

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



**SISTEMA DE RECOMENDACIÓN DE HOTELES  
SIMILARES**

TESIS

QUE PARA OBTENER EL TÍTULO DE  
MAESTRO EN CIENCIA DE DATOS

PRESENTA

FELIPE GERARD VALDÉS

ASESOR:

MÉXICO, D.F.

2015

“Con fundamento en el artículo 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada **“MÉTODOS NUMÉRICOS PARA AJUSTAR MODELOS DE REGRESION LOGÍSTICA REGULARIZADA”**, otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Bailleres Jr., autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación”

**Felipe Gerard Valdés**

---

FECHA

---

FIRMA

# Índice general

<b>Prólogo</b>	<b>IV</b>
<b>1. Análisis previo</b>	<b>2</b>
1.1. Sistemas actuales . . . . .	2
1.2. Diagnóstico . . . . .	8
1.3. Lista de deseos . . . . .	9
<b>2. Modelo propuesto</b>	<b>10</b>
2.1. Puntos a resolver . . . . .	10
2.2. Distancia . . . . .	10
2.3. Precio . . . . .	10
2.4. Similitud de hoteles . . . . .	10
2.5. Distancia revisitada . . . . .	10
2.6. Integrando todo . . . . .	10
2.7. Comentarios . . . . .	10
<b>3. Obtención de la información</b>	<b>11</b>
3.1. Formato necesario para el modelo . . . . .	11
3.2. Calidad de la información . . . . .	11
3.3. Sistema de respaldo . . . . .	11
<b>4. Desempeño</b>	<b>12</b>
<b>5. Trabajo futuro</b>	<b>13</b>

# Índice de figuras

# Índice de tablas

# Prólogo

# Capítulo 1

## Análisis previo

Antes de desarrollar el modelo de recomendación de hoteles similares, es conveniente analizar el *status quo*. Comenzaremos con un análisis de la competencia del que trataremos de sacar información valiosa de qué hay o no hay en el mercado y tomar ideas, posteriormente haremos un análisis del sistema de Best Day que pretendemos mejorar con el fin de entender en dónde hay áreas de oportunidad y finalmente sacaremos conclusiones de lo aprendido.

### 1.1. Sistemas actuales

En esta sección analizaremos la manera en que la competencia de Best Day recomienda hoteles y luego veremos qué se hace en Best Day. Los ámbitos que nos interesan son

1. La posición dentro de la página de las recomendaciones.
2. La forma en la que se eligen las recomendaciones.

Enfatizaremos en los principales competidores, de Best Day, es decir: Booking, Price Travel, Expedia y Despegar. También veremos si podemos aprender algo de TripAdvisor, que aunque no es un competidor directo, vive de recomendar hoteles y de su comunidad de comentarios. Las empresas competidoras están ordenadas de menor a mayor relevancia del sistema de recomendación que tienen.

## Price Travel (pricetravel.com)

Pusimos a Price Travel en primer lugar en la lista porque no tienen un sistema de recomendación de hoteles similares. Cuando se busca un destino, se presentan los hoteles en el Sort Order determinado por su sistema, pero ya que se entra a la página de un hotel en particular (independientemente de cómo se llegó a esa página), no se recomienda hoteles similares ni cercanos. De hecho, la única indicación de otros hoteles es en la pestaña de mapa, en la que se pueden ver todos los cercanos en un mapa interactivo. Dado que no aprendimos nada de este competidor, procederemos al siguiente.

## Booking (booking.com)

Booking es un competidor muy importante de Best Day, aunque también tiene un mercado más grande y mayor cantidad de opciones en destinos fuera de Latinoamérica. Antes de entrar en los puntos de interés general, cabe mencionar que algo que hace booking que no hacen los demás es que cuando se busca un hotel en la página con la barra inteligente, en lugar de ir directamente a su página, se muestra en el primer lugar de una lista, dando la posibilidad de elegir algún otro hotel. Ahora sí pasemos al tema en cuestión:

### ■ Hoteles populares por destino

1. Se encuentran del lado izquierdo, abajo de la barra de búsqueda y fechas. Es una posición de prominencia media-baja, puesto que no es visible inmediatamente sin bajar en la página.
2. Cuando se entra a la página (no al hotel en particular) por primera vez y se selecciona un hotel, aparece una sección de hoteles populares en el mismo destino del hotel buscado. El tema es que siempre aparece la misma selección para cada destino, independientemente del hotel que se haya elegido, pues la popularidad de los otros hoteles no tiene relación con el hotel elegido.

