

Fastbook 05

Analítica de Cliente & Predictive Analytics

Introducción a las series temporales



05. Introducción a las series temporales

En este fastbook de series temporales vamos a cubrir los conceptos necesarios para entender cómo modelizar y predecir series temporales, pero antes de meternos en detalles técnicos vamos entender el contexto en el que se sitúan las series temporales como herramienta analítica en una compañía.

Autor: Miguel Ángel Fernández

Introducción

Conceptos

Resumen y conclusiones

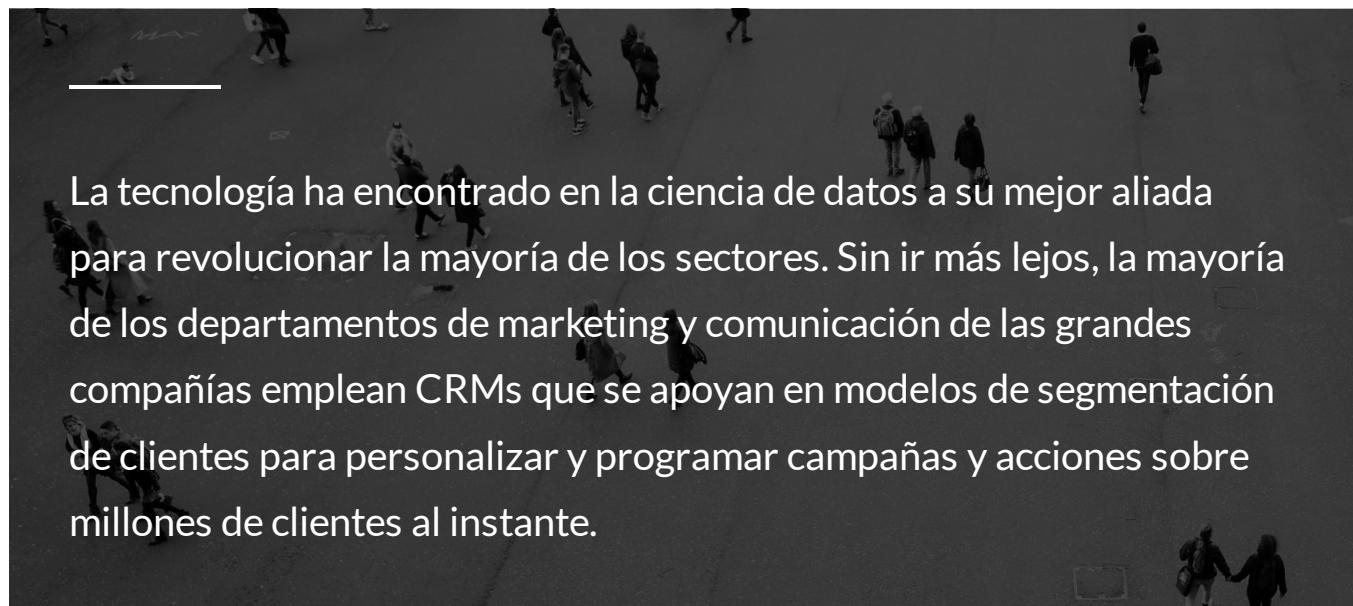
Lección 1 de 3

Introducción

X Edix Educación

A todos nos ha tocado estudiar historia en algún punto de nuestra vida y, gracias a ello, nos son familiares las llamadas Revolución Agrícola y Revolución Industrial. El día a día tal y como lo conocían las personas cambió radicalmente en esos dos momentos de la historia, muy especialmente en el ámbito laboral. Hoy ‘estamos de suerte’, porque vamos a presenciar, si no empezó ya, la siguiente revolución: la tecnológica.

Con los cuatro primeros fastbooks de esta asignatura hemos visto cómo funcionan los modelos de segmentación. Estos modelos son la base que hay detrás de muchas campañas y decisiones que se toman hoy en día en las grandes compañías y responden a muchas preguntas sobre nuestros clientes.



La tecnología ha encontrado en la ciencia de datos a su mejor aliada para revolucionar la mayoría de los sectores. Sin ir más lejos, la mayoría de los departamentos de marketing y comunicación de las grandes compañías emplean CRMs que se apoyan en modelos de segmentación de clientes para personalizar y programar campañas y acciones sobre millones de clientes al instante.

A partir de este fastbook vamos a dejar el capítulo de la segmentación de clientes a un lado, mientras que en los siguientes vamos a estudiar otra familia de modelos analíticos: **las series temporales** (time series).

Prácticamente es como si cambiáramos de asignatura, no obstante, estos dos tipos de modelos están juntos porque tienen varios puntos importantes en común:

1

Si los modelos de segmentación responden a la pregunta ‘**¿cómo son mis clientes?**’, los modelos de series temporales nos ayudan a responder ‘**¿cómo está mi negocio?**’ o, al menos, ‘**¿a dónde va mi negocio?**’. Sin duda, estas son dos preguntas que interesan a cualquier CEO y se complementan la una a la otra para tomar decisiones clave para una compañía.

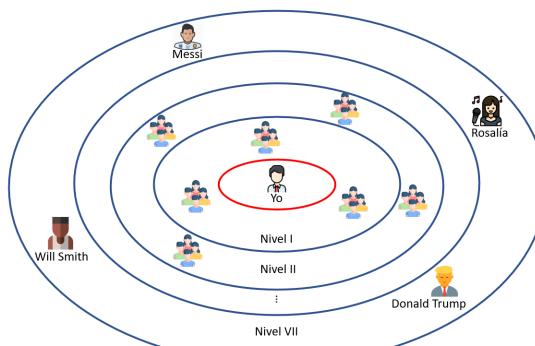
2

Además, las series temporales y los **modelos de segmentación** son **modelos de machine learning** muy especiales en cuanto a que difieren bastante en cómo funcionan del resto de modelos conocidos (regresión y clasificación).

La manera en la que vamos a estudiar las series temporales seguirá el mismo guion que hicimos con los modelos de segmentación: primero veremos los fastbooks de teoría y después los de práctica, solo que ahora, cada vez que veamos un fastbook de teoría, vamos a ver después el de práctica que le corresponde.

Para que lo tengamos claro, este será el orden de los fastbooks que nos quedan por ver:

- **Fastbook 05 (actual)**: teoría general de series temporales I.
- **Fastbook 06**: prácticas de series temporales I.
- **Fastbook 07**: teoría general de series temporales II.
- **Fastbook 08**: prácticas de series temporales II.
- **Fastbook 09**: teoría general de series temporales III.
- **Fastbook 10**: prácticas de series temporales III.



KPIs everywhere

El mundo es cada vez más complejo. Cada día que pasa las tecnologías nos permiten interactuar con más personas, realizar más acciones y almacenar más información. Tanto es así que existe una teoría que asegura que las personas estamos conectadas por siete niveles, por ejemplo...

Imaginemos que hay una única persona que conoce el secreto X y se lo cuenta a todas las personas que conoce. Si cada una de estas personas contase el secreto X de nuevo a todas las personas que conoce y así sucesivamente, se estima que todo el mundo acabaría enterándose del secreto en solo siete saltos.

