

# Caso 1 - Hoteles

Introducción a Data Mining – Virtual 2022



## Integrantes:

- De Angeli, Luis
- Fattore, Waldo
- Fiori, Ana Carla
- Roitman, Karina

# Introduccion

El presente trabajo se orienta a la resolución de un caso en donde se tienen los objetivos de conocer cuáles son los datos necesarios para que un cliente / huésped de un hotel pueda tener una buena experiencia. A su vez, que una vez que haya dejado el hotel y en función de aquellos productos y servicios que haya consumido durante su/s estadía/s, dispongamos de una mayor cantidad de información aún que permita particularizar sus necesidades y deseos.

De esta manera, se contemplan también una serie de suposiciones tendientes a originar un sistema de minería de datos (KDD), cuyo objetivo es conocer al cliente en detalle, función de generar esa satisfacción del cliente buscada, colaborando con el éxito comercial del negocio.

Particularmente, el proceso va a enfocarse en una fase de recolección de datos, una categorización de los clientes con método de clustering cuando se desconoce completamente su información, y luego un marketing post-estadía con un modelo de clasificación (probablemente de árbol de decisión) para un análisis predictivo. En las siguientes secciones detallamos este proceso

## Suposiciones

En función de cumplir con los objetivos, se realizarán una serie de suposiciones, las cuales consideran los siguientes puntos:

- Consideramos que, más allá de la base de datos de clientes y otra información aledaña, no se cuenta con absolutamente ninguna información específica del caso a explorar. Es decir, no existen encuestas previas ni se comenzó a recabar respuestas, lo que implica que las preguntas pueden definirse arbitrariamente y que una fase de recolección de datos es necesaria.
- Previo al comienzo de la utilización del modelo, se asume que las encuestas y el uso de las credenciales se efectúa durante un período de tiempo de, por ejemplo, 3 meses, sin que se utilicen para la toma de ninguna decisión, sólo a los fines de construir un grupo de datos iniciales representativo.
- El cliente tiene capacidad y actitud para brindar la información en la gran mayoría de los casos. En los casos en que el cliente no tenga la actitud de completar la información (puede ocurrir), se tomará la información pertinente a su ficha de inscripción en el hotel.
- El hotel cuenta con personal que tiene capacidad para reconocer si un cliente no está dispuesto a brindar toda la información.
- El hotel tiene acceso a base de datos externas, como Nosis, Facebook, Instagram, información de la agencia de viajes en su portal.
- El hotel cuenta con una base de datos de toda la cadena, con acceso a datos de clientes preexistentes, sucursales, eventos y atracciones por ciudad y región.
- Se presume que el hotel cuenta con todos los servicios que se están ofreciendo o, si fuera el caso, los presenta a los clientes como propios, aún cuando estén tercerizados en la realidad. De esta manera, y a título de ejemplo, aun cuando el hotel no cuente con una cancha de golf, es un servicio que ofrece al cliente como propio, y fruto de convenios con empresas u organizaciones terceras.

- Se presume que la tarjeta que tiene disponible sólo cuenta con capacidad para hacer seguimiento de las actividades del cliente durante los momentos en que el cliente se encuentra dentro del complejo.

**Información que el hotel puede recabar internamente:**

El hotel cuenta con la capacidad de recabar la información siguiente durante cada paso de la estadía del cliente:

Etapa de la estadía	Origen de los datos a relevar	Datos a relevar
Inicio de la estadía	Datos demográficos	<p>Nombre y Apellido, Teléfono, Sexo, Edad, País de origen, Dirección, Correo Electrónico</p> <p>Motivo de la estadía (Trabajo, Vacaciones, FERIA, Convención)</p> <p>Contingente (individual, pareja, familia, grupo)</p> <p>Estos datos también pueden estar disponibles de la reserva realizada a través de medios propios de la empresa hotelera, o de una agencia de viajes.</p>
Inicio de la estadía	Datos comerciales	<p>Forma de pago, Tarjeta de crédito disponibles, Cantidad de tarjetas de crédito disponibles,</p> <p>Medio de transporte.</p> <p>Destino anterior. Próximo destino</p> <p>Cantidad de acompañantes menores.</p> <p>Cantidad de acompañantes mayores</p> <p>Duración de la estadía: (1 noche, hasta 3 noches, 1 semana, mas)</p> <p>Reserva: (walk-in, reserva por internet, reserva telefónica)</p> <p>Antelación: (días, semanas, meses)</p>
Inicio de la estadía	Ficha de intereses principales	<p>Puntuando del 1 al 5, el cliente debe definir cuáles son sus intereses principales, vinculando los servicios del hotel. Esta ficha también tiene como objetivo que el cliente conozca servicios que, tal vez, no haya tenido en consideración y que existen dentro de las instalaciones:</p> <p>Spa: se detallan baño sauna, baño finlandés, jacuzzi, hidromasaje, masajes, tratamientos de belleza, etc.</p>

		<p>Gimnasio: se detallan los tipos de actividades que se pueden llevar adelante: complemento de pesas, spinning, kickboxing, artes marciales, etc.</p> <p>Actividades al aire libre: se detallan todas las actividades, a saber: cabalgata, bicicleta, espeleología, escalamiento, playa, windsurf, kayakismo, natación, buceo profesional, buceo inicial, salidas en yate, visitas a otras islas, etc.</p> <p>Deportes: se detallan todas las actividades de índole deportiva que tiene el hotel disponibles: fútbol, tenis, paddle, natación, golf (incluye alquiler de equipos), etc.</p> <p>Gastronomía: se detallan los tipos de restaurantes y bares que el cliente puede potencialmente visitar. Se detallan también si el cliente tiene restricciones en el tipo de alimento que puede definir (sea por circunstancias</p> <p>Shows: se detallan los tipos de shows que el cliente puede visitar, como ser: baile de disfraces, salsa, rumba, etc.</p> <p>Excursiones: se detallan los tipos de salidas a lugares que estén alrededor del hotel y que tengan como máximo 12 hrs de duración. Aquí se detallan: salidas en chart terrestre, salidas en chart aéreo, etc.</p> <p>Miniclub: se detallan las actividades que los niños pueden realizar en la guardería.</p>
Inicio de la estadía	Restricciones	Ficha obligatoria a completar. Aquí se añaden aquella data que implique restricciones al cliente, que conlleva lo que no puede llevar adelante (por limitaciones físicas, alergias), o que no debe llevar adelante (motivos religiosos o elecciones personales: vegetariano, veganismo, no comer cerdo, etc.).
Durante la estadía	Tarjeta de cliente	El cliente cuenta con una tarjeta que es capaz de rastrearlo por el complejo conociendo qué servicios estuvo disfrutando, cuánto dinero estuvo abonando y qué tiempo estuvo utilizando el servicio.
Al final de la estadía	Encuesta de satisfacción	El cliente completa una encuesta que tiene como objetivo conocer su grado de satisfacción y disfrute de las actividades que estuvo utilizando y que puede relevarse fruto del paso anterior.
Posterior a la estadía	Página web / mailing	Contemplando dónde el cliente manifestó mayor interés, se alimenta la base de datos.

# Flujo de operación

Pasos que sigue un huésped en toda su estadía:

1. **Check-In:** El huésped llega al hotel, se registran datos demográficos y completa el formulario.
2. **Estadía:** El huésped cursa su estadía normalmente. Se le otorga una credencial para ingreso a lugares recreativos, piscinas, bares y amenities.
  - a. Nuevo huésped
  - b. Resto de la estadía
3. **Check-Out:** el huésped se retira del hotel y completa el formulario de salida

**Check-in:** Primero determinar si es un huésped que ya haya utilizado la cadena de hoteles. Si lo es y ya completó el formulario en el pasado, se le puede efectuar una versión resumida del formulario.

