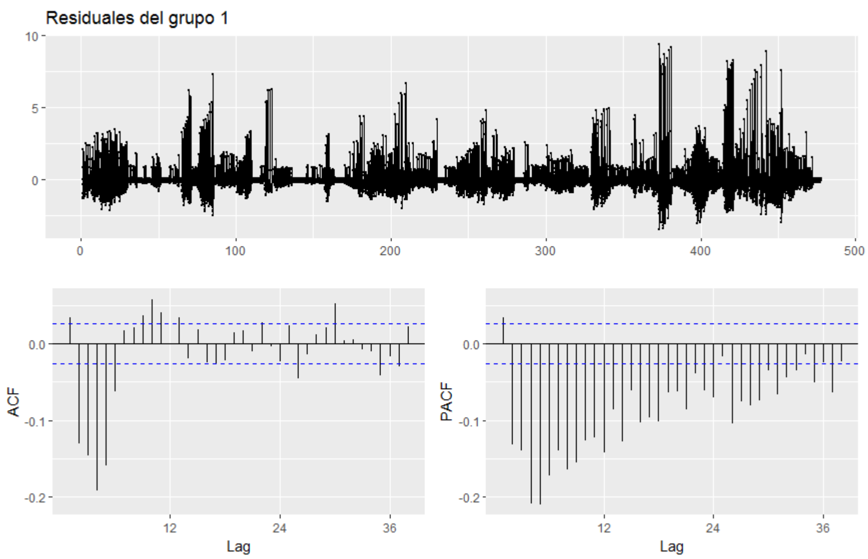

FIGURA 3.8. Residuales de la serie de tiempo de impresiones de las palabras claves del grupo 1 junto con las ACF y PACF del modelo **ARIMA**$(0, 0, 1) \times (0, 0, 1)_{12}$.

Podemos notar que algunos valores de la gráfica de la función de autocorrelación se salen de la banda de confianza.

SARIMA.

También podemos validar el modelo sugerido a parte de la función de autocorrelación con la gráfica de la probabilidad normal ó gráfica Q-Q normal y la prueba de hipótesis de Ljung-Box de los residuales.

La prueba de hipótesis **Ljung Box** es un tipo de prueba estadística que demuestra si un grupo cualquiera de autocorrelaciones de una serie de tiempo son diferentes de cero.

DEFINICIÓN 3.1. La prueba de Ljung-Box se puede definir de la siguiente manera:

$$H_o : \text{Los datos se distribuyen de forma independiente.}$$

$$H_a : \text{los datos no se distribuyen de forma independiente.}$$

El estadístico de prueba es:

$$Q = n(n+1)\sum_{k=1}^{h} \frac{\hat{p_k}^2}{n-k}$$

donde n es el tamaño de la muestra, $\hat{p_k}^2$ es la autocorrelación de la muestra en el retraso k y h es el número de retardos (Lag) que se están probando. Para un nivel de significancia α, la región de rechazo de la hipótesis es

$$Q > \chi^2_{1-\alpha,h}$$

ó

$$valor - p \leq \alpha$$

donde $\chi^2_{1-\alpha,h}$ es la α-cuantil de la distribución chi-cuadrado con m grados de libertad.

