Maestría en Explotación de Datos y Gestión del Conocimiento

IDM – Caso Nro. 1

Autores:

Agustini, Daiana
dagustini@mail.austral.edu.ar
Beloqui, Gonzalo
gebeloqui@mail.austral.edu.ar
Grondona, Mariano
mgrondona@mail.austral.edu.ar
Vargas, María Ines
mivargas@mail.austral.edu.ar

Introducción:

Cambios en situación actual:

a. Anticipar formulario de preferencias en web check in anticipado, en vez de en momento de arribo, para poder contar con información previa y entrenar datos de sugerencias desde el primer día o inclusive antes del check in.
b. Identificar aquellos huéspedes a quienes se le enviaran los videos de sugerencias en sistema de TV privado y analizar a quienes será mas conveniente hacerlo a través de un canal de red social, o una app personalizada que sea necesaria para realizar el check in previo y permita identificar el huésped por email con el cual realizó la reserva.
c. Centralizar las reservas en un único site para todos los hoteles, lo cual permite registrar los distintos pasos y movimientos en el site, para poder medir cuantas búsquedas o consultas de disponibilidad finalizan en reservas concretadas y cuales y en qué instancia se desisten en instancias previas.

Ideas a Desarrollar:

- Asociación y clustering: Analizando multidimensionalmente tipo de habitación y hotel elegido, actividades y gastos con la tarjeta del hotel, más datos personales como edad, o nivel socioeconómico en base a límite de la tarjeta de crédito, automóvil, etc. Es posible generar clusters de huéspedes y una caracterización de clientes standard para luego dirigir mejor las diferentes ofertas de servicios en base a las preferencias de cada clúster.

- Regresión Múltiple: En base a los distintos segmentos generados por la asociación y clústering, predecir el monto dinerario a gastar por cada huésped, en base a ingreso inferido por otras variables, como tipo de tarjeta de crédito, tipo de automóvil (si registra en la cochera del hotel), ocupación, género, edad.

- Saber fecha de nacimiento, preferencias de comidas, de bebidas, permite generar un programa de fidelización con gran valor percibido para los huéspedes a bajo costo.

- Clasificación: identificar huéspedes con categorías predefinidas, para predecir que segmento es mas sensible a promociones y deja de adquirir los servicios de los hoteles en primer lugar una vez terminadas ofertas.

➢ Pasos del Sistema de Data Mining aplicados a la solución del caso planteado:

1. **Limpieza de Datos:**

Considerando que una gran cantidad de los datos necesarios son provistos por los huéspedes:

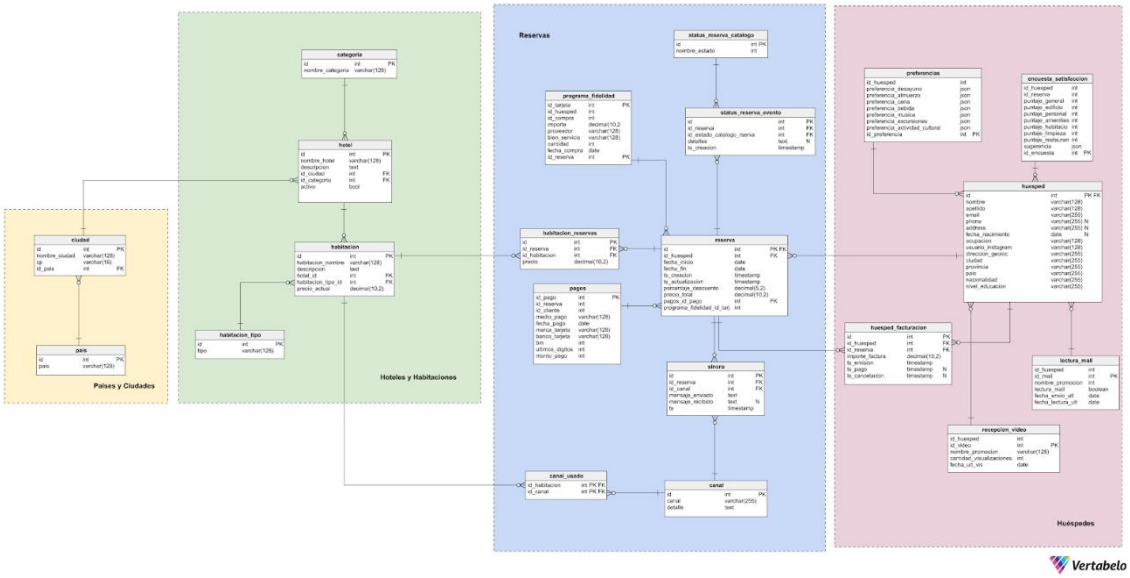
- Datos personales: fecha de nacimiento, email, teléfono, ocupación, etc.
- Datos geográficos
- Calificaciones y respuestas de texto abierto como sugerencias.

