

# Módulo 1. La analítica

La analítica es el estudio y el análisis sistemático de datos y estadísticas para aplicarlos en un proceso de toma de decisiones. A estas decisiones se arriba tras detectar patrones y desarrollar conclusiones sobre la información obtenida.

En el presente módulo, nos centraremos en este proceso y sobre la manera de trabajar en reportes de analítica web. Es de suma importancia que comprendamos de qué se trata la analítica para luego sacarle el mayor provecho en nuestros proyectos, compañías e instituciones.

En este módulo, veremos cuestiones generales de la analítica e incluso algunas que exceden al alcance regular de la analítica web. Por supuesto, nos centraremos en la analítica aplicada a negocios e instituciones que cuentan con la necesidad de tomar decisiones para mejorar sus rendimientos del sitio web.

**Figura 1: Tipos de analítica de negocios**



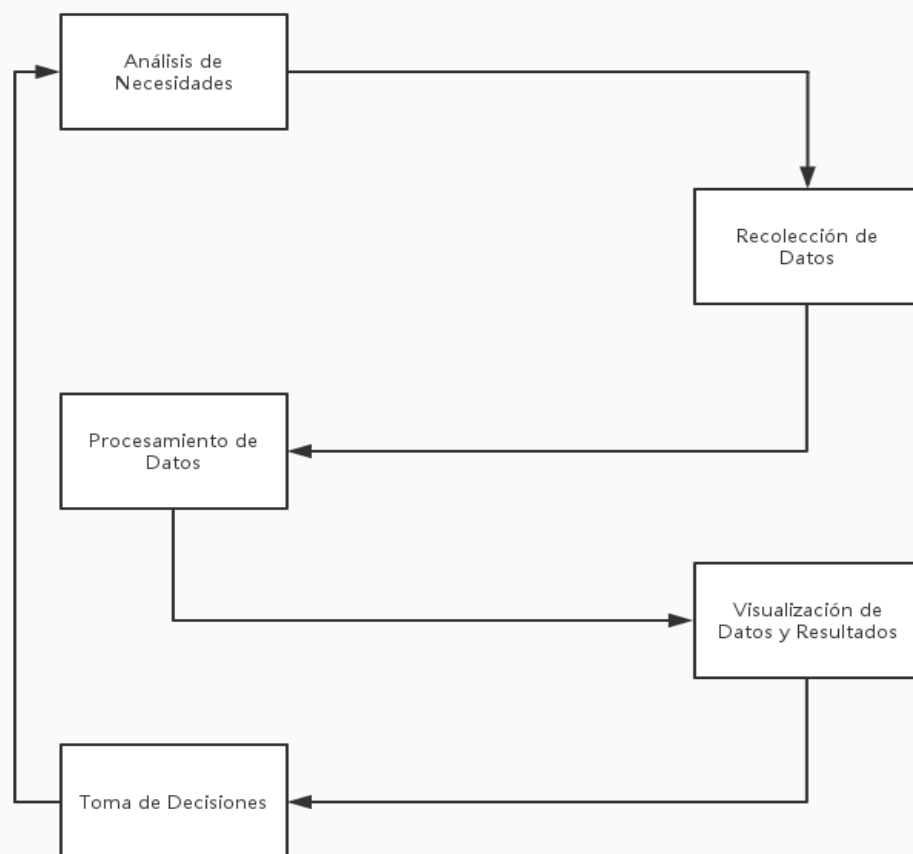
Fuente: elaboración propia.

- **Analítica descriptiva:** permite describir lo que está sucediendo en nuestro negocio para entender lo que está pasando o lo que ha pasado.
- **Analítica diagnóstica:** se enfoca en la *performance* pasada para comprender qué sucedió y por qué.
- **Analítica predictiva:** pone su énfasis en predecir comportamientos futuros o situaciones que puedan pasar basándose en la aplicación de modelos estadísticos.
- **Analítica prescriptiva:** es una rama de la analítica predictiva que se basa en recomendar un curso de acción a tomar en función de los datos analizados.

## Unidad 1.1. El proceso de la analítica

Al decidir trabajar con analítica, estamos estableciendo que una de nuestras prioridades será el estudio minucioso de lo que sucede en nuestra página web. Esto nos servirá para tomar decisiones que impacten en nuestro negocio de manera directa o indirecta y nos permitan obtener alguna mejora en los procesos o en los resultados. Es por ello que la analítica en sí misma no sirve de nada si no ayuda a este proceso. Debemos trabajar la analítica como un proceso: desde la recolección de los datos hasta la toma de decisiones.

**Figura 2: El proceso de la analítica**



Fuente: Paz, 2016, p. 6.

Para lograr implementar un proceso de mejora basado en analítica (un programa de entrenamiento, por ejemplo), se debe atender a los siguientes pasos:

- **Análisis de necesidades:** se comienza detectando qué es lo que se quiere medir, cómo se lo quiere medir y qué información esperamos que nos brinde este proceso. Siempre pensando en que nos debe permitir tomar decisiones.
- **Recolección de datos:** una vez que sabemos qué información y datos recabar, debemos comprender cómo se consiguen esos datos y cómo se confirma su validez.

- **Procesamiento de datos:** los datos obtenidos no se pueden comprender si no los procesamos apropiadamente. Aquí entran en juego las herramientas matemáticas y estadísticas para trabajar.
- **Visualización de datos y resultados:** teniendo los datos procesados, es hora de presentarlos de una manera que nos permita comprenderlos: hay que llevar lo complejo a lo sencillo como base para la toma de decisiones.
- **Toma de decisiones:** es la razón por la que hicimos todo el proceso. Este punto del proceso nos permite tomar alguna decisión para mejorar y lograr mejores resultados. Además, es el comienzo del ciclo una vez más, sentando las bases para un nuevo análisis de necesidades.

En definitiva, debemos tener en claro que “todos los procesos de analítica están orientados a poder tomar decisiones que permitan modificar algo en la manera que realizamos las cosas para obtener mejores resultados” (Paz, 2016, p. 89).

### 1.1.1. ¿Para qué medir?

Para intentar responder este interrogante, podemos revisar una definición sencilla de analítica. "La analítica es el estudio de datos históricos para investigar posibles tendencias, analizar efectos de ciertas decisiones o eventos, evaluar la *performance* de una herramienta o actividad en un escenario particular" (Paz, 2016, p. 21). En definitiva, comenzar a delinear todo este proceso de la analítica implica: planificar, recolectar, analizar y reportar como base para la toma de decisiones. Medimos para tomar decisiones.