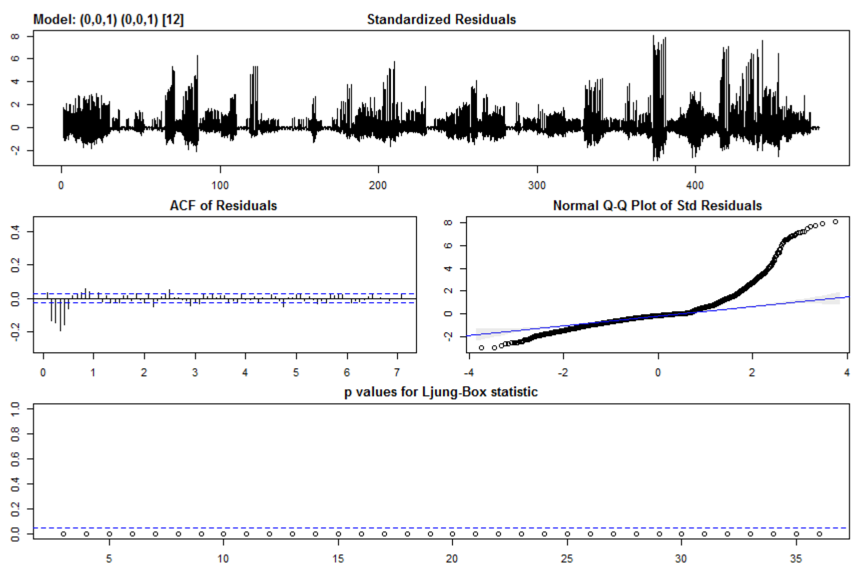

FIGURA 3.9. Residuales de la serie de tiempo de impresiones de las palabras claves del grupo 1 junto con la ACF, la gráfica Q-Q normal y la prueba Ljung-Box del modelo $\mathbf{ARIMA}(0,0,1) \times (0,0,1)_{12}$.

Podemos notar en la Figura 3.9, que los $valores - p \leq \alpha$, donde $\alpha = 0{,}05$, por lo que se rechaza la hipótesis nula, es decir, los datos no se distribuyen de forma independiente. Además, en la gráfica Q-Q se tiene una cola de valores desviados de los cuantiles de la normal, es decir, los estadísticos rechazan la hipótesis nula de similitud. Por lo que, éste modelo no es el adecuado para ajustarse a los datos.

Luego de hacer muchas pruebas para hallar un modelo que se ajuste a los datos, hemos logrado conseguir el modelo siguiente: **ARIMA**$(5,0,2) \times (0,0,2)_7$. Donde los parámetros son los siguientes:

$$\phi_1 = 1{,}3148, \phi_2 = -0{,}5060, \phi_3 = -0{,}0134, \phi_4 = -0{,}0789, \phi_5 = 0{,}0253,$$

$$\theta_1 = -1{,}8966, \theta_2 = 0{,}8966, \Theta_1 = 0{,}03780, \Theta_2 = -0{,}0299.$$

Así, el modelo puede ser reescrito como,

$$\phi(B)x_t = \Theta_2(B^7)\theta(B)w_t.$$

Donde, $\phi(B), \theta(B)$ y $\Theta_2(B^7)$ son los operadores autoregresivos, de promedio móvil y de promedio móvil estacional.

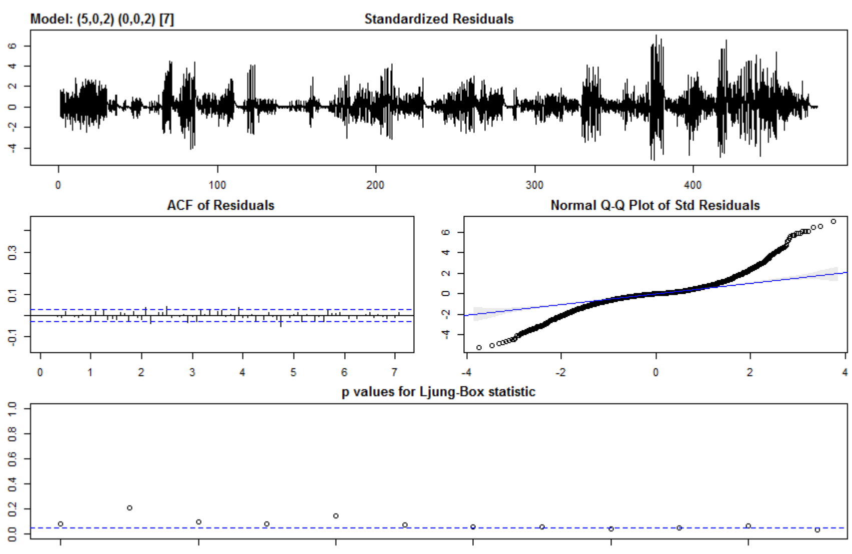

FIGURA 3.10. Residuales de la serie de tiempo de impresiones de las palabras claves del grupo 1 junto con las ACF, la gráfica Q-Q normal y la prueba Ljung-Box del modelo **ARIMA**$(5,0,2) \times (0,0,2)_7$.

Podemos observar en la gráfica de la prueba de Ljung-Box que la mayoría de los valores-p están por encima del nivel de significancia, esto quiere decir que no hay evidencias para rechazar la hipótesis nula, por lo que los datos estan distribuidos de forma independiente (es decir, las correlaciones en la población de la que se toma la muestra son cero). La mayoría de los valores de la función de autocorrelación están dentro de la banda de confianza. Sin embargo, en la gráfica Q-Q normal se tienen muchos valores desviados de los cuantiles de la normal.