Ante esta situación es clave la instancia de limpieza de los datos. Para ello, una buena práctica es limitar en el formulario de captación de datos mediante restricciones de formatos que no permitan avanzar si no se respeta la regla preestablecida, a los efectos de garantizar desde la fuente la homogeneidad de los formatos de los datos, por ejemplo:

- Formato único para fechas, teléfonos y prefijos, códigos postales, e identificadores civiles y fiscales
- Validación de email con doble ingreso o con mail de verificación
- Datos geográficos preseteados con listas que solo permitan la selección por parte del huésped.
- En cuanto al seteo de preferencias para los servicios conexos (excursiones, bares, restaurantes, etc) dar categorías predefinidas que permitan la selección, en vez de dar libertad al huésped para que escriba la respuesta en formato abierto.
- En campos de texto para respuestas abiertas, como sugerencias, es mucho más complejo la estandarización de los datos en la fuente, sino que tienen que establecerse robustos modelos de NLP que permita obtener información concreta sobre las respuestas, considerando que en el rubro hotelero puede ser común tener clientes no solo de distintas partes del país, con otros códigos de lenguaje, sino también de otras nacionalidades.

Por extensión de las respuestas, si bien pueden limitarse los caracteres, este tipo de atributos suele guardarse en bases de datos no relacionales, como archivos JSON.

El diagrama entidad relación de la base de datos de cada hotel pudiera quedar representada de la siguiente forma: https://drive.google.com/file/d/1RN3WnJeuco1j4g0Yapj7-R0MFvFn1IEc/view?usp=sharing



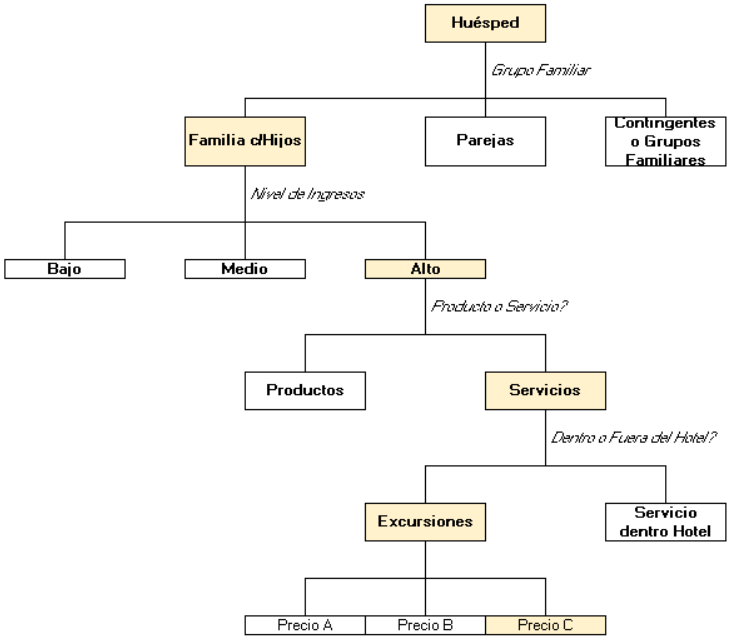
2. **Integración de Datos**:

3. Selección de Datos:

Se define como una instancia del proceso en donde se decide cuales son los datos relevantes para el análisis y luego son extraídos o recuperados de las bases con las cuales trabajamos.
La selección de los datos puede realizarse utilizando distintas técnicas:

- Redes Neuronales: se utilizan para clasificación y reconocimiento de patrones. Varios estudios indican que en algunos dominios las Redes Neuronales proveen mayor precisión predictiva que los algoritmos de aprendizaje simbólicos comúnmente utilizados. Aún así, no consideramos adecuado utilizar esta técnica para el caso de análisis por dos razones principales; los métodos de clasificación de las Redes Neuronales no son explicables, y los tiempos de aprendizaje son lentos, imprácticos para grandes volúmenes de datos.

