Maps Ranking Optimization in Airbnb

Hongwei Zhang
hongwei.zhang@airbnb.com
Airbnb, Inc.
San Francisco, CA, USA

Malay Haldar
malay.halder@gmail.com
Airbnb, Inc.
San Francisco, CA, USA

Kedar Bellare
kedar.bellare@airbnb.com
Airbnb, Inc.
San Francisco, CA, USA

Sherry Chen
sherry.chen@airbnb.com
Airbnb, Inc.
San Francisco, CA, USA

Soumyadip Banerjee
soumyadip.banerjee@airbnb.com
Airbnb, Inc.
San Francisco, CA, USA

Xiaotang Wang
xiaotang.wang@airbnb.com
Airbnb, Inc.
San Francisco, CA, USA

Mustafa Abdool
moose878@gmail.com
San Francisco, CA, USA

Huiji Gao
huiji.gao@airbnb.com
Airbnb, Inc.
San Francisco, CA, USA

Pavan Tapadia
pavan.tapadia@airbnb.com
Airbnb, Inc.
San Francisco, CA, USA

Liwei He
liwei.he@airbnb.com
Airbnb, Inc.
San Francisco, CA, USA

Sanjeev Katariya
sanjeev.katariya@airbnb.com
Airbnb, Inc.
San Francisco, CA, USA

Abstract

Search results on Airbnb are presented in two user interfaces: a list of rectangle cards, referred to as the feed result, that include photos, prices, ratings, and other details of the listings, and listing pins on the map, referred to as the map result, which can either display as price pins or appear without prices as mini-pins. The map plays a key role in Airbnb. Not only does it display the location of the listings in the search results, but it also serves as an interactive user interface that allows users to view the details of the pins and perform new searches by moving or zooming on the map. Majority of searches on Airbnb are conducted using the map, and majority of bookings are from listings shown in map search results. Limited research has been conducted within the industry to address the unique challenges of maps ranking. For years, it was assumed that showing the top K results based on the model designed for feed ranking was also optimal for the map. However, this assumption simply breaks down when we take a closer look at the NDCG (Normalized Discounted Cumulative Gain) metric. Attention is key to NDCG, and the attention flow on the feed does not apply to the map. In this paper, we will begin with the NDCG theory and redesign map-specific NDCG by introducing three types of attention factors. We conducted a series of experiments to test whether optimizing map NDCG could drive the booking rates on Airbnb, and the results strongly supported this hypothesis.

CIKM '25, Seoul, Korea

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

CCS Concepts

• **Applied computing** → **Online shopping**; • **Information systems** → *Top-k retrieval in databases*; **Learning to rank**; **Evaluation of retrieval results**.

Keywords

Search Ranking, Map Search, E-Commerce, NDCG

ACM Reference Format:
Hongwei Zhang, Malay Haldar, Kedar Bellare, Sherry Chen, Soumyadip Banerjee, Xiaotang Wang, Mustafa Abdool, Huiji Gao, Pavan Tapadia, Liwei He, and Sanjeev Katariya. 2025. Maps Ranking Optimization in Airbnb. In *Proceedings of The 34rd ACM International Conference on Information and Knowledge Management (CIKM '25).* ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Since Airbnb was started in 2007, it has grown to over 5 million hosts who have welcomed over 2 billion guest arrivals in almost every country across the globe. Users come to Airbnb with a simple search query that includes their desired travel location, dates, and the number of guests. Airbnb matches these users with the most relevant listings in our inventory. Guests can browse the ranked listing cards on the results page, referred to as feed results. Additionally, they can view a map displaying the listings as either price pins (larger rectangles showing the price) or mini-pins (smaller rectangles), known as map results. The listings ranking problem in Airbnb is a bit unique compared to other ranking systems. For example, about 70% of users book less than one listing per year, making it very challenging to collect balanced data and learn personalized signals. Furthermore, booking an unsatisfactory listing not only impacts the user's short-term experience but will also potentially lead to their departure from the platform, thus we not only optimize for user's short-term conversions but also for the

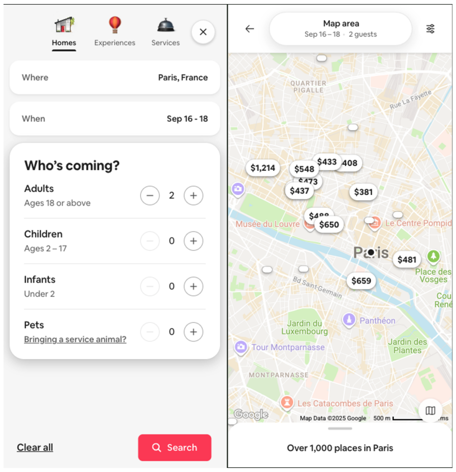

Figure 1: Searches conducted from the "search" button(left) or map zoom/move (right)

overall trip quality and retention rate. Similar challenges have also been discussed in [3] and [14] for products similar to Airbnb.

Airbnb has spent years optimizing the listing ranking system, the first deep learning model was described in [8] and there have been continuous improvements related to the model structure ([9]), diversity ([1], [7]), user journey ([16]) and so on. The ranking system has been powering not only the feed results, but also the map results.

In Airbnb, users can view a full-screen map on mobile devices and a half-screen map on web browsers(as shown in Figure 1 and Figure 8). The map plays a crucial role in helping users make booking decisions. In fact, only 20% of search requests are generated by users clicking the search button; the remaining 80% of searches come from interactions with the map, such as moving or zooming, as shown in Figure 1. Additionally, about 35% of users on Airbnb utilize the map, and about 60% of bookings are related to listings shown in map search results.