Comparaciones.

El criterio para elegir uno de los dos modelos propuestos son los criterios AIC, AICc, y BIC, el cual es elegir aquel cuyo valor es mínimo.

Modelo	Varianza estimada	Criterio AIC	Criterio AICc	Criterio BIC
ARIMA$(0,0,1)\times(0,0,1)_{12}$	1.358	1.307046	1.307397	0.310532
ARIMA$(5,0,2)\times(0,0,2)_{7}$	0.8725	0.867123	0.867480	-0.121255

De aquí, el mejor modelo que se ajustó a los datos es el **ARIMA**$(5,0,2)\times(0,0,2)_7$.

Veamos ahora cómo se ajusta éste modelo a la serie original. Datos estimados por el modelo (gráfica azul) y Serie original de impresiones del grupo 1 (gráfica negra).

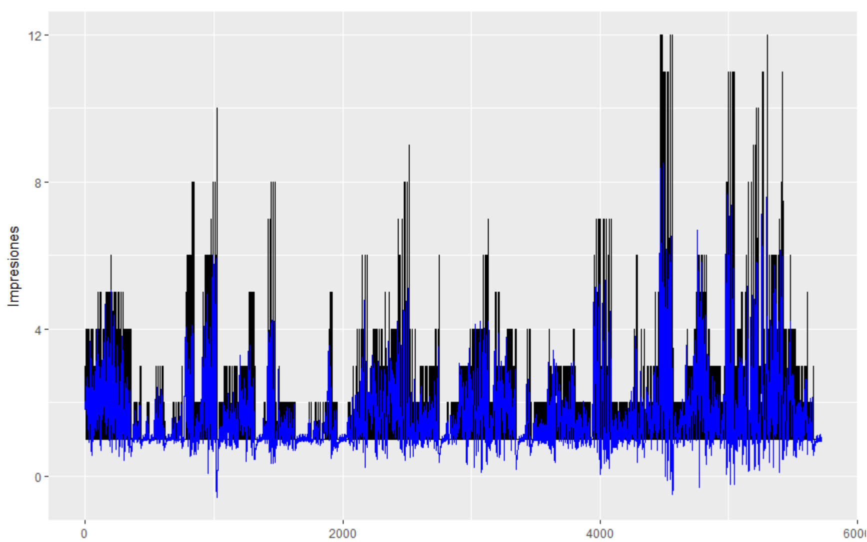

FIGURA 3.11. Serie de tiempo de impresiones de las palabras claves del grupo 1 junto con la estimacion del modelo **ARIMA**$(5,0,2)\times(0,0,2)_7$.

Podemos notar que el modelo se ajusta bien a los datos ya que mantiene el comportamiento de la serie. Éste resultado es de utilidad para el pronóstico de la serie de impresiones de dicho grupo.

Pronóstico.

Finalmente, al tener el modelo se llevó a cabo el pronóstico de las impresiones para los últimos 156 registros en el día doce del mes de octubre de 2017.

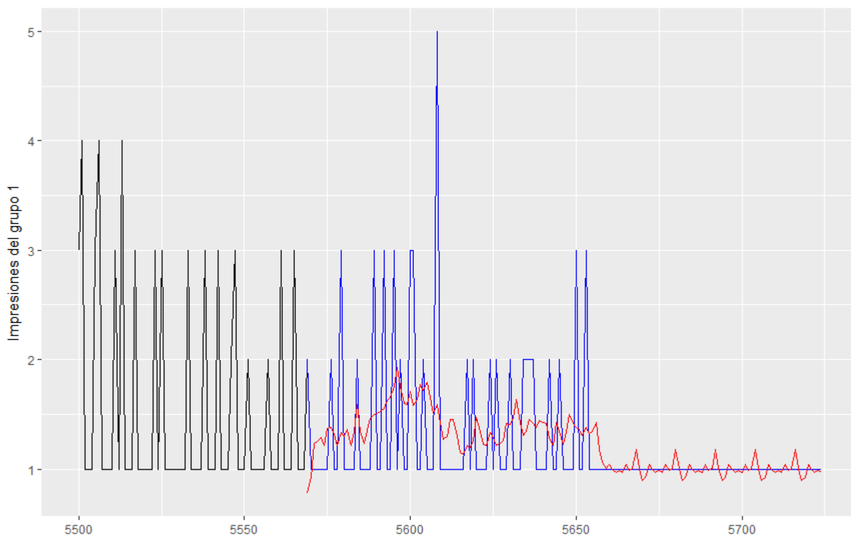

FIGURA 3.12. Pronóstico del modelo ajustado a la serie de tiempo de las impresiones para los últimos 156 registros. (Línea roja valor estimado y línea azul valor original.)

