Marketplace Data Analysis

By Giang (GiGi) Nguyen

High-level Objectives

In this analysis, we try to deep dive into the performance of the firm's 2 largest categories - **House Cleaning** and **Local Moving**. According to the dataset provided, the performance can be measured with the key metrics:

1) User engagement (clicks, contacts)
2) Visitor conversion to hire pros

By understanding the health of our product, we can make insightful recommendations to improve **monetization** and **grow our marketplace**.

We are also keen on investigating the types of pros that our visitors are more likely interested in ***(high-value pros)***. Based on these characteristics, we can optimize the **conversion of clicks, contacts and hires,** and eventually provide a **better marketplace experience** to meet more visitors needs as well as increase **retention** and **monetization** for our business.

Hypothesis Analysis

1) **Product Overview**

- When investigating the seasonality of the time-series data, we found the ***high demand for local moving near the end of the year*** since it's the cheapest time to move and more houses were highly likely to be sold within this time.
- More visitors search and hire the pros ***during the beginning of the week and during evening time***, this is surprising at first, but when we think more logically about user behaviour, they would have more time to plan ahead for cleaning & moving schedule during these periods.
- It takes a visitor ***9 clicks*** to hire a pro while he/she only needs to send out ***2 contacts*** to make the final decision.

2) **High-value Pros:**

- We defined and measured the importance of the pros metrics and found out that the most important feature to defined a high-value pros is the ***result positions*** of their listings.
- It's quite counterintuitive when we see the ***hired house-cleaning pros' average cost is higher than those of non-hired ones***, which is due to the ranking algorithm for result positions.
- We expected that hired pros will have higher rating than non-hires on average. However, the data shows that ***non-hires local-moving pros have 2% higher rating than hired pros***, which involves the problem of deflating/inflating ratings.
- While hired local-moving pros have ***2x less active time gap*** to the timestamp visitor search on the app compared to the non-hires, hired house-cleaning ***1.2x larger active time gap*** on the app than non-hire ones.