De este modo se dice que cualquier persona del mundo está a solo siete niveles de conocerte.

Esta teoría sorprende más si hablamos de personas concretas, ya que estamos a solo siete personas de distancia de, por ejemplo, Will Smith, Rosalía, Donald Trump o Lionel Messi.

En todo este mundo complejo lleno de información, las empresas no son la excepción y necesitan herramientas y métodos nuevos para organizar y gestionar todos los datos que se generan en el día a día.

Cuando hablamos de **información o datos** nos referimos concretamente a la información relacionada con el negocio de una compañía, por ejemplo, las ventas que hemos tenido, las suscripciones a nuestro servicio, las bajas de clientes, las visitas a nuestra web, etc.

Por lo tanto, no nos referimos a los mensajes de fútbol que mandamos por Teams al compañero cuando nuestro equipo gana al suyo (por mucho que sea ‘información que se genera en la empresa’).

Con toda la cantidad de mediciones posibles, acciones sobre nuestros clientes, acciones de nuestros propios clientes, movimientos en el mercado de la competencia, etc., solo podemos tener el máximo detalle de todo lo que ocurre en nuestro negocio si nuestra empresa es pequeña.

Pero entonces, ¿cómo hacen en las grandes compañías los CEOs o los altos cargos para tener el conocimiento de qué ocurre en su negocio? Pues usando los famosos **KPIs (Key Performance Indicator)**.



Los **KPIs** son esos números que resumen grandes aspectos de un negocio y que nos sirven a modo de termómetro para saber cómo está un aspecto de la compañía.



Estos indicadores ‘mágicos’ son la manera que tenemos de resumir la información de muchos procesos que ocurren a nuestro alrededor. Tanto es el poder que tiene la estadística matemática para resumir la información que hoy en día hemos sido **rodeados por estos KPIs**; queremos medir todo.

Algunos ejemplos de KPIs son:

- Satisfacción del cliente.
- ROI (return of investment).
- CTR (click through rate).
- Nivel de facturación.
- Número de visitas a la web.
- Número de descargas de la app.
- Tiempos de permanencia en la web.
- Leads conseguidos.
- Número de seguidores.
- Alcance.
- Engagement.
- Coste medio por pedido.

Los KPIs son muy relevantes para un negocio, pero **más interesante aún es su evolución**.

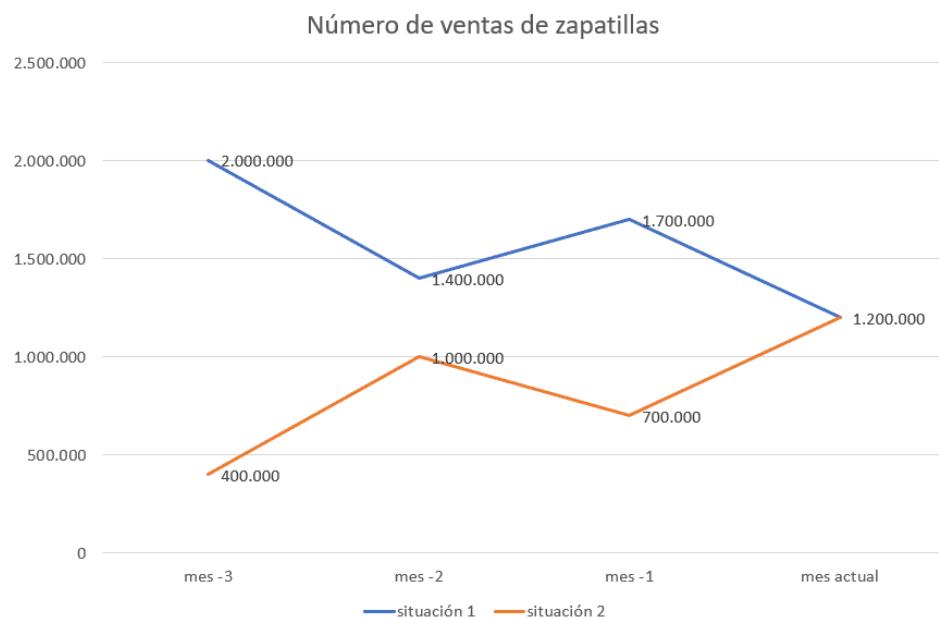
Hay situaciones en las que un valor aislado en el tiempo de un KPI es suficiente para evaluar si esa faceta del negocio va bien o no, pero en la mayoría de los casos es crucial recoger la evolución que ha tenido el KPI en el tiempo, no solo el valor que tenga en el momento actual.

Ejemplo 1

Imaginemos que nuestra compañía ha tenido un ROI de 5 puntos en la campaña de verano. Este KPI aislado en el tiempo es suficiente para valorar que ha ido bien, por cada euro que nos hemos gastado en publicidad hemos recuperado 5.

Ejemplo 2

Imaginemos ahora que vendemos zapatillas y el número de ventas del mes actual es de 1.2 millones. A priori, 1.2 millones de ventas suena muy, muy bien, pero este KPI aislado no nos sirve para evaluar si nos está yendo bien o no. No es lo mismo que el gráfico de ventas sea la situación 1 que la situación 2.



Cada uno de estos KPIs que vamos recogiendo en intervalos fijos de tiempo (cada día, cada hora, cada mes, cada año...) define lo que llamaremos una serie temporal. Estas 'líneas' de color azul y naranja en el anterior gráfico son series temporales de ventas de zapatillas.

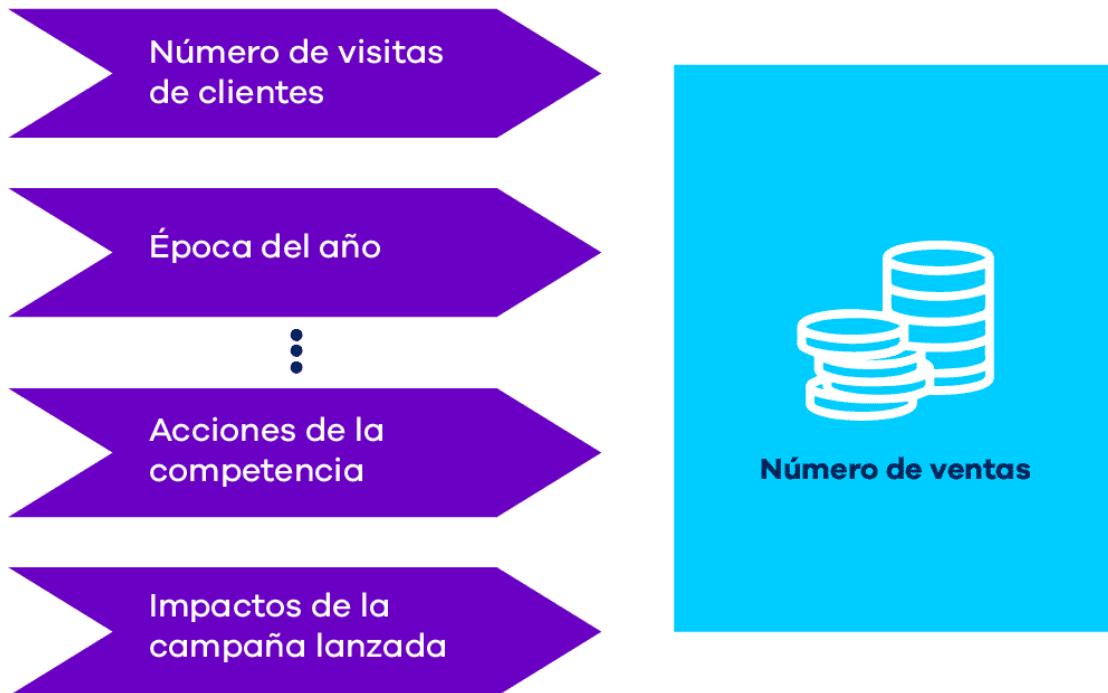
Con este enfoque temporal de los KPIs se nos abre un mundo de posibilidades que podemos explotar con técnicas analíticas, pero antes de seguir profundizando vamos a comentar la precaución que debemos tener cuando usamos KPIs.