The map has been considered as a pure user interface displaying whatever is ranked on top from the feed, and it was believed that showing the top K results based on the model designed for feed ranking was also optimal for the map as well. However, this assumption simply breaks down when we take a closer look at the NDCG (Normalized Discounted Cumulative Gain) metric.

There is evidence suggesting that NDCG in feed ranking does not directly apply to maps. Firstly, user attention plays a key role in the effectiveness of NDCG. In feed ranking, user attention is modeled through a discount term, as shown in Equation (1), where higher-ranked positions receive less attention. By aligning the relevance of listings with this attention flow, the highest NDCG can be achieved in the feed ([7]). However, on the map, there is no explicit "order," meaning the attention flow modeled for the feed cannot be directly applied to the map. Secondly, randomizing the top K results affects the user experience differently on the feed and the map. On the feed, altering the order of the top K listings directly changes what users see, potentially displaying less relevant listings, which could lead to a significant drop in NDCG and booking rates. In contrast, on the map, shuffling the rank of the top K listings does not visibly alter the user interface, meaning this randomization should not impact NDCG or booking rates. Lastly, reducing the number of results displayed to users should not have positive impact on NDCG for list-ranking. However, [10] observed that decreasing the number of pins on the map led to a significant increase in bookings.

The machine learning industry has historically placed limited emphasis on map-specific ranking optimization. While studies such as [18] and [12] explored variations in the attention term within NDCG, and [2] and [19] proposed approaches to mitigate position bias, these efforts were still restricted to list-wise ranking scenarios. [10] was the first publication to specifically address learning-to-rank methodologies for maps. It started from the booking probability theory introduced in [11] and [7], conducted experiments on map listing count optimization to demonstrate that map operates differently from the feed, and also validated the "less is more" principle. It also uncovered a distinct pattern in map user attention, establishing that attention-based assumptions are a valid hypothesis for maps.

This paper is intended to be a step forward for understanding and optimizing map-based rankings. Our goal is to bridge feed and map rankings by introducing a generalized NDCG which is compatible to feed and map specific attention factors. This generalized NDCG serves as a comprehensive optimization goal, connecting all aspects of map optimization into a single framework. We will demonstrate how this metric explains observed experimental results in pervious paper [10]. Additionally, we will conduct extensive experiments to validate that optimizing the map-specific NDCG directly translates into measurable booking gains on Airbnb.

2 Generalized NDCG

In this section, we build upon the traditional list-ranking NDCG metric and propose its generalization to accommodate map-based ranking scenarios, where patterns of user attention and discount terms differ.

To enable a more direct comparison with map-based systems where only a limited number of results are visible, we adopt NDCG@K ([18], [17]) from list-ranking problems, as shown in Equation (1).

$$\mathrm{NDCG}_{List@K} = \frac{\mathrm{DCG}_{List@K}}{\mathrm{IDCG}_{List@K}} = \frac{\sum_{i=1}^{K} \frac{\mathrm{rel}_i}{\log_2(i+1)}}{\sum_{i=1}^{K} \frac{\mathrm{rel}'_i}{\log_2(i+1)}} \tag{1}$$

Here, NDCG is defined as the quotient of discounted cumulative gain (DCG) divided by the ideal discounted cumulative gain (IDCG). For a list of results ranked from position 1 to K, rel_i represents the relevance of the listing at position i, while rel'_i denotes the relevance of the listing at position i in the ideal ranking order (sorted by relevance in descending order). The term $\frac{1}{\log_2(i+1)}$ captures the decay of user attention as the position increases, reflecting the idea that users pay less attention to listings further down the list. This positional attention decay underpins the optimization strategy of placing the most relevant listings at the top positions, where the DCG can align with the IDCG to achieve the maximum NDCG.

When transitioning from list ranking to map ranking, the concept of "position" does not directly translate to the spatial arrangement of results on a map. Therefore, adopting the decay term from list ranking is not applicable. However, we can still conceptualize decay as the term "attention"(Att) and subsequently determine attention decay patterns on maps in the next sections. Equation (2) shows the generalized form of NDCG.

$$\text{NDCG}_{@K} = \frac{\text{DCG}_{@K}}{\text{IDCG}_{@K}} = \frac{\sum_{i=1}^{K} \text{rel}_i \cdot \text{Att}(i)}{\sum_{i=1}^{K} \text{rel}'_i \cdot \text{Att}'(i)} \tag{2}$$

Here, $\text{Att}'(i)$ represents the attention a listing at position i receives in ideal result. Unlike in list-ranking problems where attention at position i remains fixed regardless of the displayed results, in the maps scenario, attention varies based on the set of displayed results. Therefore, $\text{Att}'(i)$ is used in the IDCG calculation to account for this dynamic behavior.

3 User Attention Flow on the Map

Users' attention flow on the map can be measured through Click Through Rate (CTR). On Airbnb map, users express their attention by interacting with the map components. They may drag or zoom the map to explore listings in different areas, followed by clicking on pins to open listing cards containing photos, reviews, titles, and other relevant details. Each click serves as a clear indicator of user attention, marking the beginning of their booking journey. Therefore, CTR is a robust metric for quantifying user attention. Additionally, CTR serves as a bridge between map ranking and list ranking, as it aligns with the position-based decay term used in list ranking NDCG([6], [10]).

In the following sections, we will analyze three types of user attention flow on Airbnb maps and incorporate these attention patterns into Equation (2) to construct a map-specific NDCG that can be effectively optimized in section 5 and 6.