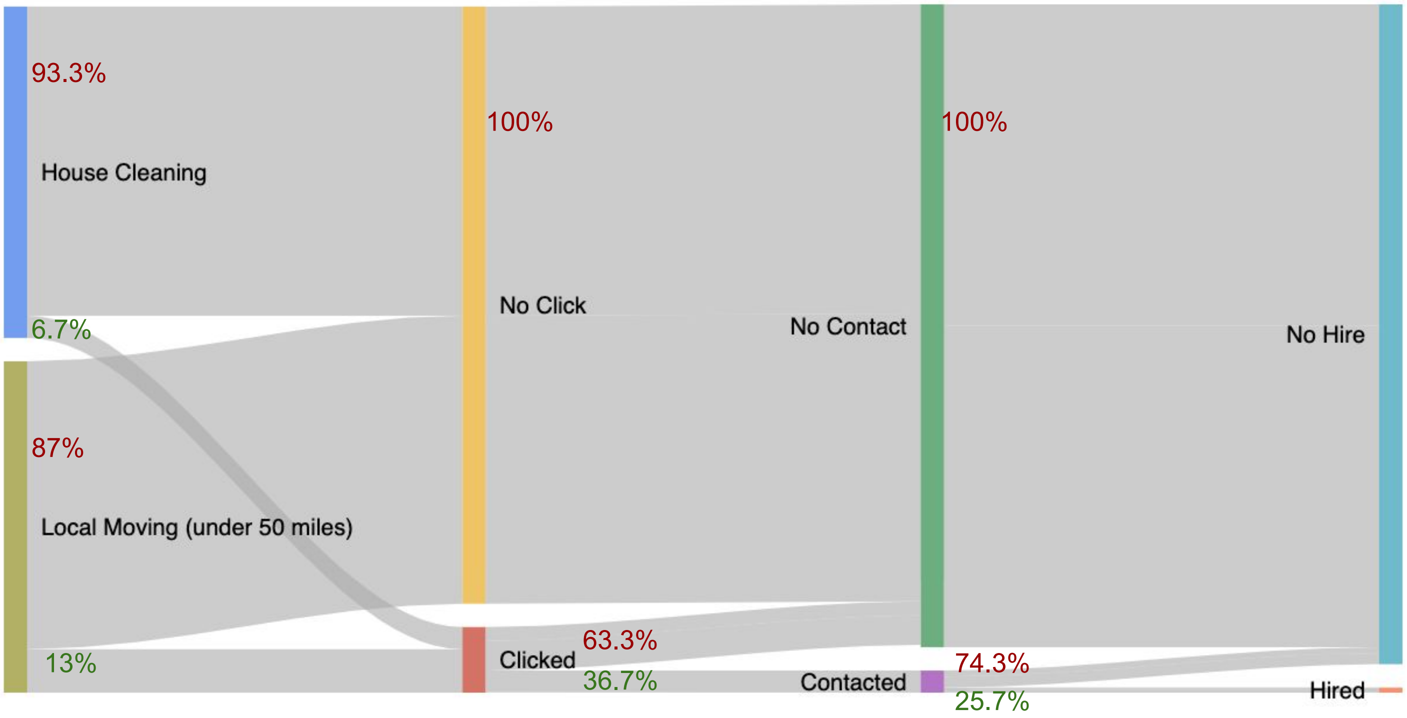
Product Overview

By looking at the product key metrics as a whole, we can identify product health, pain points, time-series and distribution analysis to improve on marketing strategies & product roadmap that impact on user growth, retention, engagement and revenue.

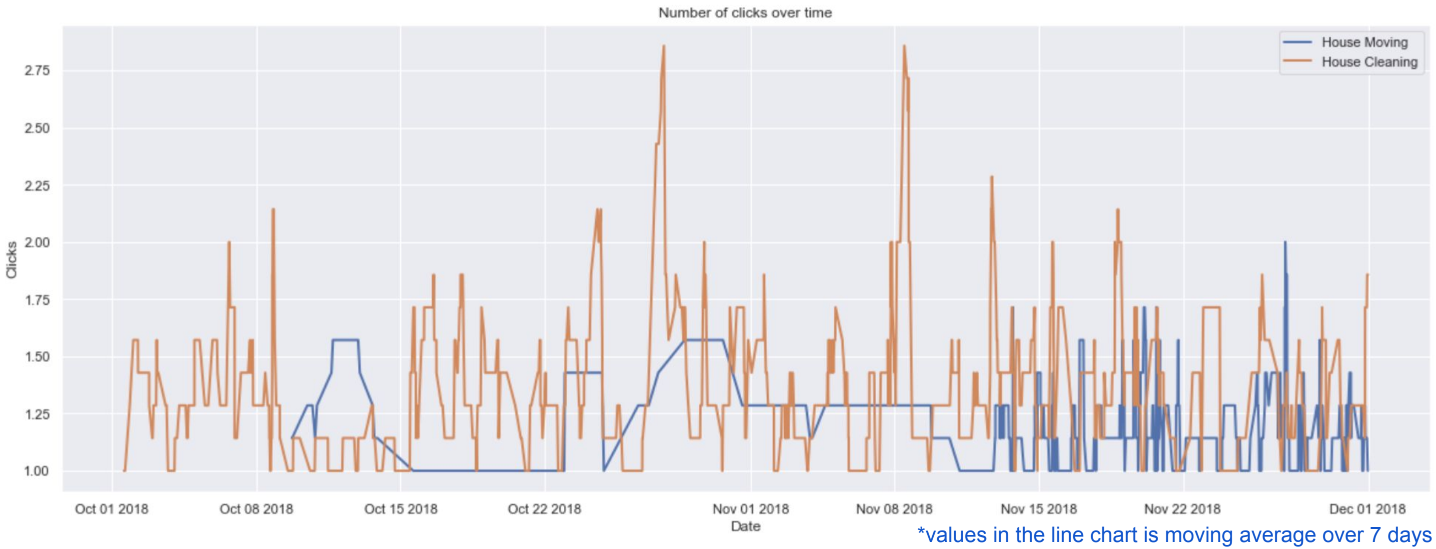
Key Analysis Insights: User Journey

In general, the conversion rate of hiring over contact is 24% and over clicks is 16%.

→ **Recommendation**: by looking at the big picture of user journey, we can identify main drop-off points is at the step where visitors click on pros profile. In order to address this churn problems, we will dig into the listing ranking algorithm and UX/UI to provide better searching experience for visitors, and ultimately increase the click through rate.



Key Analysis Insights: Time-series Analysis



*values in the line chart is moving average over 7 days

- The searches for House Cleaning peak in late October and early November. My hypothesis is that visitors are planning for upcoming Halloween parties, Thanksgiving or Christmas family gathering, etc. Therefore, there are more demand on preparing or cleaning up after the parties.
- Local Moving searches have been steady over time but increases significantly in late November. Potential hypothesis could be that (1) the price of local moving service typically drops during this time and (2) it is considered as a good time to sell houses according to Forbes so there is an uplift in demand for this service.