Cuidado con los KPIs

Acabamos de ver cómo los KPIs nos sirven para resumir todo lo que ocurre en un proceso en un número, pero para resumir es necesario 'tirar' cierta información y este es el punto que vamos a tratar en este apartado: **hay que tener mucho cuidado con usar e interpretar correctamente un KPI**. El fin de un KPI no es reducir información por el arte en sí de resumir.

El objetivo es que con el valor que tome un KPI podamos detectar de forma rápida un aspecto del negocio y podamos emprender acciones para corregirlo (en el caso de que vaya mal) o aprender qué hemos hecho bien.

Que un KPI suba o baje es debido a una serie de eventos que impactan sobre él. Por ejemplo, el KPI del número de ventas puede ser influenciado por múltiples factores:



Con el KPI del número de ventas hemos resumido un montón de eventos de nuestro negocio en un único valor, podemos detectar más fácilmente si el negocio va bien o va mal, pero perdemos todo el detalle que hay detrás de este KPI.

Por esta razón **no nos suele valer con un solo KPI**, porque si suben o bajan las ventas de zapatillas lo natural es que queramos entender por qué, y para ello necesitamos recuperar información que hemos ‘tirado’ al resumir todo en un único número.

Para recuperar esta información **solemos recurrir de nuevo a otros sub-KPIs que de forma teórica asumimos que están relacionados con el KPI principal** (el número de ventas en este caso), de modo que si bajan las ventas podamos identificar qué elementos no han ido bien y han impactado en ellas.

De esta forma, si tuviéramos los siguientes KPIs podríamos entender por dónde están yendo mal las cosas. Por ejemplo:

- **Número de visitas a la web:** con este KPI podríamos detectar que ha ocurrido una caída importante en las visitas a la web del mes actual, de modo que la caída en las ventas totales viene producida por las ventas online.
- **GRPs estimados de la campaña de la competencia:** nuestra principal competencia ha lanzado una campaña con unos GRPs estimados muy superiores a lo natural, de modo que nos ha quitado clientes este mes.
- **GRPs de nuestra publicidad:** la estimación que tenemos de nuestra campaña de publicidad indica una caída importante en el último mes que puede explicar la caída de las ventas totales.

Los principales riesgos al usar un KPI son:

- 1 La pérdida de información que acabamos de señalar.
 - 2 Que muchas veces asumimos de forma heurística, sin ninguna ciencia, la relación que existe entre varios KPIs.
-

Para ver claramente este punto observemos la siguiente situación con la que terminaremos este apartado.

Ahora se ha puesto de moda recoger la opinión de un cliente al finalizar un servicio. De forma que en un restaurante o en una tienda puedes ver que te piden que valores tu satisfacción con el servicio a través de una pantalla táctil con las siguientes caritas:

Muy insatisfecho



Muy satisfecho



Este es de nuevo un intento de las compañías para obtener más información de lo que ocurre en el negocio, solo que en vez de hacer una encuesta a los clientes (que claramente ninguno tenemos las ganas de completar), resumimos todo lo que ha ocurrido en el servicio en unos KPIs, unos pocos números, por ejemplo:

- **KPI1:** Número de Muy insatisfechos + Insatisfechos.
- **KPI2:** Número de Muy satisfechos + Satisfechos.

Llegados a este punto, la pregunta que te lanzo es la siguiente:

¿Qué pasa si uno de los dependientes recibe al finalizar un lunes un KPI1 = 5 y un KPI2 = 10?

Los números no engañan, son los que son, pero podemos cometer un gran error al interpretar estos KPIs. Muchas personas no dudarían en señalar al dependiente como el principal culpable de la insatisfacción de esos 5 clientes, pero hay muchos más factores a tener en cuenta para evaluar correctamente la decisión a tomar.

Sin duda, el peor uso que podemos hacer de este KPI es meter miedo a nuestros empleados. Por ejemplo, los KPIs son de un lunes y... ¿A quién le gustan los lunes? Directamente puede que 3 de los 5 clientes tuvieran un mal día y lo que hoy han valorado como 'Insatisfecho', un viernes podría ser un 'Neutro'.

¿Qué ocurre si la insatisfacción viene porque la tienda era de zapatillas y no había el número que buscaban los clientes? Al cliente se le pide que valore el servicio, pero **aquí se mezcla la atención recibida del empleado y el producto ofrecido al cliente.**

Con este ejemplo queremos cerrar con las dos principales lecciones que debemos tener en cuenta al utilizar KPIs:

1

Ten cuidado con la pérdida de información.

2

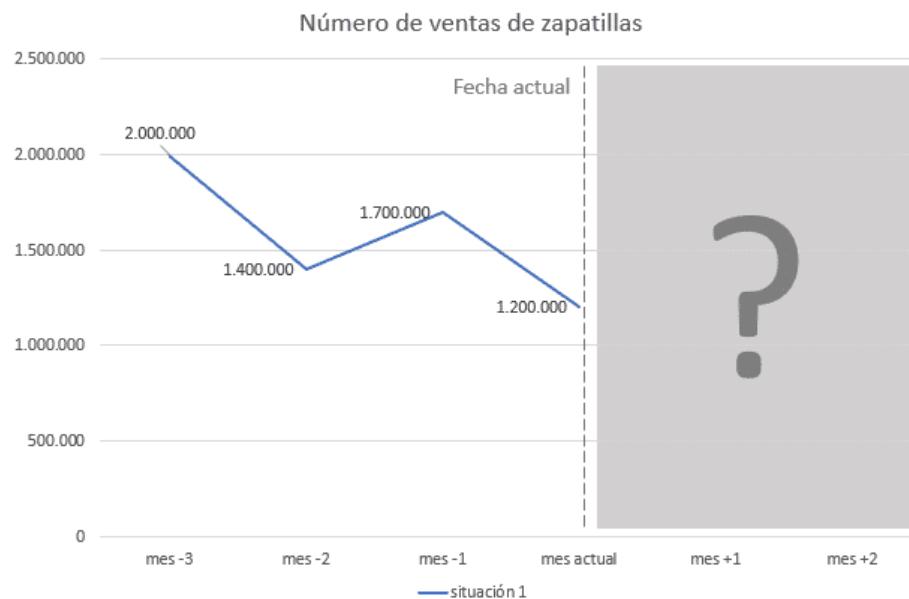
Ten cuidado con la explicación teórica que damos a los KPIs que obtenemos.