3.1 Exhaustion Attention

User attention on the map is inherently limited. In Airbnb, over 90% of map searches result in fewer than four map pin clicks, though users often perform multiple feed or map searches before making a booking. This suggests that displaying too many pins beyond the user's attention capacity can reduce the CTR for each pin and divert attention to pins with lower booking probabilities. This hypothesis is validated by an analysis of the relationship between map pin count and pin CTR, as shown in Figure 2. The data reveals a clear negative correlation between the CTR of individual pins and the total number of pins displayed. Notably, this correlation is less pronounced when the pin count is very small (e.g., 1 or 2), which supports the idea that when the pin count is within the user's attention limit, increasing the pin count does not lead to a significant CTR drop.

Based on these findings, we identify an attention pattern(Att_{exh}) associated with the number of pins displayed(N), and the number of pins that could exhaust the user's attention(N_{exh}). As shown in Equation (3), Att_{exh} remains equal to 1 when $N \leq N_{exh}$ but diminishes linearly when $N > N_{exh}$.

$$\text{Att}_{\text{exh}}(i) = \frac{min(N_{exh}, N)}{N} \tag{3}$$

This user exhaustion related attention pattern addresses the puzzle mentioned in Section 1, where significant booking gains were reported in [10] by reducing the map pin count. The underlying reason for this gain becomes clear now: prior to this adjustment, N exceeded N_{exh}. By lowering N to be closer to N_{exh}, the relevance of the N_{exh} listings with which users engage increases, resulting in an increase in DCG while the IDCG remains unchanged, as IDCG consistently represents the sum of the most relevant top N_{exh} listings. Consequently, the overall NDCG improves, which accounts for the observed booking gains.

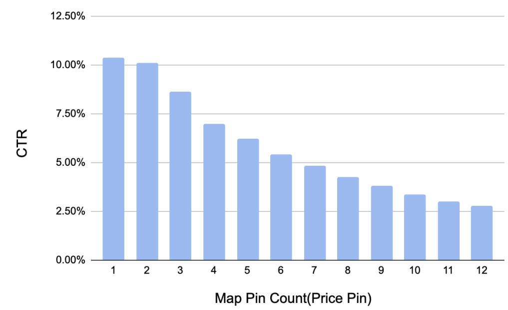

Figure 2: Map Pin CTR drops with the increase of Pin Count

3.2 Visibility Attention

On Airbnb map, pins might overlap with each other if they are close by, and listings with higher booking probabilities are shown on top of those with lower probabilities(Figure 1. right). This raises the question: do pins hidden beneath others receive less attention than those on top? To test this hypothesis, we analyzed the CTR in relation to the distance between a pin and the overlapping pin displayed above it. Our analysis shows that pins on the top has 1.6 times CTR than those hidden under other ones. We also observed a positive linear correlation between a listing's visibility ratio and its CTR, as shown in Figure 3. This means, the more visible a pin is, the higher its CTR tends to be. This insight could lead us to a user attention pattern related to map pin's visibility ratio, as shown in Equation (4),

$$\text{Att}_{\text{vis}}(i) = \max\left(\frac{min_{0\leq j<N, P_{\text{b}}(j)>P_{\text{b}}(i)} D(i,j)}{Diag \cdot \alpha} \cdot (1-\beta) + \beta, 1\right) \tag{4}$$

Here, N represents the total number of pins displayed on the map, $P_{\text{b}}(i)$ denotes the booking probability of pin i, and $D(i,j)$ represents the distance between pin i and pin j. We use a proportion(α) of the viewport diagonal($Diag$) to determine the overlapping relationship, since on Airbnb maps, a distance smaller than 5% of the viewport diagonal indicates that two pins are beginning to partially overlap, and distance of 0 means the two pins are fully overlapped. Under this configuration, user attention starts at 1 for pins that do not overlap with others and gradually decreases to β when the pins are fully overlapped, simulating the diminishing attention effect caused by overlapped pins in Figure 3.

3.3 Map Center Attention

Although not immediately apparent, our another analysis uncovered a surprising trend: the CTR of a map pin is significantly impacted by its proximity to the center of the map's viewport. Pins positioned closer to the center tend to have higher CTR, as shown

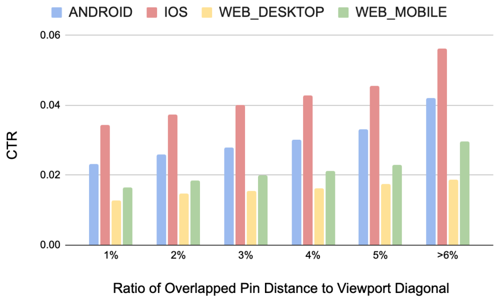

Figure 3: Distance to the the nearest overlapping Pin and CTR

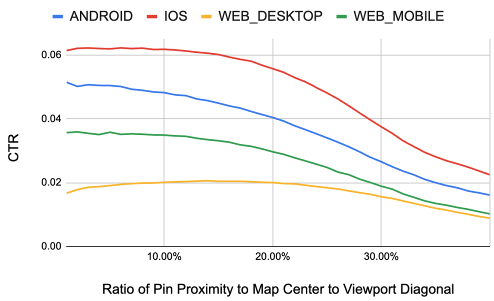

Figure 4: CTR is higher at map center

in Figure 4, where we could see that CTR decreases as the distance of a pin from the map center increases. This decay pattern is not linear: the decline is gradual for pins near the center but becomes much steeper for pins closer to the map's border. To represent this non-linear attention decay, we propose modeling it using a sigmoid-based function. This function captures the slower decay near the center and the sharper drop-off near the edges, as shown in Equation (5):

$$\mathrm{Att}_{\mathrm{ctr}}(i) = \frac{1}{1 + e^{\gamma \cdot \left(\frac{D(i,C)}{\mathrm{Diag}\cdot 0.5} - 1\right)}} \cdot (1 - \lambda) + \lambda \tag{5}$$

Here, C is the map center, $D(i, C)$ is the distance of listing i from the map center. This distance is normalized by half the viewport's diagonal($Diag * 0.5$), ensuring consistency across different map search viewports. And also, γ is a hyperparameter that controls the rate at which attention decays from the center to the edges. With this setting, attention starts at 1 for listings at the map center and gradually decreases to λ as the pin approaches the viewport's corners, simulating the diminishing attention effect caused by map center focus in Figure 4.