### ■ Búsquedas recientes

1. Se encuentran en el mismo lugar que los hoteles populares por destino, pero los sustituyen una vez que se ha visitado al menos un hotel anteriormente.



2. La información de los hoteles visitados se va almacenando en una *cookie*, de modo que para los hoteles subsecuentes, en lugar de los hoteles populares por destino, aparece un listado de unos pocos hoteles recientemente buscados por el usuario.

Al parecer en Booking sólo están interesados en lo que el cliente ha buscado y utilizan los hoteles populares por destino para no dejar vacía la sección para los clientes nuevos. El problema con los hoteles populares por destino es que no varían según el hotel buscado (siempre y cuando estén en el mismo destino), mientras que el de las búsquedas recientes es que no promueven que el cliente explore la página y encuentre la mejor opción para sus necesidades. Entraremos en detalles más adelante.

## Despegar (despegar.com)

Despegar es un competidor más directo y con un mercado más parecido al de Best Day que Booking, aunque está más enfocado en vender boletos de avión que cuartos de hotel. En este caso hay más de una estrategia y en más de un lugar de la página de cada hotel, por lo que las describiremos por separado. Cabe mencionar que no usan todas las estrategias para todos los hoteles, por lo que sospechamos que tienen información incompleta y sólo muestran lo que tienen disponible.

### ■ Búsquedas recientes

1. Se encuentra en la izquierda abajo del mapa. Igual que en booking.com, se tiene que bajar en la página para verlo, por lo que su relevancia es media-baja.
2. Son los hoteles buscados recientemente por el usuario. Curiosamente, no siempre nos aparecieron al hacer búsquedas, incluso después de haber visto varios hoteles.

### ■ Hoteles cercanos

1. Se encuentra directamente debajo de las búsquedas recientes.
2. Se listan los hoteles más cercanos al hotel en buscado, sin ningún criterio de precio, categoría o estilo del hotel.

### ■ Hoteles populares por destino

1. Se encuentra directamente abajo de los hoteles cercanos.
2. Para cada destino muestra una lista de los hoteles más populares (presumiblemente los más vendidos). La lista es la misma para todos los hoteles en un destino dado y de hecho contiene al hotel buscado en su posición correspondiente (es decir ni se omite ni aparece hasta arriba).

■ **Otros usuarios también buscaron...**

1. Aparece en la parte inferior de la página, abajo de prácticamente todo. Su relevancia es muy baja y de hecho se omitió en los primeros análisis de la competencia porque no lo vimos.
2. En la parte inferior de la página también aparece una sección de hoteles que fueron visitados por los demás usuarios que han visto el hotel en cuestión. Este análisis no es complejo, pero puede ser pesado computacionalmente si no se utilizan métodos apropiados de *Market Basket Analysis* como el algoritmo *A Priori*, por ejemplo.

En Despegar tuvieron muchas ideas para recomendar, pero todos los criterios están separados y no colaboran entre sí. Es como tener una tabla ordenada por varios campos distintos. Los hoteles recomendados con base en lo que han visto otros clientes creemos que están desperdiciados en donde están, puesto que sólo los clientes que vayan hasta abajo de la página los verán. El sistema funciona razonablemente bien y permite movilidad hacia arriba en la categoría del hotel. El problema es que es claramente afectado por la popularidad de los hoteles, de modo que tiende a atorarse en aglomerados de hoteles populares y también a subir de categoría. El resultado es que después de algunos pocos *clicks*, casi invariablemente se llega a los hoteles más caros de la zona, sin posibilidad de explorar otros.

