

# Sistema de recomendación de hoteles similares

## TESIS

Felipe Gerard Valdés



INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

Abril de 2016

# Agenda

1 Planteamiento

2 Solución

3 Implementación

4 Desempeño

# Agenda

1 Planteamiento

2 Solución

3 Implementación

4 Desempeño

- Modelo implementado en **Best Day** (agencia de viajes online)
- **Situación inicial:**
  - El sistema de recomendación es estático
    - Criterio: Destino + Estrellas
  - Los clientes potenciales no encuentran el hotel que buscan fácilmente
  - Por lo tanto, pueden abandonar la página o no encontrar la mejor opción para ellos
  - Esto impacta los KPIs de la empresa
- **Solución propuesta:** Recomendar con un criterio dinámico que tome en cuenta las principales características del hotel seleccionado

# Análisis de la competencia

- Competidores principales: Price Travel, Booking, Despegar, Expedia, Trip Advisor
- Diversas formas de recomendar:
  - Hoteles recientemente vistos
    - No son recomendaciones genuinas
  - Hoteles cercanos / en el mismo destino
    - Estilo y precios pueden ser muy distintos
  - Recomendaciones con métodos desconocidos (p.ej. “a otros usuarios les gustó...”)
    - Estilo y precios pueden ser muy distintos
- **Conclusión:** La competencia se enfrenta a los mismos retos y no los ha logrado resolver del todo

# Consideraciones

- Evitar mostrar hoteles demasiado caros
- Tomar en cuenta el perfil del hotel
- Incorporar información geográfica más detallada que el destino
- Diseñar un criterio que se adapte a zonas con diferentes densidades de hoteles
- Evitar que el sistema se estanque en las mismas recomendaciones (efecto laberinto)

# Agenda

1 Planteamiento

2 Solución

3 Implementación

4 Desempeño

# Plan de ataque

- Dados los objetivos anteriores, proponemos las siguientes medidas:
  - 1 **Precio:** Filtrar hoteles demasiado caros
  - 2 **Similitud:** Construir un *criterio integral de similitud*
    - Cantidad de servicios
    - Perfil similar
  - 3 **Distancia:** Recomendar hoteles geográficamente cercanos



# (1) Precio

- Poner hoteles demasiado caros no promueve el interés del cliente
  - Generalmente más barato es mejor (siempre y cuando tenga la categoría suficiente)
- Porcentaje máximo por encima del hotel original
- Hoteles por encima del precio establecido son *invisibles*

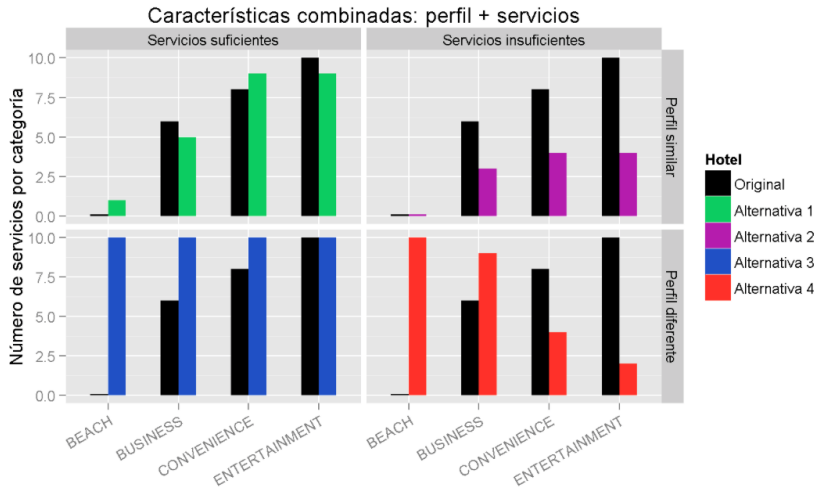
## (2) Similitud (P1)