4 Map Ranking NDCG

With the findings of the three attention patterns, the Map NDCG can be represented as,

$$\mathrm{NDCG}_{\mathrm{Map}@K} = \frac{\sum_{i=1}^{K} \mathrm{rel}_i \cdot \mathrm{Att}_{\mathrm{exh}}(i) \cdot \mathrm{Att}_{\mathrm{vis}}(i) \cdot \mathrm{Att}_{\mathrm{ctr}}(i)}{\sum_{i=1}^{K} \mathrm{rel}'_i \cdot \mathrm{Att}_{\mathrm{exh}}(i) \cdot \mathrm{Att}'_{\mathrm{vis}}(i) \cdot \mathrm{Att}'_{\mathrm{ctr}}(i)} \tag{6}$$

Now, we could explore ways to optimize the Map NDCG. Several factors influence the Map NDCG: the set of listings $L = \{l_0, l_1, \ldots, l_k\}$, the relevance of each listing rel_i, the exhaustion attention $\mathrm{Att}_{\mathrm{exh}}(i)$, the visibility attention $\mathrm{Att}_{\mathrm{vis}}(i)$ and the map center attention $\mathrm{Att}_{\mathrm{ctr}}(i)$. Unlike list-ranking problems, where the attention factor is independent of the set of results and aligning the booking probability with user attention simply produces the optimized NDCG, this approach is not directly applicable to maps. Map attention is influenced by the set of pins, rather than being predetermined. The pin count affects user exhaustion attention, the set of pins affects the map boundary, and the position of pins within the boundary impacts both map center attention and visibility attention. This makes it challenging to align booking probability with attention decay directly to find the optimal solution. However, we can still enumerate various combinations of the pin set and calculate their NDCG scores, then identify the set that achieves a relatively higher NDCG.

The three attention factors involve different ways of optimizing map NDCG. $\mathrm{Att}_{\mathrm{exh}}(i)$ relates to adjusting the map pin count, whereas $\mathrm{Att}_{\mathrm{vis}}(i)$ and $\mathrm{Att}_{\mathrm{ctr}}(i)$ do not directly involve changes to the number of pins. In the following sections, we will focus on validating the effectiveness of $\mathrm{Att}_{\mathrm{vis}}(i)$ and $\mathrm{Att}_{\mathrm{ctr}}(i)$, as prior work [10] has already explored experiments on pin count adjustments and the results align with the NDCG theory related to exhaustion attention in this paper, as discussed in Section 3.1.

For visibility and map center attention-related NDCG optimization, we can make some adjustments to the NDCG equation. First, the relevance term rel_i can be replaced with the booking probability $P_b(i)$, aligning with the feed NDCG used in Airbnb scenarios. Second, it makes sense to set K as the total number of pins displayed on the map. Lastly, since the optimization won't involve map pin count change, the IDCG term becomes unnecessary, as it was originally used for comparing sets of different lengths. With these adjustments, we could simply optimize the Map DCG as shown in Equation (7).

$$\mathrm{DCG}_{\mathrm{Map}@K} = \sum_{i=1}^{K} P_b(i) \cdot \mathrm{Att}_{\mathrm{vis}}(i) \cdot \mathrm{Att}_{\mathrm{ctr}}(i) \tag{7}$$

Since we aim to alter the listing set and select the one with the highest DCG, we should note that the baseline set where we use the top N listings with the highest booking probabilities, already achieves the maximum $\sum_{i=1}^{K} P_b(i)$. Therefore, to further optimize the DCG, we must explore ways to either enhance the attention factor or trade off $P_b(i)$ with attention. Moreover, to validate the individual contributions of visibility attention and map center attention, we will evaluate them independently in the following sections.

5 Visibility Attention Optimization

Map pin visibility can be improved through two approaches: zooming in on the map or selecting a geographically diverse set of pins. The zooming-in approach applies to searches initiated by the search button, where the map area is determined by the outer boundary of the displayed pins. Modifying edge pins by replacing them with ones closer to the viewport center can effectively zoom in the map and enhance pin visibility, though at the cost of reduced booking probability. For searches initiated by map interactions (such as zooming or moving), where the map area remains fixed, selecting a geographically diverse set of pins becomes a viable strategy. We will explore both approaches and discuss the results in the following sections.

