Re: Best Markets

Hey marketing team,

I've looked into the data to determine our 5 best markets and have some findings I'd like to share with you.

The data supports these 5 markets as being our best, which I've defined as those that have the strongest indicators of market demand:

1. **Buffalo, NY** - We have 19% market share of listings and our homes sell for 66% above the market median.
2. **Grand Rapids, MI** - We have 16% market share of listings and our homes sell for 40% above the market median.
3. **Fresno, CA -** We have 12% market share of listings and our homes sell for 11% above the market median.
4. **Richmond, VA -** We have 9% market share of listings and our homes sell for 25% above the market median.
5. **Virginia Beach, VA** - We have 6% market share of listings and our homes sell for 32% above the market median.

Please feel free to share any thoughts you have based on what I've outlined here. I'm happy to answer any other questions you may have or set up time to review these as a group.

Attachment: Best Markets.csv

Criteria

I broke this analysis into 2 key steps:

1. **Develop an index to rank markets based on strong demand, determined by 3 key factors:**
    a. **Percent of homes that saw a pricing increase in the last 30 days**. In each market, 0 to 9% of homes on the market had a listing price increase. Markets with a higher percentage were ranked higher.
    b. **Median days on market (DOM):** Markets saw a range of 12 days to 92 days on the market. Markets with a lower DOM were ranked higher.
    c. **Avg percent increase in closing price vs listing price:** Closing prices in each market ranged from 10% below the list price to 16% above the list price. Markets with a higher percent increase were ranked higher.

2. **Compare our sales data against the strongest-performing markets (60th percentile and up), looking at 2 key factors:**

a. **Percent above the market's median price that our homes sell for.** Of the markets performing above average, our homes sell anywhere from 71% below the market median up to 197% above the market median.
b. **Our share of listings compared to the total number of listings.** Of the markets performing above average, we have anywhere from a 3% to 19% share of the total listings on the market.

The result gives us the following view of **9 key markets** that have the strongest demand **and** where we perform well.

A	F	L	Q
Market	**Our Share of the Market**	**Percent Above Median Price**	**Market Index Rating**
Buffalo, NY	19%	66%	0.81
Grand Rapids, MI	16%	40%	0.60
Fresno, CA	12%	11%	0.60
Richmond, VA	9%	25%	0.66
Virginia Beach, VA	6%	32%	0.64
Baltimore, MD	4%	25%	0.73
Kansas City, MO	4%	40%	0.72
Minneapolis, MN	3%	27%	0.68
Detroit, MI	3%	197%	0.62

Area of Opportunity
This view also gives us some insight into markets we might want to tap into for home sellers:

- Baltimore, MD
- Kansas City, MO
- Minneapolis, MN
- Detroit, MI

All 4 of these markets perform above average in our market index, and we sell homes above the median in the area. We only have 3-4% market share in these markets, so I see an opportunity for growth by increasing the number of listings we have in these markets.

The attached spreadsheet shows the full view for this analysis. With this structure, we can adjust how we calculate our ranking and put more or less weight into specific factors.

Attachment: Market Analysis.csv

Market Trends

I remember you asking if there were any noticeable market trends in the data that would be appropriate to share with an external audience. I noticed some interesting data points when looking at population and wanted to share the following:

- **Smaller markets are trending toward favoring sellers, while larger markets are trending toward favoring buyers.**
    - **8 out of 10 markets** with the highest populations **ranked below the average in the market**, largely due to long median DOM ranging from 44 to 92 days (compared to the average of 46 days).
    - **Houston, TX and Washington, DC are the only two markets in the top 10 population that rank average across the market** for demand.
        - For Houston, this is due to 4% of homes seeing a price increase in the last 30 days, compared to the market average of 2%.
        - For Washington, DC, this is due to homes sitting on the market for 45 days (in line with the average) and selling for just 1% below list price (compared to the average of 2% below).
    - Meanwhile, **7 out of 10 of the markets with the highest demand have a population below the average of 3.7M**. These homes are consistently seeing price increases and low DOM, but the sell price ranges from 5% below listing to 16% above listing.