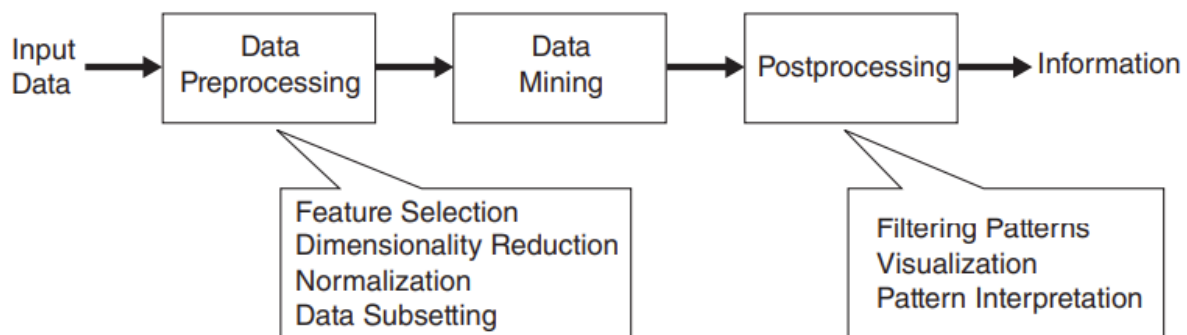
Una vez recabada la información de entrada se conoce una idea general de los intereses, es decir queda elaborado el perfil inicial

**Estadía:** a medida que transcurren las horas, una vez que el cliente hace su primera utilización de la credencial, se comienza a afinar el perfil para recomendar en base a los intereses que demuestra en la práctica por sobre los de la encuesta, y este es un proceso continuo.

**Check-Out:** luego de tomar la encuesta de salida, se elabora un perfil post-visita que utilice el feedback del cliente para acotar las recomendaciones

## KDD Knowledge Discovery in Databases (KDD)

En este apartado se llevará adelante el proceso de convertir data cruda en información útil, tal cual se puede ver en la imagen a continuación:



## Data Preprocessing

### **1. Data cleaning**

Dadas las dos encuestas, analizamos qué información es inconsistente. Se remueven datos duplicados, incompletos, formateados incorrectamente o información irrelevante, sin sacrificar la integridad de la información original.

### **2. Data integration**

Consiste en incorporar resultados de las encuestas al resto del ecosistema de la base de datos, es decir, información sobre las sucursales, la época del año, los eventos y la geografía.

Por ejemplo, si el motivo de la visita es “feria” al ser coordinado con información de eventos en la zona del hotel permite deducir los intereses del huésped con un cierto grado de confianza, según la temática de los eventos.

### **3. Data selection**

Seleccionar la información verdaderamente relevante para poder obtener conclusiones fidedignas.

### **4. Data transformation**

Transformar y consolidar los datos en formas adecuadas para la explotación minera mediante la realización de operaciones de resumen o agregación

Por ejemplo, con la información obtenida de las compras hechas con la credencial, podemos agregar una variable que contenga la sumatoria de todos los productos comprados por un cliente durante la estadía. De esta manera, luego se podría analizar alguna campaña de marketing respecto a los productos ofrecidos en el hotel.

## Data Mining

## Data Post Processing

### **5. Pattern evaluation**

Evaluaremos qué patrones se cumplen, en base al proceso de data mining realizado. El proceso de Data Mining será desarrollado en los puntos 1 y 2.

### **6. Knowledge presentation**

Se presentarán la información trabajada de manera en que podamos no sólo contar con la categorización de los clientes, sino ver el efecto que las medidas tomadas.

# Data Mining 1 - Decisión sobre nuevo Huésped

**1- Decidir cuáles anuncios de TV van a ser enviados a los huéspedes durante su estadía; un subproblema de este primer problema es decidir cuáles anuncios van a ser enviados durante los primeros días de la estadía de un huésped**

Cómo hemos mencionado en el punto anterior el esquema general del KDD, en este caso analizaremos el punto 1, mencionando el input inicial con el que contamos y la decisión respecto al paso de data mining.

## Determinación del PERFIL INICIAL

Determinamos dos escenarios posibles

- Cliente preexistente, ya se conoce a qué cluster pertenece.
- En el caso en que aparezca un nuevo cliente, entonces se utilizará un modelo predictivo, para poder estimar a qué cluster pertenecería.