5.1 Swap The Outlier

In order to trade off the visibility attention with the booking probability, we implemented a "Swap The Outlier" algorithm (as shown in Algorithm 1). This algorithm identifies the furthest pin and swaps it with a closer pin, then evaluates the $\mathrm{DCG}_{\mathrm{Map}@K}$ in Equation (8). By doing this iteratively, we select the best set of pins with optimized DCG. We also integrated a side logic with the ability to further geographically diversify the price pins on top of the zoom in affect, as a reference to check if balancing more to pin visibility would hurt booking or not.

$$\mathrm{DCG}_{\mathrm{Map}@K} = \sum_{i=1}^{K} P_{\mathrm{b}}(i) \cdot \mathrm{Att}_{\mathrm{vis}}(i) \tag{8}$$

An example of the two sets of pins, before and after optimization,

Algorithm 1: Swap The Outlier

input : candidate_pin_set $L = \{l_1, l_2, \ldots, l_M\}$
price_pin_set $L^{\mathrm{price}} = \{l_1^{\mathrm{price}}, l_2^{\mathrm{price}}, \ldots, l_{N^p}^{\mathrm{price}}\}$
mini_pin_set $L^{\mathrm{mini}} = \{l_1^{\mathrm{mini}}, l_2^{\mathrm{mini}}, \ldots, l_{N^m}^{\mathrm{mini}}\}$
map_bound $B \leftarrow \mathrm{Boundary}(L^{\mathrm{price}}, L^{\mathrm{mini}})$
$DCG_{\mathrm{max}} \leftarrow \mathrm{DCG}(L^{\mathrm{price}}, L^{\mathrm{mini}}, B)$
$maxSwap_{\mathrm{price}} = 3$, $maxSwap_{\mathrm{mini}} = 10$
price_pin_set_final $L_{\mathrm{final}}^{\mathrm{price}} = L^{\mathrm{price}}$
mini_pin_set_final $L_{\mathrm{final}}^{\mathrm{mini}} = L^{\mathrm{mini}}$
geo_diversify = $false$

output: $L_{\mathrm{final}}^{\mathrm{price}}, L_{\mathrm{final}}^{\mathrm{mini}}$

while $N^m - len(L^{mini}) < maxSwap_{mini}$ ***and*** $N^p - len(L^{price}) < maxSwap_{price}$ **do**
- $l_i \leftarrow \arg\max_{l \in L} \frac{1}{|L|-1} \sum_{l_j \in L \setminus \{l\}} \mathrm{distance}(l, l_j)$
- $L^{\mathrm{mini}} \leftarrow L^{\mathrm{mini}} \setminus \{l_i\}$, $L^{\mathrm{price}} \leftarrow L^{\mathrm{price}} \setminus \{l_i\}$
- $B \leftarrow \mathrm{Boundary}(L^{\mathrm{price}}, L^{\mathrm{mini}})$
- **if** $geo\_diversify$ **then**
  - $L^{\mathrm{price}} \leftarrow \{l \in L^{\mathrm{price}} \mid \mathrm{isVisible}(l)\}$
- $L_{\mathrm{cand}} \leftarrow \{l_i \in L \mid l_i \text{ within } B, l_i \notin L^{\mathrm{price}} \text{ and } l_i \notin L^{\mathrm{mini}}\}$
- $L_{\mathrm{new}}^{\mathrm{price}} \leftarrow L^{\mathrm{price}} \cup \left\{\text{top } N^p - \mathrm{len}(L^{\mathrm{price}}) \text{ from } L_{\mathrm{cand}}\right\}$
- $L_{\mathrm{new}}^{\mathrm{mini}} \leftarrow L^{\mathrm{mini}} \cup \left\{\text{top } N^m - \mathrm{len}(L^{\mathrm{mini}}) \text{ from } L_{\mathrm{cand}} \setminus L_{\mathrm{new}}^{\mathrm{price}}\right\}$
- $DCG_{\mathrm{new}} \leftarrow \mathrm{DCG}(L_{\mathrm{new}}^{\mathrm{price}}, L_{\mathrm{new}}^{\mathrm{mini}}, B)$
- **if** $DCG_{new} \geq DCG_{max}$ **then**
  - $DCG_{\mathrm{max}} \leftarrow DCG_{\mathrm{new}}$
  - $L_{\mathrm{final}}^{\mathrm{price}} \leftarrow L_{\mathrm{new}}^{\mathrm{price}}$, $L_{\mathrm{final}}^{\mathrm{mini}} \leftarrow L_{\mathrm{new}}^{\mathrm{mini}}$

can be visualized in Figure 5. By replacing some outlier pins with those that are closer, the map becomes more zoomed in.

An online A/B test was conducted on mobile platforms with two treatments: Treatment 1, where geo_diversify was set to false, and Treatment 2, where geo_diversify was set to true.The results are presented in Table 1. By zooming in on the map, Treatment 1

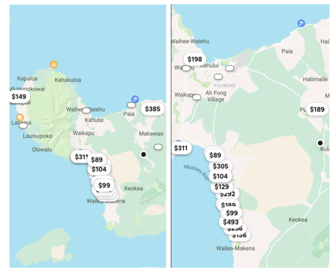

Figure 5: Improving pin visibility by swapping outliers

Table 1: "Swap the outlier" AB Test Results

Treatments strategy	Treatment 1 Zoom In	Treatment 2 +Geo-Diversify
Uncancelled Bookings	+0.28%	+0.1%
Map Zoom Changes	-1%	-0.76%
Location Outliers	-2.8%	-0.68%
Frac Within Google Bound	+0.54%	-0.36%
Avg Expansion Factor	-1.3%	-0.15%

achieved a 0.28% increase on "Uncancelled Bookings". The test also indicates significant improvements in location relevance metrics, including a 2.8% reduction in "Location Outliers", a 0.54% increase in "Fraction Within Google Bound" and a 1% reduction in "Map Zoom Changes". Compared to Treatment 1, Treatment 2 showed a smaller booking gain. This aligns with our hypothesis that excessively prioritizing attention at the expense of listing's relevance can negatively impact overall bookings.

5.2 Swap The Hidden Pin

The above "Treatment 2" to further geographically diversify the map pins provided valuable insights despite its weaker performance compared to the "Treatment 1". Specifically, it indicated that removing too many high-quality pins to increase visibility attention ultimately hurts overall bookings.