Se observa en la Figura 3.12 que el modelo predice una tendencia similar a la tendencia de los valores originales de la serie de las impresiones. Sin embargo, los valores pronosticados y los valores originales varían con respecto a las alturas, es decir, a la cantidad de impresiones.

En la tabla siguiente se muestra el valor real, el pronóstico y el error absoluto de algunos registros.

Período	Valor real	Pronóstico	Error absoluto
5569	2	0.7751	1.2249
5570	1	0.9261	0.0739
5571	1	1.2328	0.2328
5579	3	1.3368	1.6632
5608	5	1.5923	3.4077
5774	1	0.9700	0.03

4.2. Grupo 3.

Éste grupo consta de 289 registros correspondientes a una sola palabra. Las impresiones de éste grupo se transformaron en una serie de tiempo con una frecuencia de doce.

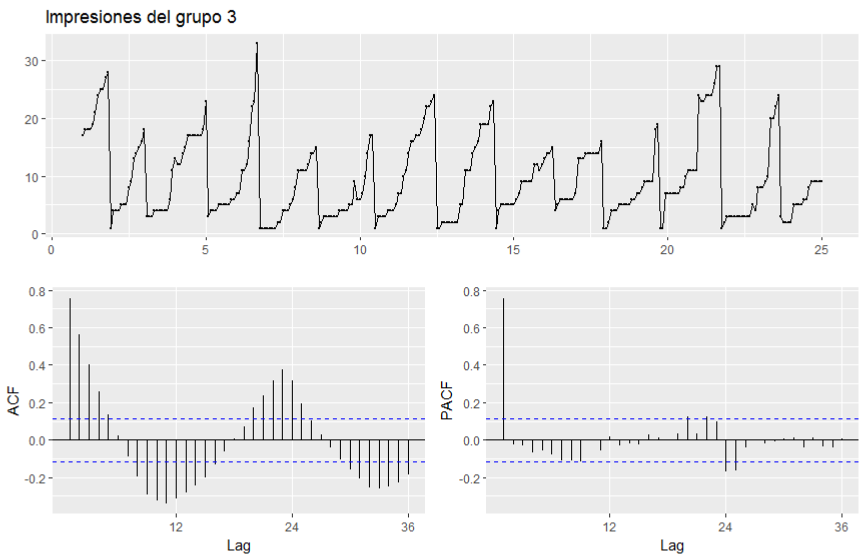

FIGURA 3.13. Serie de tiempo de las impresiones del grupo 3 junto con las ACF y PACF.

Podemos observar en la Figura 3.13 que la tendencia no ésta muy marcada, sin embargo, la estacionalidad se muestra aproximadamente en cada doce pasos y ésto lo corrobora la función ACF donde nos muestra una estacionalidad en cada 11 retardos (lags), como esta es una serie por horas del día entonces estamos capturando un patrón cada 12 horas. Descomponiendo la serie en una serie sin estacionalidad y tendencia, tenemos el siguiente resultado.

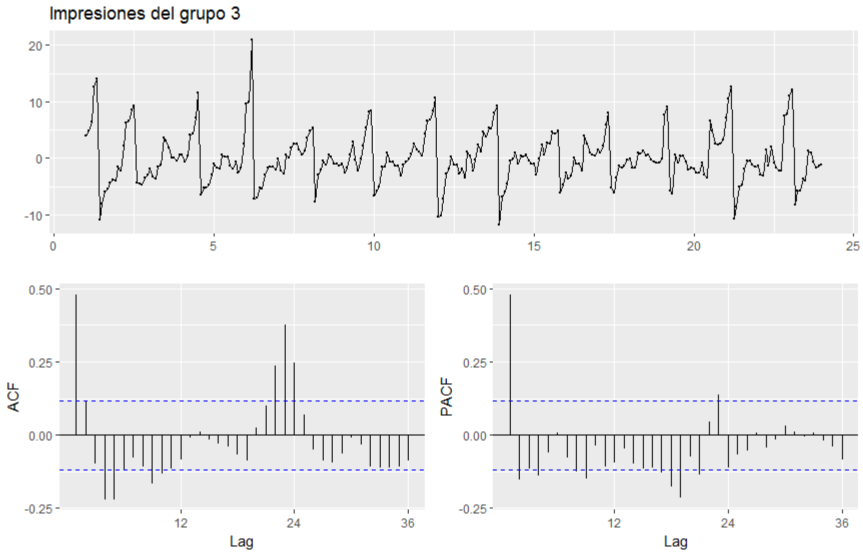

FIGURA 3.14. Serie de tiempo removida de las impresiones del grupo 3 junto con las ACF y PACF.