Profundizando un poco más en esta definición de analítica, podemos decir que implica la captación y descubrimiento, así como la interpretación de diversos patrones de conducta o resultado en un set particular de datos. Cuando hablamos de *analytics*, hacemos referencia a una acción multidisciplinar que involucra trabajos de minería y ciencia de datos, estrategia, estadística y conocimiento específico de la disciplina sobre la que se realizará la analítica. En nuestro caso, el sitio web.

El proceso de la analítica que mencionamos implica comenzar definiendo lo que es el éxito; es decir un claro objetivo investigativo que queremos obtener a partir de nuestro análisis (plan); luego medir y captar los datos que hacen falta para sentar las bases de las respuestas a las preguntas clave que se hicieron en el plan (recolección); monitorear y aprender de los datos recolectados, lograr que el bloque de datos con el que contamos sean prolijo y nos ofrezca datos libres de errores, y realizar los cruzamientos de datos necesarios (análisis); comunicar los resultados del análisis de una manera que permita tomar decisiones. De nada sirven los datos por sí mismos si no nos "cuentan una historia" que nos mueva a resolver situaciones a partir de decisiones (reporte). (Paz, 2016, p. 23).

Estamos en la era del *big data* donde detectar y procesar enormes cantidades de información puede llevar a innovaciones constantes y a optimizaciones de procesos. La clave pasa por comprender esos datos en su agregado, qué significan y cómo tomar decisiones a partir de ellos.

En otros tiempos, la información captada era posible manejarla de manera manual y procesarla mentalmente como información sencilla.

Existen dos grandes tendencias del uso de datos: la analítica y la ciencia de datos. La analítica se basa en investigar y extraer información valiosa de conjuntos de fuentes estructuradas y no estructuradas para explicar fenómenos históricos o presentes, predecir *performances* futuras y determinar los mejores modelos analíticos a utilizar para presentar y explicar soluciones posibles a problemas que ayuden a la toma de decisiones. Sobre esta línea es sobre la que trabajamos en este curso.

La ciencia de datos busca diseñar, desarrollar e implementar algoritmos a través de programación estadística que soporte la generación y el mantenimiento de herramientas para la toma de decisiones. Maneja grandes cantidades de datos y crea visualizaciones para su entendimiento. Mencionaremos algunas cuestiones sobre esta disciplina, pero no será nuestro foco.

Si volvemos al interrogante sobre para qué medir, podríamos articular cualquiera de los siguientes motivos y podríamos tener razón. Si los combinamos, tendremos la base de nuestro plan de toma de decisiones.

Algunos de los principales usos de la analítica son:

- Instigar el cambio al transformar datos en un insumo para la toma de decisiones.
- Descubrir oportunidades en conjuntos de datos a partir de poder darles sentido.
- Definir claramente problemas y proponer espacios para la búsqueda de soluciones.
- Traducir análisis estadísticos en inteligencia institucional para mejorar la *performance*.
- Crear un marco de estudio para la solución de situaciones críticas de una disciplina.
- Investigar, interpretar y visualizar datos "crudos" utilizando análisis predictivo, prescriptivo y descriptivo para que sean más fáciles de comprender.
- Manejar datos flexibles con múltiples variables a través de métodos como el análisis lineal discriminante y la selección de regresión multilínea.
- Comprender, integrar y prescribir modelos de soluciones a problemas comunes.

### 1.1.2. ¿Qué medir?

La respuesta más clara y sencilla a esta pregunta es que mediremos diferentes métricas a través de la captación y la generación de datos. Por ello, se dispara la siguiente pregunta de manera natural: ¿qué es un dato?

De la misma manera que para medir distancias utilizamos una unidad mínima y sobre ella luego vamos acumulando, en el caso de la analítica comenzamos con una unidad básica: "el dato". "Un dato se conforma como la representación simbólica (dicho símbolo puede ser representado por un número, por una palabra, de manera algorítmica, espacial) identificable y medible de un atributo cualquiera o de una variable cualquiera tanto cualitativa como cuantitativa" (Paz, 2016, p. 27). Por ejemplo, nombres y apellidos forman parte de nuestros "datos personales", ya que son una representación simbólica alfabética de un atributo de nuestra persona. Igualmente, nuestro número de pasaporte es un dato que nos identifica, pero de

manera numérica. Nuestro peso o nuestra altura, también son datos, pues representan variables que nos identifican.

Teniendo en cuenta esta definición de dato, inmediatamente podemos inferir que la cantidad de datos posibles que pueden captarse es infinita. Por ende, se vuelve más y más importante saber cuáles son los datos relevantes que debemos analizar para lo que necesitamos comprender. Por esta razón surge la minería de datos o el proceso de extracción de los datos. “La minería de datos se encarga de extraer, transformar y cargar datos dentro de nuestra base de datos. De esta manera, necesitamos de antemano generar una base de datos apropiada que pueda contener la información captada” (Paz, 2016, p. 31).

Captar solamente la información no alcanza, ya que debemos poder interactuar con ella y encontrarle sentido a través de herramientas y técnicas de análisis de datos. Además, debemos presentarlas en un formato útil. Por ejemplo, imaginemos un gráfico de torta que nos explica el fenómeno que queremos analizar de manera sencilla (existen múltiples maneras de mostrar la información).

Existen tres grandes tipos de datos desde el punto de vista del modo de recolección que hará falta trabajar.

**Datos estructurados:** los datos residen en campos fijos dentro de un mismo archivo y siempre se muestran del mismo modo. Es lo que tradicionalmente denominamos "base de datos". Puede estar en un documento de Excel, por ejemplo, con diversas columnas que representan variables y líneas que representan los datos de tales variables para cada caso. Imaginemos una planilla de clientes: es el ejemplo claro de los datos estructurados. Nos indica por cada cliente diferentes variables representadas en datos: nombre, apellido, teléfono de contacto, correo electrónico, domicilio, método de pago asociado, etcétera.

**Datos no estructurados:** son datos que no residen en locaciones fijas. Normalmente, están dentro de textos de formato libre y son ubicuos, pues pueden estar en cualquier lugar y de cualquier manera. Requieren un trabajo mayor de captación, extracción, curado y procesado de esos datos para lograr representatividad. Algunos ejemplos que podemos mencionar de datos no estructurados son variables que se encuentran en *tweets* de los usuarios o en páginas

de medios periodísticos de opinión que nos pueden ayudar a comprender principalmente variables cualitativas como sentimientos, valores asociados, etcétera.

**Datos semiestructurados:** se encuentran a mitad de camino entre ambas opciones, ya que sirven para darle algún tipo de estructura asociada a los datos no estructurados. Son ejemplos las etiquetas que se pueden agregar a los documentos o *hashtags* de Twitter. Estos elementos pueden ayudar a que organicemos los datos de una manera lógica. (Paz, 2016, p. 41).