  → **Recommendation**: Initiate marketing campaign on each category according to the “high demand” season (late October for House Cleaning and mid November for House moving pros)

Key Analysis Insights: Time-series Analysis

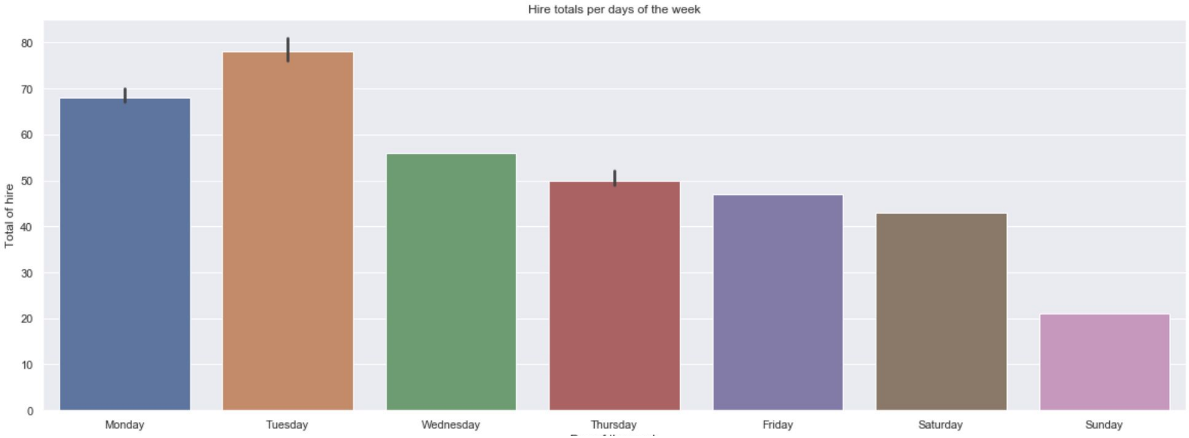
By inspecting the time-series data, we observed that:

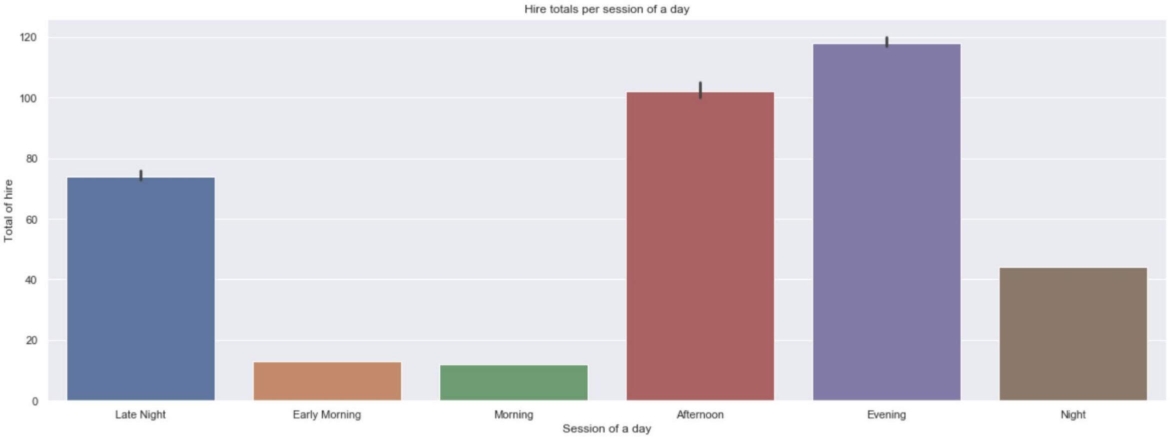
1. About 1.5x more visitors go on the app to search & hire pros on Monday and Tuesday

2. There is, on average, 2x more traffic during evening or afternoon period where visitors make decision to hire pros.

→ **Recommendation**: Hit the visitors with remarketing campaign **(*push notifications, emails*)** to remind them login the app during the high conversion windows

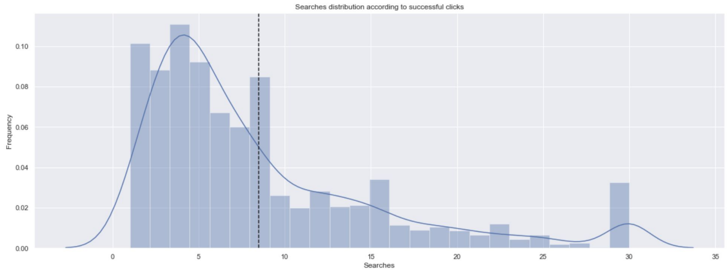
Limitation: Since the timezone is in UTC, this is a relative estimate of the seasonality since we don't have background information on the specific timezone or geo location.