### Input Data

En una primera instancia, se cuenta con información obtenida de los datos demográficos, encuesta sobre intereses e información relacionada a todos los hoteles pertenecientes a la cadena. En cuanto a este último punto, nos referimos a ubicación de los mismos, actividades que se llevarán a cabo en dicha localidad (como congresos, conciertos, ferias, etc), calificación del hotel (puntuación/estrellas). Esta información se encontrará en una tabla grande relacionada de todas las sucursales de los hoteles.

### Paso de Data Mining:

En una primera instancia utilizaremos clustering, para generar etiquetas en grupos de datos. Los objetos se agrupan según el principio de maximizar la similitud intraclase y minimizar la similitud interclase. Es decir, los clusters de objetos se forman de manera que los objetos de un cluster tengan una alta similitud en comparación entre sí, pero sean bastante diferentes a los objetos de otros clusters.

En base a esto, se generarán subpoblaciones homogéneas de los clientes según la información obtenida para enviar anuncios, lo más personalizados posible, teniendo en cuenta el consumo de TV y que las promociones deben estar sumamente dirigidas.

El Clustering es una operación no supervisada, ya que no contamos con una etiqueta respecto a los clientes. De esta manera usamos clustering a modo exploratorio con esta información.

Este proceso es dinámico, a medida que se va incorporando información al modelo (datos obtenidos por credenciales), el mismo se ajusta de manera continua. Suponemos, según la factibilidad, que utilizaremos el modelo para volver a categorizar desde una o dos veces por día, hasta inmediatamente tras cada uso de la credencial.

Esta revisión del clusters es valiosa puesto que ante la ausencia generalizada de información específica sobre el huésped, cada utilización de la credencial determina un hecho real que puede cambiar drásticamente el grupo al que pertenece, con mayor validez que el valor estimativo o meramente ideal de la encuesta inicial

## Data Mining 2 - Marketing Post Estadía

**2. Decidir sobre la selección de hoteles para el mailing privado durante el verano.**

### Determinación de un Perfil final

**Input data:** Al finalizar la estadía, contaremos con los datos de la encuesta, los obtenidos con la credencial, y los del cuestionario final.

#### Paso de Data Mining:

Dada la base de datos de clientes y el clustering que se le asignó al cliente (realizado en el punto 1) se entrenaría un modelo de clasificación (por ejemplo, un árbol de decisión) para un modelo de análisis predictivo sobre la selección de hoteles para el mailing privado.

Los modelos predictivos son supervisados, el término “supervisado” proviene de clasificar los ejemplos etiquetados en el conjunto de datos de entrenamiento.

La utilización de árboles de decisión es particularmente apropiada en este escenario porque:

- Son muy rápidos de ejecutar una vez modelados.
- Tienen una gran tolerancia a los datos faltantes.
- Tienen gran tolerancia a la multiplicidad de variables no significativas.
- Pueden tolerar la coexistencia de variables cualitativas y cuantitativas.
- Son flexibles para la evolución del modelo, es factible agregar ramas o cambiar la lógica a la luz de nuevos datos ingresando al modelo



# Conclusión

En este caso se realizó un estudio en detalle de los pasos necesarios para poner en práctica un sistema de minería de datos. En este sentido:

- Se realizaron una serie de suposiciones tendientes a establecer condiciones de borde favorables para el sistema.
- Se definieron cuáles eran los datos necesarios para que el sistema funcione, desde el ingreso del cliente al hotel, hasta luego de haberse retirado.
- Se generó un perfil inicial del cliente utilizando herramientas de KDD, desde el momento de la introducción hasta el marketing post-estadía.
- Se generaron las condiciones iniciales para realizar un modelo de predicción y clasificación del cliente.

Con esta información se dispone de una clusterización del cliente en detalle y muy particularizada, lo cual permitirá colaborar con los objetivos comerciales del negocio, brindando al cliente exactamente lo que está buscando, basado en sus gustos, preferencias, necesidades y posibilidades.