Para que nuestro proyecto de analítica funcione apropiadamente, nuestros datos deben tener integridad. Este es un concepto que hace referencia a tres aspectos sobre los datos: deben ser completos, consistentes y exactos.

**Completos:** que los datos sean completos implica que el dato en sí mismo sea una entidad. Para ejemplificar claramente, si estuviéramos analizando la repetición de nombres entre los que realizan este curso, requerimos contar con los nombres completos de todos; no alcanza con obtener las iniciales.

**Consistentes:** que los datos sean consistentes quiere decir que solamente están tomados de una manera en la que puedan ser comparados. En el ejemplo anterior, si lo que queremos es analizar la repetición de los nombres, de nada nos sirve contar con algunos nombres y con algunos números de pasaporte.

**Exactos:** el dato debe haber sido captado de la fuente original o de una copia fiel, debe ser legible, atribuible a la variable correcta y estar captado de manera aceptable. Imaginemos en el ejemplo anterior que mi nombre estuviera cargado como "Gustavo" en lugar de "Guillermo" (es un dato erróneo) o como "Guilemo" (está captado de manera inaceptable y debe ser curado) o que en la planilla donde está anotado no pueda alcanzarse a distinguirse el nombre (ilegible).

La metadata también está requerida para que la integridad del dato sea completa. La metadata es la información contextual requerida para entender los datos. Un valor de un dato en sí mismo es insignificante si no hay información sobre ese dato. Podríamos decir que la metadata está conformada por los datos acerca de los datos. Es información estructurada que describe, explica o hace más fácil de utilizar o manejar los datos. Para clarificar de qué se trata: imaginemos que tenemos el número "23" como dato. ¿A qué se

refiere? No lo sabemos; por ello requerimos metadatos que nos explique qué quiere decir ese 23: ¿se trata del número de la camiseta del jugador? (¿los datos hacen referencia a Michael Jordan?), ¿se trata de la cantidad de puntos que hizo en el partido?, ¿se trata de los minutos que jugó? La metadatos de cada dato puede ser múltiple y servir para mejorar la manera de interactuar con nuestra base de datos.

Cuando los datos no poseen integridad se dice que nuestra base de datos "está sucia" y para ello necesitamos limpiarla. Limpiar una base de datos implica aplicar procesos de curado y validación de datos, así como la desestimación de los datos irrecuperables, completar los datos incompletos, quitar los datos duplicados, etcétera. (Paz, 2016, p. 37).

Existen muchos tipos de datos y múltiples maneras de recolectar datos para responder a las preguntas que surgen de nuestros problemas. Las técnicas de recolección de datos que utilicemos pueden producir información numérica en datos cuantitativos o pueden funcionar de modo ilustrativo como los datos cualitativos. También pueden incluir una combinación de ambas opciones y trabajar con datos mixtos. Por supuesto, determinar qué tipo de dato necesitaremos para responder a las preguntas planteadas en nuestro problema, nos dirá cuál es la técnica que necesitaremos usar.

Podemos trabajar los datos de diferentes maneras. Observar los valores de las variables para el fenómeno que estamos analizando es la manera de captar o recolectar datos. Cada dato individual se denomina "observación" y la colección de las observaciones realizadas son nuestro set de datos (los valores de las variables obtenidas para una muestra de unidades) o matriz de datos (donde los valores de cada variable particular se organizan dentro de una misma columna y los valores de las variables forman las columnas de la matriz de datos).

- **Datos cuantitativos:** requieren uso de análisis estadístico. Las variables pueden ser identificadas y sus relaciones medidas. Se cuentan o expresan de manera numérica.
- **Datos cualitativos:** examinan datos no numéricos en busca de patrones y significados. Son recolectados y analizados con un mayor grado de subjetividad.

- **Datos mixtos:** pueden explicar algunos resultados inesperados (los denominados *outliers* o excepciones).

### 1.1.3. ¿Cómo medir?

Si ya decidimos qué tipo de datos debemos captar, cuáles son las variables sobre las que queremos generar interrogantes y cuál es la población o muestra sobre la que trabajaremos, llega la hora de captar los datos específicamente. Para ello, debemos definir la técnica en particular de recolección de datos que nos resulta de importancia y el lugar y momento para hacerlo.

“La observación es el chequeo del dato para una variable en particular, en una situación y un momento” (Paz, 2016, p. 41).

Por ejemplo, si estamos analizando el desarrollo de un cliente en su crecimiento en la relación con nosotros, la variable "cantidad de compras" se modificará con las distintas observaciones en el tiempo. Si tenemos un set de datos contra el que cotejarlos, podemos inferir su ritmo de crecimiento similar a un porcentaje de los clientes de diferente importancia relativa en nuestro negocio.

Para lograr responder a esa pregunta, recolectaremos los datos de diferentes maneras: para obtener el set de datos que corresponde al sujeto que estamos analizando, por ejemplo, será una observación directa en nuestra línea de cajas donde se registra la compra por parte del cliente. Ahora bien, quizá tenemos la información sobre el set de datos que utilizaremos como grupo de control en una base de datos *online* sobre la que hay que realizar consultas de minería de datos para poder hacer las inferencias finales. Como vemos, aún en casos tan sencillos, debemos interactuar de diferentes maneras con los datos. Es por esto que podemos definir múltiples maneras de recolectar los datos entre los que se destacan:

- **Observación directa:** seremos nosotros quienes haremos cada detección de la repetición de un dato que se da en la realidad a la cual tenemos acceso.
- **Observación mediada por tecnología:** se toman los datos de parte de cada sujeto de manera individual, pero sobre variables de dificultosa observación directa.
- **Minería de datos:** implica detectar en fuentes de datos existentes ciertos patrones que nos servirán para utilizar en

nuestros análisis. Por lo general, fueron generados por observación directa o mediada por tecnología por un tercero que tiene las bases de datos accesibles a consulta o como bases de datos propias que hemos desarrollado en el tiempo. En sus fundamentos, están la ley de los grandes números y la tendencia de uso del *big data*. (Paz, 2016, p. 41).

A continuación, profundizaremos en los métodos de recolección de datos para hacer un estudio.