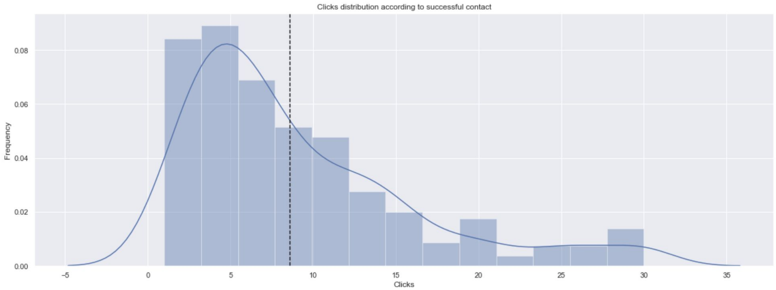


Key Analysis Insights: Distribution Analysis

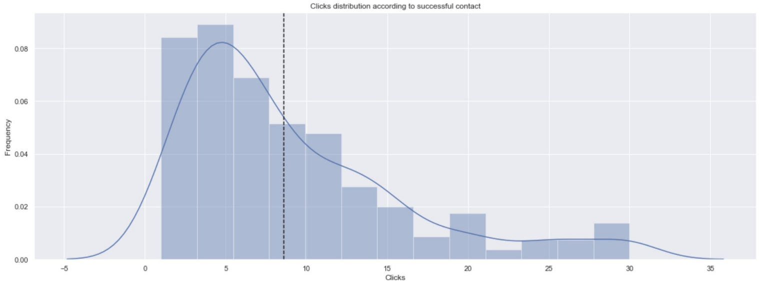
On average, there are 8.5 searches before a visitor clicks.



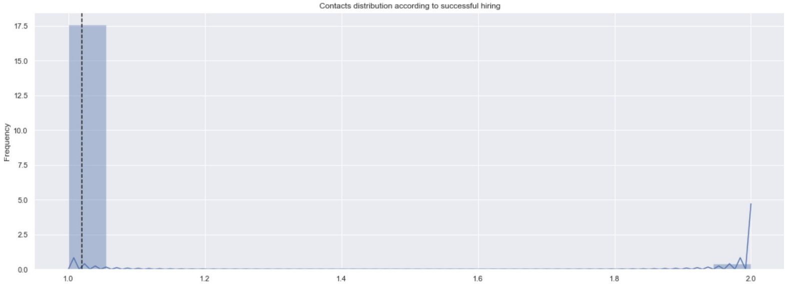
The sweet spot for hiring is after 9.1 clicks.



On average, there are 8.6 clicks before a visitor contact pros



The sweet spot for hiring is after 1.02 contacts.



By looking at this user behaviour statistics, we understand the sweet spots where more visitors are converted to hire a pro.

→ **Recommendation**:

- Send visitors call-to-action message after they perform about 9 clicks or 2 contacts.
- Improve UX/UI design to encourage visitors contacts more pros before they make decision since this will eventually help increase our revenue

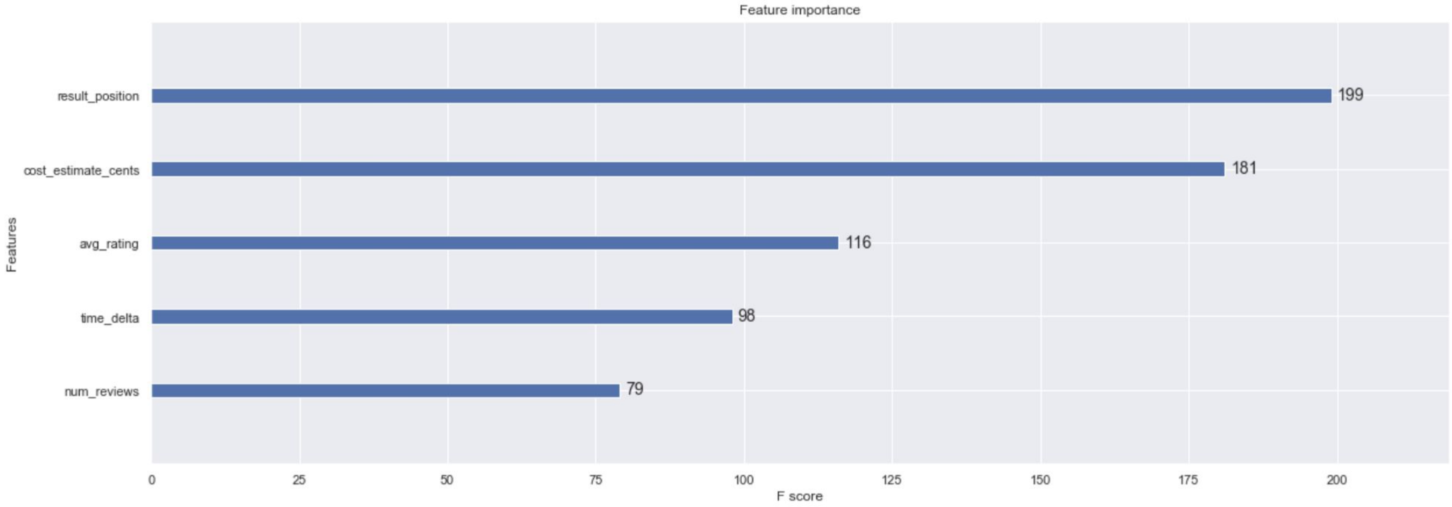
High-value pros

By identify this high-value pros, we can help pros gain the higher chance of getting contacted and hired, which eventually will generate more revenue for the company business.