Let me know if you'd like me to provide any additional context here.

Thank you!
Amy Brown

Re: Best Agents

Hey sales team,

I've done some research on the best agents in each of our markets. Using data we have on our agents and their sales, I've calculated the top agent based on the following criteria:

1. **Highest conversion rate (25% of ranking):** Total buyers and sellers closed out of total buyers and sellers sent to the agent. Our average for this sample was 16%.

2. **Highest customer rating (25% of ranking):** The average rating of customer reviews the agents received in 2024. Our average for this sample was 3.43.

3. **Highest sell price above market median (50% of ranking):** The percent above the market median that this agent's average home sells for. Our average for this sample was 20% above the market median.

I chose these factors thinking about how they affect our reputation and, ultimately, our bottom line.

Attachment: Top Agent by Market.csv

Data Assumptions
For this analysis, there are a couple judgment calls I made:

- **I've only sampled agents with at least 5 customer ratings.** Of the 869 agents we have in our records, 333 of them have at least 5 customer ratings. Since we're using the customer ratings as part of their ranking, I've narrowed down our agent pool based on having a bit stronger established base.

- **Is Sales representing just the home sellers?** I was working with a sales table to determine the average sell price for each agent, but I wasn't 100% sure if this data contained sales where we represented buyers and sellers. The data I've shared is based on us representing sellers only. If that isn't the case, let me know and I can adjust the rankings.

- **Sales outliers are normalized to level the playing field.** Some of our agents have sales much higher than the median for their respective markets, which can skew the data for the rest of our agents.
    - For example, Tyrone Thompson in Detroit, MI has sold 2 homes for a total of $9.6M, which is 2248% above the market median of $205K.
    - I've used a method that essentially curves the grade here, especially because this factor accounts for 50% of their ranking.

The attached pivot table shows all agents for each market (with more than 5 customer ratings) and where they rank in their market given the parameters above. Feel free to explore and let me know if you have any thoughts.

Market	Agent	2024 Conversion Rate	2024 Customer Rating	2025 Sell Price from Median	Agent Rank in Market
Atlanta, GA	Maria Garcia 191	17%	2.86	35%	1
	Arjun Robinson 452	22%	2.92	17%	2
	Nia Thompson 808	12%	2.97	33%	3
Austin, TX	Aisha Lopez 898	25%	3.00	7%	1
	Jose Hernandez 319	15%	2.30	25%	2
	Daniel Mohamed 821	11%	3.56	10%	3
	Tyrone Robinson 873	12%	2.31	22%	4
	Aisha Nelson 239	12%	2.17	14%	5
Baltimore, MD	Jason Allen 167	16%	3.43	35%	1
	Jamal Young 798	12%	4.41	23%	2
	Mei White 978	13%	3.74	25%	3
	Tanya Hill 978	22%	3.34	13%	4
	Lakshmi Young 759	18%	2.52	32%	5
	Matthew Rodriguez 166	23%	2.68	18%	6
	Kenji Hernandez 461	17%	3.21	21%	7
	Lucia Patel 846	13%	2.40	20%	8
Birmingham, AL	Hassan Washington 575	22%	3.40	64%	1
	Linda Jackson 955	15%	4.28	63%	2
	David Nguyen 46	26%	3.33	54%	3
	Haruka Khan 777	11%	4.41	65%	4
	Camila Nelson 261	11%	4.04	58%	5
	Mei Tran 609	16%	3.51	55%	6
	Lucia Smith 808	16%	3.54	46%	7
Boston, MA	Mary Baker 972	21%	4.99	-28%	1
	Imani White 686	12%	4.80	-36%	2
	Robert Jones 846	17%	3.23	-36%	3
	Jason Garcia 337	20%	2.10	-33%	4
	Sarah King 69	13%	3.25	-38%	5
	Arjun Williams 858	14%	3.09	-38%	6

Attachment: Agent Analysis pivot table.pdf

If you have any feedback on the process or any questions about the data, please don't hesitate to reach out.