Conocer a dónde va tu negocio

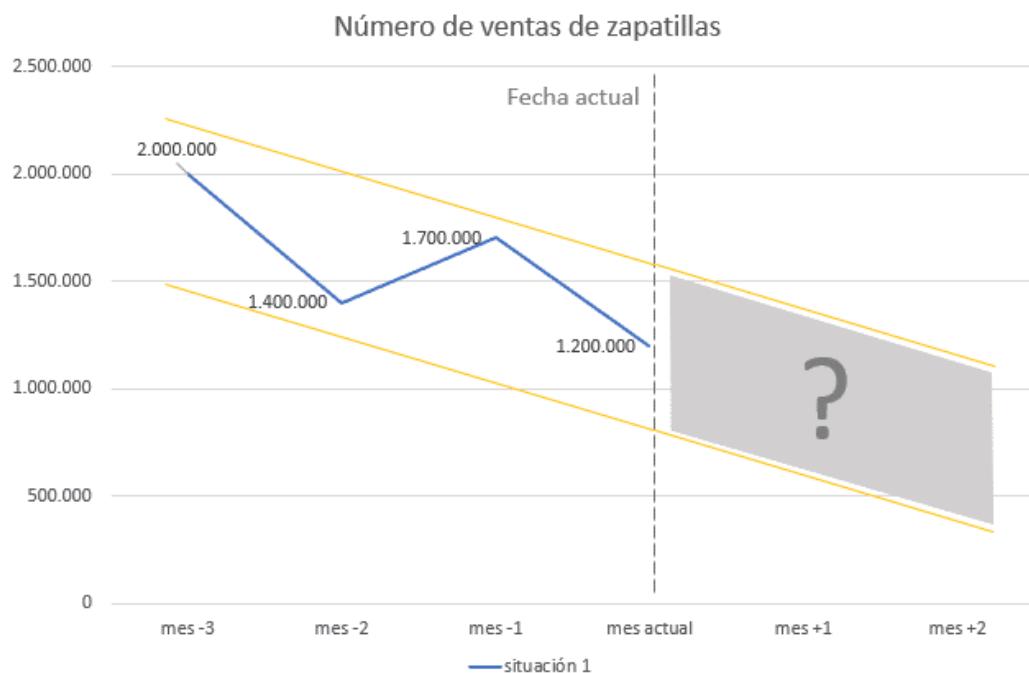
Llegados a este punto ya hemos visto qué es un KPI y cómo, al recoger varios KPIs a lo largo del tiempo, construimos una **serie temporal**. Una serie temporal puede aportarnos información en dos direcciones: pasado o futuro.

Cuando simplemente pintamos en un gráfico la serie temporal, obtenemos información de **qué ha ocurrido**. Pero esta información también nos puede servir para **predecir qué puede ocurrir en el futuro**.

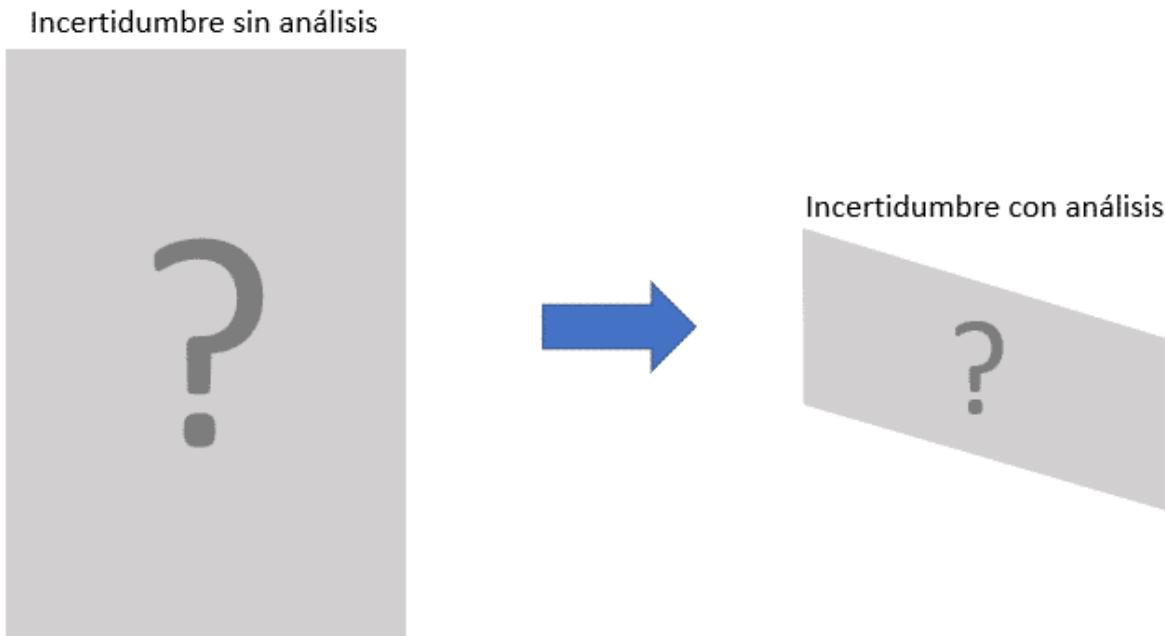
Si volvemos a visitar el ejemplo de las ventas de zapatillas, tenemos el siguiente gráfico descriptivo que nos indica qué ha ocurrido, pero es igual o más interesante aún conocer qué va a ocurrir en los próximos meses, por ejemplo:



Sería muy importante tener más histórico para saber cómo de fuerte es la tendencia decreciente del número de ventas, pero con esta información, una predicción muy naif sería la siguiente:



Siguiendo la misma tendencia, podemos tener una estimación de qué ventas vamos a tener en los meses siguientes, de modo que reducimos drásticamente la incertidumbre que tenemos sobre lo que va a ocurrir:



Naturalmente esto que hemos hecho es una estimación: viendo las ventas de los últimos meses estamos estimando qué va a ocurrir en los siguientes meses y podemos acertar o fallar en dicha estimación.

Con todo esto, se nos abren dos preguntas a responder:

- 1 ¿Cómo hago una estimación?

- 2 ¿Cómo podemos entonces tener buenas estimaciones?

La respuesta a la primera pregunta es **empleando modelos de series temporales**, algo que veremos con más detalle en la siguiente sección. Y, respecto a la segunda, para tener buenas estimaciones necesitamos dos ingredientes:

- Tantos datos como sea posible.
- Conocer muchos modelos de series temporales.

Nuestra compañía tendrá un histórico de datos determinado, de modo que los datos que tenemos son los que hay. En nuestra mano está trabajar el segundo punto, que será el foco de los fastbooks que nos quedan: **conocer los modelos de series temporales**.

¿Cómo lo hago?

La herramienta analítica que utilizaremos para dar estimaciones de series temporales son los denominados **modelos de series temporales**. En lo que queda de temario, veremos inicialmente los primeros modelos de series temporales para, más adelante, que podamos abordar los modelos más avanzados.