Observando la serie removida en la Figura 3.14, podemos notar que sigue estando una estacionalidad muy marcada, aproximadamente en cada doce pasos, en la gráfica de ACF observamos valores fuera de la banda de confianza.

Modelo.

Usando el comando *auto.arima()* en R, tenemos que el modelo que mejor se ajusta a los datos es el modelo **ARIMA**$(2,0,0) \times (2,0,0)_{12}$.

Los parámetros de éste modelo son:

$$\phi_1 = 0{,}5128, \phi_2 = -0{,}1415, \Phi_1 = -0{,}0703, \Phi_2 = 0{,}1267.$$

Así, el modelo puede ser reescrito como,

$$\Phi_2(B^{12})\phi(B)x_t = w_t.$$

Donde, $\Phi_2(B^{12})$ y $\phi(B)$ son los operadores autoregresivo estacional y autoregresivo.

Validación.

Para la validación de éste modelo se realizó un análisis de residuos usando la función de autocorrelación (ACF).

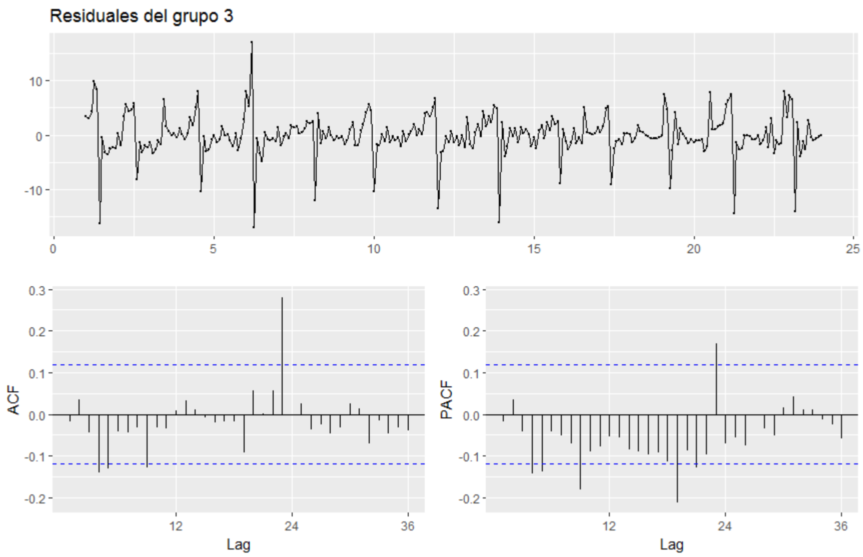

FIGURA 3.15. Residuales de la serie de tiempo de las impresiones del grupo 3 junto con las ACF y PACF del modelo **ARIMA**$(2,0,0)\times(2,0,0)_{12}$.

Podemos notar que en la gráfica de la función de autocorrelación en el retardo 23 (lag 23) el valor se sale muy notoriamente de la banda de confianza.

SARIMA.

Validando el modelo **ARIMA**$(2,0,0)\times(2,0,0)_{12}$ con la función de autocorrelación, gráfica Q-Q normal y la prueba de Ljung-Box tenemos la siguiente figura para los residuales del modelo.

