Zillow Assignment

—

Adrian Montero - Senior Data Scientist

Content

1. Problem Overview
2. Methodology Used (Python Script)
    a. Building the Dataset
3. Optimal Pricing Options
    a. Visualizations
4. Best Pricing Strategy
5. Other Pros and Cons
6. Final Remarks

Problem Overview

Zillow's New Construction Sales Team is analyzing two pricing strategies: a flat price per Community, or per Lead Received. The team agrees to:

1. Prioritize long term revenue over short term.
2. Provide the best customer experience (more listings is better)
3. We will be analyzing two pricing options:

- $40/Lead
- $400/Community/Month

Building the Dataset (2018-2020)

This dataset was built in three parts, we took into consideration:

1. The MoM growth rate % of Leads per Community
2. The MoM growth rate % of New Construction Communities (for each pricing strategy)
3. Projected Total Revenue

We wrote two main python functions:

1. **def leads_growth( )**
   a. Takes the avg. number of leads (4) per month and project its growth rate
2. **def community_growth( )**
   a. To calculate the total size of the community based on its growth rate and initial size (6174)

Sample Outputs of the Functions

def leads_growth( )

```
In [6]: df_leads = leads_growth(4, years, growth_rate1)
        df_leads
```

Out[6]:

	avg_leads	lead_growth_rate	lead_growth
2018-1	4	1.05	4.000000
2018-2	4	1.05	4.200000
2018-3	4	1.05	4.410000
2018-4	4	1.05	4.630500
2018-5	4	1.05	4.862025
2018-6	4	1.05	5.105126
2018-7	4	1.05	5.360383

def community_growth( )

	community_growth_rate1	community_growth1	community_growth_rate2	community_growth2
2018-1	1.06	6174.00	1.054	6174.00
2018-2	1.06	6544.44	1.054	6507.40
2018-3	1.06	6937.11	1.054	6858.80
2018-4	1.06	7353.33	1.054	7229.17
2018-5	1.06	7794.53	1.054	7619.55
2018-6	1.06	8262.20	1.054	8031.00
2018-7	1.06	8757.94	1.054	8464.68
2018-8	1.06	9283.41	1.054	8921.77
2018-9	1.06	9840.42	1.054	9403.54
2018-10	1.06	10430.84	1.054	9911.33

Final Dataset Output

```
leads_community_final = leads_community[['avg_leads','lead_growth', 'total_growth_method1', 'total_growth_method2',
                'revenue_method1','revenue_method2','rev1_cumulative','rev2_cumulative']]
leads_community_final
```

	avg_leads	lead_growth	total_growth_method1	total_growth_method2	revenue_method1	revenue_method2	rev1_cumulative	rev2_cumulative
2018-1	4	4.000000	6174.00	6174.00	$987,840.00	$2,469,600.00	$987,840.00	$2,469,600.00
2018-2	4	4.200000	6544.44	6507.40	$1,099,465.92	$2,602,960.00	$2,087,305.92	$5,072,560.00
2018-3	4	4.410000	6937.11	6858.80	$1,223,706.20	$2,743,520.00	$3,311,012.12	$7,816,080.00
2018-4	4	4.630500	7353.33	7229.17	$1,361,983.78	$2,891,668.00	$4,672,995.91	$10,707,748.00
2018-5	4	4.862025	7794.53	7619.55	$1,515,887.99	$3,047,820.00	$6,188,883.90	$13,755,568.00
2018-6	4	5.105126	8262.20	8031.00	$1,687,182.96	$3,212,400.00	$7,876,066.86	$16,967,968.00
2018-7	4	5.360383	8757.94	8464.68	$1,877,836.35	$3,385,872.00	$9,753,903.21	$20,353,840.00
2018-8	4	5.628402	9283.41	8921.77	$2,090,030.42	$3,568,708.00	$11,843,933.64	$23,922,548.00
2018-9	4	5.909822	9840.42	9403.54	$2,326,205.14	$3,761,416.00	$14,170,138.77	$27,683,964.00
2018-10	4	6.205313	10430.84	9911.33	$2,589,065.03	$3,964,532.00	$16,759,203.80	$31,648,496.00
2018-11	4	6.515579	11056.69	10446.55	$2,881,629.27	$4,178,620.00	$19,640,833.07	$35,827,116.00
2018-12	4	6.841357	11720.10	11010.66	$3,207,255.73	$4,404,264.00	$22,848,088.80	$40,231,380.00
2019-1	4	7.115012	12188.90	11407.04	$3,468,966.66	$4,562,816.00	$26,317,055.45	$44,794,196.00
2019-2	4	7.399612	12676.46	11817.70	$3,752,035.52	$4,727,080.00	$30,069,090.98	$49,521,276.00
2019-3	4	7.695597	13183.51	12243.13	$4,058,199.04	$4,897,252.00	$34,127,290.01	$54,418,528.00