Cada uno de estos modelos tiene un grado de dificultad distinto, por lo que haremos igual que con los modelos de segmentación: primero veremos los clásicos, que emplean conceptos más sencillos, y después terminaremos la asignatura viendo modelos complejos.

Los modelos son sin duda la parte más interesante y que más luce de todo proyecto de analítica avanzada. No obstante, vamos a necesitar familiarizarnos con distintos conceptos que van a aparecer más adelante para tratar y modelizar series temporales.

Esto es lo que vamos a ver en la siguiente sección y lo que nos ocupará hasta el final de este fastbook.

Conceptos

X Edix Educación

El mundo de la ciencia de datos está lleno de palabras raros, tenemos un nombre para todo. Una vez hemos hecho el esfuerzo de entenderlos, nos son muy útiles para trabajar conceptos más complejos que se construyen sobre ellos.

Por ejemplo, la serie de ventas de nuestra tienda de zapatillas la hemos dibujado de forma mensual, cada punto de nuestra serie temporal es el número de zapatillas que hemos vendido cada mes, pero ¿por qué contar por cada mes y no cada semana, cada día o cada trimestre?

Nos será útil ponerle un nombre a ‘cada cuánto contamos’, tener claro a qué nos referimos cuando estemos tratando las series temporales (spoiler: este concepto se denomina *granularidad* y será el primero que veamos).

En esta sección vamos a visitar los conceptos más importantes que requiere el análisis de series temporales.

Granularidad

El concepto de **granularidad** indica la periodicidad con la que se recolectan los valores de una serie temporal, y debe ser fijo para todos los valores de la serie.

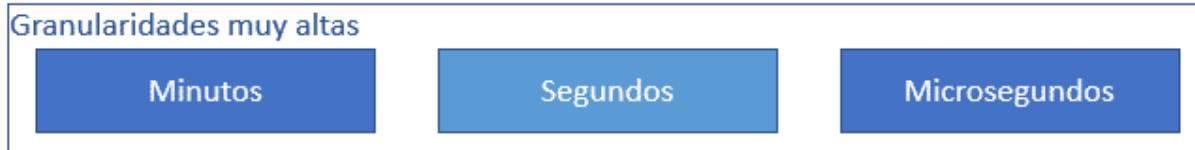
Es decir, si nos dicen que la serie de ventas tiene una granularidad mensual, quiere decir que tenemos medido el número de zapatillas que se han vendido cada mes; si nos dicen que la serie tiene una granularidad semanal, entonces quiere decir que tenemos medido el número de zapatillas que se han vendido cada semana.

De este modo, hay tantos niveles de granularidad como formas de medir intervalos de tiempo.

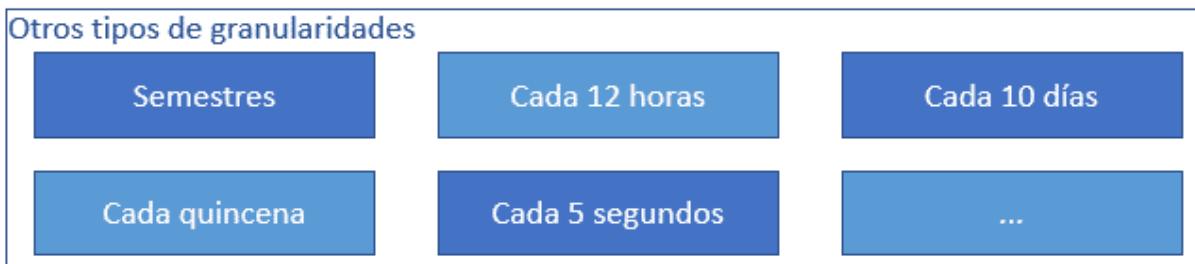
Las **granularidades más clásicas** que nos vamos a encontrar son:

Granularidades clásicas					
Anual	Trimestral	Mensual			
Semanal	Diaria	Horaria			

No obstante, nada nos impide llegar a niveles mucho más pequeños de tiempo, por ejemplo:



Y también podemos agrupar cualquiera de estos niveles para dar lugar a cualquier intervalo de tiempo que sea adecuado para nuestro problema o negocio, por ejemplo:



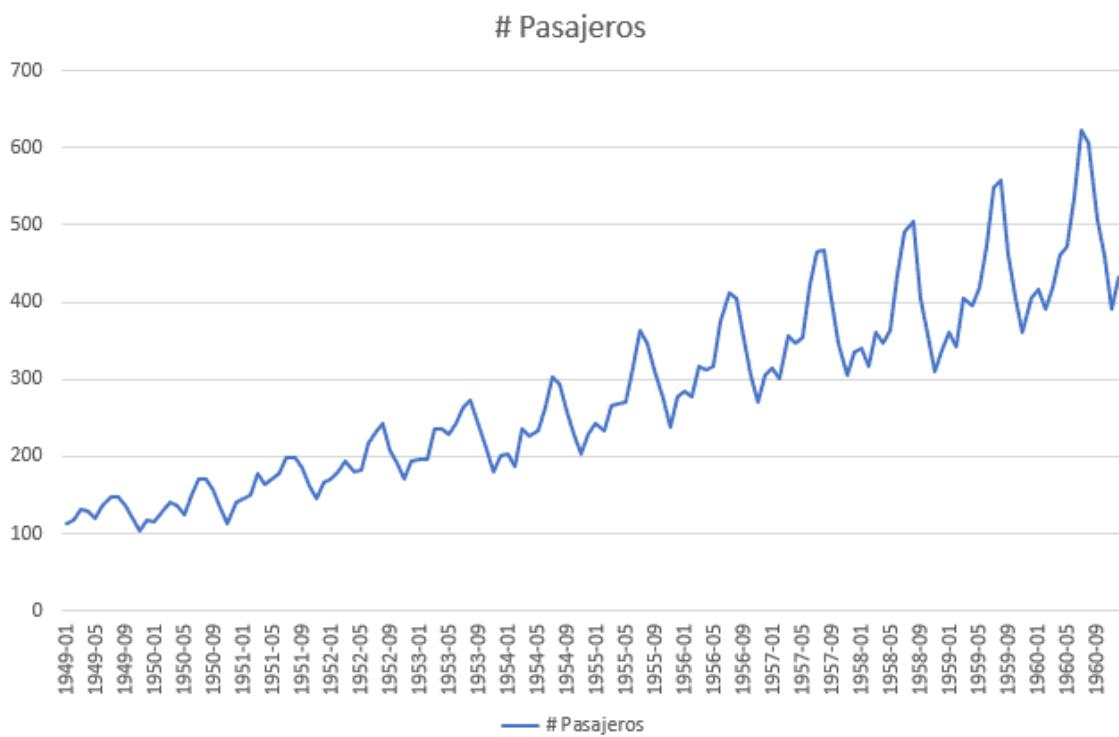
Como puedes comprobar, la lista es infinita y, dependiendo de la naturaleza de la serie temporal, es más natural medir los eventos con una cierta granularidad: todo depende del nivel de detalle que queramos tener.