- Arboles de Decisión: servirán para abordar problemas tales como la clasificación, la predicción y segmentación de datos con la finalidad de obtener información que pueda ser analizada para la toma de decisiones futuras. Al preguntarnos qué tipo de servicio o producto ofrecerle a un huésped pueden darse las siguientes situaciones (Ejemplo ilustrativo básico)



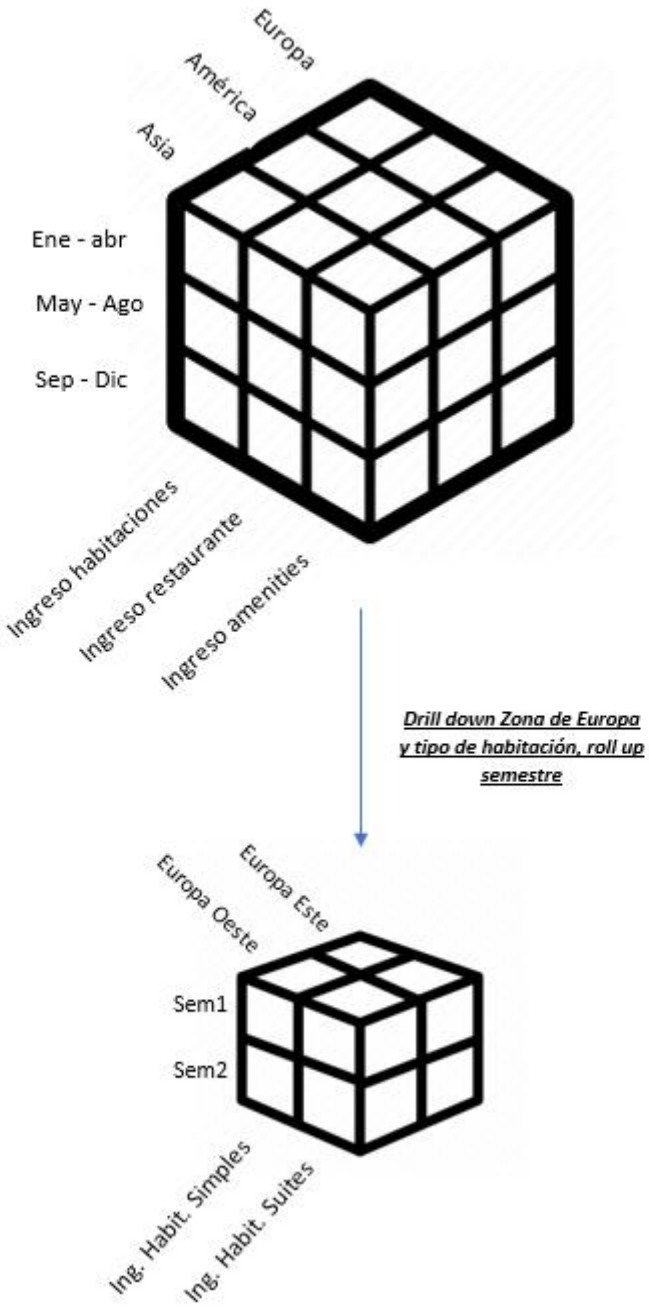
- Clasificaciones Bayesianas: se utiliza el "Naive Bayesian Classiffier" (NBC por su siglas en inglés). Es un algoritmo de aprendizaje automático simple pero potente. Se basa en la probabilidad condicional y el Teorema de Bayes. Podremos realizarnos preguntas del tipo:

    - Cuál es la probabilidad de que un huésped contrate una excursión dado que es soltero?
    - Cuál es la probabilidad de que un huésped use el Spa dado que tiene familia con dos hijos?

    Nos servirá para llevar adelante predicciones en cuanto a los niveles de consumo/gasto de cada uno de nuestros huéspedes e incentivar los mismos en función de las mayores o menores probabilidades de ocurrencia de cada uno de ellos.