Existen diversos métodos y también existen métodos mixtos que toman parte de lo que realizan cada uno de los que mencionaremos a continuación:

- **Censo:** es un tipo de estudio que obtiene los mismos datos de cada uno de los miembros de una población. Su principal beneficio es que no requiere inferir generalizaciones, ya que toma toda la población. Su principal contra es que lleva mucho más tiempo y más costo que una investigación sobre una muestra. Por supuesto que mientras la población sea más pequeña o más fácil sea la implementación tecnológica, mayores posibilidades de trabajar censos tendremos.
- **Colección de set de datos en muestra:** un estudio realizado a través de una muestra obtiene datos de un segmento de la población. Busca poder estimar atributos de la población como un todo a través de inferencias. Existen dos tipos de muestras:
  - **Las muestras probabilísticas:** cada elemento de la población tiene una chance cierta y conocida (distinta a cero) de ser elegido para la muestra.
  - **Las muestras no probabilísticas:** no sabemos las chances que cada elemento de la población tiene de ser elegido para la muestra o existe algún elemento con cero chances de ser elegido. Por supuesto que las muestras no probabilísticas tienen dos grandes ventajas: conveniencia y costo. Su gran desventaja pasa porque se vuelve más difícil extender las inferencias.

**Experimentos:** es un estudio controlado en el que el investigador intenta entender relaciones del tipo causa-efecto. Se denomina "controlado" porque quien investiga controla cómo se asignan los sujetos a los diferentes grupos y qué variables modifica en cada grupo para analizar los resultados. Se realizan a través de las relaciones entre variables independientes y variables dependientes para lograr detectar las relaciones causa-efecto.

**Estudio observacional:** similar al anterior, pero, en este caso, los investigadores no tienen control sobre los dos elementos mencionados en los experimentos. (Paz, 2016, p. 48).

Una vez que tenemos claro cómo será el método o los métodos que utilizaremos para recolectar los datos, debemos plantearnos cómo los almacenaremos. Es decir, cómo haremos el seguimiento de las variables. Si nuestra investigación será extendida en el tiempo, se vuelve un aspecto crítico de nuestro proceso.

El principal tema que hay que tener en cuenta cuando hablamos de almacenamiento de datos está dado por el mantenimiento de los mismos. Imaginemos la siguiente situación: nos hemos ido de viaje con nuestra cámara de fotos, sacamos cientos de imágenes de las vacaciones y antes de trasladarlas a nuestra computadora, perdemos la cámara. Seguramente, nunca podremos volver a ver esas fotografías y nos resulte en una pérdida de valor emocional. Desde el punto de vista de los estudios estadísticos, perder datos hace que se pierdan referencias históricas, que las inferencias que se pueden hacer no sean igualmente generalizables y que se pierda el valor comparativo. (Paz, 2016, p. 51).

Por esto es que la primera regla del almacenamiento de los datos tiene que ver con la seguridad de su alojamiento. La seguridad que tengamos de que los datos pueden preservarse en el tiempo ayudará a que el análisis y el seguimiento que realicemos nos permita ver la evolución y nos permita entender las tendencias.

La capacidad es el segundo punto. Siendo que cada vez recolectamos más datos, lo que debemos tener es no solamente un lugar seguro para almacenarlos, sino un lugar lo suficientemente grande para hacerlo.

Imaginemos una biblioteca: si cuenta con 10 000 volúmenes, deberá tener suficientes anaqueles para soportarlos. La cantidad de metros cuadrados disponibles será

fundamental. Si esa misma biblioteca realiza una compra de otros 10 000 volúmenes, deberá duplicar el espacio de alojamiento o deberá deshacerse de los volúmenes anteriores. Ahora bien, extendamos este ejemplo a los datos que tenemos guardados. Si los anotamos en un cuaderno, tendremos una cierta cantidad de páginas. Si los guardamos en una planilla de Excel, tendremos mucho más espacio y así podemos ir creciendo en cantidades de datos de acuerdo al tipo de alojamiento que utilicemos.

Luego, podemos analizar la seguridad de los datos desde el punto de vista de la privacidad. Es importante que solamente puedan acceder a los datos quienes deben hacerlo y mantenerlos alejados y a resguardo de quienes no deban tener acceso. Por supuesto, eso dependerá del tipo de dato. En algunos casos, será más importante la difusión de los datos antes que su privacidad (por ejemplo, datos de *performance* que nos permitan escalar en un *ranking*). En otros casos, serán datos sensibles que debemos preservar bajo privacidad (por ejemplo, datos de métodos de pago de nuestros clientes).

A continuación, debemos pensar en temas de productividad. Dependiendo qué tanto utilicemos los datos, podemos mejorar mucho nuestros tiempos de trabajo y productividad. Hay que tener más a mano los datos a los que debemos acceder más seguido y más alejados los datos que utilizamos eventualmente.

Cuando nos referimos a almacenamiento de datos, podemos mencionar algunos conceptos principales para lograr su preservación:

- **Archivo:** los archivos son sets de datos que tenemos guardados y que son los que debemos proteger ya que son la información que requeriremos para trabajar constantemente. En este caso, lo más importante es la capacidad de nuestro archivo; es decir, cuántos datos podemos alojar.
- **Backup:** el *backup* o respaldo de datos también es una colección de datos, pero, a diferencia del archivo, su idea no es la de preservar los datos en el tiempo, sino ofrecer protección temporal de los datos contra la corrupción o borrados accidentales y ofrecer una solución ante un inconveniente del archivo principal. Normalmente, su uso es cíclico, actualizándose sobre sí mismo y haciendo que un respaldo suplante al respaldo anterior. (Paz, 2016, p. 60).

Normalmente, se puede hacer una diferencia entre archivo profundo y archivo activo. En el caso de los archivos profundos, su intención es la de resguardar datos históricos y preservarlos por largo tiempo. En cambio, el archivo activo busca preservar datos que se usen constantemente para consulta.

A la hora de decidir qué tipo de almacenamiento será el mejor para nuestros datos, debemos respondernos a conciencia tres preguntas:

- ¿Cuántos datos estaremos alojando?
- ¿Cuánto tiempo deben estar alojados?
- ¿Con qué frecuencia se accederá a los datos?

Podemos establecer nuestras prioridades tomando como foco las respuestas que damos, sabiendo si será más importante el espacio que la funcionalidad o viceversa, entre otras posibilidades.

#### 1.1.4. El análisis de la información

Habiendo decidido cómo serán los datos que tomaremos para responder a las necesidades que tiene nuestro problema, debemos enfocarnos en la etapa del análisis en sí para poder luego realizar la descripción de los resultados y la presentación de reportes e inferencias hacia las conclusiones que se logren realizar.

Consolidar los datos habla de la necesidad de lograr consistencia en nuestros conjuntos de datos y lograr que la información que nos entregan tales datos sea válida.

Para consolidar los datos tenemos dos cuestiones principales para trabajar:

- La manera de ingreso de los datos.
- El formato de salida de los datos.

Sobre la manera de ingreso de los datos debemos comprender que, dependiendo del método de recolección que hayamos decidido tomar, cambiarán los requerimientos.