Key success identifiers

Using **XGBoost Classifier**, we can estimate the feature importance and select the best predictors that highly influence the priority that visitors evaluate pro listings. By examining the feature importance, we came to the conclusion of top 3 key metrics on the sample of pros that got hired:

1. Result Position
2. Cost Estimates
3. Number of Reviews



Our goal is to find the ***synergy*** between these key metrics to define the characteristics of pros that our visitor are interested in hiring. In order to do the analysis, we will look at the summary statistics, distribution plots and the correlation of the metrics.

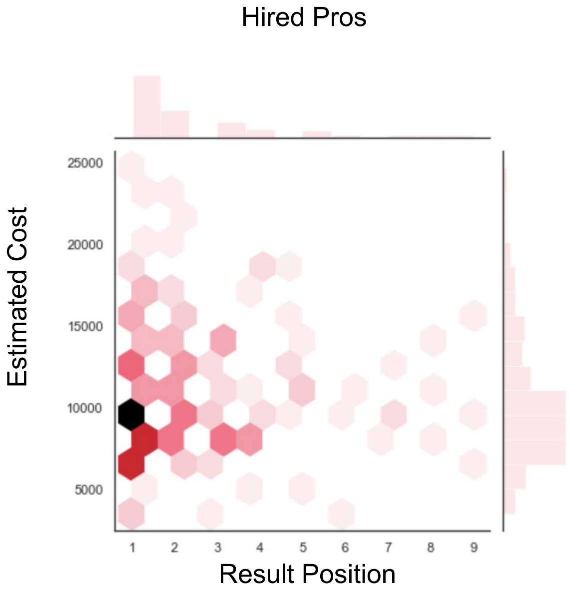
Key Analysis Insights

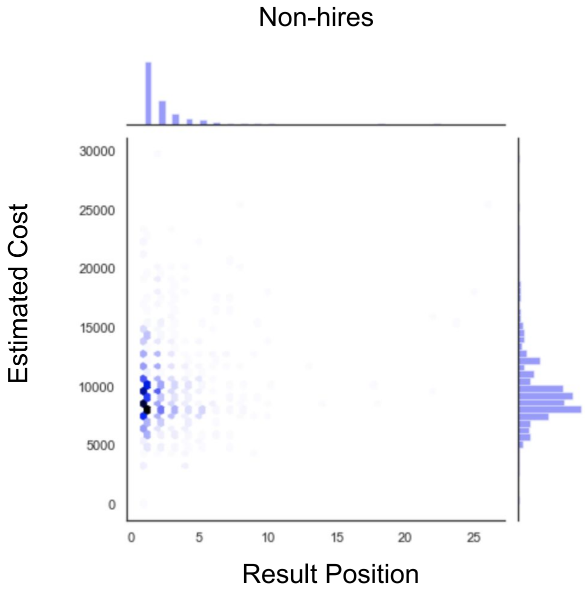
category	hired	num_reviews	avg_rating	cost_estimate_cents	result_position	time_delta
House Cleaning	False	53.87	4.77	10320.42	2.72	9917.36
	True	62.96	4.78	10617.59	2.38	11653.47
Local Moving (under 50 miles)	False	179.43	4.68	8973.03	1.72	3372.10
	True	187.95	4.66	8647.95	1.99	1319.45