## **Expedia (expedia.com)**

En Expedia se le da una prioridad bajísima a las recomendaciones, puesto que están posicionadas en un lugar casi imposible de ver. Sin embargo, tienen un sistema diferente a los demás y por eso lo pusimos en esta posición en la lista de sistemas competidores.

■ **Búsquedas recientes**

1. Se encuentran en el centro, casi hasta abajo, después de toda la información del hotel. La prioridad es muy baja.
2. Es igual que en las demás páginas. Aparecen los últimos cuatro hoteles buscados distintos del actual (en caso de repetir una búsqueda).

#### ■ Hoteles recomendados

1. Se encuentran directamente abajo de las búsquedas recientes.
2. Utilizan un método desconocido para recomendar hoteles similares que permite movilidad progresiva entre categorías. No estamos seguros si tiene alguna restricción de precios.

El sistema de recomendación de Expedia nos gustó bastante y de hecho creemos que lo están desaprovechando. Permite explorar los hoteles disponibles con hacia arriba y hacia abajo en la escala de precio y categoría y llegar a opciones completamente nuevas. Es probable que esté acotado por destino, aunque eso no es tan grave. Algo que no nos gustó tanto es que tal vez es algo rígido y después de buscar un poco se puede atorar. A diferencia del algoritmo de Booking, esto no necesariamente lleva a los hoteles más caros de la zona.

### Trip Advisor ([tripadvisor.com](http://tripadvisor.com))

Como decíamos arriba, Trip Advisor no es la competencia directa de Best Day, pero es un sitio muy popular de evaluación de hoteles y creemos que podemos aprender de ellos.

#### ■ Hoteles similares

1. Se encuentran en la derecha, abajo de la foto principal, pero al lado de otra información importante, por lo que el impacto es medio.
2. No sabemos cómo hacen las recomendaciones, pero no son por distancia estricta y sí permiten cierta movilidad de la categoría.

El sistema de TripAdvisor también funciona razonablemente bien, pero parece estar más basado en la calificación de los usuarios que en la categoría del hotel. Si bien ese enfoque tiene sentido en el contexto de su página, Best

Day no tiene información tan detallada como ellos en cuanto a lo que opinan los usuarios y además el enfoque es distinto. Algo que tampoco nos gustó es que el sistema es bastante rígido y se puede quedar atorado fácilmente en ciertos grupos de hoteles.

## **Best Day (bestday.com.mx)**

Finalmente, analizaremos lo que se hace actualmente en Best Day. Del mismo modo que con la competencia, comenzaremos con una breve descripción:

### **■ Hoteles similares**

1. Los hoteles similares están abajo de la liga al mapa del lado derecho. Como en la mayoría de los competidores, no se ve inmediatamente al entrar en la página, por lo que tiene prominencia media-baja.
2. La manera de recomendar hoteles es por categoría (estrellas) y por destino, es decir, se toman los hoteles de las mismas estrellas y en el mismo destino que el hotel buscado y se presentan en el Sort Order del momento. No se toma en cuenta el precio ni la distancia (salvo por el hecho de encontrarse en el mismo destino) de las opciones.

### **■ Búsquedas recientes**

1. Se encuentran directamente debajo de los hoteles similares.
2. Similarmente a la competencia, aparece una lista de los hoteles que el usuario visitó recientemente. Las búsquedas recientes permanecen entre sesiones dentro de una misma computadora.

Al parecer las búsquedas recientes son (casi) un estándar en la industria, por lo que es confortante saber que Best Day es competitiva en este ámbito. Por otro lado, la calidad de las recomendaciones es muy variable. En zonas homogéneas y con mucha oferta para una categoría dada, las recomendaciones pueden ser razonablemente buenas. Sin embargo, en destinos donde hay hoteles muy distintos pero con la misma categoría, como podrían ser un hotel citadino y uno de playa, las recomendaciones dejan mucho que desear. Además, si no hay o casi no hay hoteles bajo este criterio, no se muestra nada, o casi nada. Más adelante expondremos estos temas con más detalle.