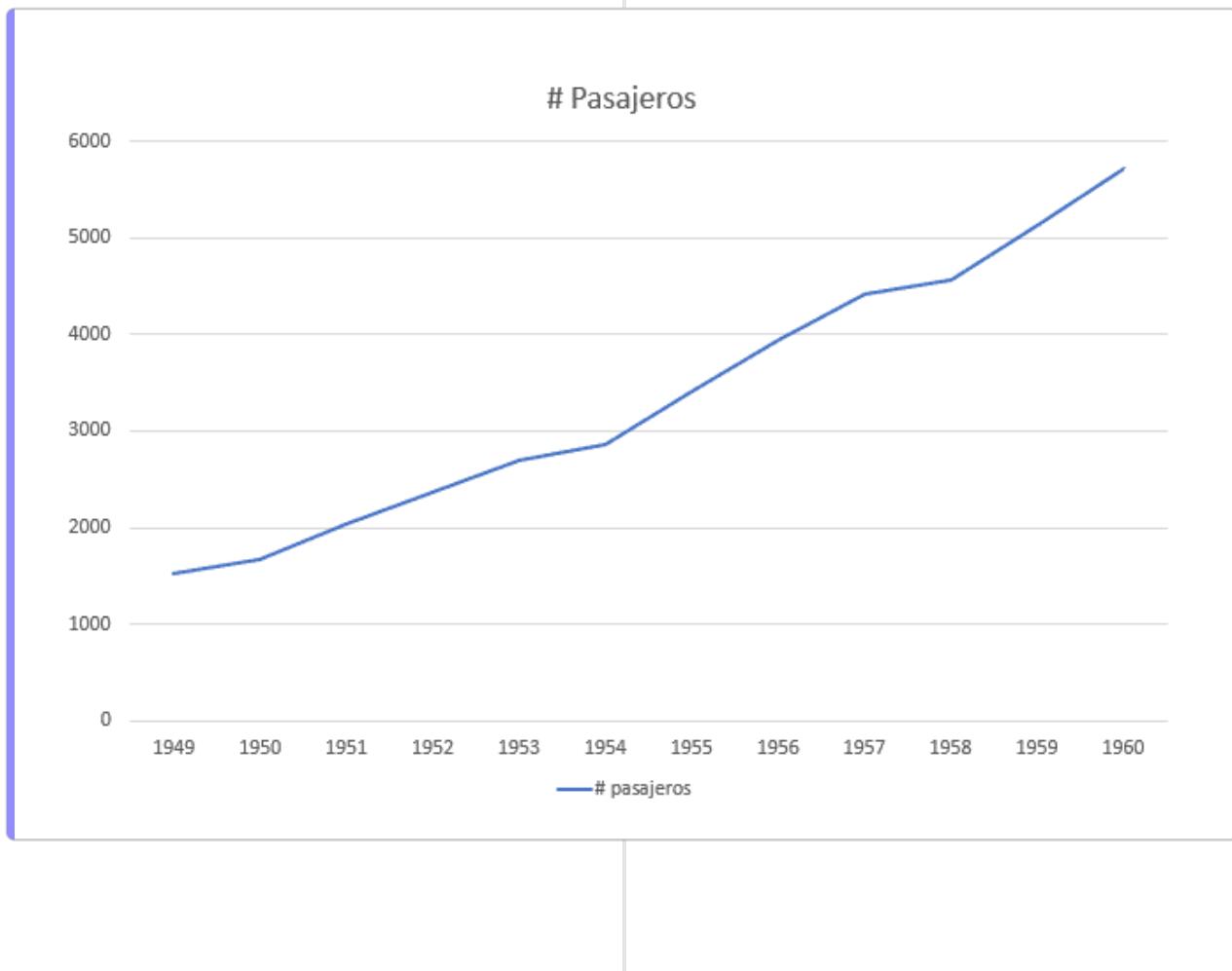
Por ejemplo, para las ventas de zapatillas de nuestras tiendas no tiene mucho sentido construir la serie temporal con granularidad de minutos; es más natural tener estas ventas por días, semanas o meses. Por el contrario, si operamos en bolsa, nos interesaría tener el detalle del valor del stock a nivel segundos o minutos en muchos casos.

**La elección del nivel de granularidad será nuestro primer paso
para tratar series temporales de forma correcta.**

Esta elección es una decisión importante, ya que la forma que tiene la serie temporal depende en gran parte de ello. Vamos a ver un ejemplo para observar el efecto que puede tener el nivel de granularidad sobre la forma de la serie.

En este ejemplo tenemos el número de clientes que tiene una aerolínea entre 1949 y 1960. Las siguientes series temporales se han construido a partir de los mismos datos, solo que en el primer gráfico verás que tiene una granularidad mensual, mientras que en el segundo tiene una granularidad anual:





Está claro que el primer gráfico tiene un patrón mucho más interesante que el segundo. Seguramente también será más difícil de modelizar, pero el patrón que tiene es tan repetitivo que nos puede dar esperanzas de poder construir un modelo con el que podamos predecir mes a mes qué va a ocurrir. Por otro lado, la serie del segundo gráfico es mucho más estable y nos será más fácil de modelizar. El problema es que la granularidad anual solo nos permitirá estimar el número de clientes que vamos a tener año a año (perdemos detalle en las predicciones).

Al margen de lo comentado, podemos ver lo distinta que pueden llegar a ser una misma serie cambiando el nivel de granularidad.

Las series de negocio que más frecuentemente nos vamos a encontrar tienen una granularidad clásica y es sobre las que trabajaremos en la mayoría de los ejemplos y prácticas en esta asignatura.

No obstante, desde el punto de vista de la modelización no nos afecta mucho trabajar con series diarias, anuales o por segundos: cualquiera de los modelos que vamos a ver pueden emplearse sobre series temporales con cualquier nivel de granularidad.

Elementos de una serie temporal

Cuando tratamos series temporales es común emplear los siguientes elementos para referirnos a características o componentes que tiene una serie temporal.

Las definiciones que vamos a ver son un poco formales, pero seguro que con los ejemplos entenderás más claramente cada uno de estos conceptos:

Tendencia

En una serie temporal existe tendencia cuando hay un crecimiento o decrecimiento durante un largo periodo de tiempo en los datos de una serie. El efecto visual es que la serie tiene forma de rampa, pero no tiene por qué ser en línea recta.