4. **Transformación de Datos:**

Ejemplo Cubo multidimensional para análisis OLAP:



5. Data Mining: Funcionalidades

a) Caracterización y discriminación de Datos:

En esta etapa el foco está en obtener un resumen de las características generales de los huéspedes. A través de análisis descriptivos, como distribuciones de frecuencias de acuerdo a características de interés principales como edad, cantidad de personas por reserva, motivo del viaje, hotel seleccionado, entre otros. De esta manera se definen diferentes targets de huéspedes, a modo de ejemplo:

- Familias (Grupo familiar promedio compuesto por dos adultos y dos niños)
- Parejas Jóvenes (Hasta 35 años de edad)
- Parejas Adultas (Entre 36 y 50 años de edad)
- Parejas de Adultos Mayores (Mayores de 50 años)
- Contingentes o Grupos Familiares Grandes (Por ejemplo: aquellos que vienen a festejar un casamiento o una celebración)

En este punto es de interés definir los huésped tipo o más frecuentes de cada hotel y los menos frecuentes. Este análisis se aplica a cada hotel, con base en los datos históricos, agrupando por las características más relevantes: ¿Cuál fue el target más frecuente del último año? ¿Cuál fue el menos frecuente?. Con esta información se podría pensar en un programa de fidelidad para aquellos huéspedes frecuentes y que alienten a volver a las distintas cadenas del hotel a través de promociones especiales como por ej. cumpleaños/ aniversario. Una vez definidas las clases se continúa segmentando en otras categorías:

- Procedencia
- Demografía (edad, sexo, ingresos, estado civil, etc.)
- Ocupación
- Actividades realizadas en la estadía
- Tipo de habitación seleccionada
- Consumos realizados
- Motivo de viaje: ocio, negocios, luna de miel, viaje familiar o en pareja
- Movilidad
- Clientes sociales: casamientos / aniversario /luna de miel
- Status de fidelidad (Cliente nuevo/Cliente recurrente)
- Valoración final (Cuestionario de satisfacción)

Es importante también realizar encuestas de satisfacción y/o permitir reseñas o comentarios en el perfil de las redes sociales, de esta forma nos permitirá entender mucho mejor cuales son las preferencias/ necesidades de nuestros clientes y de esta forma poder identificar cual es nuestro grupo demográfico ideal.

b) <u>Asociaciones y correlaciones:</u>

En base a los datos transaccionales históricos de los huéspedes y para cada target, se identifican los diferentes patrones y su frecuencia respecto a los consumos y actividades. El objetivo es poder descubrir asociaciones de interés y correlaciones que permitan evaluar hechos que ocurren de manera conjunta o secuencialmente. Frequent itemset mining es un método aplicable en esta instancia, pudiendo determinar las probabilidades de ocurrencia de evento condicionado por otro, este análisis puede orientar las diferentes ofertas a realizarle a los huéspedes con base en su target o actividades que tenga programadas y que tengan mayor probabilidad de interesarle. En este sentido, un análisis de asociación podría indicar:

Los huéspedes más jóvenes >>> Cenan fuera del hotel

> *[Support: 2% confidence: 70%]*: 2% de todas las transacciones indican que los dos eventos ocurren de manera conjunta, y que hay un 70% de probabilidad de que si el huésped es joven entonces cene fuera del hotel.

Viaje en pareja >>> Usan spa , cenan en el hotel, usan el bar.
[Support: 10% confidence 90%]

De igual manera se pueden aplicar análisis multidimensionales:

Viaje en familia ^ movilidad propia >>> Visita a sitio turístico determinado
[Support: 20% confidence: 50%]

En los atributos que representen variables continuas es posible también aplicar análisis de correlación estadística. A modo de ejemplo podría aplicarse a los consumos de la tarjeta.
Tener información sobre estos detalles ayuda al hotel a brindar un servicio personalizado a medida del cliente. Dado que buena parte de esta información se puede obtener mediante el formulario de check in, el sistema podría hacer ofertas de valor para el cliente con anticipación. También el uso de la credencial nos dice el uso de los servicios lo cual nos podría dar una pauta de aquellos productos que se podrían vender bien juntos, por ejemplo, promociones o descuentos en el bar de la pileta en determinados horarios.

c) Clasificación y Regresión para análisis predictivo:

Proceso de encontrar un modelo que describa y discrimine las distintas clases de data o conceptos. Este modelo deriva del training data y es usado para predecir las distintas clases de objetos como función de los valores de los otros atributos. En base a esta información el hotel puede tomar decisiones más precisas para llevar a cabo sus acciones de marketing e inclusive saber con certeza aquellos huéspedes a quienes contactar para ofrecerles distintos incentivos o no, y que tipo de relación establecer con ellos.
Se podrían llevar a cabo distintas reglas de clasificación que tengan en cuenta lo siguiente:

- Duración de la estadía.
- Estadías por días de semanas.
- Ingresos totales por habitación.
- Ingresos totales por cliente.
- Plazo de ejecución de la reserva.
- Porcentaje de cancelación.
- Relación de no presentación.