For this round, as we aim to improve visibility attention by swapping some pins hidden under others, we know from the previous "Treatment 2" that we shouldn't do this too aggressively, as it negatively impacts bookings. To mitigate the negative impact to bookings, we want to focus on searches with lower booking conversion rates. Our prior findings show that, for map searches, booking conversions primarily occur in smaller search areas, as shown in Figure 6. Therefore, we sliced the search traffic based on map area size and expected to see less booking drop and more attention benefits by focusing our geo-diversification at larger search areas.

The method of "Swap The Hidden Pin" is described in Algorithm 2. An example of the pin set before and after geo-diversification is shown in Figure 7. By excluding the pins that are hidden underneath others, we are able to display many more visible pins on the map.

Online A/B test was conducted based on two scales of map search area sizes. "Treatment 1" targeted large area map searches, while "Treatment 2" focused on smaller area searches. As we can see

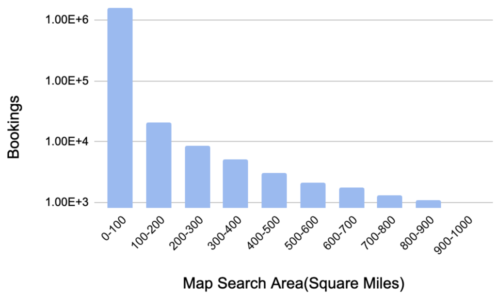

Figure 6: Bookings at different square miles for map searches

Algorithm 2: Swap The Hidden Pin

Input : candidate_pin_set $L_{all} = \{l_1, l_2, \ldots, l_M\}$,
map_pin_set $L = \{l_1, l_2, \ldots, l_K\}$,
final_pin_set $L_{final} = L$

Output : L_{final}

for $l \in L$ **do**
  if *overlap*(l, l_i) *is false for all* $l_i \in L_{final}$ **then**
    $L_{final} \leftarrow L_{final} \cup \{l\}$

while *len*$(L_{final}) <$ *len*(L) **do**
  for l *in* $L_{all} \setminus L_{final}$ **do**
    if *overlap*(l, l_i) *is false for all* $l_i \in L_{final}$ **then**
      $L_{final} \leftarrow L_{final} \cup \{l\}$

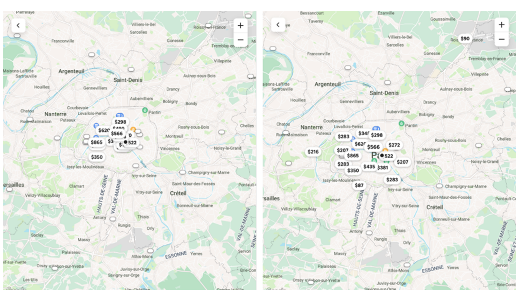

Figure 7: Improve pin visibility by "Swap The Hidden Pin"

from Table 2, both treatments demonstrated significant improvements in geographical diversity related metrics. For instance, the metric "Map Searches with Cluster Size 4 or More" was reduced by 13.3% and 29.7% for the two treatments respectively. Regarding booking related metrics, the results align with our original hypothesis: optimizing the attention factor at the cost of the relevance of the map pins has more substantial benefit for larger search areas. Specifically, we observed a 0.4% increase in "Uncancelled Nights Booked" for the large area treatment, while smaller area searches only showed a 0.17% increase.

5.3 Special Challenge on Web

The "Swap The Outlier" algorithm was initially launched on all platforms except the web browser due to negative experiment results.

Table 2: Geo-diversification AB Test Results

Treatments map area size	Treatment 1 >900 square miles	Treatment 2 other
Uncancelled Bookings	+0.26%	+0.16%
Uncancelled Nights Booked	+0.4%	+0.17%
Distinct Listings Viewed	+0.18%	+0.38%
Avg STDEV of Pins	+1.6%	+2.3%
Map Searches clustered	-13.3%	-29.7%

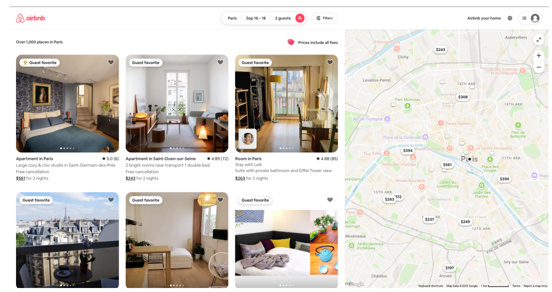

Figure 8: Search page on web desktop

The experiment showed a 0.27% drop in "Bookers That Did Not Cancel" with a p-value of 0.09, despite achieving a 12.8% reduction in the map expansion factor.

The reason behind the loss is not hard to find: the web browser is the only platform where users can simultaneously view both the feed and map results(as shown in Figure 8) and for consistency of the product, changes to the map pins must also be reflected in the feed. This creates the challenge in achieving positive bookings on the web, as the negative impact on Feed NDCG outweighs the gains on the map. The loss of NDCG in the feed is difficult to avoid, but can be mitigated. The original algorithm applied to mobile devices in Section 5.1 allows a maximum of 3 price pins and 10 mini pins to be swapped; reducing these limits can help reduce the feed-side loss, although it will also lead to a smaller gain on the map.

Two treatments are set for the A/B test. "Treatment 1" does not allow any price pin swaps and "Treatment 2" allows at most 1 price pin swap. The maximum number of mini-pin swaps is restricted to 3 for both treatments. The results are shown in Table 3. As we can see, both treatments show a significant reduction in map expansion (though not as substantial as the original version) and a neutral impact on bookings. The previously observed 0.27% loss in bookers was mitigated due to the trade-off made between the gains in the maps and the losses in the feed.