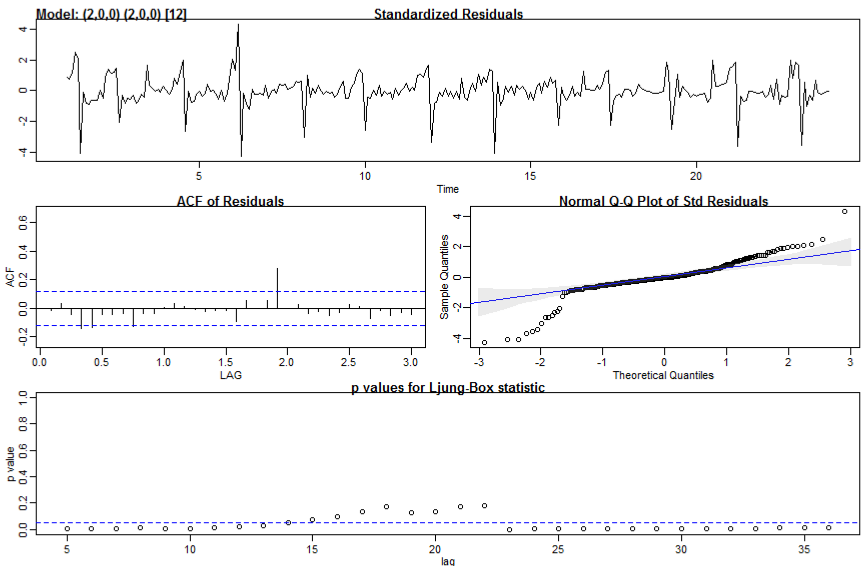

FIGURA 3.16. Residuales de la serie de tiempo de las impresiones del grupo 3 junto con la ACF, gráfica Q-Q normal y la prueba Ljung-Box del modelo **ARIMA**$(2,0,0) \times (2,0,0)_{12}$.

Podemos observar en la gráfica de la prueba de Ljung-Box que la mayoría de los valores-p están por debajo del nivel de significancia, esto quiere decir que hay evidencias para rechazar la hipótesis nula, por lo que los datos están distribuidos de forma dependiente. La mayoría de los valores de la función de autocorrelación están dentro de la banda de confianza a pesar de un valor que sobresale notoriamente. Sin embargo, en la gráfica Q-Q normal se tienen muchos valores ajustados a los cuantiles de la normal, es decir, los estadísticos aceptan la hipótesis nula de similitud.

Modelo ARIMA$(2,1,2) \times (2,1,0)_7$**.**

Después de hacer un trabajo riguroso en busca de un modelo que se ajuste mejor a los datos, hemos encontrado el modelo **ARIMA**$(2,1,2) \times (2,1,0)_7$ como candidato a modelar los datos. Los parámetros son los siguientes:

$$\phi_1 = 1{,}3593, \phi_2 = -0{,}5476, \theta_1 = -1{,}9988, \theta_2 = 0{,}9988, \Phi_1 = -0{,}6536, \Phi_2 = -0{,}3624$$

Así, el modelo puede ser reescrito como,

$$\Phi_2(B^7)\phi(B)\Delta_7\Delta x_t = \theta(B)w_t.$$

Donde, Δ_7 y Δ son las componentes de diferencias estacionales.

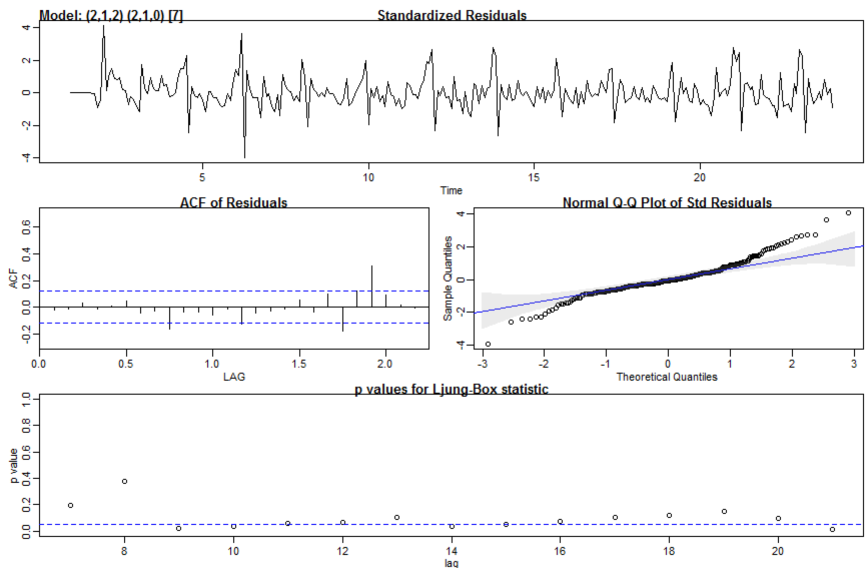

FIGURA 3.17. Residuales de la serie de tiempo de las impresiones del grupo 3 junto con la ACF, gráfica Q-Q normal y la prueba Ljung-Box del modelo **ARIMA**$(2,1,2)\times(2,1,0)_7$.