Cuando hablamos de **muestras no probabilísticas**, nos referimos a dos tipos de métodos de obtención de la muestra:

- **Muestra voluntaria:** está compuesta por sujetos que de manera voluntaria se autoseleccionan para participar del estudio.
- **Muestra conveniente:** está compuesta de sujetos que son fáciles de alcanzar.

Entre los **métodos probabilísticos** podemos encontrar cinco opciones de muestreo:

- **Muestreo aleatorio simple:** en todos los casos en los que existe n cantidad de sujetos (u objetos), y la muestra consistirá en n cantidad de sujetos (u objetos) y todos los n tienen iguales chances de suceder, entonces, el muestreo se denomina muestreo aleatorio simple.
- **Muestreo aleatorio sistemático:** en este caso, a diferencia del muestreo simple, generamos la lista de nuestra población y asignamos un número a cada sujeto. Luego, se selecciona un sujeto dentro de cierta cantidad de sujetos de la lista. Finalmente, se selecciona el mismo número de cada bloque de sujetos de la lista.
- **Muestreo estratificado:** en este método, la población se divide en grupos según alguna característica. Luego, dentro de cada grupo, se selecciona una muestra probabilística aleatoria simple. Es importante que todos los estratos estén representados en la muestra.
- **Muestreo por clústeres:** a diferencia del muestreo por estrato, en el muestreo por clústeres se asignan los sujetos en grupos y se selecciona de manera aleatoria un grupo. Los miembros de ese grupo serán los analizados.
- **Muestreo por múltiples etapas:** se trata de estudios que requieren diferentes etapas de análisis. Cada etapa puede recurrir a un método de muestreo diferente.

Cuando trabajamos los datos a través de muestras, debemos comprender que existe siempre lo que se denomina "parcialidad de la muestra" o *statistical bias*. Esto es una tendencia de la muestra a sobreestimar o subestimar los parámetros de una población. Para que nuestros análisis sean de calidad, debemos trabajar con muestras que sean representativas; esto quiere decir que la muestra represente los atributos de la población en un número aceptable.

La parcialidad se da cuando la muestra no representa adecuadamente la población. Si el error de la muestra sucede en el momento de la captación de los datos, se denomina parcialidad de selección y puede darse por:

- **Subcobertura:** sucede cuando algunos miembros de la población están representados inadecuadamente en la muestra. Suele ser el problema de las muestras por conveniencia.
- **Parcialidad por no respuesta:** en algunos casos, los individuos que buscamos para nuestra muestra no pueden participar del estudio. Esto puede producir que nuestros resultados difieran mucho entre los que responden y los que fueron seleccionados para la muestra, pero no responden.
- **Parcialidad por respuesta voluntaria:** sucede cuando los sujetos participantes de la muestra lo hacen de manera voluntaria por un interés particular. Es la contracara del error de parcialidad anterior. (Paz, 2016, p. 71).

En el momento de medir, los errores de parcialidad pueden darse también por problemas de medición y no de muestreo. Son los denominados errores por parcialidad en la respuesta (*response bias*):

- **Preguntas tendenciosas:** ocurre cuando la pregunta está realizada para favorecer alguna de las opciones. Imaginemos que queremos consultar sobre la opinión de nuestra afición sobre nuestro entrenador y en las preguntas ponemos las opciones: satisfecho, insatisfecho y muy insatisfecho. En este caso, estaremos dando dos opciones para demostrar niveles de insatisfacción y solamente uno para demostrar satisfacción. Por ende, estamos tendiendo a generar datos de opinión negativa.
- **Aceptación social:** en algunos casos, podemos estar preguntando cuestiones que hacen que el respondiente quede mal ante el investigador; especialmente, en investigaciones sin confidencialidad.
- **Errores humanos o técnicos de captación:** también puede darse que alguno de los aparatos tecnológicos que utilizamos para captar los datos deje de funcionar y, por ende, nos hagamos de datos

parciales. También puede suceder por errores humanos, por ejemplo, en la confección de las planillas. (Paz, 2016, p. 71).

Una vez que tenemos los datos, debemos integrar nuestra base de datos y validar que lo que aparece allí está correctamente indizado. Asimismo, si trabajamos con distintas fuentes debemos evitar la duplicación de datos.

Siempre debemos recordar que no podemos inferir conclusiones aceptables si los datos de los que partimos no son confiables.

Ya tenemos los datos en condiciones de ser analizados, ¿cómo debemos proseguir?

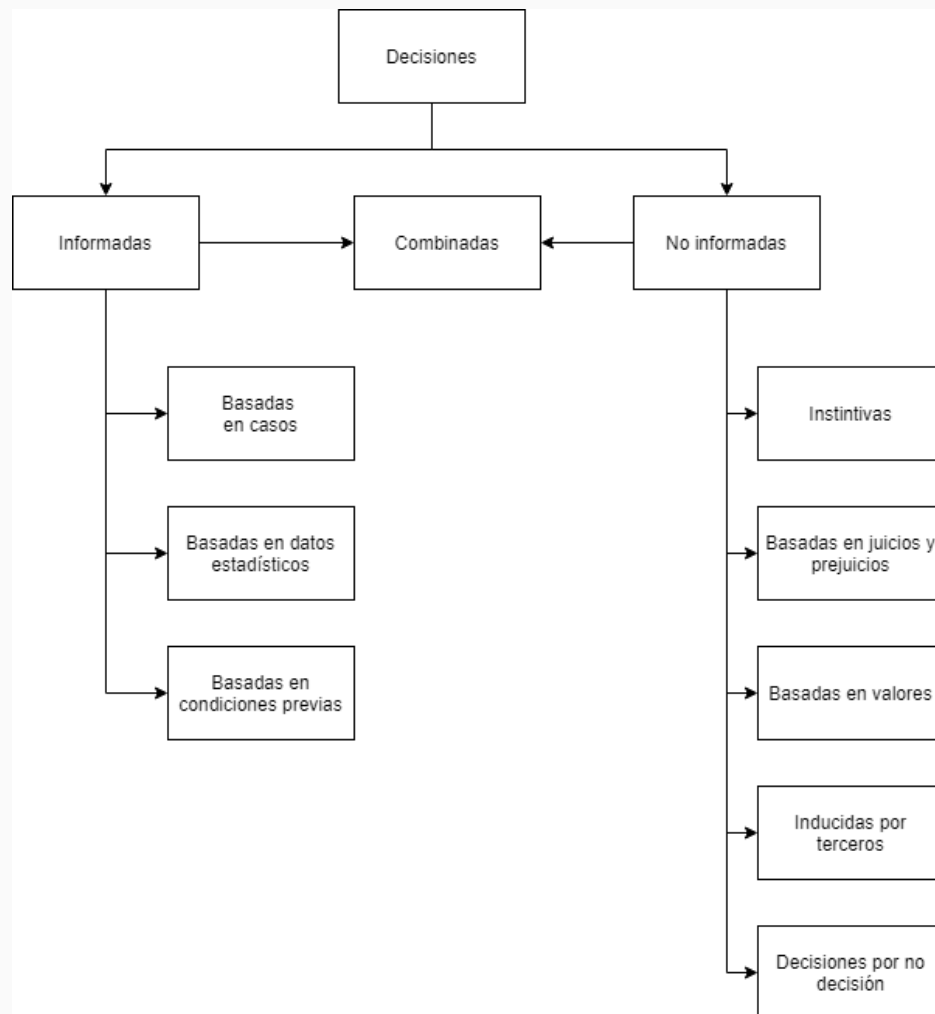
La mejor manera de comunicar nuestros resultados es a través de la simplificación de la información para que quede comprensible y se puedan tomar decisiones sobre lo analizado. Debemos intentar pensar de manera visual al generar nuestros reportes. Para ello, existen múltiples métodos. Los más utilizados son:

- **Uso de cuadros y diagramas:** para explicar un fenómeno basado en datos numéricos, utilizamos gráficos, diagramas, cuadros, etcétera. Algunas de las opciones son: gráficos de tortas, diagramas de Venn, gráficos de burbuja, gráficos de área, gráficos de línea, diagramas de Gantt, diagramas de flujo, árboles de decisión, líneas de tiempo, etcétera.
- **A través de analogías abstractas:** pirámides, embudos, escaleras, velocímetro, balanzas, mapas, anatomías, árboles genealógicos, etcétera.
- **A través de analogías directas:** iceberg, montaña, sándwich, reloj de pared, efecto dominó, etcétera.
- **A través de alegorías:** el campo de juego, la vida en la ciudad, la evolución, la maquinaria, la caja de herramientas, etcétera. (Paz, 2016, p. 82).

## Unidad 1.2. La toma de decisiones basada en datos

Como explicitamos en la unidad anterior, nuestra intención a la hora de iniciar un proceso de analítica está dada en la necesidad de tomar decisiones informadas. Esto es fundamental para tener la suficiente confianza y el respaldo necesario para tomar y defender una decisión.

**Figura 3: Tipos de decisiones**



Fuente: elaboración propia.

Existen diferentes maneras para arribar a las decisiones:

- **Decisiones informadas:** decisiones fundamentadas en el procesamiento de información y de datos basados en analítica. Podríamos decir que son decisiones de base racional. Encontramos distintos tipos de decisiones informadas.
  - Decisiones basadas en casos.
  - Decisiones basadas en datos estadísticos.
  - Decisiones basadas en condiciones previas y disparadores de situación.
- **Decisiones no informadas:** no son necesariamente decisiones que no tienen sustento, sino que su sustento está dado principalmente en elementos que no tienen que ver con información procesada siguiendo el proceso de la analítica. Por ello, decimos que no están basadas en información, sino quizás en elementos sueltos, datos aleatorios, etcétera.
- **Decisiones instintivas:** se dan por reacción a un estímulo.
- **Basadas en juicios o prejuicios:** muchas veces tomamos decisiones en función del interlocutor al que nos enfrentamos y por encima de los datos que se nos presentan.
- **Basadas en valores:** existen situaciones en que nuestros valores toman preferencia por sobre la información. No tomaríamos una decisión de optimización de venta, si hacerlo implicara romper con nuestros propios valores.
- **Inducidas por terceros:** existen ocasiones en que tercerizamos nuestro proceso de decisión. Ya sea ante alguien que toma las decisiones de manera informada o no.
- **Decisiones por no decisión:** en ciertos casos, al no tomar una decisión cualquiera, las circunstancias siguen su curso. Esto genera consecuencias con las que hay que lidiar y que quizás se podrían haber evitado si se hubiera tomado una decisión previamente.
- **Decisiones combinadas:** en la mayoría de los casos utilizamos una mezcla de las diferentes maneras, ya que revisamos la información con la que

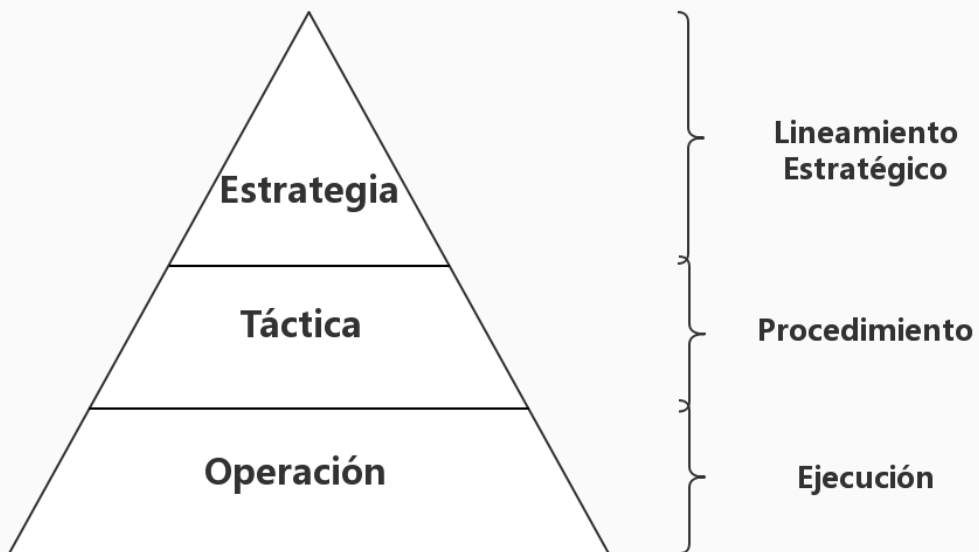
contamos y luego la pasamos por nuestro filtro de juicios, valores, emociones, instinto, etcétera.

### 1.2.1. El proceso de toma de decisiones

En nuestro sitio web, así como en otras áreas de trabajo, debemos comprender que existen diferentes niveles de toma de decisiones. Una vez que hemos comprendido esa situación, debemos también darnos cuenta de que nos enfrentamos a distintas instancias de medición. En cada caso, se medirán los diferentes KPI y serán distintos los tomadores de decisión dependiendo de cada uno de aquellos casos a los que nos enfrentemos.

Es por ello que nos resulta indispensable internalizar que existen tres diferentes niveles de toma de decisión: estrategia, táctica y operación.

**Figura 4: Niveles de toma de decisión en los planes**



Fuente: Paz y Barrionuevo, 2017, p. 26.