- Las estrellas no bastan como medida de similitud entre hoteles porque no toman en cuenta el estilo del hotel
- **Alternativa:** Usar información de servicios
- **Complicación:** Usarla directamente es ruidoso porque no toma en cuenta sustitutos
- **Solución:** Agrupar los servicios en categorías
- **Método:** Manual (criterio de negocio) + aglomerado jerárquico
- **Resultado:** Cada hotel tiene un vector que lo caracteriza con el número de servicios en cada una de las categorías

## (2) Similitud (P2)

- El siguiente paso es definir una medida de similitud en el espacio de los vectores de características
- Hay dos aspectos importantes que queremos controlar:
  - **Servicios:**
    - Al menos ciertos servicios ( $\sim$  estrellas)
    - Los servicios faltantes restan similitud
    - Los servicios extras no se penalizan
  - **Perfil:**
    - Hoteles del mismo estilo
    - Misma proporción de cada categoría
    - Ignorar la cantidad de servicios
    - Podría dar hoteles más sencillos o más lujosos

## (2) Similitud (P2)



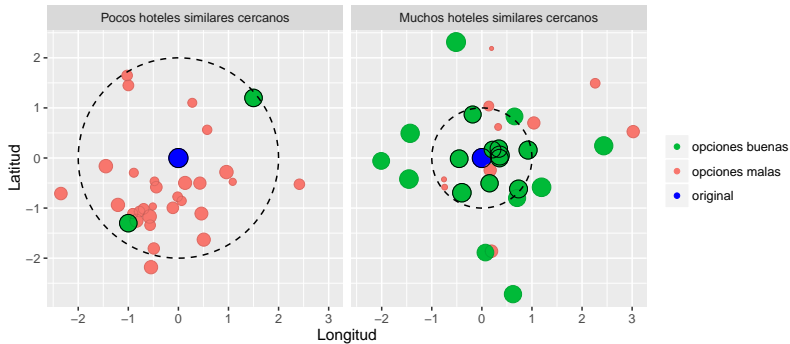
## (2) Similitud (P3)

- Medida de similitud de servicios: Garantiza que las recomendaciones tengan una cantidad razonable de servicios
- Medida de similitud de perfil: Promueve que los hoteles alternativos sean del mismo estilo que el que se está viendo
- Medida de similitud final:
  - Capturar ambos efectos
  - Promedio ponderado entre ambas medidas
  - Ponderación de acuerdo a la variabilidad de cada una (PCA)
- ¿Por qué no diferencia absoluta?
  - Es simétrica  $\implies$  Es demasiado agresiva con los hoteles mejores
  - No distingue entre perfil y servicios  $\implies$  No es ajustable

### (3) Distancia

- Hoteles cercanos (usando coordenadas)
- La definición de “cercanía” depende de la densidad de hoteles en la zona
- Se necesita un criterio que se adapte a los cambios de escala
- Hay que evitar que los hoteles distintos (pero cercanos) introduzcan demasiado ruido

### (3) Distancia



# Modelo completo

- 1 Agrupar los servicios en categorías
- 2 Calcular la ponderación entre servicios y perfil
- 3 Definir una **geocerca exterior** inicial (30 km  $\Rightarrow$  máx. 1 hr en coche) alrededor de cada hotel
  - Esto disminuye el costo computacional, pues ya no se debe evaluar la similitud con todos los hoteles
- 4 Definir una **geocerca interior** de manera dinámica:
  - Ignorar hoteles caros dentro de la geocerca exterior (precio de largo plazo)
  - Acumular una cantidad predefinida de similitud (mayor peso a hoteles parecidos, menor a los diferentes)
  - Ignorar lo que quede afuera de la geocerca interior
- 5 Recomendar dentro de la geocerca interior de acuerdo a la **similitud**, ignorando la distancia



