# Smiley Real Estate

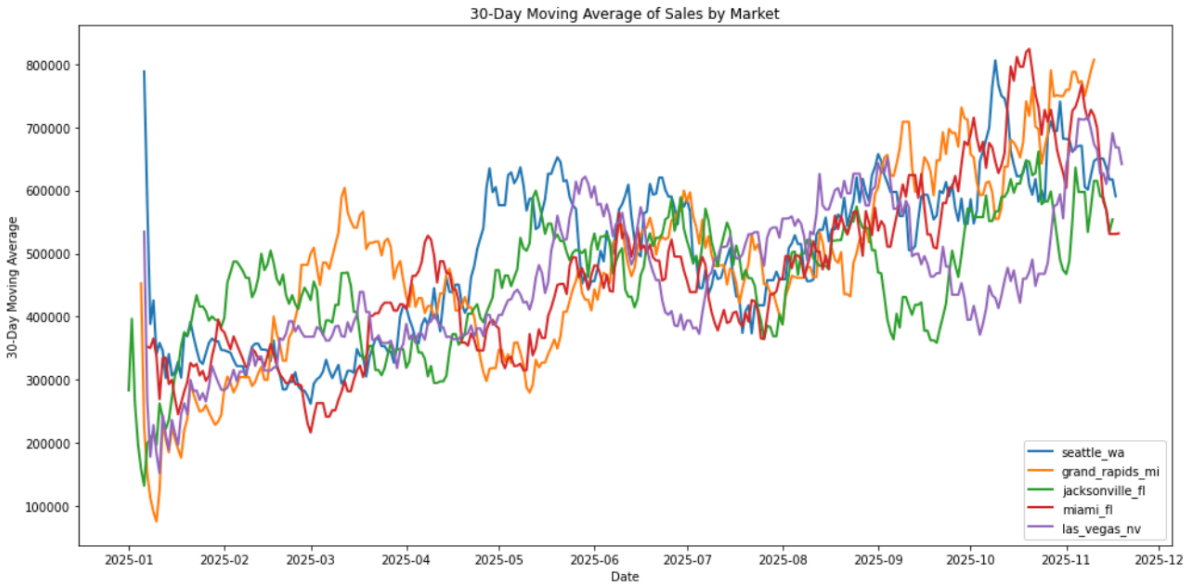
## Best Markets

By combining the sales data with the customer data, you can get a look at the top markets by total sales for the year of 2025. I am assuming that the sales data represents revenue generating activity for Smiley, and therefore using it as the driver for performance. I have also removed an outlier from Detroit, MI, that showed a $38mm selling price which skewed the data.

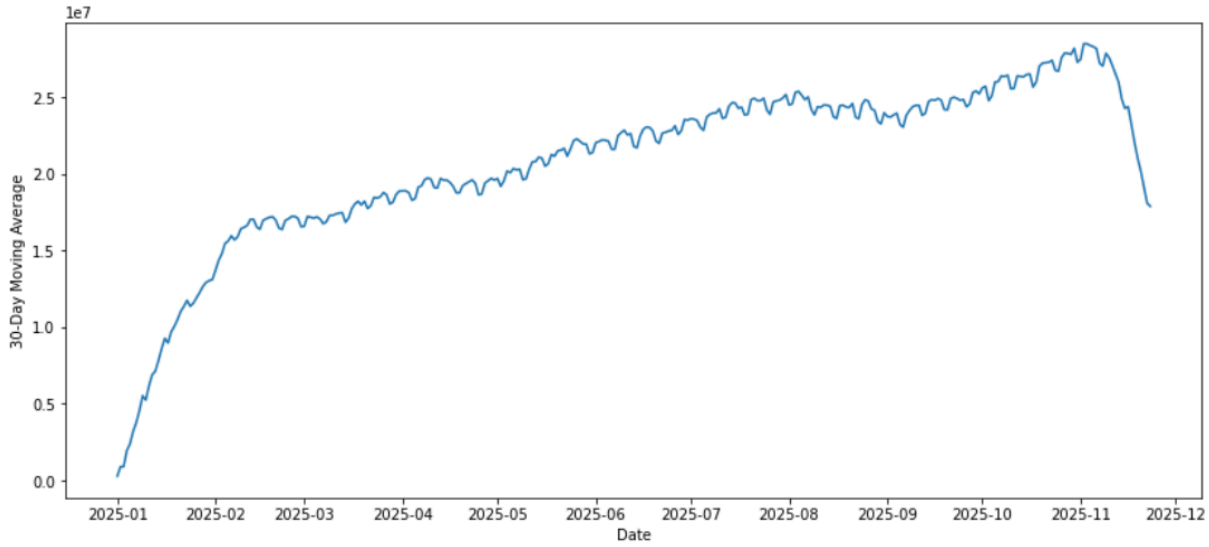
| **Market** | **Sales Volume, 2025** |
|---|---|
| Seattle, WA | $158,307,881 |
| Grand Rapids, MI | $156,174,194 |
| Jacksonville, FL | $150,300,813 |
| Miami, FL | $149,932,321 |
| Las Vegas, NV | $149,920,376 |

Overall, sales are trending upwards over the year for all of our top 5 markets. Overall, the aggregate of the 50 markets Smiley covered is also seeing growth for the year of 2025.