- **Estrategia:** cuando nos enfocamos en el nivel estratégico, la tarea más importante es clarificar nuestra visión, establecer el estado deseado y poder posicionarnos en el futuro para plantear el camino que debemos seguir. Es un posicionamiento en aspectos relacionados

con la visión de nuestro proyecto, con la propuesta de valor que tenemos para ofrecer y con el beneficio final que queremos que se genere a partir de la ejecución del plan que realizaremos. Es por ello que debemos pensar en términos de lineamientos estratégicos y de políticas de aplicación. En definitiva, implica mirar todo el contexto completo, conectar los puntos que lo unen y mirar el todo en su conjunto para extraer los ejes de trabajo que se deberán realizar para alcanzar los objetivos.

- **Táctica:** al trabajar a nivel táctico, nuestra tarea más importante es verificar la factibilidad de lo que queremos aplicar. Es fundamental desarrollar procedimientos para alcanzar nuestros planes y trabajar en mejorar los procesos que realizamos para alcanzar los objetivos. Implica mirar los diferentes elementos de nuestro plan y plantear metas y lineamientos de acción de cada uno de tales elementos para luego integrarlos a través de lo que se planteó desde la estrategia. Implica un foco en la factibilidad y la aplicación de los recursos que disponemos.
- **Operación:** a nivel operativo, nuestra tarea más importante es trabajar en el detalle. Se debe trabajar en la realización efectiva de cada una de las tareas y actividades planteadas y en la eficacia y eficiencia de procedimientos. El foco estará puesto en la calidad de la ejecución de las tareas. (Paz y Barrionuevo, 2017, p. 28).

Cada una de estas instancias requiere de diferentes perspectivas de acción. Principalmente, difieren en cuáles son los puntos de éxito de cada una de esas etapas. En el caso de la estrategia, debemos pensar en positivo y con optimismo para imaginar lo que podemos conseguir. En la táctica, debemos pensar en negativo y con pesimismo para imaginar cuáles son los obstáculos que podemos encontrar en el camino y cómo generar procesos para superarlos. Finalmente, en la operación, debemos pensar en la ejecución y el detalle para realizar actividades que contribuyan a la visión de la estrategia y que eviten los riesgos analizados en la táctica.

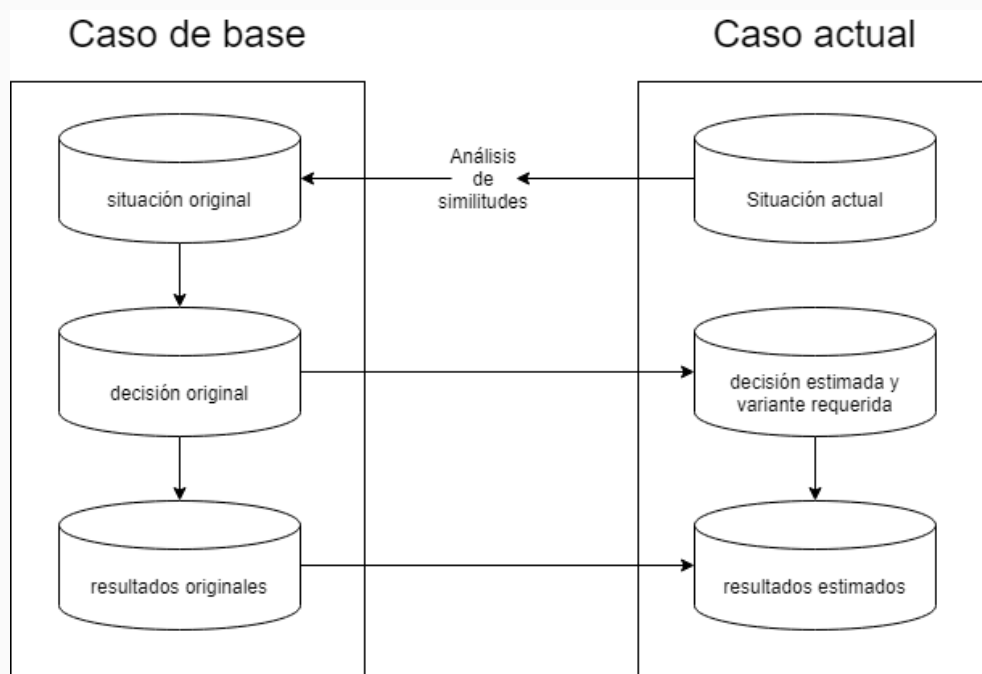
### 1.2.2. Decisiones basadas en casos

Cuando contamos con mayor cantidad de información de tipo cualitativo que cuantitativo (permitiéndonos comprender en mayor profundidad un fenómeno), podemos plantearnos la posibilidad de tomar decisiones basadas en casos. Es decir, de acuerdo a lo que sucede en una situación particular, buscaremos los elementos que nos permitan encontrarnos ante la misma situación en el futuro para poder reaccionar del mismo modo.

Normalmente, podemos plantearnos un conocimiento lo más acabado posible sobre un caso puntual. A partir de este punto, buscaremos en nuestros segmentos de público objetivo a aquellos miembros que se encuentran en la misma situación anterior.

En este sentido, nuestro trabajo de analítica estará centrado en comprender el grado de similitud que existe entre nuestro caso de base y el caso actual. Esto nos permitirá estimar qué resultado podemos esperar si tomamos la misma decisión que en el caso de base.