Thank you!
Amy Brown

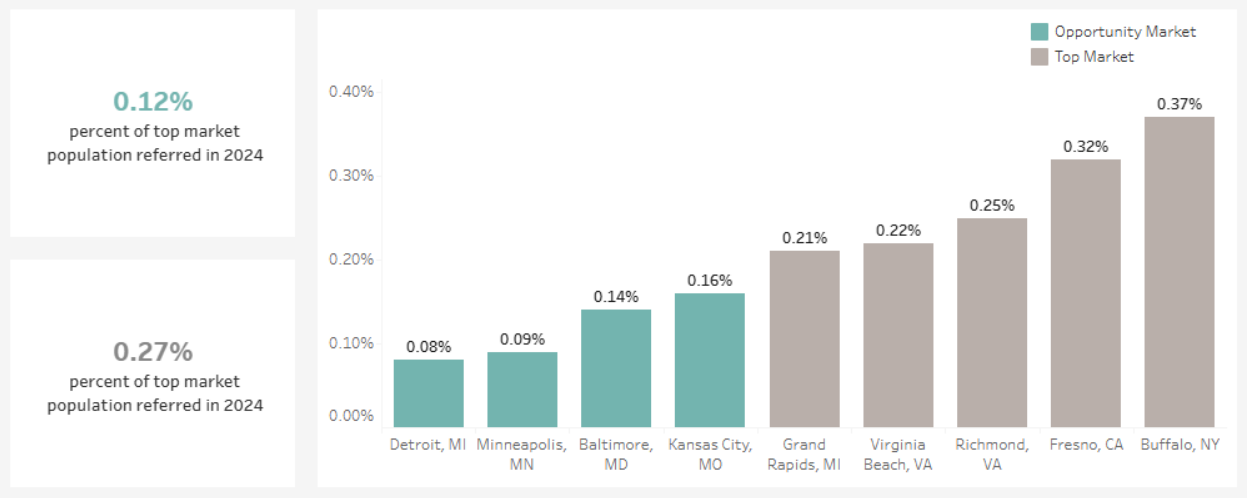
One Area for Growth

In working on the prompts for the Market Analysis and Agent Analysis, I see a compelling argument for increasing marketing efforts in the following 4 markets:

- Baltimore, MD
- Kansas City, MO
- Minneapolis, MN

- Detroit, MI

Looking at the total number of 2024 referrals as a ratio of each market's population, there's a stark contrast of .27% in our top markets compared to .12% in our opportunity markets.



The average population of these 4 markets is 3.3M, while the average population of our 5 best markets is 1.3M. Looking across the data points for all of the following, these 4 "Opportunity Market" cities consistently rank in the range of their "Top Market" counterparts:

Performance Metrics	**Top Markets**	**Opportunity Markets**
Market: Number of Agents	22	21
Market: Median DOM	27	31
Agents: % Diff from Market Median	35%	72%
Agents: Avg Conversion on Leads	16.48	15.71
Agents: Avg Customer Rating	3.58	3.34

This tells me that we already have a sales team established in our areas of opportunity that can take new leads, convert them to sales, and drive earnings for the company through selling homes above the median price in their respective market.

Our top 5 markets produced $1.4M in company earnings in 2024 (1% of home sale price), while our 4 opportunity markets produced $1.5M. Given that our top markets hold almost exactly half of the population as our opportunity markets, there's evidence that supports that our opportunity markets could bring in an additional $1.5M in 2025 if we concentrate some of our marketing efforts on bringing in more leads from these areas.