Revenue:

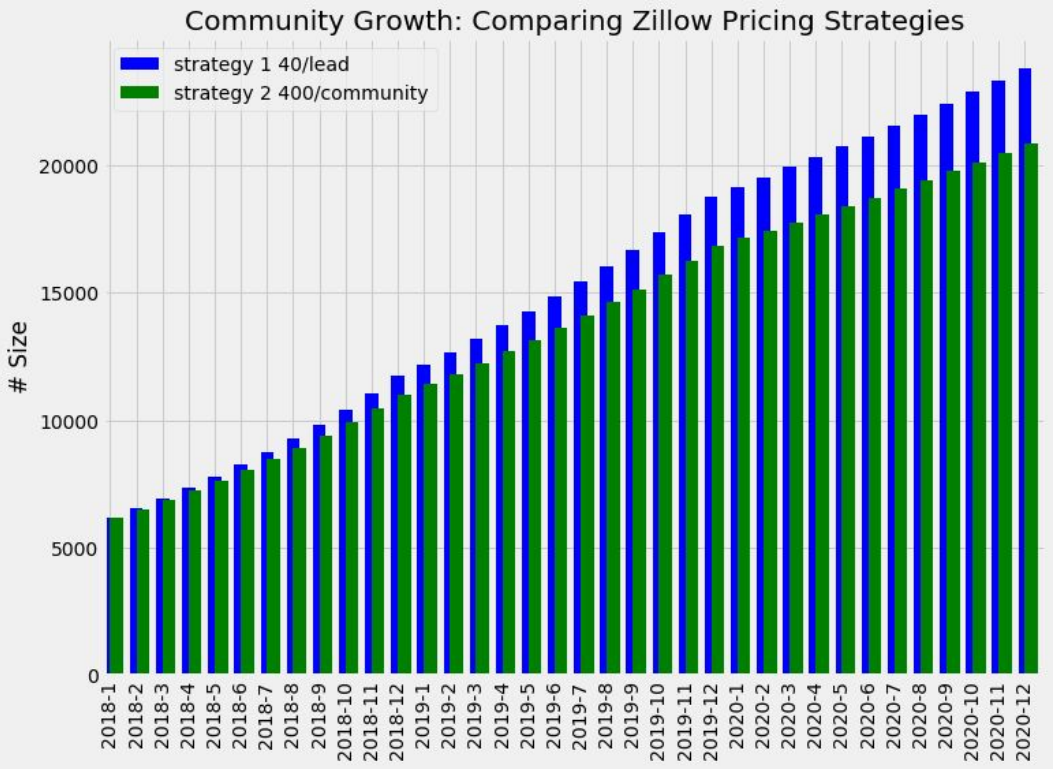
- **Strategy 1 Revenue per Month =** $40 X Total Leads per Month X Total # of Communities
- **Strategy 2 Revenue per Month =** $400 X Total # of Communities

For detailed information on how the dataset was built, please refer to the Jupyter Notebook attached in the email