6 Center Focus Attention Optimization

Based on the above experiments, we learned that trading off lower quality pins for increased user attention can benefit overall bookings. However, swapping too many pins can diminish this positive impact, especially in smaller search areas. For this section, we will explore ways to improve map center attention while minimizing the impact on $P_b(i)$. Actually, adjusting the map boundary can achieve this goal, which allows us to enhance attention without affecting

Table 3: "Swap the Outlier" AB Test Results on Web

Treatments	Treatment 1	Treatment 2
MaxPrcePinSwap	0	1
MaxMiniPinSwap	3	3
Bookers That Did Not Cancel	-0.12%	+0.18%
Avg Expansion Factor	-4.4%	-7.6%
Frac Within Google Bound	+1.3%	+2.7%

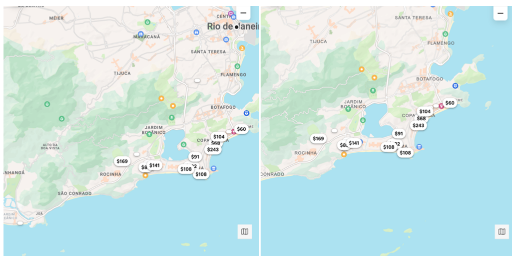

Figure 9: Improve Map Center Attention by Recentering the map

the listing set and the booking probability term. The method of "Recenter The Map" is described in Algorithm 3. The *DCG* is calculated based on the map center attention as shown in Equation (9).

$$\mathrm{DCG}_{\mathrm{Map}@K} = \sum_{i=1}^{K} P_{\mathrm{b}}(i) \cdot \mathrm{Att}_{\mathrm{ctr}}(i) \tag{9}$$

An example of the map before and after re-centering is shown in Figure 9. With this optimization, we can display more pins near the center of the viewport, enhancing user attention toward maps center with minimal impact to their relevance.

Algorithm 3: Recenter the map

Input : map_pin_set $L = \{l_1, l_2, \ldots, l_K\}$
map_bound $B = Boundary(L)$
$DCG_{max} \leftarrow DCG(L, B)$
$sw_x \leftarrow southWestLongitude(B)$
$sw_y \leftarrow southWestLatitude(B)$
$w \leftarrow width(B), h \leftarrow height(B)$
grid_size $n \leftarrow 10$
$Centers \leftarrow \{\{x_i, y_i\} \mid x_i = sw_x + iw/n, y_i = sw_y + ih/n, i, j \in \{0, 1, \ldots, n\}\}$
map_bound_final $B_{final} = B$
Output: B_{final}
for *each C in Centers* **do**
  $B_{new} \leftarrow Boundary(L, C)$;
  $DCG_{new} \leftarrow DCG(L, B_{new})$;
  if $DCG_{new} > DCG_{max}$ **then**
    $DCG_{max} \leftarrow DCG_{new}, B_{final} \leftarrow B_{new}$

Two treatments were set for online A/B testing based on the extent of the map boundary changes, measured by Intersection over Union (IoU) between the old and new viewports. "Treatment 1" targets scenarios with significant changes in the map boundary (more than 60% IoU change), while "Treatment 2" focuses on scenarios with smaller changes (less than 60% IoU diff). As shown in Table 4, there is a 0. 68% increase in "Bookers New Guests That Did Not Cancel" in "Treatment 1" and a 0. 27% increase in "Uncanceled Nights Booked" in "Treatment 2". Both of these results indicate that the optimization positively impacts bookings. There is also a slight regression in the map expansion factor, indicated by a 0.054% decrease in the "Fraction Within Google Bounding Box" for "Treatment 1" and a 0.013% decrease for "Treatment 2". This is as expected, as the map must zoom out slightly during recentering to ensure all pins fit within the viewport.

Table 4: "Recenter the map" AB Test Results

Treatments Viewport IoU	Treatment 1 IoU > 60%	Treatment 2 other
Uncancelled Bookings	+0.68%	-0.61%
Uncancelled Nights Booked	-0.08%	+0.27%
Searches low inventory	-0.58%	-1.2%
Frac Within Google Bound	-0.013%	-0.054%

7 Future Work

This paper outlines Airbnb's efforts in maps ranking, a novel research area within the recommendation track. While user attention differs significantly between feed and map ranking, both can be unified under the same NDCG metrics format. This work focuses on validating this Map NDCG theory through a series of heuristic-based experiments, and the results strongly support the hypothesis that: (1) map attention flow differs from feed attention flow, and (2) optimizing map NDCG based on map attention can optimize bookings.

The discovery of three types of map attention flow(exhaustion attention, visibility attention and center focus attention) and the corresponding map NDCG metric not only builds a standardized evaluation framework that explains previously observed results related to map pin set changes and booking outcomes but also paves the way for future research in map ranking. Continued work can focus on uncovering additional attention patterns and potentially designing attention models based on various map features. These features could include not only pin count, pin visibility and map center distance but also the information from map base layers(points of interest(POI), user engagement metrics, search context, and more). The model structure can be similar to the wide and deep framework proposed in [5] or the Bayesian network proposed in [4].

The Map NDCG metric also makes it possible to optimize the map booking rate as a whole rather than relying on existing booking probability models. Potential developments could include a graph-based booking prediction model ([13], [15]), with map pins or map base layer components (POIs, etc.) as nodes and their distances or overlapping relationships as edges. This kind of work is on our roadmap.