Figura 5: Decisiones basadas en casos



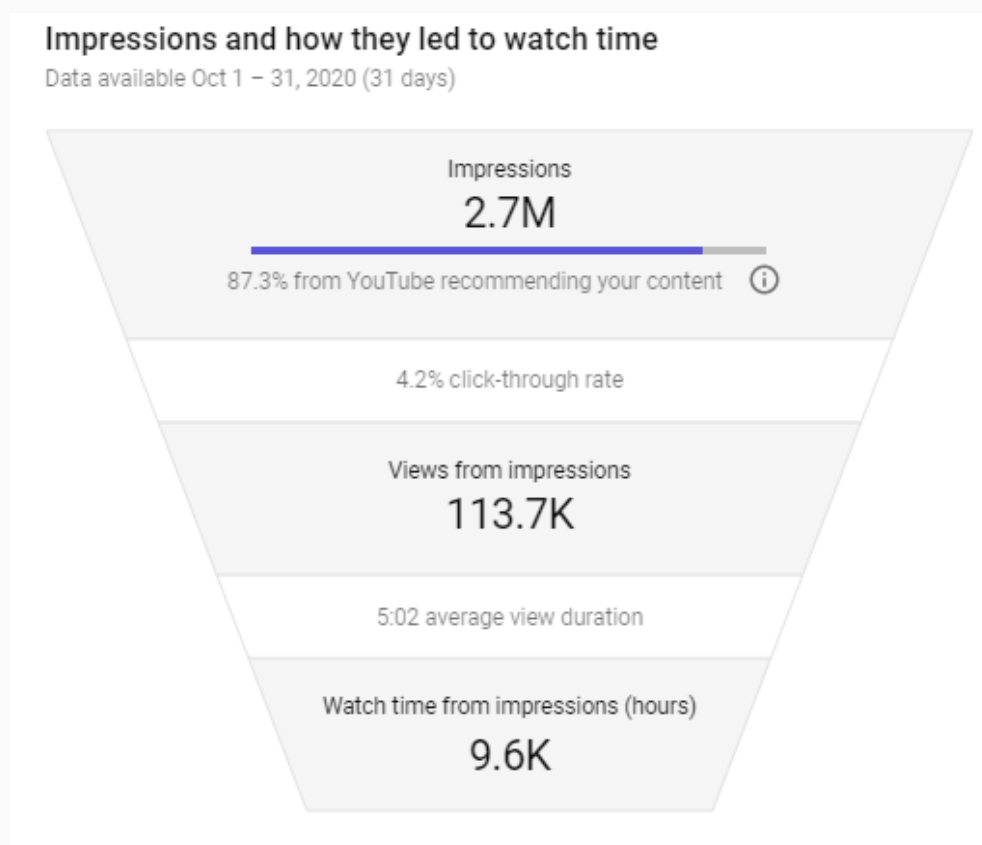
Fuente: elaboración propia.

### 1.2.3. Decisiones basadas en agregación de datos

Son las decisiones de base estadística. Nos permite tomar los casos en función de su agregación y esperar ciertos porcentajes de comportamiento en función de la estimación realizada.

En el caso del sitio web, es la manera en que podemos planificar nuestro embudo de conversión. Por ejemplo, conociendo nuestra tasa de conversión, podemos esperar un porcentaje mínimo de conversiones si logramos ingresar una cantidad específica en el embudo.

**Figura 6: Embudo de conversión basado en porcentajes**



Fuente: Captura de pantalla de video de Paz [Video de YouTube], 2020, <https://n9.cl/daso>

Por ejemplo, si tenemos un canal de video en YouTube y esperamos una *performance* regular de 4.2 % de CTR (*click through rate* o tasa de clics), requeriremos aproximadamente 2.34 millones de impresiones para alcanzar las 100 mil vistas de video. De ese modo, si sabemos que cada 5 mil vistas conseguimos un cliente, podemos plantearnos que necesitaremos 500 mil vistas para alcanzar los 100 clientes.

A partir de esa información, podemos planificar nuestros contenidos para obtener los 11 millones de impresiones que nos permitan conseguir las 500 mil vistas que nos llevarán a los 100 clientes.

Conociendo también estos valores, podemos modificar nuestro plan a partir de intentar optimizar las métricas.

De este mismo modo, si logramos una tasa de clics de mejor *performance*, necesitaremos menos impresiones para lograr la misma cantidad de clientes o para lograr más clientes con la misma cantidad de impresiones.

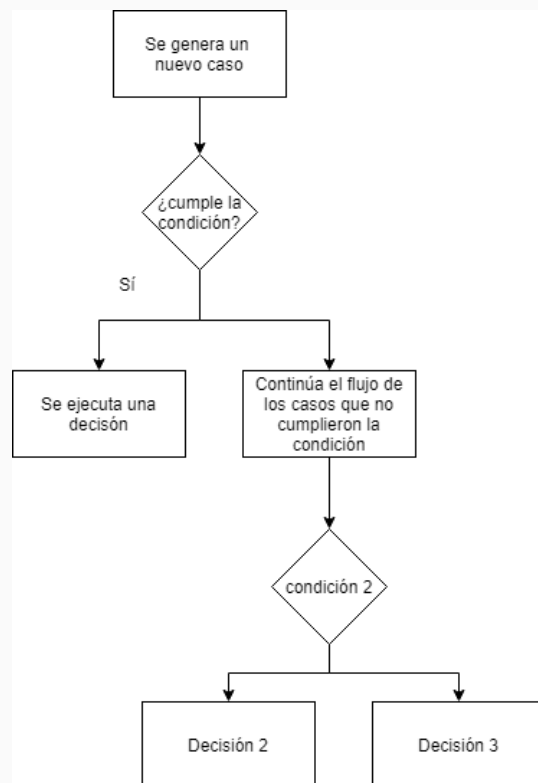
#### **1.2.4. Decisiones condicionales y triggers**

Las decisiones condicionales y basadas en *triggers* o disparadores nos permiten combinar las dos metodologías anteriores a través de decisiones que se realizan en función del cumplimiento de ciertas condiciones.

Decimos que esta metodología es una combinación por las siguientes razones:

- Contamos con información estadística y análisis integral sobre todos los casos que suceden en nuestro sitio web.
- Definimos cuáles son los casos de base que representarán los tipos de decisiones a tomar.
- Validamos el cumplimiento de las condiciones en cada uno de los casos actuales.
- Ejecutamos las condiciones que disparan las decisiones tomadas previamente.

**Figura 7: Decisiones condicionales**



Fuente: elaboración propia.

Como podemos ver en la figura anterior, estamos realizando un proceso simplificado que nos permite comprender la metodología de manera sencilla.

Para ejemplificar con un proceso de analítica web, podemos mencionar:

- 1) Un visitante llega a nuestro sitio web: caso.
- 2) El visitante supera los sesenta segundos de permanencia: condición.
- 3) Se le muestra un *pop-up* que lo invita a suscribirse a nuestro *newsletter*: decisión.
- 4) Se suscribe a nuestro *newsletter*: condición.
- 5) Enviamos un correo electrónico de bienvenida: decisión.
- 6) El usuario abre el correo electrónico: condición.

7) Agregamos al usuario a nuestra base de *e-mail marketing*: decisión.

Como vemos, hay que estratificar las condiciones y las decisiones. Del mismo modo, debemos tener casos de base sobre cómo nos comportaremos cuando las condiciones no se cumplan o cuando se cumplan condiciones alternativas entre diferentes opciones.

En definitiva, el proceso de la analítica es la base de nuestro trabajo en *web analytics* y es lo que nos llevará a la posibilidad de tomar decisiones en nuestros sitios web para que impacten de manera positiva en nuestros negocios.

## Referencias

**Paz, G.** [Canal de YouTube]. (2020). Recuperado de <https://n9.cl/daso>

**Paz, G.** (2016). *Analytics. Análisis y tratamiento de datos deportivos*. Córdoba: FC Barcelona Universitat.

**Paz, G. y Barrionuevo, D.** (2017). *Analítica y reporte en social media*. Córdoba: Social Media Trends.