Community Size, Pricing Strategy 1 vs 2

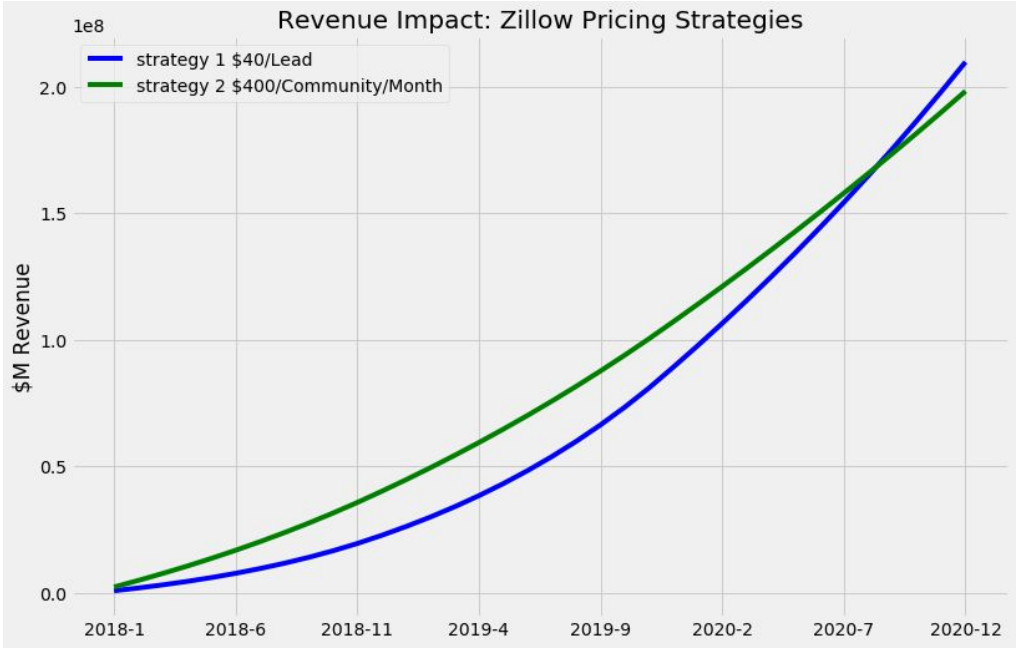
December 2020 Community Size (**S1: 23797 vs S2: 20849**)

- We can see that during the first four months the two pricing strategies grow at a similar rate.
- Each year, the size of the community grows at a higher rate for Strategy 1. **By December 2020,** the New Construction Communities Size is **1.14x** more than that of Strategy 2 (**23797, vs 20849**)
- **In the long term, Strategy 1** seems to be a better fit, due to the **increased number of listings projected.**



Revenue Impact - Strategy 1 vs 2

- Due to the pricing strategy **$400/Community/Month** and its growth rate, we can see that in the short term **Strategy 2 generates a higher revenue.**
- On **September 2020**, Strategy 1 ($40/Lead) intersects and surpasses Strategy 2 in terms of cumulative revenue.
  - $175M vs $173M
- By December 2020, **Strategy 1 is projected to generate ~ $11.5M more** than Strategy 2.



In the Long Term ($40/Lead) is a better pricing strategy

- **Assuming that market conditions** (economically and real estate) remain constant, **Strategy 1 ($40/Lead) seems to be a better option**, given that the New Construction sales team wants to prioritize long term revenue over short term and increase the number of listings.
- By December 2020 when compared to Pricing Strategy 2, Strategy 1 will lead to:
    - **~1.14x** more in the number of listings
    - **~$11.5M** more in Revenue

Other Pros and Cons

- Although Strategy 1 seems to have a higher revenue in the long term (if **market conditions** and projections remain constant - which **seems unrealistic**). **Today Zillow could be better off by choosing strategy 2** and electing **to generate a higher revenue sooner.**
    - The excess revenue over strategy 1 in the first 33 months can be used to **venture into new investments** and allow the company to diversify its portfolio
    - The **market can be very volatile at times:** the assumption for community growth rate for strategy 2 ('higher cancellation rates') could not be accurate. For example, Zillow could acquire other local companies creating other potential streams of growth.
    - Compound interest (**money invested today has more value**), whether it is a new feature that delights customers or creating a better experience for agents/builders/contractors. Zillow should be constantly innovating.

Final remarks

- This pricing model should be constantly monitored and revised to better adjust the growth and cancellation rates
- We should explore other datasets to improve our pricing models: US Census population rate, Income by Household, Age/Gender, Global Economy performance, weather data, seasonality demand/supply, urban/rural geo conditions, competitor's data (what are other charging for similar services, customer reviews
- If more time is allowed, we should do AB tests to compare different pricing strategies (if the pay per lead model is elected, we should study the best price to maximize both customer satisfaction and revenue)
    - AB tests could be done in smaller regions, then release nationally
- **Currently, the pricing model is a one size fits all**, we could explore:
    - Clustering algorithms to better assign different pricing structures (e.g. per region)
    - Machine Learning models to determine what's driving growth rates and cancellations (i.e. decision trees, regressions, knn), or to create more dynamic pricing strategies (e.g. based on seasonality and demand)