Data Processing, Transformation, and Findings

All data provided (the CSVs) were loaded into a MySQL database for all cleansing, processing, and analysis. SQL and a database is my default approach for this. Some work was also done in Python to visualize, but this was secondary to the SQL approach of working with the data. MySQL was chosen just because it's what I currently have access to easily.

A general review of the data indicated that it was very uniform in distribution across its dimensions. The categories that stood out were the **first_attribution_source** and **first_attribution_channel**. All other dimensions, such as **search_term** or **search_type** or **search_dma** were nearly perfectly distributed (likely a result of generating this data programmatically for this exercise). Because of this I focused my analysis on the source and channel dimensions to see what could be found.

Queries looking at this distribution can be found in the *data_info.sql* file in the github repo.

Funnel

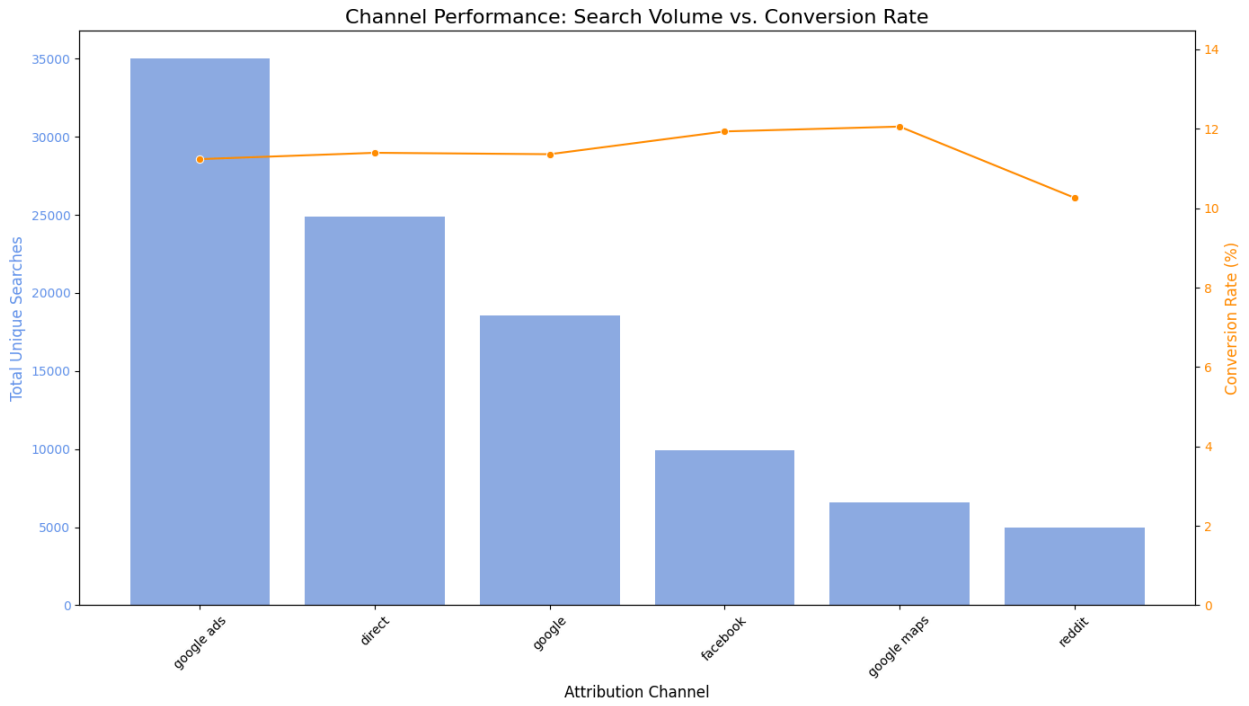
After review of the data, I attempted to create a conversion funnel from Search > Click > Reservation. It was straightforward to see the relationship from Search > Click. One search may result in a view or click on multiple results, and then to a reservation.

It appeared as though there wasn't a direct way to get from a click to a reservation by some key or indicator, but it was possible to approximate this fairly closely by comparing the reservation listing id and created date. While not a perfect solution, it was fairly accurate, with ~200 records of ~10,000 unaccounted for in the join. There were also cases where more than one reservation resulted from search, but this wasn't typical. The SQL for creating this funnel data can be found in the *conversion_funnel.sql* file.

Findings and recommendations.

Based on the above, I wanted to look at overall volume and conversion rates by channel as highlighted in the chart below.