Below is a plot of the 30-day moving average of total sales volume for each of our top 5 markets:



And the total 30 day moving average sales volume for all markets in 2025:




# Top Performing Agents

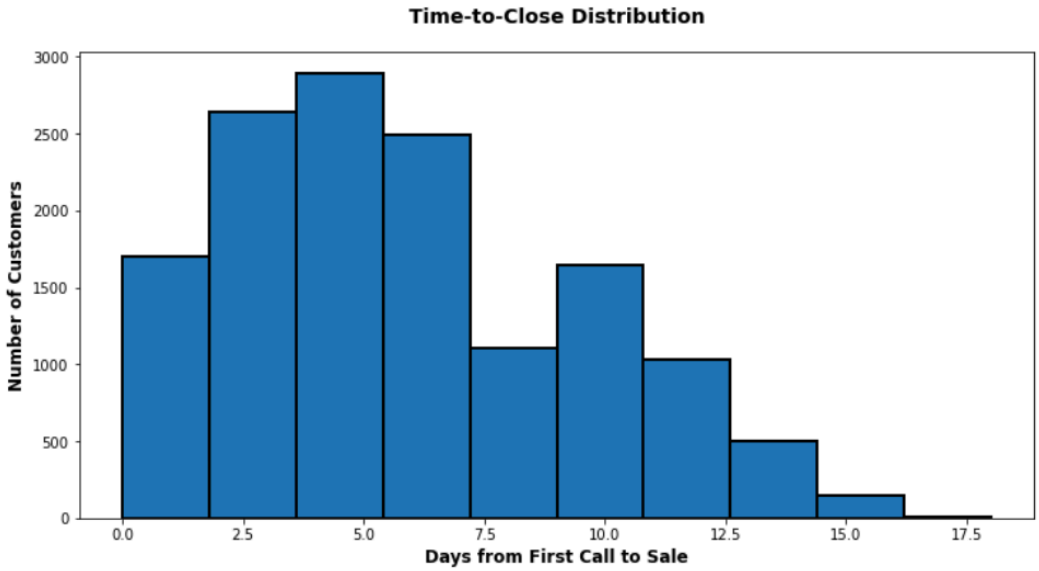
When looking at agent performance, I'm looking directly at the market share an agent manages to capture for a given year. I'm measuring "market share" as the total number of closes an agent has in a market for a given year, compared to the total number of deals that were closed in the market for that year.

I have attached an Excel file in the github which contains this presentation, please refer to that list for the top performing agent in each market.

# Growth Opportunities

For growth, we can look at potentially re-focusing our marketing efforts on being more proactive when contacting leads. Looking at the distribution of days between first contact via a call and closing with a customer, you can see the vast majority of our customers close within the first 10

days of contact.




This shows that adopting a proactive strategy where we make calls early and follow up often could help improve performance, and focusing more on leads who have made actual contact (picking up the phone) rather than those that continue to go to voicemail.

# What if I had more time and data sources?

For one, I'd have put a lot more effort into Growth Opportunities. I really want to slice it a bit more, include more data from opportunities that did not converge. Also, I'd love to load this into something I can write SQL against, because it would help me write the transformations a bit faster. When using Pandas, I have to rely on GPT a bit more than I'd like for some of the more complex transformations because I typically prefer SQL's syntax to transform data.

One thing for Agent performance that I believe is missing is extending the "total closes" to total volume closed, I don't have a clear way to link an Agent to an exact sale with this data, but adding that dimensionality to the data would help us form an even better understanding of agent performance.

# My Process

I try to be pretty explicit about what I'm thinking during my Jupyter Notebook EDA. I started before anything by just visually inspecting data in google sheets, then loading it into an ipynb file to explore more deeply. I'm a finance nerd at heart, so I really like to focus in on "revenue

adjacent” data when doing analysis that involve questions like “what is our best market”. I also do a lot of plotting because in my experience, plotting can help make data issues more prominent, and I was able to find an outlier sale which I excluded after plotting some of the sales data.