More Time & Data Sources

I took some notes as I worked through these analyses to record what ways I'd want to improve them. While this list is far from comprehensive, here are a few key areas I'd love to dig into further:

Insights & Direction

- **Collaboration with the team:** In the real world, I'd love to make a visual to help both our marketing team and sales team understand what data we have available. In my experience, teammates who have strong industry knowledge tend to have strong ideas for how the data can be used to grow and improve.
    - In that same vein, these analyses were fun challenges because of how open-ended they were. On the team, I'd like to get more context from the teams making the requests. What do they plan to do with the insights and what action(s) will they drive? This can totally pivot the approach we take in our analysis.
- **Build seasonal trends and predictions:** I'd love to dig into seasonality. We have some visibility into this in the tables provided and could start to understand the broader picture by comparing to industry trends. This could influence both our marketing and sales strategies and get us into the territory of forecasting, which I could see benefiting multiple departments across the company.
- **Days on market trends:** I'd be curious to look at industry trends for these:
    - What's the cost of a home staying on the market for longer? Can we tie a tangible cost to higher DOM vs lower DOM? Can we see a definite correlation between high DOM and pricing decreases?
    - What's the ideal and/or average amount of time for a home to be on the market? We could use this insight to adjust the index to give more or less weight to homes that deviate from a standard.
- **Look at developing trends on the sales team:** Because I filtered our Agent Analysis sample to agents with at least 5 customer ratings, we didn't dig into about 62% of our agents. I see some indications that we could uncover insights here if we sampled the full agents pool.
    - For example, our Market Analysis revealed that, on average, we sell homes for 66% above the market median in Buffalo, NY. Looking at just agents with at least 5 star ratings, we see that their average is about 59% above the market

median. This could point to some of our newer and/or less established agents driving high value for the company.

Possible Data Sets

- **More robust sales data:** For the Agent Analysis in particular, there were a few limitations in that data that I'd like to fill in and get a better picture:
    - How much were homes in our Sales table listed for?
    - When did the homes in our Sales table go on the market?
    - What were the defining features of the homes that would impact their price? Square feet, number of bedrooms, number of bathrooms, and year built are just a few fields that would give a ton more context on why homes sell above or below the market median.
- **Industry data**
    - Seasonal buying trends
    - Seasonal selling trends
    - Average DOM
    - Cost associated with higher DOM
    - Market median across home size, bedrooms, etc.
    - Avg share of market of other real estate companies
- **Additional market data:** Our Markets table contains data on 50 markets. The Market Analysis revealed that the smaller markets can perform just as well — or even better — than the markets with larger populations. There could be potential to tap into more mid-sized markets, and our Market Analysis supports that.

The Process

Overall, I enjoyed the process of digging into these questions. The data sets presented conquerable challenges and interesting data points that could lead you down many different roads. Below is a brief(ish) summary of my process.

Order of Operations

- **Create my own reference document**
    - Type up the assignment brief in my own words
    - Interpret the prompts and begin to break out the asks
    - Start notes in each section, jotting down initial strategies, questions, challenges, assumptions, etc.
- **Familiarize myself with the data**
    - Look through the data set in Google Sheets
    - Get familiar with each of the tables and their fields
    - Start brainstorming how to answer the business questions by leveraging this data
- **Register with MySQL and import tables to real_estate_data schema**