# Agenda

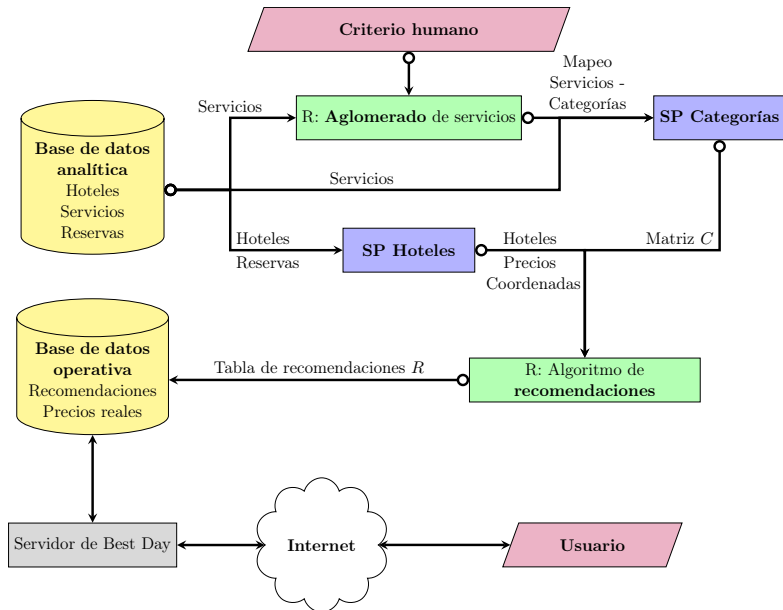
1 Planteamiento

2 Solución

**3 Implementación**

4 Desempeño

# Arquitectura: SQL + R



# Agenda

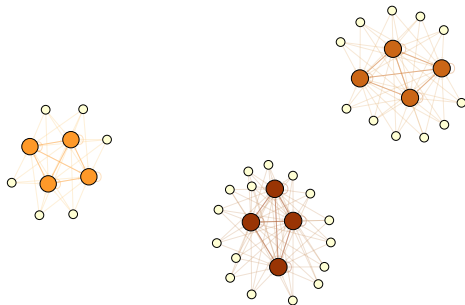
1 Planteamiento

2 Solución

3 Implementación

4 Desempeño

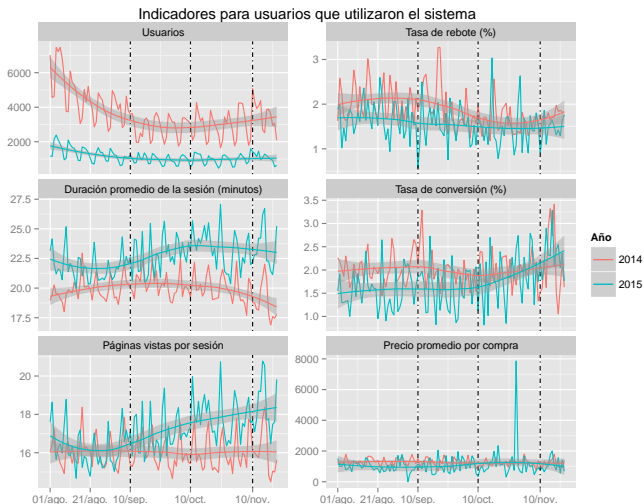
# Desempeño teórico (antes)



# Desempeño teórico (después)



# Desempeño de la página



## ■ Teóricas:

- Mejores cualidades: **precio, distancia, similitud**
- La red de recomendaciones está mejor conectada
- Mayor **flexibilidad**: se puede cambiar de categoría gradualmente

## ■ Prácticas:

- Impacto importante en los clientes que empezaron a utilizar las recomendaciones
- **Mayor uso** una vez que se utiliza el sistema
- Después de unos meses la **conversión** empezó a aumentar
- No hay aumento de entradas (diseño y posición)

## ■ Trabajo futuro:

- Precio dinámico, filtro de precio en forma de “U”
- A/B testing
- Mayor periodo de observación

¡Muchas gracias!



# Anexo: Elección de ponderación en la similitud ( $\alpha$ )