Observando la gráfica de la prueba de Ljung-Box, la mayoría de los valores-p están por encima del nivel de significancia, esto quiere decir que no hay evidencias para rechazar la hipótesis nula, por lo que los datos están distribuidos de forma independiente. La mayoría de los valores de la función de autocorrelación están dentro de la banda de confianza a pesar de que algunos valores sobresalen notoriamente. Adicionalmente, en la gráfica Q-Q normal se tienen muchos valores ajustados a los cuantiles de la normal, es decir, los estadísticos aceptan la hipótesis nula de similitud.

Comparaciones.

El criterio para elegir uno de los dos modelos propuestos son los criterios AIC, AICc, y BIC, el cual es elegir aquel cuyo valor es mínimo.

Modelo	Varianza estimada	Criterio AIC	Criterio AICc	Criterio BIC
ARIMA$(2,0,0) \times (2,0,0)_{12}$	15.61	3.784	3.7425	2.8496
ARIMA$(2,1,2) \times (2,1,0)_7$	17.34	3.8965	3.9053	2.9750

De aquí, el mejor modelo que se ajustó a los datos es el **ARIMA**$(2,0,0) \times (2,0,0)_{12}$, sin embargo, de acuerdo a la prueba de hipótesis Ljung-Box el modelo que mejor se ajusta a los datos es el **ARIMA**$(2,1,2) \times (2,1,0)_7$ ya que la distribución de los datos es independiente, es decir, las correlaciones en la población de la que se toma la muestra son cero.

Escogiendo el modelo **ARIMA**$(2,1,2) \times (2,1,0)_7$ como el mejor que se ajustó a los datos, veamos cómo se ajusta éste modelo a la serie original. Datos estimados por el modelo (gráfica azul) y Serie original de impresiones del grupo 3 (gráfica negra).

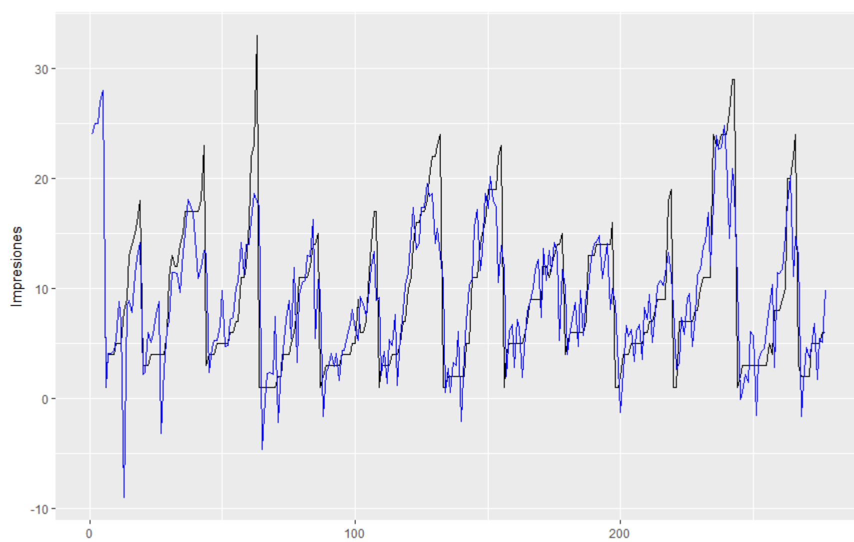

Figura 3.18. Serie de tiempo de impresiones del grupo 3 junto con la estimacion del modelo **ARIMA**$(2,1,2) \times (2,1,0)_7$.

Podemos notar que el modelo se ajusta bien a los datos ya que mantiene el comportamiento de la serie. Éste resultado es de utilidad para el pronóstico de la serie de impresiones de dicho grupo.

Pronóstico.

Finalmente, al escoger el modelo $\mathbf{ARIMA}(2,1,2)\times(2,1,0)_7$ como el mejor que se ajustó a los datos se llevó a cabo el pronóstico de las impresiones para los últimos 49 registros en el día cuarto del mes de octubre de 2017.