Estacionalidad

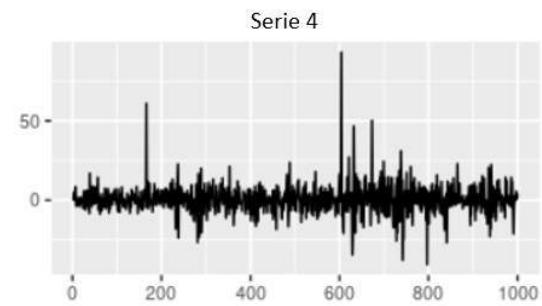
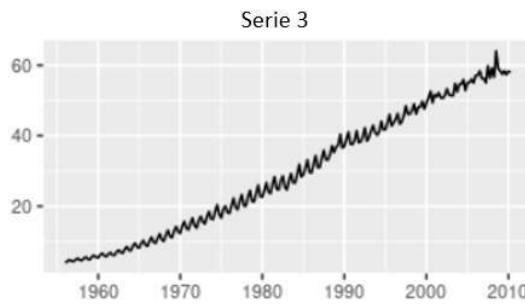
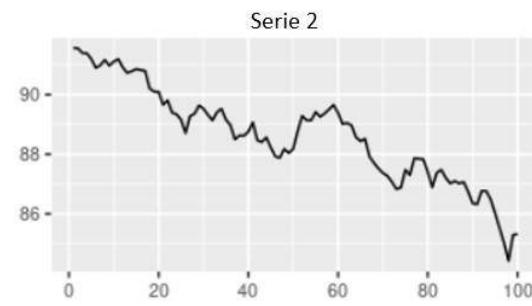
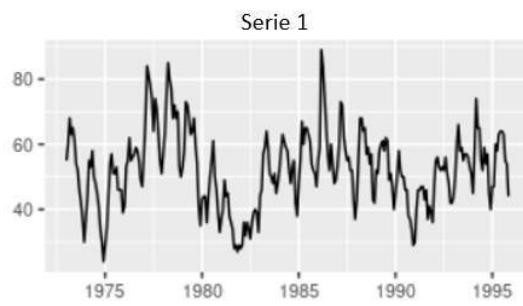
Diremos que nuestra serie tiene un patrón estacional cuando los valores que toma la serie están influidos por factores estacionales. Por ejemplo, si vendemos entradas de cine, claramente nuestras ventas se verán influidas por ser miércoles (día del espectador) o fin de semana. Además, es importante señalar que estos factores estacionales tienen una longitud fija (cada semana, cada 30 días, cada 3 meses, etc.).

Ciclo

Diremos que nuestra serie tiene un ciclo cuando los datos muestran subidas y bajadas que no son de una longitud fija (al contrario que los factores de estacionalidad). Estas fluctuaciones suelen deberse a condiciones económicas y están relacionadas con el contexto del propio negocio.

Para entender de una manera más clara los tres conceptos anteriores te propongo el siguiente ejercicio.

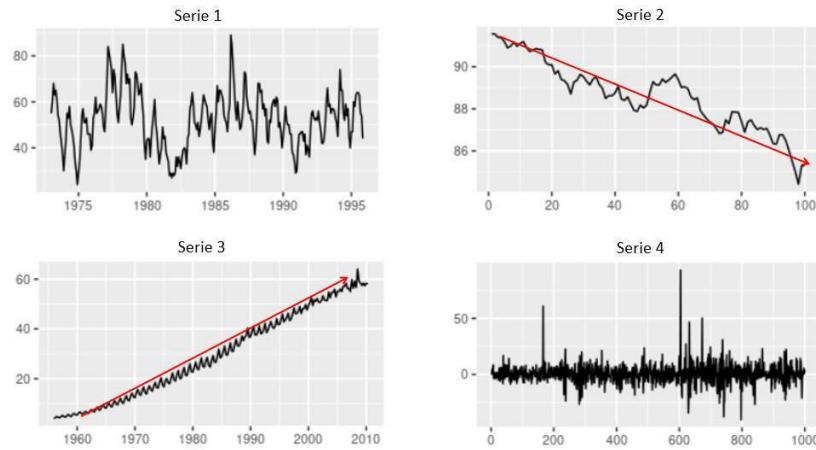
Vas a ver el gráfico de cuatro series temporales, y el ejercicio consiste en identificar cuáles de los tres anteriores factores (tendencia, estacionalidad, ciclo) muestra cada serie:



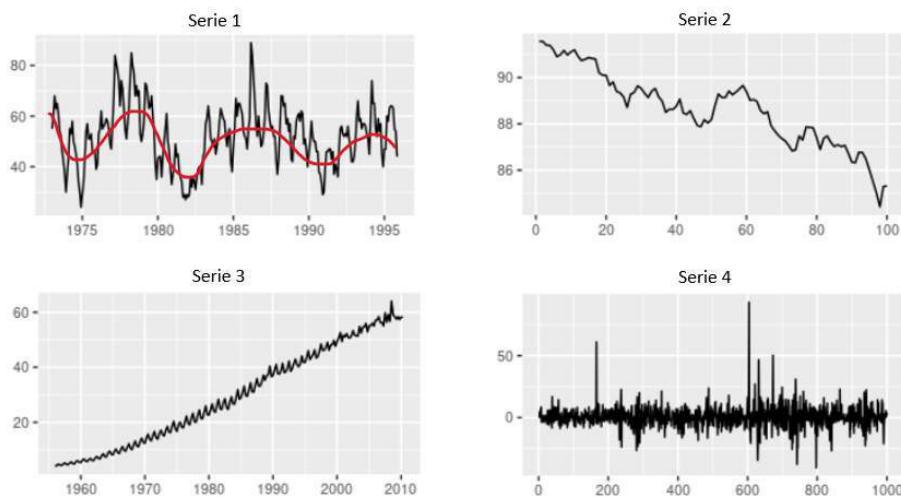
Trata de responder a las siguientes preguntas antes de leer los siguientes párrafos, donde estarán las respuestas:

- ¿Qué series tienen tendencia?
- ¿Qué series tienen estacionalidad?
- ¿Qué series tienen ciclos?

Tanto la serie 2 como la serie 3 tienen una fuerte tendencia, mientras que la primera es negativa, la segunda es positiva.



La única serie que presenta ciclos es la 1, cuyo patrón cíclico podemos ver en el siguiente gráfico en rojo (fíjate en cómo cada una de las fluctuaciones no tiene la misma duración):



Y por último, las series que tienen un patrón estacional son la 1 y la 3, donde podemos ver cómo hay un patrón que se va repitiendo en el dibujo a lo largo de la serie.

Atípicos

El concepto de atípico no es muy diferente del que vimos en los modelos de segmentación.

Hay momentos en los que nuestras series temporales toman valores muy distintos con respecto al nivel de la serie. Estos valores atípicos suelen deberse a situaciones muy excepcionales que afectan a nuestro negocio. Por ejemplo:

Ejemplos de causas de atípicos

Caídas de Web

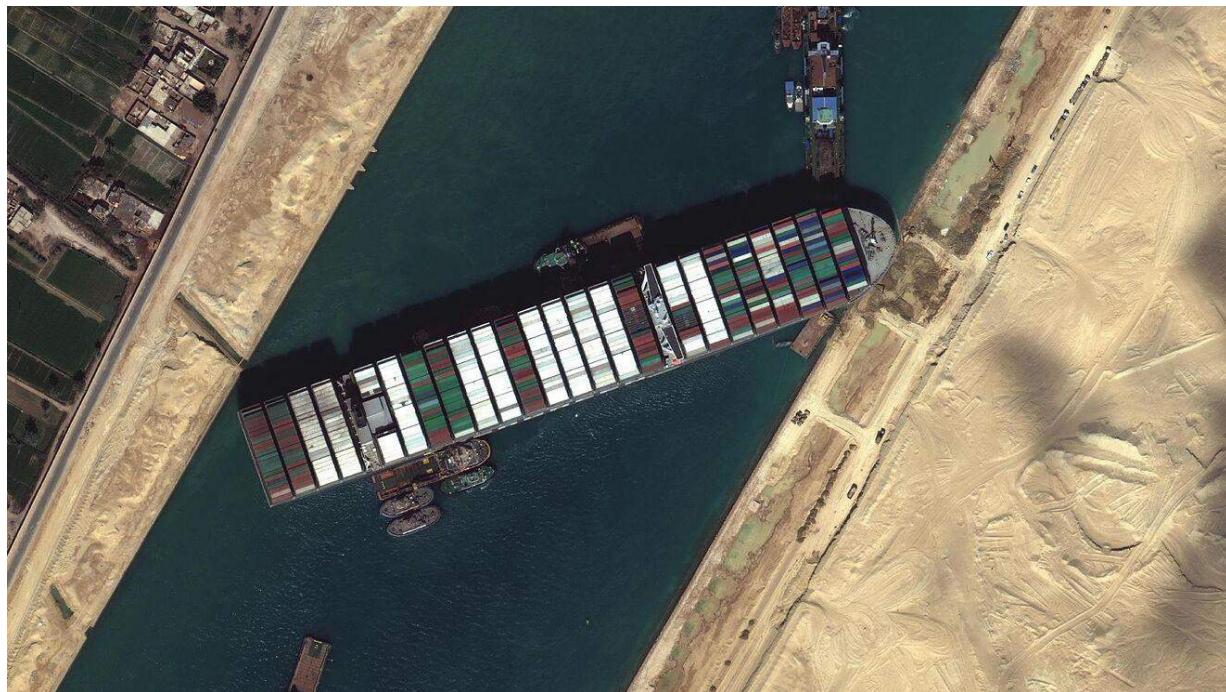
Festivos

Conflictos sociopolíticos

Desastres Naturales

Crisis económicas

Etc.

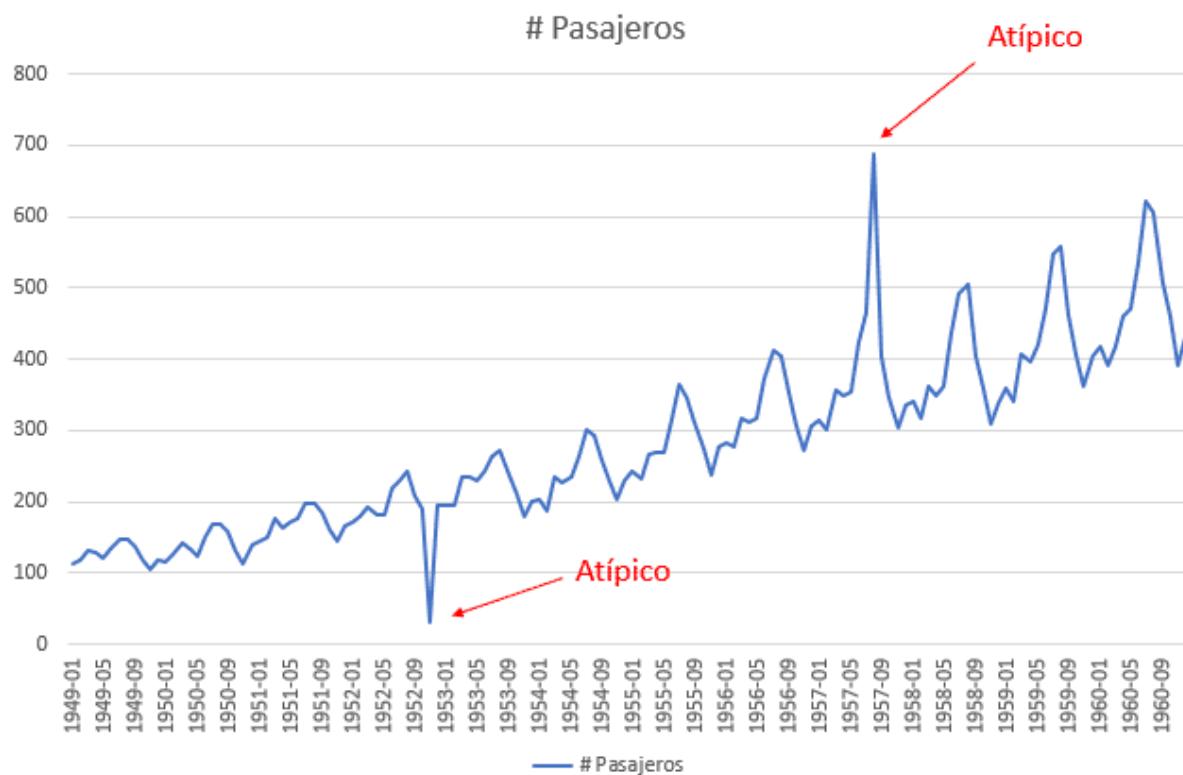


El pasado mes de marzo de 2021 todos tuvimos noticia del supercarguero que se quedó encallado en el canal de Suez bloqueando el transporte de mercancía a nivel mundial.

Si nuestra compañía requiere de este transporte nos podremos imaginar que la serie de ventas sufrirá una caída pronunciada en el periodo de tiempo comprendido hasta que se desencalle el barco.

Los días que son festivos y nuestras tiendas cierran, naturalmente, tendremos cero ventas.

Reutilizando el ejemplo del número de pasajeros en una aerolínea, podría tener atípicos como los siguientes:



Como vemos, los atípicos pueden tener valores muy por encima de la serie o muy por debajo.

En el gráfico anterior podemos ver que en noviembre de 1952 hay un atípico en el número de pasajeros muy inferior al que esperaríamos de la serie. Esto podría explicarse por un gran temporal que obligó a la cancelación de muchos vuelos en dicho mes.

Por otro lado, el segundo atípico se produce en agosto de 1957, donde la serie toma unos valores muy por encima de lo esperado, que podría deberse, por ejemplo, al éxito de una campaña comercial para volar a ver los juegos olímpicos que se producen en otro país vecino.

Antes de utilizar cualquier modelo es necesario detectar estos atípicos y tratarlos, ya que pueden hacer mucho daño a los resultados de nuestros modelos. El qué hacer con los atípicos lo veremos en el fastbook de prácticas.

Resumen y conclusiones

X Edix Educación

En este fastbook hemos tenido nuestro primer contacto con el mundo de las series temporales. Hemos visto **cómo las series temporales aparecen en las compañías de la mano de los KPIs**, esas métricas/números grandes sobre los que solemos hacer reportes y analizar mes a mes para conocer cómo va nuestro negocio.

Como no solo nos importa cómo están las cosas sino también cómo van a estar, las empresas recogen KPIs durante grandes períodos de tiempo para formar el elemento central de lo que nos queda de asignatura: las series temporales. Estas nos servirán para analizar **qué ha ocurrido** y nos darán información para que podamos predecir **qué va a ocurrir en el futuro**.

En los siguientes fastbooks veremos modelos de series temporales y empezaremos a modelizar, pero para poder entender los modelos más complejos hemos visto también los conceptos más importantes dentro de este campo: **granularidad, tendencia y atípicos**.

Prepárate porque vamos a dedicar múltiples ejemplos para saber cómo vamos a trabajar con series temporales usando el ordenador, de modo que nos entretendremos de forma práctica. Suficiente teoría hemos tenido hasta el momento, ¿cierto?

¡Enhорabuena! Fastbook superado

edix

Creamos Digital Workers