Establecer un modelo para identificar el tipo de anuncio que va a ser propuesto para cada tipo de huésped a modo de ejemplo:

cliente id	Tipo de viaje	Demografico	Recurrente	Duracion de Estadia	Edad	Nivel de Ingresos	Clase : Tipo de Anuncio
id##	Vacaciones	Viaje Familiar	si	Dias de semana / quincena	N/a	Medio / Alto	Anuncio x tv
id##	Vacaciones	Viaje en Pareja	no	Fin de semana	Hasta 35 años	Medio / Alto	Redes sociales
id##	Negocios	N/A	no	Dias de semana	Mayor 35 años	Medio / Alto	Mail

d) Clustering:

Identificar subgrupos en base a sus similitudes y separarlos de otros subgrupos diferentes o no relacionados. en base a los distintos segmentos generados por la asociación y clústering, predecir el monto dinerario a gastar por cada huésped, en base a ingreso inferido por otras variables, como tipo de tarjeta de crédito, tipo de automóvil (si registra en la cochera del hotel), ocupación, género, edad. Esto es de vital importancia para conocer de una manera detallada quienes son nuestros clientes. Por ejemplo, se podría clusterizar en función a las siguientes variables:

Por edades:

Parejas Jóvenes (Hasta 35 años de edad)

Parejas Adultas (Entre 36 y 50 años de edad)

Por grupo familiar: como por ejemplo una familia tipo de 4 personas que se quedan mas en el hotel y usan mas las instalaciones como pileta, restaurant, playroom etc.

Por duración de estadía:

Fines de semana

Dias de Semana / Quincenas

Nivel Socioeconomico:

Por ej. Aquellos que reservan habitación en suite regularmente serian considerados huéspedes Premium.

El análisis de clusters nos permite descubrir asociaciones y estructuras en los datos que no son evidentes pero que pueden ser útiles una vez que se han detectado. Para lograr esta agrupación de variables será necesario seguir cierto algoritmo, que en este caso consideramos el K- Means como el mas apropiado dado a que una de sus ventajas es que trabaja bien con datos faltantes, como también computacionalmente rápido. Es importante destacar que el dato de entrada es el numero K de conglomerados deseados y luego implementar los siguientes 4 pasos:

i. Dividir a los n objetos en k subconjuntos no vacíos. (Los cuales se podrian elegir aleatoriamente o también utilizar información previa disponible).
ii. Se asigna cada elemento a uno de los k clusters, cuando la distancia al centroide de ese cluster es minima. (asignación secuencial)
iii. Se reasigna cada objeto al centroide de cluster mas cercano
iv. Se vuelve al paso 2 y repetir hasta que no es necesario realizar nuevas asignaciones.

e) Análisis de Outliers:

Consistiría en detectar aquellas anomalías que podrian indicarnos un consumo excesivo o fuera de lo usual en un determinado tipo de cliente a partir del cual el hotel quiera agradecer o recompensar de alguna forma por esa reciente compra o gasto. En función de esto se lo podria ofrecer una determinada promoción, upgrades en futuras estadías o un regalo de cortesia.

6. Evaluación de Patrones:

Un patrón resultará relevante si posee las siguientes características:

1. Fácilmente entendido por las personas
2. Permite validar o testear nueva información con cierto nivel de certeza
3. Es potencialmente útil
4. Resulta novedoso

Habitualmente buscamos confirmar o refutar hipótesis que tenemos de antemano respecto a los targets de los clientes. En algunos casos con sesgos subjetivos por parte de los usuarios. El descubrimiento de estos patrones nos permitirá llevar adelante distintas

acciones de Marketing a los efectos de incentivar compras o contrataciones de servicios determinados.
Esta etapa guarda estrecha relación con la de "Asociaciones y Correlaciones" descubiertas en el paso previo.