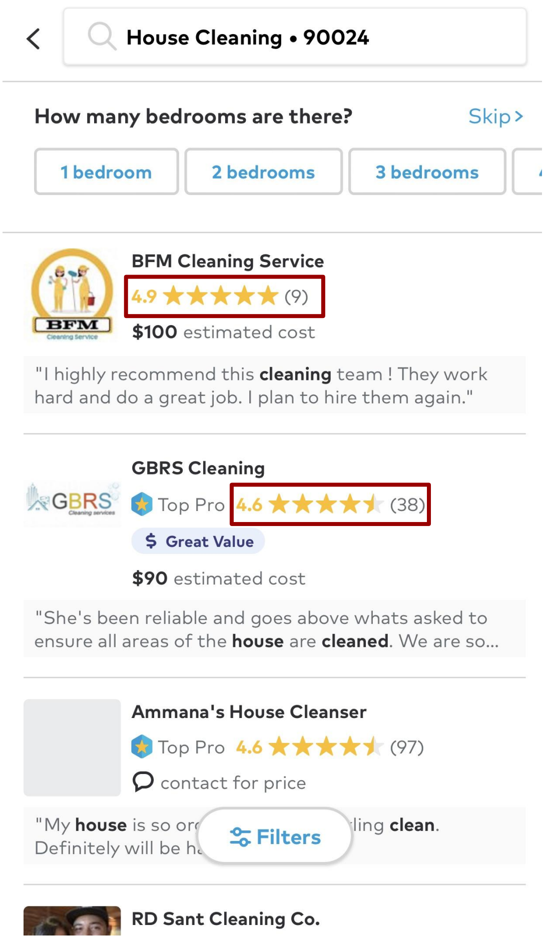
By comparing the mean of the key metrics between Hired and Non-hire pros, here are some observable insights:

- **Result Position**
  - On average, pros that were hired stands from #1 - #3 on search result.
- **Estimated Cost**
  - Even though that cost is one of the top 3 drivers for high-value pros, it's quite counterintuitive when we see the hired house-cleaning pros' average cost is $106.17 which is higher than those of non-hired ones. <u>This is due to the ranking algorithm of the listing position (review more details in the next slides)</u>
- **Number of Reviews**
  - Hired local moving pros have the mean of 180+ reviews while hired house cleaning ones only gain 55+ reviews.
- **Average Rating**
  - Despite playing an important role as top 3 key predictor, it goes against our intuition of hired pros will have higher rating on average, however, in this case, local-moving non-hires have 2% higher rating than hired pros. <u>This is due to the inflating/deflating problem where we have correlation between number of reviews and average rating (review details in the next few slides).</u>.
- **Time delta (time stamp difference between the search time and the pros last active time)**
  - While hired local-moving pros have 2x less active time gap to the timestamp visitor search on the app compared to the non-hires, hired house-cleaning 1.2x larger active time gap on the platform than non-hire ones. There is no definite explanation to this based on the given dataset, so we may investigate more into other outside behavioral factors such as comments, message frequency, etc.

Key Analysis Insights: Estimated Cost



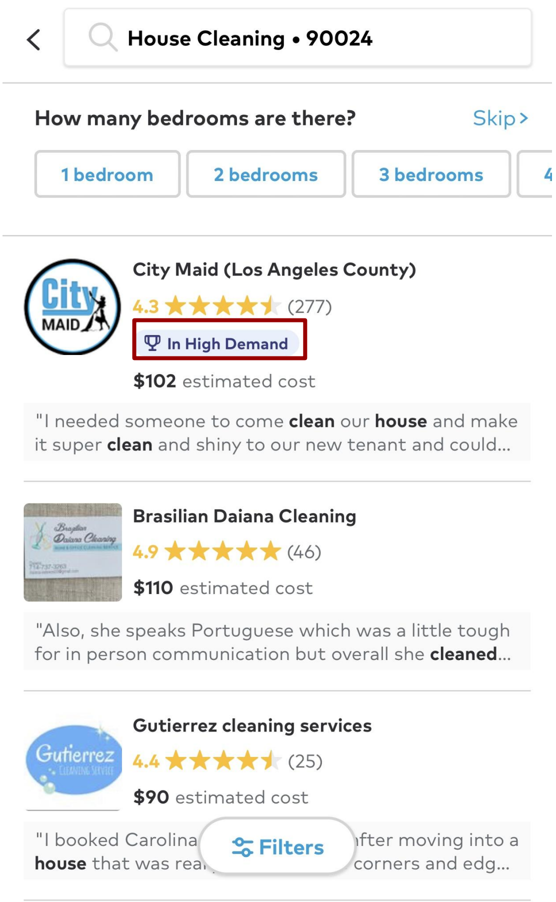




Why visitors are more interested in house-cleaning pros that charge higher price?

- This is due to the ***ranking algorithm*** of the listing. In reality, the company prioritizes avg. rating than number of reviews and estimated cost for each pro listing - as can be seen in the screenshot 👉
- As result position is the strongest predictor, visitors are more likely to interested in pros that get listed first in the search list.