1. First recommendation: **Increase volume in the best performing channels.**

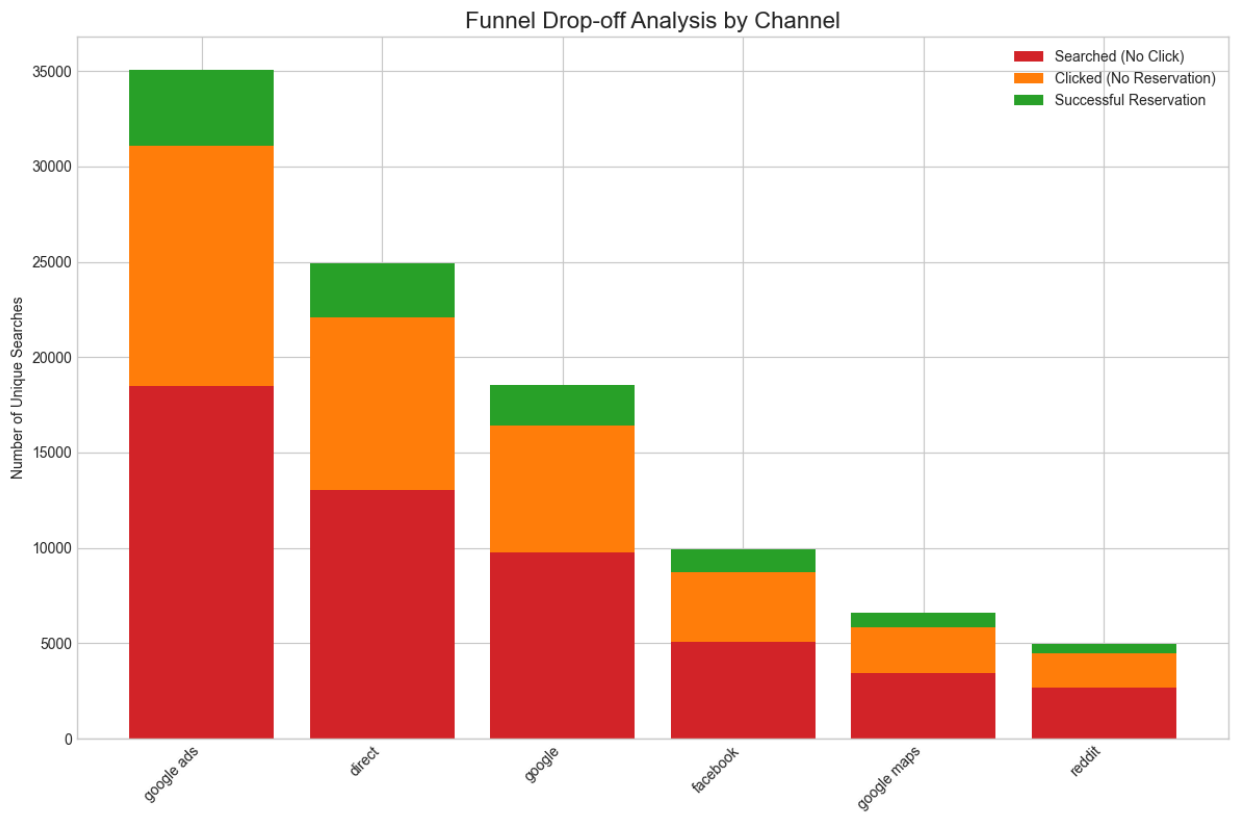


There isn't a huge difference in conversion rates, as indicated by the orange line, but the organic Google Maps channel has the highest rate of conversion with reddit being the lowest. Paid google ads dominate in volume while again reddit is lowest in volume.

This leads to my recommendation to **focus on improving organic Google Maps volume** because of its leading conversion rate. This is typically done by closely managing the Google Business Profile and actively managing reviews, using the Q&A section etc. All of this helps boost organic SEO in Google Maps. Improving the overall search results in the highest conversion category is the recommendation. I'm not exactly sure how this plays into the Neighbor model as it currently stands, however.

2. Second recommendation: **improve listing quality.**

As seen in the charts below there is significant drop off at each stage of the funnel. While not a lot of detail is available as to what could be happening, I would speculate that one area of focus would be ensuring uniform and consistent listing quality.



Funnel Performance by Channel

Channel	Total Searches	Clicks	Reservations	Overall Conversion (%)
google ads	35037	16531	3937	11.24%
direct	24913	11877	2839	11.40%
google	18547	8785	2107	11.36%
facebook	9913	4820	1183	11.93%
google maps	6611	3175	797	12.06%
reddit	4979	2327	511	10.26%

Some things I would recommend for further investigation that may lead to great conversion at each stage:

- Are there standards around listing quality such as high quality photos, descriptions, and prominent reviews?
- Pricing and fee transparency? Are reservations lost because of confusion or concern around overall pricing or commitment required?
- Is there anything in the checkout process that doesn't create a smooth flow to reservation and completion?

Because of the lack of detail, each of these points would need a deeper analysis and more data, but I think it's a good starting point.

Additional Data

To make this a more detailed and reliable analysis the following data would be helpful.

- A better way to connect reservations to listing views. In this data it isn't straightforward as far as I could tell.
- More demographics around users to help understand what might be driving behavior. Nothing is available to help understand consumer behavior. This could include things like:
    - Tenure, age, and geo of user.
    - Data specific to the relationship with Neighbor. For example: First time search vs repeat. Early indications that first time searchers perform slightly better than repeat searchers.

Funnel Performance: First-Time vs. Repeat Searches

User Segment	Total Searches	Clicks	Reservations	Overall Conversion (%)
Repeat Search	92059	43740	10338	11.23%
First-Time Search	7941	3775	1036	13.05%

- More detail on pricing of reservations. This analysis is based completely on volume metrics, but it's possible that we could consider different recommendations if we were to understand the money behind this in addition to just events. It is possible that certain channels produce much higher or lower revenue than the mean. This includes things like fee structure. Are cleaning fees or other non-rental fees contributing to final reservation dropoff?
- Ad spend. Overlaying the cost of acquisition to understand an ROI could also guide us in deciding where and how to improve certain channels.
- Listing quality metrics. As mentioned above, is there a way to understand the quality of a listing and how it performs? Number of reviews, quantity of prior reservations, or frequency of reservation?
- Host quality metrics. Being able to understand if how well a host performs (responsiveness etc) may give some indication of desirability to reserve.
- Booking platform data. Are people more likely to book on mobile vs other. How do the experiences compare on the web vs app vs mobile web?