References

[1] Mustafa Abdool, Malay Haldar, Prashant Ramanathan, Tyler Sax, Lanbo Zhang, Aamir Manaswala, Lynn Yang, Bradley Turnbull, Qing Zhang, and Thomas Legrand. 2020. Managing Diversity in Airbnb Search. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Virtual Event, CA, USA). ACM, 2952–2960.

[2] Grigor Aslanyan and Utkarsh Porwal. 2018. Direct Estimation of Position Bias for Unbiased Learning-to-Rank without Intervention. *CoRR* abs/1812.09338 (2018). arXiv:1812.09338 http://arxiv.org/abs/1812.09338

[3] Lucas Bernardi, Themis Mavridis, and Pablo Estevez. 2019. 150 Successful Machine Learning Models: 6 Lessons Learned at Booking.com. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining* (Anchorage, USA). ACM, 1743–1751. doi:10.1145/3292500.3330744

[4] Olivier Chapelle and Ya Zhang. 2009. A dynamic bayesian network click model for web search ranking. In *Proceedings of the 18th International Conference on World Wide Web* (Madrid, Spain) *(WWW '09)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/1526709.1526711

[5] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, and Hemal Shah. 2016. Wide & Deep Learning for Recommender Systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems* (Boston, MA, USA) *(DLRS 2016)*. Association for Computing Machinery, New York, NY, USA, 7–10. doi:10.1145/2988450.2988454

[6] Aleksandr Chuklin, Pavel Serdyukov, and Maarten de Rijke. 2013. Click model-based information retrieval metrics. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Dublin, Ireland) *(SIGIR '13)*. Association for Computing Machinery, New York, NY, USA, 493–502. doi:10.1145/2484028.2484071

[7] Malay Haldar, Mustafa Abdool, Liwei He, Dillon Davis, Huiji Gao, and Sanjeev Katariya. 2023. Learning To Rank Diversely At Airbnb. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (Birmingham, United Kingdom) *(CIKM '23)*. Association for Computing Machinery, New York, NY, USA, 4609–4615. doi:10.1145/3583780.3614692

[8] Malay Haldar, Mustafa Abdool, Prashant Ramanathan, Tao Xu, Shulin Yang, Huizhong Duan, Qing Zhang, Nick Barrow-Williams, Bradley C. Turnbull, Brendan M. Collins, and Thomas Legrand. 2019. Applying Deep Learning to Airbnb Search. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Anchorage, AK, USA). ACM, 1927–1935.

[9] Malay Haldar, Prashant Ramanathan, Tyler Sax, Mustafa Abdool, Lanbo Zhang, Aamir Mansawala, Shulin Yang, Bradley Turnbull, and Junshuo Liao. 2020. Improving Deep Learning for Airbnb Search. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Virtual Event, CA, USA, 2822–2830.

[10] Malay Haldar, Hongwei Zhang, Kedar Bellare, Sherry Chen, Soumyadip Banerjee, Xiaotang Wang, Mustafa Abdool, Huiji Gao, Pavan Tapadia, Liwei He, and Sanjeev Katariya. 2023. Learning To Rank For Maps At Airbnb. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Barcelona, Spain). ACM, 5061–5069. doi:10.1145/3637528.3671648

[11] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased Learning-to-Rank with Biased Feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (Cambridge, United Kingdom) *(WSDM '17)*. Association for Computing Machinery, New York, NY, USA, 781–789. doi:10.1145/3018661.3018699

[12] Evangelos Kanoulas and Javed A. Aslam. 2009. Empirical justification of the gain and discount function for nDCG. In *Proceedings of the 18th ACM conference on Information and knowledge management* (HongKong, China). ACM, 611–620. https://doi.org/10.1145/1645953.1646032

[13] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*. https://openreview.net/forum?id=SJU4ayYgl

[14] Themistoklis Mavridis, Soraya Hausl, André Mende, and Roberto Pagano. 2020. Beyond algorithms: Ranking at scale at Booking.com. In *ComplexRec-ImpactRS@RecSys*. https://api.semanticscholar.org/CorpusID:226968062

[15] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne vanden Berg, Ivan Titov, and Max Welling. 2018. Modeling Relational Data with Graph Convolutional Networks. In *The Semantic Web*, Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam (Eds.). Springer International Publishing, Cham, 593–607.

[16] Chun How Tan, Austin Chan, Malay Haldar, Jie Tang, Xin Liu, Mustafa Abdool, Huiji Gao, Liwei He, and Sanjeev Katariya. 2023. Optimizing Airbnb Search Journey with Multi-Task Learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Long Beach, CA, USA). ACM, 4872–4881. https://doi.org/10.1145/3580305.3599881

[17] Hamed Valizadegan, Rong Jin, Ruofei Zhang, and Jianchang Mao. 2009. Learning to Rank by Optimizing NDCG Measure. In *Advances in Neural Information Processing Systems*, Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta (Eds.), Vol. 22. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2009/file/b3967a0e938dc2a6340e258630febd5a-Paper.pdf

[18] Yining Wang, Liwei Wang, Yuanzhi Li, Di He, Tie-Yan Liu, and Wei Chen. 2013. A Theoretical Analysis of NDCG Type Ranking Measures. *arXiv preprint arXiv:1304.6480* (2013). doi:10.48550/arXiv.1304.6480 arXiv:1304.6480

[19] Michael Bendersky Donald Metzler Marc Najork Xuanhui Wang, Nadav Golbandi. 2018. Position Bias Estimation for Unbiased Learning to Rank in Personal Search. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Los Angeles, USA). ACM, 610–618. doi:10.1145/3159652.3159732