Key Analysis Insights: Estimated Cost

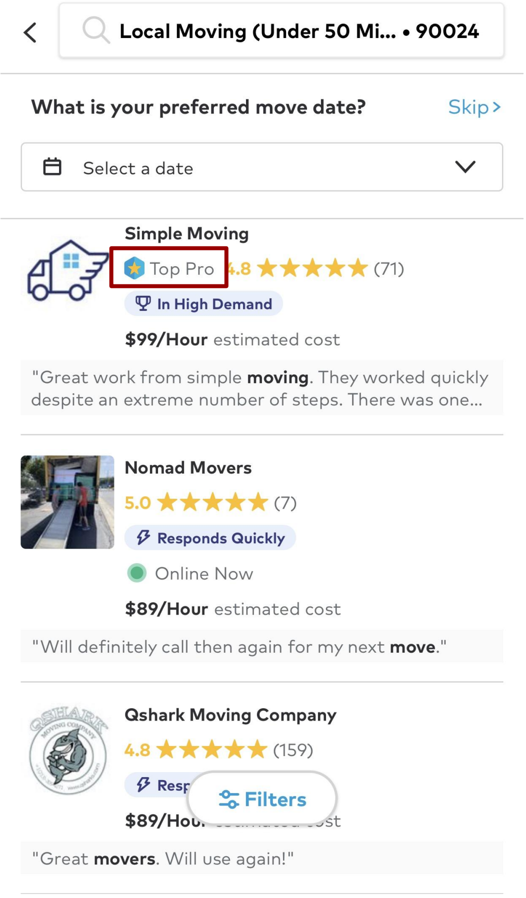


However, there are some edge cases when lower rating get listed first like these 2 situations.

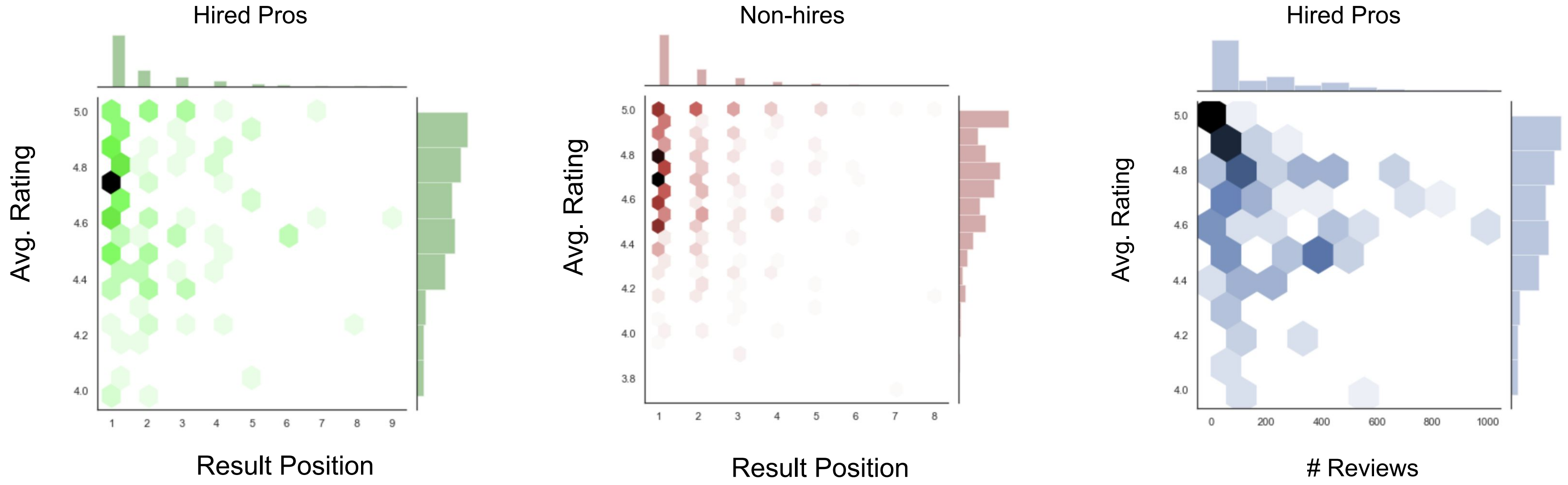
So what is lacking that our dataset does not cover?

Some pros got the special label "***In High Demand***" or achieved the honorable badge "***Top Pro***" are prioritized in the ranking in result position.

→ **Recommendation**: the company can drive pros incentive to achieve these awards on the app to yield higher conversion rate from the visitors.



Key Analysis Insights: Average Rating



Why hired and non-hire pros have equivalently the same average rating?

- As can be seen from the distribution, most hired local-moving pros have avg. rating of 4.7, while non-hires saturated on 4.7 → 5.0 ratings. My hypothesis is because of the inflating/deflating rating problem that pros with more reviews will have higher chance of receiving lower ratings.
- This hypothesis is confirmed by the average rating and number of reviews distribution chart - where we see the strong correlation between the two metrics. ***As number of reviews increases, there tends to be more lower ratings.***

→ **Recommendation**: when evaluate a high-value pros, we need to not only take into account of the avg. rating but also the number of reviews that pros have, in order to better improve the ranking algorithm for result position listing.