    - This is where I started to think that taking an index ranking approach would be interesting
    - I liked that this would allow me to outline the parameters in the analyses
    - This also translates to a flexible model if this were a real project: I can pick a direction to start, but we can make tweaks to the calculations, adjust weighting, or dive deeper into specific metrics to more accurately address the business question
- **Begin Market Analysis**
    - Clean data as I go (for example: currencies stored as strings in Markets table)
    - Build out index logic for determining our best markets
        - Looked at the metrics and thought about what might be the strongest indicators of a seller's market
        - Worked in SQL to capture calculations across markets and build index ranking
        - Took notes on the process as I go
        - Validate data with small samples (ex: compare un-aggregated San Jose, CA data with aggregated data coming from my query)
    - Write up answers to the prompt
        - I decided to take an "email" approach for this because it can be representative of how asks like this are handled across teams
        - I considered how I would communicate these findings to a real marketing team counterpart and favored casual and accessible language
    - Export data, tweak Google Sheets views, and upload components to GitHub
    - Revisit the approach and add additional context with the Sales table
        - I initially relied only on the Markets table for this prompt, but I realized that the Sales table could support or refute the 5 best markets I had picked
        - I came to find that my 5 best markets changed when I brought in this data
    - Update write-up, files, and GitHub attachments
        - Note a question:
- **Begin Agent Analysis**
    - Consider what kind of index calculations would best support "best agents," thinking about how I would approach this topic to a real sales manager and being considerate of their team
        - I wanted to avoid favoring agents with more ratings, so I ultimately landed on filtering our sample to agents with at least 5 ratings and then taking their raw rating, similar to how Google Maps handles review aggregation
        - I wanted to avoid favoring agents with the highest sales prices without more context around how agents work to represent more expensive homes, but I determined that comparing their average to

        - the median in their markets would help distribute more evenly than comparing them to all markets
        - I saw the callouts in the data table description that Smiley Real Estate wants to make money, so I weighed the home price median difference at 50% and the other two factors at 25%
    - Work with the data in a similar approach, but start by joining tables together to get the full agent picture
    - Validate data as I go and identify a few cleaning opportunities (two Michael Lewis agents, “orlando_fla” as a market_id, etc.
    - Export data, tweak Google Sheets views, upload components to GitHub
    - Write up answers to the prompt
- **Complete writeups for “More Time & Data Sources” and “The Process” sections**
- **Complete writeup and visuals for “One Area for Growth” section**
- **Update ReadMe file in GitHub with updated links and files**

Tools Used

- **MySQL & Workbench** - data storage and querying; index building for Market Analysis and export view for Agent Analysis
- **Google Docs** - write up document and keep in-progress notes
- **Google Sheets** - index building and pivot table view for Agent Analysis
- **Google** - gut-check indicators of strong real estate markets and percentile rankings
- **ChatGPT** - feedback on analysis writeups and troubleshooting one Sheets formula
- **GitHub** - upload SQL files, CSV files, images, and other components used in project
- **Tableau Public** - build visual for Executive Analysis

Challenges Faced & Solutions

- **Defining metrics and calculations**
    - I was excited by the index approach because it gives us a framework that would be editable and customizable in the real world.
    - I read through the assignment brief and data table descriptions a few times so that I could catch anything that was explicitly stated as important. Beyond that, I thought about the value that the marketing team and sales team provide to the customers and tried to build frameworks that served all three of these groups as best as possible.
- **Handling messy data**
    - When working with CTEs and subqueries, I like to make sure that the logic and aggregations are working correctly by taking small samples of the data and looking at it without aggregation.
    - I follow a similar approach for joining tables, which was especially relevant with the two different instances of a Michael Lewis in the Agents table.
- **Handling outliers**
    - For the index method I was using in the Agent Analysis, there were outliers like Tyrone Thompson that heavily skewed the data. Using a simple percentile

ranking gave him a score of 1.00 and almost all records a score of 0.01 to 0.04.
    - I researched how to adjust my method so that we could give those outliers a score of 1.00 but adjust the other rankings to a more representative value. I did some troubleshooting to create a formula that capped at the 95th percentile, and I used the pivot table view to evaluate whether the ranking system was functioning as I expected.
- **Managing ambiguity**
    - I included information on my thought process and data assumptions in the email writeups to be transparent. I think this project was intentionally open-ended, so I decided to embrace the learning opportunity and have confidence that my teammates are happy to collaborate and educate.
    - As a data analyst, a big part of my job is to be accurate. At the same time, I remind myself that strong relationships, sound logic, and process documentation can help overcome ambiguity and the unknown.

I appreciate how much thought was put into the assignment. I genuinely enjoyed working with the data and uncovering these insights.