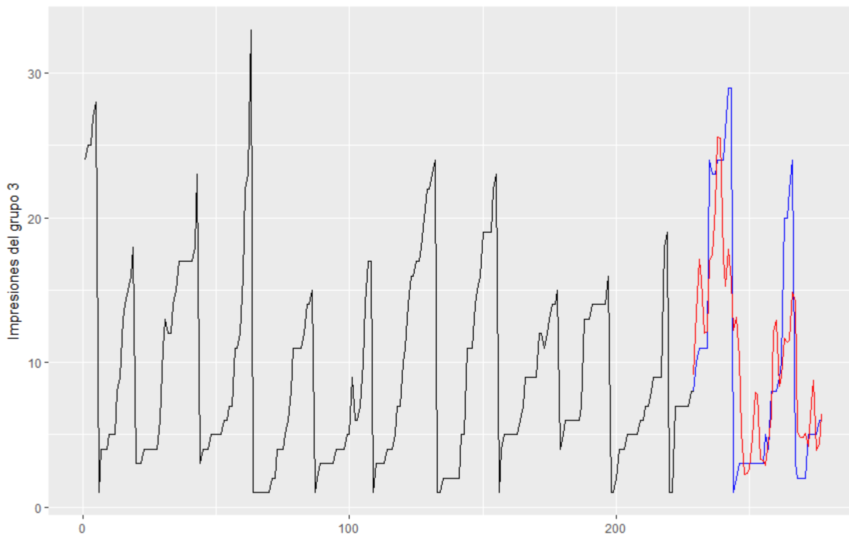

FIGURA 3.19. Pronóstico del modelo ajustado a la serie de tiempo de las impresiones para los últimos 49 registros. (Línea roja valor estimado y línea azul valor original.)

Se observa en la Figura 3.19 que el modelo predice un comportamiento parecido al comportamiento de los valores originales, siendo éste modelo un modelo confiable al que puedan hacerse análisis a futuro con respecto a las impresiones de éste grupo.

En la tabla siguiente se muestra el valor real, el pronóstico y el error absoluto de algunos registros.

Período	Valor real	Pronóstico	Error absoluto
229	8	9.1345	1.1345
230	10	13.4490	3.449
233	11	12.0148	1.0148
237	23	20.7589	2.2411
238	24	25.5817	1.5817
277	6	6.4238	0.4238

Conclusiones

Como hemos apreciado anteriormente, el conjunto de palabras se segmentó en tres grupos claramente diferenciados por las variables clics, impresiones y conversiones. El grupo 1 se caracteriza por tener menos impresiones que todos los grupos y además no se registran clics, el grupo 2 a diferencia del grupo 1 tiene al menos un clic y más impresiones, sin embargo, no tiene conversiones, por su parte el grupo 3 presenta características más ricas como, más impresiones y clics que todos los grupos y el total de las conversiones en el conjunto de palabras.

Éstas características fueron de mucha utilidad al momento de ajustarle modelos ARIMA a las impresiones de los dos grupos escogidos, como lo fueron el grupo 1 y el grupo 3. Una vez ajustado los modelos ARIMA a éstos grupos, la predicción arrojó para el grupo 1 una tendencia similar a la de los valores originales de la serie, sin embargo, éstos variaban con respecto a las alturas, es decir, a la cantidad de impresiones de la serie. La predicción del grupo 3 arrojó un comportamiento parecido a los datos originales, por lo que éste modelo es un modelo al cual se le puede realizar futuras predicciones. Finalmente, éstos modelos ajustados a los grupos 1 y 3 pueden ayudar a analizar sí cada uno de los grupos de palabras tendrán buenos rendimientos o no, y así poder usarlas para evaluar que tan atractivo son los anuncios en el desarrollo de la campaña publicitaria de sombreros.

Bibliografía

[1] A. Kassambara, (2017). Practical Guide To Cluster Analysis in R. 1era Edición. Editorial STHDA.

[2] Cluster Analysis. https://www.stat.berkeley.edu/~s133/Cluster2a.html.

[3] G.H.Duteman, (1987). Principal Components Analysis. A sage university paper 69. Editorial Sage Publications,Inc.

[4] B. Everitt, T.Hothorn, (2011). An Introduction to Applied multivariate Analysis with R. Springer.

[5] R. Shumway, D. Stoffer, (2011). Time Series Analysis and Its Aplications. 3era Edición. Springer.

[6] J.B.Hernández. Análisis de Series de Tiempo. Universidad Central de Venezuela.

[7] Google AdWors. https://support.google.com/adwords/?hl=es.