Recommendations: Product and Business

- Optimize ***listing ranking algorithm and UX/UI*** to provide better searching experience for visitors, and ultimately increase the click through rate.
- Initiate marketing campaigns on each category according to ***the "high demand" season***
- Improve UX/UI design to ***encourage visitors contact more pros before they make decision*** since this will eventually help increase the company revenue
- Hit the visitors with ***retargeting marketing campaign (push notifications, emails)*** to remind them login the app during the high conversion windows
- Send visitors ***call-to-action message*** after they perform about 9 clicks or 2 contacts - the sweet spots for conversion.

Recommendations: High-value pros

- Improve the ***ranking algorithm*** to prioritize the best matched result positions according to average rating, number of reviews, estimated cost, pro’s last-online time, and badges or awarded labels on app
- Initiate ***training/educating program for pros*** and drive their incentive to achieve badges and awards on the app to yield higher hiring rate
- Give personalized advice to pros on ***strategic pricing strategies***
- Send reminders or in-app messages to visitors for ***writing reviews about pros after each service performed*** to improve pros number of reviews

Future Work

Based on the exploratory data analysis on user engagement and conversion metrics, we can have a big picture of path to purchase a service on the firm and who are our most successful pro listings.

This analysis can be extended with more data points on geo location, comments, pro badges & labels, etc. to help us better understand the data and background context and provide a more comprehensive analysis.

In the future, this analysis can assist the work of fine-tune machine learning models for ranking algorithm and visitors/pros classifications to better match visitor searches and accurately predict the conversion rate.

Appendix

[Dataset]

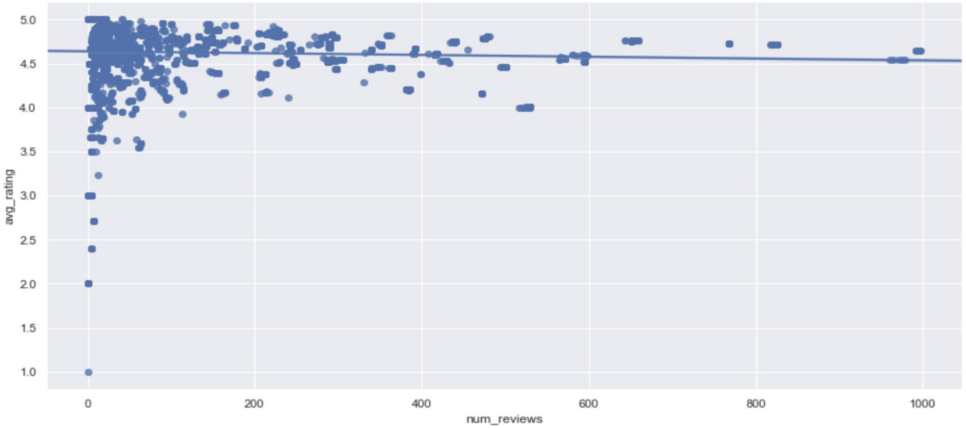
- Contacts.csv
- Visitors.csv

[Codes]

- iPython Jupyter Notebook and graphs can be found [HERE].

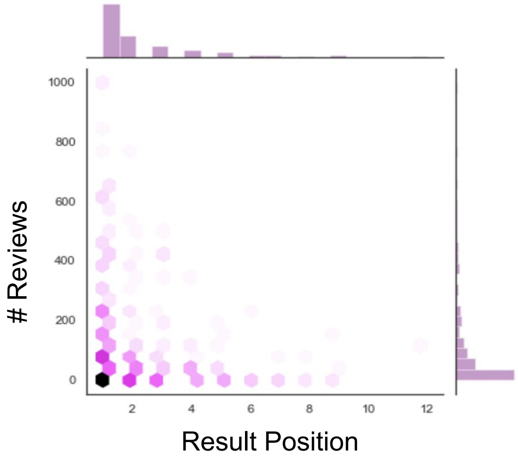
Additional Visualization

Linear Regression Analysis between number of reviews and avg. rating



We can see the in the linear model, as number of reviews increases, there tends to be more lower ratings.

Hired Pros



This strengthen our analysis of feature importance where Result Position is the most important feature. Regardless of number of reviews, visitors are more likely to hire pros with top 3 listed positions.

Analysis Timeline

Task	Date	Duration
Project Start	Oct. 8, 2019	---
Cleaning data + Feature engineering	Oct. 8, 2019	2 days
Analysis	Oct. 10, 2019	1 day
Result compilation + Writing report	Oct. 11, 2019	1 day

Thank you for your time and consideration! Hope you enjoy reading my analysis and hope this can contribute some unique and meaningful insights for the company! 🎉