## 1.2. Diagnóstico

Antes de describir lo que nos gustaría que hiciera el algoritmo de recomendación, resumiremos lo que aprendimos de analizar las diversas páginas. En primer lugar, nos dimos cuenta de que poner una sección con las búsquedas recientes es algo muy deseable y relativamente fácil de implementar. De hecho, es un estándar en la industria, pues todos los competidores lo tienen, a excepción de Price Travel. Sin embargo, recomendar hoteles que el cliente ya vio no cuenta realmente como recomendaciones y debería manejarse en un plano separado. Posiblemente se podría utilizar información de los hoteles visitados en las recomendaciones, pero por ahora proponemos dejar intacta esta sección.

Quitando las búsquedas recientes de en medio, podemos analizar los diversos sistemas de recomendación. Tal parece que utilizar el destino como punto de partida es una opción razonable. Sin embargo, lo que realmente importa en la vida real es la distancia y similaridad geográfica entre los hoteles; los destinos son particiones más o menos arbitrarias de la geografía que no deberían afectar las recomendaciones. Por ejemplo, se puede recomendar un hotel en Cancún a dado que se buscó un hotel muy cercano al norte de la Riviera Maya o uno en Playa del Carmen a una persona que buscó algo en Cozumel.

Otro punto que pensamos que es importante es el precio. Hasta donde pudimos ver, ninguna página tiene miramiento alguno con el costo de los hoteles recomendados. El resultado es que pueden recomendar hoteles del doble o incluso triple de precio que el original. Creemos que esto no promueve que el cliente explore más alternativas, ya que la mayoría basa su decisión fuertemente en el precio de los hoteles. Propondremos una forma simple de lidiar con este problema sin perder flexibilidad en las recomendaciones.

Un tercer problema con el que nos encontramos en las recomendaciones es que en la mayoría de los casos eran estáticas. Es decir, son idénticas para ciertos grupos de hoteles. Esto pasa en todas las opciones que vimos que utilizan el destino (Booking, Despegar, Best Day) como criterio, pues no dependen también del hotel. Esto en sí mismo puede ser un problema en destinos amplios, además de que el usuario agota las recomendaciones rápidamente, volviéndolas inútiles.

En cuarto lugar, nos dimos cuenta que en varias de las páginas se han dado cuenta de que quieren recomendar hoteles vía varios criterios, pero muchas veces no los integran. La estrategia tiende a ser usar un criterio para ordenar

a la vez, en lugar de integrarlos. Esto contribuye al tercer punto, pues hay poco cambio en las recomendaciones de hotel a hotel.

En quinto lugar, notamos que en general los sistemas que sí recomendaban de forma más sofisticada resultaban rígidos y/o se podían quedar atorados fácilmente en grupos aislados de hoteles. Esto hace que el sistema se vuelva inútil una vez que se llegó a uno de estos grupos absorbentes. El único que nos pareció decente en este sentido es el de Expedia, aunque sí es algo más rígido de lo que nos gustaría. Otro punto sobre los sistemas en general, pero de los de hoteles similares en particular, es que en ninguna página (salvo tal vez Trip Advisor) aparecen en posiciones de importancia alta. En el mejor de los casos aparecen una página abajo de la principal, al lado, y en los peores casos están hasta abajo. Creemos que están desaprovechando mucho el potencial de dichos sistemas.

### 1.3. Lista de deseos

## Capítulo 2

### Modelo propuesto

- 2.1. Puntos a resolver
- 2.2. Distancia
- 2.3. Precio
- 2.4. Similitud de hoteles
- 2.5. Distancia revisitada
- 2.6. Integrando todo
- 2.7. Comentarios

No es personalizado, sino de hoteles similares. [Ver presentación para qué sí/no es o podría ser]

## Capítulo 3

### Obtención de la información

- 3.1. Formato necesario para el modelo
- 3.2. Calidad de la información
- 3.3. Sistema de respaldo



## Capítulo 4

## Desempeño

## Capítulo 5

### Trabajo futuro

Porcentaje dinámico en forma de  $u$  para el tope de precio.