7. Presentación y visualización de la Información:
   Esta última última etapa consiste simplemente en aplicar y evaluar el conocimiento encontrado en los pasos anteriores al contexto y comenzar a resolver sus problemáticas. Si de lo contrario, los resultados no son satisfactorios entonces es necesario regresar a las anteriores etapas y revisar si es necesario realizar algún ajuste, analizando desde la selección de los datos hasta en la etapa de evaluación.
   Consiste en la visualización y la aplicación de distintas técnicas para presentar la información que encontramos en el proceso, como por ejemplo a través de un gráfico de Histograma en el cual se representa las ganancias en los diferentes meses de temporada o fuera de temporada por cada hotel.
   También se podria usar un gráfico de mosaicos para representar los datos de la encuesta de satisfacción y asi poder evaluar cuales son los puntos mas débiles en los cuales nos tenemos que focalizar para realizar las mejoras.
   El Grafico de Chernof para comparar entre las distintas características de los hoteles. Los datos que representan ojos, narices, orejas y otras formas de la cara, podrian representar distintos atributos como por ej: atención al registrarse en el hotel (check in), limpieza y condiciones de la habitación, comodidad, servicio al cuarto, desayuno, menu del Restaurant, Seguridad etc. Estas asociaciónes nos permiten rápidamente hacer asociaciones y detectar diferencias a través de las distintas filiales. Estos gráficos sumados a alguna técnica de Storytelling podria ayudar a dar un contexto de la información.

   <u>Conclusión</u>:

1. En base al análisis de los a) Registros históricos que se poseen de los huéspedes del hotel, b) El formulario completado y envíado previamente por cada uno de ellos con los principales datos demográficos y sus preferencias y c) El estudio del comportamiento en las principales redes sociales de los futuros huéspedes, se configurarán una serie de productos y servicios a ser promocionados durante los primeros dos días de estadía en el hotel. Estos consistirán en anuncios televisivos en las habitaciones de cada uno de ellos así como también en sus redes sociales (mayor hincapié en esta última alternativa ya que sabemos de antemano que el tiempo de encendido de la televisión en la habitación es mínimo). Esta oferta busca un impacto inmediato en la mente del futuro consumidor ya que la estadía promedio de los huéspedes, independientemente del target al cual pertenecen, no supera los 7 días. En tal sentido, se busca un efecto rápido y que no se diluya en el tiempo ya que, con el transcurso de los días, los huéspedes se acercan a la fecha de check out y disminuyen las probabilidades de venta.
   A modo de ejemplo, estás son algunas de las promociones que se ofrecerán en función del target al cual pertenecen:

   - Parejas Jóvenes (Nivel de Ingresos Alto): Noche de Cena. Menú por pasos en Restaurant de Comida Francesa del Hotel. Maridado con vinos de distintas regiones del mundo y con posibilidades de upgrades. Previamente hemos identificado patrones de consumo para este target que coinciden con esta oferta, así como también las preferencias indicadas por ellos a la hora de completar los formularios envíados

- Familia con Hijos (Niveles de Ingresos Medio y Alto): Alternativa de excursiones con actividades familiares al aire libre. Alta correlación observada dentro de este target de huéspedes en lo que a la contratación de este tipo de servicio se refiere. Los padres buscan alternativas educativas y recreativas para sus hijos fuera de las inmediaciones del hotel. Con servicios de traslado y almuerzo para grupo familiar.

Los anteriores son ejemplos simples y a modo ilustrativo del tipo de ofertas o promociones que se pueden llevar adelante y no pretenden abarcar a la totalidad de los targets de huéspedes que se poseen.

2. Finalizada la estadía y con la recolección de los datos de utilización y consumos provistos por la credencial entregada oportunamente a los huéspedes, se configuran las ofertas de mailing a ser remitidas para la temporada de Verano. Se busca la fidelización de los huéspedes con membresías y programas de beneficios del tipo "suma de millas". Nuevamente y a modo de ejemplo ilustrativo, se remiten promociones de estadía en los hoteles de playa que posee la cadena para cada uno de los targets.
En el caso de las parejas jóvenes, a los hoteles especialmente preparados para ellos en donde por ejemplo, se ofrecen servicios de barra libre las 24 horas, fiestas en el hotel, servicio de spa y traslados a bares y discotecas de la zona de influencia.
En el caso de las familias con hijos, promociones a los hoteles de la cadena que poseen parques acuáticos, servicios de guardería para los más chicos y actividades acuáticas con profesores y personal especialmente contratado para estos fines. De esta manera, se busca que el matrimonio pueda tener una alternativa de cuidado de sus hijos para poder descansar o elegir actividades para ellos solos.